TECHNISCHE UNIVERSITÄT MÜNCHEN

Fachgebiet für Bioinformatik

Tissue-specific gene (and protein) expression and its effects on protein-protein interaction networks in cancer and other complex diseases.

EVANS SIOMA KATAKA

Vollständiger Abdruck der von der Fakultät TUM School of Life Sciences
der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr.rer.nat)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Jan Baumbach

Prüfer der Dissertation: 1. Prof. Dr. Dmitrij Frishman
2. Hon.-Prof. Jürgen Cox, Ph.D.

Die Dissertation wurde am 29.04.2020 bei der Technischen Universität München eingereicht und durch die Fakultüt TUM School of Life Sciences am 12.10.2020 angenommen.

OUTLOOK

Even though researchers have made great strides in elucidating disease-causing genes (e.g., driver genes in cancer), the determination of the complete set of such genes is still an ongoing challenge. In the case of complex diseases which are a consequence of multiple underlying factors, the search for biomarkers is even arduous. Furthermore, it is now accepted that genes may not act in isolation to promote disease development, but rather, multiple genes work in tandem. These multiple gene (or protein) interactions bring about the diverse molecular processes that manifest as pathological (or disease) phenotypes. For example, in cancer and diseases of the nervous system, network subnetworks have been found to influence disease progression. Additionally, isoform switching (IS) - the differential expression of alternatively spliced gene products in healthy and in disease, has been shown to promote tumorigenesis. IS has thus been termed a cancer hallmark. IS often results to (i) translation of protein variants - proteoforms, that may be differentially expressed in healthy and disease states, and (ii) the loss or gain of protein domains (structural and functional) mediating protein-protein interactions. Consequently, this can lead to the re-wiring of the interactome. Quantitating the differential expression of genes (or transcripts) and proteins in different cell types, tissues or organs will potentially enhance our knowledge and understanding of the biology of complex diseases. Indicators of the phenotypes between healthy and disease states are termed disease biomarkers and are used to monitor disease phenotypes as well as develop therapeutic targets. However, complex diseases such as cancer and Alzheimer's disease arise as a result of the combination of both inheritable and environmental factors, therefore, the determination of the complete sets of biomarkers has proven to be challenging.

The ability to obtain high throughput data using next-generation sequencing (NGS) was a big step towards understanding the cancer complexity and heterogeneity. Recently, the integration of NGS data with biological network information (e.g. protein-protein interaction networks, PPINs) has been suggested as a novel way of studying complex diseases to discover inherent biomarkers. As such, PPINs have played a vital role in our understanding of the behavior of complex diseases as well as the development of new therapies. While minimal studies have incorporated isoform expression data in studying complex diseases at the PPIN level, few have studied the effect of alternative splicing on patient-specific protein-protein interaction networks. The research discussed in this thesis starts by describing how differential isoform expression between cancer and healthy states may result in edgetic perturbations in cancer **(**Chapters 1 and 2**).** The study further sought to find if the proteins involved in the above-mentioned perturbations may be crucial in promoting cancer initiation and growth, classification of cancer types and subtypes, or act as potential targets for cancer therapy development **(**Chapter 2**).**

Equipped with the tools emerging from the genomics revolution, it is now possible to determine perturbations that link inherent molecular states to pathological or physiological states through the reverse engineering of protein (or gene) interaction networks. Computational tools are now able to utilise domain-domain interaction data to resolve condition-specific interaction networks from RNA-Seq data by accounting for the domain content of the primary transcripts expressed. In this work, we used The Cancer Genome Atlas (TCGA) RNA-Seq datasets to generate patient-specific pairs of protein-protein interaction networks (interactomes) corresponding to both the tumor and the healthy tissues across multiple cancer types. The comparison of these interactomes provided a list of patient-specific edgetic perturbations of the interactomes associated with the cancerous state. To the best of our knowledge, this is the first time it can be shown that using patient-specific PPIN derived from corresponding mRNA

expression profiles of healthy and cancer patient samples is a novel way of identifying patient-, cancer-type and subtype as well as pan-cancer edges susceptible to perturbation during tumour growth. We found that among the identified perturbations, select sets are robustly shared between patients at the multi-cancer, cancer-specific and cancer sub-type specific levels. Interestingly, the majority of the alterations do not directly involve significantly mutated genes, nevertheless, they strongly correlate with patient survival. Our findings are freely available at EdgeExplorer: http://webclu.bio.wzw.tum.de/EdgeExplorer - Chapter 2, and are a new source of potential biomarkers for classifying cancer types. We envisage that the information in EdgeExplorer will complement the available transcriptomics, proteomics as well as clinical cancer data, and help oncologists and other biomedical researchers to further understand the cancer microenvironment. Collectively, our analyses show that the diverse proteins driving edgetic perturbations in cancer are essential biomolecules in tumorigenesis and could be used in cancer disease monitoring and in developing new cancer therapies for clinical use. Our findings present an integrated omics and protein-protein interaction network approach for the computational identification of pan-cancer, cancer type and subtype specific biomarkers with potential clinical prognostic relevance. Additionally, the robustness and reproducibility of our approach show that our framework can be readily applied to other complex diseases.

In a nutshell, the first part of this thesis (Chapters 1-2) highlights a framework to identify network biomarkers in cancer by determining (i) how domain changes associated with alternative splicing rewire the PPIN and, (ii) how the proteins significantly involved in such network rewiring events may be novel cancer biomarkers. The second part of the thesis (Chapter 3) highlights how different brain cell types uniquely secrete proteins, and we identify how the cell type specific secreted proteins interact with each other. These secreted proteins are crucial molecules in the manifestation of diseases of the nervous system. This chapter was a close collaboration with Johanna Tüshaus, a PhD student of Professor Stephan Lichtenthaler's group at the German Center for Neurodegenerative Diseases (DZNE).

Keywords: Isoform switching, edgetic perturbations, biomarker discovery, cancer, network rewiring, cell type specific protein (mRNA) expression.

ZUSAMMENFAFSSUNG

Auch wenn die Forscher bei der Aufklärung krankheitsverursachender Gene (z.B. Treibergene bei Krebs) große Fortschritte gemacht haben, ist die Bestimmung der vollständigen Sets solcher Gene vor allem bei komplexen Krankheiten immer noch eine ständige Herausforderung. Des Weiteren ist inzwischen anerkannt, dass Gene nicht isoliert wirken dürfen, um die Krankheitsentstehung zu fördern, sondern dass vielmehr mehrere Gene zusammenwirken, um die verschiedenen molekularen Prozesse zu bewirken, die sich als pathologische (oder Krankheits-) Phänotypen manifestieren (z.B. bei Krebs und Erkrankungen des Nervensystems). Darüber hinaus kommt es häufig zum Isoform-Switching - die unterschiedliche Expression von alternativ gespleißten Genprodukten im Gesunden und in der Krankheit, einem kürzlich charakterisierten Kennzeichen von Krebs, führt oft zu: (i) Translation von Proteinvarianten - Proteoformen, die in gesunden und kranken Zuständen unterschiedlich exprimiert werden können, und (ii) Verlust oder Gewinn von (strukturellen und funktionellen) Proteindomänen, die Protein-Protein-Interaktionen vermitteln, und damit die Neuverdrahtung des Interaktoms. Die Quantifizierung der differentiellen Expression von Genen (oder Transkripten) und Proteinen in verschiedenen Zelltypen, Geweben oder Organen wird unser Wissen und Verständnis der Biologie komplexer Krankheiten potenziell erweitern. Indikatoren für die Phänotypen zwischen gesundem und krankem Zustand werden als Krankheitsbiomarker bezeichnet und dienen der Überwachung von Krankheitsphänotypen sowie der Entwicklung therapeutischer Ziele. Komplexe Krankheiten wie Krebs, Alzheimer und Herzerkrankungen entstehen jedoch durch die Kombination von sowohl vererbbaren als auch Umweltfaktoren, und daher hat sich die Bestimmung der vollständigen Sets von Biomarkern als schwierig erwiesen.

Der Fortschritt zur Gewinnung von Hochdurchsatzdaten mittels Next-Generation Sequencing war ein großer Schritt zum Verständnis der Komplexität und Heterogenität von Krebs. Neulich wurde die Integration von Omics-Daten mit Informationen aus biologischen Netzwerken (z.B. Protein-Protein-Interaktionsnetzwerken, PPINs) als eine neuartige Möglichkeit zur Untersuchung komplexer Krankheiten vorgeschlagen, um inhärente Biomarker zu entdecken. Daher spielen PPINs eine wichtige Rolle für unser Verständnis des Verhaltens komplexer Krankheiten sowie für die Entwicklung neuer Therapien. Während nur wenige Studien integrieren Isoformen-Expressionsdaten in die Untersuchung komplexer Krankheiten auf PPIN-Ebene, haben aber die wenigesten die Auswirkung des alternativen Spleißens auf patientenspezifische Protein-Protein-Interaktionsnetzwerke untersucht. Die aktuelle Arbeit beginnt mit der Beschreibung, wie die differentialle Expression der Isoformen zwischen Krebs und gesunden Zuständen zu edgetischen Störungen bei Krebs führen kann (Kapitel 1 und 2). Weiter, untersucht die Arbeit ob die an den oben genannten Störungen beteiligten Proteinen, für die Förderung der Krebsentstehung und des Krebswachstums, für die Klassifizierung von Krebsarten und -subtypen oder als potenzielle Ziele für die Entwicklung der Krebstherapie von entscheidender Bedeutung sind (Kapitel 2).

Ausgestattet mit den Werkzeugen der Genomik-Revolution ist es nun möglich, durch Reverse Engineering von Protein- (oder Gen-) Interaktionsnetzwerken Störungen zu bestimmen, die inhärente molekulare Zustände mit pathologischen oder physiologischen Zuständen verbinden. Computational Tools sind jetzt in der Lage, Domäne-Domäne-Interaktionsdaten zu nutzen, um zustandspezifische Interaktionsnetzwerke aus RNA-Seq-Daten aufzulösen, indem der Domäneninhalt der exprimierten primären Transkripte berücksichtigt wird. In der in dieser Dissertation beschriebenen Arbeit verwendeten wir die RNA-Seq-Datensätze des Krebsgenom-Atlas (TCGA), um patientenspezifische Paare von Protein-Protein-

Interaktionsnetzwerken (Interaktomen) zu generieren, die sowohl dem Tumor als auch dem gesunden Gewebe über mehrere Krebsarten hinweg entsprechen. Der Vergleich dieser Interaktome lieferte eine Liste patientenspezifischer Kantenstörungen der Interaktome, die mit dem Krebszustand assoziiert sind. Nach unserem besten Wissen kann damit zum ersten Mal gezeigt werden, dass die Verwendung patientenspezifischer PPINs, die aus entsprechenden mRNA-Expressionsprofilen gesunder und Krebspatientenproben abgeleitet werden, eine neuartige Methode zur Identifizierung von Patienten-, Krebs-Typ- und Sub-Typ- sowie Pan-Krebs-Rändern ist, die während des Tumorwachstums für Störungen anfällig sind. Wir stellten fest, dass unter den identifizierten Störungen ausgewählte Sets robust zwischen Patienten auf den Ebenen Multi-Krebs, krebsspezifisch und krebssubtypspezifisch aufgeteilt sind. Interessanterweise betrifft die Mehrzahl der Veränderungen nicht direkt signifikant mutierte Gene, dennoch korrelieren sie stark mit dem Überleben der Patienten. Unsere Ergebnisse sind frei verfügbar unter EdgeExplorer: http://webclu.bio.wzw.tum.de/EdgeExplorer - Kapitel 2, und stellen eine neue Quelle potenzieller Biomarker zur Klassifizierung von Krebsarten dar. Wir gehen davon aus, dass die Informationen in EdgeExplorer die verfügbaren Transkriptomik-, Proteomik- und klinischen Krebsdaten ergänzen und Onkologen und anderen biomedizinischen Forschern dabei helfen werden, die Krebsmikroumgebung besser zu verstehen. Insgesamt zeigen unsere Analysen, dass die verschiedenen Proteine, die die Kantenstörungen bei Krebs verursachen, wesentliche Biomoleküle bei der Tumorentstehung sind und bei der Überwachung von Krebserkrankungen und bei der Entwicklung neuer Krebstherapien für die klinische Anwendung eingesetzt werden könnten. Unsere Ergebnisse präsentieren einen integrierten Omics und Protein-Protein-Interaktionsnetzwerk-Ansatz für die rechnergestützte Identifizierung von Pan-Krebs, krebstyp- und subtypspezifischen Biomarkern mit potenzieller klinischer prognostischer Relevanz. Zusätzlich zeigt die Robustheit und Reproduzierbarkeit unseres Ansatzes, dass unser Rahmenwerk ohne weiteres auf andere komplexe Krankheiten anwendbar ist.

Insgesamt beleuchtet der erste Teil dieser Arbeit (Kapitel 1-2) einen Rahmen zur Identifizierung von Netzwerk-Biomarkern bei Krebs, indem bestimmt wird, (i) wie Domänenveränderungen, die mit alternativem Spleißen verbunden sind, den PPIN neu verdrahten und (ii) wie die Proteine, die signifikant an solchen Netzwerk-Umverdrahtungsereignissen beteiligt sind, neuartige Krebs-Biomarker sein können. Die identifizierten Kandidaten-Biomarker sollten bei der experimentellen Validierung der Biomarker, der gezielten Therapie und der Krebsüberwachung in Zukunft von großer Bedeutung sein.

Im zweiten Teil der Dissertation (Kapitel 3) heben wir kurz hervor, wie verschiedene Hirnzelltypen auf einzigartige Weise Proteine sezernieren, und identifizieren, wie die zelltypspezifischen sezernierten Proteine miteinander interagieren oder entscheidende Moleküle bei der Manifestation von Erkrankungen des Nervensystems sind. Dieses Kapitel war eine enge Zusammenarbeit mit Johanna Tüshaus, einer Doktorandin der Gruppe von Professor Stephan Lichtenthaler am Deutschen Zentrum für Neurodegenerative Erkrankungen (DZNE).

Schlüsselwörter: Isoform-Switching, edgetische Störungen, Entdeckung von Biomarkern, Krebs, Netzwerk-Neuverdrahtung, Expression von zelltypspezifischem Protein (mRNA).

ACKNOWLEDGEMENTS

Undertaking PhD studies in a foreign country was a truly life-changing experience for me, and this journey would have been impossible without the help, support and guidance that I received from close friends, colleagues and family. At this point of closure, I am humbled to express my sincere gratitude to each and every one of them.

First, I am indebted to my supervisor, Prof. Dr. Dmitrij Frishman: he provided me with an excellent working environment, and was both the biggest critique and supporter of my research. I thank him for his unwavering support and patience to see me through my doctoral studies. His in-depth knowledge in bioinformatics, tolerance, and guidance helped me tremendously in the course of my research. His patience during the writing of my manuscript and this thesis was crucial. Also, I am grateful that Prof. Dr. Jan Baumbach and Prof. Dr. Jürgen Cox agreed to be part of my thesis supervision committee. I gratefully acknowledge the funding i received towards my PhD studies from the Deutscher Akademischer Austauschdienst – DAAD in cooperation with the National Research Fund of Kenya (NRF). The DAAD and NRF gave me the financial assurance and support throughout my studies. By being a DAAD scholar, i have not only been able to study at the best university in Germany, but i have also benefited from the acquisition of soft skills that i attained by attending numerous workshops. In brief, the DAAD has made me a better individual in the society.

I would also like to say many thanks to Dr. Jan Zaucha. I greatly benefited from his invaluable input and significant assistance towards the end of my PhD research and during the preparation of my manuscripts. Likewise, i am thankful to Prof. Dr Stefan Lichtenthaler and his group, in particular, Johanna Tüshaus, whom i collaborated with in two other projects. Johanna patiently taught me important elements in mass spectrometry and critical aspects in neuroscience. From these collaborations, we were able to write two manuscripts; one of which is under peer review in Nature neuroscience and the other one is still being internally reviewed.

I must also thank all my colleagues in the Dmitrij Lab for the academic discussions we had, and for creating a friendly research environment during my PhD studies. Whenever I required help or wanted my research criticized, they always gave me an ear. I will forever cherish the four years (or more) i spent with these excellent colleagues. In particular, I thank Michael Kiening, Peter Hönigschmidt, Hongen Xu, Usman Khan, Xeynab Usman, Marina Parr, Martina Weigl and Stephan Breimann. I should also mention Leonie and Martina for their invaluable support in all the administrative issues at the department, especially when it came to renewing my residence permits at the "Landratsamt" in Freising. I am greatly indebted to Leonie for taking care of me and my wife during our early days in Freising.

Last but not least, I would like to thank my family: my wife Winfred Aluoch Kataka, and my son, Karl Kataka. I always had a welcoming and warm house each day after my tedious days at the laboratory. These two individuals always gave me the will to get through on both exciting and stressful days. I sincerely thank my parents and siblings for their undying love and support throughout my career up to this point. Without each one of them, I would not be the person I am, or the individual i strive to become.

DECLARATION

This thesis is my own work. However, Chapter 3 of this thesis was the product of close collaboration with Johanna Tüshaus from Professor Stephan Lichtenthaler's group at the German Center for Neurodegenerative Diseases, DZNE. Johanna Tüshaus designed and performed the experimental set up to generate the LC-MS/MS secretomics and proteomics data while I designed and performed the bioinformatics analyses.

In chapter 3, I detail the bioinformatics analysis workflow but not the experimental work performed by Johanna Tüshaus. For example, my work involved data preprocessing followed by dataset analysis. The details of the new experimental method hiSPECS (improved secretome-protein-enrichment-with-click-sugars method) developed by Johanna Tüshaus, are described in our joint publication that has been submitted to the Nature Neuroscience.

TABLE OF CONTENTS

List of figures

Figure 8: What do cells require to become oncogenic? Six vital biological processes must be evaded by cells and form abnormal proliferative cancer cells. Adapted from Hanahan & Weinberg, 2000.

Figure 9: A more realistic disease model is one in which considers multiple sources of perturbations at the network level, e.g., a combination of genetic and environmental perturbations affecting the molecular states of networks. a - Classic genetic association approaches seek to identify variations in DNA that correlate with disease state or with quantitative traits associated with disease. The attraction of this approach is the identification of the genetic causes of disease. b- Changes in DNA on their own do not lead to disease but, instead, lead to changes in molecular traits that go on to affect disease risk. By layering in molecular phenotypes as intermediate phenotypes, causal relationships between genes and disease can be established directly. c -Disease gene networks sense constellations of genetic and environmental perturbations. Adapted with modifications from Schadt E, 2009.

Figure 10: The different experimental and computational methods used in characterizing, detecting, and predicting protein-protein interactions. Adapted from Gonzalez et. al, 2012.

Figure 11: The advantage of using the PPIN to infer cancer biomarkers is the ability to couple the analysis with multiple other OMIC datasets of patient clinical characteristics. Consequently, it is now likely that more personalized, accurate and rapid disease gene diagnostic techniques will now be devised. Adapted from Ozturk et. al, 2018.

Figure 12. Healthy and cancer PPINs significantly differ in size in 11 out of 13 cancer types (p-value <0.05). The density plots indicate the distribution of paired cancer and healthy PPIN sizes for individual cancer types (A-M) and across cancer types (N). The vertical dashed lines indicate the mean sizes of cancer PPINs (red) as compared to corresponding healthy PPINs (green). For BRCA, LUSC, PRAD, KIRP, KIRC, KICH, COAD, LIHC, HNSC and STES healthy PPINs were larger than the corresponding cancer PPINs but the difference was not significant in KIRP. For THCA and BLCA (green label), cancer PPINs were larger than the corresponding healthy PPINs, but the difference was not significant in BLCA.

Figure 13. Bar plots indicating the number of edgetic perturbations obtained as a result of gene expression changes or domain changes that come about after isoform switches between cancer and healthy states. Sky blue: edgetic gains as a result of more genes being expressed in the cancer state, dark brown (left of zero intercept): edgetic gains as a result of isoform/domain changes (left of zero intercept). Light brown: edgetic losses as a result of the depletion of genes in the cancer state (right of zero intercept), light green: edgetic losses as a result of isoform/domain changes (right of zero intercept).

Figure 14. An example showing the consequences of domain changes between the cancer state and healthy state in patients diagnosed with BRCA. The protein structures of both (P0DP23) *CALM1* and (P62140) *PP1CB* were obtained from PDB while those of *DST* were modelled using the ensemble transcript sequences in SWISS-MODEL and visualized in PyMol. Following an isoform switch from ENST00000370765 (in healthy) to ENST00000244364 (in cancer), the protein Q03001 (*DST*) gained the domain PF13499. The consequence is the gain of interactions with the genes *PPP1CB* and *CALM1*.

Figure 15. A two-dimensional scaling projection of the enriched Biological processes (A), Cellular Components (B) and Molecular functions (C) for proteins involved in cancer-specific

edgetic gains after REVIGO pruning (dispensability value < 0.005). Dispensability of a term represents both the degrees of redundancy and enrichment. The lower the dispensability of a term, the least redundant and more significant a term is. The axes show the distribution of the GO terms based on their semantic similarities. The bubble color reflects the degree of significance (p-value) with blue color indicating a higher significance than the red color. The richly colored bubbles in the foreground represent GO terms with a dispensability value of < 0.005. The bubble sizes indicate how often a GO term occurs, the bigger the size the more frequent the term is.

Figure 16. Two-dimensional scaling projections of the enriched Biological processes (A), Cellular Components (B) and Molecular functions (C) for proteins involved in cancer-specific edgetic losses after REVIGO pruning (dispensability value < 0.05). Dispensability of a term represents reduced redundancy and a high degree of enrichment. The lower the dispensability of a term, the least redundant and more significant a term is. The axes show the distribution of the GO terms based on their semantic similarities. The bubble color reflects the degree of significance (p-value) with blue color indicating a higher significance than the red color. The richly colored bubbles in the foreground represent GO terms with a dispensability value of < 0.005. The bubble sizes indicate how often a GO term occurs, the bigger the size the more frequent the term is.

Figure 17. KEGG pathways differentially enriched between the proteins engaged in edgetic gains and those involved in edgetic losses. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate how often a KEGG pathway term occurs. The bigger the size, the more frequent the term occurs.

Figure 18. Cancer types share multiple perturbation patterns: Dendrograms based on edgetic gains (A), edgetic losses (B) and both edgetic gains and losses (C) across cancer types. Gained edges revealed 2 main clusters (A) with sub-clusters consisting of (i) BRCA, BLCA and STES, (ii) LUAD and LUSC, (iii) COAD and KICH, (iii) LIHC and PRAD, and (iv) KIRC and KIRP. Lost edges identified 2 main clusters (B) with additional sub-clusters consisting of (i) KICH, KIRP, and KIRP, (ii) LUAD and LUSC, (iii) COAD, HNSC and BRCA, (iv) STES, BLCA and THCA. Clustering of both edgetic gain and loss patterns revealed 3 main clusters (C) consisting of (i) LIHC, KICH, KIRC, KIRP, (ii) PRAD, STES, BLCA, THCA and (iii) LUAD, LUSC, COAD, BRCA and HNSC. The Approximately unbiased AU (green) and Bootstrap probability BP (red) scores indicate the likelihood of observing the obtained clusters. The clusters within the red rectangles with AU scores of >99% were observed after multiscale bootstrap (n= 10000). The edge # below the AU and BP values gives the edge count within the tree. The height indicates the similarity or dissimilarity between any two observations: the lower the height of the fusion between two observations, the more similar they are.

Figure 19. A screenshot of the EdgeExplorer portal. The EdgeExplorer portal provides a resource to the scientific community to easily query proteins of interest to find out if they are involved in edgetic perturbations in 13 different cancer types.

Figure 20. Edgetic perturbations in cancer. Assuming a global PPIN with 9 edges interconnecting 9 nodes and using cancer and healthy patient-specific mRNA expression profiles, for each patient (P1, P2 and P3) perturbed edges in cancer can be identified by comparing the healthy and the corresponding cancer PPIN. Significantly Mutated Genes

(SMGs) may be involved in perturbation of edges directly interacting with them, or those interacting with their perturbed neighbors (secondary neighbors).

Figure 21: Signalling to and from the early secretory pathway. (A, B) ER, ERESs and Golgi complex with the different signalling cascades that are either directed towards these organelles (A, yellow), or emanating from them (B, green). Autochthonous Golgi signalling pathways are shown in blue. Stimuli that trigger signalling to the secretory pathway (A) or the cellular responses elicited by signalling from the secretory pathway (B) are shown in yellow or green, respectively. The long black arrows indicate direction of transport along the secretory pathway. Adapted from Farhan et. al., 2011.

Figure 22: MS-based proteomics approaches. The top part of the image shows the bottom-up MS approach, while the bottom part of the image shows the top-down approach. Adapted from Chait et. al., 2006.

Figure 23: Diagram showing the distribution of the data prior and after normalisation. Data normalisation was achieved via variance stabilisation. For the missing data, a two-step imputation approach (knn and left-shifted Gaussian distribution) was undertaken. Here we need to update with your stepwise procedure from Perseus.

Figure 24: Protein coverage ranged from approximately 450 to 750 proteins per sample. A total of 995 proteins were detected in at least 5 of the 6 replicates across the brain cell types. Microglia cell-type had the highest protein coverage while Astrocytes had the lowest coverage.

Figure 25A: PCA analysis. The secretomes of the cell types segregated based on component 1 and component 2, which accounted for 44.9% and 19% of the variability, respectively.

Figure 25B: UMAP (Uniform Manifold Approximation and Projection) plot showing brain cell type clusters based on log transformed raw LFQ intensities of quantified proteins. This indicates that the secretomes of the four cell types differ from each other.

Figure 26: Heatmap of the top 50 differentially expressed proteins (Bonferroni p.adj < 0.05) across the 4 cell types from hierarchical clustering. The rows represent the differentially expressed proteins and the columns represent the cell types (and their replicates). The colours in the Heatmap represent log-scaled (z-scores) expression levels with blue indicating the lowest expression, white indicating intermediate expression, and red indicating the highest expression. The rows represent the differentially secreted proteins and the columns represent the cell types with their replicates. The colors represent log-scaled protein levels with blue indicating the lowest, white indicating intermediate, and red indicating the highest protein levels. Proteins significantly differentiated in one cell type with respect to the other 3 cell types were analysed with regard to their biological function via GO and KEGG pathway enrichment analysis.

Figure 27: Correlation matrix showing the relationship between the different brain cell types. All replicates of a cell types showed higher correlations (>0.7) as compared to replicates from other cell types. The matrix shows the Pearson correlation coefficient (red indicates a higher,

blue a lower correlation) and the correlation plots of the log2 LFQ intensities of the secretome of astrocytes, neurons, microglia and oligodendrocytes processed with the iSPECS method.

Figure 28A: Comparison of the biological processes enriched across brain cell types. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

Figure 28B: Significantly downregulated processes were observed only in Neuron and Microglia cell types. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

Figure 29A: Comparison of the molecular functions enriched across brain cell types. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

Figure 29B: Significantly downregulated molecular functions in brain cell type. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

Figure 30A: KEGG pathways significantly enriched across brain cell types. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

Figure 30B: Significantly downregulated KEGG pathways. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

Figure 31: Interacting proteins between secreted CSF proteins and the cell lysate proteins detected by Sharma et.al. We found a total of 711 unique interacting pairs, with all the proteins secreted with a cell type having interacting partners with proteins in the lysate of the other cell types. CSF proteins from the Neuron and the proteins from the neuron cell lysate had the highest number of interacting pairs (115 interactions), while those between CSF Astrocytes and proteins from the Oligodendrocytes cell lysate were the least (2 interactions).

Figure 32. List of proteins detected in murine CSF and the iSPECS glyco-secretome resource which have human homologs that are linked to brain disease based on the DisGeNET database. Relative protein expression in the brain cell secretome is indicated with black showing the highest and white the lowest abundance. Colored gene names indicate cell type-specific secretion.

List of Tables

CHAPTER 1: GENERAL INTRODUCTION AND LITERATURE REVIEW

1.1 Complex diseases

Current understanding of many human diseases and how best they can be treated is hampered by the complexity of the underlying human molecular system in which they are manifested[2,3]. On the contrary, the genes and mutations responsible for simple Mendelian disorders are easily identifiable. Diseases with complex underlying molecular systems are referred to as complex diseases and are as a result of a plethora of changes in the DNA of the diseased as well as a broad range of environmental factors and an individuals' lifestyle[2,4]. For example, it is known that cigarette smoking is a major risk factor for nasopharyngeal cancer (NPC) especially for young smokers: compared with never smokers, current smokers and ever smokers had a 59% and a 56% greater risk of NPC, respectively[5]. Complex diseases include cancer, Neurodegenerative diseases, Diabetes and Schizophrenia among others. Due to the multiple factors at play in the pathobiology of complex diseases, a network view of these factors (e.g., proteins within the interactome) has been fronted as one of the most suitable approaches to decipher how we may have a better understanding of complex diseases and relevant bioinformatics tools are consistently being developed to reveal multiple proteins and the pathways they affect[6,7]. Another aspect of complex diseases is their heterogenous nature which then prompts the identification of biomarker proteins or genes at the patient level[8]. In this thesis, we shall delve deeper into the discovery of cancer biomarkers at the protein interaction network level and briefly highlight a new method of quantitating the secretome while using the mouse brain cell types.

1.2 Cancer

Cancer, a leading global health burden, is a complex molecular disease that involves abnormal proliferation of cells with the potential to metastasise to other healthy tissues and organs[9]**.** Cancer metastasis may result in deregulation of multiple cellular functions and pathways leading to the death of cancer patients[9]. This year, it is estimated that roughly 1,800,0000 new cases of cancer will be diagnosed in the USA, and about 600,000 deaths are projected to happen[10]. In Germany alone, about 440,000 new cases of cancer are expected each year[11]. For the effective diagnosis and treatment of cancer, a better understanding of the disease is necessary. While the scientific community has advanced our understanding of cancer, efficient cancer treatment regimens are still required to counter the number of deaths associated with cancer. The consistent and rapid growth of the field of genomics has propelled our understanding of cancer, resulting to a common consensus that dynamic changes in the genetic material (DNA) within our cells result to cancer, i.e., cancer is a disease of the genome[12–14]. Cancer can be traced from the clonal expansion of a single abnormal tumor cell that have the ability to produce metastatic colonies[15,16], Figure 1.

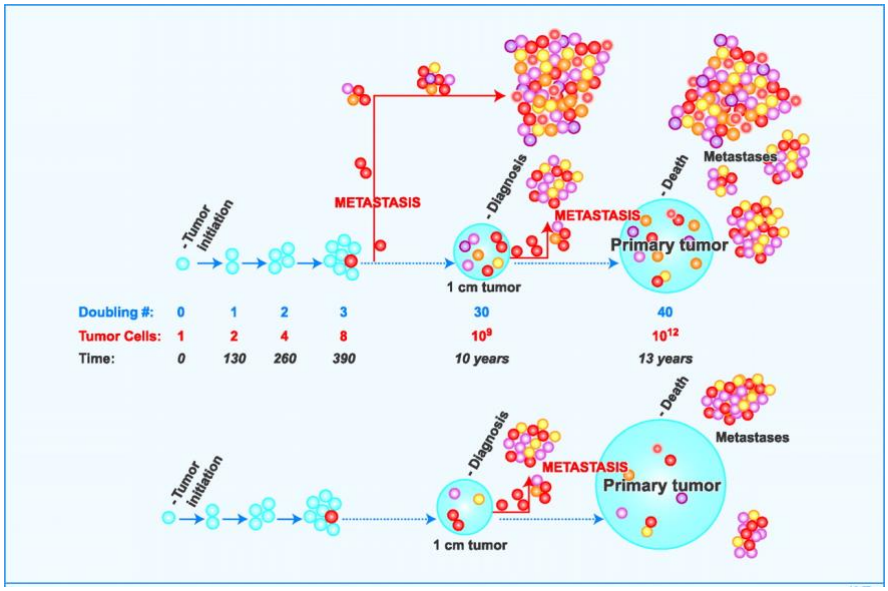


Figure 1: A schematic depiction of the clonal selection in cancer. During tumor initiation, progression, and metastasis, an estimated 10 years is required to achieve the 1-cm tumor needed for clinical diagnosis. During this period, genetic instability results in metastatic variants such that metastases occurring closer to the time of diagnosis are smaller and less heterogeneous as compared with those that occurred much earlier. This is contrasted with tumors having less genetic instability; these develop metastases late during tumor progression, are small in size at diagnosis, and have less heterogeneity. Regardless of the timing of metastases, if a 1-cm tumor is left untreated, then a lethal tumor volume, occurs within 10 doubling times, which calculates to 3 years. Adapted with modifications from Talmadge, 2007.

In 1914, the observation of chromosomal aberrations in cancer cells was among the first links between mutation and cancer[9]. The causal involvement of somatic mutations in cancer was later on supported by the discovery that multiple carcinogenic substances can also be mutagenic[17]. Substantive evidence came from studies that showed the introduction of DNA fragments from cancer cells into non-cancer healthy cells led to malignancy, and also from the identification of the responsible mutations linked to the transformation of the DNA[9]. This research then led to the discovery of the first oncogenes, whose mutations resulted to a gain of function that promotes transformation into cancer. At the same time, studies on hereditary cancers led to the discovery of tumor suppressor genes[18], which are normally inactivated by either germline or somatic mutations. Mutations are brought about by replication errors or by DNA damage that is either incorrectly repaired or left unrepaired. DNA damage may result from exogenous factors (chemicals, ultraviolet (UV) light, and ionizing radiation), or from endogenous factors (e.g., reactive oxygen species, aldehydes, or mitotic errors), or from enzymes involved in DNA repair mechanisms or genome editing, among others[19]. More saw, viruses and endogenous retrotransposons may bring about insertions of new DNA sequences in the genome[20]. Such changes in the genetic material of an individual cancer genome can now be accurately determined via multiple massively parallel sequencing technology platforms.

1.3 Next generation sequencing (NGS) technologies and genomics.

NGS is the deep, high-throughput, in-parallel DNA sequencing technology developed after the Sanger DNA sequencing method. NGS technologies resulted to a paradigm shift in the field of genomics, facilitating fast and cost-effective acquisition of genome-scale sequence data with

exquisite resolution and previously unmatched accuracies, together with bioinformatics tools for their analysis[21–24]**.** DNA sequencing technology was first developed by Frederick Sanger and Walter Gilbert and were based on either the Sanger sequencing approach (chain-termination method)[25], or the Maxam and Gilbert chemical degradation method[26]. Afterwards, first automated DNA sequencers were developed by Applied Biosystems Instruments who coupled the Sanger method together with fluorescent dye-terminator reagents[27]. Afterwards, these sequencers were enhanced by the introduction of computers to gather, store and analyze the generated sequenced data. With time, NGS technologies surpassed the conventional Sanger sequencing technique due to their ability to perform massively parallel sequencing (up to hundreds of millions of sequence reads) of short DNA fragments. For this, NGS technologies have become considerably cheaper, require significantly less DNA and are more accurate and reliable compared with Sanger sequencing. Additionally, NGS technologies have considerably increased the throughput data to several orders of magnitude as compared to Sanger sequencing[21]. NGS technologies include RNA-Seq which is used to measure transcript/Isoform expression level, and Chip-Seq which is used to study protein-DNA interactions. While there exists multiple NGS platforms and manufacturers[21,23] (e.g. Illumina, Oxford Nanopore, Pacific biosciences and Thermofischer), they all have similar sequencing steps where:

I. DNA samples to be sequenced are randomly fragmented.
II. Platform-specific adaptors are added to the flanking regions of the DNA to produce a library of arrays.
III. The library is then amplified via PCR (e.g., emulsion PCR or bridge PCR) prior to their detection.
IV. The amplified fragments are then sequenced by synthesizing the complimentary strand (e.g., via sequencing by synthesis or sequencing by ligation).
V. Base incorporation events are then detected (e.g., via image capture of fluorescent dye or light emission signal).

The resultant sequenced data is in the form of millions of reads (roughly 75-400 base pairs) and a throughput of between 1- 600 GB from a single run based on the platform used.

Table 1: A summary of NGS platforms as depicted by Levy & Myer[23] .

Manufacturer	Amplification	Detection	Chemistry	Url
Commerial				
Illumina	Clonal	Optical	Sequencing by synthesis	http://www.illumina.com
Oxford Nanopore	Single molecule	Nanopore	Nanopre	http://www.nanoporetech.com
Pacific biosciences	Single molecule	Optical	Sequencing by synthesis	http://www.pacb.com
Thermofischer Ion Torrent	Clonal	Solid state	Sequencing by synthesis	http://www.thermofisher.com/us/en/home/brands/ion-torrent.html
Precommercial				
Quantum Biosystems	Single molecule	Nanogate	Nanogate	http://www.quantumbiosystems.com
Base4	Single molecule	Optical	Pyrophosphorolysis	http://base4.co.uk
GenepSys (GENIUS)	Clonal	Solid state	Sequencing by synthesis	http://www.genapsys.com
QIAGEN (GeneReader)	Clonal	Optical	Sequencing by synthesis	http://www.qiagen.com
Roche Genia	Single molecule	Solid state	Nanopore	http://geniachip.com
Postcommercial				
Helicos BioScience (Heliscope)	Single molecule	Optical	Sequencing by synthesis	-
Roche 454 (GS FLX)	Clonal	Optical	Sequencing by synthesis	http://www.454.com
Dover (Polonator)	Clonal	Optical	Sequencing by ligation	-
ThermoFisher Applied Biosystems (SOLiD)	Clonal	Optical	Sequencing by ligation	http://www.thermofisher.com/us/en/home/brands/applied-biosystems. html
Complete Genomics	Clonal	Optical	Sequencing by ligation	http://www.completegenomics.com

Table 1: (-) indicates that no URL is available. Precommercial platforms have not been formally launched; post-commercial Platforms are no longer commercially available.

1.4 The Advent of cancer genomics

After the realization that cancer is a disease of the genome and the introduction of next generation sequencing technologies (NGS) coupled with the development of sophisticated bioinformatics and data analysis software, the era of cancer genomics was born[28]. The invention of high-throughput genomic technologies e.g., microarrays and NGS brought about the unprecedented insights into the complexity of cancer genomics. For example, Veer et. al., used microarrays derived from 117 patients to classify breast carcinomas based on gene expression variation patterns. They performed hierarchical clustering on expression data from both cancer and normal breast tissues and were able to determine that breast cancer patients had varying outcomes and treatment responses based[29]. Another study by Ahr et. al., using microarrays was able to discover breast cancer patients with high disease recurrence rates, and their results correlated with the conventional tumor staging[30]. With the introduction of NGS technologies, it was possible to obtain the DNA sequences of individual cancer genomes. This realization has not only revolutionized the scientific approach to the studies of omics data but has also heralded a paradigm shift in genomic and personalized medicine research in cancer. For instance, the ability of NGS to provide an unbiased view of the whole genome is vital in studying the cancer genome which is often consists of de novo genetic aberrations[31–33]. With NGS, the possibility to discover copy number variations, mutations (single nucleotide polymorphisms -SNPs), gene expression signatures and epigenetic changes in cancer was realized. These discoveries have not only led to the identification of novel diagnostic and prognostic cancer biomarkers but also the development of the early cancer drugs[34,35]. A continuing process is the search for such biomarkers at the patient level, cancer subtype level or even across multiple cancer types which is necessary for drug repurposing purposes[36–39].

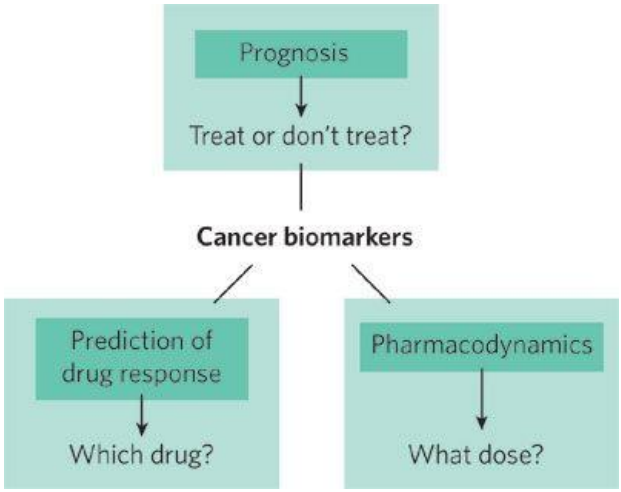


Figure 2: A schematic depiction of the importance of cancer biomarkers. Cancer biomarkers can be used for prognosis: to predict the natural course of a tumour, indicating whether the outcome for the patient is likely to be good or poor (prognosis). Biomarkers can help doctors to decide which patients are likely to respond to a given drug (prediction) and at what dosage the drug might be most effective (pharmacodynamics). Adapted with modifications from Sawyers, 2008.

1.5 RNA-Seq, tissue- and patient-specific expression of isoforms, genes and proteins in cancer.

With the advent of massively parallel sequencing platforms for NGS, the design and implementation of genetic studies protocols dramatically altered. RNA-seq is the most utilized NGS application, especially due to its coverage in determining the RNA expression content (transcriptome) of biological samples (e.g., tissues). Transcription involves making a copy of a gene to produce a precursor mRNA and processing the precursor mRNA to remove the intronic regions while fusing the exonic regions to produce a mature messenger RNA or the transcript. The transcriptome is the entire collection of transcripts available in a cell at a given time point. Transcripts serve as the link between an individuals' genotype and the observed phonotype. RNA-Seq includes the determination of RNA expression levels and alternative splicing events with highly reproduceable accuracies. Being a powerful tool for revealing the complexity of all transcriptional activities (within coding and noncoding regions), RNA-Seq is extensively used in biomedical research, clinical medicine, and in drug discovery experiments.

Precursor mRNAs may be processed in a plethora of ways to produce various transcripts via alternative splicing. Additionally, recent research shows that alternative splicing brings about expression of various transcripts as well as protein isoforms from a single gene – Figure 1 as depicted from El Marabti & Younis[40]. Furthermore, the alternative splicing events reflect a tissue-specificity or more importantly may occur in a disease related manner as is now known in cancer, and differential isoform expression or usage (isoform switching) is now considered a hallmark of cancer[41–43]. A gene can thus give rise to more than one transcript. For instance, a large number (up to 94 %) of human genes have multiple transcripts or undergo alternative splicing events[44,45].

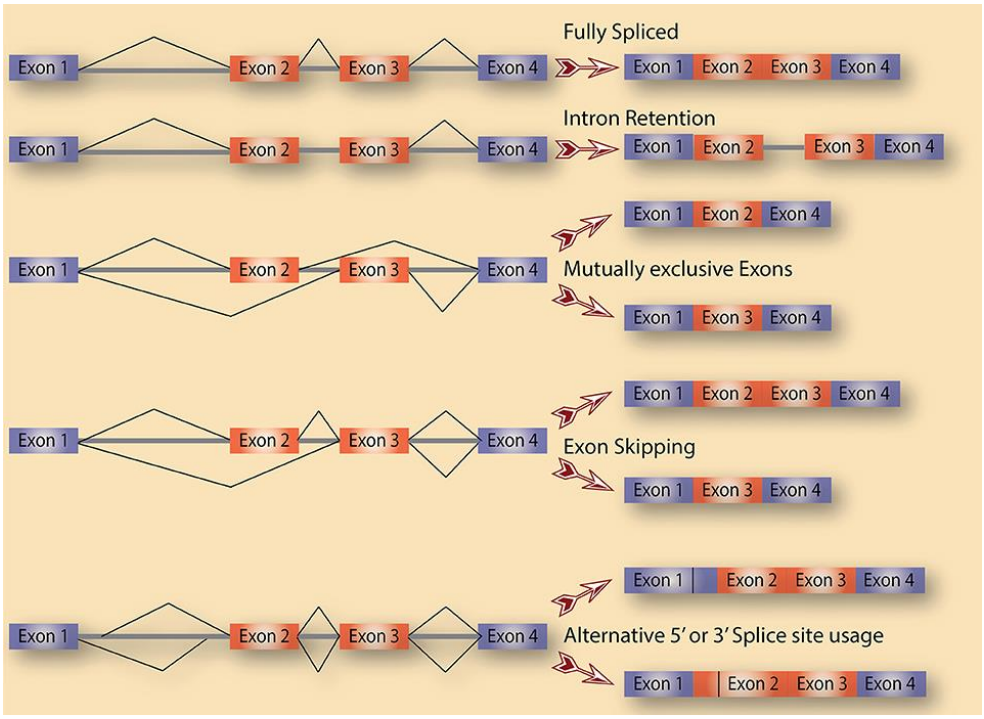


Figure 3: A schematic depiction of alternative splicing processes as a source of diversification in the expressed transcripts and protein isoforms in cells. In this image, a gene with 4 exons

may undergo different alternative splicing events such as exon skipping or intron retention to yield different types of transcripts.

The total collection of transcripts present in a cell at any given time therefore depends on the biological function and the physiological state of that cell. Profiling the transcriptome of a cell or tissue can then provide crucial molecular insights in the observed phenotype of the cell under study. In cancer, microarrays were previously extensively utilized to generate gene expression profiles of various cancer types. This allowed the classification of tumors, the prediction of overall patient survival and patient responses to therapy[46,47]. Microarrays have fast been replaced by NGS approaches since NGS studies can narrow down the analysis to the isoform level and reveal novel isoforms or even isoform usage in cancer. Bourdon, J.-C. *et al*, investigated the effects of alternative splicing events in the tumor suppressor gene *TP53* and found that *p53* mutant breast cancer patients expressing the *TP53γ* isoform had a low cancer recurrence and better prognosis (similar to breast cancer patients expressing the wild type *p53*) than patients expressing other *p53* isoforms[48].

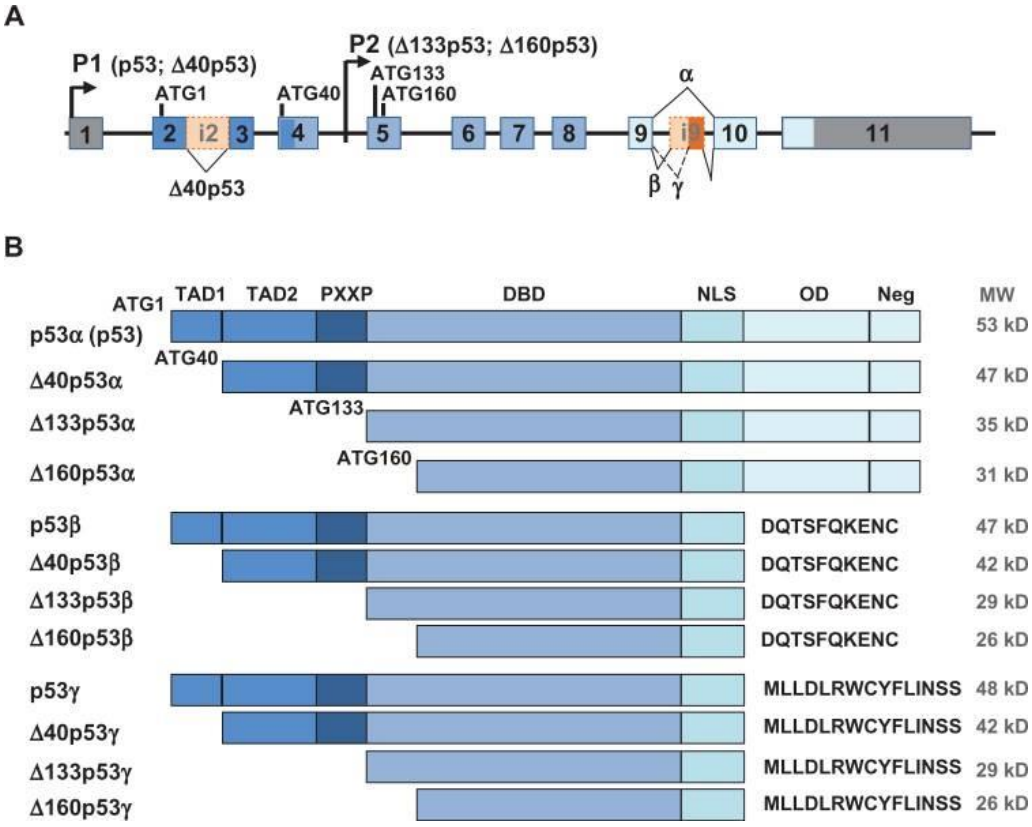


Figure 4: A diagrammatic depiction of the *TP53* gene and its isoform by Surget et. al[1]. The human *TP53* gene encodes twelve different isoforms and consists of eleven exons (A), and 2 promoters (P1: proximal promoter, P2: internal promoter). In B, *p53α* (canonical transcript), *p53β* and *p53γ* isoforms together with known *TP53* domains (TAD1, TAD2, PXXP, DBD, OD, Neg and NLS) are shown. MW: molecular weight, kD: kilo Dalton.

Also, Nagane. M. et al, while using the human glioma cell line (U87MG) found out that the expression of deltaEGFR (EGFR isoform variant without exons 2, 3, 4, 5, 6 and 7) enhanced the tumorigenicity of glioblastomas[49]. Presently, the quest for precision medicine (or personalized medicine) in cancer dictates that an individuals' cancer genome be sequenced to

facilitate the accurate diagnosis, subtyping, and appropriate treatment remedy for that patient. This is envisaged to improve the clinical outcome for cancer patient. National Cancer Institute of the National Institutes of Health, USA, coined the term "personalized medicine" to refer to a form of healthcare that considers a patients' genetic information (expressed genes or proteins) and their environment in order to prevent, diagnose and treat the cancer type the patient has been diagnosed with. Because of heterogeneity in cancer, the overlay of a personal genome with the personal medical record of cancer patients has the potential to improve patient survival prediction, monitor disease progression, and to allow for a more pro-active therapeutic strategy[50]. NGS technology has brought about genome-guided clinical care. This strategy focuses on an individual patient's genomic markers (for example, sequencing a patients' genome and determining the set of markers such as Single nucleotide polymorphism -SNPs) to help ascertain if a patient may to respond to a given therapy, avoid toxic side-effects from certain drugs that may likely not work, and adjust the pharmacological dosage of medications so as to optimize their efficacy and safety. Genetic variations (e.g., SNPs and copy number variants) in humans are recognized as an important determinant of the variability in drug responses across patients diagnosed with a particular disease[51–54].

1.6 Cancer genes: oncogenes and tumor suppressor genes.

Of utmost importance in cancer studies is the search for genes that are involved in tumor initiation and its development. Such genes mainly carry mutations, and often, the mutations are somatic in nature but can also be germline (inherited) mutations. Two different classes of genes – tumor suppressor genes (TSGs) or oncogenes – are the major targets for mutations and variations during the molecular evolution of various cancer types[55–57]. Based on whether the mutations are dominant or recessive at the cellular level, cancer genes can be classified as either oncogenes (e.g., *KRAS, BRAF, EGFR*) or tumor suppressors (e.g., *TP53, PTEN, BRCA1/2*)[58]. Oncogenes possess dominant mutations: a single altered allele is sufficient to initiate cancer. Tumor suppressor genes bear recessive mutations: both alleles need to be changed. Oncogenes often show gain-of-function mutations while TSGs carry loss-of-function mutations. The protein products of oncogenes include transcription factors, chromatin remodelers, growth factors, growth factor receptors, signal transducers, and apoptosis regulators[59]. Oncogenes are altered in ways that render them permanently active or active when they are not supposed to. Around 80% of detected cancer mutations occur in tumor suppressor genes[56]and t. TSGs normally act to inhibit inappropriate cell growth and division, stimulate apoptosis, and repair DNA. In many tumors, these genes are lost or inactivated by genetic or epigenetic alterations, including non-synonymous mutations, insertion or deletions of variable sizes, and epigenetic silencing. After the onset of sequencing technologies and cancer genomics, many research groups and consortia delved into cancer genomics research, with The Cancer Genome Atlas[60] (TCGA) being at the forefront in generating cancer patient sequence data to reveal important genes, proteins and pathways that could be linked to cancer[61].

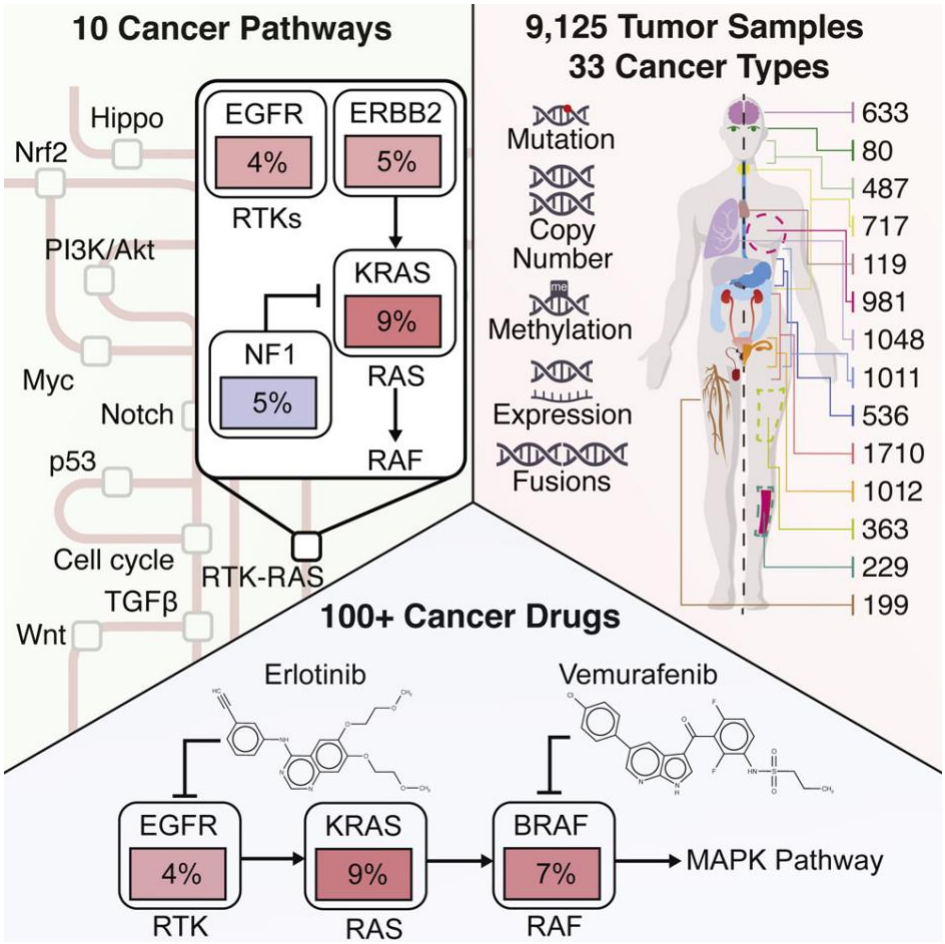


Figure 5: A diagrammatic depiction of the oncogenes and TSGs derived from TCGA data and the pathways they affect. Adapted from Sanchez-vega et al, 2018.

1.7 Significantly mutated genes, underlying pathways and targeted therapy in cancer.

Oncogenes or tumor suppressor genes that harbor significant mutations are termed as cancer driver genes. These mutations are often somatic in nature, i.e., they occur in genes after conception and cannot be passed onto the next generation of offspring, unlike germline mutations[62–65]. Germline mutations account for 5-10% of cancer - Figure 4, with the mutations in disease causing alleles showing complete penetrance. Complete penetrance is the phenomenon where all the individuals diagnosed with a particular cancer type harbor the disease-causing mutation. Somatic mutations result to the majority of the known cancer types. Cancer driver gene products are vital for tumor initiation and progression, and provide growth advantages to cancer cells. TCGA analyses together with other research utilizing the TCGA datasets (e.g. the catalog of somatic mutations in cancer - COSMIC[66]) revealed the genes driving cancer in several tissue types analyzed; a consensus list of 299 genes with KICH having the fewest driver genes (2) and UCEC with the most driver genes (55)[67]. Somatic mutations accumulate during an individuals' lifetime, with the majority of these mutations being unnoticeable. However, some of the mutations, especially those occurring in the protein coding regions[68], can alter crucial cellular functions resulting to cancer.

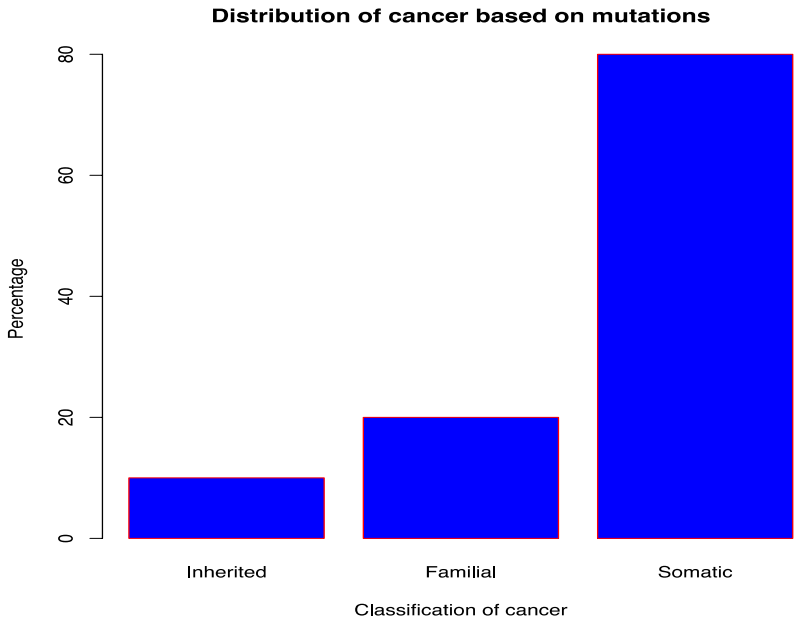


Figure 6: A diagrammatic depiction of the proportion of cancer as a result of mutations. Sporadic mutations (somatic) are responsible for the majority of cancer types, followed by familial cancer types that are as a result of low penetrance mutations coupled with environmental and lifestyle factors. Inherited cancer types account for up to 10% of all cancer types.

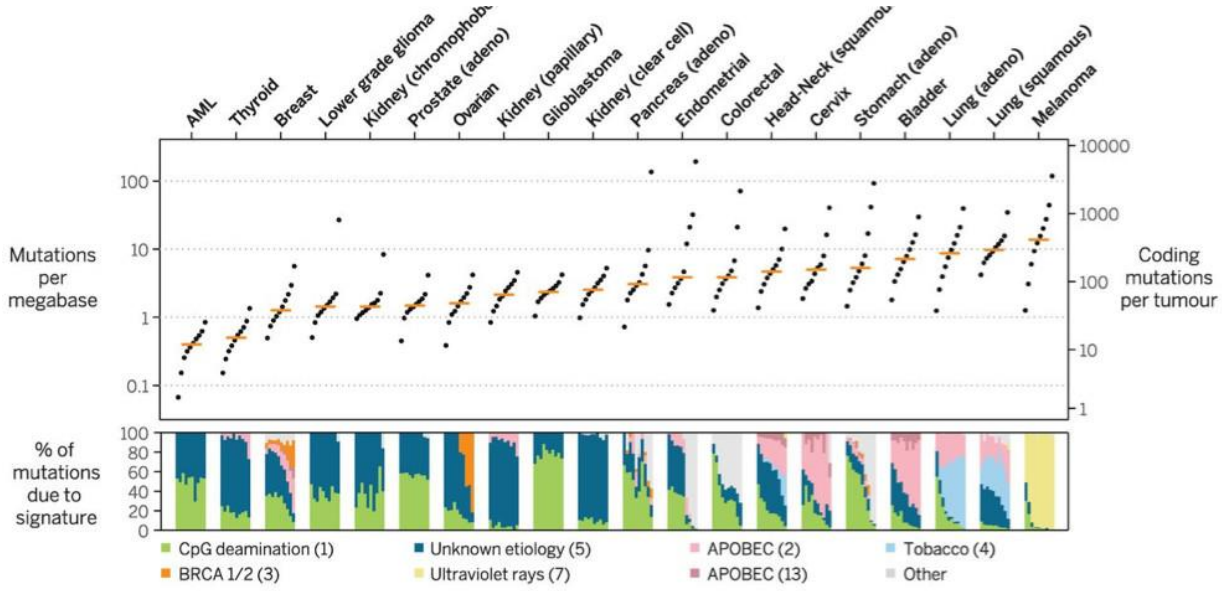


Figure 7: Mutation burden in 20 tumor types and relative contribution of different mutational processes. For each tumor type, samples were divided into deciles on the basis of their mutation burden. The median mutation burden is shown as a dot plot (substitutions and small indels); orange bars denote the median burden of all samples. AML - acute myeloid leukemia (Top). The mean percentage contribution of different mutation signatures is depicted by stacked bars (Bottom). Adapted with modifications from Martincorena, I. & Campbell, P. J, 2015.

Cancer driver genes affect a plethora of molecular functions that manifest into a cancer phenotype. Hanahan & Weinberg, termed these molecular functions or pathways as "cancer hallmarks", and they consisted of six biological processes that were seen to be vital for oncogenesis[14]. They suggested that a cell must first acquire the capability to self-sufficiently grow and become insensitive to antigrowth cell signals. Such a cell should then evade apoptosis while having a limitless replicative potential which would promote sustained angiogenesis, with the consequence of invading adjacent tissues and metastasis - Figure 7.

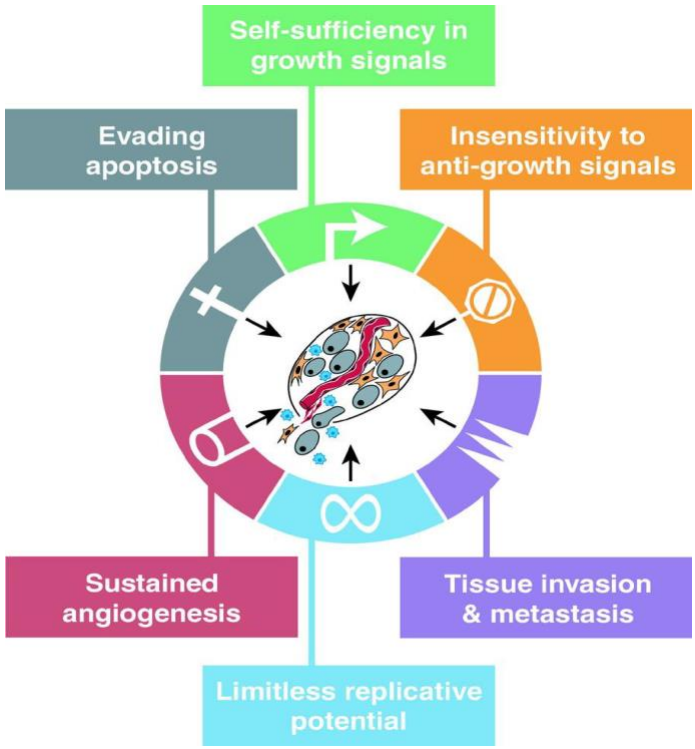


Figure 8: What do cells require to become oncogenic? Six vital biological processes must be evaded by cells and form abnormal proliferative cancer cells. Adapted from Hanahan & Weinberg, 2000.

1.8 PPINs act as sensors and critical drivers of human diseases.

Knowledge on the function and molecular characteristics of individual proteins exists, however, proteins rarely act alone; they connect with several others and build networks (PPINs), which play important roles in cellular functions[69,70]. Protein-protein interactions (PPIs) are significant players in cellular functions and pathways by possessing and transmitting necessary information within these molecular systems[2]. Analyzing PPINs could lead to the unravelling of complex molecular relationships in living systems and understanding today's complex diseases such as cancer.

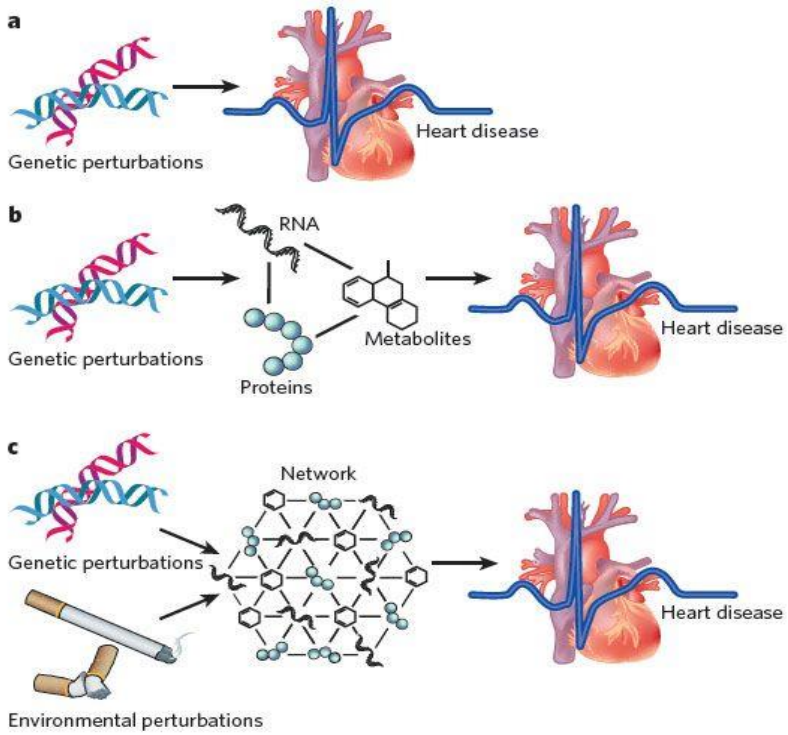


Figure 9: A more realistic disease model is one in which considers multiple sources of perturbations at the network level, e.g., a combination of genetic and environmental perturbations affecting the molecular states of networks. a - Classic genetic association approaches seek to identify variations in DNA that correlate with disease state or with quantitative traits associated with disease. The attraction of this approach is the identification of the genetic causes of disease. b- Changes in DNA on their own do not lead to disease but, instead, lead to changes in molecular traits that go on to affect disease risk. By layering in molecular phenotypes as intermediate phenotypes, causal relationships between genes and disease can be established directly. c -Disease gene networks sense constellations of genetic and environmental perturbations. Adapted with modifications from Schadt E, 2009.

In 2006, Sam et. al., showed that the associations between diseases are directly correlated to their underlying PPINs, thus providing insight into the underlying molecular mechanisms of phenotypes and biological processes disrupted in related diseases[71]. Afterwards, the task of studying, identifying and modelling the complex dynamics of biological molecular systems so as to describe various human diseases gathered immense interest, and this was also aided by advances in the design of computational tools to analyse complex networks. PPIs can either be determined computationally or experimentally. Experimental determination of PPIs is achieved via the use of yeast two hybrid (Y2H) systems or mass spectrometry to detect physical binding or interaction between proteins, whereas computational ("*in sillico*") determination involves using gene context analysis studies such as gene fusion, gene neighbourhood and gene co-occurrences or phylogenetic profiles[72–77].

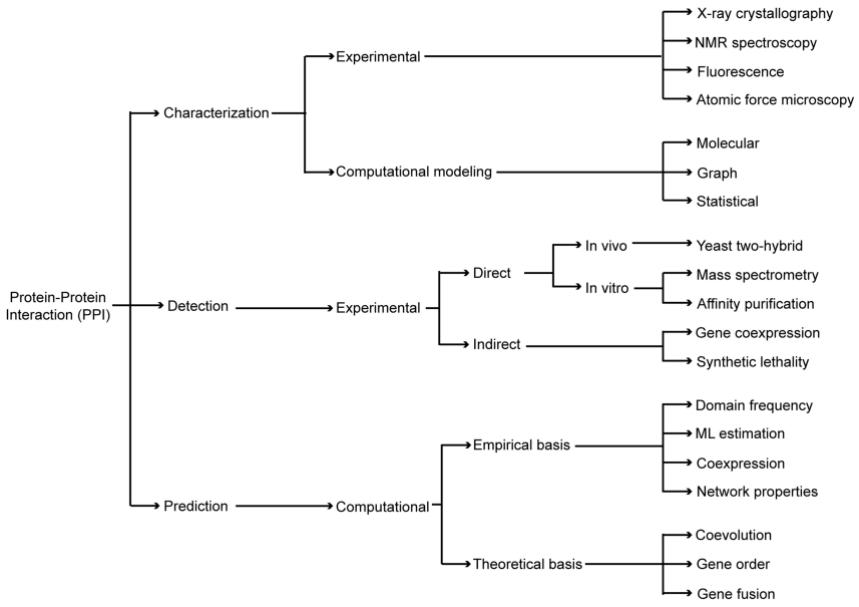


Figure 10: The different experimental and computational methods used in characterizing, detecting, and predicting protein-protein interactions. Adapted from Gonzalez et. al, 2012

While the pharmaceutical industry has invested heavily in cancer drug research and development, there still exists low numbers of new drug approvals or the translation of single biomarkers to the clinic. It is henceforth critical to question whether the single molecule (e.g targeting a single driver gene in cancer) targeted drug discovery approach is the most efficient in combating cancer. Beadle and Tatum's "one-gene/one-enzyme/one-function" hypothesis[78] has now been disapproved in the context of cancer, as evidenced by the early work of Sharma et. al. To counter complex systemic diseases such as cancer, intervention at the biological networks may be advantageous than the use of single target intervention approaches. As such, systems medicine (network pharmacology) approaches have lately been fronted as being highly promising, because they address the ability of targeting multiple proteins or the networks involved in causing the disease.

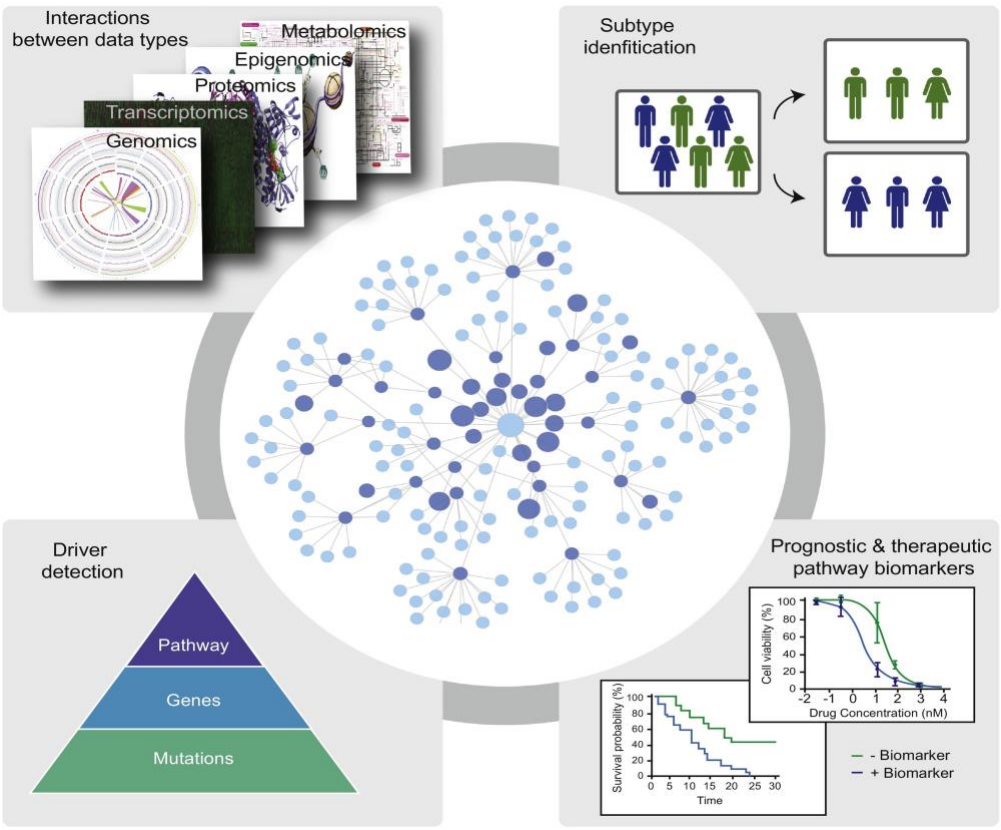


Figure 11: The advantage of using the PPIN to infer cancer biomarkers is the ability to couple the analysis with multiple other OMIC datasets of patient clinical characteristics. Consequently, it is now likely that more personalized, accurate and rapid disease gene diagnostic techniques will now be devised. Adapted from Ozturk et. al, 2018[79].

A paradigm shift in the analysis of PPIN in cancer was observed after 2010 when Vandin et. al., developed HotNet[80]: A novel algorithm that was used to identify significantly mutated pathways (subnetwork biomarkers) in cancer. HotNet assumes a mutation to be a 'heat source' on the network, the heat is then allowed to diffuse across edges, thus spreading its influence across the network dependent on the topology. After diffusion, the network can then be partitioned to reveal 'hot' regions that are likely driver pathways enriched for the influence of the said mutation. The advantage of this approach is that it naturally penalizes highly interconnected regions of the network (the heat must be divided across large numbers of edges) where mutation influence will appear more concentrated at random. Later, an improved version HotNet2[81] was developed, and it proposed an "insulated" heat diffusion model to incorporate edge direction. This allowed the algorithm to capture a sense of effects as either being upstream or downstream of causal mutations, and a damping factor that can be tuned to emphasize local topology over distant network regions. Several other algorithms to decipher cancer pathways have since been developed: MUFFIN, Multi-Dendrix, MEMo, TieDIE[82–85]. Understanding prognosis in cancer is important for clinical decision making; a tumor may be slow in progressing to malignancy, and patients may be over-treated (as is common for ductal carcinoma in situ and prostate tumors), whereas another tumor may be very aggressive and thus require aggressive clinical intervention. Another vital aspect of prognosis is post-treatment tumor progression monitoring due to inter-tumor and inter-patient heterogeneity. Consequently, Individual somatic mutations or overexpressed genes/proteins have limited value as biomarkers. An alternative is the use a panel of biomarkers; however, this comes with the risk of potential overfitting as there is a large number of possible combinations to

explore[86,87]. Networks have since been applied to optimize selection of relevant biomarkers and have even been used directly as biomarkers themselves. When using the PPIN in 2013, Sharma et. al., discovered how human Sirtuin (a group of proteins implicated in numerous biological pathways) work within subnetworks and modules which are enriched in multiple diseases such as cancer[88]. Additionally, Chuang et. al., mapped differentially expressed genes in breast cancer patients on the PPIN and discovered subnetwork biomarkers that could distinguish metastatic from non-metastatic breast tumors[89]. In brief, Chuang et al found at least one cancer susceptible gene (e.g, *TP53, PIK3CA, BRCA1*) in various subnetworks, and these results were comparable with previous results from Van de Vijver et.al., and Wang et. al[90,91]. While these studies have already found genes and PPI subnetworks strongly associated with cancer, none considered alternative splicing and the isoform specific preferences at the patient specific level. As mentioned earlier on, alternative splicing generates multiple proteins that may have different functions and structures from a single gene. The resulting splice variants play significant roles in cancer as alternative splicing can be deregulated through alterations in core spliceosomal components, in an accessory splicing factor or through genomic mutations in splicing motifs. Because of these splice variants, the cell may gain the ability to escape apoptosis[92], a key hallmark in cancer progression[41]. Furthermore, metabolic pathways may be affected due to a switch between antagonistic gene isoforms causing proliferation or affecting tumor suppressors[93].

1.9 OBJECTIVES

The work discussed in this thesis focuses on the identification of edges (connections between interacting proteins) whose interacting protein partners are involved in tumorigenesis. Such proteins that are involved in cancer progression are crucial in tumor monitoring, prognosis and in the development of therapy targets. As already mentioned, the use of a group of proteins (or genes) within a network has gained impetus in characterizing critical biomolecules at play in a variety of complex diseases, especially in cancer. Targeting of such a group of proteins with functional relevance in cancer and other complex diseases promises new paradigms in the treatment of such diseases.

In cancer, for example, tumor diversity across patients diagnosed with a particular cancer type has brought about the advent of precision oncology. In this thesis we sought to utilize publicly available patient data from TCGA to infer how the cancer interactome is modulated following tumorignesis. Use of patient centric data might provide an even more selective targeting of only specific components of the interactome in the quest for identifying druggable proteins at the network level. This work focused therefore on the rewiring of the cancer interactome to pursue a more selective identification of biomarkers at the patient-, cancer subtype, cancer type as well as multi cancer levels. Such biomarkers were identified and then characterized for their modulating properties in promoting oncogenesis via survival analysis.

We also sought to find the effects of SMGs on the cancer interactome: are SMGs responsible for the bulk of edgetic perturbations observed? SMGs have been consistently shown to be important cancer biomarkers and several of them have been used in the design of therapeutics. Such an analysis is crucial in revealing SMGs and their interacting partners that are significantly rewired at the interactome level and thus provide further insight in the determination of the complete set of cancer biomarkers. Coupling advanced NGS data with PPIN data could bolster the discovery of novel biomarkers that will facilitate quick cancer diagnosis, monitoring, and development of correct therapies for clinical use.

CHAPTER 2: EDGETIC PERTURBATION SIGNATURES REPRESENT KNOWN AND NOVEL CANCER BIOMARKERS.

Parts of this chapter have been published in: Kataka, E., Zaucha, J., Frishman, G., Ruepp, A. & Frishman, D. **Edgetic perturbation signatures represent known and novel cancer biomarkers.** ***Sci Rep*** **10, 1–16 (2020).**

Dmitrij Frishman and I conceived the project and I implemented the bioinformatics analyses. Goar Frishman, Andreas Ruepp and I undertook the biological annotation of the protein biomarkers. Jan Zaucha, Dmitrij Frishman and I interpreted the results and wrote the paper.

ABSTRACT

Recent computational tools leverage domain-domain interaction data to resolve the condition-specific interaction networks from next-generation sequencing (RNA-Seq) data accounting for the domain content of the primary transcripts expressed. In the work described in this thesis, we used The Cancer Genome Atlas RNA-Seq datasets to generate 642 patient-specific pairs of interactomes corresponding to both the tumor and the healthy tissues across 13 cancer types. The comparison of these interactomes provided a list of patient-specific edgetic perturbations of the interactomes associated with the cancerous state. We found that among the identified perturbations, select sets are robustly shared between patients at the multi-cancer, cancer-specific and cancer sub-type specific levels. Interestingly, the majority of the alterations do not directly involve significantly mutated genes, nevertheless, they strongly correlate with patient survival. Our findings, which are freely available at EdgeExplorer: http://webclu.bio.wzw.tum.de/EdgeExplorer, are a new source of potential biomarkers for classifying cancer types, and the proteins we identified as significantly being involved in edgetic perturbations are potential anti-cancer therapy targets.

2.1 INTRODUCTION

2.1.1 PPINs in Cancer

Cancer, a leading global health burden, is a complex molecular disease that involves abnormal proliferation of cells with the potential to metastasise to other healthy tissues and organs. To accurately diagnose and treat cancer, better understanding of its molecular pathology and the players involved is required. Indicators of the phenotypes between healthy and disease states are termed disease biomarkers and are used to monitor disease phenotypes as well as develop therapeutic targets. Research from large-scale cancer consortia (e.g. TCGA) have greatly enhanced our knowledge of cancer. Even though central cancer genes responsible for tumourigenesis (driver genes) are known, determination of the complete set of cancer type (or subtype specific) and pan-cancer biomarkers at the network level is a central problem in tumour research. Cancer involves the accumulation of somatic mutations[68] and epigenetic modifications[94], which drive the cells into the malignant state. Recurrent mutations implicated

in tumorigenesis affect highly connected proteins within the protein interaction network[95,96] and are enriched at the interaction interfaces[97,98] and phosphorylation sites[37] signifying their role in rewiring protein interactions[99]. For this reason, cancer has been described as the disease of the interactome[100]. Analyzing PPINs could lead to unraveling complex molecular relationships in living systems and understanding today's most complex diseases such as cancer. Indeed, the network of protein-protein interactions (PPI) has repeatedly allowed for the extraction of molecular features predictive of various phenotypic traits relevant to cancer – the so-called disease biomarkers[101].

PPINs have been proven to be important in cancer research, as perturbations in these networks can be associated with disease states. For example, Cui et al. have identified putative interaction-disrupting mutations occurring at the interfaces of protein complexes and demonstrated that their presence is prognostic of poor survival[102]. In another study, Li et al.[87] developed the "OncoPPI" network of protein-protein interactions (PPIN) relevant to lung cancer, identifying biomarkers that can inform therapeutic decisions according to the drug sensitivity in certain conditions[87]. Nevertheless, the physical disruption of interaction sites by somatic mutations is only one mode of perturbing the interactome. Another relevant cellular process is regulating the expression (and thereby the local molecular concentration) of the interacting proteins[103]; this has been utilized in mining the network of protein-protein interactions to identify modules of differentially expressed genes serving as robust biomarkers indicative of breast cancer metastasis[89] or stratifying patients from several breast cancer subtypes[104]. Furthermore, the phenomenon of "isoform switching", i.e. altering the major splice variant of the gene that is favorably expressed within the cell, has been implicated in driving tumorigenesis and several such switches have been identified as biomarkers predictive of patient survival[41]. Interestingly, the majority of isoform switches that we observed across most of the cancer types could not be explained by somatic mutations in the same genomic locus suggesting that they usually arise through other complex molecular mechanisms[105]. In the case of multi-domain proteins, isoform switching can lead to the loss or gain of a domain responsible for mediating the interaction, thus perturbing the interactome. Recently developed computational tools leverage domain-domain interaction data in order to match transcriptomes to condition-specific interactomes, accounting for the major isoform of the protein that is expressed within the cell[106,107]. This allows comparing the healthy and cancer tissue interactomes from the same patient and identifying both the lost and the gained interactions (edgetic perturbations).

In this study, we analyzed all samples from The Cancer Genome Atlas for which both the healthy and cancer tissue RNA-Seq data was available, thus generating the first large-scale set of patient- and condition-specific interactomes along with the corresponding tumor-specific edgetic perturbations. Crucially, in contrast to recurrent somatic mutations that are typically present in only a small proportion of patients, many of the edgetic perturbations are consistently shared between the vast majority of patients across multiple cancer types, while other sets of perturbations are shared explicitly between patients in a given cancer type or sub-type. We show that in most cancer types the malignant tissue interactome is smaller than the interactome of the corresponding healthy tissue – the only significant exception to this trend was thyroid carcinoma (THCA). Interestingly, even though a considerable number of significantly mutated genes are cancer driver genes, they are not directly involved in a majority of the identified perturbations. Our results show high reproducibility of the perturbed co-occurring network biomarkers within patients of a cancer type (and subtype) and some shared network biomarkers across multiple cancer types. Furthermore, we found known (e.g. *TP73, NTRK1,* and *CDC25C*) and novel cancer biomarkers at the multi-cancer, cancer type and cancer subtype levels. These

findings are a new source of robust biomarkers for detecting or classifying cancer types, may potentially point to new anti-cancer therapy targets and, owing to the extensive literature annotation we performed, they are also a comprehensive publicly available resource ready for experimental validation studies. We corroborate the relevance of the identified targets by demonstrating their strong correlation with overall patient survival and report the previously gathered insights on their role in tumorigenesis. The main goal of this work was to adopt the publicly available paired patient data from TCGA in order to assess patient specific interactomes, describe PPIN patterns specific to a cancer type, subtype or those that are multi-cancer, and enumerate molecular processes arising due to the activities of proteins bringing about differential PPIN in cancer.

Furthermore, we undertook a detailed search from scientific publications to find how the proteins involved in these processes may be linked to cancer initiation, progression or in cancer treatment responses. An advantage in this kind of research which focuses on a patients' genetic profile, is technology advancement in respect to high throughput data generation in biomedicine. This has allowed researchers to comprehensively characterize genomic, transcriptomic, proteomic, lipidomic and metabolomic changes in various cancer types[108,109]. These multiple omic data types allow the understanding of the "geno-pheno-envirotype" (geneome-pheneotype-environment) relationships and the complex biological mechanisms involved in tumorigenesis and other complex diseases. Although developing efficient computational methods for integrating multi-omics data is challenging, the analysis of these heterogeneous data could lead to the capturing of a more accurate picture of the biological processes associating, for example, cancer recurrence and development with the transcript/isoform expression[110]. Furthermore, addition of clinical/phenotype data with information such as cancer subtypes or tumor stages could be of great relevance in finding significant associations applicable in personalized medicine.

2.2 RESULTS

2.2.1 Cancer PPINs are smaller than healthy PPINs in the majority of cancer types

We analyzed 642 paired cancer and healthy PPINs covering 13 cancer types derived from the global protein interaction network using patient-specific mRNA expression profiles. First, we used PPIXPress[106] to construct cancer and healthy patient-specific PPINs. Next, using the Wilcoxon singed-rank test we tested the hypothesis that PPINs are disrupted during tumorigenesis by comparing the number of binary interactions observed in the healthy and the corresponding cancer PPIN for all patients with a given cancer type. Our results show that cancer PPINs are smaller than their corresponding healthy PPINs in 11 cancer types out of 13, and the difference is insignificant only in KIRP (Figure 12: A-K and M).

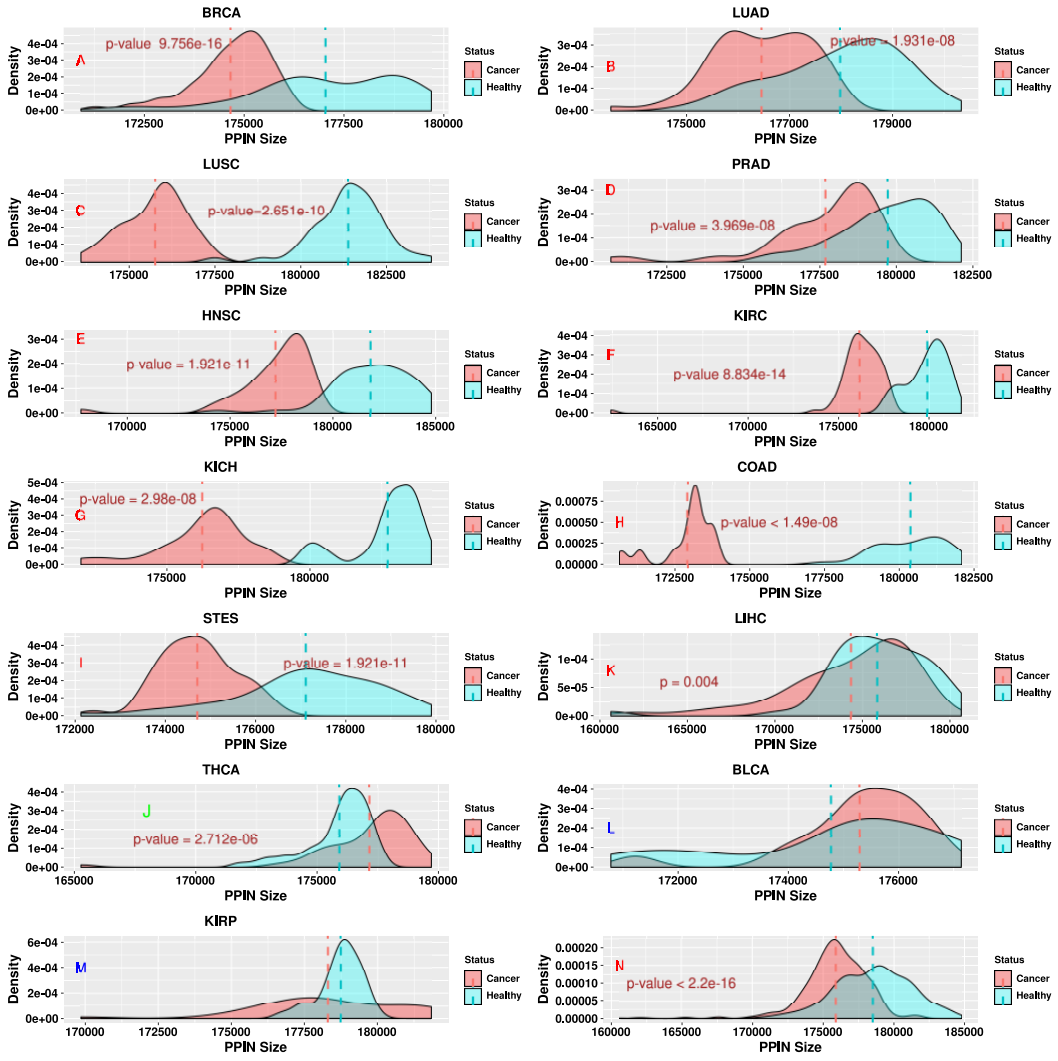


Figure 12. Healthy and cancer PPINs significantly differ in size in 11 out of 13 cancer types (p-value <0.05). The density plots indicate the distribution of paired cancer and healthy PPIN sizes for individual cancer types (A-M) and across cancer types (N). The vertical dashed lines indicate the mean sizes of cancer PPINs (red) as compared to corresponding healthy PPINs (green). For BRCA, LUSC, PRAD, KIRP, KIRC, KICH, COAD, LIHC, HNSC and STES

healthy PPINs were larger than the corresponding cancer PPINs but the difference was not significant in KIRP. For THCA and BLCA (green label), cancer PPINs were larger than the corresponding healthy PPINs, but the difference was not significant in BLCA. In the remaining 2 cases (Figure 12: J and L) cancer PPINs are larger than the corresponding healthy PPINs, but the difference is only significant for THCA (p-value < 0.05). Across all cancer types, cancer PPINs were significantly smaller than the corresponding healthy PPINs (p-value < 0.05, Figure 1N). The mean PPIN size for cancer and healthy samples was 175,888 and 178503, respectively. Similar results (apart from BLCA) were observed when using the randomized PPIN – Supplementary figure 1.

While gene expression signatures have become the mainstay of cancer research, information about global transcriptome shifts between cancer and the corresponding healthy states is only beginning to emerge. In line with our findings, Danielsson et al.[111] reported a reduction in the number of expressed genes in the course of malignant transformation. Distorted gene expression in cancer has been associated with genetic instability (e.g. chromosomal gains and losses[112]) and epigenetic control[112–113]. Anglani et al.[114] reported that gene co-expression networks associated with pancreatic, cervical, gastric and non-small cell lung cancers exhibit losses of connectivity compared with healthy samples while colorectal cancer exhibits more gains in connectivity. We also find that edgetic losses prevail in STES (a type of gastric cancer) and in both LUSC and LUAD (non-small cell lung cancer subtypes), Table 2. In contrast to Anglani et al.[114] we found that colorectal cancer experienced more edgetic losses than gains, probably because our cohort consisted of only colon cancer (but not rectal cancer) patients. However, our results are in agreement with those of Cordero et al.[115], where a significant reduction in colon tumor regulatory networks when compared with healthy samples was reported. This observation implies that colon and rectal cancer types are substantially different in terms of their network dynamics and should be analyzed separately.

Table 2: Characteristics of healthy and cancer PPINs and associated perturbations in 13 cancer types.

Cancer type [a)]	Total gained edges [b)]	Cancer-specific gained edges [c)]	Total lost edges [d)]	Cancer-specific lost edges [e)]	Healthy PPIN size [f)]	Cancer PPIN size [g)]	p-value [h)]
THCA	22831	1271	28065	797	175910	177146	0.0004
BLCA	20739	1463	19030	202	174770	175289	-
BRCA	22195	1453	25516	712	177033	174658	<2.2e-16
COAD	10065	566	21024	953	180365	172933	<2.2e-16
KIRC	18258	462	33005	887	179887	176147	<2.2e-16
KIRP	17174	1085	27141	1117	178740	178300	-
KICH	12423	627	27490	1402	182708	176236	1.037e-15
HNSC	21913	1027	27485	877	181819	177203	1.572e-15
LUAD	16622	959	21907	259	177980	176455	1.397e-11
PRAD	17529	684	22468	915	179710	177677	1.275e-08
LUSC	13108	644	23242	1049	181377	175758	<2.2e-16
STES	24326	1215	20800	835	177110	174708	3.875e-12
LIHC	36458	2019	51445	4068	175831	174337	0.001

[a)] BLCA-bladder urothelial carcinoma, BRCA-breast invasive carcinoma, COAD-colon adenocarcinoma, HNSC-head and neck squamous cell carcinoma, KICH-kidney chromophobe, KIRC-kidney renal clear cell carcinoma, KIRP-kidney renal papillary cell carcinoma, LIHC-liver hepatocellular carcinoma, LUAD-lung adenocarcinoma, LUSC-lung squamous cell carcinoma, THCA-thyroid carcinoma, PRAD-prostate adenocarcinoma, and STES-stomach and esophegael carcinoma

[b)] and [d)] Total number of all perturbed edges (gains and losses) in a cancer type

[c)] and [e)] Total number of edges only observed to be strictly gained or strictly lost in a cancer type, respectively

[f)] and [g)] Total number of edges observed in all healthy samples and cancer samples of a cancer type, respectively

[h)] p value indicating whether PPIN sizes between cancer and corresponding healthy PPINs are different, - Indicates non-significant p value (>0.05).

2.2.2 Isoform switches and resultant domain changes between cancer and healthy states result in edgetic perturbations.

The majority of the identified perturbations across the cancer types resulted from complete-protein-product losses or gains as a consequence of gene expression changes between the healthy and cancer states. Across all cancer types, the cancer state expressed slightly fewer genes than the healthy state apart from THCA, BLCA and KIRP (Dataset 1, freely accessible in the web portal). Nevertheless, we obtained additional perturbations that were attributed to differential isoform expression (resulting in domain composition changes of the majorly expressed protein transcript) between cancer and healthy states – as exemplified in Figure 13 and Figure 14.

To test whether our results were brought about by differential gene expression or were due to domain changes between the healthy and cancer state, we used the R package BiRewire first to generate a randomized network and then analyzed the resulting perturbations. We did this by building condition-specific PPINs in two randomly selected cancer types (BRCA and BLCA). We were able only to recover the prominent proteins involved in edgetic perturbations resulting from the loss or gain of genes in the cancer state (Dataset 2, freely accessible in the web portal. See Supplementary figure 2 on how to access the data from the web-portal). These results indicate that the edgetic perturbations we obtained were indeed a property of the protein interactions and also the expression landscape of the genes. In brief, the cancer state expressed slightly fewer genes than the healthy state, resulting in a reduced number of protein products available to interact with each other. The consequence of this is manifested in the PPIN, where a reduced number of interactions is observed. In BLCA, for example, the mean number of interactions in the cancer state was 98224, while the mean in the healthy state was 107387. For the obtained perturbations, the proteins involved in these disruptions were still predictive of patient survival (Supplementary figure 3).

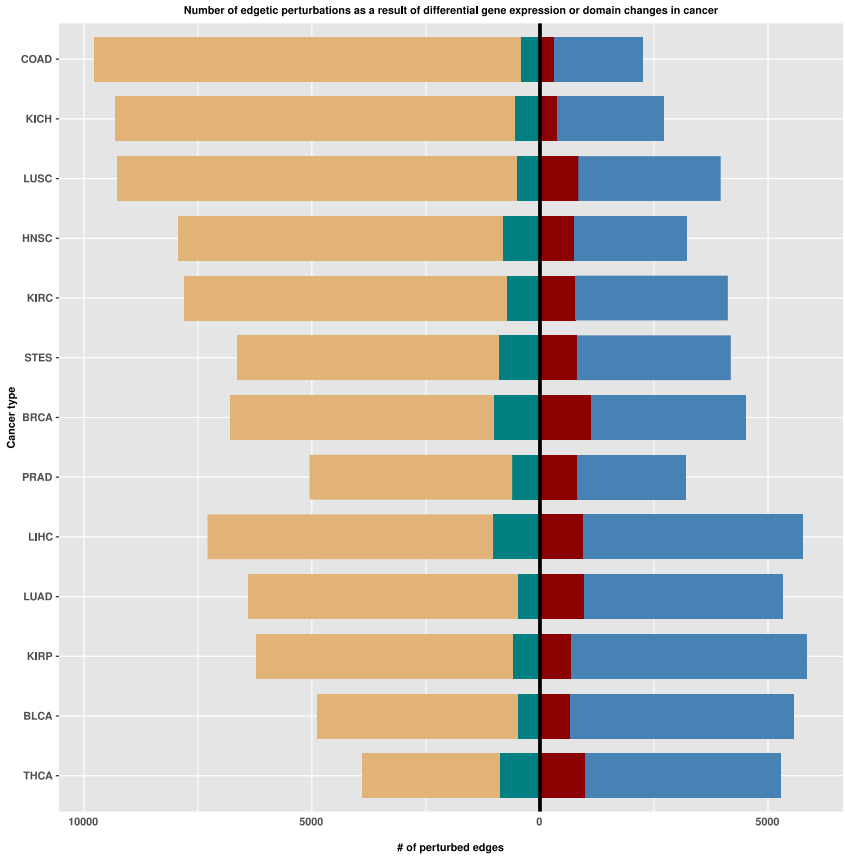


Figure 13. Bar plots indicating the number of edgetic perturbations obtained as a result of gene expression changes or domain changes that come about after isoform switches between cancer and healthy states. Sky blue: edgetic gains as a result of more genes being expressed in the cancer state, red (left of zero intercept): edgetic gains as a result of isoform/domain changes (left of zero intercept). Light brown: edgetic losses as a result of the depletion of genes in the cancer state (left of zero intercept), light green: edgetic losses as a result of isoform/domain changes (left of zero intercept).

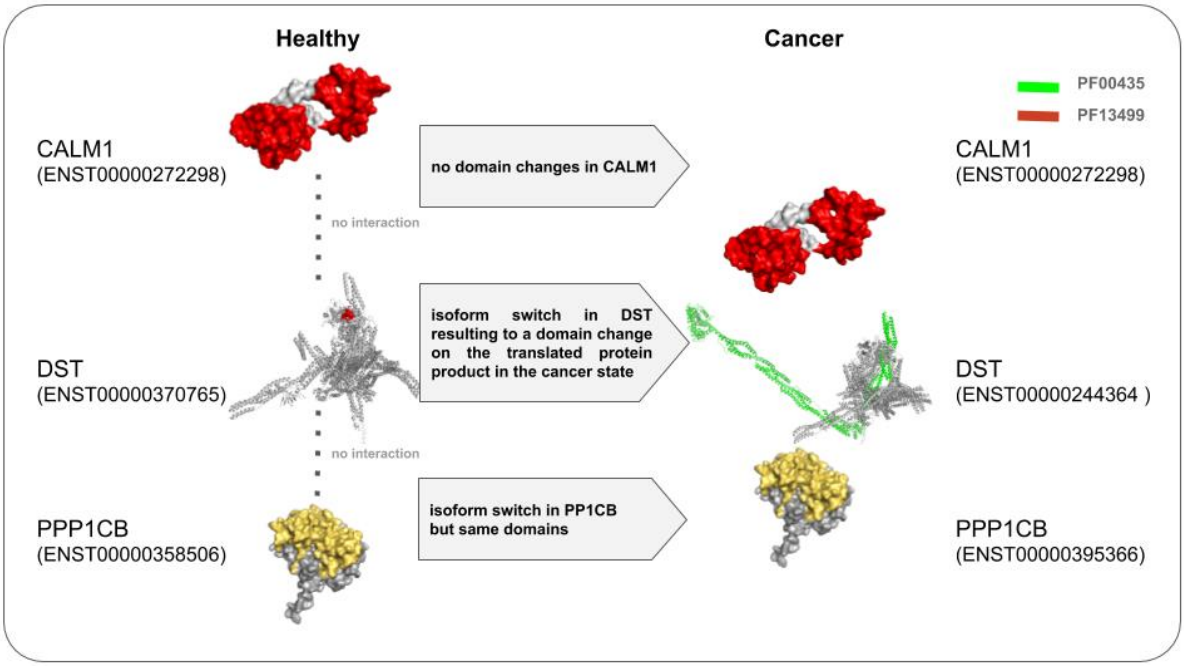


Figure 14. An example showing the consequences of domain changes between the cancer state and healthy state in patients diagnosed with BRCA. The protein structures of both (P0DP23) *CALM1* and (P62140) *PP1CB* were obtained from PDB while those of *DST* were modelled using the ensemble transcript sequences in SWISS-MODEL and visualized in PyMol. Following an isoform switch from ENST00000370765 (in healthy) to ENST00000244364 (in cancer), the protein Q03001 (*DST*) gained the domain PF13499. The consequence is the gain of interactions with the genes *PPP1CB* and *CALM1*.

Of the latter, most perturbations involved an isoform switch in either one of the interacting partners, however, we also identified cases of proteins where across patients of a given cancer type, isoform switches in both proteins were responsible for disrupting the interaction – Table S1. When using the randomized network derived from BiRewire, we were able to reobtain the prominent proteins involved in edgetic perturbations as a result of differential gene expression changes between the cancer and healthy state - Dataset 2. Our findings show that the transformation from the healthy to the cancer state results in (i) the loss or gain of gene expression, which alters the pool of proteins available within the interaction network, and (ii) differential isoform and domain expression, which further translates to the loss or gain of edges between the available proteins. Standard differential co-expression network analyses cannot detect such perturbations, thus making our approach appealing especially in the detection of the repertoire of proteins rewiring the interactome. Here, we corroborate a recent study by Climente-González et al.[116], where the authors suggested that alternative splicing events promote tumor growth by, among other ways, remodelling the protein-protein interaction network.

2.2.3 The identified edgetic perturbations are retained in the protein-abundance filtered PPIN

To test whether our approach yields reliable results, we additionally generated patient-specific PPINs using a smaller network with nodes constituted by highly abundant proteins (see Methods). The majority of the edgetic perturbations identified based on the global PPIN were retained within the reduced high-confidence set. For example, all the 134 edgetic losses involving the nitric oxide synthase inducible protein (*NOS2*) where preserved in the BRCA-specific networks. Other cancer types exhibited only minor variations in the total number of edgetic perturbations identified based on the protein-abundance filtered PPIN (see Table S2, Dataset 3 and in Dataset 4 for details- freely accessible in the web portal). For instance, among the significant edgetic losses in BLCA samples, the protein abundance-filtered PPIN recovered one less edgetic perturbation involving the actin alpha skeletal muscle protein (*ACTA1*) and one of its interactors, neurabin-2 protein (*PPP1R9B*). The protein abundance database (PaxDb) does not report any abundance data for the neurabin-2 protein, and the protein is mainly undetected in the bladder samples from the human proteome map. The results for these comparisons can be found in Table 3, Dataset 3 and in Dataset 4. Due to the modest correlations between mRNA and protein expression data[117,118], it is still challenging to infer protein levels from transcriptome studies. Nevertheless, the majority of our results involve the highly-abundant proteins, which indicates that these interactions constitute the most relevant processes occurring within the cells and corroborates the reliability of the identified perturbations.

Table 3: The identified edgetic perturbations are retained in the protein-abundance filtered PPIN

Cancer type	Total number of perturbed edges using global PPIN	Total number of perturbed edges using protein abundance filtered PPIN	Protein whose interactions are significantly gained across all patients	Number of gained edges observed using global PPIN	Number of gained edges observed using protein abundance filtered PPIN	Protein whose interactions are significantly lost across all patients	Number of lost edges observed using global PPIN	Number of lost edges observed using protein abundance filtered PPIN
BRCA	33485	32829	CDC25C	31	31	NOS2	134	134
BLCA	29867	29284	HIST2H2AC	77	77	ACTA1	95	94
KICH	30497	29860	HRK	6	6	VTN	69	68
KIRC	37773	37081	CDKN2A	124	122	ESRRB	82	82
KIRP	34034	33368	IGF2BP3	72	71	NROB2	47	45
LUAD	27778	27220	ABCC2	38	37	APOA1	57	55
LUSC	28641	28048	HIST1H2AE	50	50	USHBP1	93	92
LIHC	59886	26568	EBF2, ZNF23	1	1	AMHR2	2	2
PRAD	27112	26568	CENPA	61	61	CACNA1A	75	75
THCA	33936	33211	ALK	52	52	VEGFD	8	8
STES	31830	31186	TNFSF11	21	21	HSPA1L	115	113
COAD	24398	23844	KLC3	57	57	RPL10A	98	97
HNSC	36343	35620	FOXL2	56	56	PCK1	68	68

2.2.4 Significantly mutated genes together with proteins having high degrees of connectivity in the PPIN are crucial players in edgetic perturbations of cancer PPINs

Elevated mutation rate is a hallmark of cancer driver genes[31,119,120]. We analysed the involvement of SMGs as well as their first and second network neighbours in edgetic perturbations. Leiserson *et al.* previously suggested that somatic mutations affect subnetworks within PPINs via a heat diffusion model where "hot" nodes/SMGs propagate their heat to neighbouring nodes[81]. First, we found that not all SMGs are involved in edgetic perturbations, but only a specific number in each cancer type (Supplementary Table Ia,b,c). Also, there were significant differences in the proportion of perturbations associated with SMGs and those associated with the randomly generated genes having similar node degrees in the PPINs. A majority of the perturbations across the cancer types had more instances where the portion of the perturbations associated with random genes was more substantial than the proportion of perturbations associated with SMGs. This observation was prominent in BRCA, PRAD and STES where the portion of the perturbations associated with random genes at both the first and second neighbours was significant, while HNSC had no significant differences in the two proportions (Supplementary Table Ib). However, a look into the proteins involved in the majority of the perturbations associated with the random genes (e.g., *SKIP, HIST1H3J* and *EZH2*) revealed that the proteins function in gene expression deregulation in cancer and are potential molecules for therapeutic intervention in cancer[121–127]. With a rise in the interest of therapeutic targeting of cancer enabling proteins at the PPIN level, our findings suggest that therapeutic targeting of only SMGs involved in edgetic perturbations particularly in BRCA, PRAD, STES and HNSC may not yet be a sound idea. However, additional incorporation of epigenetic markers engaged in tumourigenesis of these cancer types may be additionally beneficial as previously suggested [128].

Nevertheless, we found that in 9 out of 13 cancer types, edgetic perturbations were associated with the SMGs ($p < 0.05$, S3 Table c) as compared to edgetic perturbations resulting from randomly generated genes with similar network topologies. Our findings correspond to those of [129–130] who pointed out that somatic mutations occurring at protein interaction interfaces may alter protein-protein interaction networks for example by resulting in loss of interactions or gain of new interactions. Besides, Cui *et al.* while analysing the effects of somatic mutations on the PPIN of liver cancer patients found that SMGs significantly rewire liver cancer PPINs when compared to random non mutated genes[102]. In these 9 cancer types listed above, the instances showing significant perturbations attributed to the SMGs provide opportunities for therapeutic targeting at the PPIN level as is in the case with BH3 like proteins[34] .

2.2.5 Proteins involved in edgetic perturbations affect the overall patient survival and can serve as cancer type biomarkers.

To find out whether changes in the expression of significantly mutated genes (SMGs) are the leading causes of the observed edgetic perturbations, we compared the proportions of perturbations involving SMGs versus those involving randomly generated proteins with a similar network degree. Surprisingly, across the majority of cancer types, more perturbations could be associated with the randomly selected genes rather than the SMGs (Supplementary Table I b-c). For example, when looking at newly gained interactions connected to SMGs in comparison to randomly selected genes of similar degrees, only BLCA and LUSC showed significant enrichment. While the SMGs in our PPINs tended to be high-degree nodes, only a

small number of their interactions exhibited frequent disruptions (in agreement with previous reports[131]), unlike the case for many other genes of a similar degree whose interactions were often perturbed. One possible explanation for this is that a majority of the randomly selected genes were house-keeping genes occupying more central positions in the PPINs[132] and thus highly prone to rewiring as detailed by Kim *et al*[133]. Also, this can mean that SMGs have subtle effects on the PPIN and affect the same interaction partner consistently across patients.

Nevertheless, among the frequently perturbed edges across patients of a cancer type, we found multiple SMGs among the perturbed edges in all cancer types except in LIHC (Supplementary Table Ia). With a rise in the interest for therapeutic targeting of cancer enabling proteins at the PPIN level[128], our findings suggest that extending the range of target proteins beyond only the SMGs may augment the efficacy of anti-cancer treatments. To gain insight into the possible roles the proteins involved in edgetic perturbations may have in tumorigenesis, we used SurvExpress (except for KICH whose data is absent in the database) to analyze if the expression changes of these proteins could predict overall patient survival (OS) and distinguish between patients with longer and shorter lifespans following tumorigenesis. For each cancer type, we selected proteins connected by the significantly perturbed edges and randomly chose a similar number of proteins from the non-perturbed edges to predict overall patient survival. In all cancer types, we found that all the significantly perturbed edges harbor proteins that significantly affect patient survival (log-rank p-value <0.05, (Supplementary Table IIa, and Figure S4) while the non-perturbed edges did not contain proteins that could predict the overall patient survival.

We carried out SurvExpress analysis on all cancer types except in kidney chromophobe (KICH) whose survival data is absent in SurvExpress database. To understand the roles these proteins play in KICH tumorigenesis, we performed text mining in PubMed using the protein identifiers plus the term cancer for each protein involved in significant edgetic perturbations[134]. The results for each individual cancer type are summarized below (and also in the webportal), with the corresponding images available in Supplementary figure 3, and Supplementary Table II. For the results of the survival analysis using the proteins obtained after network randomization, see Supplementary figure 4.

BRCA.
We found that proteins involved in both edgetic gains and losses (e.g., *CDC25C, NOS2*, and *FOXF1*) contribute to BRCA tumorigenesis as previously suggested[135–136]. We further observed that most patients showing significant edgetic perturbations were at a higher risk of BRCA than those who did not have such edgetic perturbations. For example, the edge between the regulator of nonsense transcripts 2 and heparan sulfate 3-O-sulfotransferase 3A1 (*UPF2-HS3ST3A1)* was specifically gained in BRCA patients, and all of them (110/110) were predicted to be at high risk of BRCA related death (shorter lifespan). Our findings also corroborate previous research indicating the cell and tumor specificity of *HS3ST3A1* in BRCA tumorigenesis[137].

LUAD.
The primary protein involved in LUAD specific edgetic gains, mitochondrial 2-oxodicarboxylate carrier (*SLC25A21,* an ornithine decarboxylate carrier), was more significant in predicting a majority of LUAD patients as being at a higher risk of LUAD related death than any other proteins involved in the other perturbations. Tian *et al.* have shown the existence of elevated levels of ornithine decarboxylate (ODC) and polyamines in lung cancer[138] while

Kumar *et al*. have shown that targeting ornithine decarboxylase and related pathways by the agent DMFO/Eflornithine prevents tumor and adenocarcinoma formation in mice infected with lung cancer[139]. Since we identified *SLC25A21* perturbations as being specific to LUAD, our findings suggest that *SLC25A21* and three of its interacting partners (*PPIE, FBX06,* and *NOSIP*) may be essential biomarkers in LUAD and targets for LUAD chemoprevention.

LUSC.
Proteins involved in both edgetic losses and gains may be important in LUSC tumorigenesis. For instance, our study indicates that the mediator of RNA polymerase II transcription subunit 12-like protein (*MED12L)*, a lung cancer marker previously associated with carboplatin-induced cytotoxicity in cancer patients of African descent[140] could be a multiracial lung cancer biomarker and specifically vital for LUSC subtype. While our study revealed that LUAD and LUSC shared a high proportion of edgetic losses, we also found perturbations harbouring proteins distinguishing the two non-small cell lung cancer types. For example, while previous research has linked significant mutation of the T-cell surface glycoprotein CD1b (*CD1B)* protein to non-small cell lung cancer types[141], our study further suggests that *CD1B* may be more relevant to LUSC.

PRAD.
Even though there was no data for deceased patients in the PRAD cohort, our analysis revealed at least 13 out of 52 patients that showed a higher risk of PRAD related death as a consequence of the proteins involved in edgetic perturbations. For instance, we found eight patients carrying perturbations affecting the homeobox protein DLX-2 (*DLX2)* that was explicitly gained in PRAD cancer type, as being at a high risk of PRAD related death. *DLX2* is a novel epigenetic marker used in the identification of PRAD patients for active surveillance[142]. Also, we found an additional patient predicted to be at high risk of PRAD-related death following disruptions involving the galectin-9C (*LGALS9C)* protein. While a recent study identified *galectin-9* as an anti-cancer agent[35], the authors could not confirm if *LGALS9C* or *LGALS9B* (*galectin-9* like proteins) were also anti-cancer agents. Our research suggests otherwise, and implicates *LGALS9C* in tumorigenesis.

KIRC.
Proteins involved in both edgetic gains and losses appear to be essential in KIRC tumorigenesis since a significant number of patients showed a high risk of KIRC-related death (Supplementary Table IIb and Supplementary figure 3). Additionally, we discovered KIRC specific edge losses involving the bcl-2-interacting killer protein (*BIK)*, which was previously reported as a landmark in KIRC oncogenesis[143].

KIRP.
Gene expression changes of the proteins involved in both edgetic gain and loss perturbations could predict overall patient survival, and these proteins have already been found to be critical in cancer progression. For instance, high levels of expression of the protein ribonucleoprotein IMP3 (*IGF2BP3)* is an indicator of kidney tumors more likely to undergo distant metastasis. Moreover, *IGF2BP3* is an independent prognostic marker in kidney cancers[144]. The role of *ASB14* in cancer is largely unknown; however, some ASB proteins have been shown to be

involved in cancer progression (e.g., *ASB3*, *ASB8* and *ASB16* in Kidney cancer)[145]. Our study may be the first to link *ASB14* to kidney cancer: we found *ASB14* and its interactors to be prognostic in KIRP (p =1.52e-08, S4 Table and S2 Fig), making it a viable candidate for experimental validation given the recent knowledge of the role of *ASB* proteins in other types of cancer. Also, our findings agree with those of Prestin *et al.* who showed the deregulation of the nuclear receptor subfamily 0 group B member 1 (*NROB2)* protein to be an important step in renal cancer progression[146]. Additionally, edgetic loss between dickkopf-related protein 1 and MyoD family inhibitor (*DKK1-MDFI),* proteins involved in Wnt signalling[147,148], may suggest that deregulation of the Wnt Signalling pathway is a vital event in KIRP. Since *DKK1* is a tumor suppressor[149], its perturbation in KIRP may be an indicator of why most patients showing this edgetic loss perturbation were at a higher risk of KIRP related death.

COAD.
Proteins involved in both edgetic gains and losses may be essential in COAD tumorigenesis. Our results agree with previous works linking, for instance, overexpression of the melanocyte-specific protein 1 (*CITED1)* to reduced patient survival in intestinal tumors[150] and the voltage-gated calcium channel subunit alpha protein (*CACNA1A)* to patient survival as well as drug resistance in colorectal cancer[151].

THCA.
Proteins involved in both edgetic gain and loss perturbations are engaged in THCA tumorigenesis. Also, our study revealed probable and, to the best of our knowledge, hitherto unknown THCA biomarkers (*RAB40A* and *CSAG1)*. However, the ras-related protein Rab-40A (*RAB40A)* has been shown to participate in ubiquitination and migration in high-grade breast cancer samples[152] while the expression changes of the chondrosarcoma-associated gene 1 protein (*CSAG1),* a cancer-testis antigen, has been reported to be a signature in some human cancer cell lines[153]. Additionally, other cancer testis antigens are prevalent in thyroid malignancies, but their biological roles are still unclear[154].

HNSC.
We found that proteins involved in both edgetic loss and gain perturbations participate in HNSC progression. For instance, we found an 11-gene (*WNK4, SGK1, KLHL2, HSP90AA1, YWAHQ, AKT1, BCL6, CUL3, NEDD4L, STK39, KLHL3*) HNSC-specific loss perturbation signature with the serine/threonine-protein kinase WNK4 *(WNK4)* losing interactions with all the other 10 genes. *WNK4* mutations result in hyperkalemia, cell permeability[155] and recruitment of claudin proteins which promote metastasis in cancer[156]. Additionally, cullin 3 (*CUL3)* has been linked to HNSC metastasis and drug resistance[157]. Our study, therefore, presents a multi-gene HNSC specific biomarker that may be of use in clinical monitoring and therapy decision making

STES.
Proteins involved in edgetic losses (e.g., *HSPA1L*) may be more oncogenic than those involved in edgetic gains: twice as many STES patients were predicted to be at a higher risk of STES related death by the proteins engaged in edgetic losses. The perturbation of the heat shock protein (*HSPA1L/HSP70-hom)* in our analysis supports the current knowledge of the deregulation of *HSP70* anti-apoptotic family members in gastric cancers. *HSP70* proteins are pivotal in the folding of proteins or refolding of denatured proteins and have been shown to be

prognostic in gastric cancers[158]. Our study, therefore, additionally supports that *HSPA1L* may also be a therapeutic target for STES.

LIHC.
Survival analysis revealed that proteins involved in both edgetic gains and losses may participate in tumor growth and are essential for patient stratification. Our findings corroborate previous research linking increased expression of the protein Wnt-3a (*WNT3A)* to tumor cell proliferation in LIHC[159]. We found a 14-gene edgetic gain perturbation biomarker consisting of *WNT3A, HSPA5, LRP6, CANX, TRAF2, FZD2, FZD1, KCTD1, PPP2R1B, PPP2R5D, PPP2R5A, PPP2R5B, PPP2R5E,* and *PPP2R2D*. This 14-gene signature presents a biomarker for probable therapy targeting via microRNA-195, as previously suggested[159].

BLCA.
Our results suggest that proteins involved in both edgetic gains and losses are essential biomarkers in BLCA tumorigenesis and represent candidate BLCA biomarkers. For instance, perturbations involving the histone protein *HIST2H2AC* may be responsible for the epigenetic changes in BLCA tumorigenesis. Accumulation of mutations in *HIST2H2AC* has previously been linked to tumorigenesis in cancer[160]. Additionally, Monteirro *et al*. have recently confirmed that indeed *HIST2H2AC* may be a biomarker in BRCA[161]. Since we have already shown a close relationship between edgetic gain perturbation in BRCA and BLCA, we tend to think that *HIST2H2AC* may also promote tumor proliferation in BLCA. To our knowledge, this study is the first to link *HIST2H2AC* to BLCA oncogenesis.

KICH.
To determine if proteins involved in significant edgetic perturbations in KICH play a role in oncogenesis, we searched in PubMed for publications linking these proteins to cancer and specifically to KICH oncogenesis[134]. The top gained biomarker in KICH (gained in 25/25 samples) included a 14-gene signature (*HRK, BCL2, BCL2L1, MCL1, ELAVL1, BCL2A1, GRPR, CEP250, CDK5, DCLK3, DGUOK, SLC12A5* and *NUFIP1*) with the activator of apoptosis harakiri (*HRK)* protein gaining interactions with all the other 13 genes. While *HRK* together with other pro-apoptosis BH3-only members of the Bcl2 family have been extensively linked to apoptosis[162] and possible cancer therapy[163], to our knowledge, no study has linked them directly to KICH oncogenesis. Our study uncovered deregulation of several Bcl2 family members in KICH, and this information may be critical for therapeutic targeting for the only clinically approved drug venetoclax in treating leukaemia[34].

2.2.6 Cancer subtypes exhibit unique edgetic perturbation patterns

Among the cancer subtypes, we also searched for subtype-specific disruptions (Supplimentary Table Ib). We found that subtypes differed in their edgetic perturbations and that most of the proteins involved in these network disruptions might be responsible for the observed subtype phenotypes. For example, network edgetic disturbances involving the ribonucleoprotein IMP3 (*IGF2BP3)*, which frequently occurred in ER+, PR+, HER− subtypes, and those involving the m-phase inducer phosphatase 3 protein (*CDC25C*), often observed in ER−, PR−, HER+ subtypes, revealed the mutual exclusivity nature of BRCA subtypes. Also, some of

the proteins whose edges were specifically perturbed within patients grouped in a particular cancer subtype may be novel subtype-specific biomarkers. For example, the protein cytochrome P450 1A1 (*CYP1A1)* is a probable biomarker in Classical LUSC subtypes, neuron navigator 2 protein (*NAV2)* in MSS STES subtypes and the apin protein (*ODAM)* in THCA BRAF-like subtypes. Furthermore, we found interacting proteins whose connections were differentially perturbed across cancer subtypes. For instance, while PR−/ER− BRCA and KIRP Type1 subtypes shared nearly all (71/73) edgetic gain perturbations involving the gene *IGF2BP3*, KIRP Type 1 also had two other edgetic perturbations affecting the *IGF2BP3* gene (*IGF2BP3 -KRT17* and *IGF2BP3 -SYT17*) suggesting that these proteins may have a probable role in the differential mechanisms between BRCA and KIRP tumorigenesis. Besides, we discovered biomarkers shared by several subtypes, for example, the core components of the nucleosome (*HIST1H2AB, HIST2H3A* and *HIST1H3A*) in Secretory and Classical LUSC, PRAD SPOP and BRCA HER+ subtypes. Somatic mutations in these histone proteins have previously been linked to cancer, thus suggesting the relevance of these molecules as candidate cancer subtype-specific biomarkers[122,164,165].

2.2.7 Proteins involved in cancer-specific edgetic gains and losses possess distinct functional roles.

Based on the perturbation profiles associated with each cancer type, we identified two different edgetic events – those occurring in only one patient (patient-specific perturbations) and those occurring in at least 2 samples (cancer type perturbations). In the latter, perturbed edges present in only one cancer type are cancer-specific perturbations (Supplementary Table III) while those present in at least 2 cancer types are multi-cancer perturbations.

We found that LIHC had the highest number of cancer-specific edgetic gains (2019) while KIRC had the lowest number of such gains (462). LIHC and BLCA had the highest (4068) and the lowest (202) number of cancer-specific edgetic losses, respectively (Supplementary Table III). Overall, LIHC had the highest number of both cancer-specific edgetic gains and losses (6087), meaning that LIHC is more susceptible to cancer-specific perturbations (unique perturbations) than other cancer types. On the other hand, LUAD had the least number of both cancer-specific gains and losses (1218), suggesting that LUAD is least susceptible to cancer-specific perturbations, and is more likely to share most perturbations with other cancer types. 7 of the 13 cancer types (COAD, PRAD, KICH, KIRP, KIRC, LUSC and LIHC) had more cancer-specific edgetic losses than gains while the rest (THCA, BLCA, HNSC, STES, LUAD and BRCA) had more cancer-specific edgetic gains than losses (Supplementary Table III). These findings are in line with previously published results, which suggest that the liver has a large number of genes showing tissue specific expression[166–167], while the lung has a low number of such genes[168].

To explore the biological implications of edgetic perturbations we carried out a GO enrichment analysis using topGO and then employed REVIGO to group together the enriched GO terms. Among the proteins involved in edgetic gains, REVIGO summarized their enriched GO terms into 8 biological processes (BP), 14 cellular components (CC), and 51 molecular functions (MF) (dispensability value < 0.05 after REVIGO pruning, Figure 15A-15C).

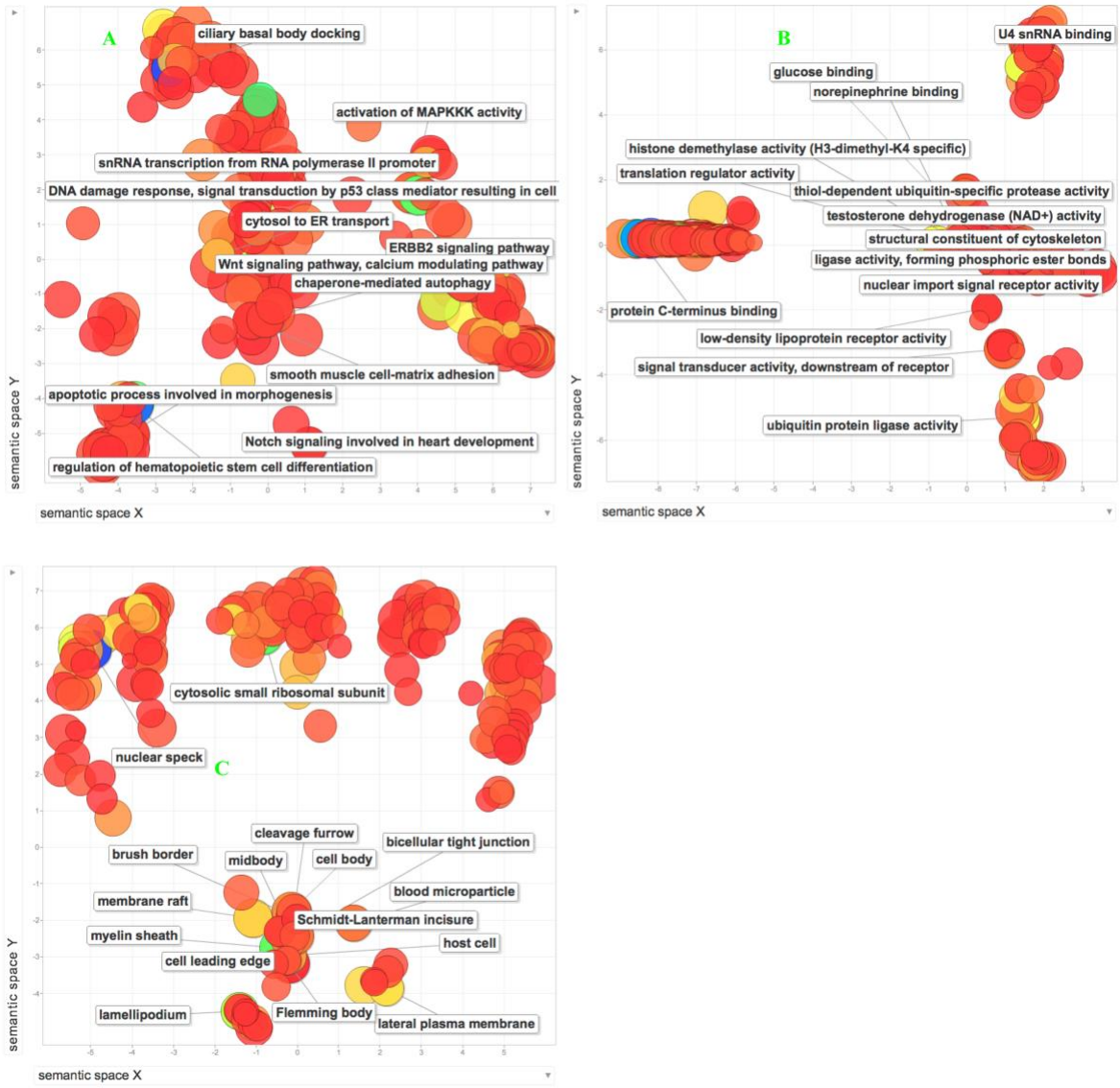


Figure 15. A two-dimensional scaling projection of the enriched Biological processes (A), Cellular Components (B) and Molecular functions (C) for proteins involved in cancer-specific edgetic gains after REVIGO pruning (dispensability value < 0.005). Dispensability of a term represents both the degrees of redundancy and enrichment. The lower the dispensability of a term, the least redundant and more significant a term is. The axes show the distribution of the GO terms based on their semantic similarities. The bubble color reflects the degree of significance (p-value) with blue color indicating a higher significance than the red color. The richly colored bubbles in the foreground represent GO terms with a dispensability value of < 0.005. The bubble sizes indicate how often a GO term occurs, the bigger the size the more frequent the term is.

Of these enriched GO terms, 2 biological processes (lysosomal transport and viral process), 5 cellular components (focal adhesion, retromer complex, nucleoid, ribbon synapse, Flemming body), and 8 molecular functions (transcription factor activity- RNA polymerase II transcription factor binding, transcriptional activator activity-RNA polymerase II core promoter proximal region sequence-specific binding, low-density lipoprotein receptor activity, signal transducer activity, downstream of receptor, structural molecule activity, protein transporter activity, ubiquitin protein ligase binding, translation regulator activity, histone

methyltransferase activity (H3-K27 specific), ubiquitin-protein transferase activator activity) had a dispensability value of 0. Our results support previous findings that suggest that lysosomal transport and viral processes mediate cell proliferation and apoptosis in cancer cells[169,170] by targeting cellular components such as the focal adhesions or retromer complex[171,172]. For the proteins involved in edgetic losses, REVIGO clustered their enriched terms into 7 biological processes, 17 cellular components, and 51 molecular functions (dispensability value < 0.05 after REVIGO pruning, Figure 16A-16C).

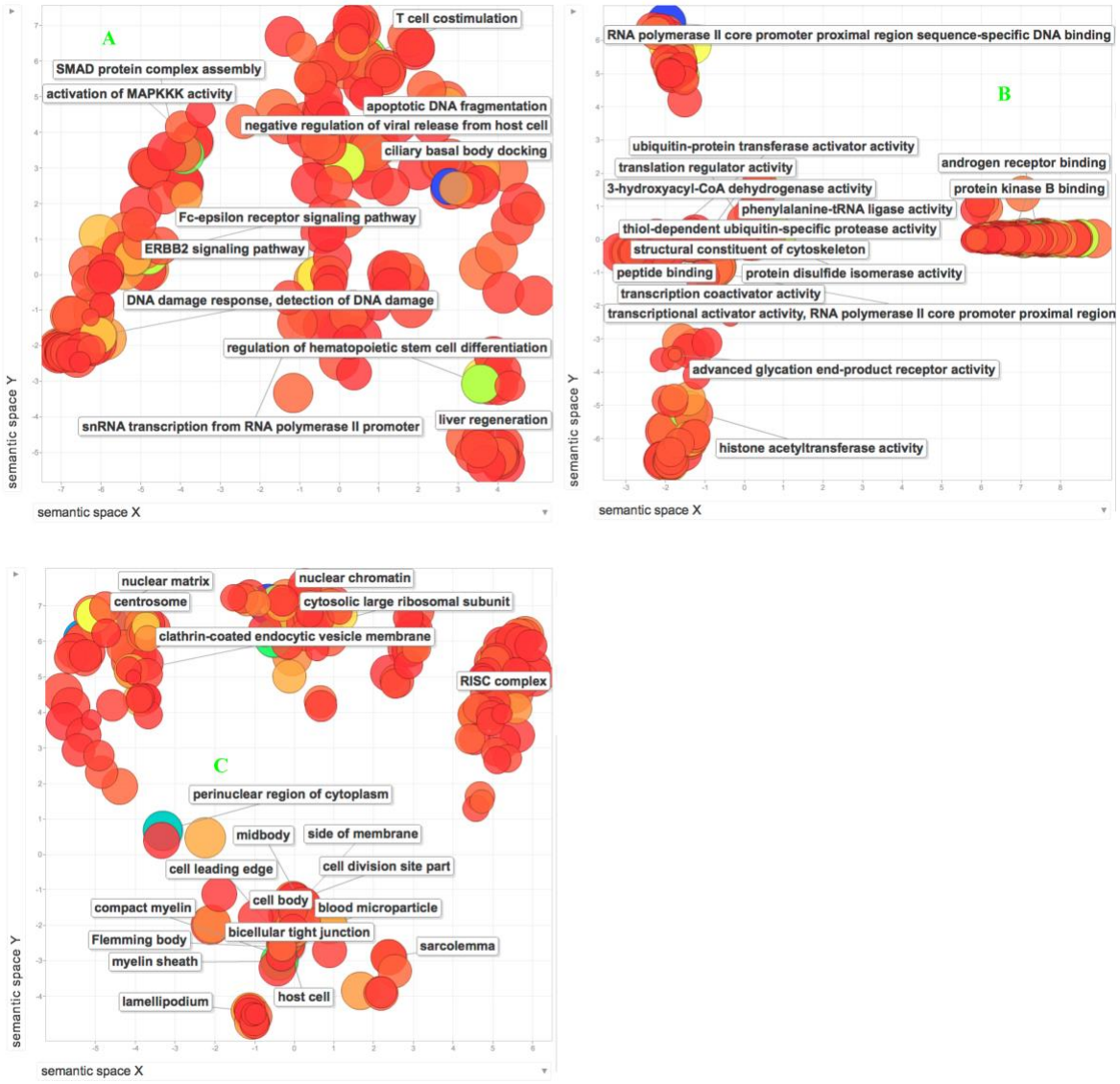


Figure 16. Two-dimensional scaling projections of the enriched Biological processes (A), Cellular Components (B) and Molecular functions (C) for proteins involved in cancer-specific edgetic losses after REVIGO pruning (dispensability value < 0.05). Dispensability of a term represents reduced redundancy and a high degree of enrichment. The lower the dispensability of a term, the least redundant and more significant a term is. The axes show the distribution of the GO terms based on their semantic similarities. The bubble color reflects the degree of significance (p-value) with blue color indicating a higher significance than the red color. The richly colored bubbles in the foreground represent GO terms with a dispensability value of < 0.005. The bubble sizes indicate how often a GO term occurs, the bigger the size the more frequent the term is.

Of these, 3 biological processes (negative regulation of transcription from RNA polymerase II promoter, anterior/posterior pattern specification, entry of bacterium into host cell), 3 cellular components (focal adhesion, RISC complex, host cell), and 11 molecular functions (see details in Dataset 5, transcription factor activity, protein binding, RNA polymerase II transcription cofactor activity, transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific binding, protein disulfide isomerase activity, protein transporter activity, ubiquitin protein ligase binding, translation regulator activity, advanced glycation end-product receptor activity, ubiquitin-protein transferase activator activity) had a dispensability value of 0. These results complement previous work that indicates the importance of pathogens and transcription deregulation via the RISC complex during tumorigenesis[173–174]. On the one hand, our results may suggest that proteins involved in edgetic gains may be recruited to upregulate cancer cell proliferation and put critical pathways under stress, as previously suggested[175]. On the other hand, edgetic losses appear to cause the deregulation of transcription activities as well as the distortion of epithelial cell polarity, an essential process in cancer cell transport membranes[176].

2.2.8 Hierarchical clustering of perturbed edges reveals cancer types sharing similar perturbation signatures

Cancer hallmarks often cut across cancer types[28], we thus sought to find out whether cancer types might also share perturbed network edges. To this end we merged all lost and gained edges to build multi-cancer loss and gain profiles, respectively. The maximum number of cancer types sharing edgetic perturbations (either gains or losses) was 9 out of 13. We found 82 and 2178 gained and lost edges shared across 9 cancer types, respectively, with the Q9BZD4 (*NUF2)* and P04629 (*NTRK1)* proteins associated with the largest number of perturbations (Table 4).

Table 4: Proteins driving pan-cancer edgetic perturbations

Type of perturbation	Gene identifier
Proteins involved in edgetic gain perturbations	*NUF2(30), CDC45(23), ZIC2(6), CANPA(3), TICRR(3), NEIL3(6), TOPBP1(2)*
Proteins involved in edgetic loss perturbations	*NTRK1(1787), TDGF1 (23), AVPR2(13), MAPK4(11), PTPN5(12), CNTN1(10), CHD5(14), ITLN1(10), PACRPG(20), MYOC(36), CA14(36), CAMK2A(39)*

The numbers in the brackets indicate the total number of edges perturbed at that protein node

For instance, edgetic gains involving the *NUF2* protein were observed in all cancer types except for COAD, KICH, KIRC and PRAD. Edgetic losses involving the *NTRK1* protein were observed in all cancer types except for STES, BLCA, LUSC and PRAD. The majority (98.78%) of edgetic gains involved 6 proteins, with the *NUF2* protein alone being a subject in 36.58% of the perturbations. Most edgetic losses (98.76%), on the other hand, involved a common set of 35 proteins, with the *NTRK1* protein being involved in 88.34% of the perturbations. Both of these proteins are known cancer drug targets and are now being considered as crucial molecules in the development of tumor-agnostic drugs to treat diverse cancer types[177]. Silencing of the *NUF2* protein has been shown to hinder tumor growth across cancer types[178,179] while deregulation of the *NTRK1* protein has been successfully targeted by

the drug Entrectinib[180]. Our results, therefore, suggest that the drug Entrectinib may be a choice in the treatment regimen of a diverse number of cancer types but may not be beneficial to patients diagnosed with STES, BLCA, LUSC (apart from ROS1-positive) and PRAD. A higher proportion of the multi-cancer edgetic losses compared to edgetic gains implies that cancer progression favors the loss of crucial protein interactions preventing the cell's safeguards from inhibiting malignant proliferation. This phenomenon was also observed in the SMGs, most of them were involved in edgetic losses rather than in edgetic gains. A subset of the edgetic losses can be attributed to the truncation of proteins leading to the loss of the domains responsible for mediating the interaction (Figure 2), while the remaining edgetic losses are due to a complete loss of expression of the specific genes, a phenomenon previously implicated in oncogenesis[111]. Analysis of the significantly enriched KEGG pathways affected by the proteins involved in multi-cancer edgetic perturbations revealed known pathways[181,182] deregulated across cancer types - (e.g., hsa05200 - pathways in cancer, hsa04120 - ubiquitin mediated proteolysis), Figure 17, Table 4, Dataset 3 and Dataset 5.

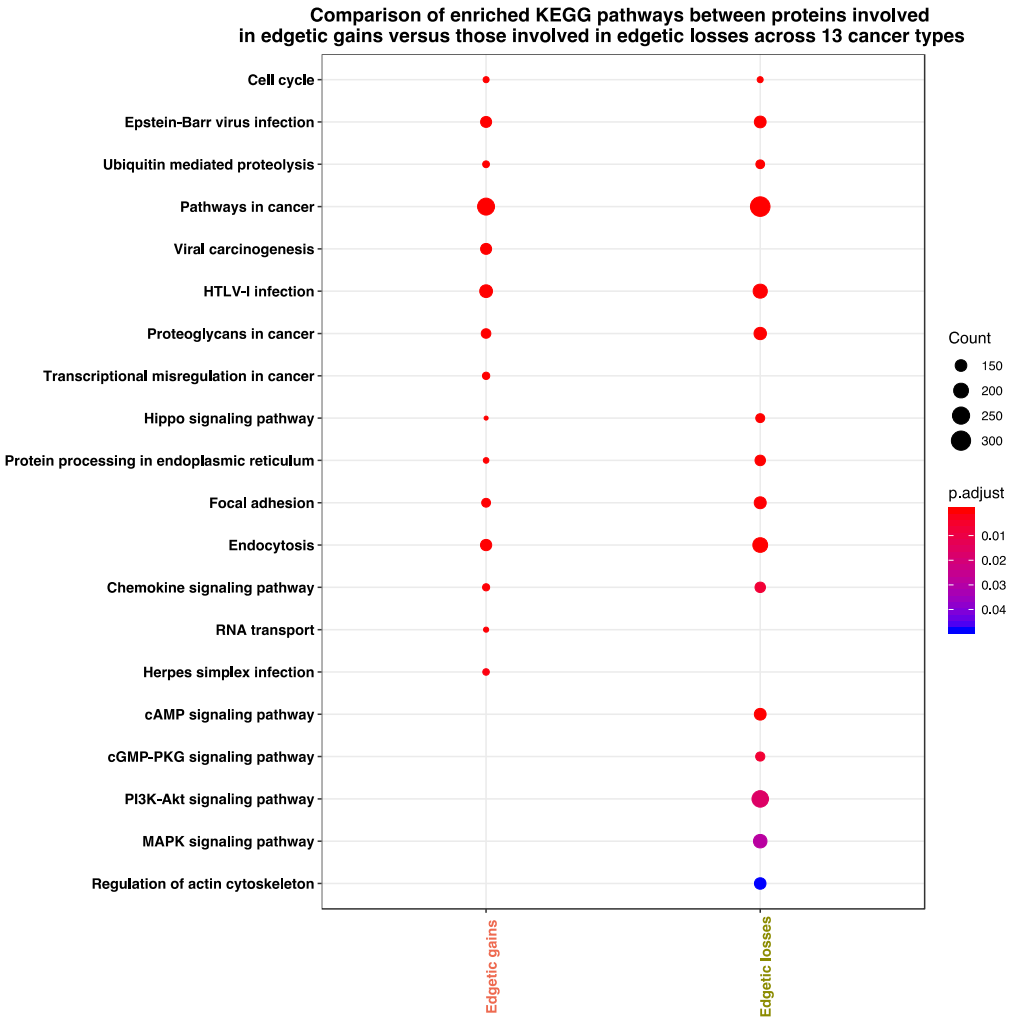


Figure 17. KEGG pathways differentially enriched between the proteins engaged in edgetic gains and those involved in edgetic losses. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate how often a KEGG pathway term occurs. The bigger the size, the more frequent the term occurs.

Additionally, we observed KEGG pathways that were unique only to the proteins involved in edgetic gains (e.g. hsa05203 - viral carcinogenesis, hsa03460 - fanconi anaemia pathway and hsa03008 - ribosome biogenesis in eukaryotes) or in edgetic losses (e.g., hsa04024 - hsa04022 cAMP signalling pathway and hsa04024 - cGMP-PKG signalling pathway) (Table 5a-b, and Dataset 5).

Table 5a: The unique KEGG pathways affected by the proteins involved in multi-cancer edgetic gains

KEGG pathway identifier and term
hsa03460: Fanconi anemia pathway
hsa03008: Ribosome biogenesis in eukaryotes
hsa03018: RNA degradation
hsa00190: Oxidative phosphorylation
hsa03420: Nucleotide excision repair
hsa04114: Oocyte meiosis
hsa03440: Homologous recombination
hsa04622: RIG-I-like receptor signaling pathway
hsa05016: Huntington's disease
hsa00310: Lysine degradation
hsa05323: Rheumatoid arthritis
hsa03430: Mismatch repair
hsa00051: Fructose and mannose metabolism
hsa04370: VEGF signaling pathway
hsa03030: DNA replication
hsa04966: Collecting duct acid secretion
hsa04730: Long-term depression
hsa05416: Viral myocarditis
hsa04960: Aldosterone-regulated sodium reabsorption
hsa04330:Notch signaling pathway
hsa04710:Circadian rhythm

Table 5b: The unique KEGG pathways affected by the proteins involved in multi-cancer edgetic losses

hsa04728: Dopaminergic synapse
hsa04922: Glucagon signaling pathway
hsa04022: cGMP-PKG signaling pathway
hsa04261: Adrenergic signaling in cardiomyocytes
hsa05031: Amphetamine addiction
hsa05030: Cocaine addiction
hsa00010: Glycolysis / Gluconeogenesis
hsa00020: Citrate cycle (TCA cycle)
hsa05020: Prion diseases
hsa04923: Regulation of lipolysis in adipocytes
hsa05410: Hypertrophic cardiomyopathy (HCM)
hsa00970: Aminoacyl-tRNA biosynthesis
hsa04340: Hedgehog signaling pathway
hsa01200: Carbon metabolism
hsa04720: Long-term potentiation
hsa01130: Biosynthesis of antibiotics
hsa01230: Biosynthesis of amino acids
hsa03015: mRNA surveillance pathway
hsa04921: Oxytocin signaling pathway
hsa05414: Dilated cardiomyopathy
hsa04713: Circadian entrainment
hsa04725: Cholinergic synapse
hsa04961: Endocrine and other factor-regulated calcium reabsorption

Even though intra-tumor heterogeneity offers crucial data during therapeutic decision making[183,184], biomarkers cutting across multiple cancer types are invaluable in the clinical research set up as they shed light on pathways shared across cancer patients and inform on inter-tumor heterogeneity[185]. Such biomarkers may be used as standard assessment biomolecules to facilitate the interpretation of laboratory test results in the clinic or help in establishing commonly distorted biological pathways in cancer, as suggested by[186]. Because the edgetic gains and losses are responsible for affecting different molecular pathways, we considered them separately in clustering cancer types based on the perturbations observed. We performed this analysis three times (for edgetic gains/losses separately and considering all data together) under the assumption that the majority of gains or losses may have some common underlying cause (for example, molecular pathways), which may persist across multiple cancer types. Hierarchical clustering of the perturbation patterns using the R package Pvclust identified high confidence (p-value < 0.05) cancer clusters based on shared perturbation signatures (Figure 18A-18C). Using the random forest algorithm (see Methods), we found sets of edgetic perturbation patterns important in grouping cancer types into the identified clusters (Figure 18).

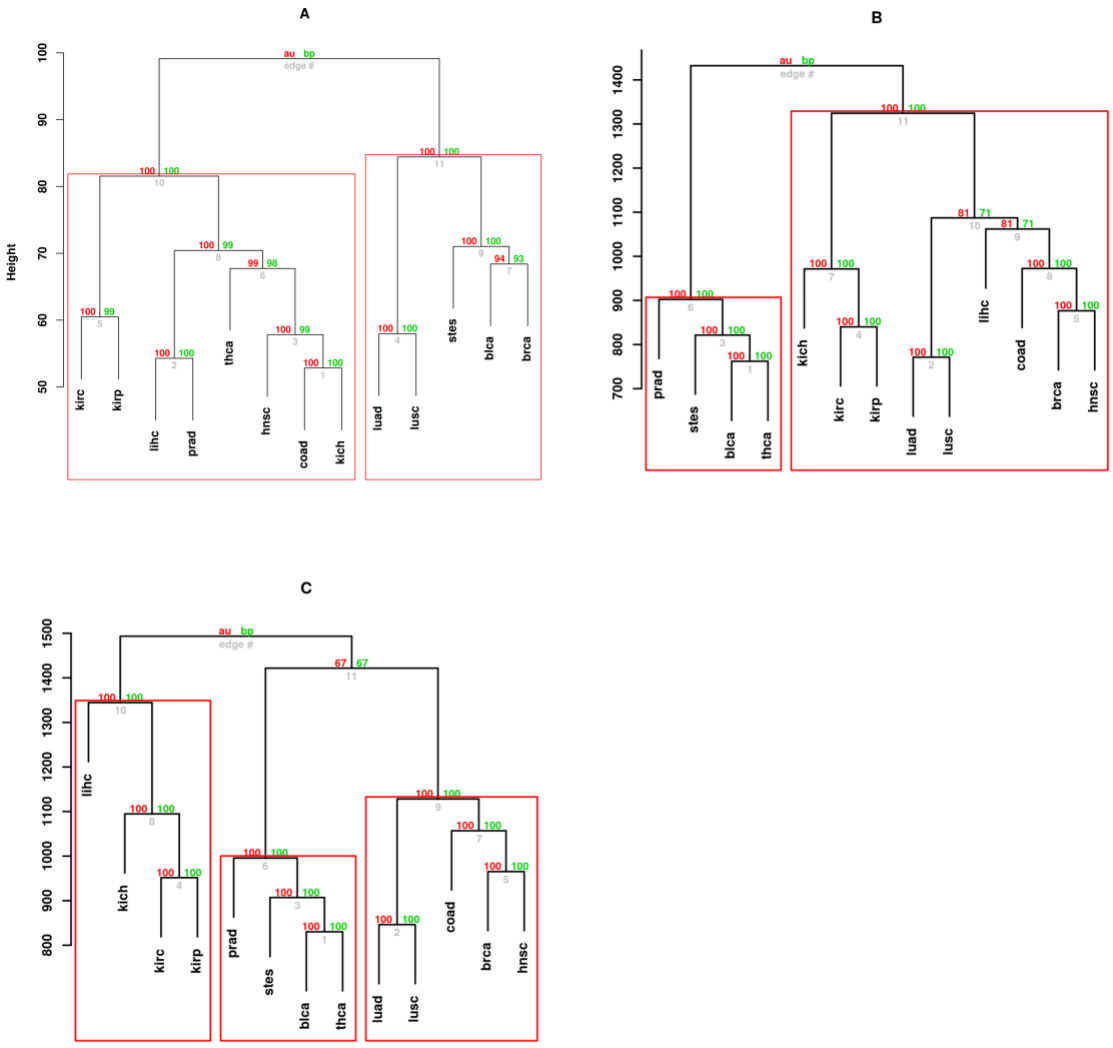


Figure 18. Cancer types share multiple perturbation patterns: Dendrograms based on edgetic gains (A), edgetic losses (B) and both edgetic gains and losses (C) across cancer types. Gained edges revealed 2 main clusters (A) with sub-clusters consisting of (i) BRCA, BLCA and STES, (ii) LUAD and LUSC, (iii) COAD and KICH, (iii) LIHC and PRAD, and (iv) KIRC and KIRP. Lost edges identified 2 main clusters (B) with additional sub-clusters consisting of (i) KICH, KIRP, and KIRP, (ii) LUAD and LUSC, (iii) COAD, HNSC and BRCA, (iv) STES, BLCA and THCA. Clustering of both edgetic gain and loss patterns revealed 3 main clusters (C) consisting of (i) LIHC, KICH, KIRC, KIRP, (ii) PRAD, STES, BLCA, THCA and (iii) LUAD, LUSC, COAD, BRCA and HNSC. The Approximately unbiased AU (green) and Bootstrap probability BP (red) scores indicate the likelihood of observing the obtained clusters. The clusters within the red rectangles with AU scores of >99% were observed after multiscale bootstrap (n= 10000). The edge # below the AU and BP values gives the edge count within the tree. The height indicates the similarity or dissimilarity between any two observations: the lower the height of the fusion between two observations, the more similar they are.

Clustering of the edgetic gain patterns identified two main groupings: one cluster comprised of BRCA, LUSC, LUAD, STES and BLCA, and another one containing KIRP, KIRC, KICH,

LIHC, THCA, HNSC, COAD and PRAD (Figure 18A). The proteins participating in the above edgetic perturbations present critical biomarkers shared across multiple cancer types. our findings are in agreement with Yuan et al.[187], who indicated that pan-cancer analyses reveal additional biomarkers that may be masked when searching for biomarkers in single tumor type studies. The first set consisted of edgetic perturbations affecting the melanoma-associated antigen 3 protein (*MAGEA3*) and the DNA repair and recombination protein RAD54-like (*RAD54L*) (Supplementary figure 5A). These perturbations were observed only in BRCA, LUAD, STES and BLCA. *Yamada et al*. have shown that deregulation of *MAGEA3* and other cancer testis antigens in BRCA and LUAD maybe a promising route for therapeutic targeting[188]. Our results further suggest that *MAGEA3* is similarly deregulated in STES, LUSC and BLCA and therapeutic targeting of this protein can also be extended to patients diagnosed with these three cancer types. Moreover, the periodic upregulation of *RAD54L* (a DNA repair protein) during the G1/S phase of the cell cycle has been shown to positively correlate with cancer proliferation by setting up feedback loops important in rapid cell multiplication processes[189].

While the TCGA consortium ranks *RAD54L* as one of the genes involved in the DNA repair pathway in some TCGA cancer types[190], our study further suggests that *RAD54L* deregulation may be an indicator of S phase expression in BRCA, LUSC, LUAD, STES and BLCA, therefore implicating *RAD54L* in the proliferation of the above cancer types . The other set of the predicted informative perturbations affected the histone-H3 like centromeric protein A (*CENPA)*, kinesin-like protein KIF14 *(KIF14)*, RPGR-interacting protein 1 (*RPGRIP1)*, and deoxyribonuclease-2-beta (*DNASE2B)* protein and were observed in KICH, KIRP, KIRC, LIHC, PRAD, STES and THCA (Supplementary figure 5B). While *CENPA* is an epigenetic marker in multiple cancer types indicating how aggressive the cancer type is , the role of *RPGRIP1* in cancer is not yet clear[193,194]. As it is a player in ciliopathy and proteasome deregulation, our results suggest that additional research should be undertaken to establish the oncogenic or tumorigenic role of *RPGRIP1* in cancer. Additionally, deregulation of *KIF14* and *DNASE2B* via p27 signaling has been observed in multiple cancer types[195,196] and may offer an opportunity for therapeutic targeting since p27 has been found to be prognostic of therapeutic response in cancer[197]. Our results, therefore, indicate that the proteins prone to gaining new interacting partners, while being relatively rare compared to proteins involved in edgetic losses, also have a role in cancer progression, may be important disease monitors and are possible candidates for therapeutic targeting.

Clustering of the edgetic loss patterns identified two main clusters: one cluster consisting of PRAD, STES, THCA, and BLCA and another cluster consisting of KIRP, KIRC, KICH, LUAD, LUSC, LIHC, COAD, BRCA and HNSC (Fig 7B). Here, we also found two sets of edgetic perturbation patterns important in distinguishing these cancer types. One set contained edgetic perturbations affecting the peripherin-2 protein (*PRPH*) together with the edges connecting the amyloid-beta precursor protein and the serine/threonine protein kinase NIM1 (*APP-NIM1K*), and the reelin protein with the very low-density lipoprotein receptor (*RELN-VLDLR*) (Supplementary figure 5C). These perturbations were frequently observed in PRAD, STES, THCA, and BLCA, suggesting a shared disease mechanism amongst these cancer types. Depletion of *APP* and *NIM1K* has been found to control the G1 or G2 phase of mitotic cells resulting in abnormal cell sizes[198,199]. The inhibition of proteins involved in the deregulation of G1/G2 checkpoint (e.g., *WEE1*) has been suggested to be a viable option for therapy development (e.g., the drug AZD1775/MK1775) against advanced malignancies[200]. First, our

results indicate that the patients having the above perturbations were at an advanced cancer stage, and secondly, AZD1775 may be viable in controlling the G1 or G2 phase of abnormal mitotic cells in patients diagnosed with advanced PRAD, STES, THCA, and BLCA. Wang *et al*. have previously shown that inactivation of alpha-internexins in gastroenteropancreatic neuroendocrine tumors (cancers affecting the pancreas, thyroid glands, gastrointestinal tract and partly the bladder) indicated poor prognosis of the patients[201]. In this study, we specifically found deregulation of *PRPH/peripherin* (an alpha-internexin) via edgetic loss perturbations, thus suggesting that it may be indicative of aggressive gastroenteropancreatic neuroendocrine tumors and consequently provide direction in therapy decision making as well as in disease monitoring. The second set contained high scoring edgetic perturbations affecting the neurotrophic tyrosine kinase receptor type 1 (*NTRK1*) and occurred in KIRP, KIRC, KICH, LUAD, LUSC, LIHC, COAD, BRCA and HNSC (Supplementary figure 5D). As already mentioned above, *NTRK1* is a target for the drug Entrectinib. Our study, therefore, implies that the drug Entrectinib is not only beneficial to non-small cell cancer types but may also be clinically relevant to KIRP, KIRC, KICH, LIHC, COAD, BRCA and HNSC.

Finally, to account for all the molecular pathways affected by the edgetic perturbations, we performed clustering based on both edgetic gains and losses that yielded 3 main groups consisting of (i) LIHC, KICH, KIRC, KIRP, (ii) PRAD, STES, BLCA, THCA and (iii) LUAD, LUSC, COAD, BRCA and HNSC (Figure 18C). Here, the random forest algorithm predicted 3 groups of perturbed edges as being highly discriminant of the cancer types. The first group of perturbed edges was observed in LIHC, KICH, KIRC, KIRP and affected the Wnt-7b protein (*WNT7B),* together with a number of edges – e.g., the edge between the homeobox protein Hox-B9 and the hepatocyte nuclear factor 3-alpha protein (*HOXB9-FOXA1)* (S5 Fig E). The perturbations affecting *WNT7B* involved edgetic gains in KICH, KIRP and LIHC and edgetic losses in BLCA, BRCA, COAD and STES. No perturbations involving *WNT7B* were found in HNSC, KIRC, LUAD, LUSC, PRAD and THCA. We suggest that *WNT* signaling may be enhanced in KICH, KIRP and LIHC while being depleted in BLCA, BRCA, COAD and STES tumor types. *WNT7B* participates in the deregulation of the beta catenin, c-Jun N-terminal and Ca^{2+} releasing pathways, and its increased expression is critical in cancer development[202]. For example, when up-regulated in BRCA, STES and some types of LIHC (cholangiocarcinoma), this abnormal expression correlates to poor prognosis and can be pharmacologically inhibited in mice[203–206]. Our results indicate that both up- and downregulation of *WNT7B* across cancer types may result in edgetic gains or losses at the protein-protein interaction network, and further suggest which human cancer types may be candidates for a *WNT7B*-based targeted therapy or disease monitoring (i.e, KICH, KIRP, LIHC, BLCA, BRCA, COAD and STES). We also found another grouping of multiple perturbed edges that were crucial in distinguishing the cancer types (Supplementary figure 5F). Of these, the edge between the phosphatase and tensin homolog protein and sialyltransferase 8F protein (*PTEN-ST8SIA6*) was predicted to have the highest score. This edgetic perturbation involved edgetic losses in BRCA, COAD, HNSC, KIRC, KIRP, LIHC, LUAD and LUSC, with an edgetic gain in PRAD, but no perturbations in THCA, BLCA and STES. Our study agrees with the current knowledge on the loss of the tumor suppressor *PTEN*, in multiple cancer types[57]. The loss of *PTEN* in cancer has been correlated to immunosuppression and reduced T cell trafficking in mice melanoma cells[207]. Our findings suggest that *PTEN* and P13-AKT pathway targeted immunotherapy may be beneficial in BRCA, COAD, HNSC, KIRC, KIRP, LIHC, LUAD and LUSC cancer types but probably not in THCA, BLCA and STES. Lastly, edgetic perturbations affecting the ciliary neurotrophic factor (*CNTF*), otoferlin (*OTOF*), and the inhibitor of CDK interacting with cyclin A1 (*INCA1*)

proteins together with the edge between cytokeratin-75 and cullin-3 (*KRT75-CUL3*) proteins were also predicted to be crucial in distinguishing the cancer types (Supplementary figure 5G). *CNTF* and *INCA1* perturbations were only observed in cluster 3 cancer types and involved edgetic losses in all the 5 cancer types in cluster 3. *OTOF* perturbations affected edgetic losses in COAD, HNSC, LUSC and BRCA, edgetic gains in BLCA, KIRP and THCA, but no perturbations in KIRC, LIHC, PRAD and STES. Our findings confirm that the above mentioned proteins are crucial in tumor progression as previously suggested[208–210] and pinpoint their important role in tumor progression in BRCA, LUAD, LUSC, COAD, BRCA and HNSC.

2.2.9 Protein nodes rewired across cancer types are involved in tumorigenesis

We used DyNet algorithm in Cytoscape to find significantly rewired proteins, that is, nodes recurrently affected by edgetic perturbations. We selected a node as considerably rewired if it had a DyNet score of at least 0.5 (see methodology section)[211]. In cancer, significantly rewired nodes were either perturbed across cancer types or were specific to a cancer type, with some being known cancer biomarkers (Table 6, Datasets 3 and 5). Nodes rewired across multiple cancers, for example, Q9HBJ0 (*PLAC1*), Q9BVV2 (*FNDC11*) and Q5T7N2 (*L1TD1*) have been suggested to influence the growth of tumor cells in various cancer types[212–213]. Furthermore, to better understand the association between the significantly rewired nodes and cancer, we used DisGeNET[214] in clusterProfiler to search for any gene-disease relationships associated with the rewired nodes. We found that most of these proteins were significantly ($p < 0.05$) associated with diseases observed during the onset of cancer (e.g bronchial and lung dysplasia) as well as the development of multiple cancer types (Table 7).

Table 6: Top 10 significantly rewired nodes per cancer type across 13 cancer types.

THCA	Q9BVV2**, Q99666, Q8N8A2, P61371, P50549, P50222, Q9Y2A9, Q92608, Q8NCP5, Q5T2W1
BRCA	Q9NQL9, Q9HBJ0**, Q86YZ3, Q76N89, P78337, P48740, P19544, Q96FW1, P04629, P12882
BLCA	Q9HBJ0**, Q96RI1, Q96FV0, Q5VYV7, P61371, P09017, P02675, Q9BVV2**, Q8NEC5
KIRC	P09651, Q12906, P35222, O60341, P46379, P12931, Q13501, Q9BXN2, Q9BVV2**, Q8TBC3
KIRP	Q92905, Q96Q40, Q96AY2, Q86SE5, P35548, O95156, P00533, Q12933, Q9NQA5, Q8WXK1
LUSC	Q13547, P78362, Q9HBJ0**, Q96Q40, Q81YR6, Q5T7N2**, P19544, P12882, Q7Z3S9, Q96FW1
LIHC	Q96FW1, Q9HBJ0**, Q96AY2, Q86SE5, P54289, P18509, P49736, Q9Y577, Q8TAK6, Q8NEC5
PRAD	Q96Q40, Q5T7N2**, Q9NZM1, Q9ULJ8, Q9BXA6, P62913, P04637, Q69YH5, Q12988, P48668
STES	Q13616, Q96PN8, Q86YZ3, Q86W54, Q5VU13, Q16769, P50222, Q9NRD1, Q9NZM1, Q9H4K1
COAD	P38398, Q9HAT0, Q96PN8, Q86SE5, Q13224, O95661, A6NK59, Q93034, P22681, Q6UXL0
HNSC	Q9UQB9, Q9HBJ0**, Q8WXK1, Q81YR6, Q5T7N2**, Q5T2W1, P19544, P05814, Q13618, Q9ULJ8
LUAD	Q99666, Q76N89, Q5T7N2**, P16520, P04629, Q93034, Q9Y2A9, Q9BVV2**, Q8NB78, Q9BX46
KICH	P02751, Q9Y2A9, Q9BVV2**, Q99666, Q86SE5, Q76N89, Q5T7N2**, P50549, P11245, P06732

** proteins involved in edgetic rewiring in at least 5 cancer types

Table 7: Proteins involved in edgetic rewiring are associated with disease

Disease ID	Disease Description	Gene Ratio	BgRatio	p.adjust	geneID	Gene Count
umls:C0346163	Endometrioid carcinoma ovary	4/96	16/17381	0.003	1956/7157/1499/4488	4
umls:C1623038	Cirrhosis	12/96	389/17381	0.003	1788/10401/1956/2335/7157/2784/1499/4488/4171/8988/2904/10987	12
umls:C1112356	Bronchial dysplasia	3/96	6/17381	0.006	1956/7157/4171	3
umls:C1334708	Metaplastic breast carcinoma	3/96	6/17381	0.006	1956/7157/672	3
umls:C1336084	Squamous Lung Dysplasia	3/96	6/17381	0.006	1956/7157/4171	3
umls:C0684337	Ewings sarcoma-primitive neuroectodermal tumor	6/96	76/17381	0.008	1956/4914/7157/7490/867/1499	6
umls:C1176475	Ductal Carcinoma	9/96	223/17381	0.008	8202/1956/2335/7157/6714/672/388697/9971/6795	9
umls:C0153579	Malignant neoplasm of fallopian tube	3/96	7/17381	0.01	1956/7157/672	3
umls:C0007133	Carcinoma, Papillary	9/96	233/17381	0.01	1956/2335/4914/7157/10/2784/672/3065/10987	9
umls:C0206629	Pulmonary Blastoma	3/96	8/17381	0.018	1956/7157/1499	3
umls:C0035335	Retinoblastoma	12/96	472/17381	0.023	1788/1956/4914/7157/6714/116/7490/1499/672/4171/3065/10987	12
umls:C0851135	In situ cancer	3/96	10/17381	0.03	1956/7157/8988	3

The GeneRatio represents the number of genes from the input (query) gene list that match the GO term / Total number of the input (query) genes.
The BgRatio represents the number of genes in the DisGeNET database associated with a disease-gene ID/ number of genes in the DisGeNET database.
p.adjust: Benjamini Hochberg adjusted p-value

2.2.10 Proteins participating in significant edgetic perturbations are implicated across all cancer stages

For a mutated gene to be tumorigenic (*i.e.* to be a driver gene), it must accumulate mutations throughout the life of the cancer cell[63–32]. Consequently, cancer driver genes are implicated from the onset of cancer and progressively increase the survival of the cancer cell as the disease progresses. We hypothesised that edgetic perturbations that harbour essential biomarkers may play an important role in tumorigenesis and cut across all cancer stages. To determine if this phenomenon applied to edgetic perturbations, we searched for the stage distribution of the patients that had significantly perturbed edges in their PPINs. In all cancer types, the significantly perturbed edges were observed across all stages albeit in varying proportions, indicating their probable role from cancer onset and in progression (Table 8). Our results mirror those from Li [215] who pointed out that essential cancer biomarkers are active in the entire life of a cancer cell.

Table 8A: Distribution of patient samples harbouring the top lost edges across cancer stages.

Cancer type	Sample size[(a)]	Number of samples grouped in Stage I	Number of samples grouped in Stage II	Number of samples grouped in Stage III	Number of samples grouped in Stage IV
THCA	59	25	6	9	3
BLCA	19	0	4	7	8
BRCA	110	18	60	21	2
KIRC	72	25	11	16	20
KICH	25	10	8	3	4
LUAD	58	26	9	12	2
LUSC	51	24	15	5	1
LIHC	40	15	7	8	0
COAD	26	3	13	4	5
KIRP	32	14	1	12	4
STES	43	5	16	7	3
HNSC	43	2	16	7	19

[(a)]number of patient samples having significantly perturbed edges.

Table 8B: Distribution of patient samples harbouring the top lost edges across cancer stages.

Cancer type	Sample size[a]	Number of samples grouped in Stage I	Number of samples grouped in Stage II	Number of samples grouped in stage III	Number of samples grouped in Stage IV
THCA	59	25	6	9	3
BLCA	19	0	4	7	8
BRCA	110	18	60	21	2
KIRC	72	25	11	16	20
KICH	25	10	8	3	4
LUAD	49	26	9	12	2
LUSC	46	24	15	5	1
LIHC	38	14	9	9	1
COAD	26	3	13	4	5
KIRP	32	14	1	12	4
STES	43	5	16	7	3
HNSC	43	2	16	7	19

[a]number of samples having the significantly perturbed edges

2.2.11 The EdgeExplorer website

The EdgeExplorer portal (http://webclu.bio.wzw.tum.de/EdgeExplorer) provides annotations for all the cancer-type specific proteins involved in edgetic perturbations (a total of 539 proteins). We annotated each protein by performing an exhaustive literature search for relevant experimental evidence linking it to the specific cancer type; if hits related to that cancer type were not found, we broadened the search to include other cancer types. The main advantage of the web portal is that it allows for easy browsing of the results and searching for information on specific proteins. Moreover, it provides the functionality to download all of the annotated data.

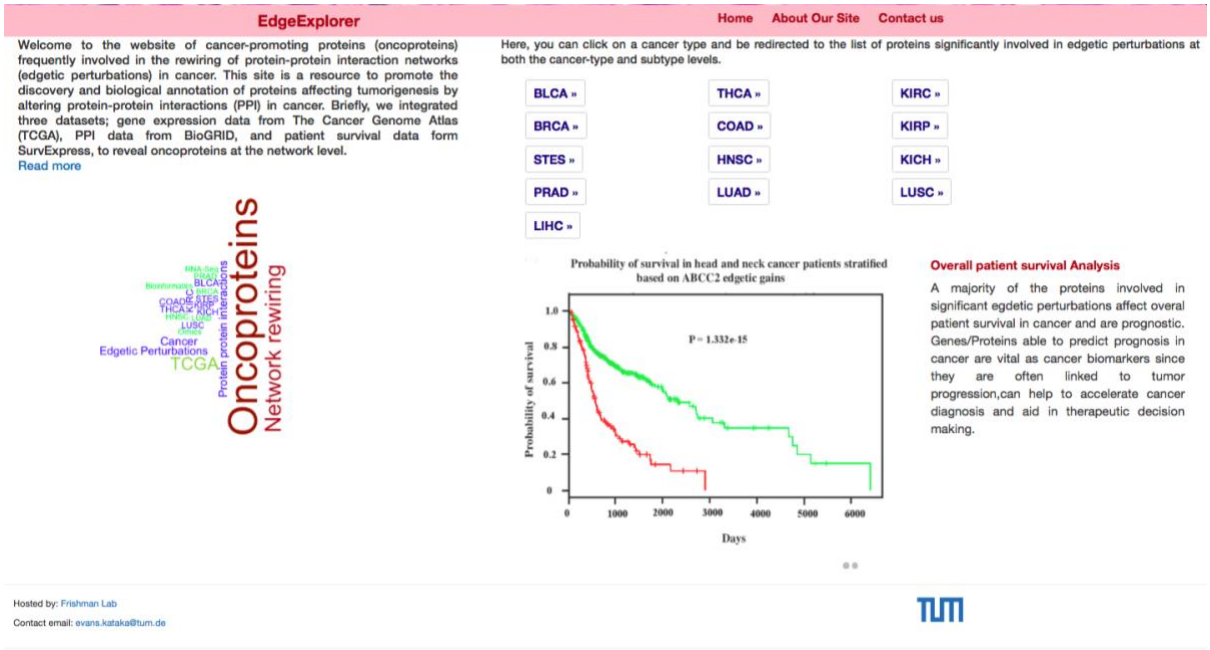


Figure 19. A screenshot of the EdgeExplorer portal homepage. The EdgeExplorer portal provides a resource to the scientific community to easily query proteins of interest to find out if they are involved in edgetic perturbations in 13 different cancer types.

2.3 DISCUSSION

Even with improving knowledge on cancer oncogenesis as well as the development of new cancer therapies aided by translation of experimental results from multiple omic data to clinical use, cancer remains a leading cause of death worldwide. However, rigorous analysis of the incomplete human PPIN can reveal essential biomarkers driving diseases such as cancer. Cancer biomarkers are crucial biomolecules because of their use in early disease detection, disease progression monitoring and advising treatment regimens, especially in personalized therapy. Identification of cellular interconnections perturbed by diseases has long been recognized as a promising avenue towards elucidating reliable biomarkers[69]. Over the recent years this general idea was being actively put into practice by using molecular networks to study the differences between healthy and diseased states in cancer[87, 81, 216]. Here, we derived 642 patient-specific PPINs from patient-specific paired healthy and cancer mRNA expression profiles and identified candidate biomarkers significantly involved in distorting PPINs during tumorigenesis. In doing so we considered shared patient edgetic perturbation profiles across tumors, within a cancer type and further distinguished edgetic perturbation signatures between cancer subtypes.

Our approach utilizes the publicly available data of paired cancer and healthy gene expression profiles from 13 cancer types and combines them with previously reported cancer-specific significantly mutated genes, and binary protein interaction data to identify proteins driving significant edgetic perturbations in cancer networks. For the first time, we show that using multiple patient-specific PPINs derived from the corresponding mRNA expression profiles of healthy and cancer patient samples is a novel way of identifying patient-, cancer-type and subtype as well as multi-cancer edges susceptible to perturbation during tumorigenesis.

Furthermore, we were able to reproduce similar perturbed edges for each cancer type when using a smaller protein abundance-filtered PPIN (Dataset 3), validating our approach. We demonstrate that perturbed edges harbor known and novel cancer biomarkers and that they also capture previously reported cancer hallmarks[28,217]. While the differential expression of genes between the cancer and healthy state dictates the availability of proteins that interact with each other, we also show that alternative splicing events causing protein domain composition changes in the cancer state have effects on the protein-protein interaction network. A gain of an interacting domain may result in the gain of a new interaction while the loss of an interacting domain may bring about edgetic losses in the cancer PPIN - Figure 3.

We found that the majority of perturbations were not attributed to SMGs either directly as first neighbors or indirectly as second neighbors within the PPIN - there were only several cancer types that did not follow this trend (Supplementary Table I). One such exception was LIHC, and indeed the interactions disrupting mutations of SMGs in this cancer type have been recently reported to strongly affect survival, which indicates that our results are in agreement with previous findings from Cui et al.[102]. While our study cannot model the effects of mutations on the PPIN as undertaken by Cui et al., our results suggest that any mutations that are prevalent in the domains do promote edgetic perturbations and consequently tumorigenesis.
We also found that cancers exhibit either a high proportion of edgetic losses or a high proportion of edgetic gains. We speculate that this may be a downstream effect of a deregulation of the components of the spliceosome resulting in a systematic truncation or elongation of transcripts during pre-mRNA processing.

However, most cancer types (9) showed more edgetic losses than edgetic gains, resulting in a reduction in the size of the cancer PPIN when compared to the corresponding healthy PPIN

(Figure 1). We also found multiple biomarkers already validated at multi-cancer, cancer type and subtype levels. When considering proteins driving significantly perturbed edges and using SurvExpress, our study confirmed most of the proteins as being biomarkers predictive of survival while those that did not show any perturbation were not prognostic in cancer. Moreover, at the multi-cancer level, known cancer drivers such as *CDC45* and *NUF2* were identified to be involved in edgetic gains while *NTRK1*, *PRPH* and *MYOC* were determined to be involved in edgetic losses and may serve as targets for widely applicable therapeutic interventions. For instance, *Liu et al.* showed that knockdown of *NUF2* may inhibit proliferation of carcinomas and may be a potential target for therapy in cancer[178]. Furthermore, our clustering analysis of the cancer perturbation profiles revealed novel relationships between cancer types.

We found that KICH, KIRP, and KIRP, LUAD and LUSC, COAD, HNSC and BRCA, as well as THCA, BLCA, and STES shared a more significant proportion of lost edges. Also, BRCA, BLCA and STES, LUAD and LUSC shared a higher portion of gained edges. Targeting of the proteins shared and perturbed in these cancer types for clinical use could benefit patients diagnosed with these cancer types. For example, developing therapy to target *UCHL1*, the protein most rewired across kidney cancers, would be an economical way of treating all kidney cancers by targeting the same molecule[218].
At the cancer type level, some of the perturbations we identified, such as *IGF2BP3* and *DKK1-MDFI*, have already been suggested to be KIRP biomarkers. Our analysis supports the roles of these molecules as KIRP biomarkers, as they were among proteins significantly perturbed in KIRP and showed prognostic value when their expression changes were analyzed for predicting overall patient survival. We also uncovered known biomarkers for specific cancer types not yet directly linked to other cancer types. For example, *TRIM15* is a tumor suppressor in colon cancers[219] , however, to our knowledge no study has linked *TRIM15* to KICH tumorigenesis. We found *TRIM15* perturbations among the proteins involved in edgetic losses in KICH.

Our study, therefore, suggests that *TRIM15* could also be an informative KICH biomarker. We also found multiple cancer-specific edgetic perturbation biomarkers such as the *SLC25A21* distortion in LUAD. Most importantly in KICH, our study is also able find perturbations of Bcl2 family proteins which are targeted by the only clinically approved drug (Venetoclax) targeting a protein-protein interaction[34]. While previous studies such as Li et al.[87] found biomarkers at the cancer network level (lung cancer), our study expands on this work to obtain cancer subtype-specific markers at the network level. Using our methodology, we identified probable subtype-specific biomarkers, including *PARVG* and *XPO4* in PR+ BRCA, *MYL1* in PRAD, *KHDRBS1-DLG2* edgetic perturbation in HNSC, and *KLF8* in STES. We also observed several cancer subtypes sharing perturbed proteins pointing to probable shared oncogenic patterns. Our findings, therefore, suggest that these subtypes could be targeted by similar therapies.

Functional and pathway enrichment analysis further revealed that proteins driving edgetic perturbations are consistent with the observed cancer phenotype, that is, we obtained known canonical oncogenic KEGG pathways involved in viral carcinogenesis, chemical carcinogenesis, *EGFR* tyrosine kinase inhibitor resistance, FoxO signalling, proteoglycans in cancer and transcriptional deregulation in cancer. Our analyses show that the diverse proteins participating in edgetic perturbations in cancer are essential biomolecules in tumorigenesis, that could be used for monitoring disease progression and developing new therapies. This integrated analysis is the first to utilize patient-specific PPIN derived from corresponding

paired cancer and healthy mRNA expression profiles to decipher essential interactions distorted at the multi-cancer, cancer type, and subtype levels. Our findings present an integrated multi-omics approach for the computational identification of multi-cancer, cancer type and subtype-specific biomarkers with potential clinical prognostic relevance. As OMICS data become more complete, our methodology will be of increasing help in determining the full extent of protein network distortion across cancer types.

2.4 Conclusion

In summary, our study presents a novel and robust scheme capable of identifying known and novel cancer-specific and multi-cancer biomarkers using patient-specific PPIN derived from mRNA expression data. Furthermore, the ability to determine uniquely distorted interactions whose participants are predictive of patient survival opens up the possibility to computationally obtain potential protein biomarkers for specific cancer types and subtypes. We also established that SMGs do not bring about the majority of perturbations in cancer PPINs. Additionally, we found probable novel biomarkers such as the THCA BRAF-like specific 4-gene signature biomarker (*ODAM, APP, IKBKG,* and *TOLLIP*). The THCA biomarkers may be essential for disease monitoring of THCA subtypes whereas the 14-gene signature (with *HRK* node perturbation) explicitly observed in KICH samples is a candidate for therapeutic targeting. Survival and functional enrichment analysis revealed that our candidate biomarkers are indeed involved in tumorigenesis. Our user-friendly portal will not only facilitate experimental research in the continued quest for druggable proteins at the protein-protein interaction network level but will also be essential for researchers to quickly mine and access the proteins involved in edgetic perturbations of cancer PPINs. We envisage that subsequent experimental validation will demonstrate the applicability of the novel biomarkers generated in this study for making informed clinical decisions as well as in developing cancer therapies. In the future, we will investigate patient-specific edgetic perturbations and determine proteins and corresponding isoforms (and protein domains) responsible for such disruptions.

2.5 MATERIALS AND METHODS

2.5.1 Cancer datasets

We obtained RSEM[220] quantified count data for healthy (non-cancer) as well as the corresponding cancer patient-specific mRNA expression profiles from the Broad Institute Web site (http://gdac.broadinstitute.org/). We further selected datasets with at least 10 paired healthy and cancer samples, covering 13 cancer types (Table 9). The corresponding cancer stage-specific annotated clinical data and subtype annotations were downloaded using the TCGABiolinks R package[221]. Stomach and esophageal carcinoma subtypes were downloaded from the supplementary materials of the TCGA consortium paper for STES[222] because the Broad Institute Web site did not include all the paired samples. The clinical dataset consisted of patient samples grouped according to stages I, II, III and IV. TCGA clinical files contain important cancer phenotype information, including patient treatment regimen, cancer staging, alive/dead status, tumor state, and age (Table 9).

Table 9: Clinical and phenotypic traits of the 639 patients diagnosed with 13 cancer types as obtained from TCGA.

Cancer type[a]	Sample size[b]	Subtypes	SMGs	StageI	StageII	StageIII	StageIV	Status (D/A)[c]
BLCA	19	NA	684	0	4	7	8	7 /12
BRCA	98	3	640	16	60	21	1	24 /74
COAD	41	2	2580	4	22	7	7	7 /34
HNSC	43	4	518	2	16	7	19	31 /11
KICH	25	2	335	10	8	3	4	4 /21
KIRC	72	4	433	25	11	16	20	25 /47
KIRP	32	2	344	14	1	12	4	6/26
LIHC	50	NA	797	18	11	12	1	30 /20
LUAD	57	3	901	28	13	13	2	22 /37
LUSC	51	4	496	27	17	6	1	24 /27
THCA	57	2	508	35	7	11	4	4 /53
PRAD	51	6	658	NA	NA	NA	NA	0 /51
STES	43	2	2065	9	21	9	3	5 /38

[b)] number of patients with paired healthy and cancer RNA-sequence data

[c)] D-Dead, A-Alive

2.5.2 Global protein-protein interaction network (PPIN)

We obtained information on 330,557 binary interactions between human proteins from BioGRID[223] and selected only those interactions whose individual interacting partners have a "reviewed" status in UniProt[224]. The resulting global network consisted of 224,223 human binary protein interactions between the total of 15,689 proteins. Based on the assumption that

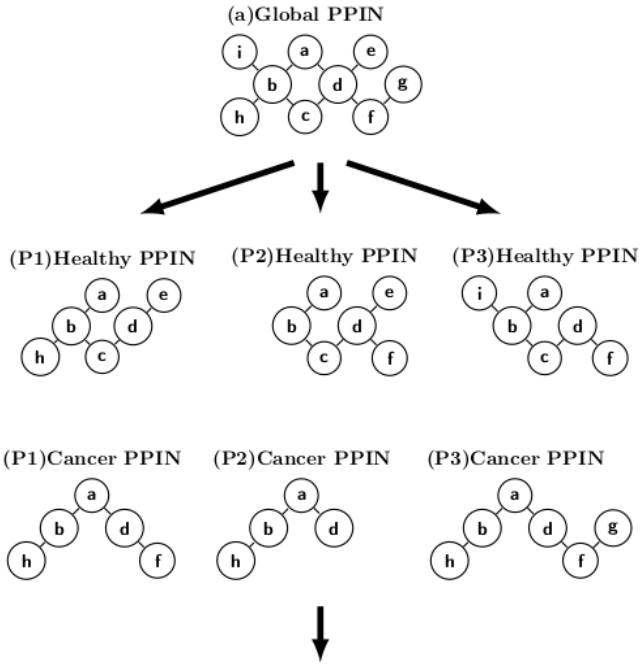
two proteins can only interact if proven to be translated, we further filtered the human interactome using protein abundance data. Whole-proteome high-confidence abundance data were obtained by combining information from PaxDb[225] and The Human Proteome map . Upon retaining only the proteins reported as translated (having non-zero abundance values) in both proteomics datasets our final PPIN consisted of 216,134 binary interactions involving 15,125 proteins. Hereafter, the total number of binary interactions in a PPIN is referred to as PPIN size.

2.5.3 Patient- and cancer-specific protein interaction networks

Patient and cancer-specific PPIN were derived from gene expression data by PPIXpress using both PPINs described above. PPIXpress adapts PPINs to specific cellular conditions at the isoform level, thus enabling identification of tumor-related alterations missed by gene-level analysis. For each tissue type, we filtered the RNA-seq data to only include the genes that were consistently expressed across most samples using the EstimateExpression function of the xseq R package[33]. The function fits a mixture-of-Gaussian distributions model on the gene expression count data to distinguish between lowly expressed genes (presumed to be transcriptional noise) and biologically relevant gene expression (Supplementary Figure 6). Furthermore, for an isoform of a selected gene to be considered as expressed, its RSEM value was required to be 0.1 or higher. If multiple isoforms of a gene are expressed, the mean expression value of all isoforms is selected (running PPIXpress with '–g' option).

2.5.4 Patient-, cancer-, subtype-specific and multi-cancer perturbed edges

For each cancer PPIN we retrieved interactions that were not present in the paired healthy PPIN (gained edges). Likewise, in each healthy PPIN we identified interactions that were absent in the corresponding cancer PPIN (lost edges). Edges occurring in both healthy and cancer PPINs were considered non-perturbed. For brevity, lost, gained, and non-perturbed edges were assigned the codes 10, 01, and 11, respectively. For each patient, the set of all perturbed and non-perturbed edges represents their individual network perturbation profile. To obtain cancer type perturbation profiles we merged perturbation profiles of patients diagnosed with a specific type of cancer (Figure 20).



Edgetic perturbation patterns in cancer

Edge	Patient1	Patient2	Patient3	Perturbation status
a-b	11	11	11	Non perturbed
b-c, d-e, c-d	10	10	10	Strict losses
a-d, b-h	00	01	01	Strict gains
d-f	01	10	11	Partly lost/gained
f-g	00	00	01	Patient specific gain
b-c-d	10	10	10	Node (c) with multiple perturbations
d-c-b	10	10	10	SMG (d) probably responsible for d-c and c-b perturbations

Figure. 20. Edgetic perturbations in cancer. Assuming a global PPIN with 9 edges interconnecting 9 nodes and using cancer and healthy patient-specific mRNA expression profiles, for each patient (P1, P2 and P3) perturbed edges in cancer can be identified by comparing the healthy and the corresponding cancer PPIN. Significantly Mutated Genes (SMGs) may be involved in perturbation of edges directly interacting with them, or those interacting with their perturbed neighbors (secondary neighbors).

Edges that were not observed in one sample but observed in other samples were assigned the code 00 in the samples where they were absent, and either 01 or 10 where they were perturbed. On a cancer PPIN (Figure 20), an edge can be gained across all patients (strict gains, edges a-d and b-h), lost across all patients (strict losses, edges b-c and d-e), partly gained or partly lost across patients (d-f), non-perturbed in all patients (a-b), or not observed in one patient but observed in others (f-g). The list of all the perturbed and non-perturbed edges in a single patient constitutes their perturbation profile. The union of all patient profiles diagnosed with a particular cancer type is referred to as a cancer perturbation profile. A cancer type perturbation profile is a list of lost, gained, and non-perturbed edges in all patients with a particular cancer type with their associated codes, as described above. For each cancer type *i*, each edge *j* was ranked depending on the percentage of samples it was gained ($PercGained_{i,j}$) and lost ($PercLost_{i,j}$) in (Dataset 3).

In a similar fashion, for each cancer type we merged all edges lost or gained in any of the samples and retrieved only those edges that are perturbed at least once in this specific cancer type and are observed in at least two samples (for example edges a-d and b-h (gained edges) and edges d-e and c-d (lost edges) in Figure 20). Note that edges perturbed in a cancer type but observed only in a single patient sample are considered patient-specific. For each cancer with a subtype *s*, we also searched for perturbations unique only to a particular subtype and ranked each perturbed edge *j* depending on the percentage of samples it was gained ($SubtypePercGained_{s,j}$) and lost ($SubtypePercLost_{s,j}$) in, with $j \geq 2$. Note that a cancer subtype perturbation profile is a subset of the corresponding cancer type perturbation profiles involving the patients diagnosed with this particular subtype.

To identify perturbations occurring across multiple cancers, we first merged all edges perturbed in each cancer type and then identified only those perturbations observed in at least two cancer types. If, instead of the three patients P1, P2, and P3, the perturbation patterns in Figure 20 corresponded to three different cancer types C1, C2, and C3, edges a-d and b-h would represent multi-cancer gained edges while edges d-e and c-d would represent multi-cancer lost edges since they are perturbed in more than two cancer types. Also, for each cancer type *i*, each perturbed edge *j* was ranked depending on the percentage of cancer types it was gained ($MultiCanGained_{ij}$) or lost ($MultiCanLost_{ij}$) in, with $i \geq 2$. We then ranked these multi-cancer perturbations based on the number of cancer types exhibiting them.

Finally, we sought to find prominent proteins frequently involved in edgetic perturbations as well as frequently perturbed edges at the multi-cancer, cancer type, and cancer subtype levels. At the cancer type and cancer subtype levels, proteins were ranked according to the number of perturbations they and their first network neighbors are involved in. Note that the perturbations associated with the second neighbors of a protein were counted if (i) the protein had at least two interacting partners and that the (ii) protein itself was associated with a perturbation. For instance, perturbation of edges b-c, c-d and d-e (Figure 20) would give a rank of 3 for node c, and a rank of 1 each for nodes b, d and e.

2.5.5 Identification of PPIN nodes associated with perturbations.

We next searched for network nodes involved in edgetic perturbations. For each cancer type we merged all observed edges in the cancer and the corresponding healthy PPIN and then used DyNet[211], a Cytoscape[227] plugin, to identify the nodes associated with gained or lost edges. DyNet compares the nodes and edges present in two networks and then computes a rewiring metric score to determine which nodes have been rewired. To consider a node as rewired, we used a DyNet rewiring score of ≥ 0.5 and an edge count of ≥ 2, which corresponds to selecting the nodes with at least a degree (number of interaction partners) of 2 and showing perturbation of at least one edge. A single edge perturbation on a 2-degree node (2 interacting partners) means 50% of the edges are perturbed and thus have a rewiring score of 0.5.

2.5.6 Clustering of cancers based on edgetic perturbation signatures

To understand the relationship between cancers in terms of their perturbation patterns, we used unsupervised clustering as implemented in the Pvclust[228] R package to group cancers based on shared perturbations. Pvclust allows assessing the uncertainty of hierarchical clustering by performing multiscale bootstrap resampling and assigning p-values (as percentages) to clusters depending on how strong the cluster is supported by data. Pvclust provides two p-values: an approximately unbiased p-value (AU) computed from multiscale bootstrap resampling and a bootstrap probability (BP) computed by regular bootstrap resampling. High percentage values

indicate a strong relationship of the cluster and the data, which may be biologically relevant. In our study, a cluster with an AU p-value >0.95 (95%) or the significance level <0.05 was selected and kept for further analysis. In order to identify the edges defining the cluster, we searched for the edges perturbed across all cancer types within each cluster.

2.5.7 Ranking of perturbed edges in terms of their importance in classifying cancer types

For the multi-cancer perturbations, after performing hierarchical clustering of the cancer types using the perturbed edges as features, we identified the edges appearing in only one cluster. Next, we used the random forest algorithm[229] to identify the features (perturbed edges), which were most informative for attributing cancer types to the detected clusters based on shared edgetic perturbations. The Pvclust algorithm is essential in accurately identifying high confidence groups in data, and thus we did not face the problem of retraining our random forest algorithm to accurately classify cancer types sharing the majority of perturbed edges together. Our interest here was only to use the random forest algorithm (using the VarImp function from the R package caret[230]) to rank the features based on their importance in classifying cancer types into the groups detected during clustering. The VarImp function outputs feature ranking based on their mean squared error (MSE). Features with high MSE scores were then chosen to be the perturbed edges (features) having the highest weight in grouping the cancer types into the categories identified during clustering.

2.5.8 Identification of Gene Ontology, KEGG pathways and disease-gene relations significantly enriched by proteins driving edgetic perturbations

To understand the biological relevance of the perturbed edges we identified statistically enriched Gene ontology (GO) terms and KEGG pathways associated with the proteins involved in edgetic perturbations. GO analysis was carried out using the R package topGO[231] with statistical significance calculated using Fisher's exact test. GO terms having a p-value of <0.05 were chosen to be considerably enhanced. Additionally, significant GO terms were clustered using REVIGO[232] to remove redundancy. Furthermore, a dispensability value (representing both the degree of redundancy and enrichment of a GO term) of <0.05 was considered significant after the REVIGO pruning step. To avoid statistical bias[233] in the enrichment analysis of the proteins involved in edgetic losses we used all the genes expressed in cancer as the background for comparison. On the other hand, to analyze the GO terms and KEGG pathways enriched among the proteins involved in edgetic gains, we used all the genes expressed in the healthy (non-tumor) condition as the background for comparison. Disease-gene relation analysis was performed using DisGeNET[214] implemented within the R package clusterProfiler[234]. KEGG pathway analysis was carried out using DAVID[235].

2.5.9 Predicting overall patient survival in cancer

Disease genes often work in concert and several studies have discovered network modules and hubs under attack in cancer[95,236,237]. To understand the importance of the proteins driving perturbations in cancer, we used SurvExpress[238] to determine multi-gene cancer signatures and to assess their prognostic value for cancer. SurvExpress is a multi-gene cancer biomarker validation and discovery tool based on a wide collection of cancer datasets, including TCGA. From the ranked lists of perturbed edges in each cancer type, we selected each edge or a group of edges lost or gained across the largest number of patients as candidate biomarkers and analyzed them in SurvExpress.

2.5.10 Implementation of the EdgeExplorer website.

The EdgeExplorer website resides on a Linux server that provides Apache 2 for web services, SQLite for relational database management, and the PHP for server-side scripting services on the backend. The portal application further utilizes additional web technologies, among them: JavaScript, CSS, and jQuery.

2.5.11 EdgeExplorer web portal annotations.

To identify gene-disease relations with experimental evidence, we did manual annotation by searching for publications implicating anomalies of each query gene in affecting cancer progression or treatment outcomes. The following annotation rules were followed:

1. Search for gene – cancer-type associations in recent experimental papers indexed by PubMed, PMC and Google Scholar while using all synonyms of a gene name. If no full text of the articles were found, we further searched in the Bavarian State Library database.
2. Priority was first given to journal papers with experimental evidence demonstrating the association of the exact gene with the particular cancer type in which edgetic perturbations were detected. The associations included: gene mutation or differential expression of a gene in tumor samples when compared to normal samples, or gene involvement in metastasis, patient survival, prognosis, therapy resistance or therapy success. Results reporting findings ased on TCGA data were excluded from the annotations unless no other hits were found for that gene (see 4).
3. If there was no such information, the second priority was given to papers with experimental evidence demonstrating the association between the specific gene and a cancer type occurring in the same somatic tissue. For example, if there was no information on KIRP but there was for KIRC, we report that association.
4. Finally, if there was no such information, as a third priority, we checked for gene-disease association in The Cancer Genome Atlas to show if indeed our study yields similar results to other studies that have previously used TCGA data.

2.5.12 Generation of genes having similar node degrees as SMGs and their associated perturbations

To comprehend whether SMGs were pivotal in edgetic perturbations, we compared the proportions of perturbations involving SMGs and those from randomly generated genes with a similar degree of interacting proteins. First, we downloaded lists of pan-cancer and cancer specific significantly mutated cancer genes from the COSMIC Cancer Gene Census (https://cancer.sanger.ac.uk/census)[58] and from the TCGA consortium (https://cancergenome.nih.gov/publications)[67].

The genes amounted to 719 and 299 cancer genes from COSMIC and Bailey *et. al*, respectively, and were classified according to their significance as cancer specific or as pan-cancer. We considered a gene to be significantly mutated in a certain cancer type if it was characterized as either cancer specific or pan-cancer, but affected that cancer type. Finally, in each cancer type, a union of the significantly mutated genes from both the above sources were considered as cancer specific significantly mutated genes (Dataset 3). To find the number of perturbations involving the SMGs, we searched for any perturbed interactions having an SMG as an interacting partner. Then, in each cancer type, we merged all the interactions observed in both cancer and healthy in all the patients to generate all possible interactions within a cancer type. Next, for each cancer type, we determined the degree of each of the proteins within all the possible interactions of a cancer type. We used the degree of each SMG involved in any

perturbation to randomly query for other proteins having a similar number of interacting partners to them (Dataset 3). Finally, we determined the number of perturbations associated with the genes having a similar degree to the SMGs.

For each cancer type, the Z-test of proportions[239] was used to estimate the statistical significance of the extent of edgetic perturbations associated with cancer-specific SMGs compared to the extent of edgetic perturbations associated with genes having a similar network topology to the SMGs. To do this, we first determined if there were significant differences in the proportion of perturbations associated with SMGs and the proportion of perturbations associated with randomly generated genes. Then, for each significant difference, we sought to find only the cancer types where the proportion of perturbations associated with SMGs were significantly larger than those associated with the randomly generated genes.

2.5.13 Randomization of the PPIN

To check whether our results were brought about by changes in differential gene expression or were due to domain changes between the healthy and cancer state, we used the R package BiRewire first to generate a randomized network and then analyzed the resulting perturbations. We did this by building condition-specific PPINs in three randomly selected cancer types (BRCA, THCA and BLCA). BiRewire has the advantage of rewiring PPINs while preserving their functional connectivity and keeping the node degrees intact[240].

2.5.14 Statistical analyses

All statistical analyses were carried out in the in Python or the R environment. The Wilcoxon singed-rank test[241] was used to determine if the mean of a healthy and the corresponding cancer PPIN sizes differed. For each cancer type, the chi-squared test[242] was used to estimate the statistical significance of the extent of edgetic perturbations associated with cancer-specific significantly mutated genes (SMGs) compared to the extent of edgetic perturbations associated with genes having a similar network topology to the SMGs. Unsupervised hierarchical clustering using the Ward.D2 method and Euclidian distance[243] was used to group cancer types. Patient stratification using Kaplan-Meier curves and log-rank test p-values for survival analysis were calculated using SurvExpress. For all analyses, p-values < 0.05 were considered significant.

CHAPTER 3: AN OPTIMIZED QUANTITATIVE PROTEOMICS METHOD ESTABLISHES THE CELL TYPE-RESOLVED MOUSE BRAIN SECRETOME.

Parts of this chapter have been published in the EMBO Journal: Johanna Tüshaus, Stephan A. Müller, **Evans Sioma Kataka**, Jan Zaucha, Laura Sebastian Monasor, Minhui Su, Gökhan Güner, Georg Jocher, Sabina Tahirovic, Dmitrij Frishman, Mikael Simons, Stefan F. Lichtenthaler. An optimized quantitative proteomics method establishes the cell type-resolved mouse brain secretome. *The EMBO Journal*, e105693 (2020).

Stephan Lichtenthaler, Stephan A. Müller and Johanna Tüshaus conceived the project, perfumed the laboratory experiments. Dmitrij Frishman, Jan Zaucha and I designed and implemented the bioinformatics analyses. All authors wrote and reviewed the paper.

Abstract

To understand intercellular communication, it is essential to define the cellular secretome; a collection of proteins including soluble secreted, unconventionally secreted and proteolytically-shed proteins. Quantitative methodologies to decipher the secretome are challenging, because of large cell numbers required and abundant serum proteins interfering with the detection of low-abundant cellular secretome proteins. Here, we miniaturized secretome analysis by developing the high performance secretome-protein-enrichment-with-click-sugars method (hiSPECS), which identifies the glyco-secretome. We applied this method to provide a cell type-resolved mouse brain glyco-secretome resource. Our data show that a surprisingly high number of secreted proteins are generated by ectodomain shedding in a cell-type specific manner. One example includes the neuronally secreted *ADAM22* and *CD200*, which we identified as new substrates of the Alzheimer-linked protease *BACE1*. Taken together, hiSPECS and the brain glyco-secretome resource can be exploited for a wide range of applications to study protein secretion and shedding.

3.1 Introduction

Most omic quantitative analyses are based on gene expression (mRNA quantification), however, few studies have performed such analyses at the protein level (protein quantification). At the protein level, fundamental aspects, for example, which proteins exist, where the proteins are expressed and in what quantities they are expressed, are not yet fully resolved. To help the scientific community resolve these questions, a quantitative proteomic analysis of mouse brain secretome was performed using high-resolution mass spectrometry. Briefly, the main aim of this project was to generate a quantitative secretome map of the glyco-secretome from different cells of the mouse brain using an improved approach termed as, high performance secretome-protein-enrichment-with-click-sugars - hiSPECS. Dysregulated protein secretion and the shedding of signaling proteins have been linked to multiple complex diseases, e.g. diabetes, obesity, inflammation, neurodegeneration, and cancer. Even though mice and humans have adapted to different environments, mouse models are still an invaluable resource for studying biological processes that have been preserved in the course of the evolution of both the rodent

and primate lineages, or for the investigation of the conserved mammalian developmental mechanisms. Consequently, the identification and quantification of the mouse brain secretome not only allows the understanding of the underlying biological processes under physiological conditions, but also, may bring forth the molecular basis of complex diseases (e.g., Alzheimers' disease) and identify potential drug targets or biomarkers. To understand intercellular communication, it is essential to define the cellular secretome. The secretome constitutes a collection of proteins secreted (either unconventionally secreted or also proteolytically-shed) into the extracellular environment of a cell, Figure 21. The study of the secretome has recently brought about the field of secretomics: a sub-field of proteomics that represents a reliable strategy for the characterization and quantification of proteins secreted by a given cell under specific conditions[244].

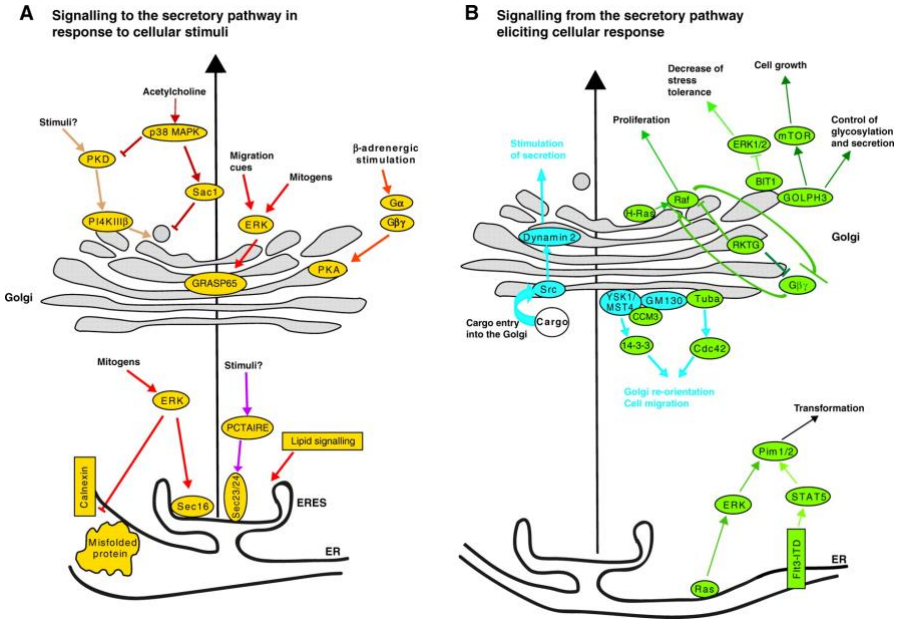


Figure 21: Signalling to and from the early secretory pathway. (A, B) ER, ERESs and Golgi complex with the different signalling cascades that are either directed towards these organelles (A, yellow), or emanating from them (B, green). Autochthonous Golgi signalling pathways are shown in blue. Stimuli that trigger signalling to the secretory pathway (A) or the cellular responses elicited by signalling from the secretory pathway (B) are shown in yellow or green, respectively. The long black arrows indicate direction of transport along the secretory pathway. Adapted from Farhan et. al., 2011.

Current approaches for the experimental detection of proteomic abundances or secretion have only revealed a fraction of the proteins expressed in a cell under a particular physiological condition. For instance, in 2001, Georgiu et. al., pointed out how difficult it is to conventionally detect low abundant plasma proteins due to the presence of more abundant proteins (e.g., albumin) that account for up to 80% of the total protein[245]. In addition, some proteins may only be secreted by specialised cell types, or are solely secreted in the course of specific developmental stages, or their secretion is only induced in response to cellular specific responses, e.g., after tumorigenesis or inflammation[246–249].

Since quantitative methodologies to decipher the secretome are challenging, in this study, we miniaturized secretome analysis by developing the improved secretome-protein-enrichment-with-click-sugars method (hiSPECS), which identifies the glyco-secretome. We then applied

this method to bring forth a cell type-resolved mouse brain glyco-secretome resource. In total, our experiment yielded 1023 proteins, with 995 proteins occurring in at least 5 of the 6 replicates per sample analysed (neurones, oligodendrocytes, microglia and astrocytes). Interestingly, we found high number of secreted proteins were shed in a cell-type specific manner. Examples include the neuronally secreted *ADAM22* and *CD200*, which we identified as new substrates of the Alzheimer-linked protease *BACE1*.
Also, GO and KEGG pathway analyses of the significantly enriched secreted proteins disclosed a high abundance of the mouse brain glyco-secretome was related to processes such as such as (i) metabolic process, gliogenesis, immune response for the astrocyte secretome, (ii) autophagy and phagocytosis for microglia, axon guidance, trans-synaptic signaling, (iii) axonogenesis and neurogenesis for neurons and (iv) lipid metabolic process and myelination for oligodendrocytes, thereby representing key cell-type specific biological functions of the four main brain cell types at the secretome level.

3.2 Mass spectrometry (MS)-based proteomics.

MS-based proteomics is the method of choice for protein identification, quantification, and characterization. Protein quantitation can be achieved via labelling-based quantitation or via label-free quantification (LFQ), with the label-free approaches more preferred[250]. Presently, two prevalent approaches in MS-based proteomics exist: 'top-down' and 'bottom-up' techniques, with bottom-up approach being the most widely used[251–254].

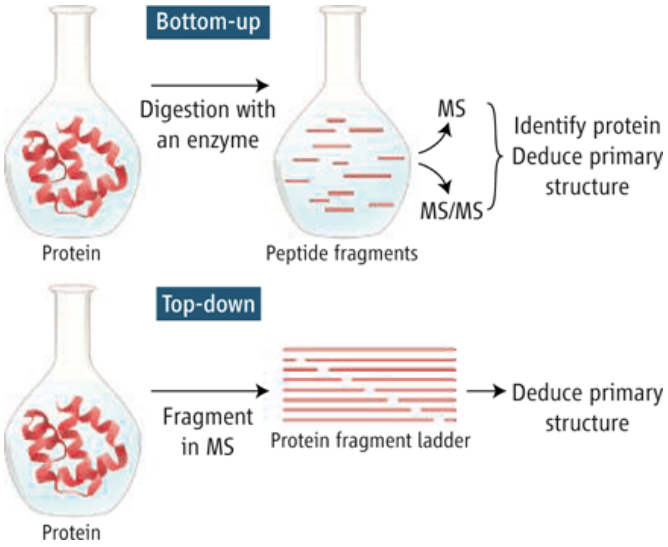


Figure 22: MS-based proteomics approaches. The top part of the image shows the bottom-up MS approach, while the bottom part of the image shows the top-down approach. Adapted from Chait et. al., 2006[251].

In the bottom-up approach, the proteins of interest are digested in solution with an enzyme, and the resulting peptides are then analysed in the gas phase by mass spectrometry. Firstly (referred to as "MS" or "MS1"), the masses of the intact peptides are determined; and secondly (referred to as "MS/MS" or "MS2"), these peptide ions are further fragmented to provide information on the identities and sequences of the proteins as well as any inherent modifications. In this way, parallel acquisition of quantitative information on thousands of

proteins and post-translational modifications from minute quantities of the input material is achieved. In the top-down proteomics experiments, intact protein ions are introduced into a gas phase and are then fragmented before being analysed in the mass spectrometer. The result is the molecular masses of the proteins and the protein ion fragment ladders. This information may then be used to characterise the complete primary structure of the protein. The top-down approach thus analyzes intact proteins and enables the identification of different protein forms, i.e., proteoforms. A setback of this approach is in its application in proteome-wide analyses due to difficulties with protein fractionation, protein ionization and fragmentation in the gas phase[254]. The data generated and used in this experiment was obtained via the bottom-up approach. Despite the different approaches, MS-based proteomics workflows involve[255]: (i) proteins extraction from the biological material under study (e.g., a tissue or cell), (ii) proteolytic digestion into peptides by site-specific proteases (e.g., trypsin), (iii) high-resolution peptide separation (liquid chromatography, also called LC-MS), (iv) peptide ionization (e.g., via electrospray ionization - ESI), (v) tandem mass spectrometry analysis (MS2). Label-free LC-MS/MS-based proteomics allow accurate peptide peak intensity identification from biological samples via Data-dependent acquisition (DDA) or Data-independent acquisition (DIA)[256,257]. In DDA acquisition, the mass spectrometers generate full-scan mass spectra in order to determine the molecular weights of the various peptide species present in a sample and then acquires MS/MS spectra only on the top "N" most intense peptides. In DIA acquisition, all ions of the entire mass range are sequentially isolated within defined and broader mass (m/z) windows and then fragmented together. The identification of spectra is then achieved via the use of library spectra previously generated from DDA approaches. At the end, bioinformatics data analyses are performed to identify or detect the differential expression patterns of the proteins secreted or expressed in the biological samples under study. The statistical elucidation of the protein expression variation in different biological conditions from different cell types, tissues or organs, and delineating the experimental factors that control such protein expression and activity are vital in biological and biomedical research.

3.3 The brain cell types.

The brain consists primarily of Neurons, Microglia, Oligodendrocytes and Astrocytes. Neurons function as the primary communication unit of the brain, and come in various shapes, sizes and specialties, for instance, motor neurons send and receive messages to muscles within the body for the purpose of movement. Microglia, which originate from macrophages, are at the center of phagocytosis. The Microglia prune synapses and help to avoid a hyper-connected brain. Oligodenrocytes, a type of glial cells, are the insulators of the brain, i.e., they wrap neuronal axons with thick fatty layers (myelin) thus helping to shield neurons from shock. A deficiency in myelin levels is associated with diseases such as Schizophrenia. Lastly, Astrocytes function as supportive layers of the brain. They also assist in clearing wastes between neurons and regulate blood flow to the brain. When the communication between astrocytes and neurons is interfered with, neurological disorders, depression or dementia can set in[258–261].

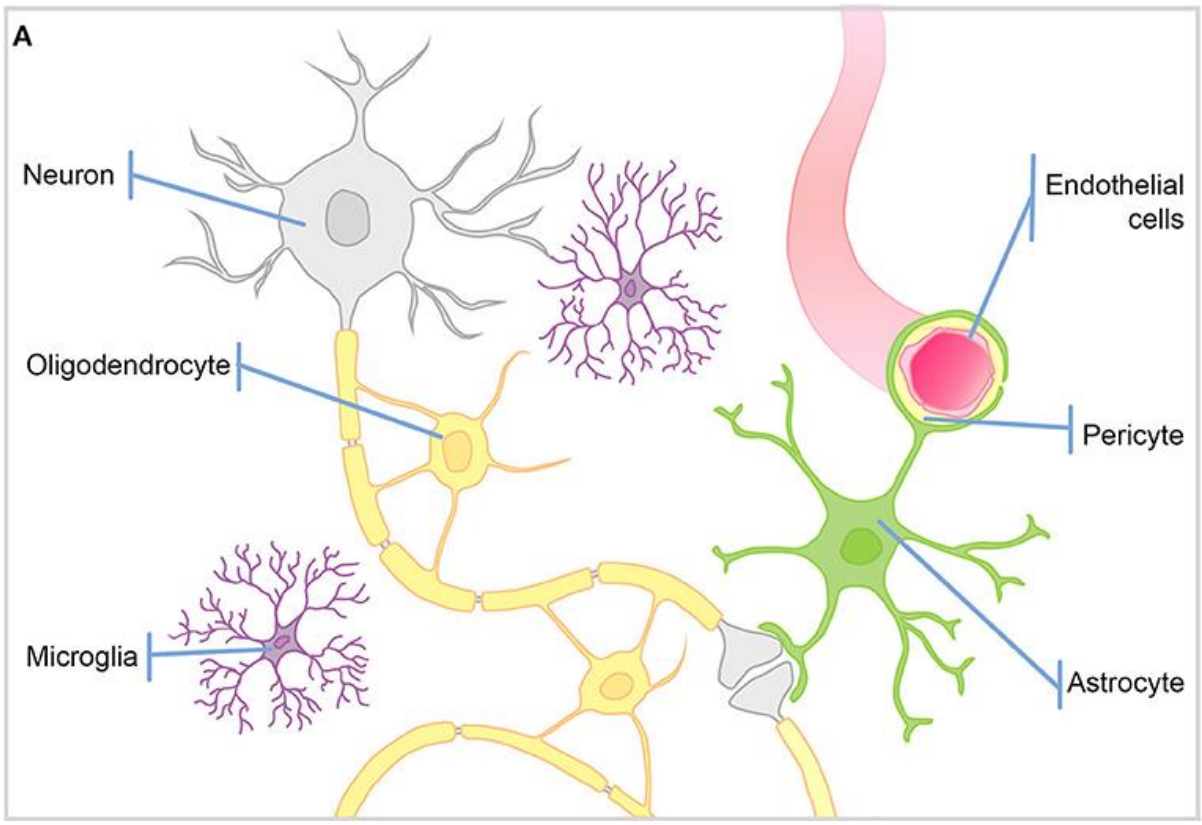


Figure 23: Brain cell types. A cartoon image depicting the 4 main brain cell types. Image adapted with modifications from Carpanini et. al., 2019.

3.4 Materials and Methods

3.4.1. Data pre-processing and normalization

Before normalisation, we selected only the proteins detected in at least 5 of the 6 replicates (5/6 or 6/6) in each of the 4 cell types. This yielded a total of 995 proteins from the pool of 1083 proteins. Data normalisation was achieved via variance stabilisation by employing the R package vsn (Figure 24). For the missing protein data, an imputation approach was undertaken based on the protocol explained in[263]. Briefly, a manually defined left-shifted Gaussian distribution (shift of 1.8 and scale of 0.3) for the data not missing at random (MNAR). After imputation, we did principal component analysis to understand the relationship between the cell types. Additionally, we downloaded the mouse protein-protein interaction network (PPIN) from BioGRID[223] and additional binary interactions data from UniProt[224].

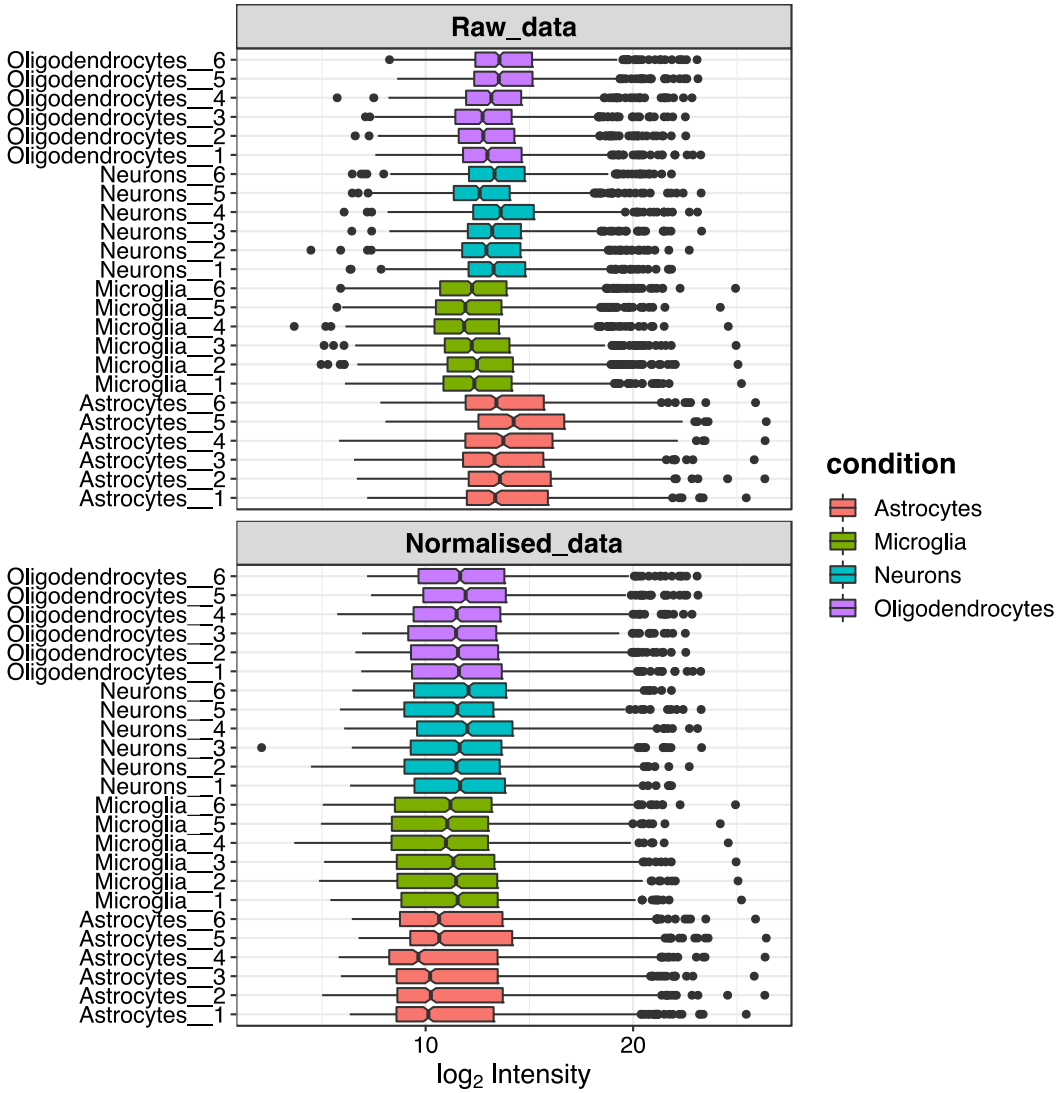


Figure 24: Diagram showing the distribution of the data prior and after normalisation. Data normalisation was achieved via variance stabilisation. For the missing data, a one-step imputation approach (left-shifted Gaussian distribution) was undertaken.

3.4.2. Detection of differentially expressed (DE) proteins.

We employed protein-wise linear models combined with empirical Bayes statistics (implemented in the R package Limma[264]) to detect differentially expressed proteins between any two cell types (pairwise comparison) as previously suggested[265]. Limma has the advantage of modelling unequal variances even for experiments with small sample sizes. For a protein to be selected as differentially expressed between any two cell types, we set a cut-off p-value of <0.05 (Bonferroni corrected) and a log fold-change (LogFC) of 2. The LogFC was set at 2 in order to reduce false positives as a result of data imputation.

3.4.3. Identification of Gene Ontology, KEGG pathways and disease-gene relations significantly enriched by proteins differentially secreted across brain cell types.

To understand the biological relevance of proteins significantly differentiated in one cell type with respect to the other 3 cell types, we identified statistically enriched Gene ontology (GO) terms and KEGG pathways. We did GO and KEGG pathway enrichment analysis using the R package ClusterProfiler[234]. To avoid statistical bias[233] in the enrichment analysis, we used all the proteins detected from our mass spectrometry analysis as the background for comparison. Statistical significance was calculated using Fisher's exact test and GO terms having a corrected (Benjamini Hochberg) p-value of <0.05 were chosen to be considerably enhanced. To determine if the proteins significantly differentiated in one cell type with respect to the other 3 cell types are linked to neurodegenerative diseases, we searched for curated gene disease associations (GDA) from DisGeNET[214]. Our search list contained 31 known diseases of the nervous system. We set the evidence index (EI) to 0.95: an EI of 1 indicates that all the available scientific literature supports the specific GDA.

3.4.4. Selection of cell type specific proteins and their interactions with cell lysate proteins detected by Sharma et. al, 2014.

To identify proteins with cell-type specific upregulation or downregulation, we employed a two-step procedure. First, we selected all proteins detected exclusively in one cell-type but not the other three cell types. Additionally, in each cell type, we chose proteins whose expressions had a fold change enrichment of >10 when compared to the other three cell types, as previously done[266]. We then searched for interacting partners between the proteins specifically enriched in the secretome of a specific cell type and the proteins from the cell's lysate as determined by Sharma et al[267]. First, we downloaded the mouse PPIN from BioGRID[223] and additional binary interactions data from UniProt[224].

3.4.5. Statistical evaluation

All statistical analyses were carried out in Python or the R environment. The moderated t-test coupled with empirical Bayes in limma was used to determine the extent of protein differential expression across the cell types. Pearson correlation was used to calculate the correlation coefficients between replicates of a cell type as well as across cell types. Fisher's exact test was used to identify significantly enriched GO terms and KEGG pathways.
For all analyses, p-values < 0.05 were considered significant.

3.6 RESULTS GENERATED FROM THE BIOINFORMATICS ANALYSES

3.6.1 Proteins Expression per sample

Mass spectrometry (LC-MS/MS) analysis of the brain glycol-secretome yielded a total of 1023 proteins from 4 different brain cell types, namely: Astrocytes, Neorons, Microglia and Oligodendrocytes. After removing the proteins that did not fit our selection criteria (protein quantification in at least 5 of the 6 replicates of each brain cell type) based on the raw LFQ intensity values, we established the protein coverage of the filtered data ranged from approximately 450 to 750 proteins per sample replicate. In total, 995 proteins were detected in at least 5 of the 6 replicates across the brain cell types. The microglia cell-type had the highest protein coverage while the astrocytes cell-type had the least protein coverage. Roughly 234 proteins were detected across all the cell types, while 377 proteins were detected in only one cell type -Figure 25.

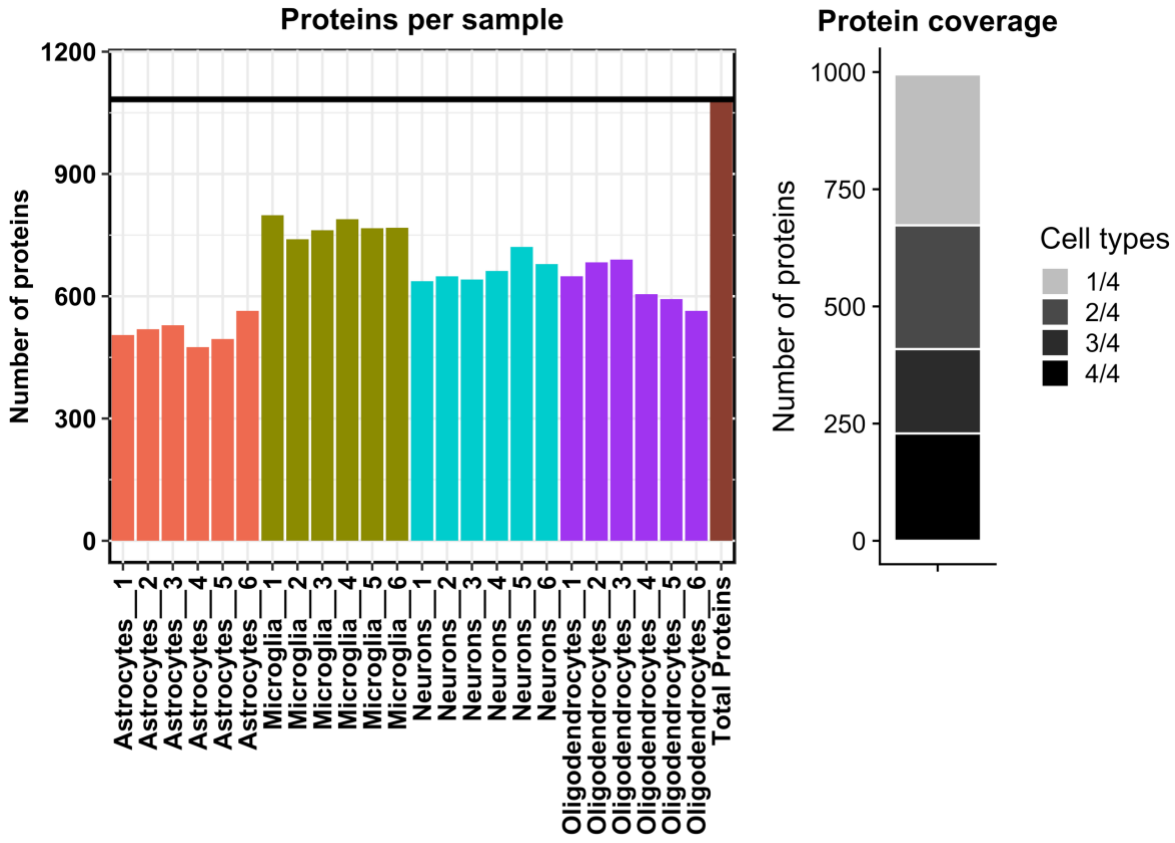


Figure 25: Protein coverage ranged from approximately 450 to 750 proteins per sample. A total of 995 proteins were detected in at least 5 of the 6 replicates across the brain cell types. Microglia cell-type had the highest protein coverage while Astrocytes had the lowest coverage.

3.6.2 Dimensionality reduction.

For enhanced visualisation and interpretation of the secretomics data, we used PCA (Principal Component Analysis) and UMAP (Uniform Manifold Approximation and Projection) for dimensionality reduction. Dimensionality reduction is advantageous while it enables researchers to speedily have a data-centric overall view of high-throughput data. While PCA has been the algorithm of choice in multiple studies, UMAP and other nonlinear dimensionality reduction algorithms have started to be the methods of choice for scientists[268,269]. On the one hand, PCA uses linear relationships of variables to build orthogonal axes that efficiently capture the variation inherent in the data with fewer variables. On the other hand, nonlinear dimensionality reduction algorithms (e.g., t-Distributed Stochastic Neighbor Embedding[270] (t-SNE) and UMAP) are able to do away with overcrowding of the data representation, wherein distinct data (or sample) clusters are represented on an overlapping area. UMAP is a recently developed non-linear dimensionality algorithm that offers higher reproducibility, meaningful organization of data clusters coupled with faster run times when compared with other algorithms[271,272].

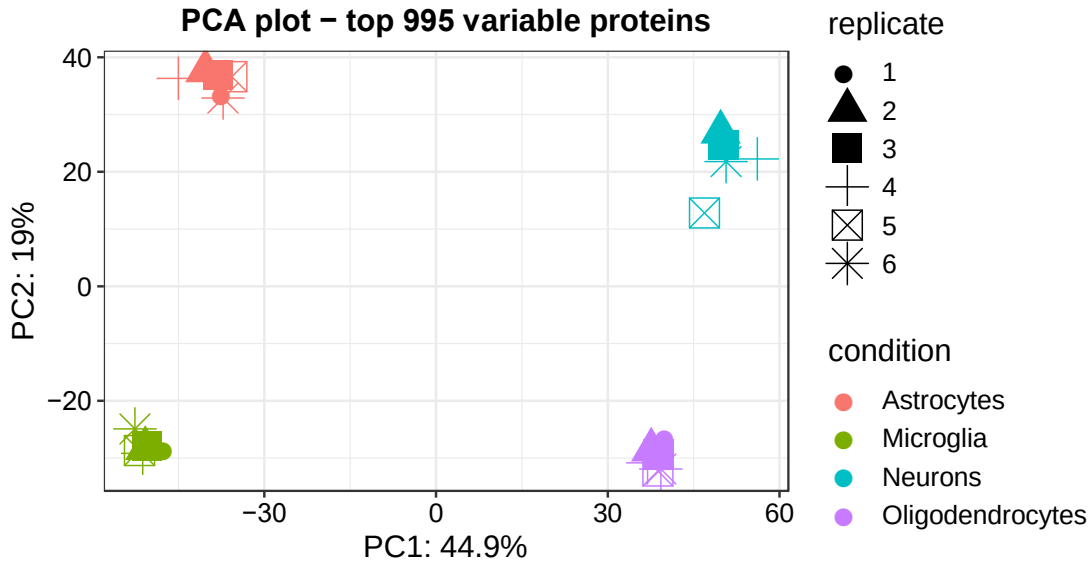


Figure 26A: PCA analysis. The secretomes of the cell types segregated based on component 1 and component 2, which accounted for 44.9% and 19% of the variability, respectively.

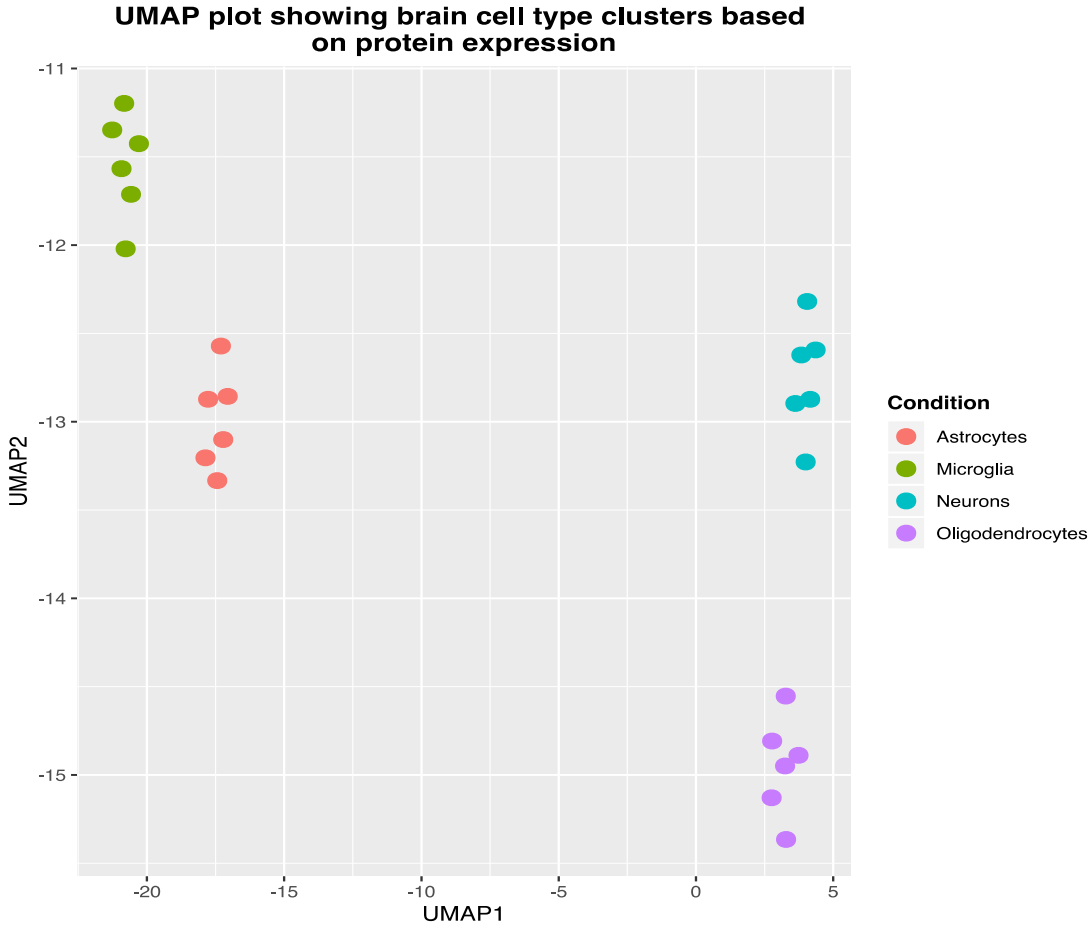


Figure 26B: UMAP (Uniform Manifold Approximation and Projection) plot showing brain cell type clusters based on log transformed raw LFQ intensities of quantified proteins. This indicates that the secretomes of the four cell types differ from each other.

Both PCA and UMAP determined that oligodendrocytes and neurones secreted a higher percentage of similar proteins. We also observed a more proteins being shared between secreted proteins from the astrocytes and microglia. Similar relationships have previously been postulated by Sharma et. al[267], thus giving us the confidence to perform further analysis.

3.6.3 Correlation between replicates of a sample and across brain cell type samples.

Inferential statistics such as the calculation of person's correlation (also known as the product moment correlation coefficient) enable a quick determination (generalization) of the relationships between samples or populations. To find the linear relationship between any of the 4 brain cell types, we computed the pearson's correlation coefficient (*r*, determined from the r distribution). The pearson's *r* ranges between -1 and 1, with -1 indicating a perfect negative correlation and 1 indicating a perfect positive correlation. A coefficient of 0 indicates no linear relationship between samples. In our dataset, all replicates of a cell types showed higher correlations (>0.7) as compared to replicates from other cell types - Figure 27.

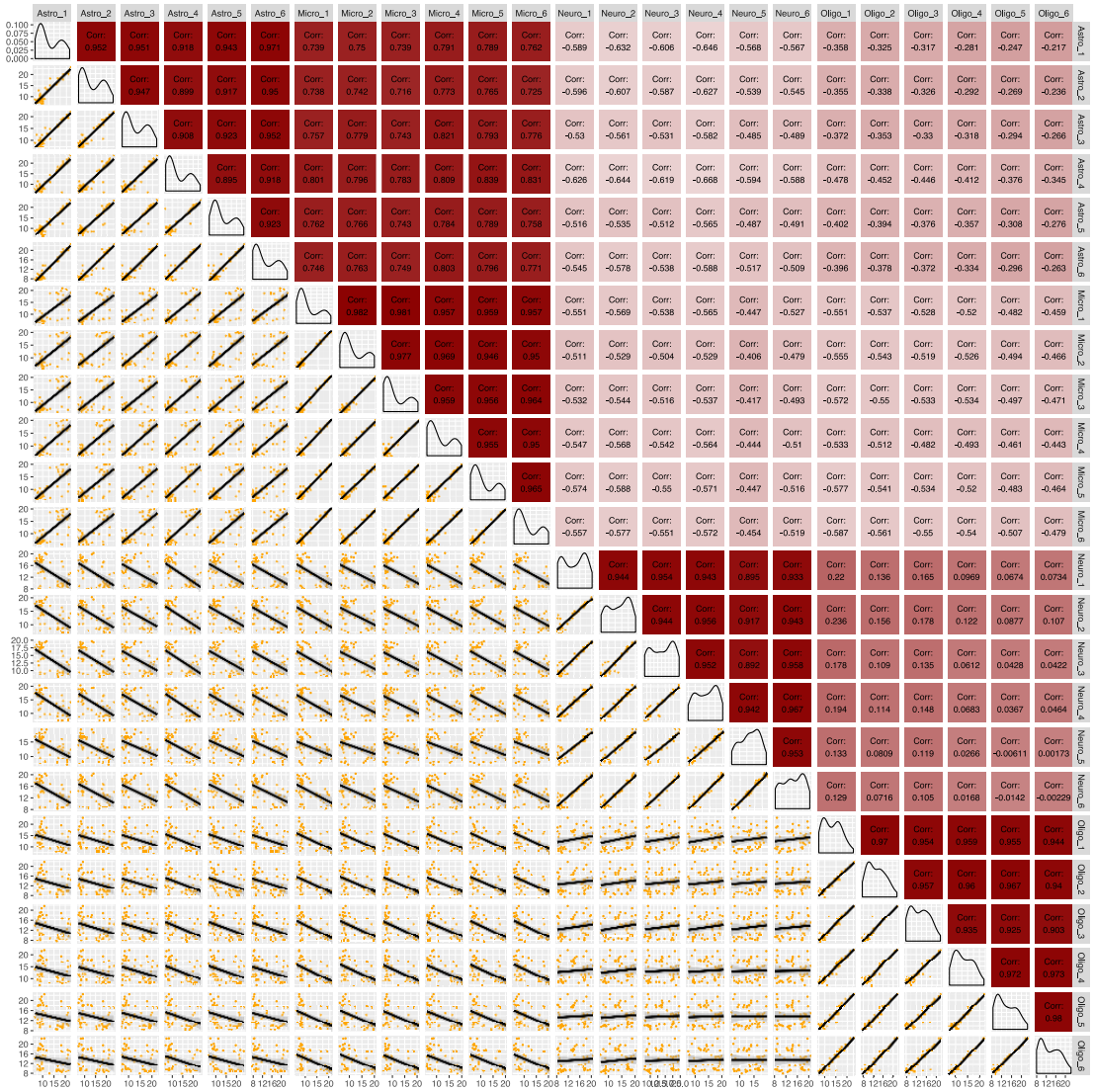


Figure 27: Correlation matrix showing the relationship between the different brain cell types. The matrix shows the Pearson correlation coefficient (dark red indicates a higher, light shade of red indicates a lower correlation) and the correlation plots of the log2 LFQ intensities of the secretome of astrocytes, neurons, microglia and oligodendrocytes processed with the iSPECS method.

3.6.4 Hierarchical clustering and detection of the cell-type specific secreted proteins

We employed protein-wise linear models combined with empirical Bayes statistics (using the R package Limma[264]) to detect differentially expressed proteins as previously suggested[265]. While the top significantly differentially expressed proteins revealed distinct secretion of the proteins across the different brain cell types – Figure 28, we observed close relations between neurons and oligondedrocytes, as well as between astrocytes and microglia. This observation is in line with previously available literature[266]. Proteins significantly differentiated in one cell type with respect to the other 3 cell types were analysed with regard to their biological function via GO and KEGG pathway enrichment analysis. Table 10 is a summary containing the number of differentially expressed proteins from the pairwise comparisons between brain cell types.

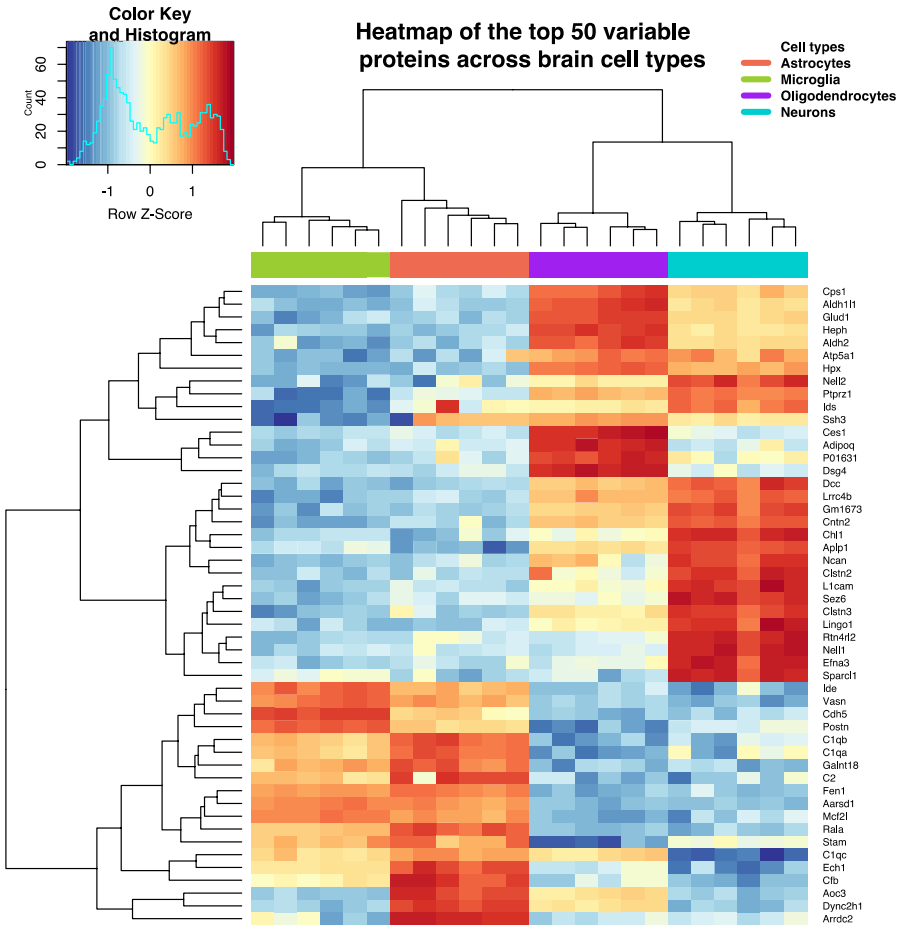


Figure 28: Heatmap of the top 50 differentially expressed proteins (Bonferroni p.adj < 0.05) across the 4 cell types from hierarchical clustering. The rows represent the differentially expressed proteins and the columns represent the cell types (and their replicates). The colours in the Heatmap represent log-scaled (z-scores) expression levels with blue indicating the lowest expression, white indicating intermediate expression, and red indicating the highest expression. The rows represent the differentially secreted proteins and the columns represent the cell types with their replicates. The colors represent log-scaled protein levels with blue indicating the lowest, white indicating intermediate, and red indicating the highest protein levels.

Table 10: Number of differentially expressed protein between brain cell types.

Cell type comparison	Number of up regulated proteins	Number of down regulated proteins
Astrocytes_vs_Microglia	172	167
Astrocytes_vs_Neurons	181	256
Astrocytes_vs_Oligodendrocytes	174	241
Microglia_vs_Neurons	220	293
Microglia_vs_Oligodendrocytes	190	276
Neurons_vs_Oligodendrocytes	177	181

3.6.5 Enriched gene ontologies (GO) associated with the mouse secretome

To explore the biological implications of differential expression across brain cell types, we did GO and KEGG pathway enrichment analysis. Generally, the GO analyses of the significantly enriched secretome proteins pointed to functional clusters corresponding to well-known functions of the four different brain cell types: metabolic process, gliogenesis, immune response for the astrocyte secretome, autophagy and phagocytosis for microglia, axon guidance, trans-synaptic signaling, neurogenesis for neurons and lipid metabolic process and myelination for oligodendrocytes (Figures 28, 29 and 30). Our results demonstrate that cell function cannot be determined only by a cell´s proteome but also by its secretome. Interestingly, we found that extracellular structure organisation, extracellular matrix structural constituent (Figures 29A, 30A and 31A), receptor ligand activity (Figure 30B) to be the most frequently enriched term underlining the quality of our secretome library.

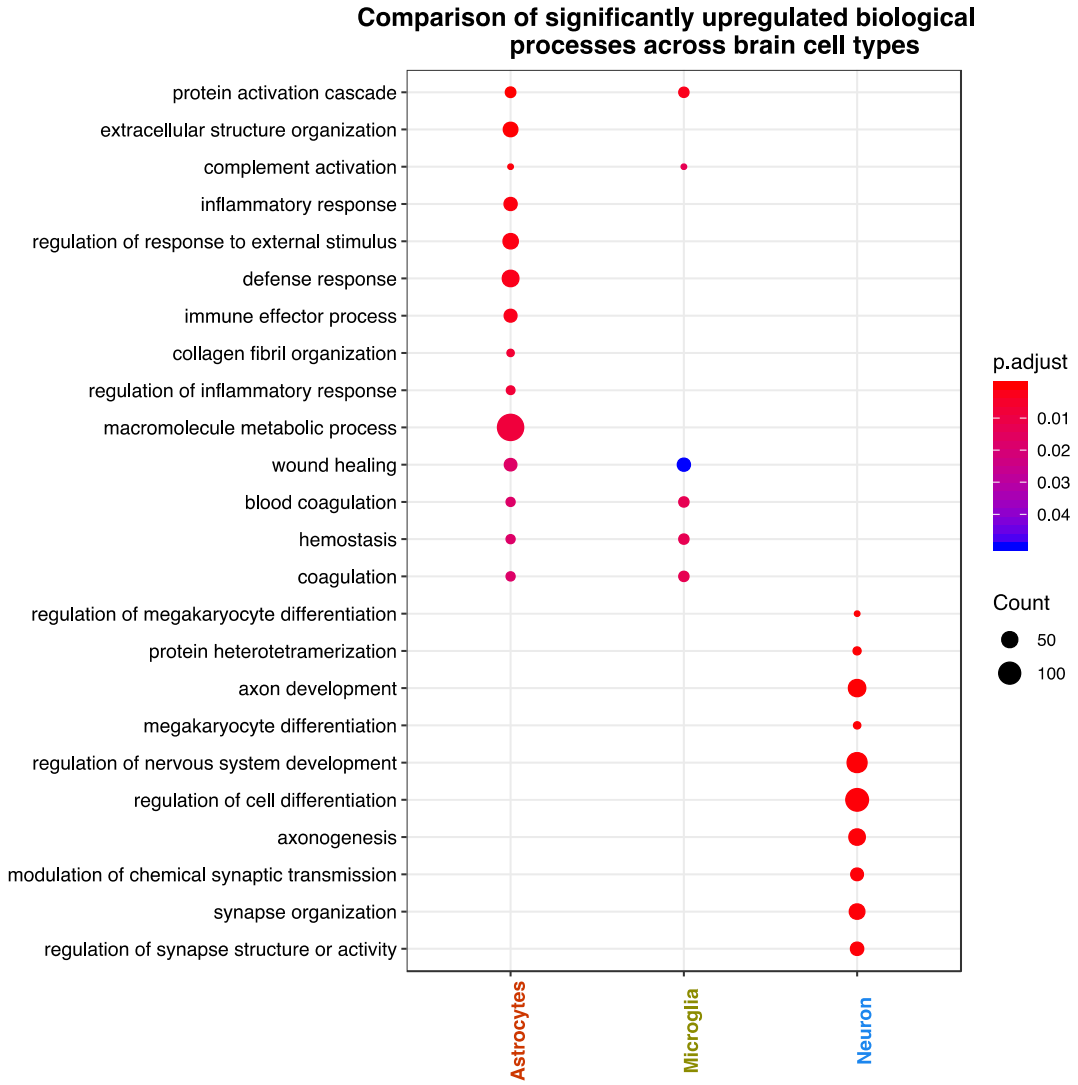


Figure 29A: Comparison of the biological processes enriched across brain cell types. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

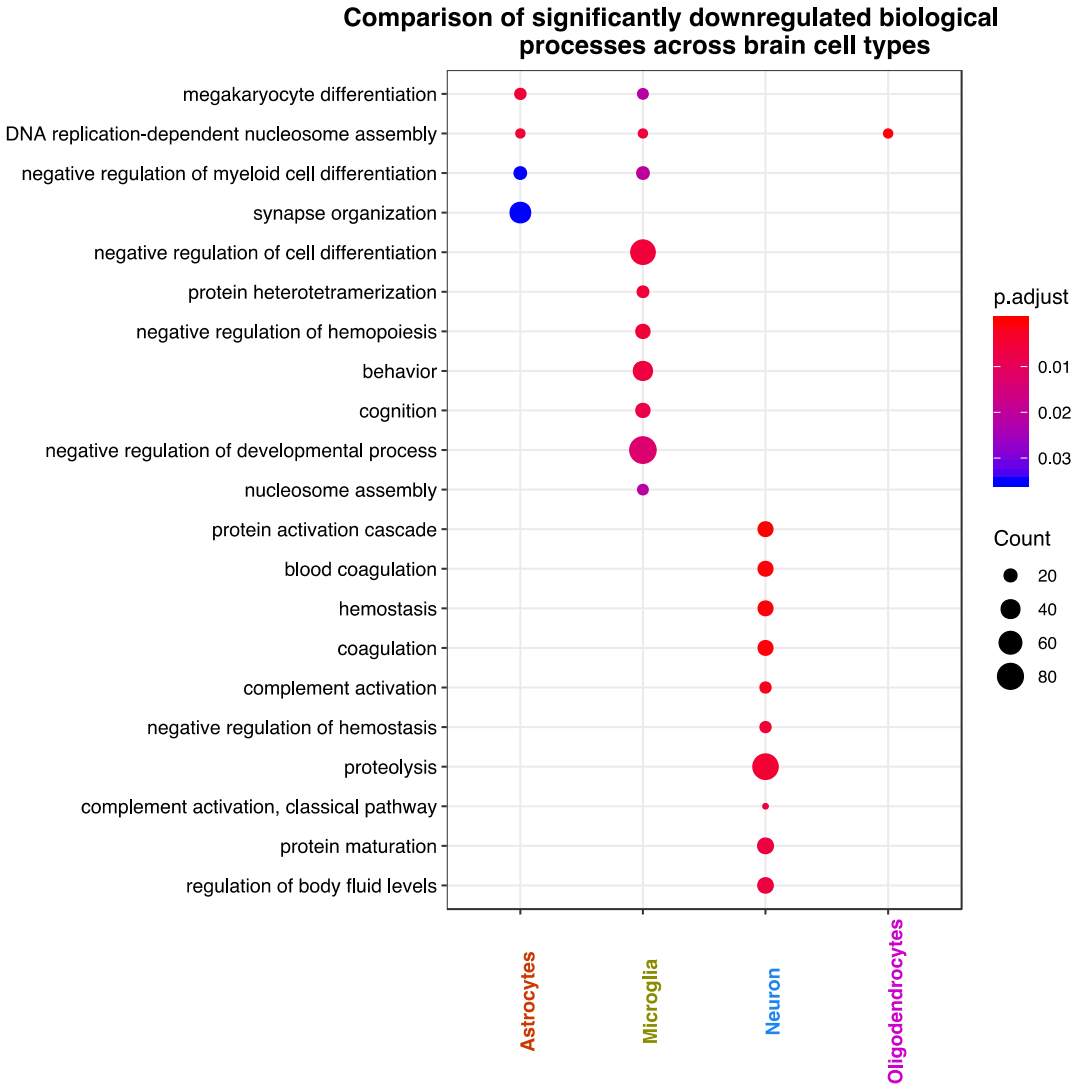


Figure 29B: Significantly downregulated processes were observed only in Neuron and Microglia cell types. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

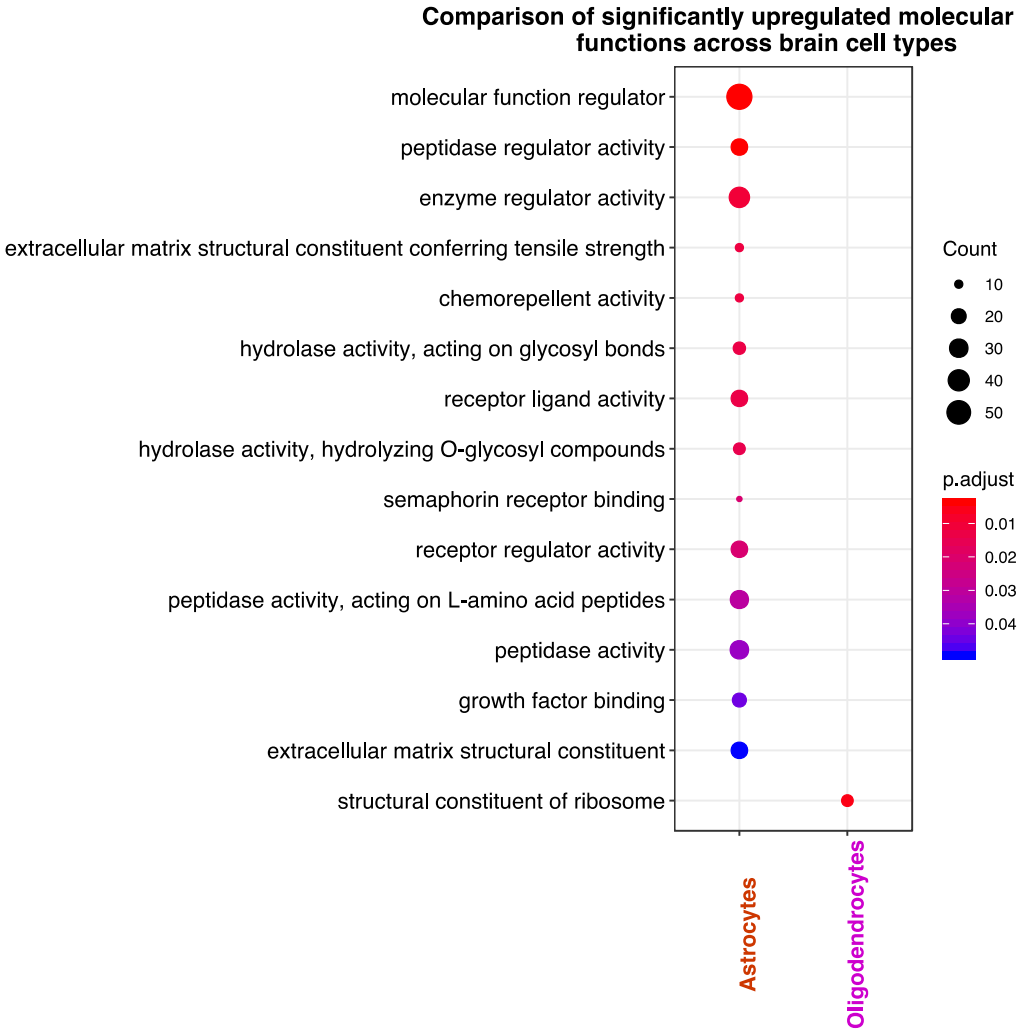


Figure 30A: Comparison of the molecular functions enriched across brain cell types. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

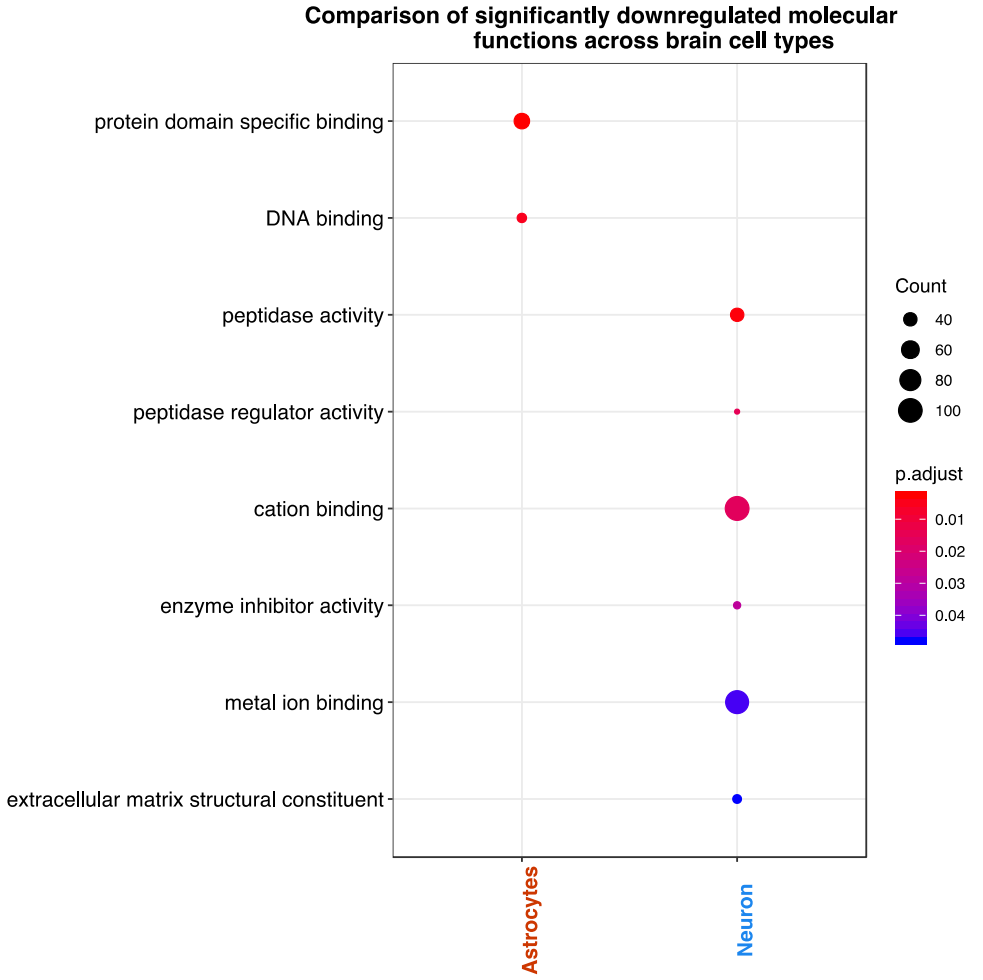


Figure 30B: Significantly downregulated molecular functions in brain cell type. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

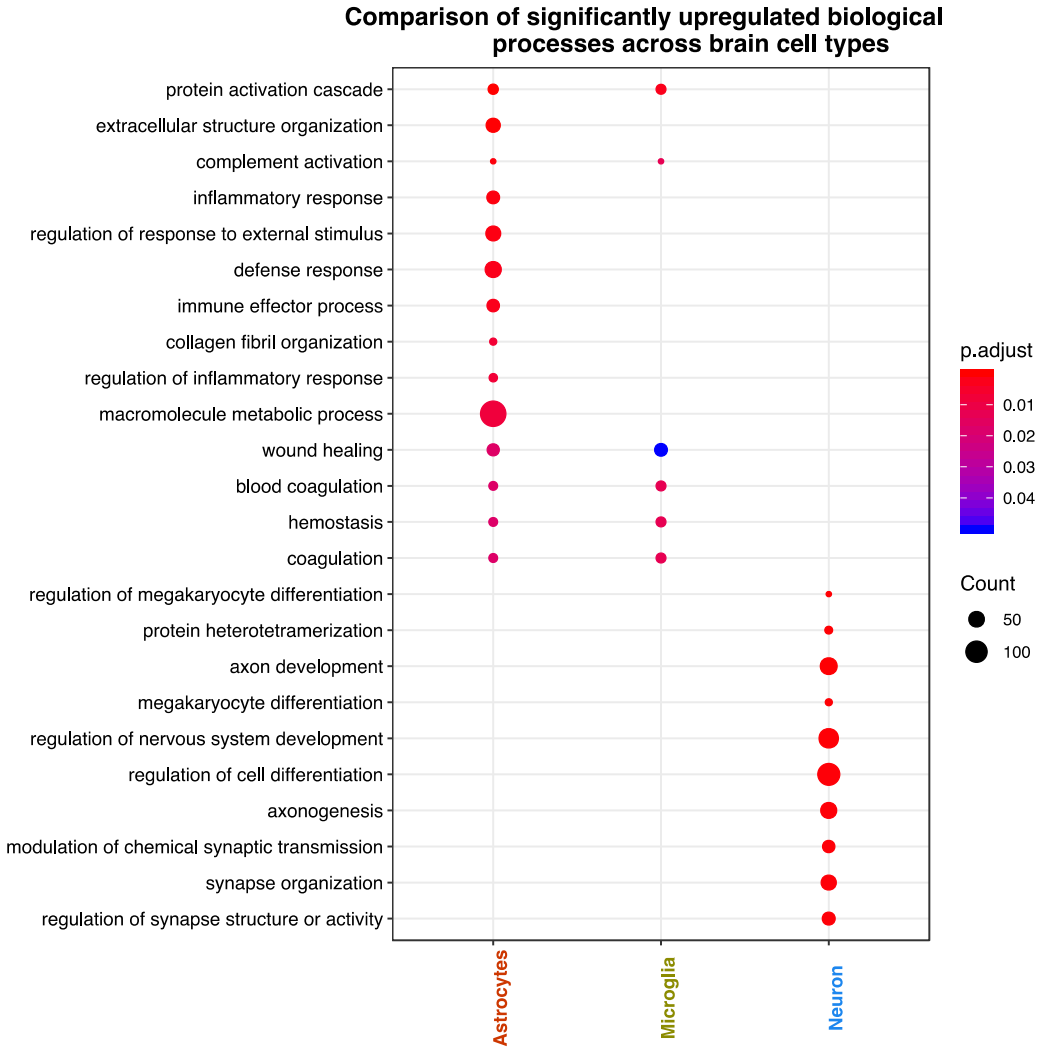


Figure 31A: KEGG pathways significantly enriched across brain cell types. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

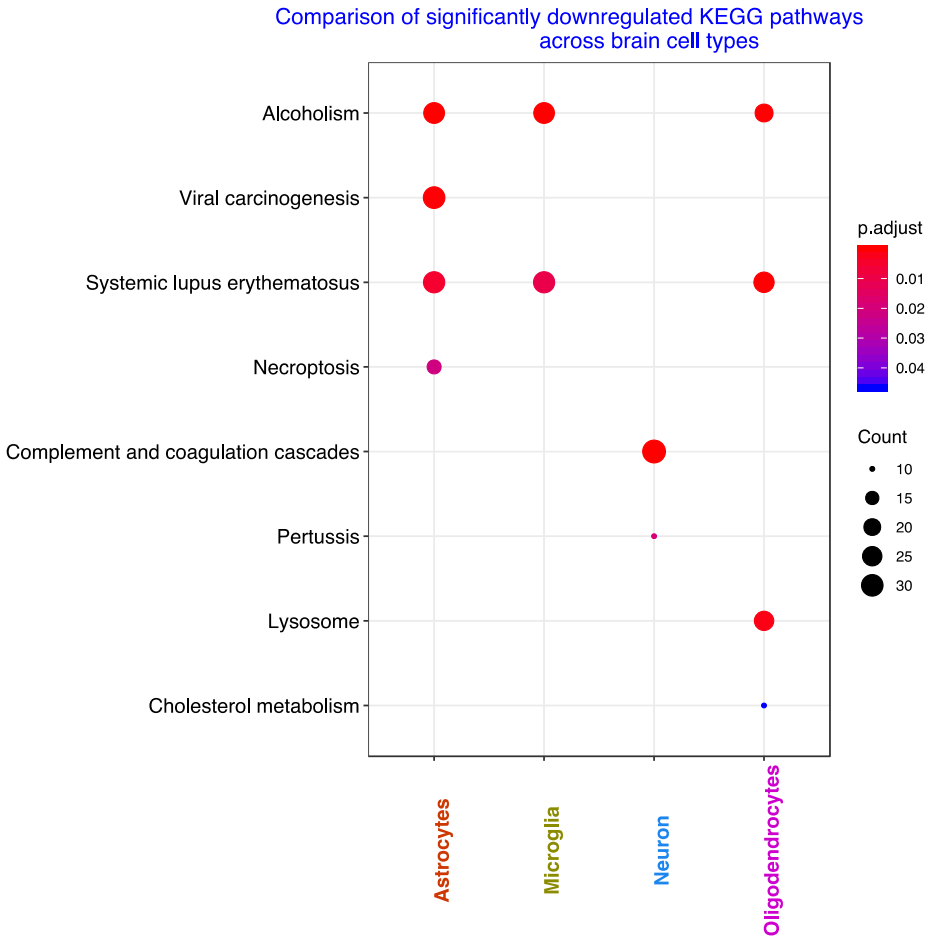


Figure 31B: Significantly downregulated KEGG pathways. The dot colour reflects the degree of significance (p-value) with red colour indicating a higher significance than the blue colour. The dot sizes indicate the number of proteins in our analysis were clustered in a particular GO term. The bigger the size of the dot, the more the number of proteins.

3.6.6. Interactions between CSF proteins and cell lysate proteins detected by Sharma et. al, 2014.

In order to unravel the complex network of inter-cellular communication between secreted proteins and transmembrane proteins (proteins from the cell lysate as determined by Sharma et al[267]) that may act as potential binding partners, we mapped known interaction partners (from UniProt and BioGRID) in a cell type resolved manner. We found a total of 711 unique interacting pairs, with all the proteins secreted with a cell type having interacting partners with proteins in the lysate of the other cell types (Figure 32 & Table 11).

Table 11: Number of protein interactions between CSF proteins and cell lysate proteins detected by Sharma et. al.

Celltype in iSPECS data	Cell type in Sharma data	Number of interactions
Astrocytes_iSPECS	Astrocytes_Sharma	9
Microglia_iSPECS	Astrocytes_Sharma	84
Neuron_iSPECS	Astrocytes_Sharma	31
Oligodendrocytes_iSPECS	Astrocytes_Sharma	44
Astrocytes_iSPECS	Microglia_Sharma	6
Microglia_iSPECS	Microglia_Sharma	111
Neuron_iSPECS	Microglia_Sharma	36
Oligodendrocytes_iSPECS	Microglia_Sharma	48
Astrocytes_iSPECS	Neuron_Sharma	11
Microglia_iSPECS	Neuron_Sharma	111
Neuron_iSPECS	Neuron_Sharma	115
Oligodendrocytes_iSPECS	Neuron_Sharma	59
Astrocytes_iSPECS	Oligodendrocytes_Sharma	2
Microglia_iSPECS	Oligodendrocytes_Sharma	23
Neuron_iSPECS	Oligodendrocytes_Sharma	16
Oligodendrocytes_iSPECS	Oligodendrocytes_Sharma	5

For example, the protein *CD200* is found exclusively in the secretome of neurons and binds to its receptor *CD200R1*, specifically expressed in microglia. CSF proteins from the neuron and the proteins from the neuron cell lysate had the highest number of interacting pairs (115 interactions), while those between CSF astrocytes and proteins from the oligodendrocytes cell lysate were the least (2 interactions). This observation suggests the presence of enhanced intracommunication between neurones with little intercommunication between neurons and oligodendrocytes.

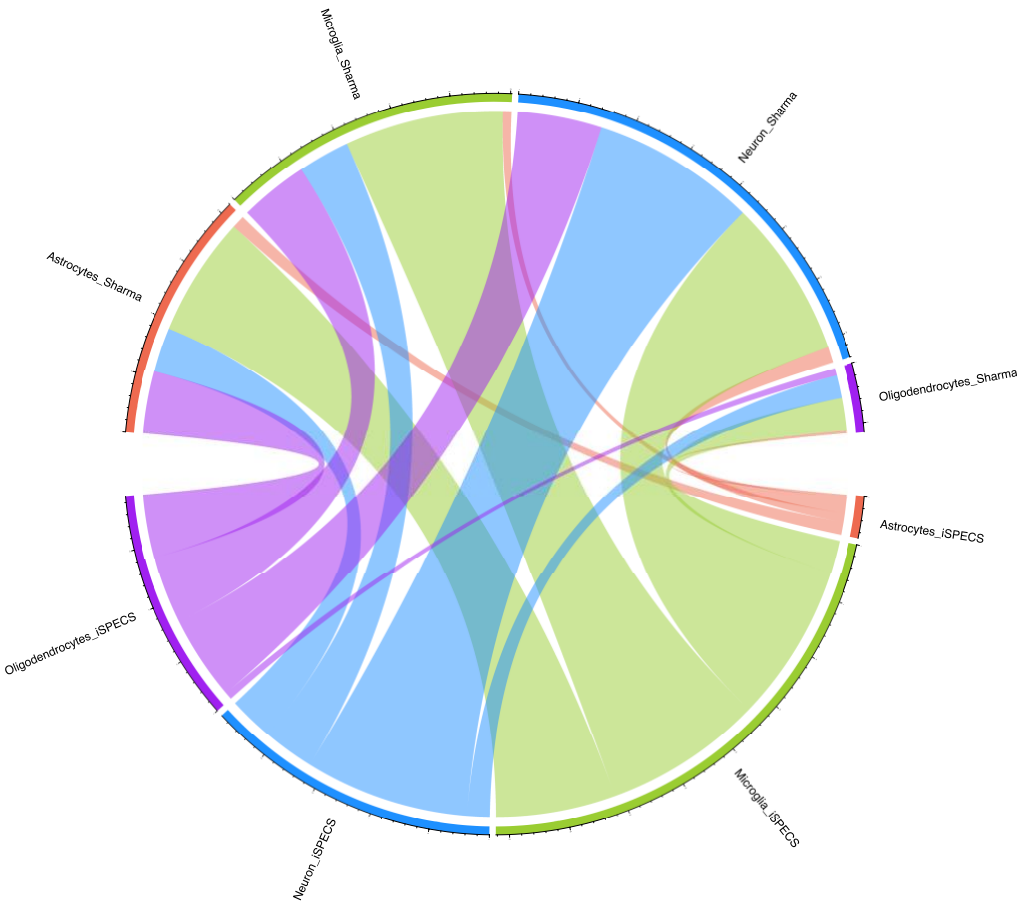


Figure 32: Interacting proteins between secreted CSF proteins and the cell lysate proteins detected by Sharma et.al.

3.6.7 Mapping of murine CSF proteins to disease association using the DisGeNET database.

From the multitude of the secreted proteins, we found at least 57 proteins being linked to neurodegenerative diseases (Figure 33), suggesting the robustness of iSPECS in resolving critical brain glycol-proteins. A majority of the proteins were involved in the Alzheimers disorder, Schizophrenia and Bipolar disorder.

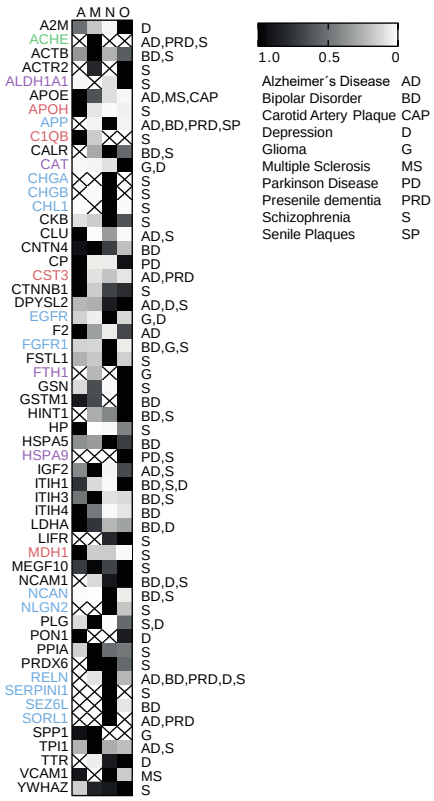


Figure 33. List of proteins detected in murine CSF and the iSPECS glyco-secretome resource which have human homologs that are linked to brain disease based on the DisGeNET database. Relative protein expression in the brain cell secretome is indicated with black showing the highest and white the lowest abundance. Colored gene names indicate cell type-specific secretion.

3.6.8 Summary

To sum up, the iSPECS approach miniaturizes secretome analyses under physiological culture conditions of cells and tissues. It also provides a highly reproducible and cost-effective way for deep secretome identification with cellular resolution. iSPECS enabled us to achieve: (i) a cell type-resolved brain glyco-secretome, (ii) further permitted the unravelling of the mechanisms involving cell-type specific protein secretion and, (iii) allowed the identification of the likely cellular origins of cell type-specifically secreted CSF-proteins. This secretome data will be available to the scientific community to complement the available genomic, transcriptomic, and proteomic data and thus facilitate further biomedical research. Taken together, iSPECS and the mouse brain glyco-secretome resource can be exploited for a wide range of applications to study cell-type specific protein secretion and shedding.

CHAPTER 4: THESIS SUMMARY

This thesis aims at providing a better understanding of the major disease-causing proteins involved in rewiring the protein-protein interaction network. Our approach has the potential application in identifying additional biomarkers that otherwise conventional differential expression analysis pipelines do not. Based on the obtained results presented in chapters 2 and 3, we developed a freely available database for ease of access to biologists and other oncologists. We hope that with this freely available online platform, the scientific community can easily and speedily access the generated data and allow them to perform further experimental validation studies. Additionally, this study highlights the importance of alternative splicing in tumorigenesis and how analyses of isoform expression in healthy versus cancer tissues lead to finding potential biomarkers that may be important targets in advancing the search of personalized cancer therapies. It is crucial at this point to state that, future differential expression analysis should put more emphasis on the importance of alternative splicing as these splice variants are the sources of the final proteoforms expressed in a particular phenotype.

To sum up, this thesis presents a novel and robust scheme capable of identifying known and novel cancer-specific and multi-cancer biomarkers using patient-specific PPIN derived from mRNA expression data. Furthermore, the ability to determine uniquely distorted interactions whose participants are predictive of patient survival opens up the possibility to computationally obtain potential protein biomarkers for specific cancer types and subtypes. We also established that SMGs do not bring about the majority of perturbations in cancer PPINs. Additionally, we found probable novel biomarkers such as the THCA BRAF-like specific 4-gene signature biomarker (*ODAM, APP, IKBKG,* and *TOLLIP*). The THCA biomarkers may be essential for disease monitoring of THCA subtypes whereas the 14-gene signature (with HRK node perturbation) explicitly observed in KICH samples is a candidate for therapeutic targeting. We were able to identify multiple protein interactions (edges) whose perturbation may have implications in tumorigenesis. Furthermore, we described gene ontologies (GO) and KEGG pathways enriched by the above-mentioned group of proteins across cancer types. Survival and functional enrichment analysis revealed that our candidate biomarkers are indeed involved in tumorigenesis. Last but not least, EdgeExplorer web portal allows for the free access of the findings by the scientific community. EdgeExplorer will not only facilitate experimental research in the continued quest for druggable proteins at the protein-protein interaction network level but will also be essential for researchers to quickly mine and access the proteins involved in edgetic perturbations of cancer PPINs. We envisage that subsequent experimental validation will demonstrate the applicability of the novel biomarkers generated in this study for making informed clinical decisions as well as in developing cancer therapies. In the future, studies should also seek to investigate and shed more light on patient-specific edgetic perturbations and determine proteins and corresponding isoforms (and protein domains) responsible for such disruptions.

Finally, iSPECS is a cost-effective and reproducible methodology that miniaturizes secretome analyses under the physiological culture conditions of cells and tissues. It has the suitability for a wide range of applications to study protein secretion and shedding as demonstrated with our experimental set-up of the cell type-resolved brain glyco-secretome. Our strategy and approach allowed us to not only unravel the mechanisms of cell-type specific protein secretion but also to identify the probable cellular origins of cell type-specifically secreted CSF-proteins. The collective results, provide a basis to elucidate the complex network of intercellular communication in organs, for example, the brain.

CHAPTER 5: PUBLICATIONS ARISING FROM THIS THESIS

5.1 Publications discussed in this thesis

1. Kataka, Evans., Zaucha, Jan., Frishman, Goar., Ruepp, Andreas. & Frishman, Dimitrij. Edgetic perturbation signatures represent known and novel cancer biomarkers. *Sci Rep* 10, 1–16 (2020).

2. **Johanna Tüshaus, Stephan A. Müller, Evans Sioma Kataka, Jan Zaucha, Laura Sebastian, Minhui Su, Gökhan Güner, Georg Jocher, Sabina Tahirovic, Dimitrij Frishman, Mikael Simons and Stefan F. Lichtenthaler.** An optimized quantitative proteomics method establishes the cell type-resolved mouse brain secretome. *The EMBO Journal*, e105693 (2020).

5.2 Other co-authored publications

3. Littmann, M., Selig, K., Cohen-Lavi, **Kataka E**., Hönigschmid P., Mösch A., Herty T., Burkhard Rost. Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. *Nat Mach Intell* **2,** 18–24 (2020). https://doi.org/10.1038/s42256-019-0139-8

4. Johanna Tüshaus, Stephan A. Müller, **Evans Sioma Kataka**, Jan Zaucha, Dimitrij Frishman, and Stefan F. Lichtenthaler. Neuronal Differentiation of LUHMES Cells Induces Substantial Changes of the Proteome. *PROTEOMICS*, 2000174 (2020).

5.3 Conference presentations

Poster presentation at the "The Seventh German-Russian Week of the Young Researcher: Computational Biology and Biomedicine", 11.09.-14.09.2017, Skolkovo Institute of Science and Technology, Moscow (Russia):
Kataka et.al.: *Edgetic perturbation signatures represent known and novel cancer biomarkers*.

Oral presentation at "Roche Future X Healthcare 2019: The digital revolution, e new era for patients", 13.11.-14.11.2019, Munich (Germany):
Kataka et.al.: *Edgetic perturbation signatures represent known and novel cancer biomarkers*.

CHAPTER 6: LIST OF SYMBOLS AND ABBREVIATIONS

6.1 Abbreviations

TCGA The Cancer Genome Atlas

RNA-Seq: RNA Sequencing technology

iSPECS: improved secr

SMGs: Significantly mutated genes

PPI: Protein Protein inetracrtion

PPIN: Protein Protein Intracrtion Networks

LUAD Lung Adenoarcinoma

LUSC Lung Squamous cell carcinoma

COAD Colon Adenocarcinoma

BRCA Breast Adenoarcinoma

HNSC Head and Neck Squamous cell Carcinoma

BLCA Bladder urothelialc Carcinoma

STES Stomach Esophegal carcinoma

THCA Thyroid Adenocarcinoma

PRAD Prostate Adenocarcinoma

KEGG Kyoto Encyclopedia of Genes and Genomes

KIRC Kidney Renal clear cell Carcinoma

KICH Kidney Chromophobe

FDR False Discovery Rate

GO Gene Ontology

NGS Next Generation Sequencing

SMGs Significantly Mutated Genes

REFERENCES

1. Surget, S., Khoury, M. P. & Bourdon, J.-C. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. *Onco Targets Ther* **7**, 57–68 (2013).
2. Schadt, E. E. Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223 (2009).
3. Libioulle, C. & Bours, V. [Complex diseases: the importance of genetics]. *Rev Med Liege* **67**, 220–225 (2012).
4. Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
5. Long, M., Fu, Z., Li, P. & Nie, Z. Cigarette smoking and the risk of nasopharyngeal carcinoma: a meta-analysis of epidemiological studies. *BMJ Open* **7**, (2017).
6. pubmeddev & al, J. L., et. Pathway-based analysis tools for complex diseases: a review. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/25462153.
7. Al-Harazi, O. *et al.* Integrated Genomic and Network-Based Analyses of Complex Diseases and Human Disease Network. *J Genet Genomics* **43**, 349–367 (2016).
8. Yu, X., Zeng, T., Wang, X., Li, G. & Chen, L. Unravelling personalized dysfunctional gene network of complex diseases based on differential network model. *J Transl Med* **13**, 189 (2015).
9. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
10. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians* **70**, 7–30 (2020).

11. Katalinic, A. The Burden of Cancer in Germany. *Dtsch Arztebl Int* **115**, 569–570 (2018).
12. Hansemann, D. Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung. *Archiv f. pathol. Anat.* **119**, 299–326 (1890).
13. Boveri, T. & Boveri, T. Zur Frage der Entstehung Maligner Tumoren. (1914).
14. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
15. Fidler, I. J. & Kripke, M. L. Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197**, 893–895 (1977).
16. Talmadge, J. E. Clonal Selection of Metastasis within the Life History of a Tumor. *Cancer Res* **67**, 11471–11475 (2007).
17. Loeb, L. A. & Harris, C. C. Advances in Chemical Carcinogenesis: A Historical Review and Prospective. *Cancer Res* **68**, 6863–6872 (2008).
18. Knudson, A. G. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U.S.A.* **68**, 820–823 (1971).
19. Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 961–968 (2010).
20. Asch, B. B. Tumor viruses and endogenous retrotransposons in mammary tumorigenesis. *J Mammary Gland Biol Neoplasia* **1**, 49–60 (1996).
21. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135–1145 (2008).
22. Hawkins, R. D., Hon, G. C. & Ren, B. Next-Generation Genomics: an Integrative Approach. *Nat Rev Genet* **11**, 476–486 (2010).
23. Levy, S. E. & Myers, R. M. Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet* **17**, 95–115 (2016).
24. Pabinger, S. *et al.* A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform* (2013) doi:10.1093/bib/bbs086.

25. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
26. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
27. Kaiser, R. J. *et al.* Specific-primer-directed DNA sequencing using automated fluorescence detection. *Nucleic Acids Res* **17**, 6087–6102 (1989).
28. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
29. Veer, L. J. van 't *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
30. Ahr, A. *et al.* Identification of high risk breast-cancer patients by gene expression profiling. *The Lancet* **359**, 131–132 (2002).
31. Pon, J. R. & Marra, M. A. Driver and passenger mutations in cancer. *Annu Rev Pathol* **10**, 25–50 (2015).
32. Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18545–18550 (2010).
33. Ding, J. *et al.* Systematic analysis of somatic mutations impacting gene expression in 12 tumour types. *Nat Commun* **6**, 8554 (2015).
34. Green, D. R. A BH3 Mimetic for Killing Cancer Cells. *Cell* **165**, 1560 (2016).
35. Fujita, K. *et al.* Cancer Therapy Due to Apoptosis: Galectin-9., Cancer Therapy Due to Apoptosis: Galectin-9. *Int J Mol Sci* **18, 18**, (2017).
36. Lage, K. *et al.* A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20870–20875 (2008).

37. Zhao, J., Cheng, F. & Zhao, Z. Tissue-Specific Signaling Networks Rewired by Major Somatic Mutations in Human Cancer Revealed by Proteome-Wide Discovery. *Cancer Res.* **77**, 2810–2821 (2017).
38. Gonzalez de Castro, D., Clarke, P. A., Al-Lazikani, B. & Workman, P. Personalized Cancer Medicine: Molecular Diagnostics, Predictive biomarkers, and Drug Resistance. *Clin Pharmacol Ther* **93**, 252–259 (2013).
39. Sawyers, C. L. The cancer biomarker problem. *Nature* **452**, 548–552 (2008).
40. El Marabti, E. & Younis, I. The Cancer Spliceome: Reprograming of Alternative Splicing in Cancer. *Front. Mol. Biosci.* **5**, (2018).
41. Vitting-Seerup, K. & Sandelin, A. The Landscape of Isoform Switches in Human Cancers. *Mol. Cancer Res.* **15**, 1206–1220 (2017).
42. Schittek, B. & Sinnberg, T. Biological functions of casein kinase 1 isoforms and putative roles in tumorigenesis. *Molecular Cancer* **13**, 231 (2014).
43. Corominas, R. *et al.* Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat Commun* **5**, (2014).
44. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
45. Castle, J. C. *et al.* Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat. Genet.* **40**, 1416–1425 (2008).
46. Kuo, I.-Y. *et al.* A prognostic predictor panel with DNA methylation biomarkers for early-stage lung adenocarcinoma in Asian and Caucasian populations. *J. Biomed. Sci.* **23**, 58 (2016).
47. Li, B., Cui, Y., Diehn, M. & Li, R. Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. *JAMA Oncol* **3**, 1529–1537 (2017).

48. Bourdon, J.-C. *et al.* p53 mutant breast cancer patients expressing p53γ have as good a prognosis as wild-type p53 breast cancer patients. *Breast Cancer Res* **13**, R7 (2011).
49. Nagane, M. *et al.* A common mutant epidermal growth factor receptor confers enhanced tumorigenicity on human glioblastoma cells by increasing proliferation and reducing apoptosis. *Cancer Res.* **56**, 5079–5086 (1996).
50. Tremblay, J. & Hamet, P. Role of genomics on the path to personalized medicine. *Metab. Clin. Exp.* **62 Suppl 1**, S2-5 (2013).
51. Choi, J. R. *et al.* Genetic Variations of Drug Transporters Can Influence on Drug Response in Patients Treated with Docetaxel Chemotherapy. *Cancer Res Treat* **47**, 509–517 (2015).
52. Hildebrandt, M. A. T. *et al.* Genetic Variations in the PI3K/PTEN/AKT/mTOR Pathway Are Associated With Clinical Outcomes in Esophageal Cancer Patients Treated With Chemoradiotherapy. *J Clin Oncol* **27**, 857–871 (2009).
53. Savas, S. & Liu, G. Genetic variations as cancer prognostic markers: review and update. *Human Mutation* **30**, 1369–1377 (2009).
54. Wu, X. *et al.* Germline Genetic variations in drug action pathways predict clinical outcomes in advanced lung cancer treated with platinum-based chemotherapy. *Pharmacogenet Genomics* **18**, 955–965 (2008).
55. Liu, Y. *et al.* Targeting tumor suppressor genes for cancer therapy. *BioEssays* **37**, 1277–1286 (2015).
56. Morris, L. G. T. & Chan, T. A. Therapeutic targeting of tumor suppressor genes. *Cancer* **121**, 1357–1368 (2015).
57. Dillon, L. M. & Miller, T. W. Therapeutic targeting of cancers with loss of PTEN function. *Curr Drug Targets* **15**, 65–79 (2014).

58. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Reviews Cancer* **4**, 177–183 (2004).

59. Croce, C. M. Oncogenes and cancer. *N. Engl. J. Med.* **358**, 502–511 (2008).

60. The Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113–1120 (2013).

61. Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-337.e10 (2018).

62. Nagy, R., Sweet, K. & Eng, C. Highly penetrant hereditary cancer syndromes. *Oncogene* **23**, 6445–6470 (2004).

63. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).

64. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucl. Acids Res.* **39**, D945–D950 (2011).

65. Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science* **349**, 1483–1489 (2015).

66. Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91**, 355–358 (2004).

67. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385.e18 (2018).

68. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).

69. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).

70. Gonzalez, M. W. & Kann, M. G. Chapter 4: Protein Interactions and Disease. *PLoS Comput Biol* **8**, (2012).

71. Sam, L., Liu, Y., Li, J., Friedman, C. & Lussier, Y. A. Discovery of protein interaction networks shared by diseases. in *Biocomputing 2007* 76–87 (WORLD SCIENTIFIC, 2006). doi:10.1142/9789812772435_0008.
72. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *PNAS* **98**, 4569–4574 (2001).
73. Gavin, A.-C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
74. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86 (1999).
75. Huynen, M., Snel, B., Lathe, W. & Bork, P. Predicting Protein Function by Genomic Context: Quantitative Evaluation and Qualitative Inferences. *Genome Res.* **10**, 1204–1210 (2000).
76. Ben-Hur, A. & Noble, W. S. Kernel methods for predicting protein–protein interactions. *Bioinformatics* **21**, i38–i46 (2005).
77. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. The use of gene clusters to infer functional coupling. *PNAS* **96**, 2896–2901 (1999).
78. Beadle, G. W. & Tatum, E. L. Genetic Control of Biochemical Reactions in Neurospora. *PNAS* **27**, 499–506 (1941).
79. Ozturk, K., Dow, M., Carlin, D. E., Bejar, R. & Carter, H. The Emerging Potential for Network Analysis to Inform Precision Cancer Medicine. *Journal of Molecular Biology* **430**, 2875–2899 (2018).
80. Vandin, F., Upfal, E. & Raphael, B. J. Algorithms for detecting significantly mutated pathways in cancer. *J. Comput. Biol.* **18**, 507–522 (2011).

81. Leiserson, M. D. M. *et al.* Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* **47**, 106–114 (2015).
82. Cho, A. *et al.* MUFFINN: cancer gene discovery via network analysis of somatic mutation data. *Genome Biology* **17**, 129 (2016).
83. Leiserson, M. D. M., Blokh, D., Sharan, R. & Raphael, B. J. Simultaneous Identification of Multiple Driver Pathways in Cancer. *PLOS Computational Biology* **9**, e1003054 (2013).
84. Ciriello, G., Cerami, E., Aksoy, B. A., Sander, C. & Schultz, N. Using MEMo to Discover Mutual Exclusivity Modules in Cancer. *Curr Protoc Bioinformatics* **CHAPTER 8**, Unit-8.17 (2013).
85. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
86. Vargas, A. J. & Harris, C. C. Biomarker development in the precision medicine era: lung cancer as a case study. *Nat. Rev. Cancer* **16**, 525–537 (2016).
87. Li, Z. *et al.* The OncoPPi network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. *Nat Commun* **8**, 14356 (2017).
88. Sharma, A., Costantini, S. & Colonna, G. The protein–protein interaction network of the human Sirtuin family. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1834**, 1998–2009 (2013).
89. Chuang, H.-Y., Lee, E., Liu, Y.-T., Lee, D. & Ideker, T. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* **3**, 140 (2007).
90. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).

91. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671–679 (2005).

92. Sah, N. K. & Seniya, C. Survivin splice variants and their diagnostic significance. *Tumour Biol.* **36**, 6623–6631 (2015).

93. Chen, J. & Weiss, W. A. Alternative splicing in cancer: implications for biology and therapy. *Oncogene* **34**, 1–14 (2015).

94. Sharma, S., Kelly, T. K. & Jones, P. A. Epigenetics in cancer. *Carcinogenesis* **31**, 27–36 (2010).

95. Jonsson, P. F. & Bates, P. A. Global topological features of cancer proteins in the human interactome. *Bioinformatics* **22**, 2291–2297 (2006).

96. Sun, J. & Zhao, Z. A comparative study of cancer proteins in the human protein-protein interaction network. *BMC Genomics* **11**, S5 (2010).

97. Nishi, H. *et al.* Cancer Missense Mutations Alter Binding Properties of Proteins and Their Interaction Networks. *PLOS ONE* **8**, e66273 (2013).

98. Buljan, M., Blattmann, P., Aebersold, R. & Boutros, M. Systematic characterization of pan-cancer mutation clusters. *Molecular Systems Biology* **14**, e7974 (2018).

99. Bowler, E. H., Wang, Z. & Ewing, R. M. How do oncoprotein mutations rewire protein-protein interaction networks? *Expert Rev Proteomics* **12**, 449–455 (2015).

100. Cheng, F. *et al.* Studying tumorigenesis through network evolution and somatic mutational perturbations in the cancer interactome. *Mol. Biol. Evol.* **31**, 2156–2169 (2014).

101. Yan, W., Xue, W., Chen, J. & Hu, G. Biological Networks for Cancer Candidate Biomarkers Discovery. *Cancer Inform* **15**, 1–7 (2016).

102. Cui, H., Zhao, N. & Korkin, D. Multilayer View of Pathogenic SNVs in Human Interactome through in-silico Edgetic Profiling. *Journal of Molecular Biology* (2018) doi:10.1016/j.jmb.2018.07.012.

103. Patil, A., Kinoshita, K. & Nakamura, H. Hub promiscuity in protein-protein interaction networks. *Int J Mol Sci* **11**, 1930–1943 (2010).

104. Alcaraz, N. *et al.* De novo pathway-based biomarker identification. *Nucleic Acids Res.* **45**, e151 (2017).

105. Sebestyén, E., Zawisza, M. & Eyras, E. Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. *Nucleic Acids Res.* **43**, 1345–1356 (2015).

106. Will, T. & Helms, V. PPIXpress: construction of condition-specific protein interaction networks based on transcript expression. *Bioinformatics* **32**, 571–578 (2016).

107. Ghadie, M. A., Lambourne, L., Vidal, M. & Xia, Y. Domain-based prediction of the human isoform interactome provides insights into the functional impact of alternative splicing. *PLoS Comput. Biol.* **13**, e1005717 (2017).

108. Zhu, J., Shi, Z., Wang, J. & Zhang, B. Empowering biologists with multi-omics data: colorectal cancer as a paradigm. *Bioinformatics* **31**, 1436–1443 (2015).

109. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biology* **18**, 83 (2017).

110. Han, S. *et al.* CAS-viewer: web-based tool for splicing-guided integrative analysis of multi-omics cancer data. *BMC Med Genomics* **11**, 25 (2018).

111. Danielsson, F. *et al.* Majority of differentially expressed genes are down-regulated during malignant transformation in a four-stage model. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6853–6858 (2013).

112. Duijf, P. H. G., Schultz, N. & Benezra, R. Cancer cells preferentially lose small chromosomes. *Int. J. Cancer* **132**, 2316–2326 (2013).

113. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, (2017).

114. Anglani, R. *et al.* Loss of connectivity in cancer co-expression networks. *PLoS ONE* **9**, e87075 (2014).

115. Cordero, D. *et al.* Large differences in global transcriptional regulatory programs of normal and tumor colon cells. *BMC Cancer* **14**, 708 (2014).

116. Climente-González, H., Porta-Pardo, E., Godzik, A. & Eyras, E. The Functional Impact of Alternative Splicing in Cancer. *Cell Rep* **20**, 2215–2226 (2017).

117. Edfors, F. *et al.* Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol Syst Biol* **12**, (2016).

118. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).

119. Merid, S. K., Goranskaya, D. & Alexeyenko, A. Distinguishing between driver and passenger mutations in individual cancer genomes by network enrichment analysis. *BMC Bioinformatics* **15**, 308 (2014).

120. Tokheim, C. J., Papadopoulos, N., Kinzler, K. W., Vogelstein, B. & Karchin, R. Evaluating the evaluation of cancer driver genes. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14330–14335 (2016).

121. Chen, Y., Zhang, L. & Jones, K. A. SKIP counteracts p53-mediated apoptosis via selective regulation of p21Cip1 mRNA splicing. *Genes Dev.* **25**, 701–716 (2011).

122. Chan, K.-M. *et al.* The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes Dev.* **27**, 985–990 (2013).

123. Liu, G. *et al.* High SKIP expression is correlated with poor prognosis and cell proliferation of hepatocellular carcinoma. *Med. Oncol.* **30**, 537 (2013).

124. Ramakrishnan, S., Ellis, L. & Pili, R. Histone modifications: implications in renal cell carcinoma. *Epigenomics* **5**, 453–462 (2013).

125. Wang, X. *et al.* Clinical and prognostic relevance of EZH2 in breast cancer: A meta-analysis. *Biomed. Pharmacother.* **75**, 218–225 (2015).

126. Wang, Y. *et al.* Overexpression of YB1 and EZH2 are associated with cancer metastasis and poor prognosis in renal cell carcinomas. *Tumour Biol.* **36**, 7159–7166 (2015).

127. Yan, K.-S. *et al.* EZH2 in Cancer Progression and Potential Application in Cancer Therapy: A Friend or Foe? *Int J Mol Sci* **18**, (2017).

128. Poornima, P., Kumar, J. D., Zhao, Q., Blunder, M. & Efferth, T. Network pharmacology of cancer: From understanding of complex interactomes to the design of multi-target specific therapeutics from nature. *Pharmacological Research* **111**, 290–302 (2016).

129. Engin, H. B., Kreisberg, J. F. & Carter, H. Structure-Based Analysis Reveals Cancer Missense Mutations Target Protein Interaction Interfaces. *PLoS ONE* **11**, e0152929 (2016).

130. Jubb, H. C. *et al.* Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. *Prog. Biophys. Mol. Biol.* **128**, 3–13 (2017).

131. Latysheva, N. S. *et al.* Molecular Principles of Gene Fusion Mediated Rewiring of Protein Interaction Networks in Cancer. *Mol Cell* **63**, 579–592 (2016).

132. Zhang, X.-F. *et al.* Comparative analysis of housekeeping and tissue-specific driver nodes in human protein interaction networks. *BMC Bioinformatics* **17**, (2016).

133. Kim, J., Kim, I., Han, S. K., Bowie, J. U. & Kim, S. Network rewiring is an important mechanism of gene essentiality change. *Scientific Reports* **2**, 900 (2012).

134. Jurca, G. *et al.* Integrating text mining, data mining, and network analysis for identifying genetic breast cancer trends. *BMC Res Notes* **9**, 236 (2016).

135. Nilsson, G. & Kannius-Janson, M. Forkhead Box F1 promotes breast cancer cell migration by upregulating lysyl oxidase and suppressing Smad2/3 signaling. *BMC Cancer* **16**, (2016).

136. Fahey, J. M. & Girotti, A. W. NITRIC OXIDE-MEDIATED RESISTANCE TO PHOTODYNAMIC THERAPY IN A HUMAN BREAST TUMOR XENOGRAFT MODEL: IMPROVED OUTCOME WITH NOS2 INHIBITORS. *Nitric Oxide* **62**, 52–61 (2017).

137. Mao, X. *et al.* The heparan sulfate sulfotransferase 3-OST3A (HS3ST3A) is a novel tumor regulator and a prognostic marker in breast cancer. *Oncogene* **35**, 5043–5055 (2016).

138. Tian, H., Li, L., Liu, X.-X. & Zhang, Y. Antitumor Effect of Antisense Ornithine Decarboxylase Adenovirus on Human Lung Cancer Cells. *Acta Biochimica et Biophysica Sinica* **38**, 410–416 (2006).

139. Kumar, G. *et al.* Simultaneous targeting of 5-LOX-COX and ODC block NNK-induced lung adenoma progression to adenocarcinoma in A/J mice. *Am J Cancer Res* **6**, 894–909 (2016).

140. Huang, R. S., Duan, S., Kistner, E. O., Hartford, C. M. & Dolan, M. E. Genetic variants associated with carboplatin-induced cytotoxicity in cell lines derived from Africans. *Mol Cancer Ther* **7**, 3038–3046 (2008).

141. Liu, P. *et al.* Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis* **33**, 1270–1276 (2012).

142. Green, W. J. *et al.* KI67 and DLX2 predict increased risk of metastasis formation in prostate cancer-a targeted molecular approach., KI67 and DLX2 predict increased risk of

metastasis formation in prostate cancer–a targeted molecular approach. *Br J Cancer* **115, 115**, 236, 236–242 (2016).

143. Sturm, I. *et al.* Loss of the tissue-specific proapoptotic BH3-only protein Nbk/Bik is a unifying feature of renal cell carcinoma. *Cell Death and Differentiation* **13**, 619–627 (2006).

144. Jiang, Z. *et al.* Oncofetal protein IMP3. *Cancer* **112**, 2676–2682 (2008).

145. Anasa, V. V., Ravanan, P. & Talwar, P. Multifaceted roles of ASB proteins and its pathological significance. *Front. Biol.* (2018) doi:10.1007/s11515-018-1506-2.

146. Prestin, K. *et al.* Modulation of expression of the nuclear receptor NR0B2 (small heterodimer partner 1) and its impact on proliferation of renal carcinoma cells. *Onco Targets Ther* **9**, 4867–4878 (2016).

147. Kim, C. H., Neiswender, H., Baik, E. J., Xiong, W. C. & Mei, L. Beta-catenin interacts with MyoD and regulates its transcription activity. *Mol Cell Biol* **28**, 2941–2951 (2008).

148. Kagey, M. H. & He, X. Rationale for targeting the Wnt signalling modulator Dickkopf-1 for oncology., Rationale for targeting the Wnt signalling modulator Dickkopf-1 for oncology. *Br J Pharmacol* **174, 174**, 4637, 4637–4650 (2017).

149. Hirata, H. *et al.* Wnt antagonist DKK1 acts as a tumor suppressor gene that induces apoptosis and inhibits proliferation in human renal cell carcinoma. *Int J Cancer* **128**, 1793–1803 (2011).

150. Méniel, V. *et al.* Cited1 deficiency suppresses intestinal tumorigenesis., Cited1 Deficiency Suppresses Intestinal Tumorigenesis. *PLoS Genet* **9, 9**, e1003638–e1003638 (2013).

151. Zheng, Y., Zhou, J. & Tong, Y. Gene signatures of drug resistance predict patient survival in colorectal cancer., Gene signatures of drug resistance predict patient survival in colorectal cancer. *Pharmacogenomics J* **15, 15**, 135, 135–143 (2015).

152. Dart, A. E. *et al.* PAK4 promotes kinase-independent stabilization of RhoU to modulate cell adhesion. *J Cell Biol* **211**, 863–879 (2015).

153. Liggins, A. P., Lim, S. H., Soilleux, E. J., Pulford, K. & Banham, A. H. A panel of cancer-testis genes exhibiting broad-spectrum expression in haematological malignancies., A panel of cancer-testis genes exhibiting broad-spectrum expression in haematological malignancies. *Cancer Immun* **10, 10**, 8–8 (2010).

154. Freitas, M. *et al.* Expression of cancer/testis antigens is correlated with improved survival in glioblastoma. *Oncotarget* **4**, 636–646 (2013).

155. Yamauchi, K. *et al.* Disease-causing mutant WNK4 increases paracellular chloride permeability and phosphorylates claudins. *Proc Natl Acad Sci U S A* **101**, 4690–4694 (2004).

156. Singh, A. B., Sharma, A. & Dhawan, P. Claudin family of proteins and cancer: an overview., Claudin Family of Proteins and Cancer: An Overview. *J Oncol* **2010, 2010**, 541957–541957 (2010).

157. Namani, A., Matiur, M. R., Chen, M. & Tang, X. Gene-expression signature regulated by the KEAP1-NRF2-CUL3 axis is associated with a poor prognosis in head and neck squamous cell cancer., Gene-expression signature regulated by the KEAP1-NRF2-CUL3 axis is associated with a poor prognosis in head and neck squamous cell cancer. *BMC Cancer* **18, 18**, 46–46 (2018).

158. Sharma, P. C. & Verma, R. Implication of HSP70 in the Pathogenesis of Gastric Cancer. in *HSP70 in Human Diseases and Disorders* (eds. Asea, A. A. A. & Kaur, P.) vol. 14 113–130 (Springer International Publishing, 2018).

159. Yang, Y. *et al.* MicroRNA-195 acts as a tumor suppressor by directly targeting Wnt3a in HepG2 hepatocellular carcinoma cells. *Molecular Medicine Reports* **10**, 2643–2648 (2014).

160. He, Q. *et al.* Genome-wide prediction of cancer driver genes based on SNP and cancer SNV data. *Am J Cancer Res* **4**, 394–410 (2014).

161. Monteiro, F. L. *et al.* The histone H2A isoform Hist2h2ac is a novel regulator of proliferation and epithelial-mesenchymal transition in mammary epithelial and in breast cancer cells. *Cancer Lett.* **396**, 42–52 (2017).

162. Chang, I. *et al.* Hrk mediates 2-methoxyestradiol-induced mitochondrial apoptotic signaling in prostate cancer cells. *Mol. Cancer Ther.* **12**, 1049–1059 (2013).

163. Adams, J. M. & Cory, S. The BCL-2 arbiters of apoptosis and their growing role as cancer targets. *Cell Death Differ.* **25**, 27–36 (2018).

164. Hsia, D. A. *et al.* KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. *PNAS* **107**, 9671–9676 (2010).

165. Lee, J.-C., Liang, C.-W. & Fletcher, C. D. Giant cell tumor of soft tissue is genetically distinct from its bone counterpart. *Modern Pathology* **30**, 728–733 (2017).

166. Ramsköld, D., Wang, E. T., Burge, C. B. & Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* **5**, e1000598 (2009).

167. Yu, N. Y.-L. *et al.* Complementing tissue characterization by integrating transcriptome profiling from the Human Protein Atlas and from the FANTOM5 consortium. *Nucleic Acids Res.* **43**, 6787–6798 (2015).

168. Uhlén, M. *et al.* Transcriptomics resources of human tissues and organs. *Mol. Syst. Biol.* **12**, 862 (2016).

169. Davidson, S. M. & Heiden, M. G. V. Critical Functions of the Lysosome in Cancer Biology. *Annual Review of Pharmacology and Toxicology* **57**, 481–507 (2017).

170. Mirzaei, H. & Faghihloo, E. Viruses as key modulators of the TGF-β pathway; a double-edged sword involved in cancer. *Rev. Med. Virol.* **28**, (2018).

171. Mosesson, Y., Mills, G. B. & Yarden, Y. Derailed endocytosis: an emerging feature of cancer. *Nature Reviews Cancer* **8**, 835–850 (2008).

172. Kingston, D. *et al.* Inhibition of retromer activity by herpesvirus saimiri tip leads to CD4 downregulation and efficient T cell transformation. *J. Virol.* **85**, 10627–10638 (2011).

173. Rajagopalan, D. & Jha, S. An epi(c)genetic war: Pathogens, cancer and human genome. *Biochim. Biophys. Acta* **1869**, 333–345 (2018).

174. Kim, J., Yao, F., Xiao, Z., Sun, Y. & Ma, L. MicroRNAs and metastasis: small RNAs play big roles. *Cancer Metastasis Rev.* **37**, 5–15 (2018).

175. Dhillon, A. S., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26**, 3279–3290 (2007).

176. Muthuswamy, S. K. & Xue, B. Cell Polarity As A Regulator of Cancer Cell Behavior Plasticity. *Annu Rev Cell Dev Biol* **28**, 599–625 (2012).

177. Jørgensen, J. T. A paradigm shift in biomarker guided oncology drug development. *Annals of Translational Medicine* **7**, (2019).

178. Liu, Q., Dai, S.-J., Li, H., Dong, L. & Peng, Y.-P. Silencing of NUF2 inhibits tumor growth and induces apoptosis in human hepatocellular carcinomas. *Asian Pac. J. Cancer Prev.* **15**, 8623–8629 (2014).

179. Hu, P., Shangguan, J. & Zhang, L. Downregulation of NUF2 inhibits tumor growth and induces apoptosis by regulating lncRNA AF339813. *Int J Clin Exp Pathol* **8**, 2638–2648 (2015).

180. Ardini, E. *et al.* Entrectinib, a Pan-TRK, ROS1, and ALK Inhibitor with Activity in Multiple Molecularly Defined Cancer Indications. *Mol. Cancer Ther.* **15**, 628–639 (2016).

181. Jin, Z., Kotera, M. & Goto, S. Virus proteins similar to human proteins as possible disturbance on human pathways. *Syst Synth Biol* **8**, 283–295 (2014).

182. Zhao, M., Kim, P., Mitra, R., Zhao, J. & Zhao, Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res* **44**, D1023–D1031 (2016).

183. Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intra-tumor heterogeneity. *Nat Med* **22**, 105–113 (2016).

184. Reiter, J. G. *et al.* Minimal functional driver gene heterogeneity among untreated metastases. *Science* **361**, 1033–1037 (2018).

185. Vandin, F. Computational Methods for Characterizing Cancer Mutational Heterogeneity. *Front Genet* **8**, (2017).

186. Martinez-Ledesma, E., Verhaak, R. G. W. & Treviño, V. Identification of a multi-cancer gene expression biomarker for cancer clinical outcomes using a network-based algorithm. *Scientific Reports* **5**, 11966 (2015).

187. Yuan, Y. *et al.* Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat. Biotechnol.* **32**, 644–652 (2014).

188. Yamada, R. *et al.* Preferential expression of cancer/testis genes in cancer stem-like cells: proposal of a novel sub-category, cancer/testis/stem gene. *Tissue Antigens* **81**, 428–434 (2013).

189. Mjelle, R. *et al.* Cell cycle regulation of human DNA repair and chromatin remodeling genes. *DNA Repair* **30**, 53–67 (2015).

190. Knijnenburg, T. A. *et al.* Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep* **23**, 239-254.e6 (2018).

191. Arimura, Y. *et al.* Crystal structure and stable property of the cancer-associated heterotypic nucleosome containing CENP-A and H3.3. *Sci Rep* **4**, (2014).

192. Athwal, R. K. *et al.* CENP-A nucleosomes localize to transcription factor hotspots and subtelomeric sites in human cancer cells. *Epigenetics & Chromatin* **8**, 2 (2015).

193. Coene, K. L. M. *et al.* The ciliopathy-associated protein homologs RPGRIP1 and RPGRIP1L are linked to cilium integrity through interaction with Nek4 serine/threonine kinase. *Hum Mol Genet* **20**, 3592–3605 (2011).

194. Gerhardt, C., Leu, T., Lier, J. M. & Rüther, U. The cilia-regulated proteasome and its role in the development of ciliopathies and cancer. *Cilia* **5**, (2016).

195. Xu, H. *et al.* Silencing of KIF14 interferes with cell cycle progression and cytokinesis by blocking the p27(Kip1) ubiquitination pathway in hepatocellular carcinoma. *Exp. Mol. Med.* **46**, e97 (2014).

196. Siddam, A. D. *et al.* The RNA-binding protein Celf1 post-transcriptionally regulates p27Kip1 and Dnase2b to control fiber cell nuclear degradation in lens development. *PLOS Genetics* **14**, e1007278 (2018).

197. Chu, I. M., Hengst, L. & Slingerland, J. M. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nature Reviews Cancer* **8**, 253–267 (2008).

198. Wu, L. & Russell, P. Nim1 kinase promotes mitosis by inactivating Wee1 tyrosine kinase. *Nature* **363**, 738–741 (1993).

199. Sobol, A., Galluzzo, P., Weber, M. J., Alani, S. & Bocchetta, M. Depletion of Amyloid Precursor Protein (APP) causes G0 arrest in non-small cell lung cancer (NSCLC) cells. *J. Cell. Physiol.* **230**, 1332–1341 (2015).

200. Yin, Y. *et al.* Wee1 inhibition can suppress tumor proliferation and sensitize p53 mutant colonic cancer cells to the anticancer effect of irinotecan. *Molecular Medicine Reports* **17**, 3344–3349 (2018).

201. Wang, Y. *et al.* Loss of expression and prognosis value of alpha-internexin in gastroenteropancreatic neuroendocrine neoplasm. *BMC Cancer* **18**, 691 (2018).

202. Kirikoshi, H. & Katoh, M. Expression of WNT7A in human normal tissues and cancer, and regulation of WNT7A and WNT7B in human cancer. *Int. J. Oncol.* **21**, 895–900 (2002).

203. Chen, J. *et al.* Up-regulation of Wnt7b rather than Wnt1, Wnt7a, and Wnt9a indicates poor prognosis in breast cancer. 11.

204. Souza-Santos, P. T. de *et al.* Mutations, Differential Gene Expression, and Chimeric Transcripts in Esophageal Squamous Cell Carcinoma Show High Heterogeneity. *Transl Oncol* **11**, 1283–1291 (2018).

205. Guest, R. V., Boulter, L., Dwyer, B. J. & Forbes, S. J. Understanding liver regeneration to bring new insights to the mechanisms driving cholangiocarcinoma. *npj Regenerative Medicine* **2**, 13 (2017).

206. Boulter, L. *et al.* WNT signaling drives cholangiocarcinoma growth and can be pharmacologically inhibited. *J Clin Invest* **125**, 1269–1285 (2015).

207. Peng, W. *et al.* Loss of PTEN promotes resistance to T cell-mediated immunotherapy. *Cancer Discov* **6**, 202–216 (2016).

208. Bäumer, N. *et al.* Inhibitor of Cyclin-dependent Kinase (CDK) Interacting with Cyclin A1 (INCA1) Regulates Proliferation and Is Repressed by Oncogenic Signaling. *J Biol Chem* **286**, 28210–28222 (2011).

209. Zenner, H. P., Pfister, M., Friese, N., Zrenner, E. & Röcken, M. Molekulare personalisierte Medizin. *HNO* **62**, 520–524 (2014).

210. Galoczova, M., Coates, P. & Vojtesek, B. STAT3, stem cells, cancer stem cells and p63. *Cell. Mol. Biol. Lett.* **23**, 12 (2018).

211. Goenawan, I. H., Bryan, K. & Lynn, D. J. DyNet: visualization and analysis of dynamic molecular interaction networks. *Bioinformatics* **32**, 2713–2715 (2016).

212. Fant, M., Farina, A., Nagaraja, R. & Schlessinger, D. PLAC1 (Placenta-specific 1): A novel, X-linked gene with roles in reproductive and cancer biology. *Prenat Diagn* **30**, 497–502 (2010).

213. Närvä, E. *et al.* RNA-Binding Protein L1TD1 Interacts with LIN28 via RNA and is Required for Human Embryonic Stem Cell Self-Renewal and Cancer Cell Proliferation. *Stem Cells* **30**, 452–460 (2012).

214. Piñero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833–D839 (2017).

215. Li, X. Dynamic changes of driver genes' mutations across clinical stages in nine cancer types. *Cancer Med* **5**, 1556–1565 (2016).

216. Shi, X. *et al.* CyNetSVM: A Cytoscape App for Cancer Biomarker Identification Using Network Constrained Support Vector Machines. *PLoS ONE* **12**, e0170482 (2017).

217. Fouad, Y. A. & Aanei, C. Revisiting the hallmarks of cancer. *Am J Cancer Res* **7**, 1016–1036 (2017).

218. Pfoh, R., Lacdao, I. K. & Saridakis, V. Deubiquitinases and the new therapeutic opportunities offered to cancer. *Endocr Relat Cancer* **22**, T35–T54 (2015).

219. Lee, O.-H. *et al.* Role of the focal adhesion protein TRIM15 in colon cancer development. *Biochim. Biophys. Acta* **1853**, 409–421 (2015).

220. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

221. Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* (2015) doi:10.1093/nar/gkv1507.

222. The Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175 (2017).

223. Chatr-Aryamontri, A. *et al.* The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* **43**, D470-478 (2015).

224. Bateman, A. *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158–D169 (2017).

225. Wang, M. *et al.* PaxDb, a database of protein abundance averages across all three domains of life. *Mol. Cell Proteomics* **11**, 492–500 (2012).

226. Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509**, 575–581 (2014).

227. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **13**, 2498–2504 (2003).

228. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).

229. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).

230. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **28**, 1–26 (2008).

231. Alexa, A. & Rahnenfuhrer, J. Gene set enrichment analysis with topGO. 26.

232. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).

233. Timmons, J. A., Szkop, K. J. & Gallagher, I. J. Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome Biology* **16**, 186 (2015).

234. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **16**, 284–287 (2012).

235. Huang, D. W. *et al.* DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* **35**, W169-175 (2007).

236. Patel, V. N. *et al.* Network Signatures of Survival in Glioblastoma Multiforme. *PLOS Computational Biology* **9**, e1003237 (2013).

237. Cao, Z. & Zhang, S. An integrative and comparative study of pan-cancer transcriptomes reveals distinct cancer common and specific signatures. *Sci Rep* **6**, (2016).

238. Aguirre-Gamboa, R. *et al.* SurvExpress: An Online Biomarker Validation Tool and Database for Cancer Gene Expression Data Using Survival Analysis. *PLoS One* **8**, (2013).

239. Zou, K. H., Fielding, J. R., Silverman, S. G. & Tempany, C. M. C. Hypothesis Testing I: Proportions. *Radiology* **226**, 609–613 (2003).

240. Gobbi, A. *et al.* Fast randomization of large genomic datasets while preserving alteration counts. *Bioinformatics* **30**, i617-623 (2014).

241. Dexter, F. Wilcoxon-Mann-Whitney test used for data that are not normally distributed. *Anesth. Analg.* **117**, 537–538 (2013).

242. Yates, F. Contingency Tables Involving Small Numbers and the $\chi 2$ Test. *Supplement to the Journal of the Royal Statistical Society* **1**, 217–235 (1934).

243. Nagahashi, M. *et al.* Genomic landscape of colorectal cancer in Japan: clinical implications of comprehensive genomic sequencing for precision medicine. *Genome Med* **8**, (2016).

244. Song, P., Kwon, Y., Joo, J.-Y., Kim, D.-G. & Yoon, J. H. Secretomics to Discover Regulators in Diseases. *Int J Mol Sci* **20**, (2019).

245. Georgiou, H. M., Rice, G. E. & Baker, M. S. Proteomic analysis of human plasma: Failure of centrifugal ultrafiltration to remove albumin and other high molecular weight proteins. *PROTEOMICS* **1**, 1503–1506 (2001).

246. Brandi, J. *et al.* Proteomic approaches to decipher cancer cell secretome. *Seminars in Cell & Developmental Biology* **78**, 93–101 (2018).

247. Warmoes, M. *et al.* Secretome proteomics reveals candidate non-invasive biomarkers of BRCA1 deficiency in breast cancer. *Oncotarget* **7**, 63537–63548 (2016).

248. Farhan, H. & Rabouille, C. Signalling to and from the secretory pathway. *J. Cell. Sci.* **124**, 171–180 (2011).

249. Meissner, F., Scheltema, R. A., Mollenkopf, H.-J. & Mann, M. Direct proteomic quantification of the secretome of activated immune cells. *Science* **340**, 475–478 (2013).

250. Megger, D. A., Bracht, T., Meyer, H. E. & Sitek, B. Label-free quantification in clinical proteomics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1834**, 1581–1590 (2013).

251. Chait, B. T. Mass Spectrometry: Bottom-Up or Top-Down? *Science* **314**, 65–66 (2006).

252. Zhang, Z., Wu, S., Stenoien, D. L. & Paša-Tolić, L. High-throughput proteomics. *Annu Rev Anal Chem (Palo Alto Calif)* **7**, 427–454 (2014).

253. Floris, F. *et al.* Bottom-Up Two-Dimensional Electron-Capture Dissociation Mass Spectrometry of Calmodulin. *J. Am. Soc. Mass Spectrom.* **29**, 207–210 (2018).

254. Patrie, S. M. Top-Down Mass Spectrometry: Proteomics to Proteoforms. *Adv. Exp. Med. Biol.* **919**, 171–200 (2016).

255. Takemori, N. *et al.* Top-down/Bottom-up Mass Spectrometry Workflow Using Dissolvable Polyacrylamide Gels. *Anal. Chem.* **89**, 8244–8250 (2017).

256. Gillet, L. C. *et al.* Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics* **11**, (2012).

257. Koopmans, F., Ho, J. T. C., Smit, A. B. & Li, K. W. Comparative Analyses of Data Independent Acquisition Mass Spectrometric Approaches: DIA, WiSIM-DIA, and Untargeted DIA. *PROTEOMICS* **18**, 1700304 (2018).

258. Wolf, S. A., Boddeke, H. W. G. M. & Kettenmann, H. Microglia in Physiology and Disease. *Annu. Rev. Physiol.* **79**, 619–643 (2017).

259. Vasile, F., Dossi, E. & Rouach, N. Human astrocytes: structure and functions in the healthy brain. *Brain Struct Funct* **222**, 2017–2029 (2017).

260. Kulkarni, A., Chen, J. & Maday, S. Neuronal autophagy and intercellular regulation of homeostasis in the brain. *Curr. Opin. Neurobiol.* **51**, 29–36 (2018).

261. Huang, J. K. & Káradóttir, R. T. Oligodendrocytes in health and disease. *Neuropharmacology* **110**, 537–538 (2016).

262. Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18 Suppl 1**, S96-104 (2002).

263. Zhang, X. *et al.* Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nature Protocols* **13**, 530–550 (2018).

264. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

265. Kammers, K., Cole, R. N., Tiengwe, C. & Ruczinski, I. Detecting significant changes in protein abundance. *EuPA Open Proteomics* **7**, 11–19 (2015).

266. Sharma, K. *et al.* Cell type- and brain region-resolved mouse brain proteome. *Nat. Neurosci.* **18**, 1819–1831 (2015).

267. Sharma, K. *et al.* Cell type- and brain region-resolved mouse brain proteome. *Nat. Neurosci.* **18**, 1819–1831 (2015).

268. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* **37**, 38–44 (2019).

269. Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S. & Trapnell, C. Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nat Commun* **11**, 1–6 (2020).

270. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).

271. Bowman, N. *et al.* Advanced Cell Mapping Visualizations for Single Cell Functional Proteomics Enabling Patient Stratification. *PROTEOMICS* **n/a**, 1900270.

272. Sánchez-Rico, M. & Alvarado, J. M. A Machine Learning Approach for Studying the Comorbidities of Complex Diagnoses. *Behav Sci (Basel)* **9**, (2019).

Supplementary Information Legends

FigureS1: Differences between the number of interactions in the healthy and cancer states for BLCA, BRCA and THCA. Healthy and cancer PPINs significantly differ in size even in the condition specific networks obtained from the randomised PPIN (p-value <0.05). The density plots indicate the distribution of paired cancer and healthy PPIN sizes in BLCA, BRCA and THCA. For BRCA and BLCA, the healthy PPIN was larger than the corresponding cancer PPIN while for THCA, the cancer PPIN was larger than the corresponding healthy PPIN.
FigureS2: The EdgeExplorer portal hosts all the Datasets produced in the study. Due to their size limitations, we stored them in the portal for ease of access.

Figure S3 (A-M): Kaplan-Meier survival analysis plots of multigene cancer biomarkers involved in edgetic perturbations. The x axes indicate the number of days until patient death whereas the y axes indicate the probability of patient survival. In all the figures, the green lines indicate better survival (longer life-span) after cancer diagnosis while the red lines indicate poor survival (shorter life-span) after cancer diagnosis as a result of the proteins involved in edgetic gains or losses. In all the cases, the proteins involved in edgteic perturbations predicted poor survival of the patients (Logrank test p-value < 0.05), indicating their importance in cancer monitoring and prognosis. (i) Overall survival predicted from gene signatures involved in edgetic gains across most patients of a cancer type (except for LIHC), (ii) Overall survival predicted from gene signatures involved in edgetic losses across most patients of a cancer type, (iii) Overall survival predicted from gene signatures involved in edgetic gains across patients showing cancer-specific perturbations, (iv) Overall survival predicted from gene signatures involved in edgetic losses across patients showing cancer-specific perturbations. The names of the prominent proteins with multiple perturbations responsible for the above observations can be found in S4 Table and in EdgeExplorer website.

Figure S4 (a - b): Kaplan-Meier survival analysis plots of multigene cancer biomarkers involved in edgetic perturbations from the randomised PPIN. The x axes indicate the number of days until patient death whereas the y axes indicate the probability of patient survival. In both the figures, the green lines indicate better survival (longer life-span) after cancer diagnosis while the red lines indicate poor survival (shorter life-span) after cancer diagnosis as a result of the proteins involved in edgetic gains or losses. In all the cases, the proteins involved in edgetic perturbations predicted poor survival of the patients (Log-rank test p-value < 0.05), indicating their importance in cancer monitoring and prognosis. (A) Overall survival predicted from gene signatures involved in edgetic gains across most patients in BRCA, (B) Overall survival predicted from gene signatures involved in edgetic gains across most patients in BLCA.

Figure S5A-G: Top ranked features (edges) from the Random Forest algorithm that distinguish cancer types based on the identified groups from hierarchical clustering (Figure 5). The x axes indicate the percentage (%) Mean Squared Error (MSE2). The higher the %MSE of the feature (perturbed edge), the more important the perturbed edge is in identifying a cluster.

Figure S6A-Z: For each plot, the left blue curve represents the lowly-expressed genes while the grey curve represents the highly-expressed genes across patients

of a cancer type for both healthy and cancer samples, respectively. We used these characteristic peaks as a threshold and only kept the genes with an all-samples probability score of greater than 0.8 for subsequent analysis

Supplementary Table Ia: The proportions of edgetic perturbations associated with SMGs and those associated with random genes with a similar degree significantly differ in size.

Supplementary Table Ib: 9 cancer types show a significantly larger proportion of edgetic perturbations associated with SMGs when compared to the proportion of edgetic perturbations associated with random genes with similar degrees.

Supplementary Table Ic: Specific cancer SMGs are involved in edgetic perturbations of cancer PPINs.

Supplementary Table IIa: Importance of the proteins involved in multiple edgetic perturbations or edges frequently perturbed across patients of a cancer type and their significance in predicting overall patient survival

Supplementary Table IIb: Proteins involved in subtype and subtype specific edgetic perturbations (SubtypePercLost and SubtypePercGained) in 11 cancer types.

Supplementary Table III: Table showing a subset of cancer-specific edgetic perturbations in 13 cancer types.

Supplementary Table IV: Table showing the protein-protein interactions (711) between secreted proteins (iSPECS) and those from the cell lysate (Sharma). The proteins are either differentially expressed or their expression is 2.5-fold increased in a cell type compared to the other three cell types (termed cell tyope-specific proteins). As- astrocytes, MI- microglia, Ne- neurons and Ol- oligodendrocytes.

Supplementary Figures and Tables

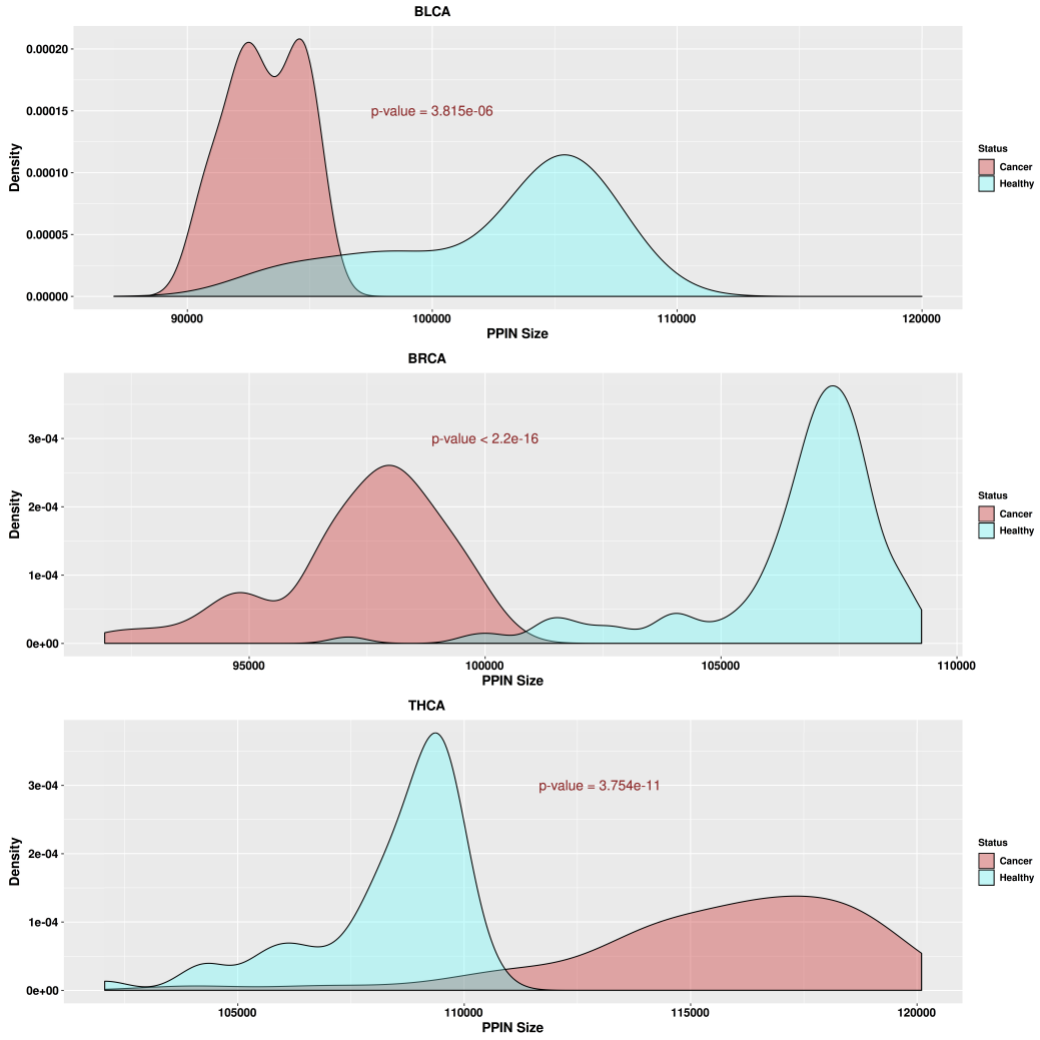


Figure S1: Differences between the number of interactions in the healthy and cancer states for BLCA, BRCA and THCA. Healthy and cancer PPINs significantly differ in size even in the condition specific networks obtained from the randomised PPIN (p-value <0.05). The density plots indicate the distribution of paired cancer and healthy PPIN sizes in BLCA, BRCA and THCA. For BRCA and BLCA, the healthy PPIN was larger than the corresponding cancer PPIN while for THCA, the cancer PPIN was larger than the corresponding healthy PPIN.

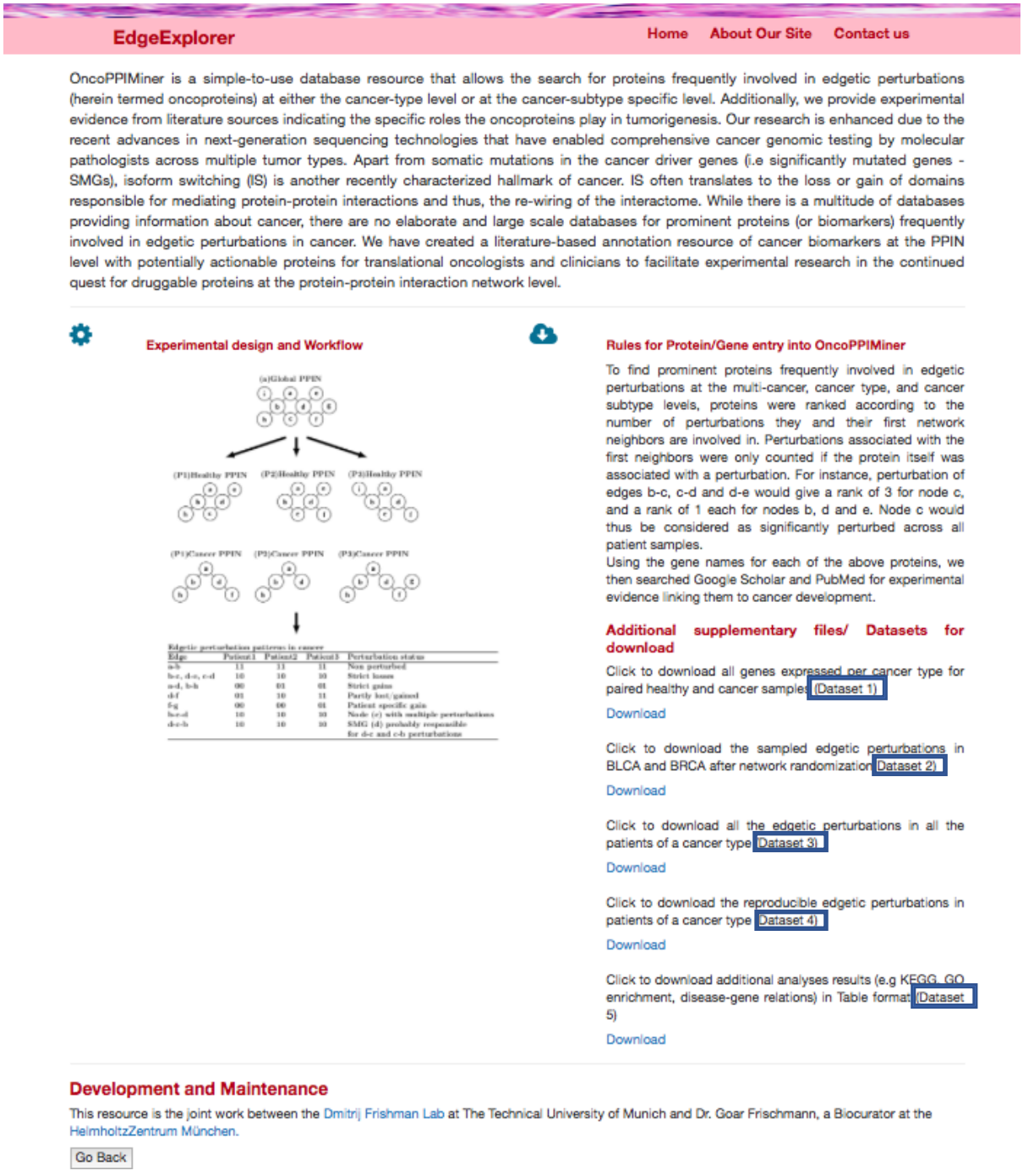


Figure S2: The EdgeExplorer portal hosts all the Datasets (1-5, shown above in blue rectangles) produced in the study. Due to their size limitations, we stored them in the portal for ease of access.

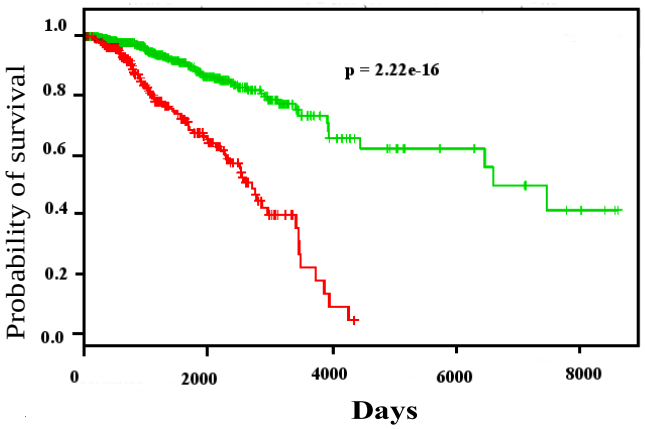


Figure Ai: BRCA Top Gain

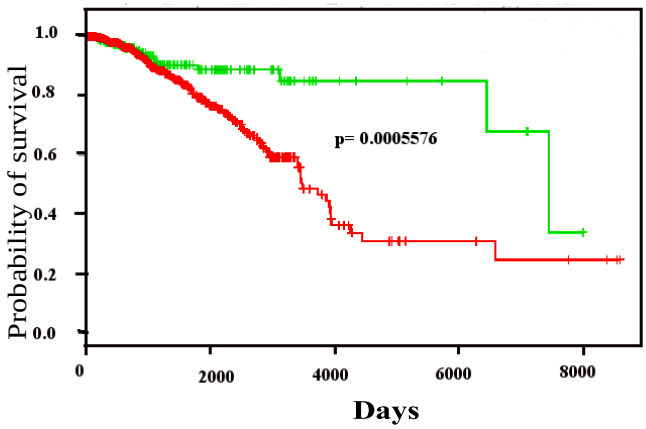


Figure Aii: BRCA Specific Top Gain

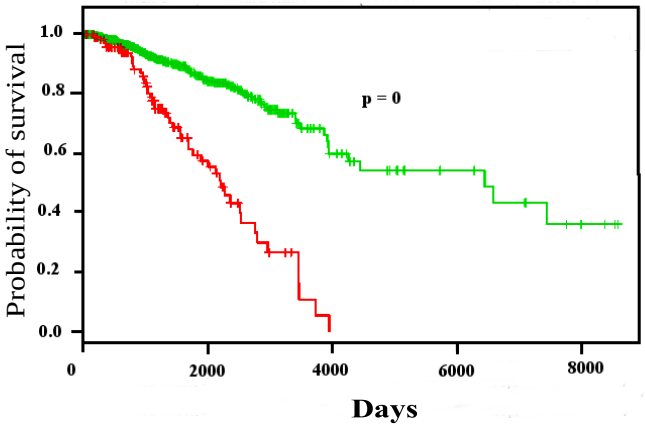


Figure Aiii: BRCA Top Lost

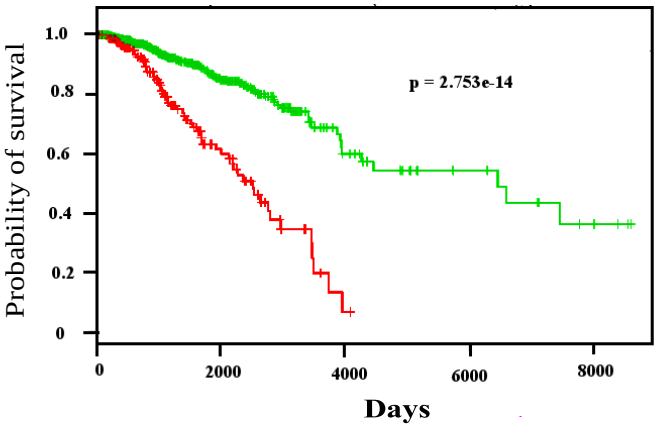


Figure Aiv: BRCA Specific Top Lost

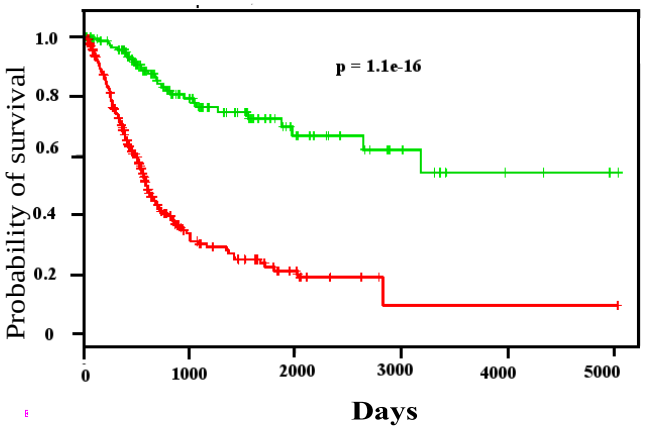


Figure Bi: BLCA Top Gain

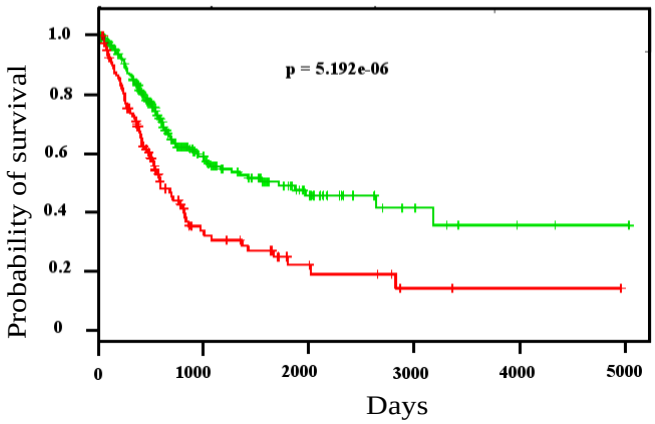


Figure Biii: BLCA Specific Top Gain

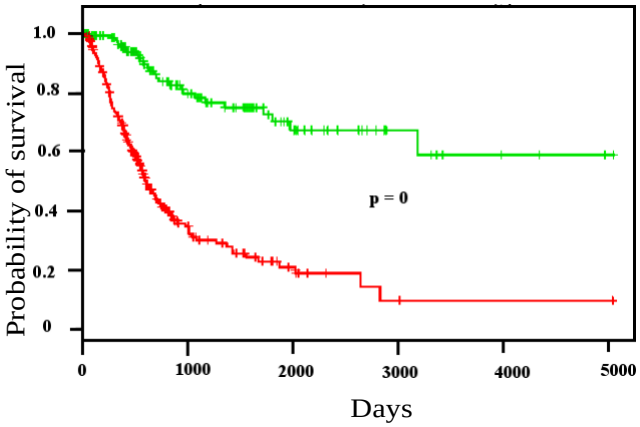


Figure Biii: BLCA Top Lost

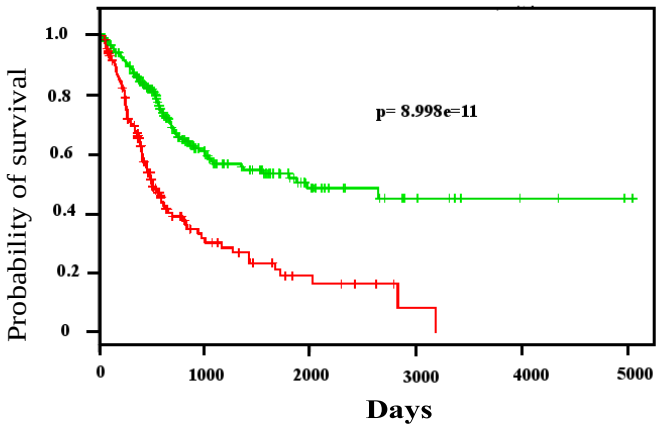


Figure Biv: BLCA Specific Top Lost

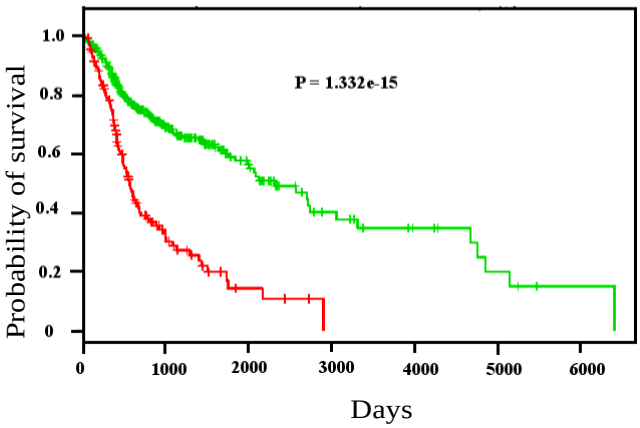


Figure Ci: HNSC Top Gain

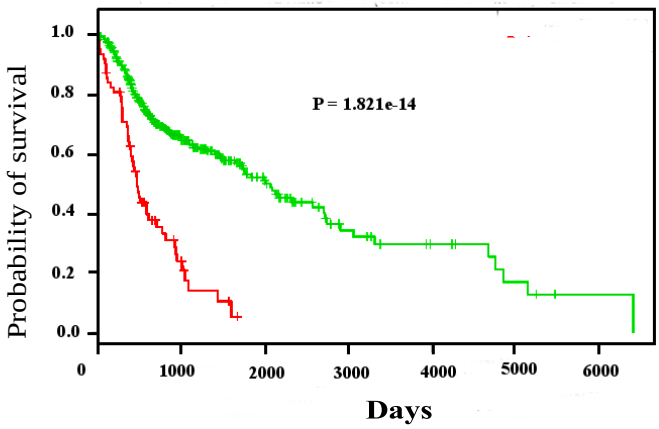


Figure Cii: HNSC Specific Top Gain

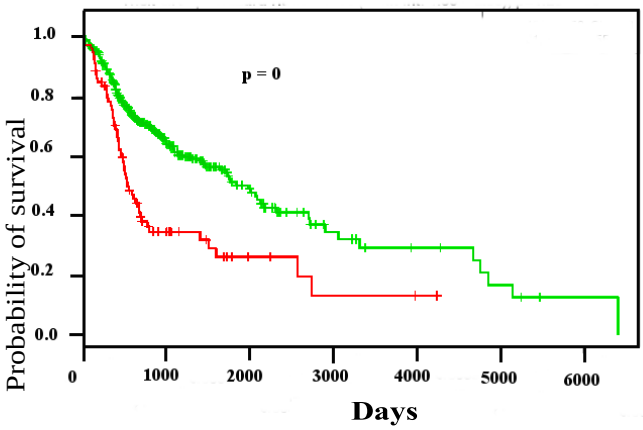


Figure Ciii: HNSC Top Lost

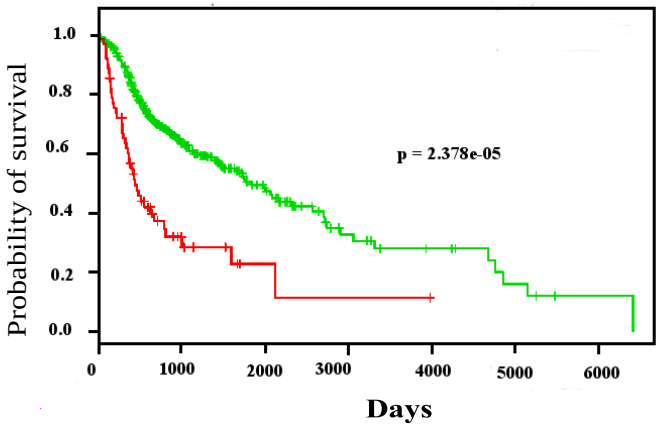


Figure Civ: HNSC Specific Top Lost

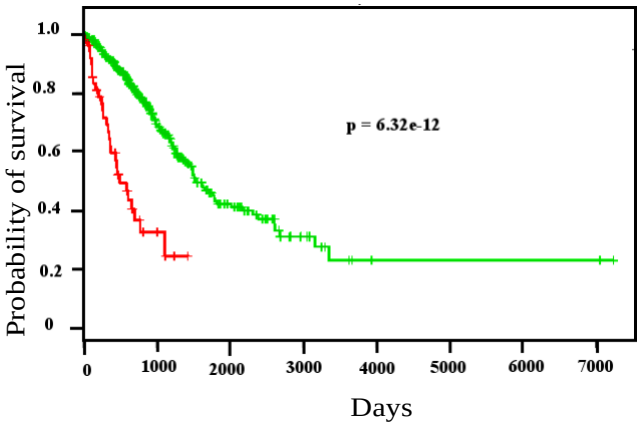
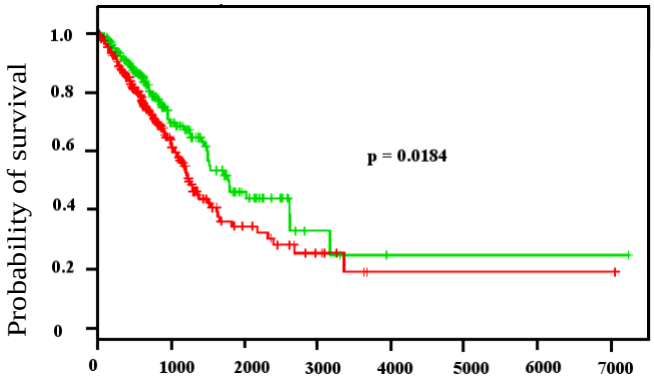


Figure Di: LUAD Top Gain



Days Figure Dii: LUAD Specific Top Gain

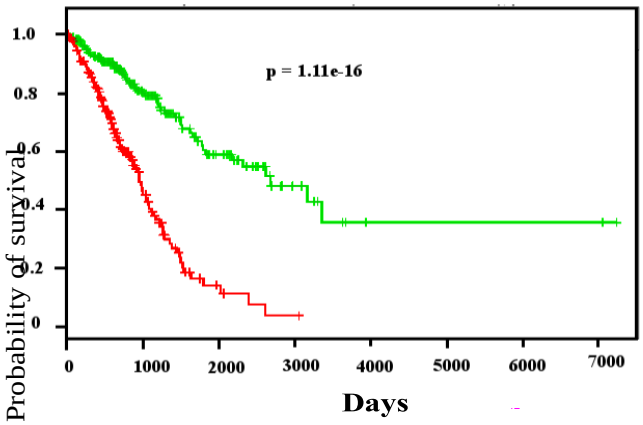


Figure Diii: LUAD Top Lost

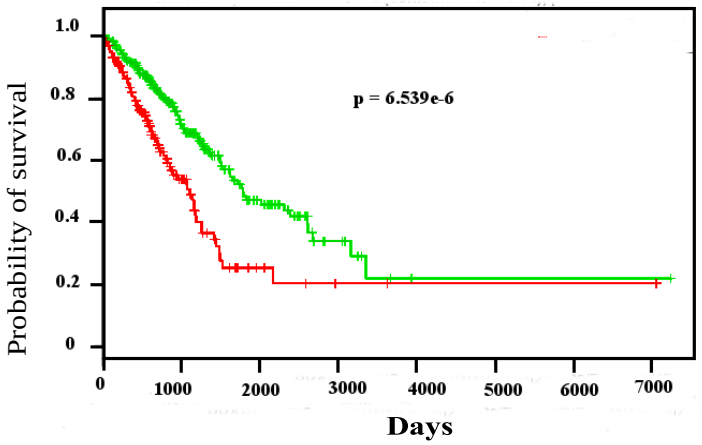


Figure Div: LUAD Specific Top Lost

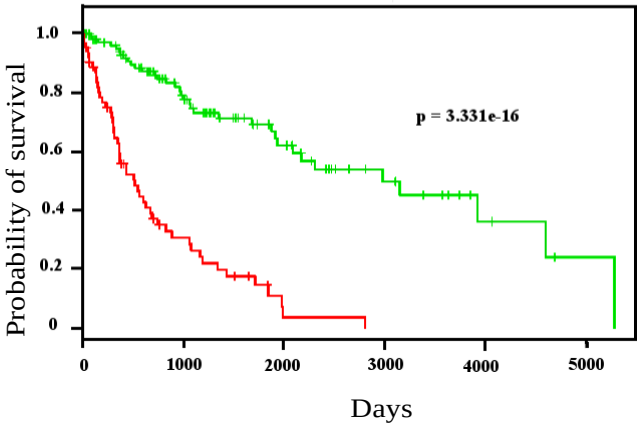


Figure Ei: LUSC Top Gain

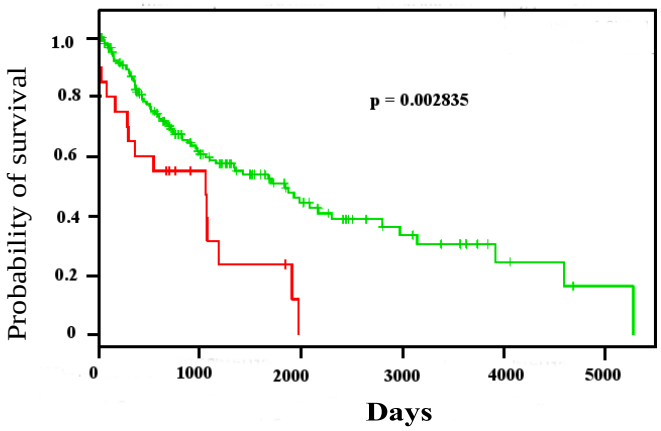


Figure Eii: LUSC Specific Top Gain

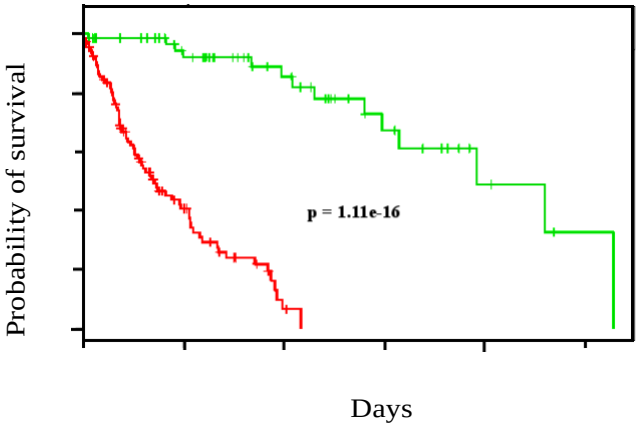


Figure Eiii: LUSC Top Lost

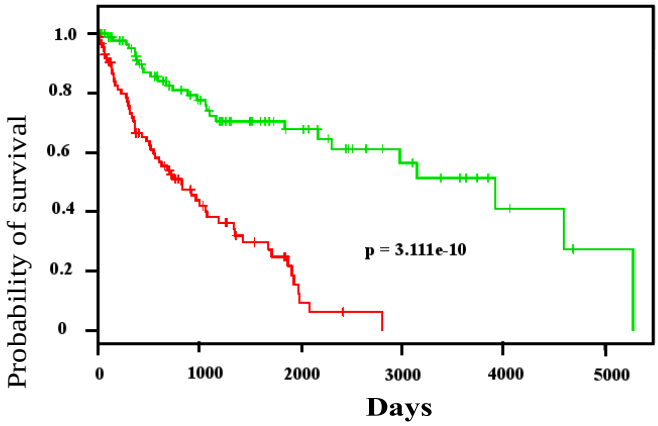


Figure Eiv: LUSC Specific Top Lost

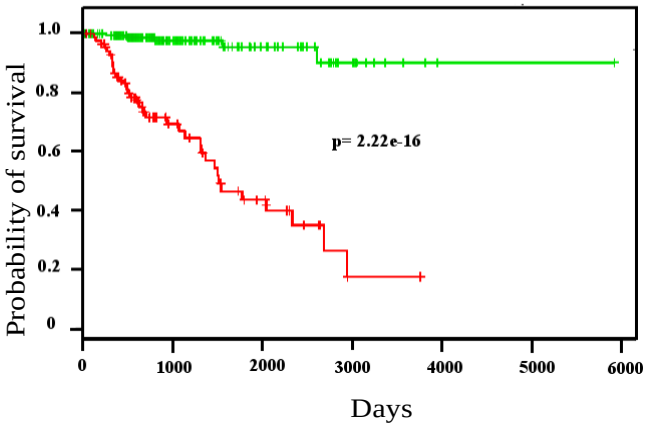


Figure Fi: KIRP Top Gain

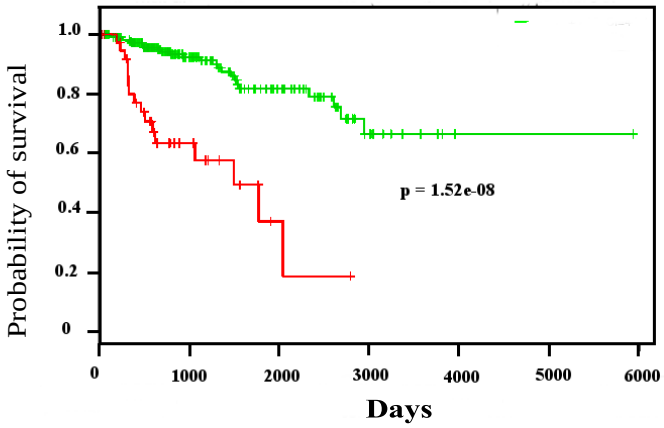


Figure Fii: KIRP Specific Top Gain

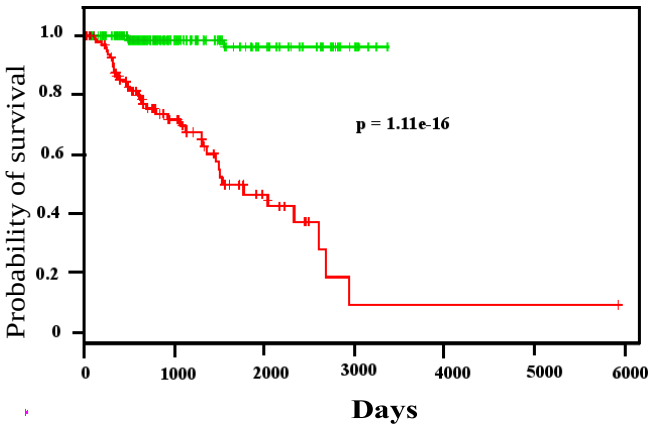


Figure Fiii: KIRP Top Lost

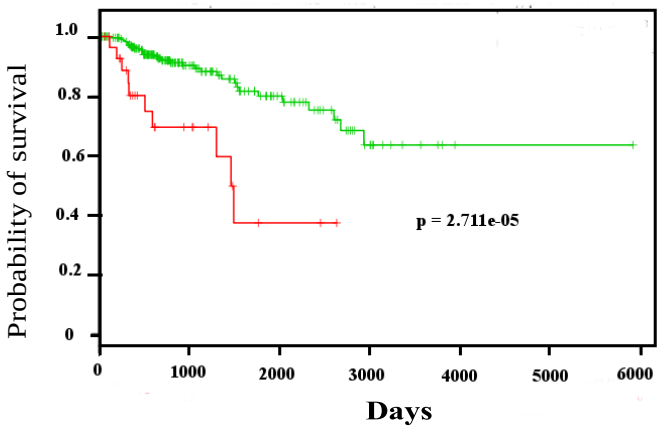


Figure Fiv: KIRP Specific Top Lost

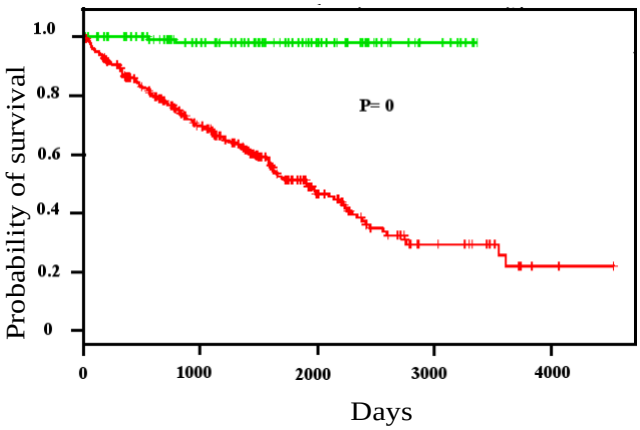


Figure Gi: KIRC Top Gain

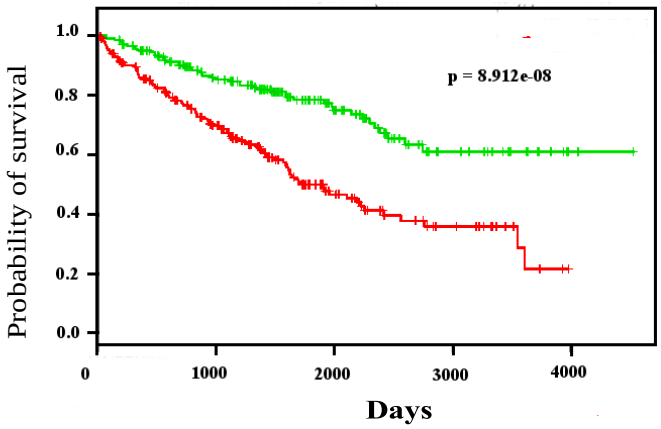


Figure Gii: KIRC Specific Top Gain

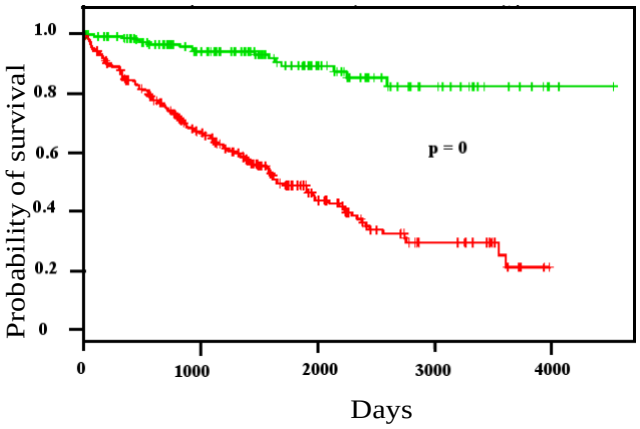


Figure Giii: KIRC Top Lost

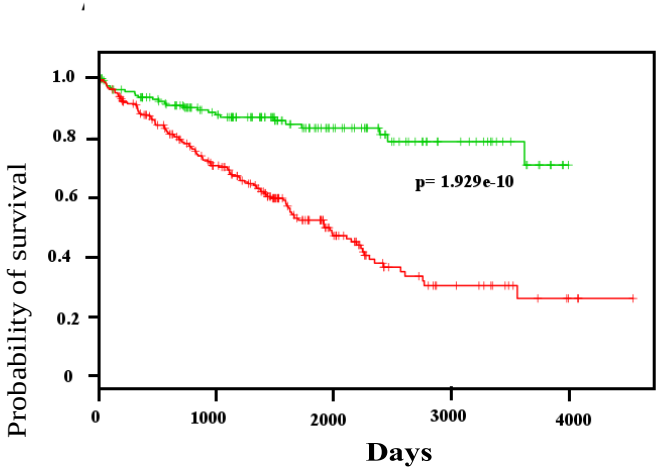


Figure Giv: KIRC Specific Top Lost

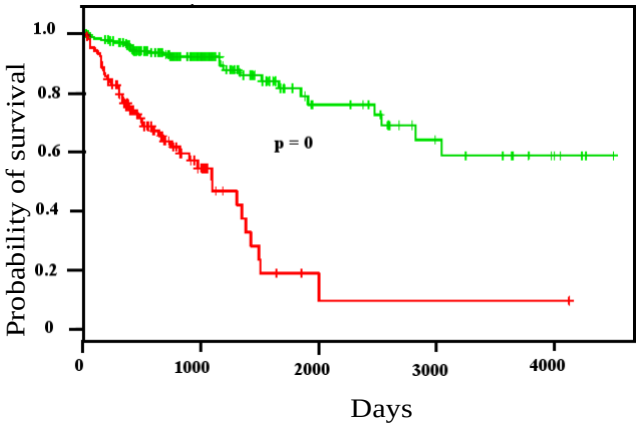


Figure Hi: COAD Top Gain

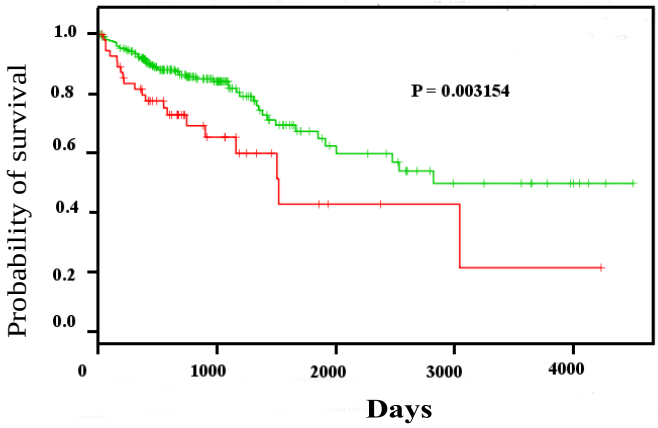


Figure Hii: COAD Specific Top Gain

1

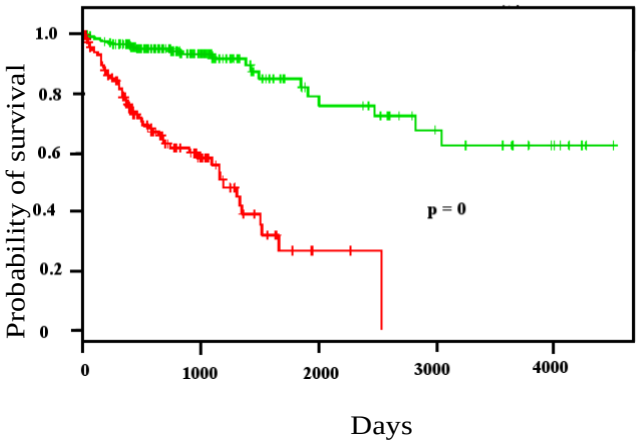


Figure Hiii: COAD Top Lost

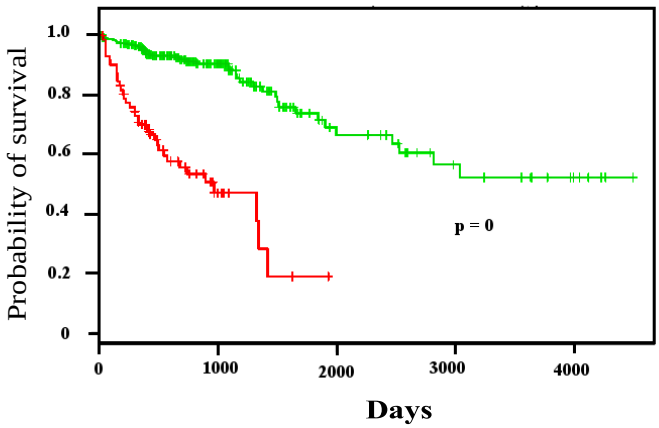


Figure Hiv: COAD Specific Top Lost

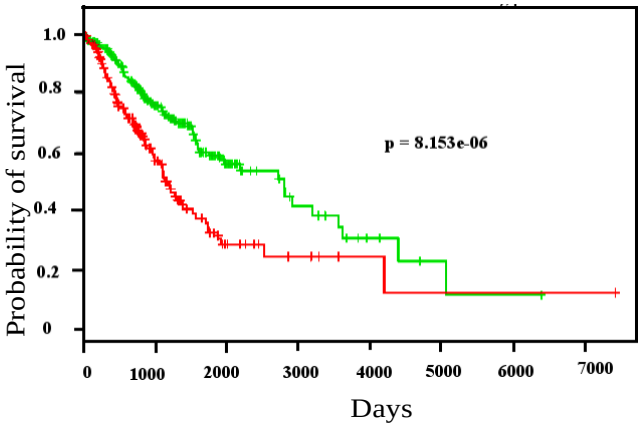


Figure Ii: STES Top Gain

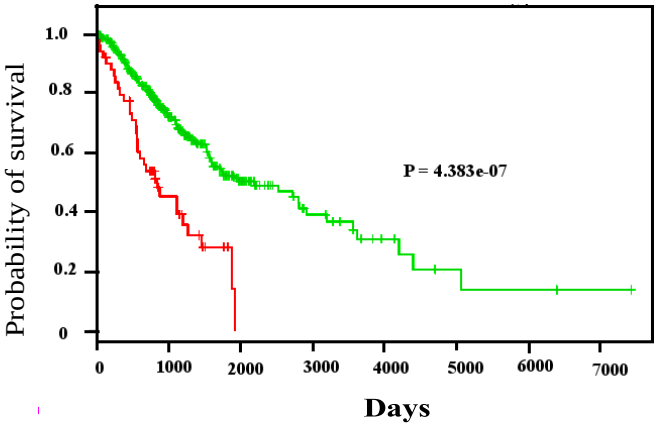


Figure Iii: STES Specific Top Gain

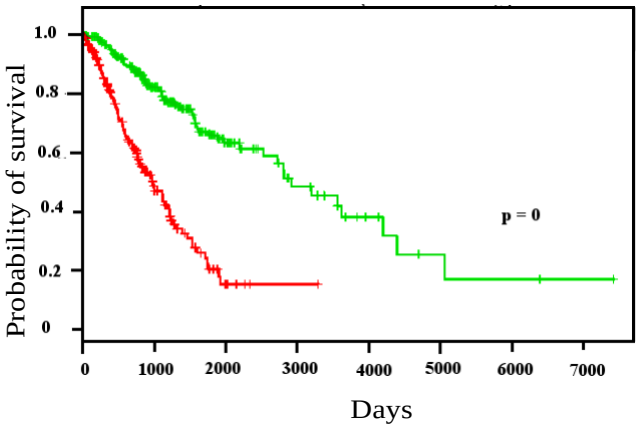


Figure Iiii: STES Top Lost

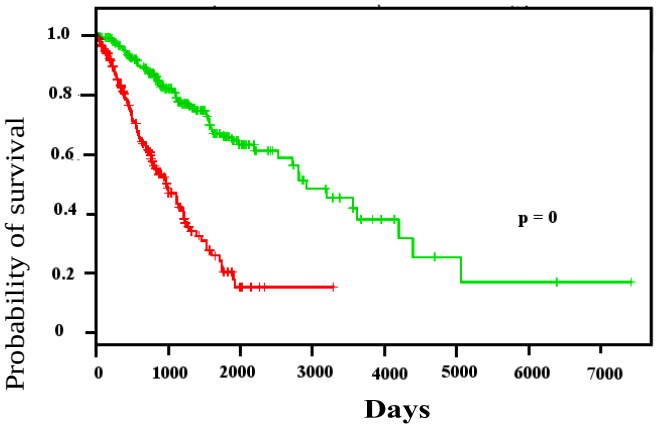


Figure Iiv: STES Specific Top Lost

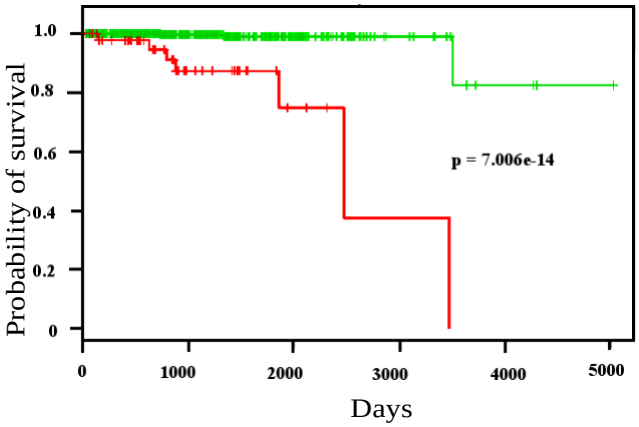


Figure Ji: PRAD Top Gain

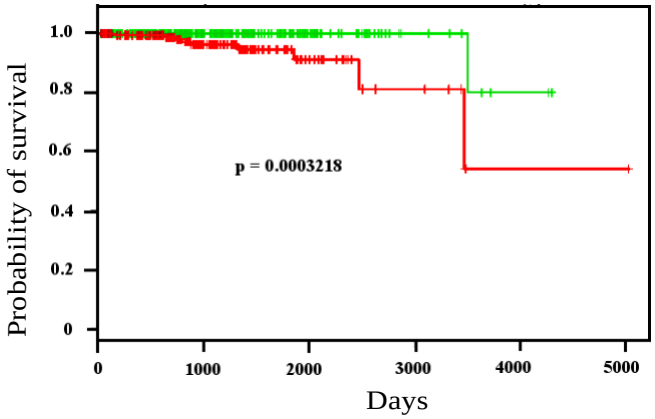


Figure Jii: PRAD Specific Top Gain

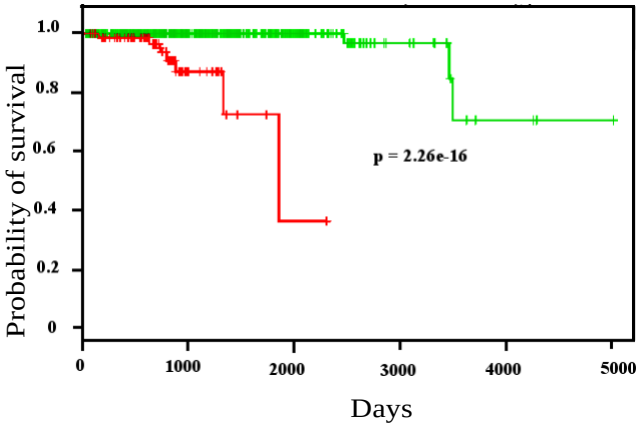


Figure Jiii: PRAD Top Lost

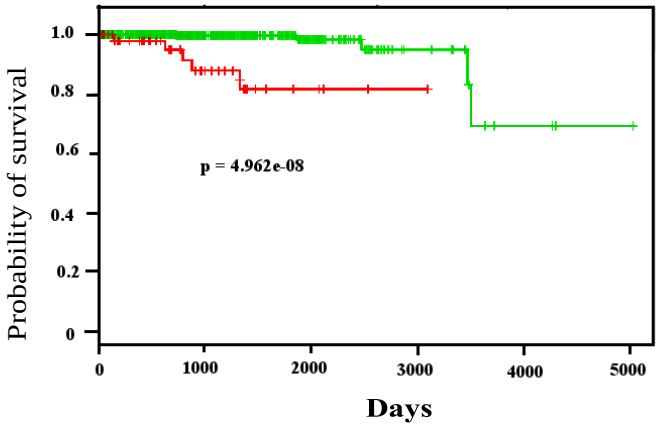


Figure Jiv: PRAD Specific Top Lost

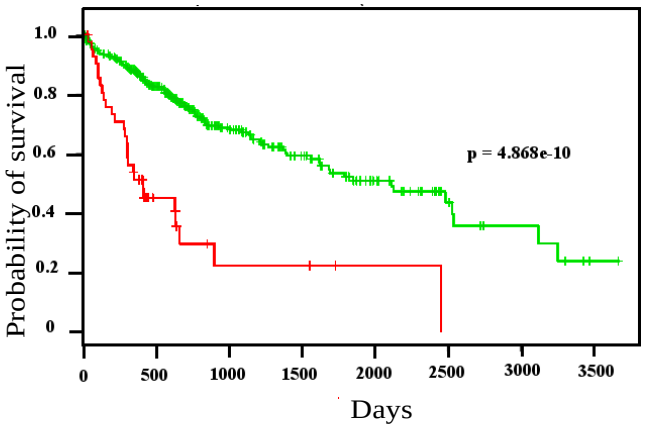


Figure Li: LIHC Specific Top Gain

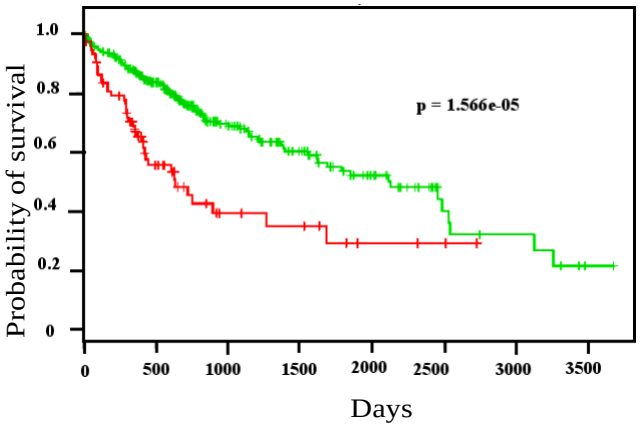


Figure Liii: LIHC Top Lost

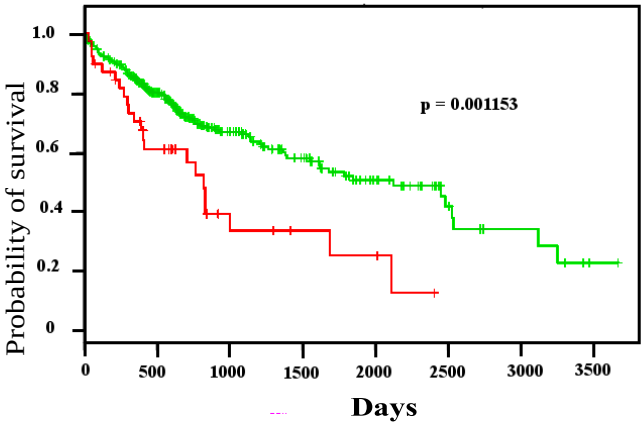


Figure Liv: LIHC Specific Top Lost

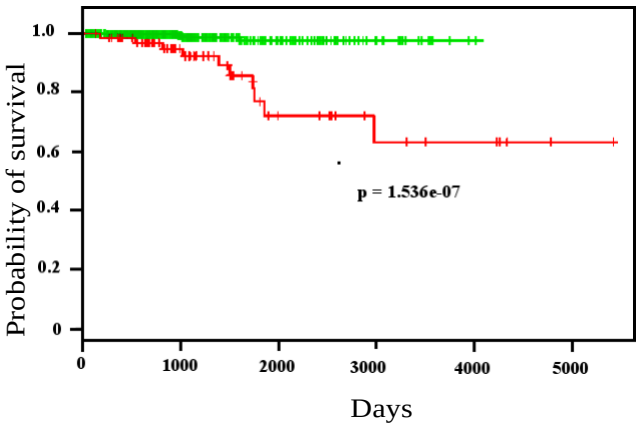
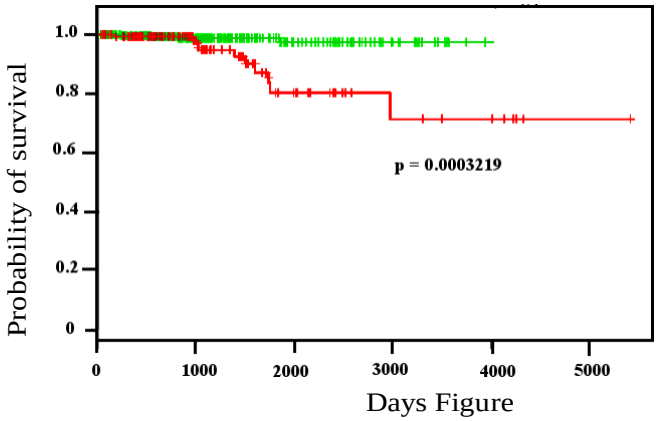


Figure Mi: THCA Top Gain



Mii: THCA Specific Top Gain

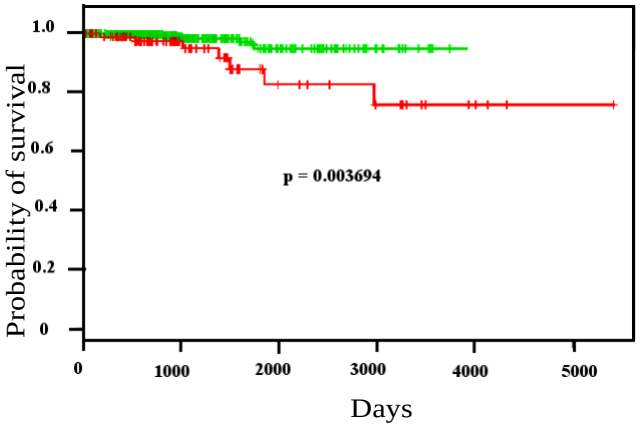


Figure Miii: THCA Top Lost

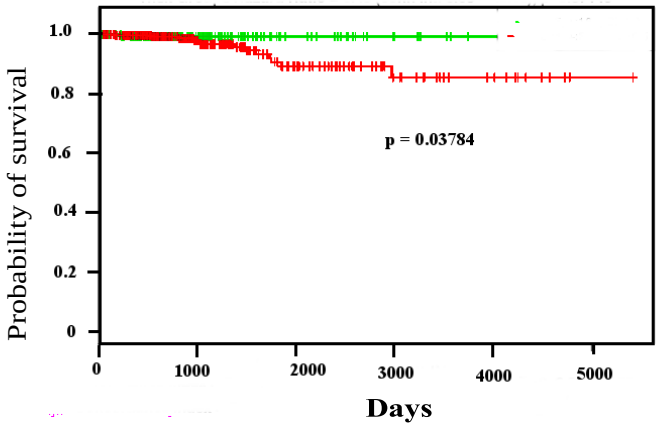


Figure Miv: THCA Specific Top Lost

Figure S3 (A-M): Kaplan-Meier survival analysis plots of multigene cancer biomarkers involved in edgetic perturbations. The x axes indicate the number of days until patient death whereas the y axes indicate the probability of patient survival. In all the figures, the green lines indicate better survival (longer life-span) after cancer diagnosis while the red lines indicate poor survival (shorter life-span) after cancer diagnosis as a result of the proteins involved in edgetic gains or losses. In all the cases, the proteins involved in edgteic perturbations predicted poor survival of the patients (Logrank test p-value < 0.05), indicating their importance in cancer monitoring and prognosis. (i) Overall survival predicted from gene signatures involved in edgetic gains across most patients of a cancer type (except for LIHC), (ii) Overall survival predicted from gene signatures involved in edgetic losses across most patients of a cancer type, (iii) Overall survival predicted from gene signatures involved in edgetic gains across patients showing cancer-specific perturbations, (iv) Overall survival predicted from gene signatures involved in edgetic losses across patients showing cancer-specific perturbations. The names of the prominent proteins with multiple perturbations responsible for the above observations can be found in S4 Table and in EdgeExplorer website.

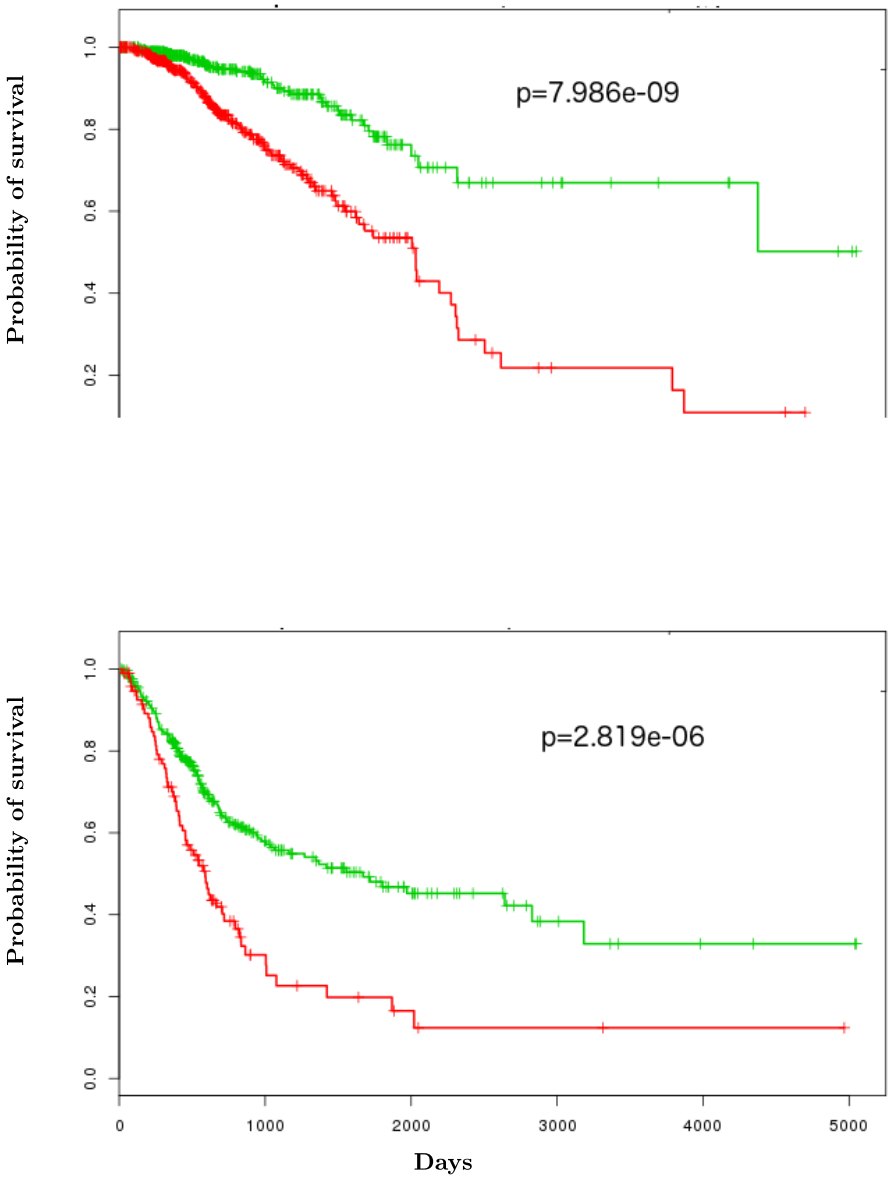


Figure b: BLCA

Figure S4 (a - b): Kaplan-Meier survival analysis plots of multigene cancer biomarkers involved in edgetic perturbations from the randomised PPIN. The x axes indicate the number of days until patient death whereas the y axes indicate the probability of patient survival. In both the figures, the green lines indicate better survival (longer life-span) after cancer diagnosis while the red lines indicate poor survival (shorter life-span) after cancer diagnosis as a result of the proteins involved in edgetic gains or losses. In all the cases, the proteins involved in edgetic perturbations predicted poor survival of the patients (Log-rank test p-value < 0.05), indicating their importance in cancer monitoring and prognosis. (A) Overall survival predicted from gene signatures involved in edgetic gains across most patients in BRCA, (B) Overall survival predicted from gene signatures involved in edgetic gains across most patients in BLCA.

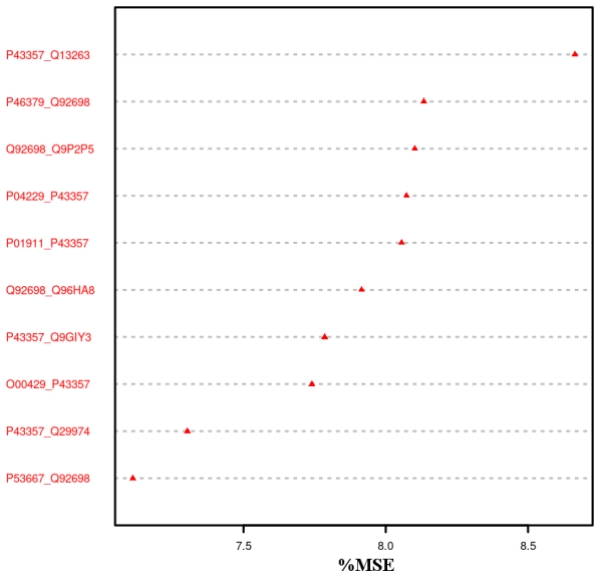


Figure A: Set one of important gained edges across cancer types

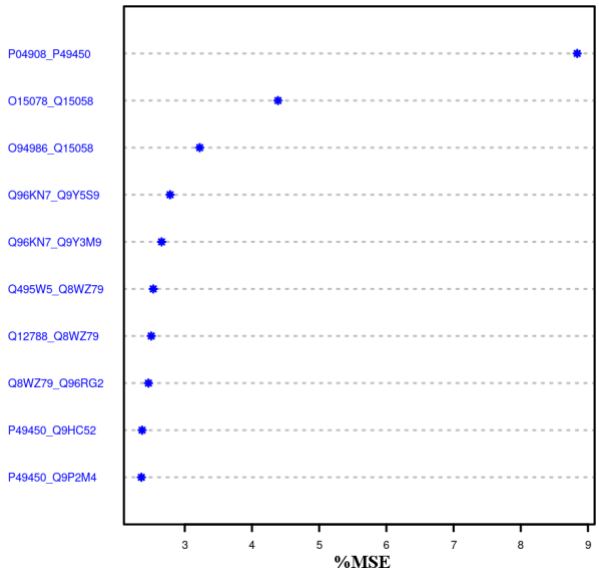


Figure B: Set two of important gained edges across cancer types

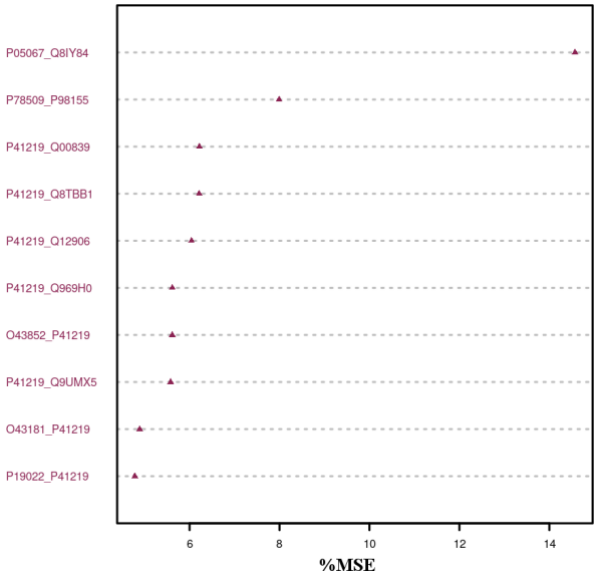


Figure C: Set one of important egdetic losses across cancer types

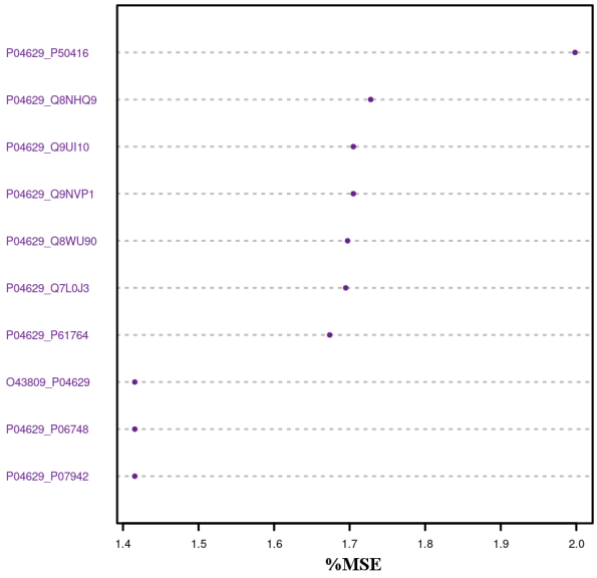


Figure D: Set two of important egdetic losses across cancer types

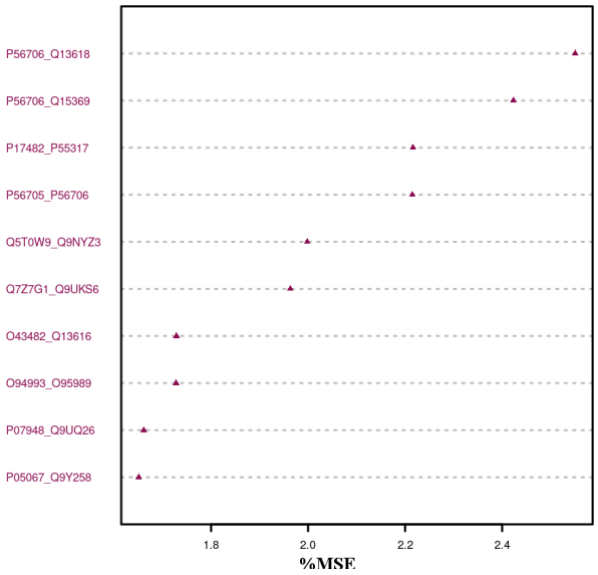


Figure E: Set one of important egdetic losses and gains across cancer types

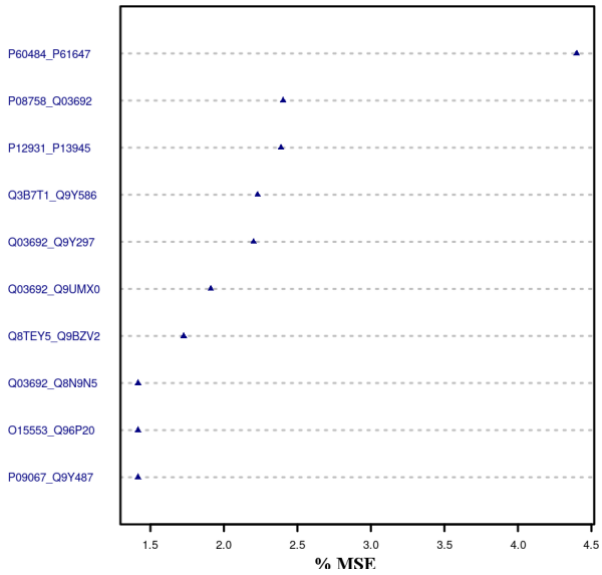


Figure F: Set two of important egdetic losses and g across cancer types

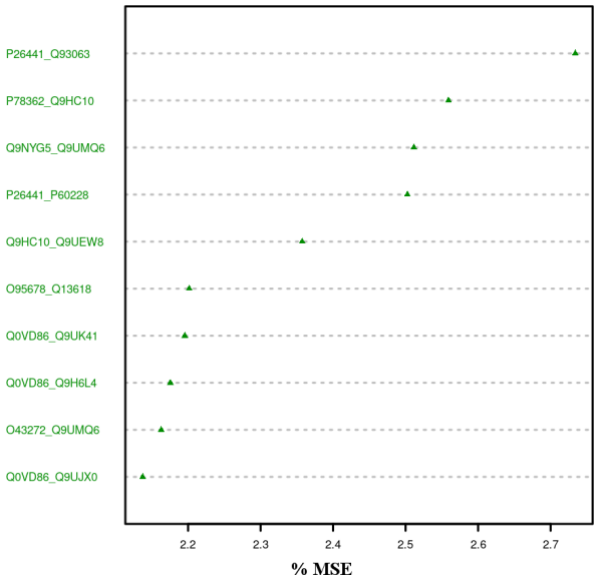


Figure G: Set three of important egdetic losses and gains across cancer types

Figure S5A-G: Top ranked features (edges) from the Random Forest algorithm that distinguish cancer types based on the identified groups from hierarchical clustering (Figure 5). The x axes indicate the percentage (%) Mean Squared Error (MSE2). The higher the %MSE of the feature (perturbed edge), the more important the perturbed edge is in identifying a cluster.

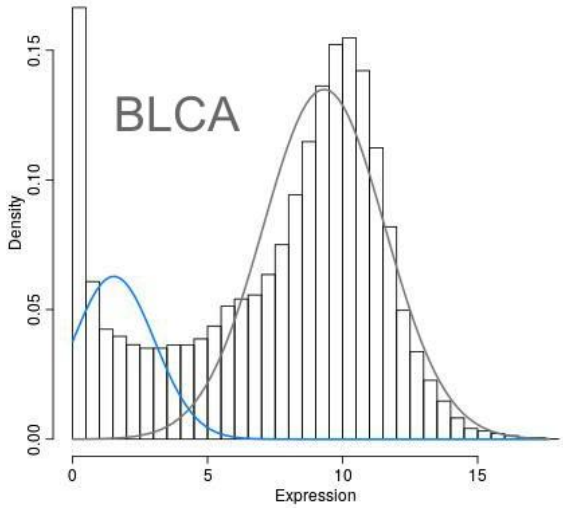


Figure H: Distribution of gene expression data in BLCA cancer samples

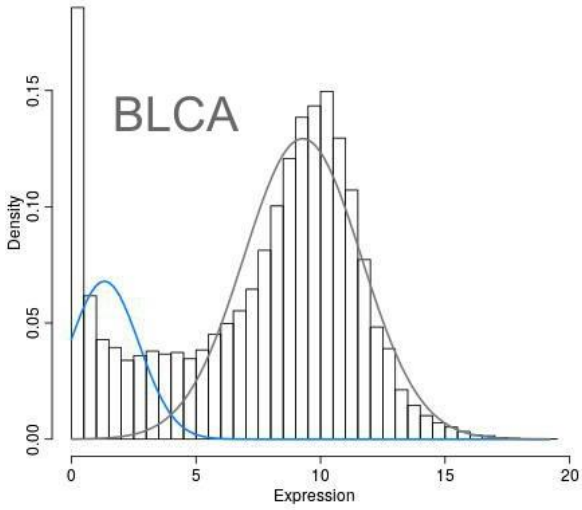


Figure I: Distribution of gene expression data in BLCA paired healthy samples

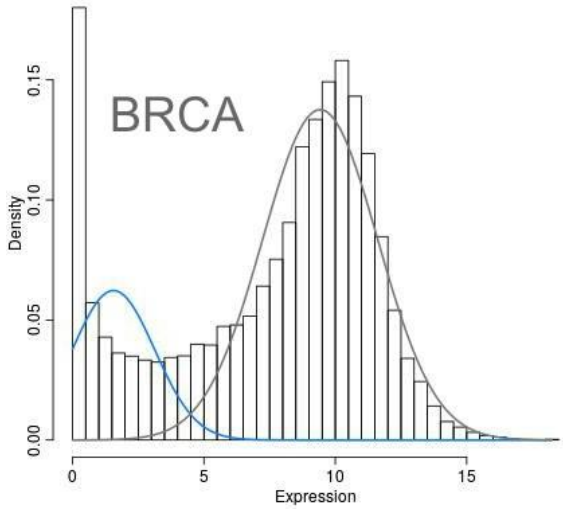


Figure J: Distribution of gene expression data in BRCA cancer samples

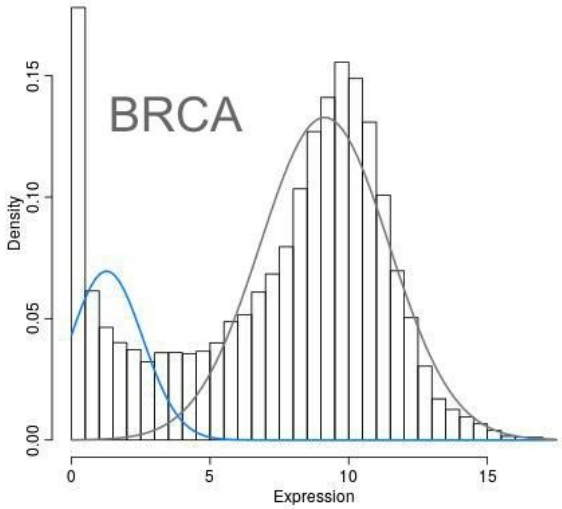


Figure K: Distribution of gene expression data in BRCA paired healthy samples

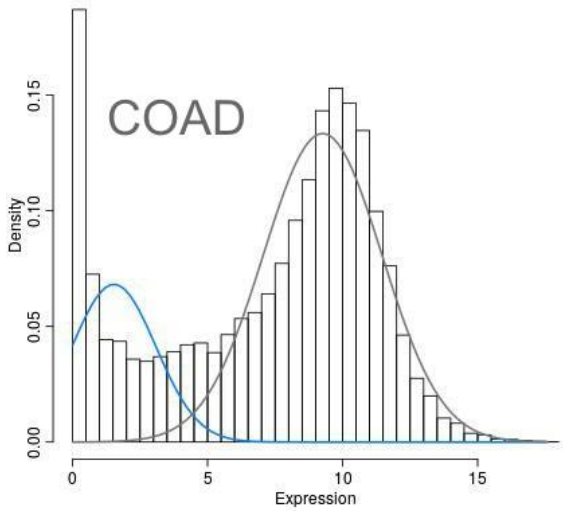


Figure L: Distribution of gene expression data in COAD cancer samples

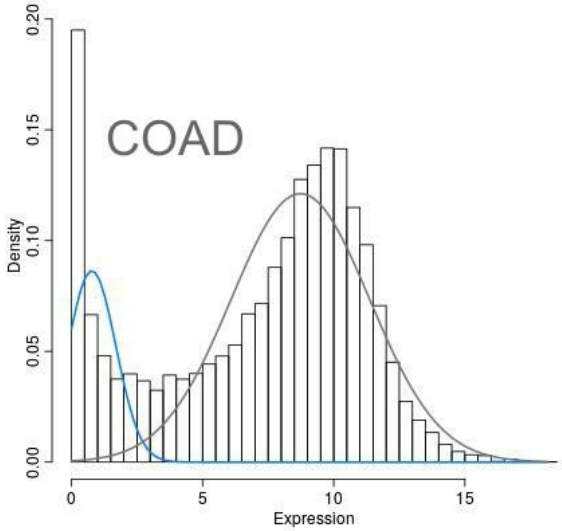


Figure M: Distribution of gene expression data in COAD paired healthy sample

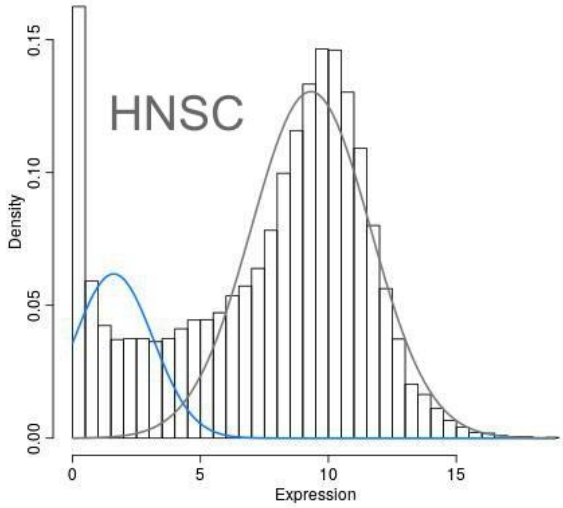


Figure N: Distribution of gene expression data in HNSC cancer samples

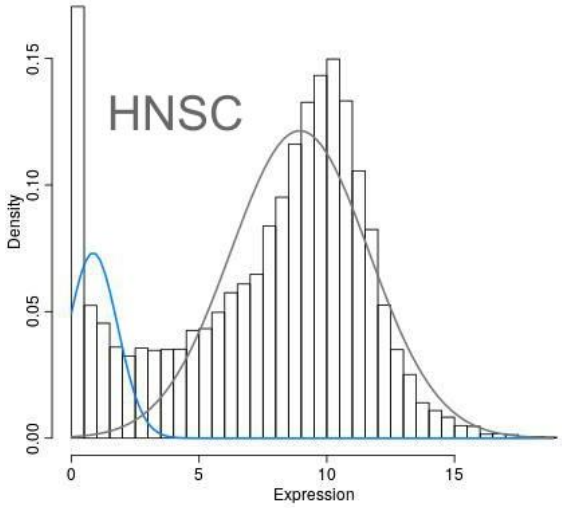


Figure O: Distribution of gene expression data in HNSC paired healthy samples

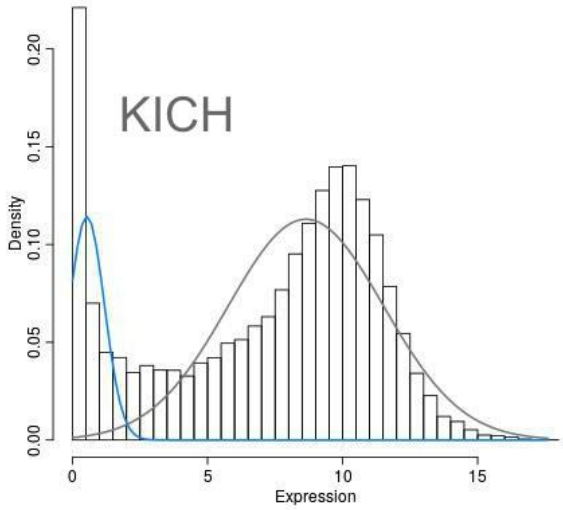


Figure P: Distribution of gene expression data in KICH cancer samples

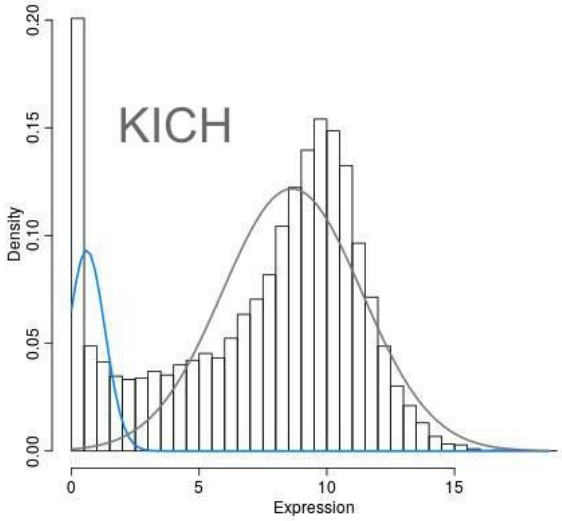


Figure Q: Distribution of gene expression data in KICH paired healthy samples

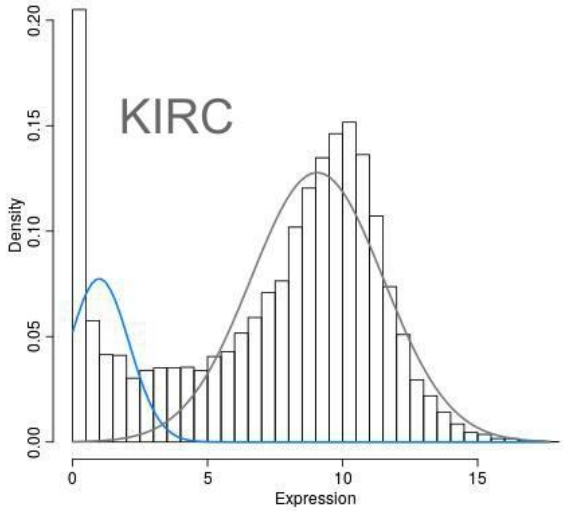


Figure A: Distribution of gene expression data in KIRC cancer samples

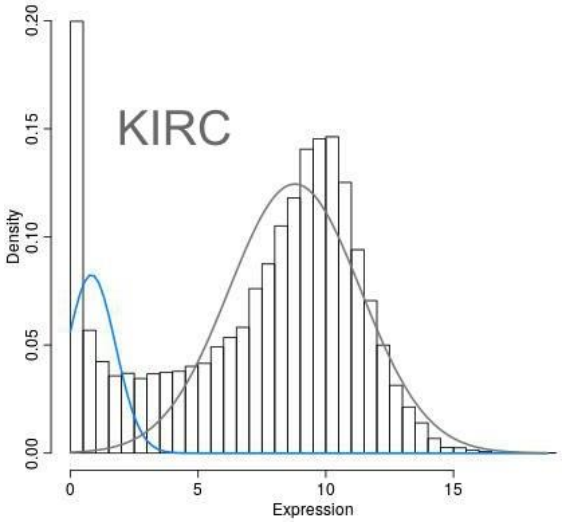


Figure B: Distribution of gene expression data in KIRC paired healthy samples

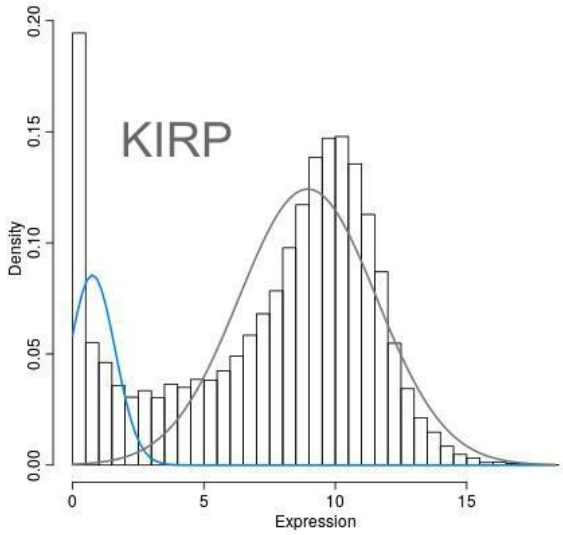


Figure C: Distribution of gene expression data in KIRP cancer samples

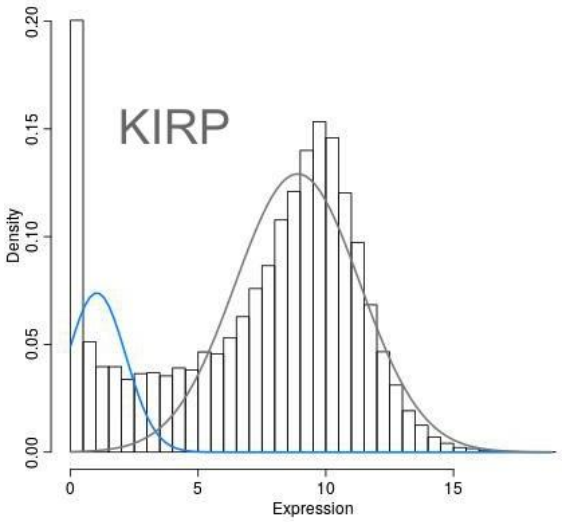


Figure D: Distribution of gene expression data in KIRP paired healthy samples

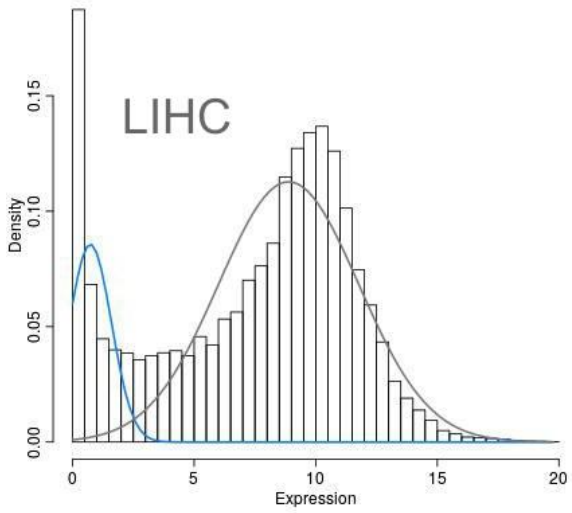


Figure E: Distribution of gene expression data in LIHC cancer samples

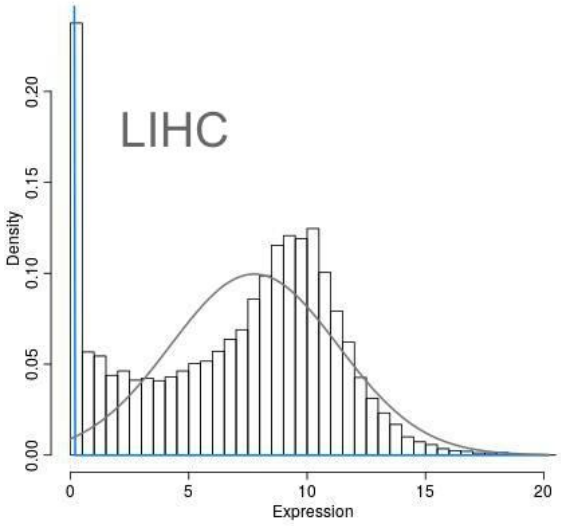


Figure F: Distribution of gene expression data in LIHC paired healthy samples

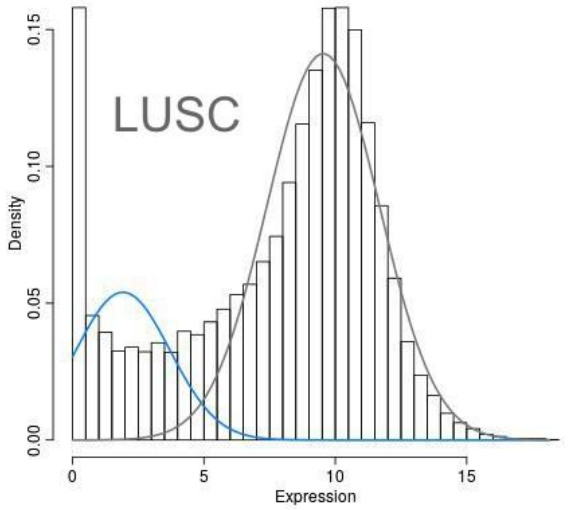


Figure G: Distribution of gene expression data in LUSC cancer samples

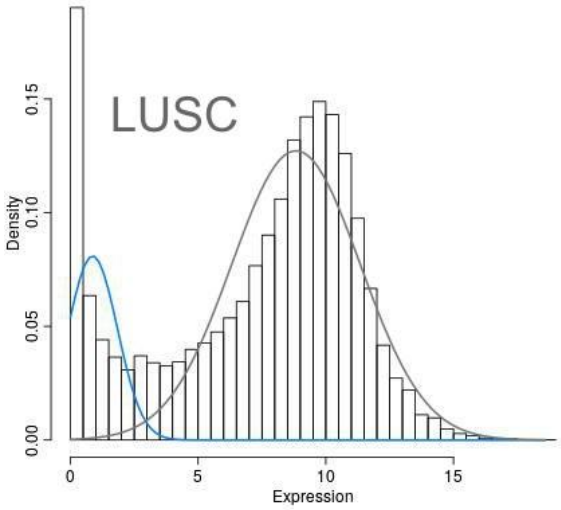


Figure H: Distribution of gene expression data in LUSC paired healthy samples

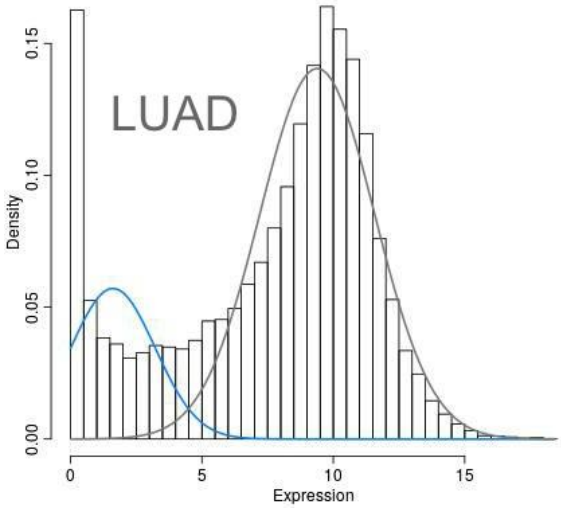


Figure I: Distribution of gene expression data in LUAD cancer samples

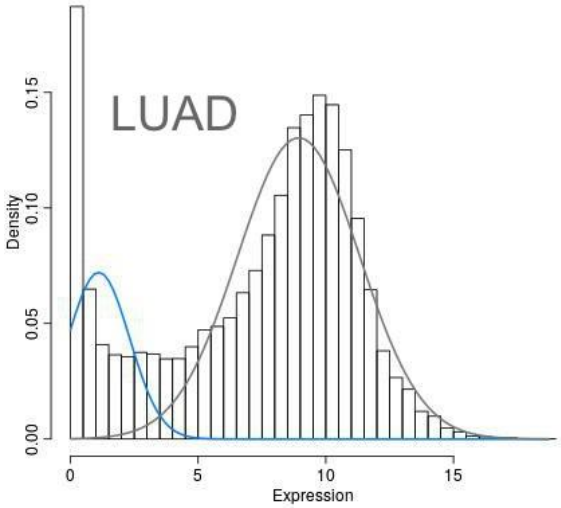


Figure J: Distribution of gene expression data in LUAD paired healthy samples

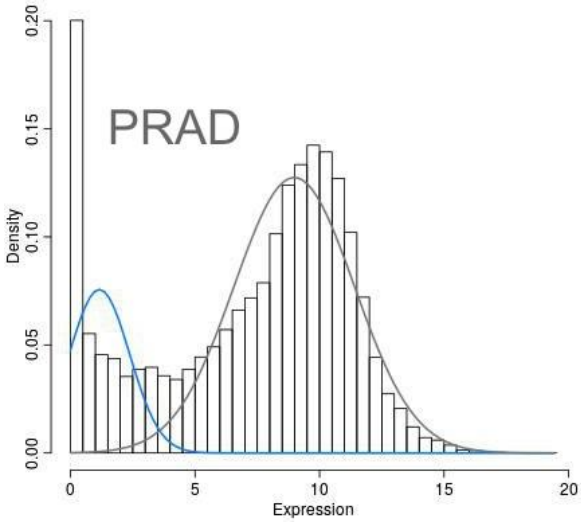


Figure K: Distribution of gene expression data in PRAD cancer samples

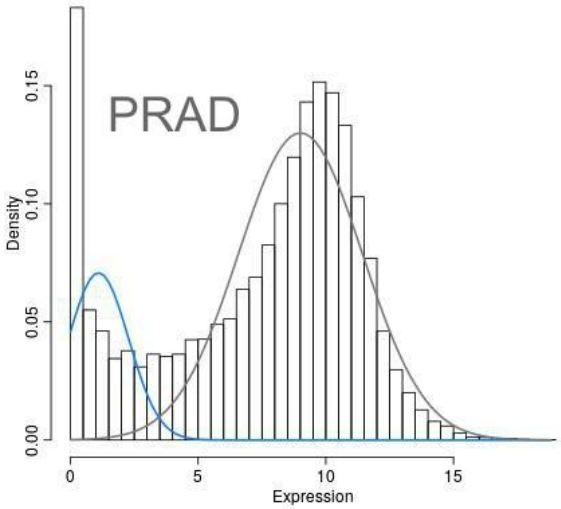


Figure L: Distribution of gene expression data in PRAD paired healthy samples

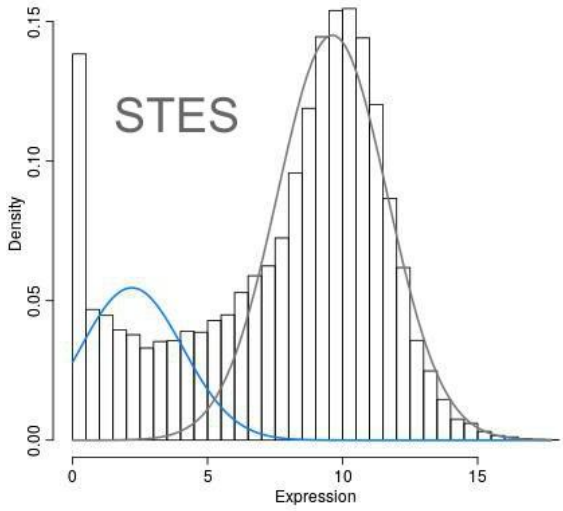


Figure M: Distribution of gene expression data in STES cancer samples

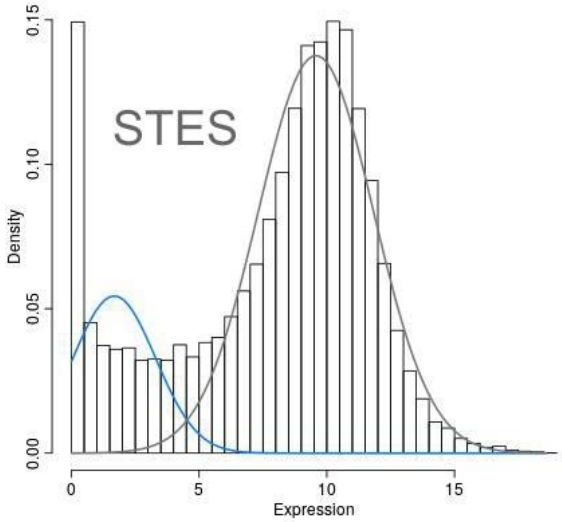


Figure N: Distribution of gene expression data in STES paired healthy samples

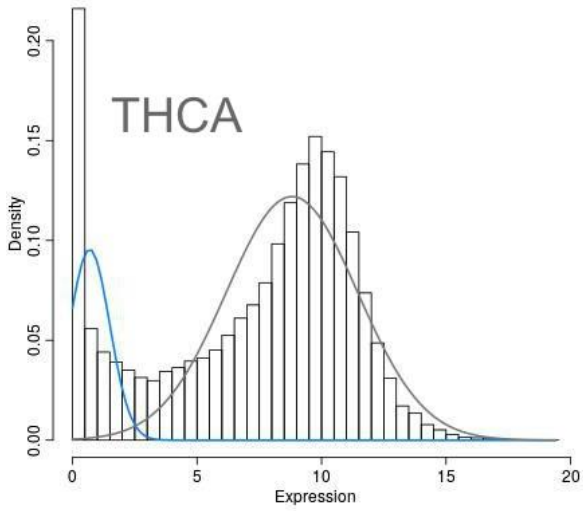


Figure O: Distribution of gene expression data in THCA cancer samples

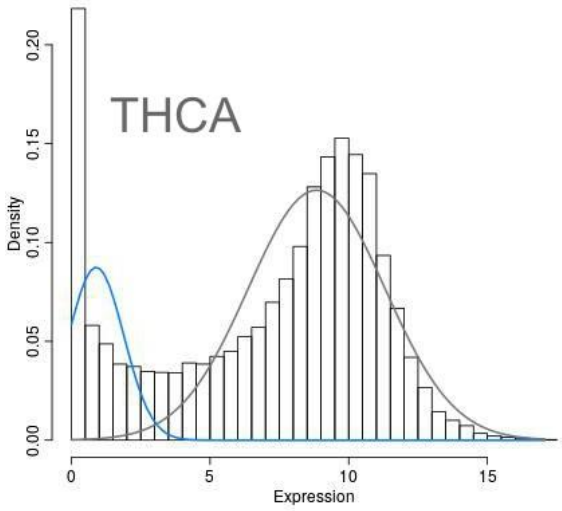


Figure P: Distribution of gene expression data in THCA paired healthy samples

Figure S6A-Z: For each plot, the left blue curve represents the lowly-expressed genes while the grey curve represents the highly-expressed genes across patients of a cancer type for both healthy and cancer samples, respectively. We used these characteristic peaks as a threshold and only kept the genes with an all-samples probability score of greater than 0.8 for subsequent analysis.

Supplementary Table Ia: The proportions of edgetic perturbations associated with SMGs and those associated with random genes with a similar degree significantly differ in size.

Cancer type	Number of edges gained as 1st neighbours of SMGs	Number of edges gained as 1st neighbours of random genes	A	Number of edges gained as 1st or 2nd neighbours of SMGs	Number of edges gained as 1st or 2nd neighbours of random genes	B	Number of edges lost as 1st neighbours of SMGs	Number of edges lost as 1st neighbours of random genes	C	Number of edges lost as 1st or 2nd neighbours of SMGs	Number of edges lost as 1st or 2nd neighbours of random SMGs	D
THCA	254	461	6.06e-15	7741	7282	5.84e-07	334	609	1.69e-19	9788	11057	1.47e-28
BLCA	759	662	0.009	11046	10818	0.02	626	851	2.35e-09	9695	10191	3.58e-07
BRCA	638	935	2.44e-14	11627	12997	4.09e-39	553	927	5.86e-23	12385	13776	7.16e-35
COAD	449	406	0.1	5966	5345	1.13e-18	679	805	0.0008	10533	11817	4.06e-36
KIRC	215	477	8.69e-24	7001	8001	2.01e-26	380	452	0.012	14418	12681	5.58e-43
KIRP	154	341	2.53e-17	5402	4567	3.19e-23	233	354	5.12e-07	9048	9168	0.27
KICH	102	89	0.34	3409	2799	3.93e-19	246	187	0.004	9337	7824	4.43e-44
HNSC	626	693	0.06	11439	11589	0.15	817	814	0.9	13947	13941	0.9
LUAD	480	431	0.09	9109	8375	7.49e-16	608	450	1.18e-06	11819	11082	1.12e-06
PRAD	261	247	0.53	8173	7109	2.1e-30	320	467	1.25e-05	9060	9265	0.05
LUSC	426	261	1.78e-10	5295	5127	0.03	430	470	0.17	10307	10132	0.1
STES	160	166	0.74	4685	5067	1.51e-05	182	200	0.35	3827	4868	3.85e-36
LIHC	1293	1508	3.43e-05	22998	23235	0.06	2161	1705	7.68e-14	32554	30974	3.86e-24

A: P-value showing if the difference between the proportion of edgetic gain perturbations associated with SMGs significantly differs from the proportion of edgetic perturbations associated with random genes at the first degree neighbours.
B: P-value showing if the difference between the proportion of edgetic gain perturbations associated with SMGs significantly differs from the proportion of edgetic perturbations associated with random genes at both the first and second degree neighbours.
C: P-value showing if the difference between the proportion of edgetic loss perturbations associated with SMGs significantly differs from the proportion of edgetic perturbations associated with random genes at the first degree neighbours
D: P-value showing if the difference between the proportion of edgetic gain perturbations associated with SMGs significantly differs from the proportion of edgetic perturbations associated with random genes at the first degree neighbours

Supplementary Table Ib: 9 cancer types show a significantly larger proportion of edgetic perturbations associated with SMGs when compared to the proportion of edgetic perturbations associated with random genes with similar degrees.

Cancer type	genes in gains	Number of edges gained as 1st neighbours of random genes	E	Number of edges gained as 1st or 2nd neighbours of random genes	F	genes in losses	Number of edges lost as 1st neighbours of random genes	G	Number of edges lost as 1st or 2nd neighbours of random genes	H
THCA	33	461	-	7282	2.42e-06	33	609	-	11057	-
BLCA	42	662	0.004	10818	0.01	46	851	-	10191	-
COAD	45	406	-	5345	5.64e-19	65	805	-	11817	-
KIRC	24	477	-	8001	-	23	452	-	12681	2.79e-43
KIRP	13	341	-	4567	-	19	354	-	9168	-
KICH	16	89	-	2799	1.97e-19	18	187	0.002	7824	2.21e-44
LUAD	34	431	-	8375	3.75e-16	46	450	4.39e-07	11082	8.99e-13
PRAD	28	247	-	7109	1.05e-30	30	467	-	9265	-
LUSC	20	261	8.9e-11	5127	0.01	25	470	-	10132	-
LIHC	44	1508	-	23235	-	45	1705	3.84e-14	30974	1.93e-24

E: P-value showing how significantly large the proportion of edgetic gains associated with SMGs is when compared to edgetic gain perturbations associated with randomly generated genes on the first degree neighbours.
F: P-value showing how significantly large the proportion of edgetic gains associated with SMGs is when compared to edgetic gain perturbations associated with randomly generated genes on both the first and second degree neighbours.
G: P-value showing how significantly large the proportion of edgetic losses associated with SMGs is when compared to edgetic loss perturbations associated with randomly generated genes on the first degree neighbours.
H: P-value showing how significantly large the proportion of edgetic losses associated with SMGs is when compared to edgetic loss perturbations associated with randomly generated genes on both the first and second degree neighbours-: Indicates no significant differences in the proportion of perturbations associated with SMGs and those associated with random genes, or the proportion of perturbations associated with SMGs is not larger than the proportion of perturbations associated with random genes
Genes in gains: number of randomly generated genes with similar degrees to the SMGs and involved in edgetic gain perturbations. Genes in losses: number of randomly generated genes with similar degrees to the SMGs and involved in edgetic loss perturbations.

Supplementary Table Ic: Specific cancer SMGs are involved in edgetic perturbations of cancer PPINs.

Cancer type	Cancer type SMGs	Gained edges	Number of edges gained as 1st neighbours of SMGs	Number of edges gained as 1st or 2nd neighbours of SMGs	SMG protein products involved in edgetic gains	% of gains linked to SMGs	Lost edges	Number of edges lost as 1st neighbours of SMGs	Number of edges lost as 1st or 2nd neighbours of SMGs	SMG protein products involved in edgetic losses	% of losses linked to SMGs
THCA	35(36)	22831	254	7741	32	33.9	28065	334	9788	32	34.8
BLCA	52(54)	20739	759	11046	47	53.26	19030	626	9695	42	50.94
BRCA	49(52)	22195	638	11627	47	52.4	25516	553	12385	46	48.5
COAD	82(87)	10065	449	5966	54	59.27	21024	679	10533	60	50.1
KIRC	24(26)	18258	215	7001	24	38.34	33005	380	14418	25	43.7
KIRP	18(21)	17174	154	5402	19	31.5	27141	233	9048	20	33.33
KICH	16(18)	12423	102	3409	17	27.4	27490	246	9337	17	34
HNSC	50(52)	21913	626	11439	46	52.2	27485	817	13947	44	50.7
LUAD	52(58)	16622	480	9109	39	54.8	21907	608	11819	46	54
PRAD	39(40)	17529	261	8173	28	46.63	22468	320	9060	30	40.32
LUSC	34(35)	13108	426	5295	26	40.39	23242	430	10307	28	44.35
STES	24(27)	24326	160	4685	20	19.26	20800	182	3827	20	18.4
LIHC	41(49)	36458	1293	22998	44	63.1	51445	2161	32554	46	63.27

The numbers in the brackets next to the Cancer type SMGs indicate the protein products of the SMGs.

Supplementary Table IIa: Importance of the proteins involved in multiple edgetic perturbations or edges frequently perturbed across patients of a cancer type and their significance in predicting overall patient survival

Cancer type	Proteins involved in significant edgetic gains		Proteins involved in significant edgetic losses			
	Top Gains	**Top cancer-specific gains**	**Top losses**	**Top cancer-specific losses**	**SMGs in edgetic gains or losses**	**Interesting Genes from Survival analysis**
BRCA	CDC25C**, TDO2, DST***	UPF2-HS3ST3A1	ALK, APOB, APP, ASB14, AURKC, AVPR2, C1QTNF9, CCDC36, DMRT3, ENPP6, ESR2, GFAP, HOXD4, HRNR, INCA1, KCNA5, LURAP1, MAP1LC3C, MASP1, MYH7B, MYOC, MYOCD, NOS2**, PPP2R2B, USP44	FOXF1**, F8***, VWF***	AKT1, BAP1, BRCA1, EP300, ESR1, FBLN2, FOXA1, KRAS, MAP3K1, NCOR1, NTRK3, PIK3R1, SALL4	CDK1, DYNLT1, HSP90AA1, MAPK3, MARK3, PLK3, HS3ST3A1, CALU, CCT5, CCT6A, CDIPT, C17orf79, CUL5, DNAJA1, DNAJA3, DNAJC7, EMD, HSPA5, LGALS3BP, NAP1L1, PLEC, PPM1B, PPP2R1A, PSMA1, PSMA2, PSMA4, PSMA6, PSMC5, PSMD14, PSMD2, PSMD4, PTBP1, PTGES3, RAC1, RPS18, RPS4X, SERPINE1, SLC25A5, STRAP, TCP1, UCHL5, YWHAQ, STRBP
	HR (84/110) LR (26/110)	**HR (110/110, 100%)**	**HR (74/110) LR (34/110)**	**HR (75/110) LR (35/110)**		
BLCA	BCAN, CATSPER1, FOXD4, FOXH1, GUCY1A2, HIST1H2AB, HIST1H2AE, HIST1H2BJ, HIST2H2AC**, HMGA2, KCNJ10, LRRC46, MAST1, OTX1, SLX4IP, TMEM52B, LONRF3***, RNF146***, SH3PB2***	DLG1, DLG3, DLG4, ERBIN, GUCY1A2**, GUCY1B1, HSP90AA1, HSPA4, LIN7A, SNTA1, STUB1	ACTA1**, ADCYAP1, ADRA1D, APP, ASB16, AURKC, AVPR2, BEX1, BMX, CLEC4G, CMTM5, EFHC2, ENPP6, FAM124A FOXD3, GPM6A, GRIN2A, ISL1, KCNA3, KCNA5, MEFV, MYH7B, P2RY12, PRPH, RUNX1T1, RXRG, SCN2B, SOX5, STX1B, BTB20	KCNA3, FAM124A**, PIPOX, GNAQ***, ADHFE1***	ATM, CDKN2A, CTNNB1, CUL1, EP300, ERBB3, FBXW7, KRAS, MDM4, PIK3CA, PSIP1, PTEN, RB1, RBM10, SPTAN1, TP53, TSC1	CTCFL, EP300, ERCC6, HIST1H3G, INO80, MLLT1, NCL, NPM1, PRMT7, RBBP7, RNF20, TAF15, TAF1B, DLG3, GUCY1B3, LIN7A, ABL1, ABLIM2, ANXA1, BTK, CAPZA2, CDH2, CSNK1A1, DNASE1, ERBB3, ETV6, HDAC4, LRCH3, MIB2, PPP1CA, PPP1R1A, SCIN, SMARCB1, XPO6, FAM124A, STAC3, THAP1, ZBTB44, ZNF165, ZNF250, ZZZ3
	HR (16/19) LR (3/19)	**HR (12/19) LR (7/19)**	**HR (18/19) LR (1/19)**	**HR (14/19) LR (5/19)**		

HNSC	HOXC9, FOXL2**, KHDRBS1***, DLG2***	CEBPE**	APP, ASB14, ASGR1, AURKC, BMX, CDKL3, FLT3, GDF9, GLIPR1L2, INCA1, PCK1**, RXRG, TSSK3, TUBB1, USP49, WNK4, ZNF396	WNK4**, GOSR2***, REPIN1***, BCAN***	AJUBA, CUL3, EP300, FBXW7, HLA-B, HUWE1, TP53	CCDC59, DDX52, DNAJC9, ELAVL1, HNRNPUL2, IARS, MOV10, NSDHL, NUP205, P4HA1, RPLP1, BATF, BATF3, E2F1, FOS, JUN, KDM2B, PSAT1, ACTB, APP, BAT3, FASN, GNAS, MAGED2, NUP62, PSMB1, PSMB4, PSMD12, RPL23, RPS6, TCP1, UBR5, WNK4, ABCC2, CHCHD2, MMGT1, SLC30A5, SSR1, TOMM20
	HR (15/42) LR (27/42)	**HR (6/42) LR (36/42)**	**HR (15/42) LR (27/42)**	**HR (10/42) LR (32/42)**		
LUAD	ABCC2**, AGMAT, CDC25C, HPDL, KCNJ10, NEIL3, RGS17, SGO1, SOX30, SPC24, SPECC1***	SLC25A21**	ADRA1D, APOA1**, APP, ASB14, ASB16, AVPR2, BEX1, BMX, BTNL8, CAV3, CCDC36, COLEC10, DNASE2B, ENPP6, GATA1, GNMT, HSPB3, IL9R, INCA1, L1TD1, MYOC, OLIG1, PPARGC1B PTPN5, RAB40A, SH3GL2, SH3GL3, SLC2A4, TCAP, WNT3A	PPARGC1B**, SUPT3H***	AKT1, ARID1A, BAP1, CTNNB1, CUL3, DROSHA, EGFR, FGFR2, KRAS, MET, PIK3CA, SMARCA4, STK11, TP53	ABCC2, CHCHD2, MMGT1, MRPL10, SLC30A5, SSR1, TOMM20, TOMM22, NOSIP, APOL1, C1QC, DGAT1, GDPD1, NLRP1, PDE4B, PLTP, SPEF2, TNS3, APOA1BP, UCHL5, ZNRD1, PIAS1, THRB, ZFP64.
	LR (58/58, 100%)	**HR (55/58) LR (3/58)**	**HR (20/58) LR (38/58)**	**HR (10/58) LR (48/58)**		
LUSC	HIST1H2AB, HIST1H2AE**, HPDL, NEIL3, HECW2***, SPECC1***	MED12L**	ACTN2, AGTR1, APOA1, APP, BIRC7, BTNL8, C1QTNF2, CCDC36, CD1B, CEBPE, CMTM5, DNASE2B, ESR2, FAM124B, FOXA3, GATA1, GDF9, GFI1B, IL9R INCA1, KCNA5, KHDRBS2, L1TD1, MAP1LC3C, MYH7B, MYOC, NR0B2, P2RY12, PACRG, RXRG, SH3GL2, TRIM55, TRIM69, TUBB1, USHBP1**, USP49, VTN	CD1B**, FAM124B, USP22***, SUPT3H***	ARID1A, CUL3, EP300, FAT1, FBXW7, FGFR2, KLF5, LEPROTL1, NOTCH1, PIK3CA, PTEN, RASA1, RB1, TP53, USP44	ACO1, ACTR2, BRCA1, CBX4, CBX8, CENPA, EIF2AK2, ERCC6, H2AFY, HNRNPA2B1, MAPK3, MCM2, MOV10, SHMT2, SUZ12, TRIP13, BET1, NGFRAP1, C1orf109, C1orf216, CCDC121, CCDC146, CCDC148, CCDC87, CEP63, CHCHD3, CNNM3, COPS4, CTNNBIP1, DTNB, EXOC7, EXOC8, FAM107A, FAM110A, FTL, GATAD2B, GCC1, GFI1B, GPSM1, HAUS1, HGS, IFT20, ING3, INTS4, KIAA0753, KLC3, KLC4, KRT19, LENG1, MCM7, MCRS1, MED28, MED4, MRPS23, NDE1, NOC4L, PMF1, PRKAA2, SYNJ2BP, THADA, THOC1, UBE2W, USHBP1, ZFYVE26, ADAM9, CPD, FKRP, ITFG1, ULBP3
	HR (12/16), LR (4/16)	**HR (13/16) LR (3/16)**	**HR (15/16) LR (1/16)**	**HR (13/16), LR (3/16)**		

	ASB14, CCNE2, CENPA, E2F2, ESCO2, HJURP IGF2BP3**, MSX2, PBK, POLE2, POTEF, SGO1 SKA3, TTK, RNF146***, SH3PB2***	ASB14**, CCNE2, MSX2 POTEF	ASB5, ATP12A, CALML3, CLNK, DKK1, FBN3, FOXI2, FOXN1, HEPACAM2, MYOZ2 NOS1, NR0B2**, PRMT8, RALYL, SOX30 TTR, VSIG2	DKK1-MDFI, REPIN1***, BCAN***	CUL3, KMT2D, MET, RNF2, SMARCB1	CAND1, CDK2, CNBP, COPS5, CRYZ, EIF2B2, EPRS, FBXO6, HNRNPA1, HNRNPU, IGF2BP3, KRT17, MDM2, MRE11A, NDUFAF4, COBRA1, OBSL1, RFC4, RPL23A, RPL26L1, RPL37A, RPL38, SIRT7, STAU1, UBC, ARF4, GALK1, SLC25A5, TUFM, CHRD, ESR1, ESRRG, FN1, HDAC1, HDAC3, HNF4G, HNRNPA1, IL3RA, KLF6, PLSCR1, PPARG, RBP5, SIRT6, SIRT7, SMAD4, SMARCA2, SMARCB1, SMARCC1, SNW1, SP2, TRAF6, MDFI
KIRP	**HR (17/32), LR (15/32)**	**HR (3/32), LR (29/32)**	**HR (13/32) LR (29/32)**	**HR (9/32) LR (23/32)**		
KIRC	CDC25C, CDC45, CDKN2A**, EME1, CENPA, FASLG, KIF14, LGALS9C, MCM10, NEIL3, SGO1, GRAMD2B***, GRAMD1C***	MLC1**	ASB14, BIK, BNIPL, CLNK, ESRRB**, FBN3, FOXA3, GRIK2, HAP1, OLFM4, RAB40A, SOX30, TCAP, USP44	BIK**, VSIG8***, SHC1***	ELOC, HIF1A, KAT7, MAGI1, PIK3CA, RNF2, TP53, VHL	ACLY, ARFIP2, ATR, AURKA, C1QBP, CASC3, CCND2, CCNG1, CDC7, CDK11A, CDK4, CDK5RAP3, COMMD1, CRELD2, CTBP2, DYRK1B, E2F1, EEF2, GGA1, HDAC1, HSP90AB1, HSPA8, IQGAP1, NAA38, MCM2, MDM2, MIS12, MOV10, MTR, NCL, WHSC1L1, ORC4L, PA2G4, PPP1CB, PPP1CC, PRKCA, RPP38, SNRPA, SNRPB, TP53, TTF1, TUBB, UBE2A, UBE2I, UBE4B, CAV1, DTNB, MYLK, SNTA1, ACSL3, ANXA2, ATP2A2, CANX, DNAJA1, DNAJB6, EGFR, ERRFI1, GFAP, GNB4, GRB2, HNRNPH1, LRPPRC, NCOA3, PHB, PPP2CA, PARK2, RPL23, RPS27A, S100A16, SEC61A1, SLC25A3, SLC25A5, SLC25A6, TNFAIP2, TUBB, UBASH3B, VAPA, BCL2, BCL2L2, BIK, SOCS3
	HR (48/63) LR (15/63)	**HR (28/63) LR (35/63)**	**HR (43/63) LR (20/63)**	**HR (39/63) LR (24/63)**		

	CST4, OT, KLC3**, MAPK15, CCNO***, CDKN1A***	CITED1**	APP, CA14, CACNA1A, CHST8, CMTM5, ENPP6, ESR2, FAM90A1, FOXH1, GDF9, GFI1B, HAP1, HIST1H2AG, HIST1H2AI, HIST1H2AK, HIST1H2AL, HIST1H2AM, KHDRBS2, MYOC, P2RY12, PPP2R2B, PTH1R RPL10L, SCN2B, SH3GL2, SUSD4, TAGLN3 TEX11, TP63, TRIM9, TTR, TUBB1, USP49 ZBTB16**	FAM184A, FOXH1, HIST1H2AG, HIST1H2AI, HIST1H2AK, HIST1H2AL, HIST1H2AM, PDIA2, RIMBP3, RPL10L**, ATP1B2, CMYA5, FBX027, HTR3E, ISL2, RIBC1, RNF152, TLX2, FLT1***, SHC1***	ACVR2A, AKT1, APC, AXIN1, AXIN2, B2M, BAX, CTNNB1, CUX1, EIF3E, EP300, EPHA7, ERBB3, FAT3, FBXW7, GRIN2A, HIF1A, KRAS, LEPROTL1, MDM2, MLH1, MSH2, PCBP1, PIK3CA, PIK3R1, PTEN, RSPO2, SALL4, SFRP4, SMAD2, SMAD3, SMAD4, SRC, TGFBR2, TGIF1, TP53, UBR5, USP44, ZNRF3	C3orf19, CCNB1, CD99L2, CENPP, CUL2, KLC3, ORC4L, PCMT1, RBBP6, TRIM26, HSPA8, LNX1, ANAPC5, ANXA7, BMI1, CDK4, DPM1, EEF1A1, HDAC3, HDAC7, IL32, MTDH, MX1, NCOR1, PAFAH1B3, PSMD11, SMN2, TERF1, THNSL2, TOLLIP, UBE2I, WDR33, CAND1, RPL31, RPL35, RPL36, RPL4, RPS21, RPS28, RPS3A, RPS7, TP53
COAD	**HR (8/18), LR (10/18,)**	**HR (14/18) LR (4/18)**	**HR (8/18) LR (10/18)**	**HR (6/18) LR (12/18)**		
STES	CLEC5A, HOXA9, MAPK15, STAC3, TNFRSF9, TNFSF11**, ANK1***, TTN***	STAC3**	ADCYAP1, AGTR1, APP, ASB16, BMX, CAMK2B, CCT6B, EID3, ENPP6, ESR2, GPM6A, HSPA1L**, PACRG, RXRG, INA, LURAP1, SCN2B, SPIN2B, TCEANC, TRIM46, TRIM9, TSSK3, USHBP1, SPATA24, ZNF396	HSPA1L**, TCEANC, SPIN2B, EDA2R, CCT6B, USP6, ETFBKMT, CEP170B***	ATR, CDH1, ERBB3, FAT3, GRIN2A	B4GALT7, LMO4, MBTPS1, SBF1, SNRNP35, TRMT2A, PPARA, AP2M1, ARRB2, BAG4, CBL, CDC5L, CEP250, DCUN1D1, PTPLAD1, HSPA1L, MAP3K1, MRPS36, NDUFB9, NDUFV1, NFKB1, NUCB2, STUB1, TAB1, TBC1D22A, TRAF3IP1, TXN2, UBASH3B, XPO1, ZBTB1
	HR (11/37) LR (26/37)	**HR (1/37), LR (36/37)**	**HR (22/37) LR (15/37)**	**HR (22/37) LR (15/37)**		
	CDC25C, CDCA2, CENPA**, DLX2, FOXD4, HASPIN, HIST3HWBB, SGO1, SPC25 CDC25C, DIDO1***, RPA1***	DLX2**	CACNA1A**, SOX30, DRD2, FABP4, KCNJ10, KIF5A, CCDC158, LGALS9C, FAM90A1, HIPK4, CMTM5, RSPH9, TUBB1, SGK2, RSPH14	LGALS9C**, BCAR3***, HOXC6***	ACSL3, DDX5, RAF1, TP53	DGCR6, MSX1, ATG9A, CD47, LGALS9, LGALS9C, RRAGB, SLC12A7, SLC38A9
PRAD	**HR (4/52), LR (48/52)**	**HR (8/52) LR (44/52)**	**HR (13/52) LR (39/52)**	**HR (1/52) LR (51/52)**		
LIHC	EBF2-ZNF23, KANK2***, WDR83***	WNT3A**	AMHR2**	BMP10**, HR***, RARB***, DDX11***, SLAMF7***	NONE	
LIHC (continued from above cells)	**NS**	**LR (50/50)**	**HR (1/50) LR (49/50)**	**HR (19/50) LR (31/50)**		FZD1, PPP2R1B, PPP2R5B, TRAF2, HSP90AA1

THCA	ALK**, HIPK4, RAB40A, PDZK1, GRAMD2B***, GRAMD1C***	RAB40A**	VEGFD**	CSAG1**, ANKS1B***, ERBB4***	AKT1, PRKAR1A	ACTB, ACTN4, ALK, BCAR1, BICD2, CDK13, CENPF, CORO1C, EIF4B, EPHA1, EPHB2, ERRFI1, FLII, GAK, GRB2, HSP90AA1, HSPD1, IKBKG, IRF7, IRS1, JAK2, JAK3, KRT18, MAP2K7, MAP3K1, MAP3K4, MAP3K5, MAPK1, MAPK8IP3, MTIF2, MYH10, MYH9, MYO6, PDLIM3, PIK3CB, PIK3R1, PLCB2, PLCG1, PRKCQ, PTN, PXN, RAB35, RAD17, SHC1, SMC6, SOCS1, SOCS5, SRC, STAT3, TNK2, TUBB2C, TUBGCP2, ZC3HC1, HSP90AA1, ISCA1, LYRM7, PSME3, ZER1
	HR (27/58), LR (38/58)	**HR (31/58) LR (27/58)**	**HR (34/58) LR (24/58)**	**HR (38/58) LR (20/58)**		
KICH	HRK**, SLC12A5, BRCA1***, OBSCN***	IL12B**	FOXE1, HIST1HIE, MYOC, PLG, TRIM15, VTN**	TRIM15**, PLG, SLC22A11, ASPH***, FBLN7***	RNF2, TP53, VHL	
	(25/25)	**(24/25)**	**(25/25)**	**(25/25)**		

The ratios in the brackets indicate the number of patients showing a particular perturbation (in KICH) or the classification of HR and LR patients. LR: Low Risk of cancer related death. HR: High Risk of cancer related death.

** Protein nodes having the highest number of edgetic perturbations.

Note: All corresponding Survival analysis plots for each of the multi-gene biomarkers are available in Supplementary Figure 3. The cox interesting genes were found to exert significant effects on the predicted survival responses than other genes.

NS: Non-significant survival analysis results. Interesting Genes from Survival analysis: Genes important in discriminating the patients based on their probabilities of survival. *** Protein nodes participating in edgetic perturbations due to isoform/domain switches.

Supplementary Table IIb: Proteins involved in subtype and subtype specific edgetic perturbations (SubtypePercLost and SubtypePercGained) in 11 cancer types.

Cancer type	**Cancer Subtype and ratio of samples showing a specific perturbation**									
BRCA	PR+ top gains ***CDC25C**** **(61/61)**	PR+ top losses ***NOS**** **(61/61)**	ER+ top gains ***CDC25C**** **(70/70)**	ER+ top losses ***NOS2*** **(70/70)**	HER+ top gains ***HIST1H3A, IGF2BP3**(13/13)**	HER+ top losses ***CFTR*** **(13/13)**	PR- top gains ***IGF2BP3**** **(29/29)**	PR- top losses ***CFTR*** **(29/29)**	ER- top gains ***IGF2BP3**(* **20/20)**	ER- top losses ***CFTR*** **(20/20)**
	PR+ specific gains ***PARVG**** ***(24/61)***	PR+ specific losses ***XPO4*(21/61)***	ER+ specific gains ***SPATS**** ***(3/70)***	ER+ specific losses ***PIH1D2**** **(5/70)**	HER+ specific gains (0)	HER+ specific losses ***ATP6V1B1**** **(2/13)**	PR- specific gains ***CORO6**** ***(2/29)***	PR- specific losses ***PTN**** ***(2/29)***	ER- specific gains ***GTPBP2**** (2/20)	ER- specific losses ***YAF2**** (2/20)
PRAD	SPOP top gains ***HIST2H3A**** ***(6/6)***	ERG top gains ***CENPA**** ***(22/22)***	Others top gains ***HIST2H3A**** ***(12/12)***	SPOP top losses ***TNF**** ***(6/6)***	Others top losses ***CACNA1A**** ***(12/12)***	ERG top losses ***CACNA1A**** ***(22/22)***				
	SPOP specific gains ***TRIM5**** ***(3/6)***	ERG specific gains ***ADAM2**** ***(21/22)***	Others specific gains ***ADAM2**** ***(9/12)***	SPOP specific losses ***COL2A1**** ***(4/6)***	Others specific losses ***KCNA4-KCNA5*** ***(8/12)***	ERG specific losses ***MYL1**** ***(11/22)***				
HNSC	Atypical top gains ***IGF2BP1**** ***(6/6)***	Mesenchymal top gains ***IGF2BP1**** ***(8/8)***	Basal top gains ***FOXL2**** ***(10/10)***	Classical top gains ***IGF2BP1**** ***(13/13)***	Atypical top losses ***GDF9**** ***(6/6, 100%)***	Basal top losses ***GDF9**** ***(10/10)***	Classical top losses ***GDF9**** ***(13/13)***	Mesenchymal top losses ***GDF9**** ***(7/8)***		
	Atypical specific gains ***FEZF1-GABRR1*** ***(5/6)***	Mesenchymal specific gains ***CSAG1**** ***(7/8)***	Basal specific gains ***CSAG1**** ***(9/10)***	Classical specific gains ***FGF19-FGFR4,*** ***KHDRBS1-DLG2*** ***(11/13)***	Atypical specific losses ***PRG2**** ***(5/6,)***	Mesenchymal specific losses ***F12**** ***(7/8)***	Basal specific losses ***PAK5**** ***(9/10)***	Classical specific losses ***ADIPOQ**** ***(9/13)***		

KIRC	Cluster 1 top gains ***CDKN2A* (16/16)***	Cluster 2 top gains ***CDKN2A* (16/16)***	Cluster 3 top gains ***TP73* (15/15)***	Cluster 4 top gains ***CDKN2A* (21/21)***	Cluster 1 top losses ***ESRRB* (16/16)***	Cluster 2 top losses ***ESRRB* (16/16)***	Cluster 3 top losses ***ESRRB* (15/15)***	Cluster 4 top losses ***ESRRB* (21/21)***
	Cluster 1 specific gains ***GRIA4*, KHDRBS1-DLG2 (7/16)***	Cluster 2 specific gains ***RUFY4*** (10/16)	Cluster 3 ***MKL2* (5/15)***	Cluster 4 specific gains ***PLA2G12B* (10/21)***	Cluster 1 specific losses ***AFP-PHB2 (9/16)***	Cluster 2 specific losses ***F8-VWF, EWSR1-SLC22A24 (10/16)***	Cluster 3 specific losses ***FHL3-FHL2 (8/15)***	Cluster 4 specific losses ***EZH2-TDRD1 (9/21)***
LUSC	Secretory top gains ***HIST1H2AB* (5/5)***	Classical top gains ***HIST1H2AB* (7/7)***	Basal top gains ***RPS17* (2/2)***	Primitive top gains ***GNAS* (2/2)***	Secretory top losses ***NTRK1* (5/5)***	Classical top losses ***GDF9* (7/7)***	Basal top losses NTRK1* (2/2)	Primitive top losses ***NTRK1* (2/2)***
	Secretory specific gains ***PAEP* (5/5)***	Classical specific gains ***MPPED1* (7/7)***	Basal specific gains ***RPS17* (2/2)***	Primitive specific gains ***HEPACAM2* (2/2)***	Secretory specific losses ***CCDC33* (5/5)***	Classical specific losses ***CYP1A1*(7/7)***	Basal specific losses ***VIM* (2/2)***	Primitive specific losses ***POU2F11* (2/2)***
LUAD	TRU top gains ***CACNA1A**** **(15/15)**	PP top gains ***SGO1**** **(9/9)**	PI top gains ***GFAP**** **(21/21)**	TRU top losses ***APOA1**** **(15/15)**	PP top losses ***APOA1**** **(9/9)**	PI top losses ***APOA1**** **(21/21)**		
	TRU specific gains ***AMBP-FHL3*** **(6/15)**	PP specific gains ***PTTG1-NT5M, POU2F1-HOXB13 (5/9)***	PI specific gains ***MAST4-SMAD1*** **(12/21)**	TRU specific losses ***BCAR3**** **(9/15)**	PP specific losses ***MCOLN3-ST7L (7/9)***	PI specific losses ***POLE2-KLK5*** **(12/21)**		

STES	MSI top gains ***TNFSF11* (16/16)***	MSS top gains ***FOXH1* (21/21)***	MSI top losses ***SCN2B* (16/16)***	MSS top losses ***SCN2B* (21/21)***
	MSI specific gains ***KLF8* (7/16)***	MSS specific gains ***APOC3* (11/21)***	MSI specific losses ***SCGB3A1* (9/16)***	MSS specific losses ***NAV2* (10/21)***
KIRP	Type 1 top gains ***TP73*, IGF2BP3* (7/7)***	Type 2 top gains ***IGF2BP3*, CENPA* (16/16)***	Type 1 top losses ***CYP1A1* (7/7)***	Type 2 top losses ***NROB2* (16/16)***
	Type 1 specific gains ***PRKCG* (5/7)***	Type 2 specific gains ***CERS1* (13/16)***	Type 1 specific losses ***KCNG* (6/7)***	Type 2 specific losses ***TDRD1* (12/16)***
THCA	BRAF-like top gains ***KRT15*, ALK* (34/34)***	RAS-like top gains ***ALK*, FATE1* (12/12)***	BRAF-like top losses ***KRT85* (34/34)***	RAS-like top losses ***LYPD6* (12/12)***
	BRAF-like specific gains ***FTR-ADCY8 (31/34)***	RAS-like specific gains ***KLK7* (4/12)***	BRAF-like specific losses ***ODAM* (29/34)***	RAS-like specific losses ***JAKMIP2* (8/12)***

Supplementary Table III: Table showing a subset of cancer-specific edgetic perturbations in 13 cancer types.

Cancer type	Total number of gained edges	Number of cancer-specific gained edges	Total number of lost edges	Number of cancer-specific lost edges	Healthy PPIN size	Cancer PPIN size	p-value[a]
THCA	22831	1271	28065	797	175910	177146	2.71E-06
BLCA	20739	1463	19030	202	174770	175289	-
BRCA	22195	1453	25516	712	177033	174658	9.76E-16
COAD	10065	566	21024	953	180365	172933	1.49E-09
KIRC	18258	462	33005	887	179887	176147	8.83E-14
KIRP	17174	1085	27141	1117	178740	178300	-
KICH	12423	627	27490	1402	182708	176236	2.98E-08
HNSC	21913	1027	27485	877	181819	177203	1.92E-11
LUAD	16622	959	21907	259	177980	176455	1.93E-08
PRAD	17529	684	22468	915	179710	177677	3.97E-08
LUSC	13108	644	23242	1049	181377	175758	2.65E-10
STES	24326	1215	20800	835	177110	174708	1.92E-11
LIHC	36458	2019	51445	4068	175831	174337	0.004

Supplementary Table IV: Table showing the protein-protein interactions (711) between secreted proteins (iSPECS) and those from the cell lysate (Sharma). The proteins are either differentially expressed or their expression is 2.5-fold increased in a cell type compared to the other three cell types (termed cell type-specific proteins). As- astrocytes, MI- microglia, Ne- neurons and Ol- oligodendrocytes. (-): not enriched, (+): enriched, (1): cell type-specific enriched, (0): not cell type-specific enriched.

Gene in Sharma	Gene in iSPECS	Enriched in Astrocytes	Enriched in Microglia	Enriched in Neuron	Enriched in Oligodendrocytes	As_2.5fold	Mi_2.5fold	Ne_2.5fold	Ol_2.5fold	iSPECS	Sharma
Ablim1	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ablim1	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Acad10	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Acadm	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Acadvl	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Acox1	Hsd17b10	-	-	-	+	0	0	0	1	Oligodendrocytes	Oligodendrocytes
Acta2	Lasp1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Actbl2	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Actr2	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Actr3	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Adcy8	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Add1	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Add1	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Agap2	Cdc42	-	+	-	-	0	0	1	0	Microglia	Neuron
Agap2	Crmp1	-	-	+	-	0	0	1	0	Neuron	Neuron
Agap2	Ppp1cb	-	+	-	-	0	0	1	0	Microglia	Neuron
Agap2	Nptn	-	-	+	-	0	0	1	0	Neuron	Neuron
Agap2	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Agap2	Nlgn3	-	-	+	-	0	0	1	0	Neuron	Neuron
Agap2	Ncam2	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Agap2	Syne1	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Agap2	Nrcam	-	-	+	-	0	0	1	0	Neuron	Neuron
Agrn	Lamb1	-	-	-	+	0	0	0	1	Oligodendrocytes	Oligodendrocytes
Ahnak	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ahnak	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Akap2	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Alox5	Cotl1	-	+	-	-	0	1	0	0	Microglia	Microglia
Amotl1	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Apbb1ip	Tln1	-	+	-	-	0	1	0	0	Microglia	Microglia
Apc	Mapre2	-	+	-	-	0	0	1	0	Microglia	Neuron
Apc	Plau	-	+	-	-	0	0	1	0	Microglia	Neuron
Apc	Sparc	-	+	-	-	0	0	1	0	Microglia	Neuron

Arg2	Hsd17b10	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Arhgap33	Cdc42	-	+	-	-	0	0	1	0	Microglia	Neuron
Arpc1b	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Arpc2	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Arpc2	Uso1	-	+	-	-	0	1	0	0	Microglia	Microglia
Arpc4	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Arpc5	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Arrb2	Ppp1cb	-	+	-	-	0	1	0	0	Microglia	Microglia
Arvcf	Cdh15	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Arvcf	Cdh2	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Asb3	Pcbd1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Aspg	Hsd17b10	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Atad2	Hist1h4a	-	-	+	-	0	1	0	0	Neuron	Microglia
Atg16l1	Clic1	-	+	-	-	0	1	0	0	Microglia	Microglia
Atg16l1	Mat2a	-	+	-	-	0	1	0	0	Microglia	Microglia
Atg16l1	Ppp2r1a	-	+	-	-	0	1	0	0	Microglia	Microglia
Atg16l1	Hspe1	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Atg16l1	Vcp	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Atg16l1	Phgdh	-	+	-	-	0	1	0	0	Microglia	Microglia
Atg16l1	Vasp	-	+	-	-	0	1	0	0	Microglia	Microglia
Atg16l1	Ahcy	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Atg16l1	Grn	-	+	-	-	0	1	0	0	Microglia	Microglia
Atg16l1	Ppp1cb	-	+	-	-	0	1	0	0	Microglia	Microglia
Atg16l1	Fkbp3	-	+	-	-	0	1	0	0	Microglia	Microglia
Atg16l1	Hspd1	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Atg16l1	Ezr	-	+	-	-	0	1	0	0	Microglia	Microglia
Atg16l1	Nap1l1	-	+	-	-	0	1	0	0	Microglia	Microglia
Atg16l1	Tuba1b	-	-	+	-	0	1	0	0	Neuron	Microglia
Atg16l1	Cbln4	-	-	+	-	0	1	0	0	Neuron	Microglia
Atg16l1	Hsd17b10	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Atg16l1	Aldh2	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Atg16l1	Hspa9	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Atg16l1	Farsa	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Atg16l1	Lasp1	-	+	-	-	0	1	0	0	Microglia	Microglia
Atp6v0d1	Hsd17b10	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Atxn3	Vcp	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Atxn3	Tuba1a	-	-	+	-	0	1	0	0	Neuron	Microglia
Bag3	Capza1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Bclaf1	Fhl1	-	+	-	-	0	1	0	0	Microglia	Microglia

Bclaf1	Erh	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Bclaf1	Uso1	-	+	-	-	0	1	0	0	Microglia	Microglia
Becn1	(Sept11)	-	+	-	-	0	1	0	0	Microglia	Microglia
Bend3	(Sept11)	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Blzf1	Uso1	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Braf	Hspa9	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Brwd3	Rpl3	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Bsg	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Cald1	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Camk1	Ptpn6	-	+	-	-	0	1	0	0	Microglia	Microglia
Capn2	Tln1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Casp12	Vcp	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Casp8	Ppp2r1a	-	+	-	-	0	1	0	0	Microglia	Microglia
Casp9	Vcp	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Cav1	Vcp	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Cav1	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Cbl	Met	-	-	+	-	0	1	0	0	Neuron	Microglia
Cbl	Pdgfra	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Cbl	Egfr	-	-	+	-	0	1	0	0	Neuron	Microglia
Ccdc50	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Ccnb1	Ppp1cb	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Cd200r1	Cd200	-	-	+	-	0	1	0	0	Neuron	Microglia
Cd22	Ptpn6	-	+	-	-	0	1	0	0	Microglia	Microglia
Cd2ap	Prdx3	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Cd2ap	Capza1	-	+	-	-	0	1	0	0	Microglia	Microglia
Cd40	Ube2n	-	+	-	-	0	1	0	0	Microglia	Microglia
Cdh13	Adipoq	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Cdk2	Hist1h1c	-	-	+	-	0	1	0	0	Neuron	Microglia
Cdkn2a	Tuba1a	-	-	+	-	0	1	0	0	Neuron	Microglia
Cdkn2a	Rpl9	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Cdkn2aip	Hspd1	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Cebpb	Serpinh1	+	-	-	-	0	1	0	0	Astrocytes	Microglia
Cebpb	Sfpq	-	-	+	-	0	1	0	0	Neuron	Microglia
Cebpb	Hspa9	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Cep131	Ppp2r1a	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Cep131	Uso1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Cep290	Tuba1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Chn1	Epha4	-	-	+	-	0	0	1	0	Neuron	Neuron
Ckm	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Clint1	Uso1	-	+	-	-	0	1	0	0	Microglia	Microglia

Clta	Vcp	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Clta	Uso1	-	+	-	-	0	1	0	0	Microglia	Microglia
Clta	Ppp1cb	-	+	-	-	0	1	0	0	Microglia	Microglia
Clta	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Cltb	Ppp1cb	-	+	-	-	0	0	1	0	Microglia	Neuron
Cltb	Uso1	-	+	-	-	0	0	1	0	Microglia	Neuron
Cltb	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Clu	Plxna4	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Cntn6	Chl1	-	-	+	-	0	0	1	0	Neuron	Neuron
Cobl	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Cobl	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Cobll1	Dync2h1	+	-	-	-	1	0	0	0	Astrocytes	Astrocytes
Col1a1	Serpinf1	+	-	-	-	0	0	0	1	Astrocytes	Oligodendrocytes
Col1a1	Tgfb1	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Col5a1	Ppp2r1a	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Cpeb1	Aplp2	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Cpeb1	Aplp1	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Cpeb1	App	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Cpm	Coro1c	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Cpne2	Rdx	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Cpne4	Rdx	-	+	-	-	0	0	1	0	Microglia	Neuron
Cpt2	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Csf1r	Ptpn6	-	+	-	-	0	1	0	0	Microglia	Microglia
Csnk1e	Gpi	-	+	-	-	0	0	1	0	Microglia	Neuron
Csnk1e	Cps1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Csnk1e	Uso1	-	+	-	-	0	0	1	0	Microglia	Neuron
Ctsc	Ppp2r1a	-	+	-	-	0	1	0	0	Microglia	Microglia
Cux1	Ppp2r1a	-	+	-	-	0	1	0	0	Microglia	Microglia
Cybrd1	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Cyfip1	Crmp1	-	-	+	-	0	0	1	0	Neuron	Neuron
Cyfip1	Crmp1	-	-	+	-	0	1	0	0	Neuron	Microglia
Cyfip1	Ncan	-	-	+	-	0	0	1	0	Neuron	Neuron
Cyfip1	Ncan	-	-	+	-	0	1	0	0	Neuron	Microglia
Dab1	Notch1	-	-	+	-	0	0	1	0	Neuron	Neuron
Dab1	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Dab1	Lrp2	-	+	-	-	0	0	1	0	Microglia	Neuron
Dab1	Aplp1	-	-	+	-	0	0	1	0	Neuron	Neuron
Dab2	Efnb2	-	-	+	-	0	1	0	0	Neuron	Microglia
Dab2	Ppp1cb	-	+	-	-	0	1	0	0	Microglia	Microglia
Dab2	Lrp2	-	+	-	-	0	1	0	0	Microglia	Microglia
Dag1	Egflam	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Dbn1	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Dbn1	Ppp1cb	-	+	-	-	0	0	1	0	Microglia	Neuron

Dctn1	Capza1	-	+	-	-	0	1	0	0	Microglia	Microglia
Dctn1	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Dctn4	Capza1	-	+	-	-	0	0	1	0	Microglia	Neuron
Decr2	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Dhtkd1	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Dlg1	Nrcam	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Dlg1	Adam10	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Dlgap1	Nrxn1	-	-	+	-	0	0	1	0	Neuron	Neuron
Dlgap1	Lgi1	-	-	+	-	0	0	1	0	Neuron	Neuron
Dlgap1	Lrrtm1	-	-	+	-	0	0	1	0	Neuron	Neuron
Dlgap1	(Sept11)	-	+	-	-	0	0	1	0	Microglia	Neuron
Dlgap1	Syne1	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Dlgap1	Ube2n	-	+	-	-	0	0	1	0	Microglia	Neuron
Dlgap4	B4galt1	-	+	-	-	0	0	1	0	Microglia	Neuron
Dnaja3	Hspa9	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Dnajb2	Serpinb5	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Dnm1	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Dock11	Cdc42	-	+	-	-	0	0	1	0	Microglia	Neuron
Dock11	Cdc42	-	+	-	-	0	1	0	0	Microglia	Microglia
Dock9	Cdc42	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Dok1	Phgdh	-	+	-	-	0	1	0	0	Microglia	Microglia
Dok1	Ptpn6	-	+	-	-	0	1	0	0	Microglia	Microglia
Dpysl3	Crmp1	-	-	+	-	0	0	1	0	Neuron	Neuron
Dpysl5	Dpys	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Ech1	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Eef1d	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Efnb2	Epha4	-	-	+	-	0	0	1	0	Neuron	Neuron
Elavl4	Sfpq	-	-	+	-	0	0	1	0	Neuron	Neuron
Elf1	Ephb4	-	+	-	-	0	1	0	0	Microglia	Microglia
Elp3	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Epha2	Epha4	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Epha3	L1cam	-	-	+	-	0	0	1	0	Neuron	Neuron
Epha4	Efnb2	-	-	+	-	0	0	1	0	Neuron	Neuron
Epha7	Chl1	-	-	+	-	0	0	1	0	Neuron	Neuron
Erbb2	Egfr	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Ezr	Rdx	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Fbxl16	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Fbxl16	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Fbxo2	Vcp	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Fbxo21	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Fbxw7	Deptor	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Fech	Ppp2r1a	-	+	-	-	1	0	0	0	Microglia	Astrocytes

Fhl1	Tln1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Flii	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Flii	Ppp1cb	-	+	-	-	0	1	0	0	Microglia	Microglia
Flnb	Rala	+	-	-	-	1	0	0	0	Astrocytes	Astrocytes
Flnb	Cd44	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Flnb	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Flnb	Capza1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Flrt2	Fgfr1	-	-	+	-	0	0	1	0	Neuron	Neuron
Flrt3	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Flrt3	Fgfr1	-	-	+	-	0	0	1	0	Neuron	Neuron
Fmnl2	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Fmnl2	Coro1c	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Fmnl3	Ppp2r1a	-	+	-	-	0	1	0	0	Microglia	Microglia
Fyn	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
G6pdx	Tagln	-	+	-	-	0	1	0	0	Microglia	Microglia
Gab1	Met	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Gabra1	Nlgn3	-	-	+	-	0	0	1	0	Neuron	Neuron
Gabra1	Nrxn1	-	-	+	-	0	0	1	0	Neuron	Neuron
Gak	Ppp2r1a	-	+	-	-	0	1	0	0	Microglia	Microglia
Gak	Uso1	-	+	-	-	0	1	0	0	Microglia	Microglia
Gapdh	Hsd17b10	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Gcg	Hist1h4a	-	-	+	-	0	0	1	0	Neuron	Neuron
Gcg	Fkbp3	-	+	-	-	0	0	1	0	Microglia	Neuron
Gcg	Mdh1	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Gcg	Tubb4a	-	+	-	-	0	0	1	0	Microglia	Neuron
Gcg	Hist1h2bf	-	-	+	-	0	0	1	0	Neuron	Neuron
Gcg	Grn	-	+	-	-	0	0	1	0	Microglia	Neuron
Gcg	Stmn2	-	-	+	-	0	0	1	0	Neuron	Neuron
Gcg	Phgdh	-	+	-	-	0	0	1	0	Microglia	Neuron
Gcg	Tuba1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Gcg	Scg2	-	-	+	-	0	0	1	0	Neuron	Neuron
Gcg	Hspd1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Gcg	Hspe1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Gcg	Pcsk2	-	-	+	-	0	0	1	0	Neuron	Neuron
Gcg	Hist1h1c	-	-	+	-	0	0	1	0	Neuron	Neuron
Gcg	Vcp	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Gcg	Hspa9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Gcg	Mthfd1	-	+	-	-	0	0	1	0	Microglia	Neuron
Gga1	Scg2	-	-	+	-	0	1	0	0	Neuron	Microglia
Gga1	App	-	-	+	-	0	1	0	0	Neuron	Microglia
Gga1	Adipoq	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Ghitm	Sarm1	+	-	-	-	1	0	0	0	Astrocytes	Astrocytes

Glud1	Ppp2r1a	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Gna11	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Gnb2	Uso1	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Gria2	Cspg4	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Grin1	Pfkl	-	+	-	-	0	0	1	0	Microglia	Neuron
Grin1	Cdh2	-	-	+	-	0	0	1	0	Neuron	Neuron
Grin1	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Grin2b	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Grin2b	Cdh2	-	-	+	-	0	0	1	0	Neuron	Neuron
Grin2b	Rala	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Grip1	Fras1	-	-	+	-	0	0	1	0	Neuron	Neuron
Grip1	Cspg4	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Gsta4	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Gstk1	Adipoq	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Gstm1	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Gstm2	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Gstm5	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Gstt1	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Hace1	Tubb4a	-	+	-	-	0	0	1	0	Microglia	Neuron
Hace1	Tpm2	-	+	-	-	0	0	1	0	Microglia	Neuron
Hace1	Capza1	-	+	-	-	0	0	1	0	Microglia	Neuron
Hace1	Serpinh1	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Hace1	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Hace1	Fhl1	-	+	-	-	0	0	1	0	Microglia	Neuron
Hace1	Hspa9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Hace1	Hspd1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Hace1	Csrp1	-	+	-	-	0	0	1	0	Microglia	Neuron
Hecw1	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Hibadh	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Hk2	Pea15	-	+	-	-	0	1	0	0	Microglia	Microglia
Hmg20a	Kif20b	-	+	-	-	0	0	1	0	Microglia	Neuron
Hmgcl	Hsd17b10	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Hmgcs1	Hsd17b10	-	-	-	+	0	0	0	1	Oligodendrocytes	Oligodendrocytes
Hnrnpk	Rpl3	-	-	+	-	0	0	1	0	Neuron	Neuron
Hnrnpk	Phgdh	-	+	-	-	0	0	1	0	Microglia	Neuron
Hnrnpk	Hspd1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Hnrnpk	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Hnrnpk	Ugdh	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Hnrnpk	Tuba1b	-	-	+	-	0	0	1	0	Neuron	Neuron

Hnrnpk	Sfpq	-	-	+	-	0	0	1	0	Neuron	Neuron
Hras	Hspd1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Hspb6	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Htra1	Tgfb1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Icam2	Rdx	-	+	-	-	0	1	0	0	Microglia	Microglia
Immt	Mat2a	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Inadl	Tuba1a	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Inpp5d	Tjp2	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Iqcb1	Nap1l1	-	+	-	-	0	0	1	0	Microglia	Neuron
Iqcb1	Sort1	-	-	+	-	0	0	1	0	Neuron	Neuron
Iqcb1	Kpnb1	-	+	-	-	0	0	1	0	Microglia	Neuron
Iqcb1	Glud1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Iqcb1	Tuba1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Iqcb1	Hspa9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Iqcb1	Nucb1	-	+	-	-	0	0	1	0	Microglia	Neuron
Iqcb1	Map1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Iqcb1	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Iqgap1	Cdc42	-	+	-	-	0	1	0	0	Microglia	Microglia
Irak1	Il1rap	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Irak2	Sarm1	+	-	-	-	0	1	0	0	Astrocytes	Microglia
Itga3	Ptprm	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Itga5	Ppp1cb	-	+	-	-	0	1	0	0	Microglia	Microglia
Itga5	Actr3	-	+	-	-	0	1	0	0	Microglia	Microglia
Itga5	Tpm2	-	+	-	-	0	1	0	0	Microglia	Microglia
Itgb4	Egfr	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Itpr3	Hspa9	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Ivd	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Jag2	Notch1	-	-	+	-	0	0	1	0	Neuron	Neuron
Jak2	Ptpn6	-	+	-	-	0	0	1	0	Microglia	Neuron
Jak2	Egfr	-	-	+	-	0	0	1	0	Neuron	Neuron
Kalrn	Cdh10	-	-	+	-	0	0	1	0	Neuron	Neuron
Kat2a	Hist1h4a	-	-	+	-	0	0	1	0	Neuron	Neuron
Kat2a	Notch1	-	-	+	-	0	0	1	0	Neuron	Neuron
Kat5	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Kat5	Sfpq	-	-	+	-	0	0	1	0	Neuron	Neuron
Kcnma1	Capg	-	+	-	-	0	0	1	0	Microglia	Neuron
Kcnma1	Actr3	-	+	-	-	0	0	1	0	Microglia	Neuron
Kcnma1	Tagln	-	+	-	-	0	0	1	0	Microglia	Neuron
Kcnma1	Sparc	-	+	-	-	0	0	1	0	Microglia	Neuron
Kcnma1	Kng1	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Kcnma1	Hspa9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron

Kcnma1	Hspd1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Kcnma1	Lrpap1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Kcnma1	Eno3	-	+	-	-	0	0	1	0	Microglia	Neuron
Kcnma1	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Kcnma1	Apoa1	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Kcnma1	Nudc	-	+	-	-	0	0	1	0	Microglia	Neuron
Kcnma1	Glud1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Kcnma1	Vcp	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Kcnma1	Apoh	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Kcnma1	Hpx	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Kcnma1	Nucb1	-	+	-	-	0	0	1	0	Microglia	Neuron
Kcnma1	Rcn3	-	+	-	-	0	0	1	0	Microglia	Neuron
Kcnma1	Phgdh	-	+	-	-	0	0	1	0	Microglia	Neuron
Kiaa0196	Capza1	-	+	-	-	0	1	0	0	Microglia	Microglia
Kiaa1033	Capza1	-	+	-	-	0	1	0	0	Microglia	Microglia
Kiaa2013	Ppp2r1a	-	+	-	-	0	1	0	0	Microglia	Microglia
Kif13b	Icam5	-	-	+	-	0	1	0	0	Neuron	Microglia
Kif5c	Cdh2	-	-	+	-	0	0	1	0	Neuron	Neuron
Kifap3	Cdh2	-	-	+	-	0	0	1	0	Neuron	Neuron
Kifc5b	Kpnb1	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Kras	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Kras	Egfr	-	-	+	-	0	0	1	0	Neuron	Neuron
Ksr1	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Ksr1	Hspa9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Ksr1	Nap1l1	-	+	-	-	0	0	1	0	Microglia	Neuron
Ksr1	Rps3a	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Ksr1	Rpl9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Ksr1	Tubb4a	-	+	-	-	0	0	1	0	Microglia	Neuron
Ksr1	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Ktn1	Cdc42	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Lama1	Lamb1	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Lama1	Ache	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Lamc1	Lamb1	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Ldha	Hsd17b10	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Ldhb	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Lgals3bp	Cps1	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Lgals3bp	Uso1	-	+	-	-	0	1	0	0	Microglia	Microglia
Lgals9	Cd44	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Lilrb4	Ptpn6	-	+	-	-	0	1	0	0	Microglia	Microglia

Lin7c	Nrxn1	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Lin7c	Sfpq	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Lingo1	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Lmtk2	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Lnpep	Uso1	-	+	-	-	0	1	0	0	Microglia	Microglia
Lrp2	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Lrp6	Igfbp4	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Lrp6	Cdh2	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Lrr1	Sugt1	-	+	-	-	0	1	0	0	Microglia	Microglia
Lrr1	Tuba1a	-	-	+	-	0	1	0	0	Neuron	Microglia
Lrrc4c	Ntng1	-	-	+	-	0	0	1	0	Neuron	Neuron
Lrrfip1	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Lrrk2	Tubb4a	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Lrrk2	Cdc42	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ltbp3	Tgfb1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Lyn	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Lzts2	Uso1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Lzts2	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Mad2l1	Nap1l1	-	+	-	-	0	1	0	0	Microglia	Microglia
Mad2l1	Cps1	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Mad2l1	Pfkl	-	+	-	-	0	1	0	0	Microglia	Microglia
Map1lc3a	Map1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Mapk14	Egfr	-	-	+	-	0	1	0	0	Neuron	Microglia
Mapk8	Sarm1	+	-	-	-	0	1	0	0	Astrocytes	Microglia
Mapk8	Sarm1	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Mapk8ip1	Lrp1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Mapk8ip1	Lrp2	-	+	-	-	0	0	1	0	Microglia	Neuron
Mapk8ip1	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Mapre3	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Mapre3	Map1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Mapt	Cand1	-	+	-	-	0	0	1	0	Microglia	Neuron
Mapt	Mapre2	-	+	-	-	0	0	1	0	Microglia	Neuron
Mapt	Otub1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Mapt	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Mapt	Pccb	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Mapt	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Mapt	Hspd1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Mapt	Pygb	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Mapt	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Mapt	Hsd17b10	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Mapt	Tuba1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Mapt	Vcp	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron

Mapt	Phgdh	-	+	-	-	0	0	1	0	Microglia	Neuron
Mapt	Pfkl	-	+	-	-	0	0	1	0	Microglia	Neuron
Mapt	Hist1h4a	-	-	+	-	0	0	1	0	Neuron	Neuron
Max	Prdx3	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Max	Ppp1cb	-	+	-	-	0	1	0	0	Microglia	Microglia
Mbp	Ptpn6	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Mdm2	Rpl3	-	-	+	-	0	1	0	0	Neuron	Microglia
Mdm2	Rps3a	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Mdm2	Ezr	-	+	-	-	0	1	0	0	Microglia	Microglia
Mdm2	Rpl9	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Mdm2	Fkbp3	-	+	-	-	0	1	0	0	Microglia	Microglia
Mef2c	Acly	-	+	-	-	0	0	1	0	Microglia	Neuron
Mks1	Hist1h4a	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Mks1	Hspa9	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Mks1	Tuba1b	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Mme	Adam10	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Mprip	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Mprip	Uso1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Mprip	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Mpzl1	Ppp2r1a	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Mycbp2	Hist1h4a	-	-	+	-	0	0	1	0	Neuron	Neuron
Mycbp2	Hspa9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Mycbp2	Ppp1cb	-	+	-	-	0	0	1	0	Microglia	Neuron
Mycbp2	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Mycbp2	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Myd88	Il1rap	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Myo18a	Coro1c	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Myo18a	Ppp1cb	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Myo1b	Ppp1cb	-	+	-	-	0	0	1	0	Microglia	Neuron
Myo1b	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Myo1e	Ppp1cb	-	+	-	-	0	1	0	0	Microglia	Microglia
Myo5a	Ppp1cb	-	+	-	-	0	0	1	0	Microglia	Neuron
Myo5a	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Nbr1	Map1b	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Ncam1	Cntn2	-	-	+	-	0	0	1	0	Neuron	Neuron
Ncam1	Fgfr1	-	-	+	-	0	0	1	0	Neuron	Neuron
Ncoa3	Ptpn6	-	+	-	-	0	1	0	0	Microglia	Microglia
Ndn	Cdh4	-	-	+	-	0	0	1	0	Neuron	Neuron
Ndn	Cdh5	-	+	-	-	0	0	1	0	Microglia	Neuron
Ndn	Otub1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Ndn	Lamb1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Ndn	Cdh2	-	-	+	-	0	0	1	0	Neuron	Neuron

Ndn	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Ndn	Nucb1	-	+	-	-	0	0	1	0	Microglia	Neuron
Ndufa7	Hsd17b10	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Nexn	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Nfasc	Fgfr1	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Nid2	Prelp	+	-	-	-	1	0	0	0	Astrocytes	Astrocytes
Nlrx1	Sarm1	+	-	-	-	1	0	0	0	Astrocytes	Astrocytes
Nphp1	Hspa9	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Nphp1	Hist1h4a	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Nphp1	Tuba1b	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Nphp1	Hspd1	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Nrp2	Sema3f	+	-	-	-	0	1	0	0	Astrocytes	Microglia
Ntrk2	Sort1	-	-	+	-	0	0	1	0	Neuron	Neuron
Oat	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Ogg1	Lamb1	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Osmr	Egfr	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Ostm1	Ppp2r1a	-	+	-	-	0	1	0	0	Microglia	Microglia
Ostm1	Kpnb1	-	+	-	-	0	1	0	0	Microglia	Microglia
Pacsin1	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Pacsin3	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Pacsin3	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Pacsin3	Tuba1a	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Pafah1b2	Nudc	-	+	-	-	0	0	1	0	Microglia	Neuron
Pafah1b3	Pafah1b2	-	+	-	-	0	0	1	0	Microglia	Neuron
Palm	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Papss1	Apoa1	+	-	-	-	0	0	0	1	Astrocytes	Oligodendrocytes
Papss1	Ppp2r1a	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Park2	Map1b	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Park2	Egfr	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Parp10	B4galt1	-	+	-	-	0	1	0	0	Microglia	Microglia
Parp14	Gpi	-	+	-	-	0	1	0	0	Microglia	Microglia
Parva	Lims1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Pawr	Pcbd1	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Pcgf2	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Pdlim5	Tpm2	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Pfn1	Vcp	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Pick1	Lrp1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Pkd2	Vcp	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Pkd2	Ppp2r1a	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Pkm	Hsd17b10	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Pkp2	Uso1	-	+	-	-	1	0	0	0	Microglia	Astrocytes

Plcb1	Ola1	-	+	-	-	0	0	1	0	Microglia	Neuron
Plcb1	Rpl3	-	-	+	-	0	0	1	0	Neuron	Neuron
Plcb1	Tln1	-	+	-	-	0	0	1	0	Microglia	Neuron
Plcb1	Hspa9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Plcb1	Kpnb1	-	+	-	-	0	0	1	0	Microglia	Neuron
Plcb1	Cdc42	-	+	-	-	0	0	1	0	Microglia	Neuron
Plcb1	Rpl9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Plcb1	Sub1	-	+	-	-	0	0	1	0	Microglia	Neuron
Plcd3	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Plk1	Sugt1	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Plxna1	Trem2	-	+	-	-	0	0	1	0	Microglia	Neuron
Plxna2	Sema6b	-	-	+	-	0	0	1	0	Neuron	Neuron
Plxna4	Sema6b	-	-	+	-	0	0	1	0	Neuron	Neuron
Plxnb3	Sema4c	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Plxnd1	Sema4c	-	-	+	-	0	0	1	0	Neuron	Neuron
Ppfia3	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Ppp1r12a	Mylk	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ppp1r12a	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ppp1r12a	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ppp1r12a	Uso1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ppp1r12a	Erh	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Ppp1r12b	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ppp1r13l	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ppp2r2b	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Ppp2r2d	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Ppp2r3c	Ppp2r1a	-	+	-	-	0	1	0	0	Microglia	Microglia
Ppp2r5c	Ppp2r1a	-	+	-	-	0	0	1	0	Microglia	Neuron
Prdx1	Hsd17b10	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Prdx5	Hsd17b10	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Prdx6	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Prkaca	Lasp1	-	+	-	-	0	0	1	0	Microglia	Neuron
Prkaca	Uso1	-	+	-	-	0	0	1	0	Microglia	Neuron
Prkcdbp	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Prkcz	Lasp1	-	+	-	-	0	0	1	0	Microglia	Neuron
Prkg1	Lasp1	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Prss23	Ppp2r1a	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Psap	Sort1	-	-	+	-	0	1	0	0	Neuron	Microglia
Psen2	Notch1	-	-	+	-	0	1	0	0	Neuron	Microglia
Pten	Vcp	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Pten	Rps4x	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Pten	Capn2	-	+	-	-	0	0	1	0	Microglia	Neuron

Pten	Idh1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Pten	Rpl3	-	-	+	-	0	0	1	0	Neuron	Neuron
Pten	Map1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Pten	Acly	-	+	-	-	0	0	1	0	Microglia	Neuron
Pten	Aldh1a1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Pten	Actr3	-	+	-	-	0	0	1	0	Microglia	Neuron
Pten	Tuba1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Pten	Serpinh1	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Pten	Hspd1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Pten	Slc9a3r1	-	+	-	-	0	0	1	0	Microglia	Neuron
Pten	Nap1l1	-	+	-	-	0	0	1	0	Microglia	Neuron
Pten	Ugdh	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Ptgs1	Nucb1	-	+	-	-	0	1	0	0	Microglia	Microglia
Ptgs2	Nucb1	-	+	-	-	0	1	0	0	Microglia	Microglia
Ptgs2	Vcp	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Ptpn1	Cdh2	-	-	+	-	0	1	0	0	Neuron	Microglia
Ptprr	Ntm	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Ptprs	Ptprm	-	+	-	-	0	0	1	0	Microglia	Neuron
Pygb	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Rab11fip5	Ppp2r1a	-	+	-	-	0	1	0	0	Microglia	Microglia
Rab3a	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Rabl3	Uso1	-	+	-	-	0	1	0	0	Microglia	Microglia
Racgap1	Pak2	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Ralb	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Rap1b	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Rbpj	Fhl1	-	+	-	-	0	1	0	0	Microglia	Microglia
Rbpj	Notch1	-	-	+	-	0	1	0	0	Neuron	Microglia
Rcor2	Notch1	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Rin1	Epha4	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Rmi1	Kpnb1	-	+	-	-	0	1	0	0	Microglia	Microglia
Rnf31	Vcp	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Rtn4	Rtn4r	-	-	+	-	0	1	0	0	Neuron	Microglia
Rxra	Cxadr	-	-	+	-	0	1	0	0	Neuron	Microglia
Rxra	Sfpq	-	-	+	-	0	1	0	0	Neuron	Microglia
Sass6	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Sass6	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Scai	Pfkl	-	+	-	-	0	0	1	0	Microglia	Neuron
Sema3a	Cntn2	-	-	+	-	0	0	1	0	Neuron	Neuron
Sema3a	Plxna4	-	-	+	-	0	0	1	0	Neuron	Neuron
Sema3e	Plxnd1	-	+	-	-	0	0	1	0	Microglia	Neuron
Sema4f	Plxnb3	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron

Sema4g	Plxnb3	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Sema5a	Plxna4	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Sema5a	Plxna2	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Sema5a	Plxnb3	-	-	-	+	0	0	0	1	Oligodendrocytes	Oligodendrocytes
Sema5b	Plxna2	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Sema5b	Plxna4	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Sema6a	Plxna4	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Sema6a	Plxna2	-	-	+	-	0	0	0	1	Neuron	Oligodendrocytes
Shank3	Mdk	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Shank3	Nrcam	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Shank3	Syne1	+	-	-	-	1	0	0	0	Astrocytes	Astrocytes
Shank3	Ncan	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Shc3	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Sidt2	Lims1	-	+	-	-	0	1	0	0	Microglia	Microglia
Slc17a7	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Slc26a6	Slc9a3r1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Slc35b1	B4galt1	-	+	-	-	0	0	1	0	Microglia	Neuron
Slc4a10	Slc9a3r1	-	+	-	-	0	0	1	0	Microglia	Neuron
Slc8a1	Fbln5	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Slc8a1	Fbln5	-	+	-	-	0	0	1	0	Microglia	Neuron
Smarcd3	Notch1	-	-	+	-	0	0	1	0	Neuron	Neuron
Smn1	Glud1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Snap25	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Snap25	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Snca	Tubb4a	-	+	-	-	0	0	1	0	Microglia	Neuron
Snca	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Snca	Hspa9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Snca	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Snca	Map1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Snca	Icam5	-	-	+	-	0	0	1	0	Neuron	Neuron
Snca	Ntm	-	-	+	-	0	0	1	0	Neuron	Neuron
Sncb	Tuba1a	-	-	+	-	0	0	1	0	Neuron	Neuron
Sod1	Chgb	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Sorbs2	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Sorbs2	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Sox9	Kpnb1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Specc1l	Uso1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Specc1l	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Srebf1	Pcbd1	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Srgap1	Cdc42	-	+	-	-	0	0	1	0	Microglia	Neuron
Ssh2	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
St5	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Stat1	Egfr	-	-	+	-	0	1	0	0	Neuron	Microglia

Stat1	Ptpn6	-	+	-	-	0	1	0	0	Microglia	Microglia
Stmn1	Vcp	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Ston2	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ston2	Uso1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Stt3a	Sarm1	+	-	-	-	0	1	0	0	Astrocytes	Microglia
Sugct	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Sumf1	Sumf2	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Sun1	Syne1	+	-	-	-	1	0	0	0	Astrocytes	Astrocytes
Suox	Hsd17b10	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Sv2a	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Sv2b	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Svip	Vcp	-	-	-	+	0	0	0	1	Oligodendrocytes	Oligodendrocytes
Syn1	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Syn2	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Syngap1	Cand1	-	+	-	-	0	0	1	0	Microglia	Neuron
Syngap1	Egfr	-	-	+	-	0	0	1	0	Neuron	Neuron
Syngap1	Ncan	-	-	+	-	0	0	1	0	Neuron	Neuron
Synj2bp	Lrp2	-	+	-	-	0	1	0	0	Microglia	Microglia
Syt1	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Syt1	Lrp1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Tagln	G6pdx	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tfe3	Ahcy	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Tfe3	Vcp	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Tfe3	Rpl3	-	-	+	-	0	1	0	0	Neuron	Microglia
Tfe3	Hist1h4a	-	-	+	-	0	1	0	0	Neuron	Microglia
Tfe3	Sfpq	-	-	+	-	0	1	0	0	Neuron	Microglia
Tfe3	Idh1	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Tfe3	Tuba1b	-	-	+	-	0	1	0	0	Neuron	Microglia
Tfe3	Hspe1	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Tfe3	Acly	-	+	-	-	0	1	0	0	Microglia	Microglia
Tfe3	Hspd1	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Tfe3	Rps4x	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Tfe3	Hist1h1c	-	-	+	-	0	1	0	0	Neuron	Microglia
Tfe3	Kif20b	-	+	-	-	0	1	0	0	Microglia	Microglia
Tfe3	Alpl	-	+	-	-	0	1	0	0	Microglia	Microglia
Tfe3	Atic	-	+	-	-	0	1	0	0	Microglia	Microglia
Tfe3	Hspa9	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Tfe3	Rpl9	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Tgfb1	Emilin1	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia

Tgfbr2	Tgfbr3	-	+	-	-	0	1	0	0	Microglia	Microglia
Tgfbr2	Cdh5	-	+	-	-	0	1	0	0	Microglia	Microglia
Thra	Sfpq	-	-	+	-	0	0	1	0	Neuron	Neuron
Tia1	Mdh1	+	-	-	-	0	0	1	0	Astrocytes	Neuron
Tia1	Hspa9	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Tia1	Otub1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Tia1	Ppp1cb	-	+	-	-	0	0	1	0	Microglia	Neuron
Tia1	Glud1	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Tia1	Rps4x	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Tiam1	Cd44	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Tiam2	Cd44	-	-	-	+	0	0	1	0	Oligodendrocytes	Neuron
Timp2	Pcsk5	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Tln1	Fhl1	-	+	-	-	0	1	0	0	Microglia	Microglia
Tln1	Capn2	-	+	-	-	0	1	0	0	Microglia	Microglia
Tlr4	Cd14	+	-	-	-	0	1	0	0	Astrocytes	Microglia
Tlr4	Ube2n	-	+	-	-	0	1	0	0	Microglia	Microglia
Tmem237	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tmod1	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tmod1	Uso1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tmod1	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tmsb4x	Lims1	-	+	-	-	0	1	0	0	Microglia	Microglia
Tmsb4x	Aplp1	-	-	+	-	0	1	0	0	Neuron	Microglia
Tnfaip3	Ube2n	-	+	-	-	0	1	0	0	Microglia	Microglia
Tnks1bp1	Capza1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tpm1	Rala	+	-	-	-	1	0	0	0	Astrocytes	Astrocytes
Tpm1	Uso1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tpm1	Tpm2	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tpm1	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tpm1	Capza1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tpm1	Actr3	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tpm2	Coro1c	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Tprn	Coro1c	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Traf1	Ahcy	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Traf3ip1	Uso1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Trim59	Capza1	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Trim67	Dcc	-	-	+	-	0	0	1	0	Neuron	Neuron
Ttc23	Tuba1a	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Ttc23	Sarm1	+	-	-	-	1	0	0	0	Astrocytes	Astrocytes
Ttc23	Sfpq	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Ttc23	Erh	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes
Ttc23	Hspa9	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes

Ttc23	Tuba1b	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Ttc23	Egfr	-	-	+	-	1	0	0	0	Neuron	Astrocytes
Ttc7a	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Tuba1a	Uso1	-	+	-	-	0	0	1	0	Microglia	Neuron
Tubb3	Map1b	-	-	+	-	0	0	1	0	Neuron	Neuron
Tubb4b	Phgdh	-	+	-	-	0	0	1	0	Microglia	Neuron
Twf2	Coro1c	-	+	-	-	0	1	0	0	Microglia	Microglia
Uaca	Ppp1cb	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Ube2o	App	-	-	+	-	0	0	1	0	Neuron	Neuron
Ube2q1	B4galt1	-	+	-	-	0	1	0	0	Microglia	Microglia
Ube2v1	Ube2n	-	+	-	-	0	1	0	0	Microglia	Microglia
Unc5b	Ppp2r1a	-	+	-	-	0	0	0	1	Microglia	Oligodendrocytes
Vasp	Lasp1	-	+	-	-	0	1	0	0	Microglia	Microglia
Vav1	Egfr	-	-	+	-	0	1	0	0	Neuron	Microglia
Vcl	Tln1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Wdfy4	(Sept11)	-	+	-	-	0	1	0	0	Microglia	Microglia
Wdfy4	Cd44	-	-	-	+	0	1	0	0	Oligodendrocytes	Microglia
Wdfy4	Tpm2	-	+	-	-	0	1	0	0	Microglia	Microglia
Wdfy4	Sort1	-	-	+	-	0	1	0	0	Neuron	Microglia
Wdfy4	Actr3	-	+	-	-	0	1	0	0	Microglia	Microglia
Xpr1	Coro1c	-	+	-	-	0	0	1	0	Microglia	Neuron
Zbtb25	B4galt1	-	+	-	-	1	0	0	0	Microglia	Astrocytes
Zbtb7b	Hist1h1c	-	-	+	-	0	1	0	0	Neuron	Microglia
Zfp36l1	Pcbd1	-	-	-	+	1	0	0	0	Oligodendrocytes	Astrocytes