UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE
CURSO DE GRADUAÇÃO EM BIOMEDICINA

Thiago Steindorff

ANÁLISE TRANSCRIPTÔMICA PÂN-CANCER SUGERE IMPORTÂNCIA LISOSSOMAL E REVELA NOVOS BIOMARCADORES PARA GLIOBLASTOMA MULTIFORME

Porto Alegre
2023

Thiago Steindorff

ANÁLISE TRANSCRIPTÔMICA PÂN-CANCER LISOSSOMAL SUGERE IMPORTÂNCIA LISOSSOMAL E REVELA NOVOS BIOMARCADORES PARA GLIOBLASTOMA MULTIFORME

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel(a) em Biomedicina.

Orientador(a): Profa Dra Ursula da Silveira Matte

Porto Alegre

2023

Dados Internacionais de Catalogação na Publicação

CIP - Catalogação na Publicação

Steindorff, Thiago
ANÁLISE TRANSCRIPTÔMICA PAN-CANCER LISOSSOMAL SUGERE IMPORTANCIA LISOSSOMAL E REVELA NOVOS BIOMARCADORES PARA GLIOBLASTOMA MULTIFORME / Thiago Steindorff. -- 2023.
77 f.
Orientadora: Ursula da Silveira Matte.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Ciências Básicas da Saúde, Curso de Biomedicina, Porto Alegre, BR-RS, 2023.

1. Genética. 2. Lisossomos. 3. Transcriptômica. 4. Câncer. 5. Bioinformática. I. da Silveira Matte, Ursula, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Thiago Steindorff

ANÁLISE TRANSCRIPTÔMICA PÂN-CANCER LISOSSOMAL SUGERE IMPORTÂNCIA LISOSSOMAL E REVELA NOVOS BIOMARCADORES PARA GLIOBLASTOMA MULTIFORME

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel(a) em Biomedicina.

Aprovado em: ____ de _______ de _____.

BANCA EXAMINADORA

____

Prof. Dr. Eduardo Cremonese Filippi Chiela - UFRGS

____

Dra. Thayne Woycinck Kowalski - UFRGS

____

Profa. Dra. Ursula da Silveira Matte - UFRGS (orientadora)

RESUMO

Estudos pan-câncer tentam revelar semelhanças e diferenças entre as características moleculares de diferentes tipos de câncer, podendo levar a descoberta de novos biomarcadores, terapias e informações sobre a biologia tumoral. Estudos anteriores já revelaram informações sobre interações de células tumorais e células imunes, caracterizaram mutações e diferenças epigenéticas entre os principais tipos de cânceres. Devido aos grandes tamanhos amostrais e a primazia pela uniformidade metodológica na coleta de dados, o *The Cancer Genome Atlas* (TCGA) é a principal fonte de dados moleculares para a pesquisa em câncer. Análises de expressão diferencial realizadas com dados do TCGA revelaram inúmeros possíveis biomarcadores de diagnóstico para os mais diversos tipos de cânceres e também trouxeram novas informações sobre seus mecanismos patológicos. O modelo de regressão de Cox integrado com dados de RNA-seq amplia o poder de estudos transcriptômicos e já permitiu a descoberta de muitos genes e assinaturas gênicas marcadores de prognóstico para tumores humanos. Nas últimas décadas, lisossomos têm cada vez mais sido associados ao estabelecimento e patologia de cânceres humanos. Nesse estudo nós utilizamos análise de expressão gênica diferencial e análise de Kaplan-Meier associada a regressão de Cox para a avaliar a possibilidade de utilização de genes lisossomais como biomarcadores em 29 cânceres humanos a partir de dados do TCGA. Nós construímos um conjunto de genes lisossomais a partir das bases Kyoto Encyclopedia of Genes and Genomes (KEGG) e The Human Lysosomal Gene Database (hLGDB) e investigamos o estado transcricional destes genes nos diferentes tipos de câncer. O único tipo de câncer para o qual encontramos fortes indícios do papel dos genes lisossomais foi o Glioblastoma Multiforme (GBM). Genes lisossômicos envolvidos na resposta imune apresentaram expressão aumentada em GBM no nosso estudo e o aumento da expressão de genes lisossômicos relacionados à degradação de macromoléculas parece associado à pior sobrevida. De forma interessante, o gene *GUSB* foi associado, de forma inédita, ao GBM tanto pelo aumento de expressão, quanto pela redução da sobrevida. Os mecanismos celulares envolvidos são discutidos e podem envolver tanto aspectos de metabolismo energético, quanto alterações da matriz extracelular, além da autofagia.

Palavras-chave: Lisossomos; Câncer; Transcriptômica; Bioinformática

ABSTRACT

Pan-cancer studies are attempts to unveil important similarities and differences between molecular features of cancer types that could lead to discovery of novel biomarkers, therapeutics and tumoral biology insights. Previous notorious pan-cancer studies have revealed information about the tumor-immune cell interactions, characterized the mutational landscape of major cancer types and the resemblances and distinctions of DNA methylation across cancer types. Given the large sample sizes and strict methodological uniformity regarding the data collection, The Cancer Genome Atlas (TCGA) is the primary source of molecular data for cancer research. Differential expression analysis performed on TCGA data has already revealed countless promising putative diagnostic biomarkers for all sorts of cancer types and also brought insights into it's pathological mechanisms. The Cox regression model integrated with RNA-seq data expands the power of transcriptomic studies and has already enabled the discovery of plenty of putative prognostic gene and gene signature biomarkers. In the past few decades, lysosomes are increasingly beeing associated with human cancers establishment and pathology. In this study we used differential gene expression analysis and Kaplan-Meier associated Cox regression analysis to asses the utilization of lysosomal genes as biomarkers for 29 human cancers with data deposited in the TCGA database. We built a lysosomal gene set gathered from Kyoto Encyclopedia of Genes and Genomes (KEGG) and The Human Lysosomal Gene Databse (hLGDB) and assessed the transcriptional status of these genes in these different cancer types. The only cancer type for which we found strong evidences of roles of lysosomal genes was Glioblastoma Multiforme (GBM). Lysosomal genes related to immune responses were up-regulated in GBM in our study and the up-regulation of lysosomal genes related to macromolecule turnover seems associated with worse prognosis. Interestingly, the *GUSB* gene was found associated, unprecedented, to GBM for both its up-regulation and correlation with worse outcomes. The molecular mechanisms are discussed and may involve energetic metabolism, extracellular matrix remodelling and autophagy.

Keywords: Lysosome; Cancer; Transcriptomics; Bioinformatics

LISTA DE FIGURAS

Figura 1 – Representação moderna de um tumor.. 9
Figura 2 – Esquematização de 10 habilidades tumorais comuns.................................... 10
Figura 3 – Fusão de lisossomos com a membrana plasmática.. 11
Figura 4 – v-ATPase na membrana lisossomal... 12
Figura 5 – Microscopia eletrônica de lisossomos... 13

SUMÁRIO

1 INTRODUÇÃO 9

1.1 JUSTIFICATIVA 14

1.2 OBJETIVOS 14

1.2.1 Objetivo geral 14

1.2.2 Objetivos específicos 15

2 ARTIGO CIENTÍFICO 16

3 CONCLUSÕES E PERSPECTIVAS 60

REFERÊNCIAS 61

ANEXO A – NORMAS DE PUBLICAÇÃO DA REVISTA FRONTIERS IN ONCOLOGY 63

1 INTRODUÇÃO:

1.1 TUMORES E CÂNCERES

Tumores foram observados inicialmente pela humanidade como crescimentos anormais em tecidos e o nome “tumor” vem do latin *tumor*, que significa inchaço ou massa. De forma mais precisa, eles são um conjunto de células descendentes de uma célula comum que sofreu uma ou mais mutações que à possibilitaram crescer e se dividir de forma de forma mais rápida do que células similares em um tecido estável [HANAHAN, D., WEINBERG, R.]. Esse conjunto de células se dividindo e crescendo passa a formar, literalmente, um novo tecido. Imperfeito, porém muitas vezes tão complexo, dinâmico e integrado ao organismo quanto qualquer outro de nossos tecidos [HANAHAN, D; WEINBERG, R]. Esses tecidos costumeiramente incluem células imunes recrutadas e integradas ao ecossistema de células tumorais, células tronco tumorais, células endoteliais que formam capilares e nutrem esse ecossistema e até matrizes celulares próprias secretadas por fibroblastos (Figura 1) [HANAHAN, D; WEINBERG, R].

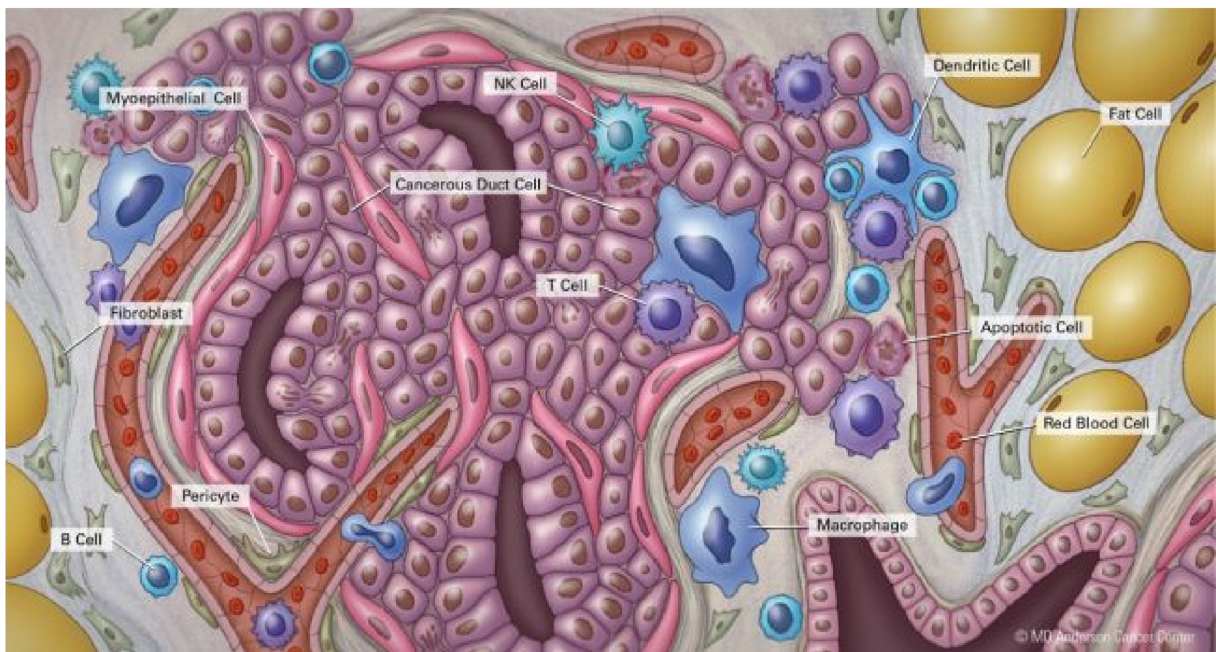


Figura 1: Caracterização moderna de tumores e cânceres como tecido complexos. Além das células tumorais, estão representadas diferentes células imunes e células constituintes de vasos sanguíneos, compostos principalmente por células endoteliais. Crédito: MD Anderson Cancer Center

Tumores que passam a invadir tecidos vizinhos ao seu local de origem são classificados como tumores malignos, os populares cânceres. As características individuais dos cânceres e a forma como eles podem levar a óbito varia de acordo com o órgão e a célula do qual ele surgiu [HANAHAN, D; WEINBERG, R]. Decorrente disso, a forma de tratamento de cada tumor e câncer também varia de acordo com esses parâmetros. Em geral, tumores de baixo grau encontrados em determinados tecidos podem ser tratados de forma mais conservadora e muitas vezes a cirurgia é curativa [MANISH, A.; BAKER LI, F.]. Os

tumores de graus mais altos são cânceres; crescem rapidamente e eventualmente algumas de suas células acabará avançando para dentro de um vaso sanguíneo ou linfático próximo e se espalharem pelo corpo, o que é conhecido como metástase [HANAHAN, D; WEINBERG, R]. A figura 2 mostra 10 características e habilidades identificadas em todos os tipos de cânceres conhecidas, capazes de aumentar sua agressividade. Uma explicação mais aprofundada dessas habilidades se encontra no Apêndice A.

Figura 2. Habilidades que virtualmente todas as células tumorais podem adquirir através de mutações ao longo do seu desenvolvimento e que aumentam sua sobrevivência e/ou agressividade. No centro há uma outra representação moderna de tumores também destacando a heterogeneidade celular e agora mostrando capilares estruturados com as células endoteliais ilustradas. (Adaptado de Hannahan, D. e Weinberg, R., 2011)

Essas células irão agir como sementes, sendo depositadas em alguns tecidos após viajarem pela circulação, podendo encontrar um tecido vulnerável em que ela consiga se adaptar e gerar o crescimento de novos tumores em outros órgãos. Por isso é essencial o acompanhamento a pacientes mesmo após tratamentos bem sucedidos [SUNG, H. et al., 2020]. Segundo a estimativa da Organização Mundial da Saúde (OMS) publicada em 2019, o câncer é a principal ou segunda principal causa de morte evitável em 112 de 183 países analisados, e é a terceira ou quarta maior causa em outros 23 países. O aumento da proporção de mortes causadas por câncer é uma tendência global e reflete avanços no tratamento de outras patologias importantes como doenças cardiovasculares e infecciosas e o aumento da expectativa de vida em praticamente todos os países [SUNG, H. et al., 2020]. O câncer normalmente mata ao afetar direta ou indiretamente a função de órgãos vitais pelo roubo de nutrientes, compressão ou obstrução de cavidades.

1.2 LISOSSOMOS

Lisossomos são uma das mais estudadas organelas de células animais. Eles foram descobertos em 1955 [de DUVE, et al., 1955] mas até hoje são conhecidos por muitos como uma mera organela de descarte e reciclagem de componentes celulares. Proteínas, polissacarídeos, lipídeos e matrizes extracelulares, são recebidos pelos lisossomos, cuja notável diferença para outras organelas é o pH ácido de seu interior, para degradação [BOUHAMDANI, N, et al., 2021]. Com o avanço de técnicas de bioquímica e biologia molecular foi descoberto que a composição proteica [BRAULKE, T, et al., 2009] e lipídica [TAPPEL, A. L., et al., 1965] dos lisossomos também é particular, que eles possuiam um complexo proteico próprio denominado v-ATPase [HARIKUMAR, P., et al., 1983], que é essencialmente uma ATP-sintase inversa que consome ATP para bombear H+ contra o gradiente para dentro do lúmen lisossomal. Mais recentemente, ainda foi descoberto que os lisossomos tem um comportamento dinâmico podendo se movimentar dentro das células [PU, J., et al., 2016], fundir suas membranas com as de outras organelas [LUZIO, J. P., et al., 2007], com a membrana celular para acidificar o meio extracelular e exocitar suas enzimas (Figura 3) [TANCINI, B, et al., 2020] e atuar em dezenas da rotas de sinalização intra e extracelulares [SETTEMBRE, C, et al., 2013].

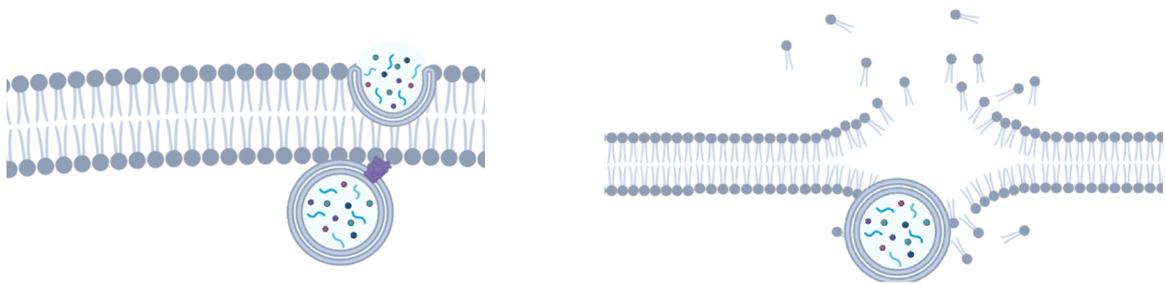


Figura 3: Lisossomos costumeiramente se fundem com a membrana plasmática para (A) exocitar suas moléculas. (B) reparar a membrana rompida cedendo os lipídeos de sua membrana, entre outras funções.

1.3 PROTEOMA LISOSSOMAL

Mesmo tanto tempo após seu descobrimento, a determinação precisa da composição proteica dos lisossomos ainda é desafiadora, já que ela difere, por exemplo, ao longo do ciclo celular [STAHL-MEYER, J., et al., 2022], com o tipo celular [WATTS, C. 2022] e com a localização intracelular [PU, J., et al., 2016]. Existe um conjunto de proteínas cuja localização lisossomal é muito bem determinada, incluindo a bomba de prótons (V-ATPase) (Figura 4) [AKTER, F., et al., 2023], necessária para a acidificação característica da organela, e diversas proteínas envolvidas nas doenças de armazenamento lisossomal (DAL) [AKTER, F., et al., 2023]. Porém mesmo essas proteínas variam de função em diferentes tecidos como pode ser visto na variedade de fenótipos em cada DAL [AKTER, F., et al., 2023]. A presença da V-ATPase é ubíqua e, portanto, a mensuração da presença de suas subnidades é utilizada como padrão para a normalização em estudos proteômicos sobre o lisossomo [AKTER, F., et al., 2023]. Outro problema é que, devido a baixa contribuição dos lisossomos para o proteoma total das células, é necessário que se realize um processo de purificação de lisossomos nas

amostras mas que adiciona artefatos, resultando em um pool de proteínas muito mais diverso do que o que representaria de fato o conjunto de proteínas lisossomais reais [AKTER, F., et al., 2023].

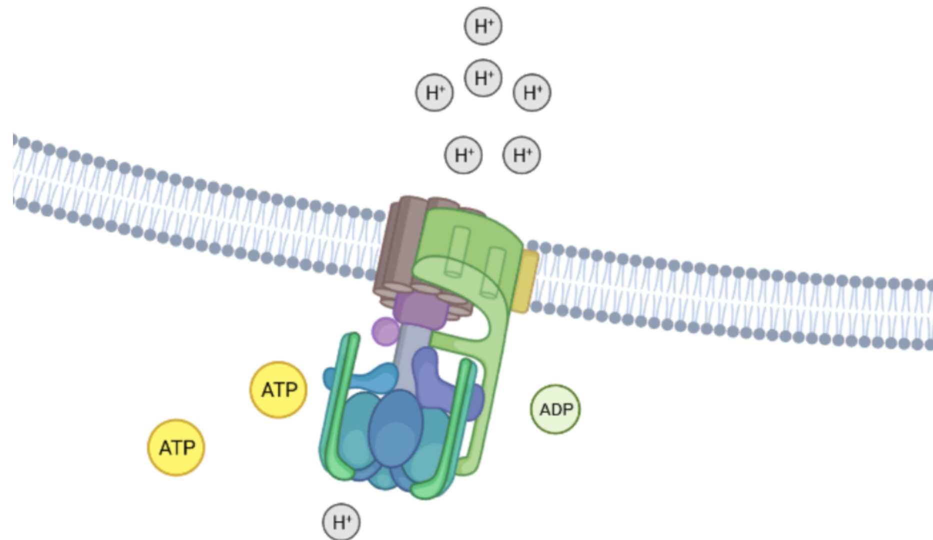


Figura 4: O complexo v-ATPase tem basicamente a mesma estrutura de uma ATP-sintase porém ele atua de forma inversa: hidrolisa ATP para bombear prótons para dentro do lisossomo e abaixar o pH.

1.4 GENES LISOSSOMAIS

À exemplo da descrição do proteoma, e também por depender dele, o genoma lisossomal ainda é pouco descrito. Por um lado, bases de dados mais curados como a rota lisossomal na *Kyoto Encyclopedia of Genes and Genomes* (KEGG) [KANEHISA, M., et al., 2000] e o hLGDB [BROZZI, A., et al., 2013] contém um número de genes que não representa a totalidade de proteínas lisossomais já estabelecida. O grupo das hidrolases lisossomais é especialmente bem coberto nessas bases de dados, porém a presença de transportadores transmembrana de resíduos de carboidratos e lipídeos é pouco documentada. Por outro lado, muitas das assinaturas gênicas relacionadas aos lisossomos depositadas no T*he Molecular Signatures Database* (mSigDB) [SUBRAMANIAN, A, et al., 2005], apesar de incluirem genes faltantes nas outras bases, carecem de curatela e incluem proteínas de background detectadas em estudos proteômicos ou proteínas raramente associadas aos lisossomos e ainda sem evidências de função lisossomal. Assim, é prudente que se revise os genes obtidos dessas bases de dados para formar um conjunto gênico novo que aumente a confiança nos resultados finais dos experimentos.

1.5 LISOSSOMOS NO CÂNCER

Talvez a função lisossomal mais óbvia a ser estudada no contexto do câncer seja a degradação de macromoléculas, já que muitas enzimas lisossomais degradam glicosaminoglicanos ou são proteases [BOUHAMDANI, N, et al., 2021], sendo assim capazes de degradar e remodelar os constituintes majoritários da matriz extracelular (MEC) e acelerar o processo de invasão de metástase de tumores. Além disso, ao remodelar a MEC, elas interferem na distribuição de moléculas de sinalização extracelular, interferindo nessa comunicação. CTSL, CTSB e CTSX/Z [RUDZINSKA, M., et al., 2019] são exemplos bem estabelecidos desses casos. Porém, com tantas novas funções lisossomais descritas, foi possível desvendar novos mecanismos pelos quais os lisossomos influenciam no estabelecimento e na malignidade de tumores. A segunda função lisossomal com importância no câncer mais bem descrita talvez seja a autofagia, já que em condições normais o complexo proteico mTORC1 se localiza justamente na membrana lisossomal [RABANAL-RUIZ, Y, et al., 2018]. Estando ligado na membrana lisossomal, as subunidades do complexo tanto interagem com outras proteínas lisossomais quanto desfosforilam e inativam proteínas de cascatas de sinalização intracelulares relacionadas à ativação da autofagia [RABANAL-RUIZ, Y, et al., 2018]. No contexto tumoral, mTORC1 tem importância terapêutica já que o controle do complexo e de suas proteínas lisossomais associadas pode ser utilizado para induzir a autofagia e cessar o desenvolvimento e a replicação de células tumorais [RABANAL-RUIZ, Y, et al., 2018]. De modo inverso, é possível que células tumorais se beneficiem desse sistema de sobrevivência celular, já que células refratárias podem sobreviver a quimioterapias graças a autofagia e causar a reincidência do tumor após o término do tratamento [CHANG, H, et al., 2020]. A figura 5 mostra duas imagens de microscopia eletrônica de lisossomos.

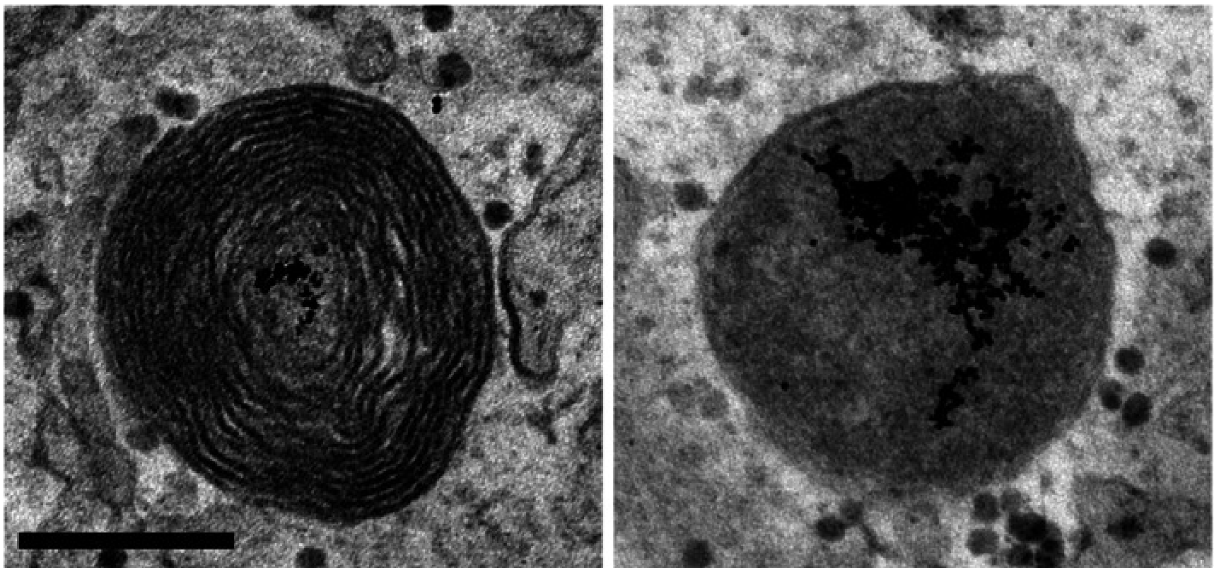


Figura 5: Microscopia eletrônica de lisossomos contendo partículas de ouro endocitadas. Na imagem da esquerda é possível notar várias lamelas de membrana no interior do lisossomo, importantes no reparo de membrana celular por fusão lisossomal [BARRAL, D. C., et al, 2022].

1.6 ESTUDOS TRANSCRIPTÔMICOS

Partindo da base desses estudos proteômicos e genômicos, a completa elucidação da composição, dinâmica e função de componentes celulares, incluindo os lisossomos, passa também pelos estudos transcriptômicos. Nestes estudos se mede o nível de atividade de genes através da detecção dos seus respectivos mRNAs. As moléculas de RNA criadas ao molde de partes da molécula de DNA que contém o gene e que saem do núcleo são utilizadas para organizar a síntese de proteínas junto com os ribossomos. A partir desses dados sobre os níveis de produção de mRNA é possível obter caracterizações dos níveis de expressão de todos os genes de um tecido ou colônia celular [WANG, Z., et al., 2009], podendo dar assim uma noção de quais genes são importantes tanto na fisiologia quanto na patologia. Para tumores, por exemplo, é comum a utilização de técnicas de análise de expressão diferencial [ANDERS, S., et al., 2010] para identificar genes mais ou menos expressos em tumores em comparação com os seus tecidos de origem, revelando genes desregulados em tumores e que podem, portanto, estar envolvidos no seu desenvolvimento. Em análises de sobrevida, como a análise de regressão de Cox [COX, D. R., 1972], é possível dividir pacientes com determinada doença em dois grupos, de acordo com os níveis de expressão de um gene (alto ou baixo). O tempo de sobrevida dos pacientes dos dois grupos após o diagnóstico da doença é comparado e, assim, pode-se identificar genes cuja alteração na expressão esteja relacionada com casos mais graves da doença.

1.7 JUSTIFICATIVA

Lisossomos são organelas essenciais para a viabilidade de células animais e a expressão de seus genes tem sido implicada na patologia de diferentes tumores humanos em estudos individuais nos últimos anos. Porém, ainda não existe um estudo que unifique e compare dados de expressão lisossomal entre diferentes tipos de tumores. Tal estudo poderia confirmar achados anteriores quanto ao envolvimento dos lisossomos em cânceres e revelar semelhanças e diferenças entre os mecanismos patogênicos de lisossomos nesses tumores, além de revelar novos biomarcadores lisossomais.

1.8 OBJETIVOS

1.8.1 Objetivo geral

O objetivo desse trabalho é analisar a expressão diferencial de genes lisossomais em tumores humanos e sua correlação com prognóstico.

1.8.2 Objetivos específicos

- Determinar um conjunto de genes que codifiquem proteínas residentes de lisossomos.
- Realizar análise de expressão diferencial para cada um dos genes e dos tipos de câncer comparando o tecido tumoral com o tecido normal adjacente.
- Realizar análise de regressão de Cox para cada um dos genes e dos tipos de câncer a fim de determinar aqueles em que possivelmente haja um maior envolvimento de lisossomos em sua malignidade.

2 ARTIGO CIENTÍFICO

O artigo intitulado “Pan-Cancer Transcriptomic Analysis Suggests Lysosomal Functional Importance and Reveals Novel Biomarker for Glioblastoma Multiforme” foi editado de acordo com as normas de publicação do periódico *Frontiers in Oncology* (Anexo A)

Pan-Cancer Trascriptomic Analysis Suggests Lysosomal Importance and Reveals Novel Biomarker for Glioblastoma Multiforme

Thiago Steindorff[1,2], Ursula Matte[1,2,3*]

[1]Gene Therapy Center, Hospital de Clínicas de Porto Alegre, Porto Alegre , Brazil

[2]Biomedical Sciences School, Institute of Health Sciences, UFRGS, Porto Alegre , Brazil

[3]Genetics and Molecular Biology Graduate Program, UFRGS, Porto Alegre, Brazil

*** Correspondence:**
Ursula Matte
umatte@hcpa.edu.br

Keywords: lysosome, cancer, transcriptomics, bioinformatics

Abstract

With the increase in life expectancy, cancer has emerged as one of the leading causes of fatality worldwide in recent decades. Pan-cancer studies aim to unveil significant similarities and differences among molecular features of distinct cancer types, which could potentially lead to the discovery of novel biomarkers, therapeutics, and insights into tumoral biology. In this study, we sought to asses the involvement of lysosome coding genes in human canceres. We conducted a literature review to filter lysosomal genes and utilized differential gene expression analysis (DEA) and Kaplan-Meier-associated Cox regression survival analysis to analize 129 lysosomal genes in 29 human cancers, using deposited data from The Cancer Genome Atlas (TCGA) database. Our differential expression analysis results unveiled large differences between the differential expression of lysosomal genes between human tumours, and we found evidences of its up-regulation in glioblastoma multiforme in comparison with adjacent normal tissue. This was the sole cancer in which lysosomal genes exhibited greater differential expression in TCGA data compared to random gene sets. The Cox regression results indicated that such overexpression is strongly associated with worse prognosis in gliblastomas, and the comparison with random gene sets suggests that this influence is specific to this cancer type. We identified *GUSB,* "beta-glucuronidase", as a novel differentially expressed and survival related gene in glioblastoma multiforme, hence making it a promising biomarker for this cancer type.

1. Introduction

Pan-cancer studies are attempts to unveil important similarities and differences between molecular features of cancer types that could lead to discovery of novel biomarkers, therapeutics and tumoral biology insights, and ever since 2012, the publishing of these studies has grown yearly [1]. Previous notorious pan-cancer studies have revealed information about the tumor-immune cell interactions [2], characterized the mutational landscape of major cancer types [3] and the resemblances and distinctions of DNA methylation across cancer types [4]. Given the large sample sizes and strict methodological uniformity regarding the data collection, The Cancer Genome Atlas (TCGA) [5] is the primary source of molecular

data for cancer research. Differential expression analysis [6] performed on TCGA data has already revealed countless promising putative diagnostic biomarkers for all sorts of cancer types and also brought insights into its pathological mechanisms [7, 8, 9]. The Cox regression model [10] integrated with RNA-seq data expands the power of transcriptomic studies and has already enabled the discovery of plenty of putative prognostic gene and gene signature biomarkers [11, 12, 13]. In the past few decades, lysosomes are increasingly being associated with human cancers pathology mainly due to its role in autophagy regulation and macromolecules turnover [14, 15]. In this study we used differential gene expression analysis and Kaplan-Meier associated Cox regression analysis to assess the extension of the involvement of human lysosome coding genes in the development and aggressiveness of 29 human cancers with data deposited in the TCGA database. We built a lysosomal gene set gathered from Kyoto Encyclopedia of Genes and Genomes (KEGG) [16] and The Human Lysosomal Gene Databse (hLGDB) [17] and assessed the transcriptional status of these genes in these different cancer types.

2. Materials and Methods

2.1 Lysosome related genes

Our initial lysosomal gene set was gathered from two databases: the KEGG lysosomal pathway [16] and the Lysosomal Gene Database [17]. The KEGG pathway has 132 genes while LGDB has 450 genes, of which 82 genes intersect. Therefore, the joined gene set was made out of 500 genes.

2.2 Lysosomal localized genes screening

We conducted a literature review using the Pubmed database to investigate the actual presence of these genes' products in lysosomes. For each lysosomal gene, we searched using two schemes: "Gene symbol + Lysosome" and "Protein name (according to gene cards) [18] + Lysosome". Articles were reviewed in search of evidence of lysosomal localization, and one such evidence was considered enough. We've considered as "lysosomal" the proteins primarily found in fully formed human lysosomes. The same review was used to group genes based on their main functions.

2.3 Protein-protein interaction (PPI) and Functional Enrichment

The PPI of lysosomal proteins was conducted using String-db [19]. All 129 proteins were screened for interactions observed in *Homo sapiens*. The MCL clustering was performed with inflation parameters of 1.5, which is the lower possible parameter. We also used the String-db with standard configuration to perform functional enrichments.

2.4 Differential expression

There are 24 TCGA projects with available adjacent normal tissue transcriptomic data, that were used for differential expression analysis, which was performed using the "TCGAanalyze_DEA" function from "TCGA_biolinks" package [20] in R programming language with false discovery rate (FDR) of 0.01 and log2FC cut-off = 2. Outliers preprocessing and count data normalization was performed using the "TCGAanalyze_preprocessing" and the "TCGAanalyze_Normalization" functions,

respectively. The lower expressing mRNA quartile was removed using the "TCGAanalyze_Filtering" function. The number of differentially expressed genes in the curated dataset was compared to the mean number of differentially expressed genes in 10,000 iterations of 129 genes from a random set of 16,000 genes.

2.5 Survival analysis

We used the GEPIA2 portal [21] for the Cox regression analysis and the Kaplan-Meier plots generation for every gene in every dataset/cancer type with n >= 100. The chosen group cutoff was quartile for tumours with n > 300 samples and tercile for tumors with n <= 300. The significance p-Value cutoff used was 0.05. The data for lysosomal genes was compared to the results from a random set of genes generated as described above.

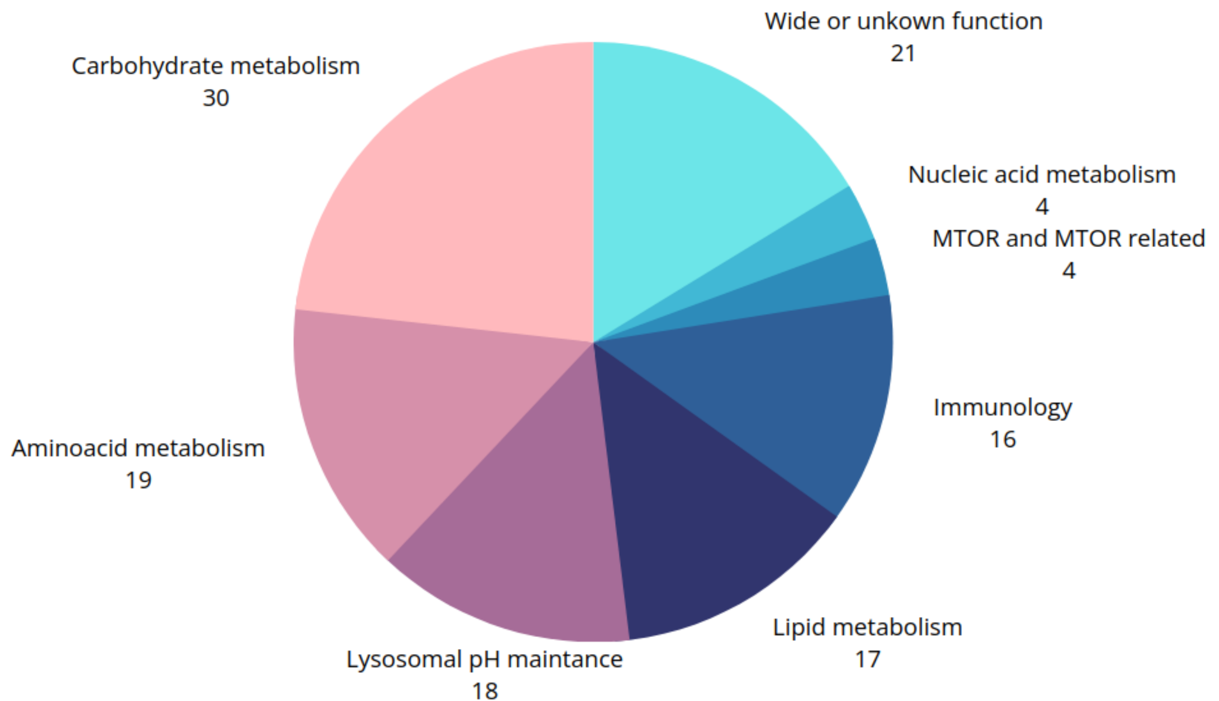
3. Results

3.1 Determination of lysosome resident proteins and their functions:

In order to evaluate the putative role of lysosomes on the different types of tumors, we first established a list of genes whose proteins are primarily found within human lysosomes. According to our review, out of the 500 genes from the initial set (KEGG lysosomal pathway [16] and the Lysosomal Gene Database [17]), only 129 genes codify proteins primarily found on human lysosomes. From now on we will refer to these as lysosomal genes.

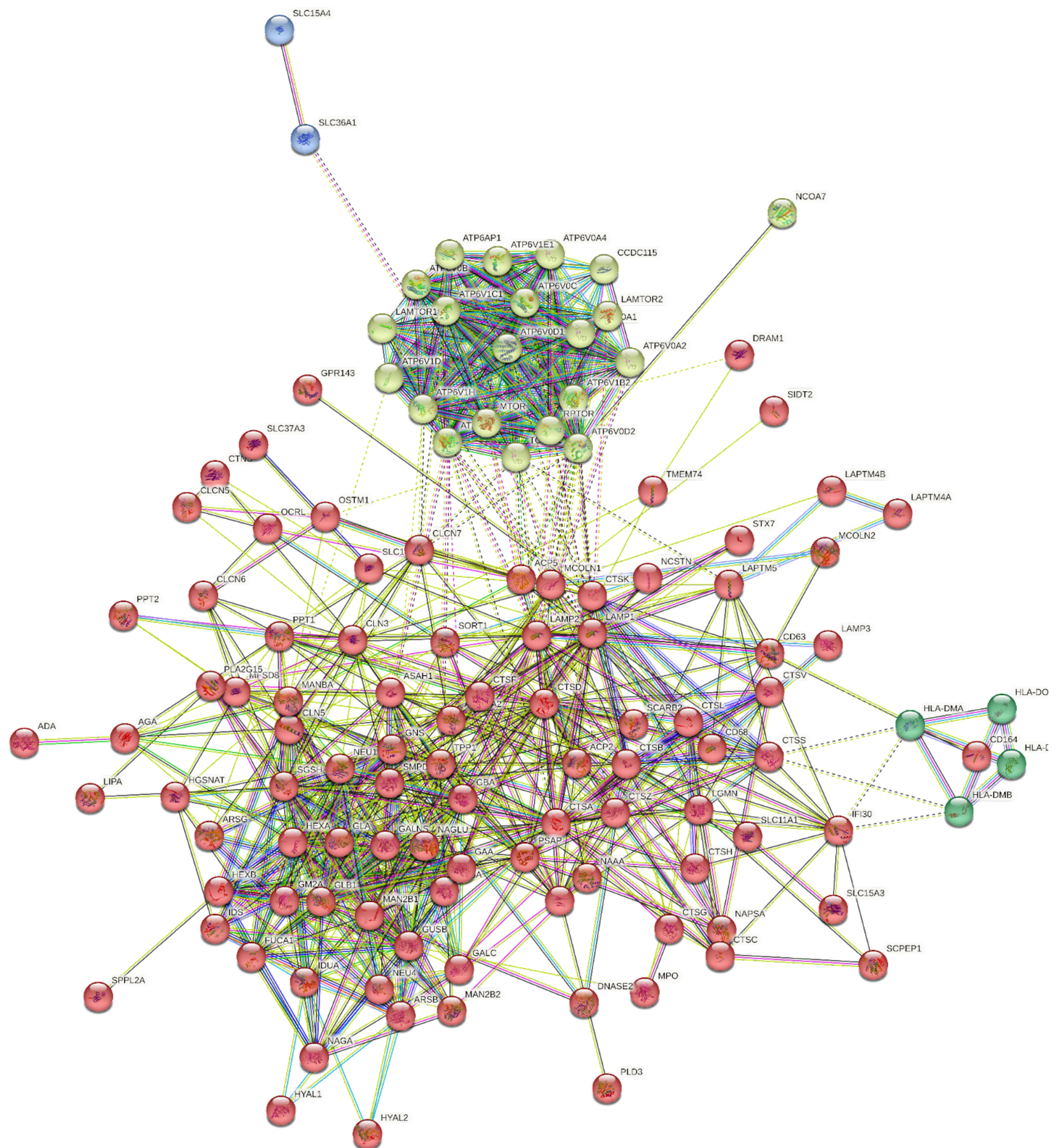
Following their confirmation as primarily lysosomal genes, their proteins were classified into eight groups according to their reported functions as by PubMed review of the literature [Supplementary table 1]. Lysosomal proteins directly involved in the catabolism of macromolecules were further divided into four groups based on the biomolecular nature of their targets (namely: Aminoacid, Carbohydrate, Lipid and Nucleic acid metabolism). Lysosomal proteins required for the function and/or specifically expressed in immune cells were clustered in the Immunology group. v-ATPase proton pump components, modulators and other proton transporters were clustered on the Lysosomal pH Maintenance group.The *MTOR* gene and other primarily lysosomal proteins related to mTORC's signalling were clustered in the seventh group. Lastly, proteins that are either globally involved in lysosomal function, or for which lysosomal function remains elusive, were clustered on Wide or unknown lysosomal function (Figure 1).

Figure 1: Depiction of functional groups of lysosomal proteins and their relative sizes. The number of proteins is shown below their respective group.



We then used the String-db to analyze known relationships between the proteins in this lysosomal gene set. The PPI enrichment p-value was lower than 10^{-9}, which was expected since this is a refined dataset derived from two already curated lysosomal gene databases. The MCL clustering with a more restricted inflation parameter resulted in 4 protein clusters (figure 2). Cluster 1 (n = 95) is composed of canonical lysosomal proteins such as lysosomal hydrolases and lysosomal associated membrane proteins (LAMPs). Cluster 2 (n = 21) is composed of v-ATPase subunits and mTOR and its related lysosomal proteins. Cluster 3 (n = 4) is composed of the subunits of HLA-DM and HLA-DO, two highly related dimeric proteins involved in antigen presentation. HLA-DO is highly unstable in the absence of HLA-DM and it's required for HLA-DM's function [22]. This might be the reason for their separate clustering. Cluster 4 is a pair of solute carriers recently described as lysosomal, which might explain why they were not included in Cluster 1.

Figure 2: Clustering of lysosomal proteins on the PPI network. Cluster 1 is shown in red, Cluster 2 in light green, Cluster 3 in dark green and Cluster 4 in blue.



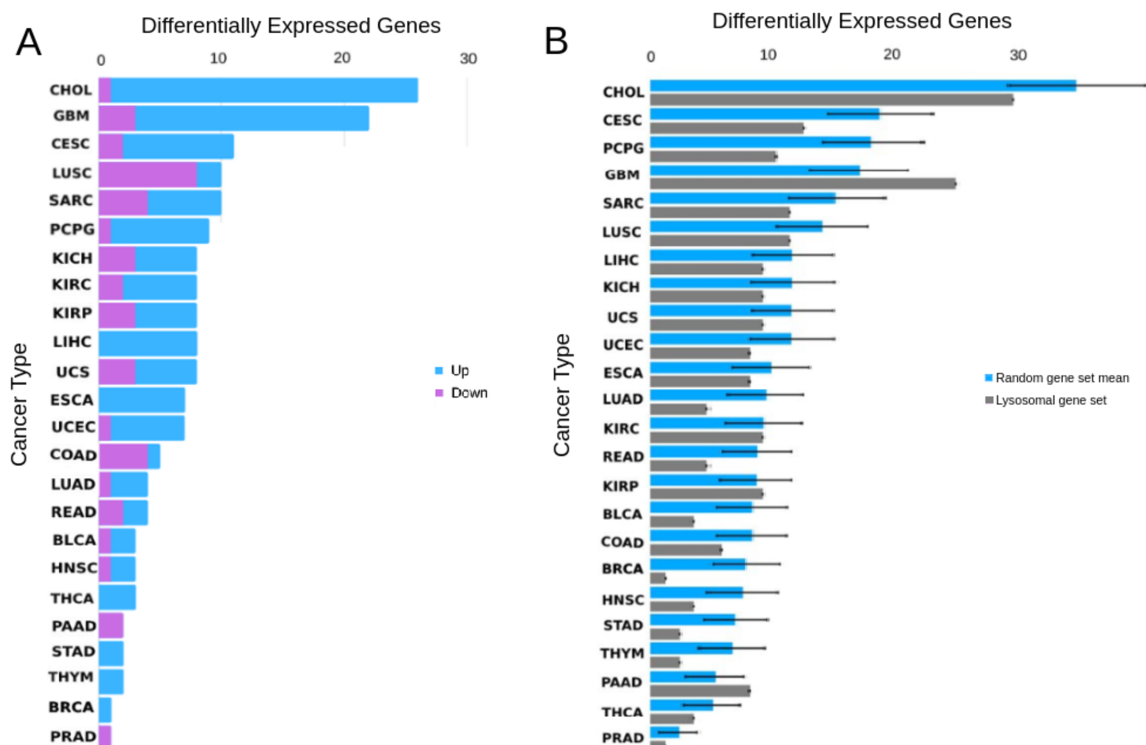
3.2 Lysosomal genes differential expression on TCGA

The number of differentially expressed lysosomal genes (DELG) was analyzed per TCGA project using a log2FC = 2 cutoff (figure 3A). The complete list of DELG per project is found in Supplementary table 2. Two projects stood out as the ones with the highest number of differentially expressed lysosomal genes: cholangiocarcinoma (CHOL) and glioblastoma multiforme (GBM). Both of them had more up than down-regulated lysosomal genes.

To test if this difference in expression is intrinsic to lysosomal genes in certain tumors or a function of sample variability, we generated 10,000 random sets of 129 genes and assessed their expression in all tumor types (Figure 3B). In most projects, including CHOL, the lysosomal gene set DEGs followed the trend of the random gene sets. On the other hand, the lysosomal gene set had more DEGs in GBM than what would be expected for random gene sets, suggesting that there might be a global increase in lysosomal genes' expression associated with the establishment of GBM.

Figure 3: **(A)** - Number of differentially expressed lysosomal genes per The Cancer Genome Atlas (TCGA) project. Light blue represents the number of up-regulated genes, pink are the down-regulated genes. **(B)** - Comparison between the number of differentially expressed lysosomal genes (gray bars) and the mean number of random differentially expressed genes (blue bars) per TCGA project.

BLCA (Bladder Urothelial Carcinoma), BRCA (Breast invasive carcinoma), CESC (Cervical squamous cell carcinoma and endocervical adenocarcinoma), CHOL (Cholangiocarcinoma), COAD (Colon adenocarcinoma), ESCA (Esophageal carcinoma), HNSC (Head and Neck squamous cell carcinoma), GBM (Glioblastoma), 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), PAAD (Pancreatic adenocarcinoma), PRAD (Prostate adenocarcinoma), PCPG (Pheochromocytoma and Paraganglioma), READ (Rectum adenocarcinoma), SARC (Sarcoma), STAD (Stomach adenocarcinoma), THCA (Thyroid carcinoma), THYM (Thymoma), UCEC (Uterine Corpus Endometrial Carcinoma), UCS (Uterine Carcinosarcoma)



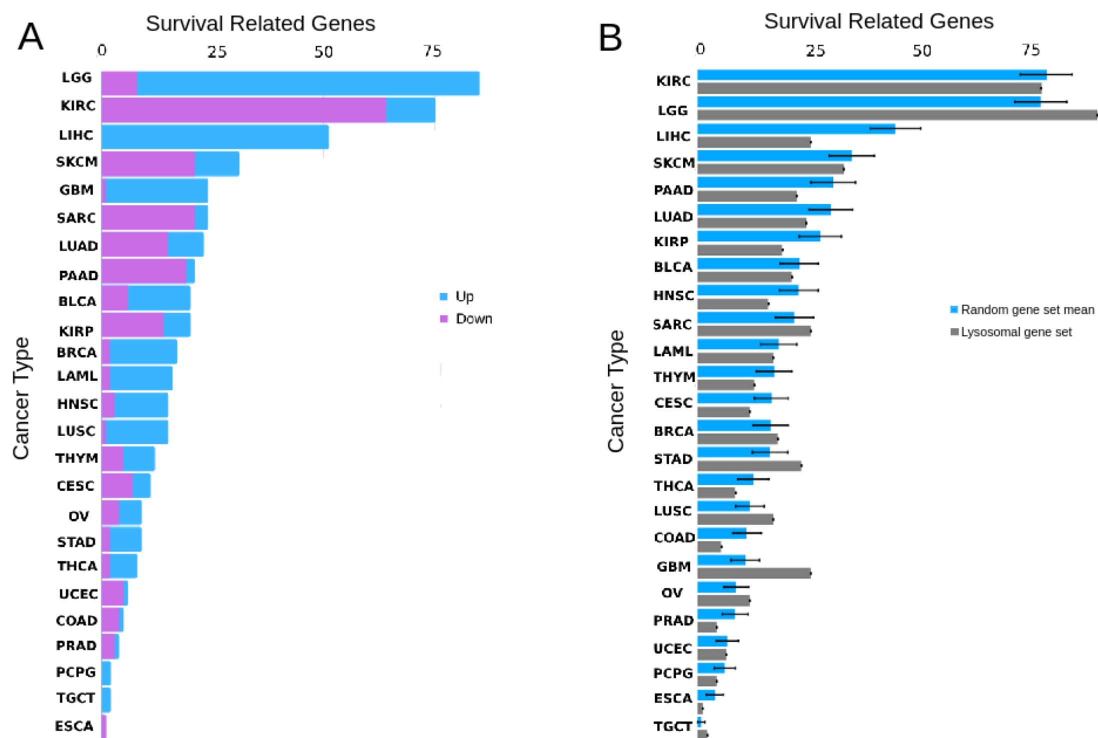
3.3 Effect of lysosomal genes in cancer survival

In order to assess the putative effects of lysosomal genes in cancer survival, we performed cox-regression and Kaplan-Meier survival analysis using the GEPIA 2 online tool for the 24 TCGA projects with sample size >= 100. The number of lysosomal genes significantly correlated with negative outcomes per TCGA project is shown in figure 4A while the complete list of cox-regression p-Values and hazard ratios for each gene and project is found on Supplementary table 3.

Again, when compared to random gene sets (figure 4B), most lysosomal survival related genes were similar in proportion to the random gene sets, including the two top cancres: KIRC and LGG. Importantly, this was not the case for GBM, as the total lysosomal survival related genes more than double the random gene set mean (24 and 9.73 genes, respectively). Together with the DEG results, this indicates that lysosomes could be involved as a whole in the aggressiveness of GBM.

Figure 4: **(A)** - Number of lysosomal genes significantly correlated with differential outcomes per TCGA project. Light blue represents genes correlated with bad prognosis when high expressed. Purple represents when low expressed. ; **(B)** - Comparison between the number of survival related lysosomal genes (gray bars) and the mean number of random survival related genes (blue bars) per TCGA project.

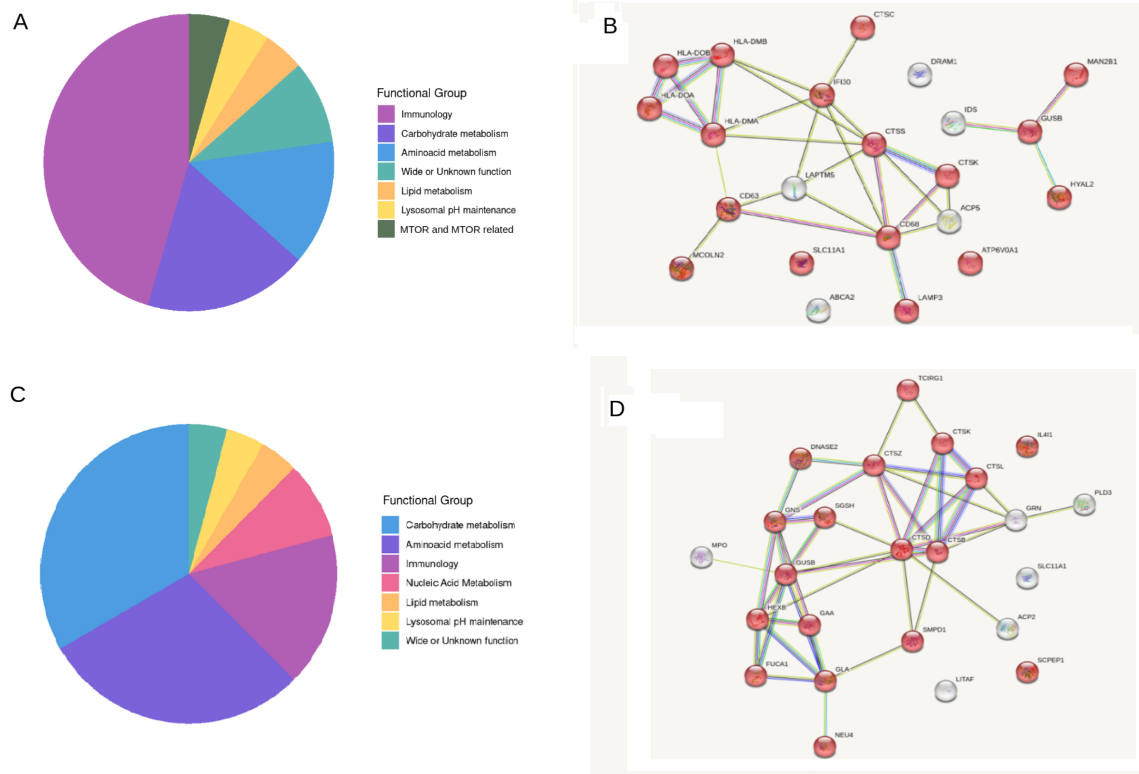
BLCA (Bladder Urothelial Carcinoma), BRCA (Breast invasive carcinoma), CESC (Cervical squamous cell carcinoma and endocervical adenocarcinoma), COAD (Colon adenocarcinoma), ESCA (Esophageal carcinoma), GBM (Glioblastoma), HNSC (Head and Neck squamous cell carcinoma), KIRC (Kidney renal clear cell carcinoma), KIRP (Kidney renal papillary cell carcinoma), LAML (Acute Myeloid Leukemia), LGG (Low Grade Glioma), LIHC (Liver hepatocellular carcinoma), LUAD (Lung adenocarcinoma), LUSC (Lung squamous cell carcinoma), OV (Ovarian serous cystadenocarcinoma PAAD (Pancreatic adenocarcinoma), PCPG (Pheochromocytoma and Paraganglioma), PRAD (Prostate adenocarcinoma), SARC (Sarcoma), SKCM (Skin Cutaneous Melanoma), STAD (Stomach adenocarcinoma), TCGT (Testicular Germ Cell Tumors), THCA (Thyroid carcinoma), THYM (Thymoma), UCEC (Uterine Corpus Endometrial Carcinoma)



3.4 DELG and SRLG functional enrichment

According to our functional classification, 10 out of the 22 lysosomal DEGs are immune-related lysosomal genes (~ 45%) (Figure 5A). Consistently, the functional enrichment (supplementary table 4) of the DELG shows that most biological processes are related to immune function. The figure 5B depicts the DELG that functional enrichment claims to be involved in immune system processes (GO:0002376). Also, according to our functional classification, 75% of the survival related genes (SRGs) (18/24) are lysosomal hydrolases (Figure 5C). Consequently, the functional enrichment shows many lysosomal and catabolic pathways as enriched in this gene set. (supplementary table 5). Figure 5D shows the network for SRLG highlighting the proteins in the organic substance catabolic process (GO:1901575).

Figure 5: **(A)** - Proportion of functional groups of lysosomal DEGs in glioblastoma. **(B)** - PPI network of lysosomal DEG in glioblastoma highlighting those related to immune system function. **(C)** - Proportion of functional groups of survival related genes (SRG) in glioblastoma. **(D)** - PPI network of SRG in glioblastoma highlighting those that codify lysosomal hydrolase.



4. Discussion

The relationship between lysosomes and glioblastoma has already been assessed, given the extent of autophagy's implications in this cancer type [71] and the findings of GBM's vulnerability to lysosomal disruption [72]. As far as we know, out of the 22 DEG found in this study, 8 had not been implicated in glioblastoma's pathology yet. Only 3 DEG found here were already characterized in GBM: *CTSK* [24, 25], *HLA-DMA* [29] and *MAN2B1* [32] and are, therefore, promising diagnostic biomarkers. Others, like *CD63* [52], *IFI30* [30, 34] and *SLC11A1* [34, 35] have been implicated as prognostic factors for GBM but not yet as DEG. Some other genes not yet found as DEG but already studied as mechanistically involved in GBM include *ABCA2* [26], *ACP5* [23], *CTSS* [27], *DRAM1* [28], *LAPTM5* [31] and *MCOLN2* [33]. Nonetheless, *ATP6V0A1*, *CTSC, HLA-DMB, HLA-DOA, HLA-DOB, HYAL2, IDS* and *LAMP3*, are novel DEG in GBM and have not been implied in its pathology in any way. We suggest these genes could serve as prognostic biomarkers and may provide insights into the lysosomal function in GBM's progression.

Regarding the 24 survival related genes, 10 were not implicated in glioblastomas in any way previously and are novel potential lysosomal prognostic markers for this cancer: *ACP2, DNASE2, FUCA1, GLA, GNS, GUSB, MPO, PLD3, SCPEP1* and *SGSH*. Others, like *CTSB* [36, 37], *CTSD* [38, 39], *CTSK* [24], *CTSL* [40], *GRN* [44], *HEXB* [45], *IL4I1* [46], *LITAF* [47, 48], *NEU4* [49], *SLC11A1* [34, 35] and *TCIRG1* [51] were already reported as potential prognosis and/or therapy resistance markers. Curiously, *GAA* and *SMPD1* protein expressions were found as a protective factor [43, 50] while our analysis shows the opposite

at the mRNA level. *CTSZ* [41, 42] was already found as both a diagnosis and prognostic biomarker for GBM, although we did not find it as a DEG.

Altogether, three genes stood up as being both differentially expressed and survival related in GBM: *CTSK, GUSB* and *SLC11A1*. As shown in table 2, they all are both up-regulated and markers of bad prognosis when highly expressed. *CTSK* codifies the cysteine proteinase cathepsin K, a protease widely known for its expression in osteoclasts and its role in bone resorption [53]. It seems to contribute to tumor's malignancy through extracellular matrix degradation and the disruption of several signaling pathways [54] . It is believed that its expression inhibits the oxaliplatin-induced apoptosis in GBM [55]. Its overexpression was formerly validated in TCGA independent GBM cohorts [24]. *SLC11A1* is a well known divalent metal solute carrier that resides in macrophagic lysosomes [56]. It shows up as a potential immune-related biomarker in several cancer types [57, 58, 59, 60] and is possibly implied in ferroptosis activation [34] and immune check-point blockade therapy [35] in GBM.

Differently from *CTSK* and *SLC11A1*, the *GUSB* gene, which codifies the lysosomal hydrolase *GUSB* was, until this day, not yet reported as differentially expressed or correlated with worse GBM's prognosis, and neither had its function studied in this cancer type. Recently it was implied in low grade gliomas (LGG) by Jiacheng Xu, et al. [61], that included it in a glucose metabolism-related gene-based model. The regular function of *GUSB* is the beta-glucuronidase activity in the glycosaminoglycans (GAGs) heparan, dermatan and chondroitin sulfate catabolic pathways and mutations in this gene lead to Mucopolysaccharidosis type VII [62]. Given its function, it is possible that it acts in the glucose metabolism in GBM as indicated for LGG [61], increasing the carbohydrate availability through the turnover of GAGs to support the higher metabolic needs of rapidly growing and proliferating tumor cells [64]. It can also be important to GBM's progression by degrading the extracellular matrix and allowing for the tumor's expansion, migration and metastasis [65] or even modulating the distribution of extracellular signaling molecules [66]. Interestingly, it can also affect tumor cell responses to therapies, since its high expression was found to be associated with resistance to anti-PD1 therapy in hepatocellular carcinoma (HCC) and its inhibition sensitized the HCC cells to the treatment [63].

When analyzing the functional groups, we found that ~45% of DELG codify immune-related lysosomal proteins, such as the macrophage marker CD63, all of the HLA-DM-HLA-DO complex subunits and the aforementioned solute carrier SLC11A1. Many of the lysosomal genes found in our study as DELGs or SRLGs and that were already related to GBM were discovered in immunology-centered studies [23, 30, 32, 35]. As all of the DELG in GBM were up-regulated and since there are so many evidences of their relationships with the immune function, maybe the lysosomal genes up-regulated in GBM reflect it's abundant immune infiltration [67], and their expression could be used to characterize it. Regarding the SRLG, 75% of the genes (18/24) codify lysosomal hydrolases, enzymes necessary for the turnover of macromolecules [68] and responsible for the main function of lysosomes. This could suggest that the lysosomal composition is important for glioblastoma's malignancy, either for maintaining the regular lysosomal function, enabling proper autophagy [69] or even for remodeling the extracellular matrix through the exocytosis of it's luminal contents [70].

Taken together our results reinforce the already known importance of individual lysosomal proteins and genes in glioblastoma, validate the expression of *CTSK* and *SLC11A1* as potential biomarkers for this cancer type and reveals *GUSB* as a novel promising lysosomal

diagnostic and prognosis biomarker for GBM. We also found that lysosomal gene's expression could be markers of cancer related immune function.

Supplementary table 1 - Functional classification of lysosomal genes

Main lysosomal function	Genes
Aminoacid metabolism	*CTNS, CTSA, CTSB, CTSC, CTSD, CTSF, CTSG, CTSH, CTSK, CTSL, CTSS, CTSV, CTSZ, LGMN, LITAF, SCPEP1, SLC15A3, SLC15A4 e SLC36A1*
Carbohydrate metabolism	*AGA, ARSB, ARSG, FUCA1, GAA, GALNS, GLA, GLB1, GNS, GUSB, HEXA, HEXB, HGSNAT, HYAL1, HYAL2, IDS, IDUA, MAN2B1, MAN2B2, MANBA, NAGA, NAGLU, NEU1, NEU4, SGSH, SLC17A5 e SLC37A3*
Immunology	*CD63, CD68, GRN, HLA-DMA, HLA-DMB, HLA-DOA, HLA-DOB, IFI30, IL4I1, LAPTM5, MCOLN2, MPO, OSTM1, SLC11A1, SPPL2A e STX7*
Lipid metabolism	*ABCA2, ABCA3, ABCA5, ABHD5, ARSA, ASAH1, GALC, GBA, GM2A, LIPA, MFSD8, NAAA, PLA2G15, PPT1, PPT2, PSAP, SCARB2, SMPD1, SORT1 e TPP1*
Lysosomal pH maintance	*ATP6AP1, ATP6V0A1, ATP6V0A2, ATP6V0A4, ATP6V0B, ATP6V0C, ATP6V0D1, ATP6V0D2, ATP6V1A, ATP6V1B2, ATP6V1C1, ATP6V1D, ATP6V1E1, ATP6V1H, CLCN5, CLCN6, CLCN7 e TCIRG1*
MTOR and related proteins	*LAMTOR1, LAMTOR2, MTOR e RPTOR*
Nucleic acid metabolism	*ADA, DNASE2, PLD3 e SIDT2*
Wide or unknown function	*ACP2, ACP5, ATP13A2, CCDC115, CD164, CLN3, CLN5, DRAM1, DRAM2, GPR143, LAMP1, LAMP2, LAMP3, LAPTM4A, LAPTM4B, MCOLN1, NAPSA, NCOA7, NCSTN, OCRL* e *TMEM74*

Supplementary table 2 - Differentially expressed lysosomal genes per TCGA project

2.1 - BLCA

Gene	Log2FC
CTSL2	2.82
ATP6V0A4	2.75
CTSG	-2.76

2.2 - BRCA

Gene	Log2FC
IL4I1	2.49

2.3 - CESC

Gene	Log2FC
CTSL2	5.22
LAMP3	5.19
NAPSA	4.74
IFI30	3.03
ATP6V0A4	2.86
NEU4	2.56
ADA	2.12
CD68	2.08
GM2A	2.06
CTSF	-2.44
CTSK	-2.59

2.4 - CHOL

Gene	Log2FC
CTSL2	6.39
MCOLN3	5.86
ABCA3	4.62
LAMP3	4.29
NAPSA	3.92
HLA-DOB	3.85
CTSC	3.67
DRAM1	3.49
MCOLN2	3.3
ATP13A2	3.12
IL4I1	3.1
IDUA	2.97
CTSH	2.86
CTSK	2.83
SLC36A1	2.81
LAPTM4B	2.64
SLC37A3	2.36
SORT1	2.32
GLA	2.25
IDS	2.2
HLA-DMB	2.18
PPT1	2.17
SLC11A1	2.11
TCIRG1	2.08
ATP6V1E1	2.04
NEU4	-3.38

2.5 - COAD

Gene	Log2FC
SLC11A1	2.61
NAAA	-2.12
ATP6V0D2	-2.12
NEU4	-2.63
CTSG	-3.19

2.6 - ESCA

Gene	Log2FC
LAMP3	4.38
NEU4	3.85
ATP6V0D2	3.25
IL4I1	2.5
IFI30	2.33
CTSL2	2.21
CTSC	2.08

2.7 - GBM

Gene	Log2FC
IFI30	4.03
MCOLN2	3.68
LAMP3	3.29
HLA-DMB	3.22
SLC11A1	3.12
ACP5	2.98
CTSC	2.87
HLA-DMA	2.84
DRAM1	2.83
CTSK	2.79
CTSS	2.75
HLA-DOB	2.62
CD68	2.59
HYAL2	2.47
HLA-DOA	2.45
LAPTM5	2.39
GUSB	2.35
CD63	2.28
MAN2B1	2.08
ATP6V0A1	-2.04
IDS	-2.44
ABCA2	-2.95

2.8 - HNSC

Gene	Log2FC
CTSL2	2.09
SLC15A3	2.05
ATP6V0A4	-3.8

2.9 - KICH

Gene	Log2FC
ATP6V0D2	3.24
ATP6V0A4	2.39
LAMP3	2.25
GPR143	2.08
LGMN	2.07
CTSL2	-2.98
NEU4	-5.0

2.10 - KIRC

Gene	Log2FC
LAPTM5	2.51
SLC11A1	2.41
CD68	2.39
ADA	2.26
HLA-DOB	2.24
IL4I1	2.14
ATP6V0A4	-2.27
CTSL2	-2.9

2.11 - KIRP

Gene	Log2FC
CTSK	3.0
GPR143	2.57
CD68	2.34
SLC11A1	2.23
LAPTM5	2.12
CTSL2	-2.43
ATP6V0D2	-3.25
ATP6V0A4	-3.52

2.12 - LIHC

Gene	Log2FC
CTSL2	5.97
ATP6V0D2	3.78
CTSK	2.98
TMEM74	2.32
ABCA3	2.29
LAPTM4B	2.29
IL4I1	2.13
GBA	2.0

2.13 - LUAD

Gene	Log2FC
CTSL2	2.64
ATP6V0A4	2.44
IL4I1	2.16
LAMP3	-2.62

2.14 - LUSC

Gene	Log2FC
CTSL2	4.58
ADA	2.11
CTSH	-2.16
CTSG	-2.43
SLC11A1	-2.43
DRAM1	-2.45
LAMP3	-2.71
HYAL1	-2.91
ABCA3	-3.54
NAPSA	-3.66

2.15 - PAAD

Gene	Log2FC
MCOLN2	-2.43
HLA-DOB	-2.67

2.16 - PCPG

Gene	Log2FC
MPO	3.55
IDS	3.48
ABCA2	2.48
SLC36A1	2.19
PLD3	2.09
SLC17A5	2.08
CLCN5	2.08
ABCA3	2.02
MCOLN2	-4.62

2.17 - PRAD

Gene	Log2FC
ATP6V0A4	-2.51

2.18 - READ

Gene	Log2FC
GPR143	3.18
SLC11A1	2.64
ATP6V0D2	-2.08
CTSG	-3.2

2.19 - SARC

Gene	Log2FC
MPO	3.65
IL4I1	2.5
LAPTM5	2.46
CTSK	2.44
SLC11A1	2.15
HLA-DOB	2.06
CTSH	-2.24
HYAL1	-3.37
ATP6V0D2	-4.97
NAPSA	-8.52

2.20 - STAD

Gene	Log2FC
MPO	4.99
SLC11A1	2.3

2.21 - THCA

Gene	Log2FC
NAPSA	4.62
CTSH	2.38
CTSC	2.08

2.22 - THYM

Gene	Log2FC
GPR143	2.76
NAPSA	2.58

2.23 - UCEC

Gene	Log2FC
CTSL2	4.33
NAPSA	3.53
ATP6V0A4	3.06
IFI30	2.95
LAMP3	2.88
IL4I1	2.01
CTSK	-2.85

2.24 - UCS

Gene	Log2FC
ATP6V0D2	3.24
ATP6V0A4	2.39
LAMP3	2.25
LGMN	2.07
CTSL2	-2.98
NEU4	-5.0

Supplementary table 3 - Survival related lysosomal genes per TCGA project

3.1 - BLCA

Gene	P-Value	HR	Worse prognosis
IDUA	0,0026	0,5	Low
TCIRG1	0,004	0,53	Low
TMEM74	0,0044	1,9	High
CTSB	0,005	1,8	High
RPTOR	0,0089	1,8	High
ATP6V0A1	0,0095	1,8	High
SORT1	0,01	1,7	NA
ATP6V0D1	0,012	1,8	NA
CTSV	0,012	1,8	High
NAPSA	0,014	0,59	Low
TPP1	0,014	1,7	High
LAMP2	0,017	1,7	High
PLD3	0,019	1,7	High
CTSH	0,021	0,62	Low
AGA	0,024	1,6	High
HYAL1	0,025	1,6	High
ATP13A2	0,026	1,6	High
PPT2	0,03	1,6	High
CTSA	0,032	1,6	High
ARSB	0,037	1,6	High
CTSS	0,046	0,66	Low
FUCA1	0,048	0,66	Low

3.2 - BRCA

Gene	P-Value	HR	Worse prognosis
ATP6V1H	0,0019	1,7	High
ATP6AP1	0,0021	2	High
CTSF	0,0034	1,9	High
GLB1	0,0039	2	High
STX7	0,0047	2	High
SCARB2	0,005	1,8	High
HLA-DOB	0,0059	0,54	Low
OCRL	0,0071	1,8	High
ATP6V0D1	0,0076	1,9	High
CTSA	0,0085	1,8	High
LAMP2	0,0088	1,8	High
NAAA	0,015	1,8	High

ATP6V1C1	0,017	1,8	High
PLA2G15	0,02	1,7	High
OSTM1	0,025	1,7	NA
ATP6V1E1	0,031	1,6	High
CTNS	0,049	1,6	High
MAN2B1	0,05	0,63	Low

3.3 - CESC

Gene	P-Value	HR	Worse prognosis
SLC15A3	6,30E-05	0,31	Low
STX7	0,00041	3,1	High
IFI30	0,00048	0,35	Low
MAN2B1	0,0012	0,36	Low
ABCA2	0,013	0,49	Low
MCOLN1	0,027	0,53	Low
LAPTM5	0,034	0,54	Low
MPO	0,041	2,6	High
SCARB2	0,047	1,9	High
NCOA7	0,048	1,8	High
NEU4	0,048	0,57	Low

3.4 - COAD

Gene	P-Value	HR	Worse prognosis
SLC11A1	0,027	2	High
ASAH1	0,048	0,51	Low
ATP13A2	0,041	0,54	Low
FUCA1	0,031	0,53	Low
SORT1	0,033	0,54	Low

3.5 - ESCA

Gene	P-Value	HR	Worse prognosis
ATP6V0D2	0,0049	2,3	High

3.6 - GBM

Gene	P-Value	HR	Worse prognosis
SGSH	0,00028	2,3	High
SMPD1	0,0018	2	High
CTSB	0,0022	2	High
CTSL	0,0022	2	High
FUCA1	0,0027	1,9	High
GRN	0,0072	1,8	High
CTSD	0,0076	1,8	High
PLD3	0,0086	1,8	High
MPO	0,0096	1,8	High
HEXB	0,0097	1,8	High
TCIRG1	0,014	1,7	High
CTSZ	0,018	1,7	High
CTSK	0,021	1,7	High
GLA	0,022	1,7	High
GNS	0,022	1,7	High
SLC11A1	0,022	1,7	High
DNASE2	0,025	1,6	High
SCPEP1	0,025	1,6	High
GUSB	0,028	1,7	High
ACP2	0,033	1,6	High
NEU4	0,033	0,62	Low
LITAF	0,039	1,6	High
IL4I1	0,043	1,6	High
GAA	0,05	1,5	High

3.7 - HNSC

Gene	P-Value	HR	Worse prognosis
CTSL	0,0023	1,8	High
NEU1	0,013	1,6	High
ATP6AP1	0,014	1,6	High
GPR143	0,015	0,63	Low
GALNS	0,019	1,6	High
AGA	0,023	1,6	High
ADA	0,027	1,5	High
CD63	0,028	1,5	NA
LAPTM4B	0,029	1,6	High
PLA2G15	0,029	1,5	High
TPP1	0,029	1,5	High
LAPTM4A	0,032	1,5	High

CTSG	0,036	0,67	Low
HLA-DOB	0,041	0,68	Low
GRN	0,047	1,5	High
ATP6V1E1	0,049	1,5	High

3.8 - KIRC

Gene	P-Value	HR	Worse prognosis
FUCA1	3,30E-11	0,24	Low
TCIRG1	2,20E-09	3,9	High
CLCN5	3,90E-09	0,23	Low
HYAL1	4,20E-09	0,26	Low
CLN5	1,50E-08	0,28	Low
CTSF	2,10E-07	0,29	Low
ATP6AP1	2,60E-07	0,3	Low
ASAH1	4,90E-07	0,34	Low
NCOA7	5,40E-07	0,32	Low
STX7	8,00E-07	0,33	Low
ATP6V1A	1,10E-06	0,34	Low
ATP6V1D	1,10E-06	0,33	Low
CCDC115	1,70E-06	0,33	Low
PPT2	3,60E-06	0,35	Low
ATP6V1B2	4,40E-06	0,35	Low
NAPSA	4,80E-06	0,34	Low
IDUA	7,40E-06	2,6	High
LIPA	8,10E-06	0,38	Low
CD164	8,40E-06	0,37	Low
ATP6V1C1	8,90E-06	0,37	Low
PPT1	1,50E-05	0,4	Low
HYAL2	2,40E-05	0,37	Low
GALC	2,70E-05	0,38	Low
LAMP2	3,40E-05	0,41	Low
SLC17A5	5,70E-05	0,43	Low
ATP6V1E1	6,40E-05	0,42	Low
NAGLU	7,30E-05	0,41	Low
AGA	7,40E-05	0,41	Low
CTSH	8,20E-05	0,43	Low
SLC11A1	8,70E-05	2,3	High
SORT1	0,00013	0,39	Low
MFSD8	0,00028	0,43	Low
LAMP1	0,00038	0,44	Low
ADA	0,00043	2,1	High

ATP6V0D1	0,00062	0,46	Low
CTSG	0,00067	0,46	Low
ARSB	0,00069	0,48	Low
MAN2B1	0,00069	2,1	High
MANBA	0,00084	0,48	Low
LAPTM4A	0,0009	0,49	Low
CTSL	0,001	0,5	Low
CTSD	0,0012	0,5	Low
ABHD5	0,0014	0,51	Low
GM2A	0,0017	0,5	Low
PLD3	0,0024	0,52	Low
CLCN6	0,0029	0,48	Low
CTNS	0,0036	0,54	Low
LAMTOR2	0,0041	1,8	High
LAMTOR1	0,0045	0,54	Low
ATP6V0A1	0,0048	0,46	Low
TPP1	0,0058	0,55	Low
SPPL2A	0,006	0,55	Low
ACP5	0,0061	0,56	Low
ARSG	0,0067	0,55	Low
CTSZ	0,007	1,7	High
ABCA5	0,0074	0,55	Low
SGSH	0,0075	1,8	High
SCARB2	0,0078	0,59	Low
GBA	0,0082	0,57	Low
DRAM1	0,0084	0,56	Low
SLC36A1	0,0096	0,56	Low
PSAP	0,01	0,59	Low
ATP6V0D2	0,011	0,58	Low
GLA	0,012	1,8	High
GALNS	0,014	1,7	High
IDS	0,014	0,58	Low
HLA-DMB	0,016	0,6	Low
LGMN	0,017	0,59	Low
ATP6V0C	0,019	0,6	Low
CTSS	0,019	0,61	Low
ARSA	0,028	1,6	High
MAN2B2	0,029	0,61	Low
HLA-DOA	0,031	0,64	Low
NEU1	0,037	0,65	Low
NAGA	0,039	0,66	Low

3.9 - KIRP

Gene	P-Value	HR	Worse prognosis
ADA	1,80E-04	4,4	High
SLC15A3	2,20E-03	0,28	Low
ATP6V0D1	3,10E-03	0,3	Low
ATP6V1E1	3,40E-03	0,31	Low
SMPD1	7,80E-03	0,36	Low
LAMP3	9,60E-03	2,7	High
HYAL1	1,10E-02	0,34	Low
ATP6V0A1	1,30E-02	0,39	Low
IL4I1	1,40E-02	2,5	High
NEU4	1,50E-02	2,8	High
NCSTN	2,00E-02	2,5	High
FUCA1	2,80E-02	0,48	Low
CTSF	3,10E-02	0,48	Low
ATP6V1D	3,30E-02	0,47	Low
ATP6V0B	3,80E-02	0,47	Low
SLC36A1	4,10E-02	2,1	High
ACP2	4,20E-02	0,47	Low
ATP6V0D2	4,20E-02	0,42	Low
ARSG	4,90E-02	0,49	Low
ATP6V1B2	4,90E-02	0,47	Low

3.10 - LAML

Gene	P-Value	HR	Worse prognosis
ATP13A2	4,80E-06	4,7	High
CLCN5	2,10E-04	3,6	High
ATP6V1H	6,00E-03	2,6	High
MPO	6,00E-03	0,35	Low
HLA-DMA	6,50E-03	2,8	High
MCOLN1	1,40E-02	2,4	High
ACP2	1,50E-02	2,2	High
SLC36A1	1,70E-02	0,46	Low
LIPA	2,00E-02	2,4	High
ATP6V1A	3,10E-02	2,2	High
MCOLN2	3,40E-02	2,1	High
MTOR	4,20E-02	2	High
ABCA5	4,30E-02	2	High
SIDT2	4,40E-02	1,9	High
LAPTM5	4,80E-02	2	High
CTSD	4,90E-02	2,1	High

3.11 - LGG

Gene	P-Value	HR	Worse prognosis
HEXB	6,70E-13	8	High
GLA	3,40E-11	5	High
GUSB	3,40E-11	5,4	High
TCIRG1	5,00E-10	4,4	High
CD63	5,80E-09	4,3	High
NEU4	3,70E-08	0,25	Low
SLC11A1	6,00E-08	3,9	High
GNS	6,10E-08	4,3	High
GLB1	8,60E-08	4,1	High
SGSH	1,10E-07	3,6	High
CD164	2,00E-07	3,8	High
FUCA1	4,80E-07	3,8	High
LAMP3	7,20E-07	3,7	High
CTSC	1,10E-06	3,8	High
CTSZ	1,20E-06	3,7	High
GRN	1,60E-06	3,5	High
SLC15A3	2,10E-06	3,9	High
CTSA	2,20E-06	3,1	High
OSTM1	2,40E-06	3,5	High
IFI30	4,30E-06	3,2	High
HLA-DMB	5,00E-06	3,3	High
LITAF	5,10E-06	3	High
GALNS	6,50E-06	2,9	High
LAMTOR2	7,30E-06	3	High
CTSL	8,40E-06	3	High
SCPEP1	1,00E-05	2,8	High
MCOLN2	1,40E-05	3,1	High
CLCN5	1,50E-05	3,1	High
HLA-DMA	1,70E-05	3,2	High
HLA-DOA	1,90E-05	3,1	High
ATP6V0B	2,00E-05	3,1	High
ABCA5	2,10E-05	2,8	High
MAN2B1	2,10E-05	3	High
SPPL2A	3,50E-05	2,9	High
LAMP2	3,90E-05	2,8	High
HEXA	4,70E-05	2,8	High
PLA2G15	4,70E-05	2,9	High
NCSTN	5,30E-05	2,9	High
PLD3	5,30E-05	3,1	High

IL4I1	6,20E-05	2,9	High
LAPTM4A	7,00E-05	2,8	High
ATP6V0A1	0,00011	0,36	Low
NAGA	0,00011	2,9	High
SMPD1	0,00011	2,7	High
DRAM2	0,00012	2,9	High
ACP5	0,00016	2,7	High
CTSS	0,00029	2,6	High
LAMP1	0,00049	2,7	High
CTSB	0,0005	2,4	High
MTOR	0,0005	2,6	High
AGA	0,00072	2,3	High
HGSNAT	0,00079	2,3	High
SLC15A4	0,00089	2,5	High
ATP6AP1	0,0011	2,4	High
CTSK	0,0011	2,3	High
IDUA	0,0012	2,2	High
SLC37A3	0,0013	2,6	High
CD68	0,0014	2,3	High
MAN2B2	0,0018	2,2	High
NAGLU	0,002	2,2	High
NAAA	0,0021	2,1	High
CTNS	0,0024	2,2	High
HYAL2	0,0024	2,1	High
LAPTM5	0,003	2,2	High
ATP6V0A2	0,0049	2,1	High
GAA	0,0055	2,1	High
CLN5	0,0083	2,4	High
CTSD	0,0092	1,9	High
ABCA3	0,0093	0,52	Low
OCRL	0,0097	2,2	High
CCDC115	0,011	0,54	Low
ACP2	0,012	2	High
ATP6V0D1	0,015	0,52	Low
HLA-DOB	0,015	1,9	High
CLN3	0,016	1,8	High
ARSA	0,019	1,7	High
LAMTOR1	0,02	0,53	Low
PPT1	0,025	1,8	High
SORT1	0,025	1,9	High
ATP13A2	0,027	1,9	High
DNASE2	0,028	1,7	High

MCOLN1	0,035	1,7	High
SLC36A1	0,041	1,8	High
CTSF	0,042	0,59	Low
PSAP	0,046	0,61	Low

3.12 - LIHC

Gene	P-Value	HR	Worse prognosis
CD63	1,90E-05	2,9	High
ADA	3,30E-05	3,2	High
LAPTM4B	5,60E-05	2,9	High
ATP6V0B	2,80E-04	2,6	High
SLC36A1	3,70E-04	2,3	High
CTSC	6,60E-04	2,4	High
DRAM2	6,90E-04	2,3	High
PPT1	7,10E-04	2,3	High
ATP13A2	9,50E-04	2,2	High
ATP6V1E1	1,20E-03	2,1	High
CTSA	1,30E-03	2,2	High
CTNS	1,60E-03	2,2	High
GALNS	2,00E-03	2,1	High
NCSTN	2,50E-03	2,1	High
GBA	2,90E-03	2,1	High
SLC37A3	3,10E-03	2,1	High
IL4I1	3,60E-03	2	High
GLB1	3,70E-03	2,1	High
PSAP	4,20E-03	2,1	High
CTSV	4,40E-03	2	High
NEU1	6,10E-03	2	High
SCPEP1	6,80E-03	1,9	High
SLC11A1	7,80E-03	1,9	High
TPP1	8,10E-03	1,9	High
GNS	8,30E-03	1,9	High
HEXB	9,10E-03	1,8	High
GM2A	9,40E-03	1,9	High
ACP2	9,80E-03	1,9	High
SORT1	9,80E-03	1,9	High
ATP6V0D2	9,90E-03	2	High
DRAM1	1,10E-02	1,9	High
LAPTM4A	1,10E-02	1,9	High
GLA	1,20E-02	1,8	High
ATP6V1C1	1,30E-02	1,9	High

OSTM1	1,30E-02	1,8	High
SLC15A4	1,30E-02	1,9	High
ATP6V0D1	1,90E-02	1,8	High
CTSB	2,00E-02	1,8	High
ATP6V1H	2,20E-02	1,8	High
CD68	2,30E-02	1,7	High
LAMTOR1	2,30E-02	1,8	High
GRN	2,40E-02	1,8	High
ATP6V1B2	0,026	1,7	High
CD164	0,027	1,7	High
ABCA3	0,028	1,8	High
CTSL	0,029	1,7	High
CLCN7	0,03	1,7	High
IFI30	0,031	1,7	High
RPTOR	0,035	1,7	High
CLCN6	0,038	1,6	High
CLN3	0,043	1,6	High

3.13 - LUAD

Gene	P-Value	HR	Worse prognosis
HLA-DOB	4,90E-05	0,42	Low
CTSL	4,30E-04	2,1	High
CTSV	1,40E-03	2	High
HLA-DMA	2,10E-03	0,5	Low
CTSH	2,80E-03	0,51	Low
ABCA3	3,30E-03	0,52	Low
FUCA1	3,60E-03	0,53	Low
HLA-DOA	4,60E-03	0,54	Low
MCOLN2	5,10E-03	0,55	Low
HLA-DMB	5,80E-03	0,55	Low
NAPSA	8,00E-03	0,58	Low
SCPEP1	8,30E-03	0,58	Low
SIDT2	1,30E-02	0,59	Low
ATP6V1C1	1,40E-02	1,7	High
ADA	1,60E-02	1,6	High
PPT2	2,20E-02	1,6	High
ATP6V0C	2,80E-02	1,6	High
CLCN6	2,80E-02	0,62	Low
CTSF	3,40E-02	0,63	Low
SGSH	4,50E-02	0,64	Low
SLC36A1	4,50E-02	1,5	High

MAN2B1	4,80E-02	0,65	Low
SCARB2	5,00E-02	1,5	High

3.14 - LUSC

Gene	P-Value	HR	Worse prognosis
PLA2G15	3,40E-03	1,7	High
ABCA3	6,60E-03	1,7	High
ATP6V0C	9,80E-03	1,6	High
CLCN7	9,90E-03	1,7	High
CTSL	1,30E-02	1,6	High
CTSD	1,40E-02	1,6	High
STX7	1,40E-02	1,6	High
NAPSA	1,50E-02	1,6	High
CTSA	1,60E-02	1,6	High
MPO	1,80E-02	1,6	High
PSAP	2,60E-02	1,5	High
DRAM2	3,10E-02	0,65	Low
DRAM1	3,40E-02	1,6	High
CTSS	4,40E-02	1,5	High
ABCA2	4,60E-02	1,5	High

3.15 - OV

Gene	P-Value	HR	Worse prognosis
HLA-DOB	3,80E-05	0,48	Low
STX7	6,30E-03	1,6	High
ATP6V1H	2,00E-02	0,66	Low
ATP6V1B2	2,40E-02	1,5	High
CTSD	2,50E-02	1,5	High
SLC11A1	3,10E-02	1,5	High
CLCN7	3,70E-02	1,4	High
MCOLN2	4,00E-02	0,7	Low
LAMP3	4,20E-02	0,7	Low

3.16 - PAAD

Gene	P-Value	HR	Worse prognosis
ARSG	7,00E-05	0,34	Low
MCOLN1	4,20E-04	0,39	Low
CTSV	9,90E-04	2,4	High
CTNS	1,10E-03	0,43	Low
ABCA5	2,90E-03	0,37	Low

SMPD1	6,90E-03	0,49	Low
NEU1	7,70E-03	0,5	Low
ACP2	9,80E-03	0,52	Low
ATP6V0A1	9,90E-03	0,51	Low
NAGLU	1,20E-02	0,52	Low
RPTOR	1,20E-02	0,54	Low
ATP6V0D1	1,60E-02	0,55	Low
SGSH	2,00E-02	0,55	NA
SIDT2	2,50E-02	0,56	Low
CLCN5	3,00E-02	0,59	Low
NAAA	3,10E-02	0,59	Low
ABCA3	3,50E-02	0,59	Low
CLCN6	3,50E-02	0,58	Low
LAPTM4A	3,50E-02	1,7	High
ATP6AP1	3,80E-02	0,59	Low
SCPEP1	3,90E-02	0,6	Low
ATP6V0B	4,50E-02	0,6	Low

3.17 - PCPG

Gene	P-Value	HR	Worse prognosis
ARSA	1,60E-02	1,40E-09	Low
NCSTN	2,20E-02	6300000000	High
HYAL2	2,70E-02	1,30E-09	Low
OCRL	4,50E-02	23000000000	High

3.18 - PRAD

Gene	P-Value	HR	Worse prognosis
CD164	3,00E-02	1,80E-09	Low
IDUA	3,50E-02	7,1	High
PSAP	4,00E-02	1,60E-09	Low
CTSD	4,50E-02	6,90E-09	Low

3.19 - READ

Gene	P-Value	HR	Worse prognosis
CD63	2,20E-03	7,90E+00	High
OCRL	2,00E-02	0,25	Low
ATP6V0B	3,20E-02	3,70E+00	High
NEU1	3,70E-02	2,50E-01	Low
GPR143	4,30E-02	0,26	Low
LGMN	4,90E-02	3,2	High

3.20 - SARC

Gene	P-Value	HR	Worse prognosis
ASAH1	1,90E-04	3,80E-01	Low
CTSG	2,00E-04	0,39	Low
CTSH	2,30E-04	3,90E-01	Low
LAPTM4B	1,10E-03	2,30E+00	High
CTSS	2,00E-03	0,45	Low
ATP6V0A1	3,60E-03	0,49	Low
MTOR	5,80E-03	2	High
LGMN	8,00E-03	0,5	Low
DNASE2	9,20E-03	0,5	Low
GPR143	9,80E-03	1,9	High
PSAP	1,10E-02	0,49	Low
SLC15A3	1,20E-02	0,52	Low
HLA-DMA	1,50E-02	0,55	Low
MCOLN1	1,50E-02	0,54	Low
GAA	2,00E-02	0,57	Low
ATP6V0D1	2,50E-02	0,57	Low
GM2A	2,50E-02	0,54	Low
TPP1	2,70E-02	0,59	Low
HLA-DOB	3,00E-02	0,59	Low
SCPEP1	4,30E-02	0,59	Low
ARSA	4,60E-02	0,61	Low
HLA-DMB	4,60E-02	0,61	Low
HYAL1	4,90E-02	0,61	Low
SIDT2	5,00E-02	0,62	Low

3.21 - SKCM

Gene	P-Value	HR	Worse prognosis
LAPTM5	4,40E-06	4,20E-01	Low
IL4I1	8,40E-06	0,41	Low
LAMP3	2,70E-05	4,40E-01	Low
CTSV	1,30E-04	2,10E+00	High
CTNS	3,70E-04	2	High
CTSS	6,70E-04	0,52	Low
NCOA7	7,80E-04	0,53	Low
SCPEP1	9,70E-04	0,52	Low
NAAA	1,50E-03	0,55	Low
SLC15A3	1,50E-03	0,55	Low
MCOLN2	2,80E-03	0,57	Low
LGMN	4,00E-03	0,58	Low

CTSC	5,40E-03	0,59	Low
IFI30	5,70E-03	0,59	Low
NAPSA	7,30E-03	0,61	Low
HLA-DMA	7,40E-03	0,46	Low
CLCN7	1,30E-02	1,6	High
HLA-DOB	1,50E-02	0,5	Low
TCIRG1	1,50E-02	0,62	Low
IDUA	2,00E-02	0,64	Low
LAPTM4B	2,00E-02	1,5	High
ATP6V0D1	2,10E-02	1,5	High
ATP6V0A4	2,30E-02	1,6	High
RPTOR	2,30E-02	1,6	High
GPR143	2,40E-02	2	High
ACP2	2,60E-02	0,65	Low
LAMTOR2	2,70E-02	1,5	High
SPPL2A	2,80E-02	0,66	Low
ATP6V0A1	3,00E-02	1,5	High
CD164	3,20E-02	0,67	Low
HEXB	3,20E-02	0,66	Low

3.22 - STAD

Gene	P-Value	HR	Worse prognosis
GLA	3,30E-03	5,00E-01	Low
CTSF	4,00E-03	2	High
TPP1	6,00E-03	1,90E+00	High
ARSB	1,00E-02	1,80E+00	High
CTSK	1,10E-02	1,8	High
TCIRG1	1,80E-02	0,59	Low
SGSH	2,00E-02	1,7	High
LGMN	2,10E-02	1,7	High
GUSB	3,30E-02	1,6	High

3.23 - TGCT

Gene	P-Value	HR	Worse prognosis
DRAM1	3,30E-02	1,90E+09	High
NAGA	4,80E-02	23000000000	High

3.24 - THCA

Gene	P-Value	HR	Worse prognosis
CLCN5	5,50E-03	1,10E+01	High
GAA	1,80E-02	5,2	High
IDUA	2,50E-02	5,00E+00	High
RPTOR	3,00E-02	7,20E+00	High
ATP6V0D1	3,70E-02	4,6	High
CTSH	3,70E-02	0,14	Low
GUSB	4,40E-02	6,6	High
CD63	5,00E-02	0,24	Low

3.25 - THYM

Gene	P-Value	HR	Worse prognosis
GBA	1,90E-03	2,40E+10	High
SMPD1	3,00E-03	4,80E-10	Low
GUSB	7,60E-03	6,00E-10	Low
ADA	1,10E-02	1,70E-10	Low
PLA2G15	1,20E-02	4,80E-10	Low
CTSD	2,10E-02	9	High
ATP6V0A4	2,80E-02	0,13	Low
SLC11A1	3,40E-02	7,3	High
HLA-DMA	3,90E-02	7,8	High
CD68	4,50E-02	7	High
ABCA5	4,60E-02	6,7	High
CTSC	5,00E-02	6,3	High

3.26 - UCEC

Gene	P-Value	HR	Worse prognosis
CTSA	5,30E-03	3,80E+00	High
HLA-DMB	6,60E-03	2,80E-01	Low
HEXA	1,10E-02	3,40E-01	Low
HLA-DMA	1,70E-02	3,10E-01	Low
ARSA	3,10E-02	3,40E-01	Low
CTSB	4,40E-02	0,39	Low

3.27 - UCS

Gene	P-Value	HR	Worse prognosis
ATP6V0D2	2,30E-02	2,20E+00	High
CCDC115	2,40E-02	2,20E+00	High
ATP6AP1	3,00E-02	2,10E+00	High

Supplementary table 4 - Enriched Biological Processes in DELG

#term ID	term description	observed gene count	background gene count	strength	false discovery rate
GO:0006955	Immune response	16	1588	0.95	2.27e-09
GO:0002376	Immune system process	17	2481	0.78	5.32e-08
GO:0002250	Adaptive immune response	9	317	1.4	1.55e-07
GO:0019886	Antigen processing and presentation of exogenous peptide antigen via MHC class II	6	96	1.74	3.85e-06
GO:0048002	Antigen processing and presentation of peptide antigen	7	191	1.51	3.85e-06
GO:0002274	Myeloid leukocyte activation	9	585	1.14	1.02e-05
GO:0002577	Regulation of antigen processing and presentation	4	20	2.25	1.60e-05
GO:0043312	Neutrophil degranulation	8	484	1.17	4.10e-05
GO:0002578	Negative regulation of antigen processing and presentation	3	9	2.47	15
GO:0002252	Immune effector process	9	969	0.92	25
GO:0050896	Response to stimulus	20	8046	0.34	71
GO:0002587	Negative regulation of antigen processing and presentation of peptide antigen via MHC class II	2	2	2.95	29
GO:0002503	Peptide antigen assembly with MHC class II protein complex	2	4	2.65	64
GO:0006027	Glycosaminoglycan catabolic process	3	61	1.64	143
GO:0060586	Multicellular organismal iron ion homeostasis	2	9	2.3	184
GO:0002604	Regulation of dendritic cell antigen processing and presentation	2	11	2.21	255
GO:0010033	Response to organic substance	11	3011	0.51	416
GO:0030214	Hyaluronan catabolic process	2	16	02.05	478

Supplementary table 5 - Enriched Biological Processes in SRLG

#term ID	term description	observed gene count	background gene count	strength	false discovery rate
GO:0009056	Catabolic process	19	2042	0.88	3.84e-11
GO:0043312	Neutrophil degranulation	13	484	1.34	3.84e-11
GO:0044248	Cellular catabolic process	18	1758	0.92	3.84e-11
GO:1901565	Organonitrogen compound catabolic process	16	1070	01.09	3.84e-11
GO:1901575	Organic substance catabolic process	18	1750	0.92	3.84e-11
GO:0009057	Macromolecule catabolic process	15	1058	01.06	7.18e-11
GO:0006955	Immune response	15	1588	0.89	1.67e-08
GO:0002376	Immune system process	17	2481	0.75	4.01e-08
GO:0046466	Membrane lipid catabolic process	5	33	02.09	3.97e-07
GO:1901136	Carbohydrate derivative catabolic process	7	183	1.49	9.64e-07
GO:0019377	Glycolipid catabolic process	4	13	2.4	1.75e-06
GO:0006027	Glycosaminoglycan catabolic process	5	61	1.82	6.24e-06
GO:0030149	Sphingolipid catabolic process	4	29	02.05	2.62e-05
GO:0030163	Protein catabolic process	9	694	01.02	2.77e-05
GO:0006664	Glycolipid metabolic process	5	106	1.58	7.43e-05
GO:0016139	Glycoside catabolic process	3	8	2.49	8.96e-05
GO:0030574	Collagen catabolic process	4	43	1.88	9.83e-05
GO:0007033	Vacuole organization	5	131	1.49	18
GO:0008152	Metabolic process	22	8298	0.33	20
GO:0009311	Oligosaccharide metabolic process	4	53	1.79	20
GO:0046479	Glycosphingolipid catabolic process	3	12	2.31	22
GO:0044265	Cellular macromolecule catabolic process	9	917	0.9	23
GO:0006687	Glycosphingolipid metabolic process	4	62	1.72	33

GO:0007040	Lysosome organization	4	63	1.71	34
GO:0071704	Organic substance metabolic process	21	7755	0.34	44
GO:0046514	Ceramide catabolic process	3	18	2.13	51
GO:0007035	Vacuolar acidification	3	23	02.03	95
GO:0006672	Ceramide metabolic process	4	87	1.57	97
GO:0044237	Cellular metabolic process	20	7513	0.34	19
GO:0016052	Carbohydrate catabolic process	4	115	1.45	26
GO:1901135	Carbohydrate derivative metabolic process	8	987	0.82	32
GO:0005975	Carbohydrate metabolic process	6	467	01.02	36
GO:0006689	Ganglioside catabolic process	2	6	2.43	75
GO:0006807	Nitrogen compound metabolic process	18	6852	0.33	139
GO:0043170	Macromolecule metabolic process	17	6137	0.35	149
GO:0042445	Hormone metabolic process	4	194	1.23	162
GO:0044257	Cellular protein catabolic process	6	633	0.89	166
GO:0044281	Small molecule metabolic process	9	1684	0.64	178
GO:0009313	Oligosaccharide catabolic process	2	12	2.13	219
GO:0042340	Keratan sulfate catabolic process	2	12	2.13	219
GO:0097067	Cellular response to thyroid hormone stimulus	2	15	02.04	315
GO:0044238	Primary metabolic process	18	7332	0.3	338
GO:0030214	Hyaluronan catabolic process	2	16	02.01	345
GO:0006590	Thyroid hormone generation	2	17	1.98	376
GO:0002224	Toll-like receptor signaling pathway	3	100	1.39	403
GO:0045730	Respiratory burst	2	18	1.96	408
GO:1903510	Mucopolysaccharide metabolic process	3	108	1.35	483

9 References

1 - Zhang, X., Lai, H., Zhang, F., Wang, Y., Zhang, L., Yang, N., et al. (2021). Visualization and Analysis in the Field of Pan-Cancer Studies and Its Application in Breast Cancer Treatment. *Frontiers in medicine*, *8*, 635035. https://doi.org/10.3389/fmed.2021.635035

2 - Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell reports*, *18*(1), 248–262. https://doi.org/10.1016/j.celrep.2016.12.019

3 - Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, *502*(7471), 333–339. https://doi.org/10.1038/nature12634

4 - Yang, X., Gao, L., & Zhang, S. (2017). Comparative pan-cancer DNA methylation analysis reveals cancer common and specific patterns. *Briefings in bioinformatics*, *18*(5), 761–773. https://doi.org/10.1093/bib/bbw063

5 - Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics*, *45*(10), 1113–1120. https://doi.org/10.1038/ng.2764

6 - Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

7 - Zhang, S., Jiang, H., Gao, B., Yang, W., & Wang, G. (2022). Identification of Diagnostic Markers for Breast Cancer Based on Differential Gene Expression and Pathway Network. *Frontiers in cell and developmental biology*, *9*, 811585. https://doi.org/10.3389/fcell.2021.811585

8 - Han, J., Chen, M., Wang, Y., Gong, B., Zhuang, T., Liang, L., et al. (2018). Identification of Biomarkers Based on Differentially Expressed Genes in Papillary Thyroid Carcinoma. *Scientific reports*, *8*(1), 9912. https://doi.org/10.1038/s41598-018-28299-9

9 - Wei, F. Z., Mei, S. W., Wang, Z. J., Chen, J. N., Shen, H. Y., Zhao, F. Q., et al. (2020). Differential Expression Analysis Revealing CLCA1 to Be a Prognostic and Diagnostic Biomarker for Colorectal Cancer. *Frontiers in oncology*, *10*, 573295. https://doi.org/10.3389/fonc.2020.573295

10 - Cox, D. R.(1972) Regression Models and Life-Tables. *Journal of The Royal Statistical Society*. Vol. 34, No. 2

11 - Li, B., Cui, Y., Diehn, M., & Li, R. (2017). Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. *JAMA oncology*, *3*(11), 1529–1537. https://doi.org/10.1001/jamaoncol.2017.1609

12 - Pan, J. H., Zhou, H., Cooper, L., Huang, J. L., Zhu, S. B., Zhao, X. X., et al. (2019). LAYN Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric and Colon Cancers. *Frontiers in immunology*, *10*, 6. https://doi.org/10.3389/fimmu.2019.00006

13 - Tian, Z., Tang, J., Liao, X., Yang, Q., Wu, Y., & Wu, G. (2020). Identification of a 9-gene prognostic signature for breast cancer. *Cancer medicine*, *9*(24), 9471–9484. https://doi.org/10.1002/cam4.3523

14 - Davidson, S. M., & Vander Heiden, M. G. (2017). Critical Functions of the Lysosome in Cancer Biology. *Annual review of pharmacology and toxicology*, *57*, 481–507. https://doi.org/10.1146/annurev-pharmtox-010715-103101

15 - Piao, S., & Amaravadi, R. K. (2016). Targeting the lysosome in cancer. *Annals of the New York Academy of Sciences*, *1371*(1), 45–54. https://doi.org/10.1111/nyas.12953

16 - Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27.

17 - Brozzi, A., Urbanelli, L., Germain, P. L., Magini, A., & Emiliani, C. (2013). hLGDB: a database of human lysosomal genes and their regulation. *Database : the journal of biological databases and curation*, *2013*, bat024. https://doi.org/10.1093/database/bat024

18 - Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current protocols in bioinformatics*, *54*, 1.30.1–1.30.33. https://doi.org/10.1002/cpbi.5

19 - Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research*, *43*(Database issue), D447–D452. https://doi.org/10.1093/nar/gku1003

20 - Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic acids research*, *44*(8), e71. https://doi.org/10.1093/nar/gkv1507

21 - Tang, Z., Kang, B., Li, C., Chen, T., & Zhang, Z. (2019). GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic acids research*, *47*(W1), W556–W560. https://doi.org/10.1093/nar/gkz430

22 - Mellins, E. D., & Stern, L. J. (2014). HLA-DM and HLA-DO, key regulators of MHC-II processing and presentation. *Current opinion in immunology*, *26*, 115–122. https://doi.org/10.1016/j.coi.2013.11.005

23 - Ji, H., Ba, Y., Ma, S., Hou, K., Mi, S., Gao, X., et al. (2021). Construction of Interferon-Gamma-Related Gene Signature to Characterize the Immune-Inflamed Phenotype of Glioblastoma and Predict Prognosis, Efficacy of Immunotherapy and Radiotherapy. *Frontiers in immunology*, *12*, 729359. https://doi.org/10.3389/fimmu.2021.729359

24 - Verbovšek, U., Motaln, H., Rotter, A., Atai, N. A., Gruden, K., Van Noorden, C. J., & Lah, T. T. (2014). Expression analysis of all protease genes reveals cathepsin K to be overexpressed in glioblastoma. *PloS one*, *9*(10), e111819. https://doi.org/10.1371/journal.pone.0111819

25 - Hira, V. V., Verbovšek, U., Breznik, B., Srdič, M., Novinec, M., Kakar, H., et al. (2017). Cathepsin K cleavage of SDF-1α inhibits its chemotactic activity towards glioblastoma stem-like cells. *Biochimica et biophysica acta. Molecular cell research*, *1864*(3), 594–603. https://doi.org/10.1016/j.bbamcr.2016.12.021

26 - Soichi, O., Masanori, N., Hideo, T., Kazunori, A., Nobuya, I., & Jun-ichi, K. (2007). Clinical significance of ABCA2' a possible molecular marker for oligodendrogliomas. *Neurosurgery*, *60*(4), 707–714. https://doi.org/10.1227/01.NEU.0000255395.15657.06

27 - Wei, L., Shao, N., Peng, Y., & Zhou, P. (2021). Inhibition of Cathepsin S Restores TGF-β-induced Epithelial-to-mesenchymal Transition and Tight Junction Turnover in Glioblastoma Cells. *Journal of Cancer*, *12*(6), 1592–1603. https://doi.org/10.7150/jca.50631

28 - Galavotti, S., Bartesaghi, S., Faccenda, D., Shaked-Rabi, M., Sanzone, S., McEvoy, A., et al. (2013). The autophagy-associated factors DRAM1 and p62 regulate cell migration and invasion in glioblastoma stem cells. *Oncogene*, *32*(6), 699–712. https://doi.org/10.1038/onc.2012.111

29 - Wang, L., Wei, B., Hu, G., Wang, L., Bi, M., Sun, Z., & Jin, Y. (2015). Screening of differentially expressed genes associated with human glioblastoma and functional analysis using a DNA microarray. *Molecular medicine reports*, *12*(2), 1991–1996. https://doi.org/10.3892/mmr.2015.3659

30 - Zhu, C., Chen, X., Guan, G., Zou, C., Guo, Q., Cheng, P., et al. (2020). IFI30 Is a Novel Immune-Related Target with Predicting Value of Prognosis and Treatment Response in Glioblastoma. *OncoTargets and therapy*, *13*, 1129–1143. https://doi.org/10.2147/OTT.S237162

31 - Berberich, A., Bartels, F., Tang, Z., Knoll, M., Pusch, S., Hucke, N., et al. (2020). *LAPTM5*-CD40 Crosstalk in Glioblastoma Invasion and Temozolomide Resistance. *Frontiers in oncology*, *10*, 747. https://doi.org/10.3389/fonc.2020.00747

32 - Lin, H., Wang, K., Xiong, Y., Zhou, L., Yang, Y., Chen, S., et al. (2022). Identification of Tumor Antigens and Immune Subtypes of Glioblastoma for mRNA Vaccine Development. *Frontiers in immunology*, *13*, 773264. https://doi.org/10.3389/fimmu.2022.773264

33 - Morelli, M. B., Nabissi, M., Amantini, C., Tomassoni, D., Rossi, F., Cardinali, C., et al. (2016). Overexpression of transient receptor potential mucolipin-2 ion channels in gliomas: role in tumor growth and progression. *Oncotarget*, *7*(28), 43654–43668. https://doi.org/10.18632/oncotarget.9661

34 - Tian, Y., Liu, H., Zhang, C., Liu, W., Wu, T., Yang, X., et al. (2022). Comprehensive Analyses of Ferroptosis-Related Alterations and Their Prognostic Significance in Glioblastoma. *Frontiers in molecular biosciences*, *9*, 904098. https://doi.org/10.3389/fmolb.2022.904098

35 - Yan, Z., Chu, S., Zhu, C., Han, Y., Liang, Q., Shen, S., et al. (2021). Development of a T-cell activation-related module with predictive value for the prognosis and immune checkpoint blockade therapy response in glioblastoma. *PeerJ*, *9*, e12547. https://doi.org/10.7717/peerj.12547

36 - Ho, K. H., Cheng, C. H., Chou, C. M., Chen, P. H., Liu, A. J., Lin, C. W., et al. (2019). miR-140 targeting CTSB signaling suppresses the mesenchymal transition and enhances temozolomide cytotoxicity in glioblastoma multiforme. *Pharmacological research*, *147*, 104390. https://doi.org/10.1016/j.phrs.2019.104390

37 - Zhang, X., Wang, X., Xu, S., Li, X., & Ma, X. (2018). Cathepsin B contributes to radioresistance by enhancing homologous recombination in glioblastoma. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, *107*, 390–396. https://doi.org/10.1016/j.biopha.2018.08.007

38 - Wu, Y., Huang, Y., Zhou, C., Wang, H., Wang, Z., Wu, J., et al. (2022). A Novel Necroptosis-Related Prognostic Signature of Glioblastoma Based on Transcriptomics Analysis and Single Cell Sequencing Analysis. *Brain sciences*, *12*(8), 988. https://doi.org/10.3390/brainsci12080988

39 - Zheng, W., Chen, Q., Wang, C., Yao, D., Zhu, L., Pan, et al. (2020). Inhibition of Cathepsin D (CTSD) enhances radiosensitivity of glioblastoma cells by attenuating autophagy. *Molecular carcinogenesis*, *59*(6), 651–660. https://doi.org/10.1002/mc.23194

40 - Tong, S., Xia, M., Xu, Y., Sun, Q., Ye, L., Yuan, F., et al. (2023). Identification and validation of a novel prognostic signature based on mitochondria and oxidative stress related genes for glioblastoma. *Journal of translational medicine*, *21*(1), 136. https://doi.org/10.1186/s12967-023-03970-6

41 - Majc, B., Habič, A., Novak, M., Rotter, A., Porčnik, A., Mlakar, J., et al. (2022). Upregulation of Cathepsin X in Glioblastoma: Interplay with γ-Enolase and the Effects of Selective Cathepsin X Inhibitors. *International journal of molecular sciences*, *23*(3), 1784. https://doi.org/10.3390/ijms23031784

42 - Wang, K., Lu, Y., Liu, Z., Diao, M., & Yang, L. (2022). Establishment and External Validation of a Hypoxia-Derived Gene Signature for Robustly Predicting Prognosis and Therapeutic Responses in Glioblastoma Multiforme. *BioMed research international*, *2022*, 7858477. https://doi.org/10.1155/2022/7858477

43 - Zois, C. E., Hendriks, A. M., Haider, S., Pires, E., Bridges, E., Kalamida, D., et al. (2022). Liver glycogen phosphorylase is upregulated in glioblastoma and provides a metabolic vulnerability to high dose radiation. *Cell death & disease*, *13*(6), 573. https://doi.org/10.1038/s41419-022-05005-2

44 - Xu, S. M., Xiao, H. Y., Hu, Z. X., Zhong, X. F., Zeng, Y. J., Wu, Y. X., et al. (2023). GRN is a prognostic biomarker and correlated with immune infiltration in glioma: A study based on TCGA data. *Frontiers in oncology*, *13*, 1162983. https://doi.org/10.3389/fonc.2023.1162983

45 - Jia, M., Zhang, W., Zhu, J., Huang, C., Zhou, J., Lian, J., et al. (2021). Microglia-Specific Expression of *HEXA* and *HEXB* Leads to Poor Prognosis in Glioblastoma Patients. *Frontiers in oncology*, *11*, 685893. https://doi.org/10.3389/fonc.2021.685893

46 - Ji, H., Ba, Y., Ma, S., Hou, K., Mi, S., Gao, X., et al. (2021). Construction of Interferon-Gamma-Related Gene Signature to Characterize the Immune-Inflamed Phenotype of Glioblastoma and Predict Prognosis, Efficacy of Immunotherapy and Radiotherapy. *Frontiers in immunology*, *12*, 729359. https://doi.org/10.3389/fimmu.2021.729359

47 - Yu, J., Shi, J., Yuan, F., Yin, W., Zeng, H., Ge, L., et al. (2023). Kavain ablates the radio-resistance of IDH-wildtype glioblastoma by targeting LITAF/NF-κB pathway. *Cellular oncology (Dordrecht)*, *46*(1), 179–193. https://doi.org/10.1007/s13402-022-00743-z.

48 - Lai, W., Li, D., Kuang, J., Deng, L., & Lu, Q. (2022). Integrated analysis of single-cell RNA-seq dataset and bulk RNA-seq dataset constructs a prognostic model for predicting survival in human glioblastoma. *Brain and behavior*, *12*(5), e2575. https://doi.org/10.1002/brb3.2575

49 - Silvestri, I., Testa, F., Zappasodi, R., Cairo, C. W., Zhang, Y., Lupo, et al. (2014). Sialidase NEU4 is involved in glioblastoma stem cell survival. *Cell death & disease*, *5*(8), e1381. https://doi.org/10.1038/cddis.2014.349

50 - Bi, J., Khan, A., Tang, J., Armando, A. M., Wu, S., Zhang, W., et al. (2021). Targeting glioblastoma signaling and metabolism with a re-purposed brain-penetrant drug. *Cell reports*, *37*(5), 109957. https://doi.org/10.1016/j.celrep.2021.109957

51 - Qi, C., Lei, L., Hu, J., Wang, G., Liu, J., & Ou, S. (2021). T cell immune regulator 1 is a prognostic marker associated with immune infiltration in glioblastoma multiforme. *Oncology letters*, *21*(4), 252. https://doi.org/10.3892/ol.2021.12514

52 - Kase, M., Adamson, A., Saretok, M., Minajeva, A., Vardja, M., Jõgi, T., et al. (2016). Impact of tumor infiltrating CD63 positive cells on survival in patients with glioblastoma multiforme. *Journal of neurosurgical sciences*, *60*(4), 417–423.

53 - Lotinun, S., Kiviranta, R., Matsubara, T., Alzate, J. A., Neff, L., Lüth, A., et al. (2013). Osteoclast-specific cathepsin K deletion stimulates S1P-dependent bone formation. *The Journal of clinical investigation*, *123*(2), 666–681. https://doi.org/10.1172/JCI64840

54 - Qian, D., He, L., Zhang, Q., Li, W., Tang, D., Wu, C., et al. (2022). Cathepsin K: A Versatile Potential Biomarker and Therapeutic Target for Various Cancers. *Current oncology (Toronto, Ont.)*, *29*(8), 5963–5987. https://doi.org/10.3390/curroncol2908047

55 - Ding, X., Zhang, C., Chen, H., Ren, M., & Liu, X. (2022). Cathepsins Trigger Cell Death and Regulate Radioresistance in Glioblastoma. *Cells*, *11*(24), 4108. https://doi.org/10.3390/cells11244108

56 - Blackwell, J. M., Goswami, T., Evans, C. A., Sibthorpe, D., Papo, N., White, J. K., Searle, S., Miller, E. N., Peacock, C. S., Mohammed, H., & Ibrahim, M. (2001). SLC11A1 (formerly NRAMP1) and disease resistance. *Cellular microbiology*, *3*(12), 773–784. https://doi.org/10.1046/j.1462-5822.2001.00150.x

57 - Xu, D., Wang, Y., Wu, J., Zhang, Y., Liu, Z., Chen, Y., & Zheng, J. (2021). Systematic Characterization of Novel Immune Gene Signatures Predicts Prognostic Factors in Hepatocellular Carcinoma. *Frontiers in cell and developmental biology*, *9*, 686664. https://doi.org/10.3389/fcell.2021.686664

58 - Yao, Z. Y., Xing, C., Liu, Y. W., & Xing, X. L. (2021). Identification of Two Immune Related Genes Correlated With Aberrant Methylations as Prognosis Signatures for Renal Clear Cell Carcinoma. *Frontiers in genetics*, *12*, 750997. https://doi.org/10.3389/fgene.2021.750997

59 - Berciano-Guerrero, M. A., Lavado-Valenzuela, R., Moya, A., delaCruz-Merino, L., Toscano, F., Valdivia, J., et al. (2022). Genes Involved in Immune Reinduction May Constitute Biomarkers of Response for Metastatic Melanoma Patients Treated with Targeted Therapy. *Biomedicines*, *10*(2), 284. https://doi.org/10.3390/biomedicines10020284

60 - Xu, H., Zhang, A., Fang, C., Zhu, Q., Wang, W., Liu, Y., et al. (2022). SLC11A1 as a stratification indicator for immunotherapy or chemotherapy in patients with glioma. *Frontiers in immunology*, *13*, 980378. https://doi.org/10.3389/fimmu.2022.980378

61 - Xu, J., Guo, Y., Ning, W., Wang, X., Li, S., Chen, Y., et al. (2022). Comprehensive Analyses of Glucose Metabolism in Glioma Reveal the Glioma-Promoting Effect of GALM. *Frontiers in cell and developmental biology*, *9*, 717182. https://doi.org/10.3389/fcell.2021.717182

62 - Tomatsu, S., Montaño, A. M., Dung, V. C., Grubb, J. H., & Sly, W. S. (2009). Mutations and polymorphisms in GUSB gene in mucopolysaccharidosis VII (Sly Syndrome). *Human mutation*, *30*(4), 511–519. https://doi.org/10.1002/humu.20828

63 - Kong, X., Zheng, Z., Song, G., Zhang, Z., Liu, H., Kang, J., et al. (2022). Over-Expression of GUSB Leads to Primary Resistance of Anti-PD1 Therapy in Hepatocellular Carcinoma. *Frontiers in immunology*, *13*, 876048. https://doi.org/10.3389/fimmu.2022.876048

64 - Keibler, M. A., Wasylenko, T. M., Kelleher, J. K., Iliopoulos, O., Vander Heiden, M. G., & Stephanopoulos, G. (2016). Metabolic requirements for cancer cell proliferation. *Cancer & metabolism*, *4*, 16. https://doi.org/10.1186/s40170-016-0156-6

65 - Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J., & Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nature communications*, *11*(1), 5120. https://doi.org/10.1038/s41467-020-18794-x

66 - Wei, J., Hu, M., Huang, K., Lin, S., & Du, H. (2020). Roles of Proteoglycans and Glycosaminoglycans in Cancer Development and Progression. *International journal of molecular sciences*, *21*(17), 5983. https://doi.org/10.3390/ijms21175983

67 - Wang, J., & Chi, S. (2022). Characterization of the Immune Cell Infiltration Landscape and a New Prognostic Score in Glioblastoma. *Journal of healthcare engineering*, *2022*, 4326728. https://doi.org/10.1155/2022/4326728

68 - von Figura, K., & Hasilik, A. (1986). Lysosomal enzymes and their receptors. *Annual review of biochemistry*, *55*, 167–193. https://doi.org/10.1146/annurev.bi.55.070186.001123

69 - Yim, W. W., & Mizushima, N. (2020). Lysosome biology in autophagy. *Cell discovery*, *6*, 6. https://doi.org/10.1038/s41421-020-0141-7

70 - Machado, E. R., Annunziata, I., van de Vlekkert, D., Grosveld, G. C., & d'Azzo, A. (2021). Lysosomes and Cancer Progression: A Malignant Liaison. *Frontiers in cell and developmental biology*, *9*, 642494. https://doi.org/10.3389/fcell.2021.642494

71 - Jandrey, E. H. F., Bezerra, M., Inoue, L. T., Furnari, F. B., Camargo, A. A., & Costa, É. T. (2021). A Key Pathway to Cancer Resilience: The Role of Autophagy in Glioblastomas. *Frontiers in oncology*, *11*, 652133. https://doi.org/10.3389/fonc.2021.652133

72 - Jacobs, K. A., Maghe, C., & Gavard, J. (2020). Lysosomes in glioblastoma: pump up the volume. *Cell cycle (Georgetown, Tex.)*, *19*(17), 2094–2104. https://doi.org/10.1080/15384101.2020.1796016

73 - Hanif, F., Muzaffar, K., Perveen, K., Malhi, S. M., & Simjee, S.hU. (2017). Glioblastoma Multiforme: A Review of its Epidemiology and Pathogenesis through Clinical Presentation and Treatment. *Asian Pacific journal of cancer prevention : APJCP*, *18*(1), 3–9. https://doi.org/10.22034/APJCP.2017.18.1.3

74 - Nagy, Á., Munkácsy, G., & Győrffy, B. (2021). Pancancer survival analysis of cancer hallmark genes. *Scientific reports*, *11*(1), 6047. https://doi.org/10.1038/s41598-021-84787-5

75 - Birzu, C., French, P., Caccese, M., Cerretti, G., Idbaih, A., Zagonel, V., et al. (2020). Recurrent Glioblastoma: From Molecular Landscape to New Treatment Perspectives. *Cancers*, *13*(1), 47. https://doi.org/10.3390/cancers13010047

3 CONCLUSÕES E PERSPECTIVAS

Neste estudo, revisando genes de duas bases de dados, identificamos 129 genes primordialmente encontrados em lisossomos. A expressão de diversos desses genes estava associada a patologia de tumores humanos. No entanto, destacamos evidências significativas de seu papel crucial em glioblastoma multiforme, um dos tumores cerebrais mais agressivos, onde observamos um aumento na expressão desses genes e a correlação desse aumento com um pior prognóstico. Descobrimos três genes previamente tanto diferencialmente expressos como relacionados a sobrevida no glioblastoma, dos quais dois, *CTSK* e *SLC11A1*, já eram biomarcadores conhecidos e validados em nossos dados, enquanto o terceiro, *GUSB*, apareceu como um novo potencial biomarcador. Esses resultados são de grande importância, uma vez que o glioblastoma é um tumor de prognóstico muito negativo, e a compreensão do papel dos lisossomos nesse contexto pode abrir caminho para novas abordagens terapêuticas. Nós ainda descobrimos que a utilização de um mesmo cutoff para diferentes tumores em estudos transcriptômicos talvez não seja o mais apropriado, sendo necessário o desenvolvimento de um método para a comparação direta de conjuntos gênicos em estudos pan-câncer.

A partir das evidências do envolvimento dos genes lisossomais na patologia de glioblastomas, pretendemos realizar análises em dados de scRNA-seq a fim de determinar quais tipos celulares são responsáveis pelo aumento da expressão dos genes lisossomais. Caso esses genes sejam expressos por células de glioblastoma, é possível realizar estudos funcionais em linhagens celulares para tentar desvendar como, de fato, esses genes funcionam nesses tumores, como, por exemplo, através da superexpressão dos genes em ensaios de invasão e migração celular. Também seria interessante a análise do efeito de reguladores mestres da biogênese lisossomal, como a cinase mTOR e os fatores de transcrição TFEB e TFE3, sobre os níveis de expressão dos genes lisossomais diferencialmente expressos e dos relacionados à sobrevida em linhagens de glioblastomas. Quanto à possibilidade dos níveis de

expressão de *GUSB* serem biomarcadores para diagnóstico, isso pode ser confirmado por meio da aferição dessa expressão em uma coorte de amostras independentes daquelas do TCGA. A correlação da alta expressão de *GUSB* com um pior prognóstico também pode ser aferida em uma coorte independente através de análise de sobrevida.

REFERÊNCIAS

AGHI, M. & BARKER, F., Benign adult brain tumors: an evidence-based medicine review. **Progress in Neurological Surgery.** Vol. 19. p. 80-96. 2006

AKTER, F., BONINI, S., PONNAIYAN, S., BLEIBAUM, F., et al., Multi-Cell Line Analysis of Lysosomal Proteomes Reveals Unique Features and Novel Lysosomal Proteins. **Molecular & Cellular Proteomics** Vol. 22. Mar, 2023

ANDERS, S. & HUBER, W., Differential expression analysis for sequence count data. **Genome Biology**. Vol. 11. Oct, 2010

BALLABIO, A. & BONIFACINO, J., Lysosomes as dynamic regulators of cell and organismal homeostasis. **Nature Reviews Molecular Cell Biology**. Vol. 21., p. 101-118. Nov, 2019

BIRZU, C., FRENCH, P., CACCESE, M., CERRETI, G., et al., Recurrent Glioblastoma: From Molecular Landscape to New Treatment Perpsectives. **Cancers**. Vol 13, 2020.

BOUHAMDANI, N., COMEAU, D. & TURCOTTE, S., A compedium of information on the lysosome. **Frontiers in Cell Development Biology**. Vol. 9. Dec, 2021.

BRAULKE, T. & BONIFACINO, J., Sorting of Lysosomal Proteins. **Biochemical Biophysics Acta.** Vol. 1793. Apr, 2009.

BROZZI, A., URBANELLI, L., GERMAIN, P., MAGINI, A. & EMILIANI, C., hLGDB: a database of human lysosomal genes and their regulation. **Database (Oxford)**. Vol. 2013. Apr, 2013.

CHANG, H. & ZHOU, Z., Targeting autophagy to overcome drug resistance: further developments. **Journal of Hematology & Oncology**. Vol. 13., Nov, 2020.

COX, D. R., Regression Models and Life-Tables. **Journal of the Royal Statistical Society. Series B (Methodological).** Vol. 34. p. 187-220. 1972

DUARTE, C. B., STAIANO, L., ALMEIDA, C., CUTLER, D., et al., Current methods to analyze lysosome morphology, positioning, motility and function. **Traffic**. Vol. 23., p. 238-269. May, 2022

de DUVE, C., PRESSMAN, B., GIANETTO, R., WATTIAUX, R. & APPELMANS, F., Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. **Biochemical Journal**. Vol. 60, p. 604-617. Aug, 1955

HANAHAN, D. & WEINBERG, R., Hallmarks of Cancer: the next generation.**Cell**. Vol 144,5. p 646-674. Mar, 2011.

HARIKUMAR, P. & REEVES, J., The lysosomal proton pump is electrogenic. **Journal of Biological Chemistry**. Vol. 258. Sep, 1983

HANIF, F., MUZAFFAR, K., PERVEEN, K., MATHI, S. M., et al., Glioblastoma Multiforme: A Review of Its Epidemiology and Pathogenesis through Clinical Presentation and Treatment. **Asian Pacific Journal of Cancer Prevention**. Vol 18, p. 3-9. Jan, 2017
KANEHISA, M. & GOTO, S., KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acid Research**. Vol. 28, p. 27-30. Jan, 2000
LUZIO, J. P., PRYOR, P. & BRIGHT, N., Lysosomes: fusion and function. **Nature Reviews Molecular Cell Biology.** Vol. 8., p. 622-632. Aug, 2007.

NAGY, A., MUNKACSY, G. & GYORFFY, B. Pancancer survival analysis of cancer hallmark genes. **Scientific reports.** Vol. 11. 2021
PU, J., GUARDIA, C., KEREN-KAPLAN, T. & BONIFACINO, J., Mechanisms and functions of lysosome positioning. **Journal of Cell Science**. Vol. 129., p. 4329-4339. Dec, 2016.
RABANAL-RUIZ, Y. & KOROLCHUK, V., mTORC1 and Nutrient Homeostasis: The Central Role of The Lysosome. **International Journal of Molecular Sciences**. Vol. 19. Mar, 2018
RUDZINSKA, M., PARODI, A., SOOND, S., VINAROV, A., et al., The Role of Cysteine Cathepsins in Cancer Progression and Drug Resistance. **International Journal of Molecular Sciences.** Vol. 20. Jul, 2019.
SETTEMBRE, C., FRALDI, A., MEDINA, D. & BALLABIO, A., Signals from the lysosome: a control center for cellular clearance and energy metabolism. **Nature Reviews Molecular Cell Biology**. Vol. 14., p. 283-296. Apr, 2013
STAHL-MEYER, J., HOLLAND, L., LIU, B., MAEDA, K. & JAATTELA, M., Lysosomal Changes in Mitosis. **Cells**. Vol. 11. Mar, 2022
SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V., MUKHERJEE, S., et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. **Proceedings of the National Academy of Sciences (PNAS)**. Vol. 102. p. 15545-15550. Sep, 2005.
SUNG, H., FERLAY, J., SIEGEL, R., LAVERSANNE, M., et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries". **CA: A Cancer Journal for Clinicians**, Vol. 71. p. 209-249. May, 2021.
TANCINI, B., BURATTA, S., DELO, F., SAGINI, K., et al., Lysosomal Exocytosis: The Extracellular Role of an Intracellular Organelle. **Membranes (Basel)**. Vol. 10. Dec, 2020
TAPPEL, L., SHIBKO, S., STEIN, M. & SUSZ, P., Studies on the composition of Lysosomes. **Journal of Food Sciences**. Vol. 30. p. 498-503. May, 1965
WANG, Z., GERSTEIN, M. & SNYDER, M., RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics.** Vol. 10., p. 57-63. Jan, 2009
WATTS, C. Lysosomes and lysosome-related organelles in immune responses. **FEBS Open Bio**. Vol. 12., p. 678-693. Mar, 2022
WEINSTEIN, J. N., COLLISON, E., MILLS, G., SHAW, K., et al., The Cancer Genome Atlas Pan-Cancer Analysis Project. **Nature Genetics**. Vol. 45. p. 1113-1120. Feb, 2014

ANEXO A – NORMAS DE PUBLICAÇÃO DA REVISTA FRONTIERS IN ONCOLOGY

General standards

Article type

Frontiers requires authors to select the appropriate article type for their manuscript and to comply with the article type descriptions defined in the journal's 'Article types' page, which can be found under the 'About journal' menu in 'For authors' on every Frontiers journal page. Please pay close attention to the word count limits.

Templates

If working with Word please use our Word templates. If you wish to submit your article as LaTeX, we recommend our LaTeX templates.

For LaTeX files, please ensure all relevant manuscript files are uploaded: .tex file, PDF, and .bib file (if the bibliography is not already included in the .tex file).

During the interactive review, authors are encouraged to upload versions using track changes. Editors and reviewers can only download the PDF file of the submitted manuscript.

Manuscript length

We encourage you to closely follow the article word count lengths given in the 'Article types' page of the journals. The manuscript length includes only the main body of the text, footnotes, and all citations within it, and excludes the abstract, section titles, figure and table captions, funding statement, acknowledgments, and references in the bibliography. Please indicate the number of words and the number of figures and tables included in your manuscript on the first page.

Language editing

Frontiers requires manuscripts submitted to meet international English language standards to be considered for publication.

For authors who would like their manuscript to receive language editing or proofreading to improve the clarity of the manuscript and help highlight their research, we recommend the language-editing services provided by the following external partners.

Note that sending your manuscript for language editing does not imply or guarantee that it will be accepted for publication by a Frontiers journal. Editorial decisions on the scientific content of a manuscript are independent of whether it has received language editing or proofreading by these partner services or other services.

Editage

Frontiers recommends the language-editing service provided by our external partner Editage. These services may be particularly useful for researchers for whom English is not the primary language. They can help to improve the grammar, syntax, and flow of your manuscript prior to submission. Frontiers authors will receive a 10% discount by visiting the following link: editage.com/frontiers.

The Charlesworth Group

Frontiers recommends the Charlesworth Group's author services, who has a long-standing track record in language editing and proofreading. This is a third-party service for which Frontiers authors will receive a 10% discount by visiting the following link: www.cwauthors.com/frontiers.

Language style

The default language style at Frontiers is American English. If you prefer your article to be formatted in British English, please specify this on the first page of your manuscript. For any questions regarding style, Frontiers recommends authors to consult the Chicago Manual of Style.

Search engine optimization (SEO)

There are a few simple ways to maximize your article's discoverability and search results.

- Include a few of your article's keywords in the title of the article
- Do not use long article titles
- Pick 5-8 keywords using a mix of generic and more specific terms on the article subject(s)
- Use the maximum amount of keywords in the first two sentences of the abstract
- Use some of the keywords in level 1 headings

CrossMark policy

CrossMark is a multi-publisher initiative to provide a standard way for readers to locate the current version of a piece of content. By applying the CrossMark logo Frontiers is committed to maintaining the content it publishes and to alerting readers to changes if and when they occur.

Clicking on the CrossMark logo will tell you the current status of a document and may also give you additional publication record information about the document.

Title

The title should be concise, omitting terms that are implicit and, where possible, be a statement of the main result or conclusion presented in the manuscript. Abbreviations should be avoided within the title.

Witty or creative titles are welcome, but only if relevant and within measure. Consider if a title meant to be thought-provoking might be misinterpreted as offensive or alarming. In extreme cases, the editorial office may veto a title and propose an alternative.

Authors should avoid:

- titles that are a mere question without giving the answer
- unambitious titles, for example starting with 'Towards,' 'A description of,' 'A characterization of' or 'Preliminary study on'
- vague titles, for example starting with 'Role of', 'Link between', or 'Effect of' that do not specify the role, link, or effect
- including terms that are out of place, for example the taxonomic affiliation apart from species name.

For Corrigenda, General Commentaries, and Editorials, the title of your manuscript should have the following format.

- 'Corrigendum: [Title of original article]'
- General Commentaries:
  'Commentary: [Title of original article]'
  'Response: Commentary: [Title of original article]’
- 'Editorial: [Title of Research Topic]'

Authors and affiliations

All names are listed together and separated by commas. Provide exact and correct author names as these will be indexed in official archives. Affiliations should be keyed to the author's name with superscript numbers and be listed as follows:

- Laboratory, Institute, Department, Organization, City, State abbreviation (only for United States, Canada, and Australia), and Country (without detailed address information such as city zip codes or street names).

Example: Max Maximus1
1 Department of Excellence, International University of Science, New York, NY, United States.

Correspondence

The corresponding author(s) should be marked with an asterisk in the author list. Provide the exact contact email address of the corresponding author(s) in a separate section.

Example: Max Maximus*
maximus@iuscience.edu

If any authors wish to include a change of address, list the present address(es) below the correspondence details using a unique superscript symbol keyed to the author(s) in the author list.

Equal contributions

The authors who have contributed equally should be marked with a symbol (†) in the author list of the doc/latex and pdf files of the manuscript uploaded at submission.

Please use the appropriate standard statement(s) to indicate equal contributions:

- **Equal contribution:** These authors contributed equally to this work
- **First authorship:** These authors share first authorship
- **Senior authorship:** These authors share senior authorship
- **Last authorship:** These authors share last authorship
- **Equal contribution and first authorship:** These authors contributed equally to this work and share first authorship
- **Equal contribution and senior authorship:** These authors contributed equally to this work and share senior authorship
- **Equal contribution and last authorship:** These authors contributed equally to this work and share last authorship

Example: Max Maximus 1†, John Smith2† and Barbara Smith1
†These authors contributed equally to this work and share first authorship

Consortium/group and collaborative authors

Consortium/group authorship should be listed in the manuscript with the other author(s).

In cases where authorship is retained by the consortium/group, the consortium/group should be listed as an author separated by a comma or 'and'. The consortium/group name will appear in the author list, in the citation, and in the copyright. If provided, the consortium/group members will be listed in a separate section at the end of the article.

For the collaborators of the consortium/group to be indexed in PubMed, they do not have to be inserted in the Frontiers submission system individually. However, in the manuscript itself, provide a section with the name of the consortium/group as the heading followed by the list of collaborators, so they can be tagged accordingly and indexed properly.

Example: John Smith, Barbara Smith and The Collaborative Working Group.

In cases where work is presented by the author(s) on behalf of a consortium/group, it should be included in the author list separated with the wording 'for' or 'on behalf of.' The consortium/group will not retain authorship and will only appear in the author list.

Example: John Smith and Barbara Smith on behalf of The Collaborative Working Group.

Artificial intelligence

These guidelines cover acceptable uses of generative AI technologies such as Large Language Models (ChatGPT, Jasper) and text-to-image generators (DALL-E 2, Midjourney, Stable Diffusion) in the writing or editing of manuscripts submitted to Frontiers.

AI use by authors

Authors should not list a generative AI technology as a co-author or author of any submitted manuscript. Generative AI technologies cannot be held accountable for all aspects of a manuscript and consequently do not meet the criteria required for authorship.

If the author of a submitted manuscript has used written or visual content produced by or edited using a generative AI technology, this use must follow all Frontiers guidelines and policies. Specifically, the author is responsible for checking the factual accuracy of any content created by the generative AI technology. This includes, but is not limited to, any quotes, citations or references. Figures produced by or edited using a generative AI technology must be checked to ensure they accurately reflect the data presented in the manuscript. Authors must also check that any written or visual content produced by or edited using a generative AI technology is free from plagiarism.

If the author of a submitted manuscript has used written or visual content produced by or edited using a generative AI technology, such use must be acknowledged in the acknowledgements section of the manuscript and the methods section if applicable. This explanation must list the name, version, model, and source of the generative AI technology.
We encourage authors to upload all input prompts provided to a generative AI technology and outputs received from a generative AI technology in the supplementary files for the manuscript.

Abstract

As a primary goal, the abstract should make the general significance and conceptual advance of the work clearly accessible to a broad readership. The abstract should be no longer than a single paragraph and should be structured, for example, according to the IMRAD format. For the specific structure of the abstract, authors should follow the requirements of the article type or journal to which they're submitting. Minimize the use of abbreviations and do not cite references, figures or tables.

For clinical trial articles, please include the unique identifier and the URL of the publicly-accessible website on which the trial is registered.

Keywords

All article types require a minimum of five and a maximum of eight keywords.

Text

The entire document should be single-spaced and must contain page and line numbers in order to facilitate the review process. The manuscript should be written using either Word or LaTeX. See above for templates.

Nomenclature

The use of abbreviations should be kept to a minimum. Non-standard abbreviations should be avoided unless they appear at least four times, and must be defined upon first use in the main text. Consider also giving a list of non-standard abbreviations at the end, immediately before the acknowledgments.

Equations should be inserted in editable format from the equation editor.

Italicize gene symbols and use the approved gene nomenclature where it is available. For human genes, please refer to the HUGO Gene Nomenclature Committee (HGNC). New symbols for human genes should be submitted to the HGNC here. Common alternative gene aliases may also be reported, but should not be used alone in place of the HGNC symbol. Nomenclature committees for other species are listed here. Protein products are not italicized.

We encourage the use of Standard International Units in all manuscripts.

Chemical compounds and biomolecules should be referred to using systematic nomenclature, preferably using the recommendations by the International Union of Pure and Applied Chemistry (IUPAC).

Astronomical objects should be referred to using the nomenclature given by the International Astronomical Union (IAU) provided here.

Life Science Identifiers (LSIDs) for ZOOBANK registered names or nomenclatural acts should be listed in the manuscript before the keywords. An LSID is represented as a uniform resource name (URN) with the following format: urn:lsid:<Authority>:<Namespace>:<ObjectID>[:<Version>]

For more information on LSIDs please see the 'Code' section of our policies and publication ethics.

Sections

The manuscript is organized by headings and subheadings. The section headings should be those appropriate for your field and the research itself. You may insert up to 5 heading levels into your manuscript (i.e.,: 3.2.2.1.2 Heading Title).

For Original Research articles, it is recommended to organize your manuscript in the following sections or their equivalents for your field.

Introduction
Succinct, with no subheadings.

Materials and methods
This section may be divided by subheadings and should contain sufficient detail so that when read in conjunction with cited references, all procedures can be repeated. For experiments reporting results on animal or human subject research, an ethics approval statement should be included in this section (for further information, see the 'Bioethics' section of our policies and publication ethics.)

Results
This section may be divided by subheadings. Footnotes should not be used and must be transferred to the main text.

Discussion
This section may be divided by subheadings. Discussions should cover the key findings of the study: discuss any prior research related to the subject to place the novelty of the discovery in the appropriate context, discuss the potential shortcomings and limitations on their interpretations, discuss their integration into the current understanding of the problem and how this advances the current views, speculate on the future direction of the research, and freely postulate theories that could be tested in the future.

For further information, please check the descriptions defined in the journal's 'Article types' page, in the 'For authors' menu on every journal page.

Acknowledgements

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors. Should the content of the manuscript have previously appeared online, such as in a thesis or preprint, this should be mentioned here, in addition to listing the source within the reference list.

Scope statement

When you submit your manuscript, you will be required to summarize in 200 words your manuscript's scope and its relevance to the journal and/or specialty section you're submitting to. The aim is to convey to editors and reviewers how the contents of your manuscript fit within the selected journal's scope.

This statement will not be published with your article if it is accepted for publication. The information will be used during the initial validation and review processes to assess whether the manuscript is a suitable fit for the chosen journal and specialty.

We encourage you to consider carefully where to submit your manuscript, as submissions to an unsuitable journal or specialty will result in delays and increase the likelihood of manuscript rejection.

If you are submitting to a Research Topic, please also clarify how your submission is suited to the specific topic.

Figure and table guidelines

CC-BY license

All figures, tables, and images will be published under a Creative Commons CC-BY license, and permission must be obtained for use of copyrighted material from other sources (including re-published/adapted/modified/partial figures and images from the internet). It is the responsibility of the authors to acquire the licenses, follow any citation instructions requested by third-party rights holders, and cover any supplementary charges.

For additional information, please see the 'Image manipulation' section of our policies and publication ethics.

Figure requirements and style guidelines

Frontiers requires figures to be submitted individually, in the same order as they are referred to in the manuscript; the figures will then be automatically embedded at the end of the submitted manuscript. Kindly ensure that each figure is mentioned in the text and in numerical order.

For figures with more than one panel, panels should be clearly indicated using labels (A), (B), (C), (D), etc. However, do not embed the part labels over any part of the image, these labels will be replaced during typesetting according to Frontiers' journal style. For graphs, there must be a self-explanatory label (including units) along each axis.

For LaTeX files, figures should be included in the provided PDF. In case of acceptance, our production office might require high-resolution files of the figures included in the manuscript in EPS, JPEG or TIF/TIFF format.

To upload more than one figure at a time, save the figures (labeled in order of appearance in the manuscript) in a zip file and upload them as 'Supplementary Material Presentation.'

Please note that figures not in accordance with the guidelines will cause substantial delay during the production process.

Captions

Captions should be preceded by the appropriate label, for example 'Figure 1.' Figure captions should be placed at the end of the manuscript. Figure panels are referred to by bold capital letters in brackets: (A), (B), (C), (D), etc.

Image size and resolution requirements

Figures should be prepared with the PDF layout in mind. Individual figures should not be longer than one page and with a width that corresponds to 1 column (85 mm) or 2 columns (180 mm).

All images must have a resolution of 300 dpi at final size. Check the resolution of your figure by enlarging it to 150%. If the image appears blurry, jagged, or has a stair-stepped effect, the resolution is too low.

The text should be legible and of high quality. The smallest visible text should be no less than eight points in height when viewed at actual size.

Solid lines should not be broken up. Any lines in the graphic should be no smaller than two points wide.

Please note that saving a figure directly as an image file (JPEG, TIF) can greatly affect the resolution of your image. To avoid this, one option is to export the file as PDF, then convert into TIFF or EPS using a graphics software.

Format and color image mode

The following formats are accepted: TIF/TIFF (.tif/.tiff), JPEG (.jpg), and EPS (.eps) (upon acceptance). Images must be submitted in the color mode RGB.

Chemical structures

Chemical structures should be prepared using ChemDraw or a similar program. If working with ChemDraw please use our ChemDraw template. If working with another program please follow the guidelines below.

- Drawing settings: chain angle, 120° bond spacing, 18% width; fixed length, 14.4 pt; bold width, 2.0 pt; line width, 0.6 pt; margin width, 1.6 pt; hash spacing, 2.5 pt. Scale 100% Atom Label settings: font, Arial; size, 8 pt

- Assign all chemical compounds a bold, Arabic numeral in the order in which the compounds are presented in the manuscript text.

Table requirements and style guidelines

Tables should be inserted at the end of the manuscript in an editable format. If you use a word processor, build your table in Word. If you use a LaTeX processor, build your table in LaTeX. An empty line should be left before and after the table.

Table captions must be placed immediately before the table. Captions should be preceded by the appropriate label, for example 'Table 1.' Please use only a single paragraph for the caption.

Ensure that each table is mentioned in the text and in numerical order.

Large tables covering several pages cannot be included in the final PDF for formatting reasons. These tables will be published as supplementary material.

Tables which are not according to the above guidelines will cause substantial delay during the production process.

Accessibility

We encourage authors to make the figures and visual elements of their articles accessible for the visually impaired. An effective use of color can help people with low visual acuity, or color blindness, understand all the content of an article.

These guidelines are easy to implement and are in accordance with the W3C Web Content Accessibility Guidelines (WCAG 2.1), the standard for web accessibility best practices.

Ensure sufficient contrast between text and its background

People who have low visual acuity or color blindness could find it difficult to read text with low contrast background color. Try using colors that provide maximum contrast.

WC3 recommends the following contrast ratio levels:

- Level AA, contrast ratio of at least 4.5:1
- Level AAA, contrast ratio of at least 7:1

You can verify the contrast ratio of your palette with these online ratio checkers:

- WebAIM
- Color Safe

Avoid using red or green indicators
More than 99% of color-blind people have a red-green color vision deficiency.

Avoid using only color to communicate information
Elements with complex information like charts and graphs can be hard to read when only color is used to distinguish the data. Try to use other visual aspects to communicate information, such as shape, labels, and size. Incorporating patterns into the shape fills also make differences clearer; for an example please see below:

Supplementary material

Data that are not of primary importance to the text, or which cannot be included in the article because they are too large or the current format does not permit it (such as videos, raw data traces, and PowerPoint presentations), can be uploaded as supplementary material during the submission procedure and will be displayed along with the published article. All supplementary files are deposited to figshare for permanent storage and receive a DOI.

Supplementary material is not typeset, so please ensure that all information is clearly presented without tracked changes/highlighted text/line numbers, and the appropriate caption is included in the file. To avoid discrepancies between the published article and the supplementary material, please do not add the title, author list, affiliations or correspondence in the supplementary files.

The supplementary material can be uploaded as:

- data sheet (Word, Excel, CSV, CDX, FASTA, PDF or Zip files)
- presentation (PowerPoint, PDF or Zip files)
- image (CDX, EPS, JPEG, PDF, PNG or TIF/TIFF),
- table (Word, Excel, CSV or PDF)
- audio (MP3, WAV or WMA)
- video (AVI, DIVX, FLV, MOV, MP4, MPEG, MPG or WMV).

Technical requirements for supplementary images:

- 300 DPIs
- RGB color mode.

For supplementary material templates (LaTeX and Word), see our supplementary material templates.

References

Frontiers' journals use one of two reference styles, either Harvard (author-date) or Vancouver (numbered). Please check our help center to find the correct style for the journal to which you are submitting.

- All citations in the text, figures, or tables must be in the reference list and vice-versa
- The names of the first six authors followed by et al. and the DOI (when available) should be provided
- Given names of authors should be abbreviated to initials (e.g., Smith, J., Lewis, C.S., etc.)
- The reference list should only include articles that are published or accepted
- Unpublished data, submitted manuscripts, or personal communications should be cited within the text only, for article types that allow such inclusions
- For accepted but unpublished works use 'in press' instead of page numbers
- Data sets that have been deposited to an online repository should be included in the reference list. Include the version and unique identifier when available
- Personal communications should be documented by a letter of permission
- Website URLs should be included as footnotes
- Any inclusion of verbatim text must be contained in quotation marks and clearly reference the original source
- Preprints can be cited as long as a DOI or archive URL is available, and the citation clearly mentions that the contribution is a preprint. If a peer-reviewed journal publication for the same preprint exists, the official journal publication is the preferred source. See the preprints section for each reference style below for more information.

Harvard reference style (author-date)

Many Frontiers journals use the Harvard referencing system; to find the correct reference style and resources for the journal you are submitting to, please visit our help center. Reference examples are found below, for more examples of citing other documents and general questions regarding the Harvard reference style, please refer to the Chicago Manual of Style.

In-text citations

- For works by a single author, include the surname, followed by the year
- For works by two authors, include both surnames, followed by the year
- For works by more than two authors, include only the surname of the first author followed by et al., followed by the year
- For humanities and social sciences articles, include the page numbers.

Reference list examples

Article in a print journal
Sondheimer, N., and Lindquist, S. (2000). Rnq1: an epigenetic modifier of protein function in yeast. Mol. Cell. 5, 163-172.

Article in an online journal
Tahimic, C.G.T., Wang, Y., Bikle, D.D. (2013). Anabolic effects of IGF-1 signaling on the skeleton. Front. Endocrinol. 4:6. doi: 10.3389/fendo.2013.00006

Article or chapter in a book
Sorenson, P. W., and Caprio, J. C. (1998). "Chemoreception," in The Physiology of Fishes, ed. D. H. Evans (Boca Raton, FL: CRC Press), 375-405.

Book
Cowan, W. M., Jessell, T. M., and Zipursky, S. L. (1997). Molecular and Cellular Approaches to Neural Development. New York: Oxford University Press.

Abstract
Hendricks, J., Applebaum, R., and Kunkel, S. (2010). A world apart? Bridging the gap between theory and applied social gerontology. Gerontologist 50, 284-293. Abstract retrieved from Abstracts in Social Gerontology database. (Accession No. 50360869)

Website
World Health Organization. (2018). E. coli. https://www.who.int/news-room/fact-sheets/detail/e-coli [Accessed March 15, 2018].

Patent
Marshall, S. P. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. U.S. Patent No 6,090,051. Washington, DC: U.S. Patent and Trademark Office.

Data
Perdiguero P, Venturas M, Cervera MT, Gil L, Collada C. Data from: Massive sequencing of

Ulms minor's transcriptome provides new molecular tools for a genus under the constant threat of Dutch elm disease. Dryad Digital Repository. (2015) http://dx.doi.org/10.5061/dryad.ps837

Theses and dissertations

Smith, J. (2008) Post-structuralist discourse relative to phenomological pursuits in the deconstructivist arena. [dissertation/master's thesis]. [Chicago (IL)]: University of Chicago

Preprint

Smith, J. (2008). Title of the document. Preprint repository name [Preprint]. Available at: https://persistent-url (Accessed March 15, 2018).

Vancouver reference style (numbered)

Many Frontiers journals use the numbered referencing system; to find the correct reference style and resources for the journal you are submitting to, please visit our help center.

Reference examples are found below, for more examples of citing other documents and general questions regarding the Vancouver reference style, please refer to Citing Medicine.

In-text citations

- Please apply the Vancouver system for in-text citations
- In-text citations should be numbered consecutively in order of appearance in the text – identified by Arabic numerals in the parenthesis (use square brackets for physics and mathematics articles).

Reference list examples

Article in a print journal

Sondheimer N, Lindquist S. Rnq1: an epigenetic modifier of protein function in yeast. Mol Cell (2000) 5:163-72.

Article in an online journal

Tahimic CGT, Wang Y, Bikle DD. Anabolic effects of IGF-1 signaling on the skeleton. Front Endocrinol (2013) 4:6. doi: 10.3389/fendo.2013.00006

Article or chapter in a book

Sorenson PW, Caprio JC. "Chemoreception". In: Evans DH, editor. The Physiology of Fishes. Boca Raton, FL: CRC Press (1998). p. 375-405.

Book

Cowan WM, Jessell TM, Zipursky SL. Molecular and Cellular Approaches to Neural Development. New York: Oxford University Press (1997). 345 p.

Abstract

Christensen S, Oppacher F. An analysis of Koza's computational effort statistic for genetic programming. In: Foster JA, editor. Genetic Programming. EuroGP 2002: Proceedings of the 5th European Conference on Genetic Programming; 2002 Apr 3–5; Kinsdale, Ireland. Berlin: Springer (2002). p. 182–91.

Website

World Health Organization. E. coli (2018). https://www.who.int/news-room/fact-sheets/detail/e-coli [Accessed March 15, 2018].

Patent

Pagedas AC, inventor; Ancel Surgical R&D Inc., assignee. Flexible Endoscopic Grasping and Cutting Device and Positioning Tool Assembly. United States patent US 20020103498 (2002).

Data

Perdiguero P, Venturas M, Cervera MT, Gil L, Collada C. Data from: Massive sequencing of Ulms minor's transcriptome provides new molecular tools for a genus under the constant threat of Dutch elm disease. Dryad Digital Repository. (2015) http://dx.doi.org/10.5061/dryad.ps837

Theses and dissertations

Smith, J. (2008) Post-structuralist discourse relative to phenomological pursuits in the deconstructivist arena. [dissertation/master's thesis]. [Chicago (IL)]: University of Chicago

Preprint

Smith, J. Title of the document. Preprint repository name [Preprint] (2008). Available at: https://persistent-url (Accessed March 15, 2018).