


Article

Cancer Metabolic Subtypes and Their Association with Molecular and Clinical Features

Enrico Moiso ^{1,2,*}  and Paolo Provero ^{3,4,*} 

¹ Department of Molecular Biotechnology and Health Sciences, University of Turin, 10126 Turin, Italy

² Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

³ Department of Neurosciences “Rita Levi Montalcini”, University of Turin, 10126 Turin, Italy

⁴ Center for Omics Sciences, San Raffaele Scientific Institute IRCSS, 20132 Milan, Italy

* Correspondence: emoiso@mit.edu (E.M.); paolo.provero@unito.it (P.P.)

† Current address: Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Simple Summary: The metabolic alterations characteristic of cancer cells play a significant role in tumors’ natural history and response to therapy. Recent technological advances have allowed the production of unprecedented amounts of data on many types of cancers. We exploited the most comprehensive collection of such data, The Cancer Genome Atlas (TCGA), to systematically investigate the associations between metabolic alterations and other tumor features. We used sets of genes known to be associated with specific metabolic pathways to classify patients into “metabolic subtypes”. Then, we systematically looked for associations between the metabolic subtypes and other tumor features, including histological classification, patient survival, and genome alterations. Our results, while correlative in nature, can provide a guide to the formulation of specific mechanistic hypotheses to be tested experimentally so as to improve our understanding of the biology of cancer and our ability to tailor therapeutic interventions to the specific features of each patient.

Abstract: The alterations of metabolic pathways in cancer have been investigated for many years, beginning long before the discovery of the role of oncogenes and tumor suppressors, and the last few years have witnessed renewed interest in this topic. Large-scale molecular and clinical data on tens of thousands of samples allow us to tackle the problem from a general point of view. Here, we show that transcriptomic profiles of tumors can be exploited to define metabolic cancer subtypes, which can be systematically investigated for associations with other molecular and clinical data. We find thousands of significant associations between metabolic subtypes and molecular features such as somatic mutations, structural variants, epigenetic modifications, protein abundance and activation, and with clinical/phenotypic data, including survival probability, tumor grade, and histological types, which we make available to the community in a dedicated web resource. Our work provides a methodological framework and a rich database of statistical associations, which will contribute to the understanding of the role of metabolic alterations in cancer and to the development of precision therapeutic strategies.

Keywords: cancer metabolism; genomics; computational biology



Citation: Moiso, E.; Provero, P. Cancer Metabolic Subtypes and Their Association with Molecular and Clinical Features. *Cancers* **2022**, *14*, 2145. <https://doi.org/10.3390/cancers14092145>

Academic Editor: J. Chad Brenner

Received: 11 February 2022

Accepted: 21 April 2022

Published: 25 April 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The last few years have witnessed a renewed interest in the metabolism of cancer, with the reprogramming of energy metabolism by cancer cells having been recently added, together with immune evasion, to the fundamental oncogenic processes collected under the heading of “Hallmarks of Cancer” [1].

Today, next-generation sequencing and other high-throughput techniques provide us with the unprecedented opportunity to take the study of the role of metabolic alterations in cancer progression to a new level, complementing the in-depth study of specific instances

of metabolic changes with unbiased assessments of cancer metabolism and its relationship with the other hallmarks. In particular The Cancer Genome Atlas (TCGA) provides us with a systematic repository of genomic, epigenomic, transcriptomic, and clinical data on tens of thousands of samples of many tumor types. For example, a recent study [2] demonstrated the close relationship between patterns of copy-number alterations and metabolic phenotypes. Another study [3] showed the clinical relevance of metabolic pathways, as reflected by gene expression profiles of metabolic genes, to patient survival.

In this work, we undertake a systematic analysis of the associations between metabolic profiles of human tumors and their molecular and clinical features. Our aim is to underline the role of metabolism in precision cancer medicine by showing that metabolic profiling can classify tumors into classes that are significantly different in both their molecular-level features (somatic mutations, structural genomic variants, epigenetic modifications) and in their clinical aspects (survival probability, tumor grade, histological type). While intrinsically correlative in nature, this analysis can serve as a guide to mechanistic studies, linking metabolism to other molecular features, and to the exploration of precision therapeutic approaches informed by tumor metabolism.

While, ideally, such a study should be based on direct assays of metabolite abundance in primary tumors, these are not yet available on a large scale. Therefore, we use transcriptomic profiles of tens of thousands of samples of many tumor types, as a proxy for metabolomic assays, to classify them into metabolic subtypes. We then systematically correlate such subtypes with molecular and clinical data using appropriate statistical tests, thus generating a database of thousands of associations.

2. Results

2.1. Transcriptome-Based Classification of Tumors into Metabolic Subtypes

For each tumor type with transcriptomic data Available online: the TCGA, and for each metabolic pathway defined as a set of genes by annotation databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) [4,5], Reactome [6], and the Gene Ontology [7,8], we divided the patients into metabolic subtypes by clustering the samples using the expression profiles of the genes belonging to the pathway. Supplementary Figure S1 shows the KEGG pathways used in the analysis and the genes assigned to each pathway as a bipartite network.

Compared with differential expression analyses, e.g., [3], clustering by metabolic pathways takes into account the fact that the genes associated with a pathway are often involved in either anabolic or catabolic processes. Therefore the activation of the pathway in a subset of patients is typically reflected by the transcriptional activation of the former and repression of the latter. For example, in Figure 1, we show the clustering of low-grade glioma samples based on the “arachidonic acid metabolism” KEGG pathway: each cluster is characterized by both up- and downregulated genes, a structure that would not be captured by simple differential expression of the pathway genes as a whole.

To perform the clustering we used partitioning around medoids (PAM) [9], a more robust method compared to k-means [10] with respect to the presence of outliers, which has been shown to be among the best-performing methods for the classification of cancer samples based on bulk RNA-seq data [11]. A total of 345 metabolic gene sets were used, as described in the Methods, on 38 tumor classes, for a total of 13,110 gene sets/tumor pairs. We used a Duda–Hart test [12] to select the pairs characterized by cluster separation (12,074 pairs, 92.1%). For these pairs, cluster silhouette analysis allowed an unsupervised choice of the number of clusters. In the majority of cases (9,157, 75.8%), the samples were subdivided into $k = 2$ clusters, while in other cases we obtained up to 10 clusters. These clusters will be referred to as “metabolic subtypes”. The tumor types analyzed with the number of samples for which expression data were available are shown in Supplementary Table S1.

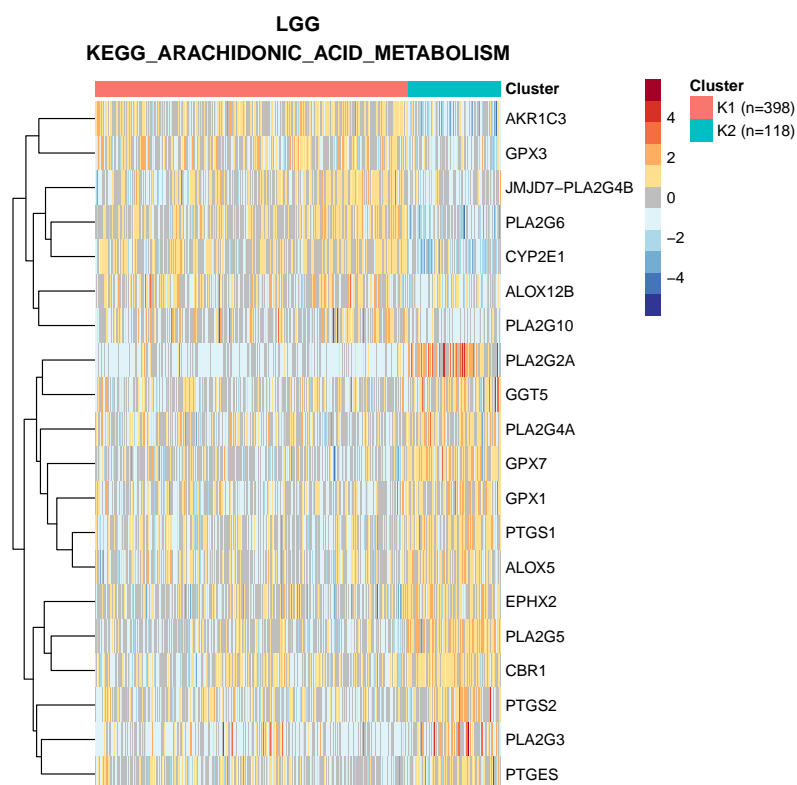


Figure 1. Clustering of low-grade glioma patients using the genes involved in arachidonic metabolism according to KEGG. Only the 20 genes most differentially expressed between the two clusters are shown.

It is worth asking whether different metabolic gene sets cluster the patients of the same tumor in significantly different ways. To answer this question, we computed the normalized mutual information between the cluster assignments obtained in each tumor with each metabolic gene set. The results, for the KEGG pathways analyzed are shown in Figure 2, and suggest that, in most cases, different metabolic gene sets cluster tumor patients in clearly distinct ways.

2.2. Associations between Metabolic Subtypes and Molecular and Clinical Features

We then proceeded to systematically explore the statistical associations between metabolic subtypes and several molecular and clinical features of the corresponding tumors, classified into 9 classes of variables (see Supplementary Table S2): 6 classes of molecular variables (copy number alterations, global DNA methylation, microRNA expression, specific point mutations, gene-level mutations, and protein expression); and 3 classes of phenotypic ones (clinical and histological parameters appropriate for each tumor type, overall survival, and recurrence-free survival). Statistical tests appropriate to the nature of the variables describing such features were chosen as described in the Methods. Bonferroni correction was applied separately to each variable class but to all tumor types together. Thus, for example, the Bonferroni correction for the association with miRNA expression takes into account all tests of association between all miRNAs and all metabolic gene sets across all tumor types.

A total of $\sim 4 \times 10^8$ statistical tests were performed, resulting in 878,655 significant associations after Bonferroni correction (0.21% of the tests performed). An overall Bonferroni correction (taking into account together all the tests made for all variable types) would have given 663,100 significant associations. In the rest of this section, to limit redundancy, we focus on the results obtained with the KEGG annotation database (a total of 99,462 significant associations). Figure 3A shows the distribution of these associations among tumor types. Note that the number of significant associations for a given tumor type depends on

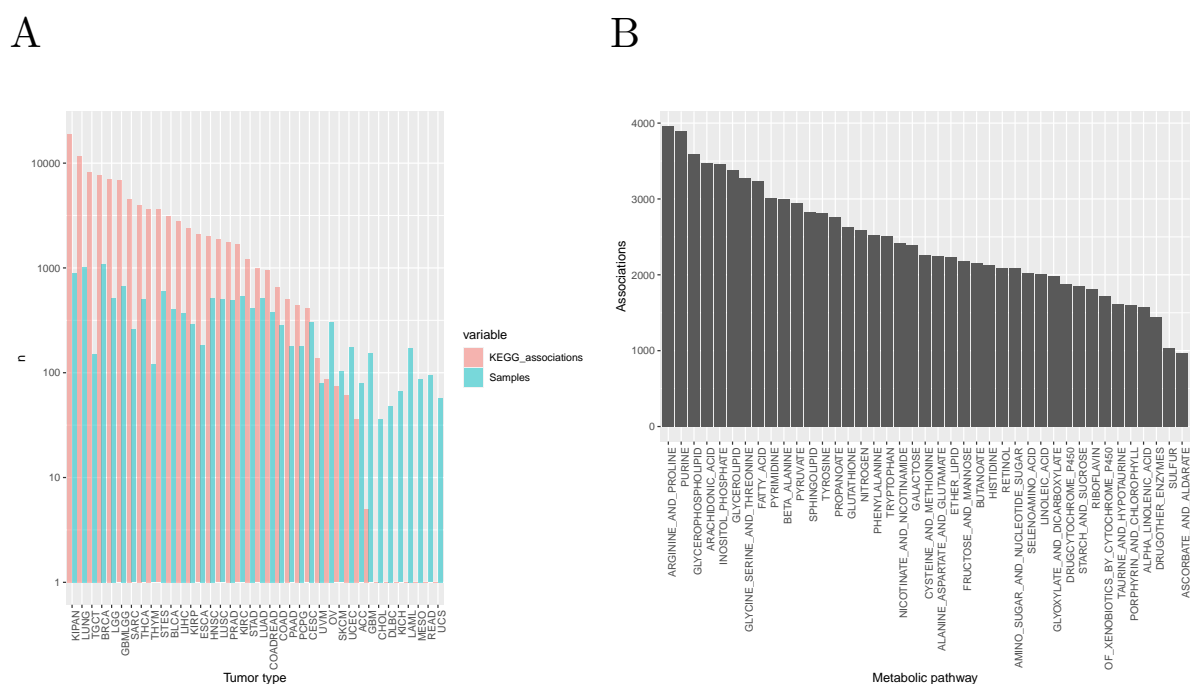


Figure 3. (A) Number of samples and of significant associations with molecular and clinical features by tumor type. **(B)** Number of significant associations by KEGG metabolic pathway.

2.3. Recurrent Associations

Of particular interest are recurrent associations, i.e., those that are statistically significant in more than one tumor type (not considering the tumor types built from the union of smaller types, such as LUNG, KIPAN, etc.). We found a total of 14,899 such recurrent associations. Supplementary Table S3 shows the most recurrent associations for each class of clinical and molecular variables. Notably, the only variable classes with no recurrent associations are point mutations: all 45 associations with specific mutations and 102 associations with gene-level mutations are specific of a single tumor type.

The most recurrent association was found between the expression of microRNA hsa-miR-222 and the metabolic subtypes based on the “inositol phosphate metabolism” gene set defined by KEGG, which recurred in 12 tumor types. Figure 4 shows the expression of this microRNA in the clusters of breast cancer and lung adenocarcinoma patients obtained with such a gene set.

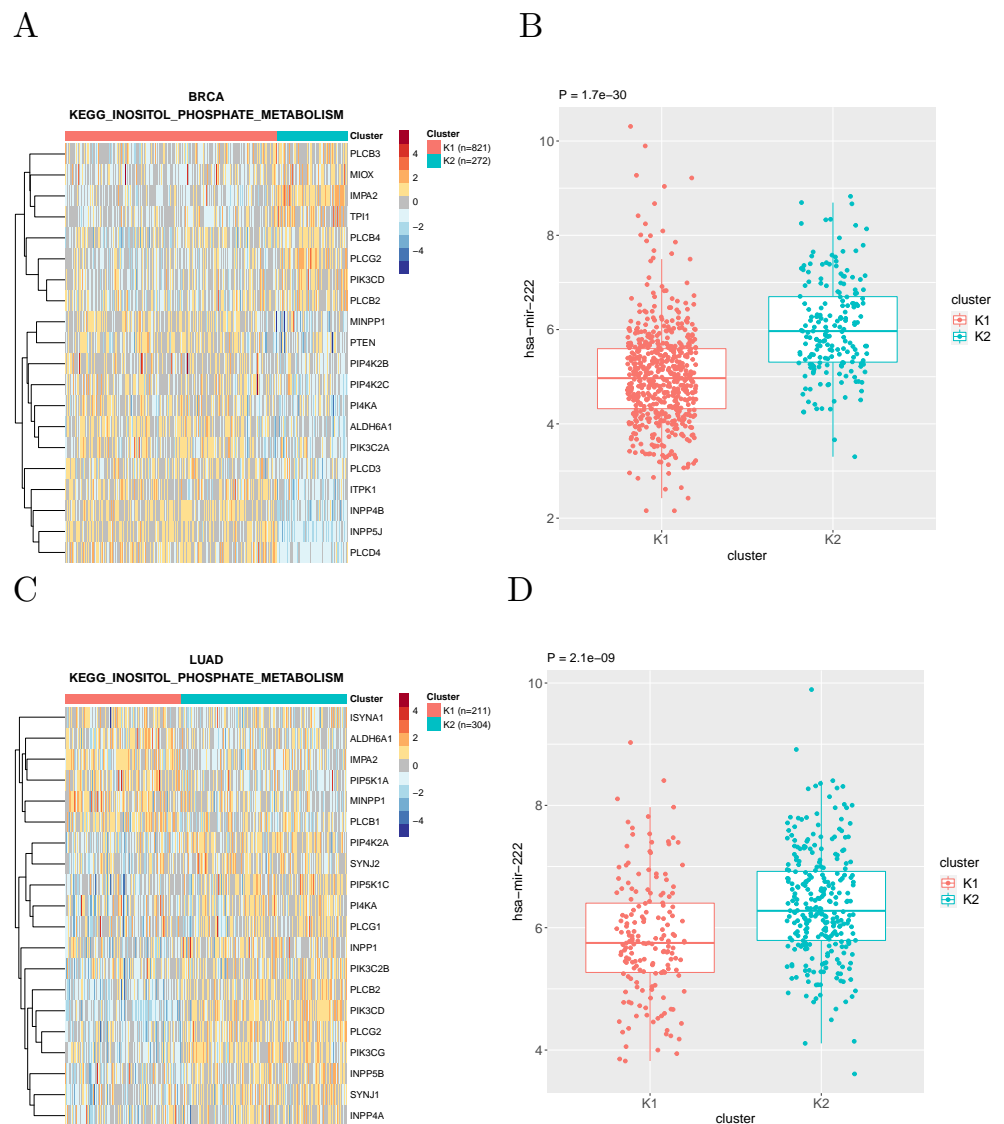


Figure 4. (A) Clustering of breast cancer samples using the genes associated to inositol phosphate metabolism according to KEGG. (B) Expression of miR-222 in the two clusters. (C,D) Same for lung adenocarcinoma samples.

2.4. Specific Examples

In the following, we show a few examples of associations, some confirming known results and some suggesting new lines of investigation, especially in less widely studied tumor types.

2.4.1. TP53 Mutations and Glucose Metabolism

Mutant TP53 has been shown to promote the Warburg effect, possibly the most notorious metabolic hallmark of cancer, in breast and lung cancer cell lines [13]. Indeed, clustering breast and lung tumors using the genes associated with glucose metabolism according to REACTOME shows, in both cancer types, the appearance of a cluster characterized by overexpression of several Warburg effect signature genes (such as GAPDH, PGK1, and PKM2) and by an enrichment in TP53 mutations (Figure 5).

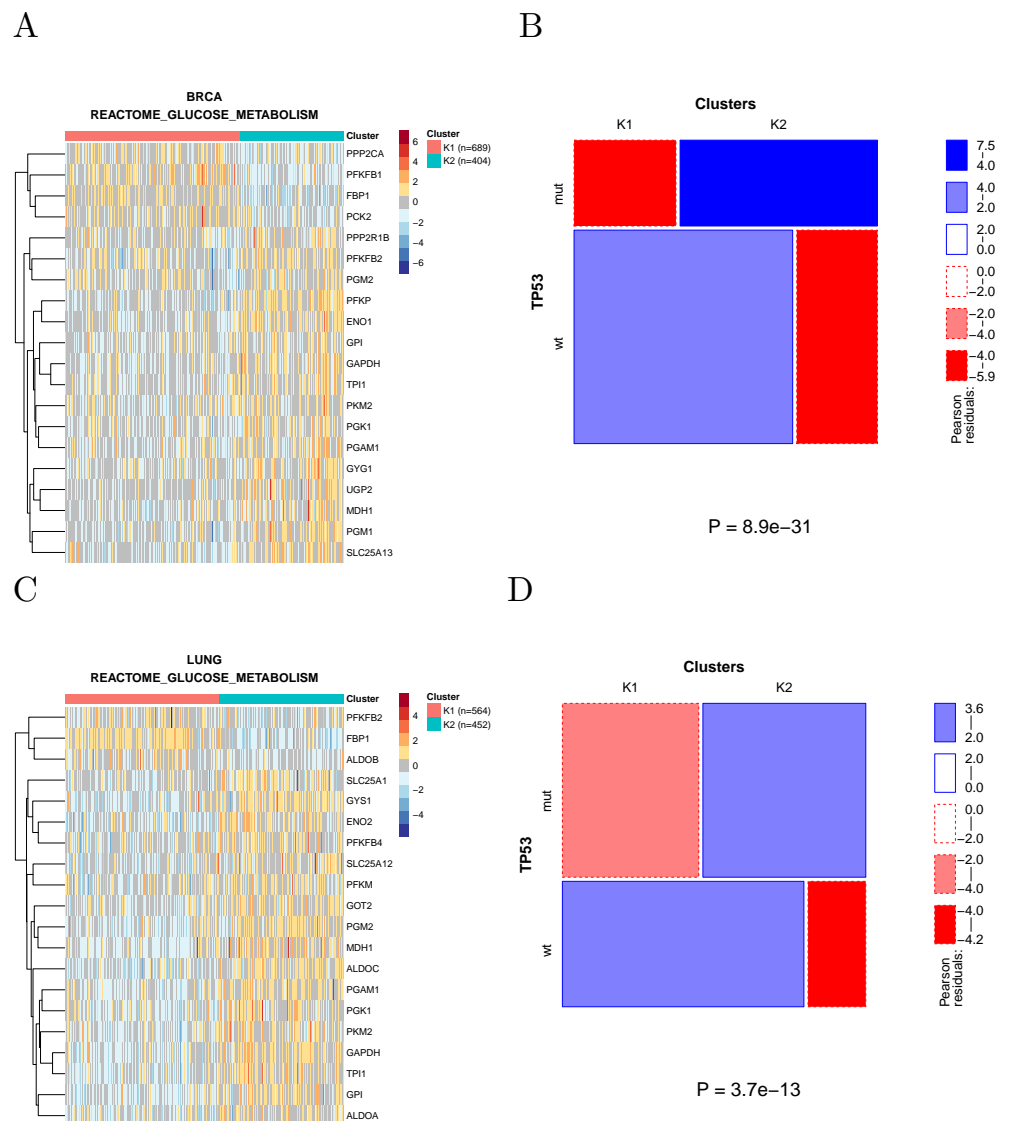


Figure 5. (A) Clustering of breast cancer samples using the genes associated with glucose metabolism according to REACTOME. Cluster K2 shows upregulation of Warburg effect genes, such as GAPDH, PGK1, and PKM2. (B) The prevalence of TP53-mutated samples is significantly higher in cluster K2. (C,D) Same for lung cancer samples.

2.4.2. Pyruvate Metabolism in Thyroid Cancer

The clustering of thyroid cancer samples according to the genes associated with KEGG to pyruvate metabolism is shown in Figure 6A. These two clusters show many significant differences in both clinical and molecular parameters. In particular (Figure 6B,C), the prevalence of BRAF and NRAS-mutated samples was strikingly different in the two clusters, with all NRAS-mutated samples found in Cluster 2, and the large majority of BRAF-mutated samples found in Cluster 1. The two clusters also strongly differed in the respective prevalence of histological types.

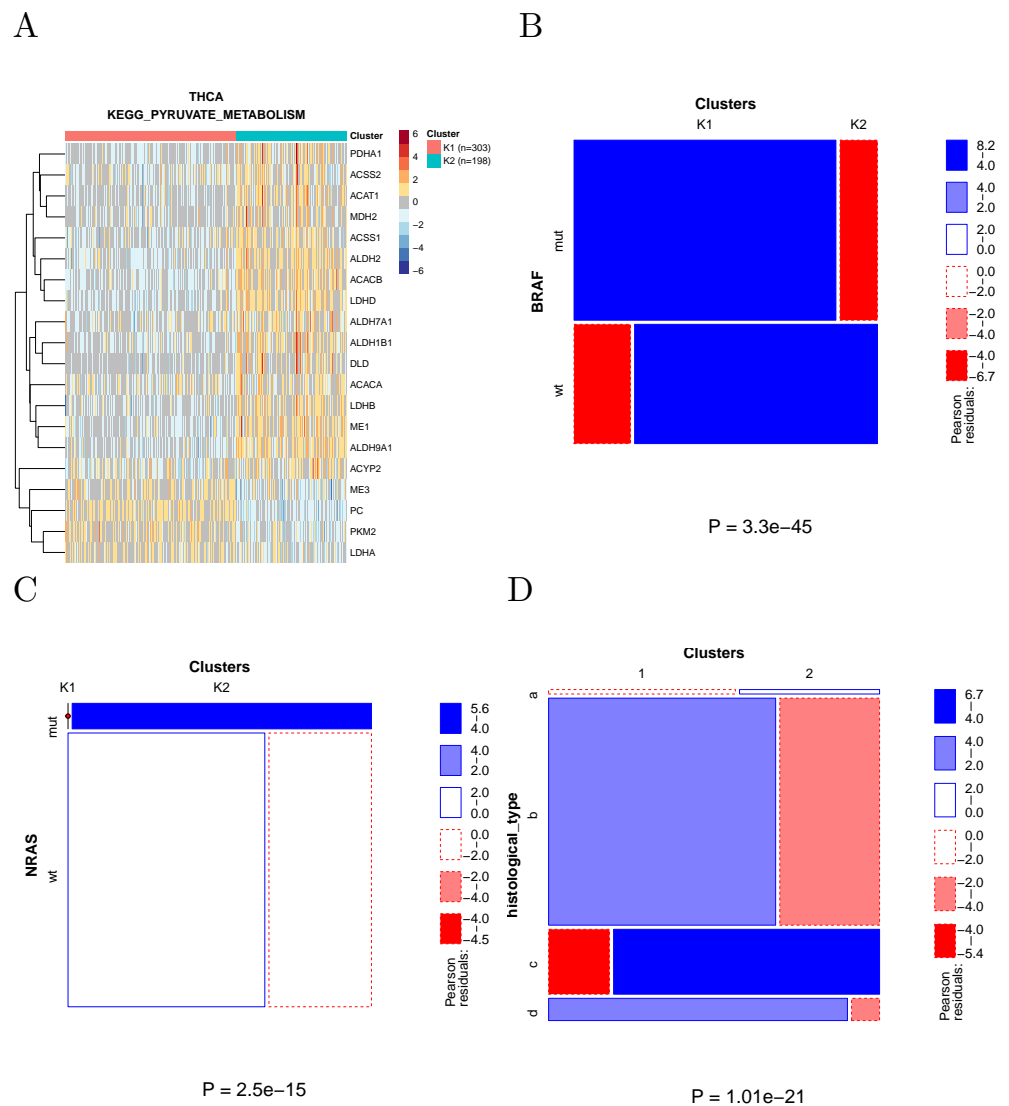


Figure 6. (A) Clustering of thyroid cancer samples using the genes associated with KEGG to pyruvate metabolism. (B) Most BRAF-mutated samples fell into Cluster 1. (C) All NRAS-mutated samples fell into Cluster 2. (D) The two clusters significantly differed in the prevalence of histological types; (a): other type; (b): thyroid papillary carcinoma - classical/usual; (c): thyroid papillary carcinoma - follicular ($\geq 99\%$ follicular patterned); (d): thyroid papillary carcinoma - tall cell ($\geq 50\%$ tall cell features)

2.4.3. Fatty Acid Metabolism in Thymoma

Clustering of thymoma samples using the genes associated with fatty acid metabolism divides the samples into two clusters (Figure 7A) largely overlapping the known histological types (Figure 7B) and differing in the prevalence of a recurrent GTF2I mutation [14].

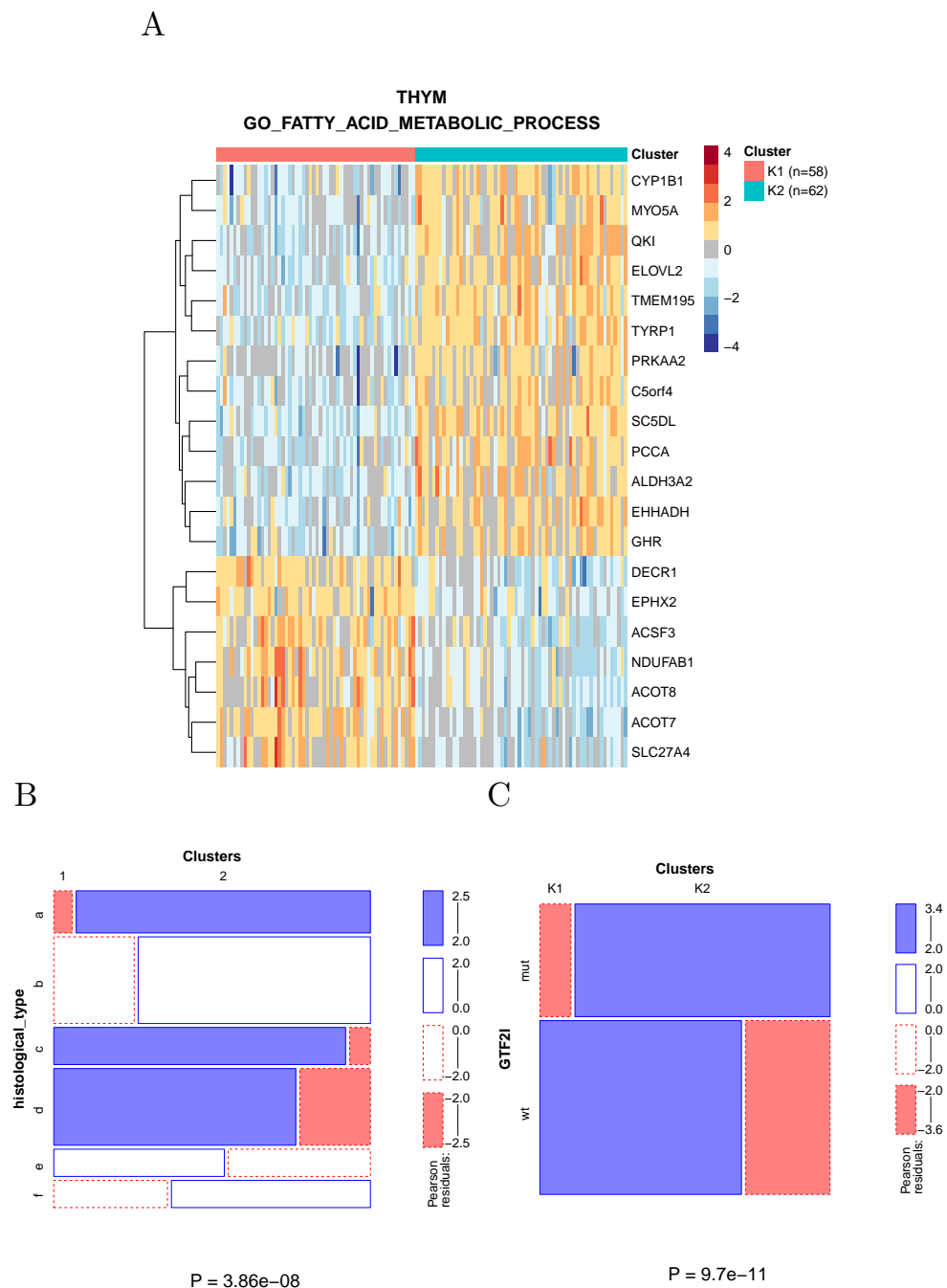


Figure 7. (A) Clustering of thymoma samples using the genes associated by KEGG to fatty acid metabolism. (B) The clusters significantly overlap the known histological types; (a): thymoma, type a; (b): thymoma, type ab; (c): thymoma, type b1; (d): thymoma, type b2; (e): thymoma, type b3; (f): thymoma, type c. (C) GTF2I mutations significantly segregate between the two clusters.

2.5. Analysis of Cancer Cell Lines

We performed the same analysis on the cancer cell lines included in the cancer cell line encyclopedia (CCLE) [15]. These cell lines are classified in terms of TCGA tumor types (22 types are represented by a total of 723 CCLE cell lines). Within each tumor type, cell lines were clustered based on their transcriptomic data, using the expression of the genes in the same metabolic gene sets used to cluster TCGA samples. The clusters were then

correlated with molecular and phenotypic data, including copy number alterations, specific point mutations, gene-level mutations, protein expression, microRNA expression, and drug response data (IC50 values).

Since the number of cell lines in the CCLE is about one order of magnitude smaller than the number of TCGA samples, the power to detect significant associations is also much smaller. Nevertheless, we found 460 significant associations at a Bonferroni-corrected significance level of 0.05. For example, KEAP1 mutations in lung adenocarcinoma are associated with xenobiotic metabolism in both TCGA samples and CCLE cell lines (Figure 8), in agreement with the known role of the NRF2/KEAP1 pathway in the regulation of xenobiotic response [16]. Note that KEAP1 mutations appear to be associated with the activation of the pathway, as observed in [17].

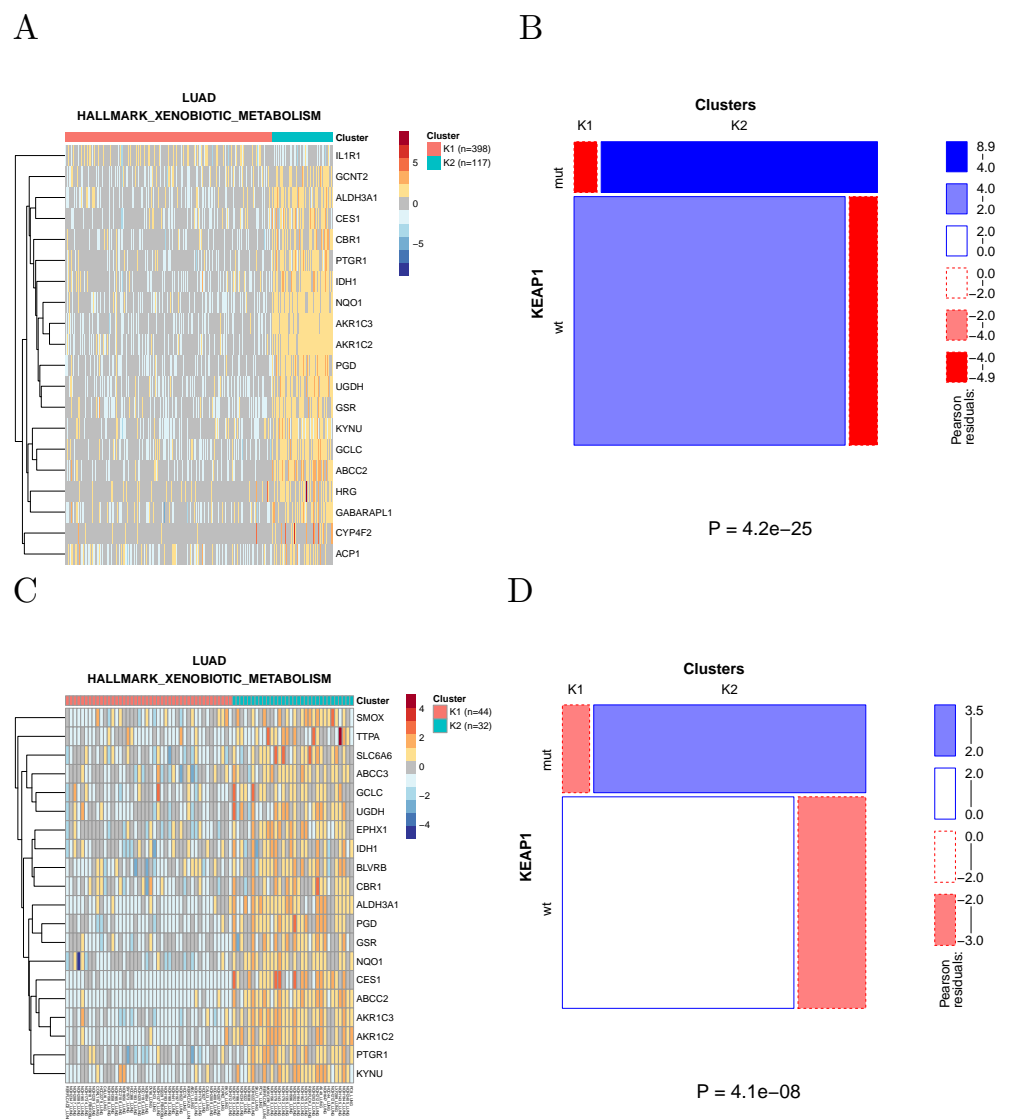


Figure 8. (A) Clustering of lung adenocarcinoma samples using the genes in the xenobiotic metabolism hallmark gene set. (B) Mutations of KEAP1 are strongly enriched in the K2 cluster where the pathway appears to be activated. (C) Clustering of lung adenocarcinoma cell lines using the same gene set. (D) Also in cell lines mutations are enriched in the cluster showing pathway activation.

3. Discussion

In this work, we systematically analyzed cancer metabolic subtypes, defined by clustering the patient samples using the transcription profiles of metabolic gene sets. These subtypes are highly specific for each individual metabolism used to define the gene sets, and show a myriad of significant associations with both molecular and phenotypic features of the tumors.

We present these associations as a resource for the community, Available online: <https://metaminer.unito.it> (accessed April 22, 2022) which, we believe, will be useful in two related but distinct ways. On the one hand, our results highlight the key role of metabolic profiles in classifying individual patients in a way that is potentially useful to devise precision treatments. This is particularly compelling for tumors for which no established classification exists, but in which metabolic subtypes strongly associated with phenotypic characteristics can be found.

On the other hand, our results can help formulate hypotheses on the mechanisms underlying the associations between metabolic subtypes on one hand, and molecular and phenotypic features on the other. These can then be experimentally tested in cell lines or organoids. For example, while both miR-222 (see, e.g., ref. [18] for a recent review) and inositol phosphate metabolism [19,20] have been often associated with cancer progression, their association, which recurs in a large number of tumor types, to the best of our knowledge, has not been previously reported, and deserves a mechanistic investigation.

Some limitations of our study should be pointed out. The strongest is the correlative nature of our results, which would require experimental analyses to be interpreted in terms of causality; this limitation is shared by all retrospective analyses of primary tumors. Moreover, the results pertaining to some specific tumor types need to be interpreted with caution. In particular, tumor types obtained by the combination of more specific types (such as GBMLGG, LUNG, etc.) might generate many statistically significant results of limited biological interest: for example, since GBM and LGG strongly differ in survival probability, any gene set able to distinguish through clustering between these two types of tumors will be associated with survival. Moreover, some tumor types are characterized by mutations leading to widespread metabolic reprogramming (such as IDH1 mutations in glioma [21] and inactivation of the von Hippel–Lindau tumor suppressor in kidney cancer [22]). Therefore, in these tumors, many or most metabolic gene sets are associated with clinical or molecular features. Such associations, while statistically correct, might not lead to a manageable number of specific, testable mechanistic hypotheses. Finally, while k-medoid clustering has been shown to perform well in the classification of cancer samples based on bulk RNA-seq data [11], we did not explore how other clustering strategies would perform in detecting associations between metabolic subtypes and molecular/clinical features.

4. Methods

4.1. Data

4.1.1. TCGA

All TCGA data (release 28 February 2016) were obtained from the Broad TCGA GDAC site (<https://gdac.broadinstitute.org/>, accessed on January 27, 2021), by means of `firehose_get`, version: 0.4.1. The gene expression data clustered to generate the metabolic subtypes were the normalized gene-level transcript per million (TPM) values obtained with the RNA-Seq by the Expectation-Maximization (RSEM) method.

Copy number data were obtained from the segmented data using the `cghMCR` [23], `DNAcopy` [24], and `CNTools` [25] Bioconductor packages. The `cghMCR` package allows the calculation of segment gain or loss (SGOL) from segmented data, by means of a modified version of the Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm. The `segment` function of the `DNAcopy` package was used to segment the normalized data so that chromosome regions with the same copy number had the same segment mean values. Then, using the `getRS` function from `CNTools`, the data returned by `segment` were organized in a matrix format. Finally, the `SGOL` function of `cghMCR` was used to compute gene-level

SGOL scores by calculating the sum (parameter method) of all the positive and negative values, respectively, above and below a set threshold (0.5, −0.5). Finally, SGOL scores for all genes included in each Broad Institute positional gene set were averaged to obtain the SGOL score of each region.

Genome-wide DNA methylation was obtained from the gene-level summarized Human Methylome 450 k data, as provided by GDAC, then summed over all genes to associate a global DNA methylation value to each sample. miRNA expression was obtained from the Illumina HiSeq data, as deposited in GDAC. Specific point mutations for each sample were obtained from the “mutation packager oncotated calls” provided by GDAC, considering only non-synonymous mutations. For gene-level mutations, a sample was considered as mutated in a gene if it carried one or more specific point mutations associated with the gene.

Protein expression data produced by Reverse-Phase Protein microArrays (RPPA) were obtained from the corresponding files annotated with gene names from the GDAC site. Survival and other clinical data were obtained from the “Tier 1 clinical pick” files Available online: GDAC.

4.1.2. CCLE

Gene expression data of 723 cell lines associated with one of the TCGA tumor types were obtained, as gene-level TPMs computed with RSEM, from the CCLE web site. We also retrieved from the CCLE web site the following data to be correlated with metabolic subtypes:

- Gene-level copy number data: the absolute score provided by CCLE was processed with the same steps used for the SGOL score reported for the TCGA data to obtain region-based copy number data;
- miRNA expression data (which were log-transformed for display in the figures);
- Somatic mutation data (here we considered only non-synonymous mutations);
- IC50 values for the available drugs;
- Metabolite levels.

4.1.3. Metabolic Gene Sets

Gene sets were obtained from the c2.KEGG, c2.REACTOME, c5.BP, and Hallmark MSigDB v5.2 collections. Metabolic gene sets were defined as those whose name matched the string “metabol” but not the string “regul” (to include only gene sets directly involved in metabolic pathways rather than in their regulation). In this way, we obtained 345 metabolic gene sets (41, 35, 265, and 4 for the c2.KEGG, c2.REACTOME, c5.BP, and Hallmark collections, respectively).

4.2. Generation of Metabolic Clusters

For each tumor type and each metabolic gene set, a Duda–Hart test [12] was performed to detect cluster separation. This was conducted on both log-transformed and rank-transformed expression values. Since the latter method turned out to be more conservative, it was used in the following. For the gene set/tumor pairs for which the Duda–Hart test revealed the existence of a cluster structure ($p < 0.001$), the patients were divided into metabolic subtypes using partition around medoid (PAM) clustering [9] on the Spearman rank correlation coefficient-based distance matrix obtained from the expression of the genes in the gene set. The number of clusters (the maximum being set at 10) generated by each metabolic gene set was based on optimizing the average silhouette width. The normalized mutual information between two clusterings of the same patients based on different gene sets was computed with the R package *infotheo* [26], using the entropy of the empirical probability distribution and normalizing to the geometric mean of the entropies. Genes differentially expressed among clusters were identified using a Kruskal–Wallis test. The top 20 genes by *p-value* are shown in the heatmaps.

4.3. Statistical Tests of Association

The statistical tests used for establishing the association between each clinical/molecular feature and metabolic subtypes depend on the type of variable describing the feature. The metabolic subtype is always a categorical variable indicating the cluster membership of each patient.

- For continuous variables (microRNA expression, DNA methylation, CNA SGOL scores) we used Mann–Whitney U test (Kruskal–Wallis for $k > 2$). In the case of microRNA expression, given the large number of microRNAs with expression equal to zero in each sample, the expression data were rank-transformed with random resolution of ties before the statistical test.
- For categorical variables (e.g., presence of a given mutation) we used Fisher’s exact test (χ^2 test for tables larger than 2×2 , removing levels for which the expected count was less than 5).
- For survival (RFS or OS), we used the log-rank test. The test was not considered valid when the expected number of events in any metabolic subtype was less than 5.

Multiple Testing

Bonferroni correction was applied separately to each variable type and to all tumor types together. Thus, for example, the Bonferroni correction for the association with miRNA expression takes into account all tests of association between all miRNAs and all metabolic gene sets across all tumor types. We chose to perform the correction separately for the feature types because these contain very different numbers of variables (hundreds of variables for miRNA expression and a single one for overall survival). Therefore, an overall Bonferroni correction would unduly penalize the variable types containing few variables.

5. Conclusions

The systematic analysis of the associations between transcriptomic-based metabolic subtypes and other clinical and molecular tumor features can serve as a guide to the formulation of mechanistic hypotheses for experimental validation, with the potential to improve our understanding of the role of metabolic alterations in cancer progression and to exploit metabolism-based stratification to develop new precision medicine strategies.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/cancers14092145/s1>. Figure S1: The KEGG pathways (red squares) used to cluster tumor samples and the respective genes (grey triangles) represented as a bipartite network. The size of the squares is proportional to the number of associated genes. Table S1: Number of samples used for each tumor type. Table S2: Molecular and phenotypic variables analyzed for association with metabolic subtypes. Table S3: Top recurrent associations for each class of variables. Data S1: Clustering of TCGA patients into metabolic subtypes. NA indicates no cluster separation according to the Duda–Hart test. Data S2: Associations between metabolic subtypes and molecular/phenotypic features in TCGA. Data S3: Clustering of CCLE cell lines into metabolic subtypes. Data S4: Associations between metabolic subtypes and molecular/phenotypic features in CCLE.

Author Contributions: Conceptualization, E.M. and P.P.; methodology, E.M. and P.P.; software, E.M.; formal analysis, E.M. and P.P.; investigation, E.M. and P.P.; writing—original draft preparation, P.P.; writing—review and editing, E.M. and P.P.; visualization, E.M. and P.P.; supervision, P.P.; project administration, P.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: This study relied exclusively on publicly available anonymized data and thus did not require informed consent.

Data Availability Statement: The results of the analysis are included as Supplementary Material and can be browsed interactively at <https://metaminer.unito.it/>, accessed on 22 April 2022.

Acknowledgments: The results of this analysis are in whole or part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>, accessed on 27 January 2021. We would like to thank Davide Cittaro, Chiara Riganti, and Giovanni Tonon for insightful comments and discussions, and Alessandro Greganti, Ivan Molineris, and Sergio Rabellino for IT support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674, doi:10.1016/j.cell.2011.02.013.
2. Graham, N.A.; Minasyan, A.; Lomova, A.; Cass, A.; Balanis, N.G.; Friedman, M.; Chan, S.; Zhao, S.; Delgado, A.; Go, J.; et al. Recurrent patterns of DNA copy number alterations in tumors reflect metabolic selection pressures. *Mol. Syst. Biol.* **2017**, *13*, 914, doi:10.15252/msb.20167159.
3. Gaude, E.; Frezza, C. Tissue-specific and convergent metabolic transformation of cancer correlates with metastatic potential and patient survival. *Nat. Commun.* **2016**, *7*, 13041, doi:10.1038/ncomms13041.
4. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30, doi:10.1093/nar/28.1.27.
5. Kanehisa, M.; Sato, Y.; Furumichi, M.; Morishima, K.; Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **2019**, *47*, D590–D595, doi:10.1093/nar/gky962.
6. Fabregat, A.; Jupe, S.; Matthews, L.; Sidiropoulos, K.; Gillespie, M.; Garapati, P.; Haw, R.; Jassal, B.; Korninger, F.; May, B.; et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **2018**, *46*, D649–D655, doi:10.1093/nar/gkx1132.
7. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29, doi:10.1038/75556.
8. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338, doi:10.1093/nar/gky1055.
9. Kaufman, L.; Rousseeuw, P. Clustering by means of Medoids. In *Statistical Data Analysis Based on the L1 Norm and Related Methods*; Dodge, Y., Ed.; North-Holland, The Netherlands, 1987; pp. 405–416.
10. Arbin, N.; Suhaimi, N.S.; Mokhtar, N.Z.; Othman, Z. Comparative Analysis between K-Means and K-Medoids for Statistical Clustering. In *Proceedings of the 2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)*, Sabah, Malaysia, 2–4 December 2015; pp. 117–121, doi:10.1109/AIMS.2015.82.
11. Jaskowiak, P.A.; Costa, I.G.; Campello, R.J.G.B. Clustering of RNA-Seq samples: Comparison study on cancer data. *Methods* **2018**, *132*, 42–49, doi:10.1016/j.ymeth.2017.07.023.
12. Duda, R.; Hart, P. *Pattern Classification and Scene Analysis*; Wiley: New York, NY, USA, 1973.
13. Zhang, C.; Liu, J.; Liang, Y.; Wu, R.; Zhao, Y.; Hong, X.; Lin, M.; Yu, H.; Liu, L.; Levine, A.J.; et al. Tumour-associated mutant p53 drives the Warburg effect. *Nat. Commun.* **2013**, *4*, 2935, doi:10.1038/ncomms3935.
14. Petrini, I.; Meltzer, P.S.; Kim, I.K.; Lucchi, M.; Park, K.S.; Fontanini, G.; Gao, J.; Zucali, P.A.; Calabrese, F.; Favaretto, A.; et al. A specific missense mutation in GTF2I occurs at high frequency in thymic epithelial tumors. *Nat. Genet.* **2014**, *46*, 844–849, doi:10.1038/ng.3016.
15. Ghandi, M.; Huang, F.W.; Jané-Valbuena, J.; Kryukov, G.V.; Lo, C.C.; McDonald, E.R.; Barretina, J.; Gelfand, E.T.; Bielski, C.M.; Li, H.; et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **2019**, *569*, 503–508, doi:10.1038/s41586-019-1186-3.
16. Motohashi, H.; Yamamoto, M. Nrf2-Keap1 defines a physiologically important stress response mechanism. *Trends Mol. Med.* **2004**, *10*, 549–557, doi:10.1016/j.molmed.2004.09.003.
17. Gong, M.; Li, Y.; Ye, X.; Zhang, L.; Wang, Z.; Xu, X.; Shen, Y.; Zheng, C. Loss-of-function mutations in KEAP1 drive lung cancer progression via KEAP1/NRF2 pathway activation. *Cell Commun. Signal.* **2020**, *18*, 98, doi:10.1186/s12964-020-00568-z.
18. Ravegnini, G.; Cargnin, S.; Sammarini, G.; Zanolli, F.; Bermejo, J.L.; Hrelia, P.; Terrazzino, S.; Angelini, S. Prognostic Role of miR-221 and miR-222 Expression in Cancer Patients: A Systematic Review and Meta-Analysis. *Cancers* **2019**, *11*, 970, doi:10.3390/cancers11070970.
19. Hakim, S.; Bertucci, M.C.; Conduit, S.E.; Vuong, D.L.; Mitchell, C.A. Inositol polyphosphate phosphatases in human disease. *Curr. Top. Microbiol. Immunol.* **2012**, *362*, 247–314, doi:10.1007/978-94-007-5025-8_12.
20. Benjamin, D.I.; Louie, S.M.; Mulvihill, M.M.; Kohnz, R.A.; Li, D.S.; Chan, L.G.; Sorrentino, A.; Bandyopadhyay, S.; Cozzo, A.; Ohiri, A.; et al. Inositol phosphate recycling regulates glycolytic and lipid metabolism that drives cancer aggressiveness. *ACS Chem. Biol.* **2014**, *9*, 1340–1350, doi:10.1021/cb5001907.
21. Pirozzi, C.J.; Yan, H. The implications of IDH mutations for cancer development and therapy. *Nat. Rev. Clin. Oncol.* **2021**, *18*, 645–661, doi:10.1038/s41571-021-00521-0.
22. Weiss, R.H. Metabolomics and Metabolic Reprogramming in Kidney Cancer. *Semin Nephrol.* **2018**, *38*, 175–182, doi:10.1016/j.semnephrol.2018.01.006.
23. Zhang, J.; Feng, B. *cghMCR: Find Chromosome Regions Showing Common Gains/Losses*; 2020. Available online: <https://bioconductor.org/packages/release/bioc/html/cghMCR.html> (accessed on 15 February 2021)
24. Seshan, V.E.; Olshen, A. *DNAcopy: DNA Copy Number Data Analysis*; 2020. Available online: <https://bioconductor.org/packages/release/bioc/html/DNAcopy.html> (accessed on 15 February 2021)

-
25. Zhang, J. *CNTools: Convert Segment Data into a Region by Sample Matrix to Allow for Other High Level Computational Analyses*; 2020. Available online: <https://bioconductor.org/packages/release/bioc/html/CNTools.html> (accessed on 15 February 2021)
 26. Meyer, P.E. *Infotheo: Information-Theoretic Measures*; 2014.