

**UNIVERSITY OF TURIN**

**Phd School in life and Health Sciences**  
*Molecular Medicine*



*Metabolic Subtypes of Human Tumors with Distinct Genomic and  
Clinical Features*

**Enrico Moiso**

**UNIVERSITY OF TURIN**

Phd School in life and Health Sciences  
*Molecular Medicine*

*XXIX Cycle*  
*Academic Years: 2014-2017*



***Metabolic Subtypes of Human Tumors with Distinct Genomic and Clinical Features***

Tutor: Prof. Paolo Provero  
Co-tutor: Prof. Mara Brancaccio  
Coordinator: Prof. Francesco Novelli

Enrico Moiso



*To my family, to my friends  
and to those who believed in it*

# Contents

<b>1</b>	<b>Abstract</b>	<b>4</b>
<b>2</b>	<b>Introduction</b>	<b>6</b>
2.1	Normal And Cancer Cell Metabolism At A Glance . . . . .	7
<b>3</b>	<b>Experimental Approach</b>	<b>10</b>
<b>4</b>	<b>Materials And Methods</b>	<b>12</b>
4.1	Cancer Samples Data . . . . .	13
4.2	CNA SGOL Scores . . . . .	14
4.3	Gene Sets . . . . .	14
4.4	Methylation Data . . . . .	15
4.5	Differential Gene Expression . . . . .	15
4.6	Gene Set Enrichment Analysis . . . . .	15
4.7	Metabolic Clusters Generation . . . . .	15
4.8	Correlation Between Metabolic Clusters And Clinical-Genomic Variables . . . . .	16
4.9	Pvalues And Multiple Tests Correction . . . . .	18
4.10	R And RSTUDIO Software . . . . .	19
<b>5</b>	<b>Results</b>	<b>20</b>
5.1	General Results . . . . .	21
5.1.1	Clinical Relevance Of Metabolic Clusters . . . . .	21
5.1.2	Genomic Relevance Of Metabolic Clusters . . . . .	26
5.1.2.1	Mutation Data . . . . .	26
5.1.2.2	Copy Number Alteration Data . . . . .	27
5.1.2.3	Methylation Data . . . . .	31
5.1.2.3.1	All Genes Methylation Level . . . . .	31
5.1.2.3.2	Gene Specific Methylation Level . . . . .	34
5.1.2.4	RPPA Data . . . . .	35
5.1.2.5	miRNA Data . . . . .	38

5.2	Context Specific Results . . . . .	41
5.2.1	Unsaturated Fatty Acids Metabolism In LGG . . . . .	41
5.2.2	Hexose Metabolism in BRCA . . . . .	55
<b>6</b>	<b>Discussion</b>	<b>68</b>
<b>7</b>	<b>Supplementary Figures And Tables</b>	<b>70</b>
7.1	Clinical Data . . . . .	71
7.1.1	Overall Survival . . . . .	71
7.1.2	Relapse Free Survival . . . . .	75
7.1.3	Clinical Data Elements . . . . .	79
7.2	Genomic Data . . . . .	86
7.2.1	Mutation Data . . . . .	87
7.2.1.1	Protein Coding Gene Level Mutations . . . . .	87
7.2.1.2	Protein Level Mutations . . . . .	94
7.2.1.3	Genomic Specific Nucleotide Variations . . . . .	101
7.2.2	Copy Number Alteration Data . . . . .	108
7.2.3	Methylation Data . . . . .	115
7.2.3.1	All Genes Methylation Level . . . . .	115
7.2.3.2	Gene Specific Methylation Level . . . . .	119
7.2.4	Reverse Phase Protein Array Data . . . . .	126
7.2.5	miRNA Expression Data . . . . .	133
<b>8</b>	<b>Publications List</b>	<b>140</b>
<b>9</b>	<b>References</b>	<b>142</b>

# 1 Abstract

Metabolic studies provided and provide, the most complete qualitative and quantitative picture of specific cellular pathways, so far. Before the advent of genomic high-throughput studies, which provided a terrific advancement in the study of gene regulation and fostered the discovery and comprehension of transcriptional networks, metabolic pathways were already text-book knowledge, described at the stoichiometric level.

In order to obtain a better understanding on the role of cellular metabolism in cancer progression, we focused our attention on TCGA cohort, which represents one of the widest repository of genomic and clinical data, publicly available.

Unsupervised classification of patients based on their metabolic gene expression profile creates the so called “metabolic clusters”. Expression levels of genes involved in metabolic pathways have been used as “proxy” for the metabolic status of the tumors due to the lack of data derived from large comparative metabolomic studies of different tumor types.

Afterwards these clusters were correlated to clinical and genomic data. As previously reported, a potent association exists between metabolic genes and prognosis. Interestingly we found that different tumours show a different correlation strength between metabolism and prognosis, as well as that some metabolisms are more prone to be associated to prognosis than other.

We then focused our attention on the molecular characterization of the clusters. Notably, specific metabolisms correlate with CNA, methylation levels, molecular classification (e.g. PAM50, St. Gallen), RPPA data, miRNAs expression levels and mutation data.

This study represents the widest, so far, quantification of the associations between metabolism and both clinical and molecular features, in more than 11000 samples derived from more than 30 different human tumor types. The results obtained could be relevant for basic research, i.e. the identification of novel mechanisms linking metabolic rewiring to specific genomic lesions, and possibly also for clinical practice, allowing the identification of both novel oncometabolites as well as new therapeutic targets.

## 2 Introduction

## 2.1 Normal And Cancer Cell Metabolism At A Glance

*“Both the body and its parts are in a continuous state of dissolution and nourishment, so they are inevitably undergoing permanent change”* Ibn al-Nafis, The Treatise of Kamil on the Prophet’s Biography (XIII century)

Metabolism (*μεταβολη*, *metabole*, “change”) is the term, in cell biology, used to describe the set of chemical reactions occurring in living systems. The classification of those reactions, based on the biochemical nature of their products, gives rise to the distinction between catabolism (*κατω*-*kato*, “downward” and *βαλλειν*-*ballein*, “to throw”); the processes which extracts energy and small compounds from higher level compounds, a.k.a degradation, e.g. from a protein to amino acids and anabolism (*ανα*-*ana*, “upward” and *βαλλειν*-*ballein*, “to throw”); the processes that foster the transformation of energy and low level compounds into higher level compound, a.k.a. synthesis, e.g. from hydrocarbons to membrane phospholipids. Anabolic and catabolic pathways define the best described network of chemical reactions occurring in living systems, containing thousands of reactions. The reconstruction of this complex network led to the creation of models in order to computationally study its dynamic behaviour (Thiele et al. 2013). The proper balance between this sets of reactions allows the cell to maintain its entropy under control, allowing its existence as an highly organized and ordered system. For what concerns multicellular organisms, cell metabolism has direct consequences on the cell itself, the tissue which it forms and on the organism that contains it. As long as metabolic reactions are taking place, the cell avoid the thermodynamic equilibrium, even apoptosis (*απα*-*apo*, “separation” and *πτωσιζ*-*ptosis*, “falling off”), the programmed cell death processus, preserve the enhancement of entropy of the surrounding tissue. The metabolism has a central role in both physiological and pathologic conditions and the capability of the cell to adapt its metabolism to face evolving environmental conditions is the key to cell survival itself.

Cancer is a pathologic state characterized by a massive cell fate reprogramming, a.k.a. neoplastic transformation, to detriment to the nearby tissue homoeostasis and definitely organism survival. In this pathology, the cancer cell avoids growth and immune control and is able not only to foster its own replication rate but also to further modify the surrounding environment to its own advantage. Indeed the peculiar chemical and physical conditions, such as  $O_2$  levels, low pH and nutrients disposability, of tumor microenvironment could partially explain the deregulation of cellular energetics, proper of cancer cells. Even if metabolic alterations have been recognised as an hallmark of cancer in 2011 (Hanahan and Weinberg 2011), cancer metabolism studies predates oncogene and tumor suppressor revolution by some 50 years. Indeed first evidences of the relevance of metabolic rewiring

during tumorigenesis were found by Otto Warburg in 1924 (Otto Warburg, Negelein, and Posener 1924, and @Warburg1956, reviewed in Hsu and Sabatini (2008) and Vander Heiden, Cantley, and Thompson (2009)). Briefly This phenomenon involves the propensity for proliferating cells ((Curi, Newsholme, and Newsholme 1988) and (Colombo et al. 2010)), including cancer cells, to take up glucose and produce lactate even when oxygen is present, the so called Warburg effect a.k.a. aerobic glycolysis. This process results far less efficient (~18-fold) in terms of net ATP production per molecule of glucose compared to mitochondrial oxidative phosphorylation, however it provides substrates for biosynthetic pathways. In tandem with glucose, glutamine has been found to be an additionally critical nutrient that foster proliferation (Cantor and Sabatini 2012), both by means of energy production and its ability to fuel biosynthetic pathways. Moreover glutamine has an key role in the generation of cellular antioxidant glutathione, which in turn is a master buffer of reactive oxygen species ((Eagle 1955), (Kvamme and Svenneby 1960), (Matés et al. 2002) and (Heyland et al. 2013)). As reviewed in (Vander Heiden and DeBerardinis 2017), during the course of the disease, the metabolic rewiring always occurs, but with different impacts on tumorigenesis: 1) it can or it can't be the leading cause of the pathology i.e. when the transforming cause is the mutation of a metabolic gene e.g. IDH1 and IDH2 mutations in acute myeloid leukemia and low grade gliomas, 2) it can be consequent to specific pathways iper-activation i.e. oncogene addiction e.g. BRAFV600E in melanoma and in turn it can be the natural adaptation to the stressing tumor microenvironment as reviewed in (Hsu and Sabatini 2008).

In the first case, dramatic consequences occur at the epigenetic level, resuting in a block of cellular differentiation (Dang, Yen, and Attar 2016). In the second case the oncogenic drivers impose an enhancement of anabolic metabolism ((Vander Heiden, Cantley, and Thompson 2009) and (Haq et al. 2013)).

Indeed metabolic enzymes are finely tuned by post translational modifications such as phosphorylation. Phosphorylation is a quite energetically expensive tool in protein activity regulation since it requires ~ 11 Kcal/mol per phosphorylation site, the amount of energy produced by ATP hydrolysis. Acetylation is the second mechanism of protein regulation in terms of energy consumption (~ 7Kcal/mol per acetylation site, due to the consumption of a molecule of Acetyl-CoA). It is worth noticing that usually in cell signalling pathways, phosphorylation induced, protein sterical rearrangements usually foster protein activation, e.g. receptor tyrosine kinase (RTK); conversely two key hubs of the glycolytic pathway, the phosphofruktokinase (PFK2) and the pyruvate kinase (PK) are inactivated by phosphorylation (Berg, Tymoczko, and Stryer 2002), in a fascinating energy saving and monitoring fashion. However, due to the reduced availability of metabolomics data, such as direct metabolites measurements,



metabolic gene expression profile has been used as a proxy for metabolic pathway activity.

Since the development of high throughput methods for studying gene expression, the abundance of publicly available, transcriptomic data dramatically increased. Moreover, for what concerns cancer biology, the project of the Cancer Genome Atlas (TCGA) consortium, represents, so far, one of the widest repository of transcriptional, genomic and clinical data, publicly available of human tumor samples. For this reason we decided to use TCGA transcriptional data for the creation of the “metabolic clusters” and their subsequent further genomic and clinical classification. Clustering approach represents one of the oldest and, so far, widely used mathematical approaches in order to identify, in an unsupervised fashion, gene signatures, and molecular subtypes of human tumors starting from transcriptomic data.

Bottom line, in order to identify metabolic differences between cancer samples, we took advantage of metabolic gene signatures. In turn, this allowed us to use transcriptomic data, instead of direct metabolic data, in order to identify differences in metabolic profiles among different samples. This goal has been achieved by means of a clustering algorithm and finally, by means of statistical methods adapted to each type of genomic and clinical feature of interest, we have been able to discover a large number of strong and sometimes unexpected correlations between metabolism and molecular and phenotypic features of human tumors.

### 3 Experimental Approach

In order to exploit transcriptome profiling to identify metabolic differences in cancer samples, for each metabolic gene set and each tumor type, the sample cohort has been analysed by means of a partitioning around medoids (PAM) approach (L. Kaufman and Rousseeuw 1987), on metabolic genes expression levels.

This analysis allowed the identification of clusters of patients characterized by differences in the expression profile of metabolism related genes. In contrast to the *k-means* clustering algorithm, which uses centroids (clusters mean values), PAM uses  $k$  representative objects (medoids that respects to centroids are always member of the data set) making it a more robust approach (Arbin et al. 2015), as not affected by outliers. The unsupervised learning algorithm is then coupled with the clusters silhouette analysis in order to allow a (almost) completely unsupervised choice of the number of clusters. The application of the following procedure allowed a selection of samples subgroups characterized by a robust and unbiased approach. Moreover compared to differential gene expression approaches, this analysis allows a better identification of synchronous transcriptional changes of the catabolic and anabolic genes of the metabolism under exam. In order to further evaluate the proper metabolic nature of the cluster, the random shuffle of gene name labels of the expression data samples has been performed. This procedure generated a novel clustering result that avoids the possibility that the results of the correlation analysis could be caused by random clusters. The results obtained with these clusters didn't show any statistically significant correlation at a Bonferroni corrected  $p$ value  $< 0.05$  (data not shown).

As a result of this classification, the difference in metabolic genes expression is translated into a categorical variable that has been tested for association with different clinical and genomic features. From a mathematical point of view these features can be broadly divided in:

1. Numeric continuous, i.e. methylation, copy number alteration (CNA, quantified as Significant Gain Or Loss [SGOL] scores), RPPA data (protein abundance levels) and miRNA expression levels.
2. Categorical nominal, i.e. clinical data elements (CDEs, such as histological type and primary therapy outcome), mutation data (the presence/absence of a specific mutation).
3. Survival, overall and relapse free.

And for each of them the most specific and appropriate statistical test was applied:

1. Mann–Whitney–Wilcoxon test (Mann and Whitney 1947) or Kruskal-Wallis rank sum test (Kruskal and Wallis 1952) for more than 2 clusters. These tests have been used to test the independence of the observations belongings to different groups.
2. Pearson’s  $\chi$ -squared,  $\chi^2$ , test (K. Pearson 1900) or Fisher’s exact test (R. A. Fisher 1922) for 2x2 tables. These tests have been developed for the analysis of contingency tables, in order to measure the association between categorical variables.
3. Cox proportional hazards regression model (Cox 1972). This is a survival model that relates the time that passes before some event (in this case patient’s death) occurs to one covariate (cluster) that may be associated with that quantity of time. Since in this case the covariate is categorical a Kaplan-Meier model could have been chosen as well, however hazard regression model also returns the hazard ratio showing the percentage of increase/decrease in the hazard.

The summary of genomic and clinical features is presented in the following table.

Feature Name	Data Type	Variable Type
Survival	Clinical	Survival
Clinical Data Element, (CDE)	Clinical	Categoric-Nominal
Mutation	Molecular	Categoric-Binary
Copy Number Alteration, (CNA)	Molecular	Numeric-Continuous
Methylation	Molecular	Numeric-Continuous
Reverse Phase Protein Array, (RPPA)	Molecular	Numeric-Continuous
miRNA expression data	Molecular	Numeric-Continuous

## 4 Materials And Methods

## 4.1 Cancer Samples Data

The results shown here are in whole or part based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>. Clinical and genomic data of TCGA cancer samples and patients (release 2016\_02\_28) were downloaded from the Broad TCGA GDAC site (<https://confluence.broadinstitute.org/display/GDAC/Home>), by means of *firehose\_get* Version: 0.4.1. The data refers to a cohort of 11158 cancer patients, of 34 different tumor types, the list of all their acronyms is presented in the following table. In 5 cases, it has been possible to merge assimilable tumors, i.e. COADREAD (union of COAD and READ), GBMLGG (union of GBM and LGG), KIPAN (union of KICH, KIRC, and KIRP), LUNG (union of LUAD and LUSC) and STES (union of ESCA and STAD).

Tumor Type	Acronym	Tumor Type	Acronym
Acute Myeloid Leukemia	LAML	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	DLBC
Adrenocortical carcinoma	ACC	Mesothelioma	MESO
Bladder Urothelial Carcinoma	BLCA	Ovarian serous cystadenocarcinoma	OV
Brain Lower Grade Glioma	LGG	Pancreatic adenocarcinoma	PAAD
Breast invasive carcinoma	BRCA	Pheochromocytoma and Paraganglioma	PCPG
Cervical squamous cell carcinoma and endocervical adenocarcinoma	CESC	Prostate adenocarcinoma	PRAD
Cholangiocarcinoma	CHOL	Rectum adenocarcinoma	READ
Colon adenocarcinoma	COAD	Sarcoma	SARC
Esophageal carcinoma	ESCA	Skin Cutaneous Melanoma	SKCM
Glioblastoma multiforme	GBM	Stomach adenocarcinoma	STAD
Head and Neck squamous cell carcinoma	HNSC	Testicular Germ Cell Tumors	TGCT
Kidney Chromophobe	KICH	Thymoma	THYM
Kidney renal clear cell carcinoma	KIRC	Thyroid carcinoma	THCA
Kidney renal papillary cell carcinoma	KIRP	Uterine Carcinosarcoma	UCS
Liver hepatocellular carcinoma	LIHC	Uterine Corpus Endometrial Carcinoma	UCEC
Lung adenocarcinoma	LUAD	Uveal Melanoma	UVM
Lung squamous cell carcinoma	LUSC		

In details, the data (version: `stdata_2016_01_28`) are:

- **Clinical data:**

- Overall and Relapse-free survival from `Clinical.Level_1`
- Clinical parameters (manually selected) from `Clinical_Pick_Tier1.Level_4`

- **Coding Genes Expression:** RNAseqV2\_RSEM\_genes\_normalized\_\_data\_Level\_3
- **Methylation data:** Methylation.Preprocess\_Level\_3
- **Reverse Phase Protein Data:** RPPA\_AnnotateWithGene.Level\_3
- **Copy Number Alteration (CNA) data:** derived from  
Level\_3\_segmented\_scna\_minus\_germline\_cnv\_hg19\_\_seg.Level\_3
- **Mutation data:** Mutation\_Packager\_Oncotated\_Calls.Level\_3
- **miRNAs expression data:** Mirnaseq\_\_illumihiseq\_mirnaseq\_\_bcgsc\_ca\_\_Level\_3\_miR\_gene\_expression

## 4.2 CNA SGOL Scores

CNA SGOL scores at the single gene level were generated by means of **cghMCR**, **DNACopy** and **CNTools** *Bioconductor* packages. **cghMCR** package allows the calculation of segment gain or loss (SGOL) starting by segmented data, by means of a modified version of GISTIC algorithm. The *segment* function of **DNACopy** package is used to segment the normalized data so that chromosome regions with the same copy number have the same segment mean values. Then, by means of **CNTools**, *getRS* function, the data returned by *segment* are organized in a matrix format. *SGOL* function of **cghMCR** is ultimately used to compute the SGOL scores for genes by calculating the summations (parameter method) for all the positive and all the negative values, over and below respectively, a set threshold (-0.5, 0.5). To avoid redundant informations linked to the fact the it is very unlikely for a certain CNA to occur in a single gene region only, the data were regrouped at the chromosomal region level using the positional gene sets of Broad Institute.

## 4.3 Gene Sets

The datasets of 345 metabolic pathways were selected from *c2.KEGG*, *c2.REACTOME*, *c5.BP* and *hallmark-MSigDB v5.2* collections. The positional datasets used to collapse CNA data at cytogenetic band level were taken from *c1-MSigDB v5.2* major collection. Hallmark datasets, used in GSEA were taken from *hallmark-MSigDB v5.2* (Liberzon et al. 2011, and @Thiele2013) All datasets were manually downloaded from

<http://software.broadinstitute.org/gsea/downloads.jsp>

#### 4.4 Methylation Data

Methylation data are the result of GDAC from Human Methylation 450k platform and already collapsed at gene level. Methylation at all genes level, for every tumor type, was calculated as the mean  $\beta$  score of all the genes analysed in every patient.

#### 4.5 Differential Gene Expression

Analysis of differentially expressed genes (DEGs) between two clusters has been performed by means of **limma** R package, with standard settings.

#### 4.6 Gene Set Enrichment Analysis

Gene Set Enrichment Analysis ((Mootha et al. 2003) and (Subramanian et al. 2005)) was performed by running the GSEAPreranked tool from command line (gsea2-2.2.0.jar, with the following parameters: -mode Max\_probe,-norm meandiv,-nperm 1000, -rnd\_seed timestamp,-set\_max 500 and-set\_min 15). The list's ranking metric was calculated on the LogFC of the differentially expressed genes between the two clusters.

#### 4.7 Metabolic Clusters Generation

The metabolic clusters were defined by means of the partition around medoid (PAM) cluster algorithm (1) on the Spearman rank correlation coefficient (2),  $\rho_{rg}$  or  $\rho_s$ , -based distance matrix of metabolic genes expression in cancer samples. The number of clusters (between a 1:10 range) generated by each metabolic gene set has been then estimated by optimum average silhouette width (3). The Duda-Hart test was applied to verify significance of 1 cluster estimation.

- (1) The goal of the algorithm is to minimize the average dissimilarity of objects to their closest selected object (medoid) (L. Kaufman and Rousseeuw 1987)

(2)

$$\rho_{rgX,rgY} = \frac{cov(rgX,rgY)}{\sigma_{rgX}\sigma_{rgY}}$$

where  $rgX$  and  $rgY$  are the ranked converted equivalent of  $X$  and  $Y$  raw scores.

Or by

$$\rho_s = 1 - \frac{6 \sum_i D_i^2}{N(N^2 - 1)}$$

where  $D_i = r_i - s_i$  is rank difference between first and second variable rank of  $i^{th}$  observation and  $N$  is total observations number

(3)

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

where  $a(i)$  is the average dissimilarity of  $i$  with all other data within the same cluster and  $b(i)$  is the lowest average dissimilarity of  $i$  to any other cluster. The resulting average  $s(i)$  overall data of a cluster is the measure of how tightly grouped all the data in the cluster are.

The clusters and the distance matrix were computed by means of the *pamk* function, from **fpc** R package and *Dist* function, from **amap** R packages, respectively.

## 4.8 Correlation Between Metabolic Clusters And Clinical-Genomic Variables

The nominal variable resulted from the clustering analysis has been correlated with the three kinds of variables analysed in this study:

1. Numeric continuous, i.e. methylation, CNA (Segment Gain Or Loss, SGOL, scores), and RPPA.
2. Categorical nominal, i.e. clinical data, mutation data and clusters.
3. Survival, overall and relapse free.



And for each of them a specific statistical analysis was applied:

1. Mann–Whitney–Wilcoxon test (4) or Kruskal-Wallis rank sum test (5) for more than 2 clusters.

(4)

$$M = \frac{1}{c(c-1)} \sum AUC_{k,l}$$

where

$$AUC_1 = \frac{U_1}{n_1 n_2}$$

where

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

where  $n_1$  is sample size for sample 1 and  $R_1$  is the sum of ranks in sample 1;  $c$  is the number of classes and the  $R_{k,l}$  are the ranking of the items belonging to classes  $k$  and  $l$ , specifically.

(5)

$$P(\chi_{g-1}^2 \geq K)$$

where

$$K = (N - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2}$$

where  $n_g$  amount of observations in group  $g$ .  $r_{ij}$  is the rank of observation  $j$  in group  $i$ .  $N$  is the total number of observations in all the groups.  $\bar{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$

2. Pearson's  $\chi$ -squared,  $\chi^2$ , test (6) or Fisher's exact test (7) for 2x2 tables.

(6)

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} = N \sum_{i=1}^n \frac{(O_i/N - p_i)^2}{p_i}$$

where  $\chi^2$  = Pearson's cumulative test statistic.  $O_i$  = the number of observations of type  $i$ .  $N$  = total number of observations.  $E_i = Np_i$  = the theoretical frequency of type  $i$  under the  $H_0$  of frequency of type  $i = p_i$  in the population.  $n$  is the number of the cells in the table

(7)

$$p = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}}$$

where  $a + b$  and  $c + d$  are the row totals and  $a + c$  and  $b + d$  are the column totals of a 2x2 table and  $n = a + b + c + d$

3. Cox proportional hazards regression model (8).

(8)

$$\lambda(t; z) = \lambda_0(t) \exp\{\beta_0' z(t)\}$$

where  $\beta_0$  is a p-vector of unknown regression coefficient and  $\lambda_0(t)$  the underlying hazard function ((Therneau 2015) and (Terry M. Therneau and Patricia M. Grambsch 2000)).

## 4.9 Pvalues And Multiple Tests Correction

All pvalues were calculated at a 95% confidence level, so with an  $\alpha = 0.05$ . Due to the elevated number of different comparisons performed during the correlative analysis and the willingness to summarize and to compare the results, Bonferroni (9, in order to control *family wise error ratio*, FWER) and Benjamini-Hochberg in order to control *false discovery rate*, FDR, from pvalues) pvalue corrections were applied at the tumor level for intra tumor comparisons and globally for inter tumor comparisons. Unless otherwise stated, global Bonferroni corrected pvalues  $< 0.05$  were considered as significant.

(9)

$$\alpha_B = \frac{\alpha}{m}$$

where  $\alpha_B$  is the Bonferroni corrected significance level and  $m$  is the number of tested hypotheses. For global comparisons,  $m = \text{number of features} * \text{number of metabolisms} * \text{number of tumors}$ .

#### 4.10 R And RSTUDIO Software

The data and figures were generated by means of R (version 3.2.1) and RStudio (version 1.0.143).

## 5 Results

## 5.1 General Results

The results of the associations are briefly and systematically presented below and more in details in the supplementary tables and figures section. For every clinical and genomic feature, a representative example of the association with a certain metabolism will be presented herein.

### 5.1.1 Clinical Relevance Of Metabolic Clusters

An example of the results of the correlative analysis between metabolic clusters and survival data is shown in Figure 1. In this example the GMP metabolism genes divides the KIPAN cohort in three clusters with different survival rate, with a significant difference between samples from cluster 1 (K1) having a better prognosis compared to samples in cluster 3 (K3). More broadly, the clusters obtained by the 345 metabolic pathways analysed, were tested for being predictors of Overall survival (OS), in all the 38 tumor types and of Relapse free survival (RFS) in 37 tumor types (no RFS data were available for LAML). Almost 40% of the analysed tumors have at least one metabolism affecting its survival. Interestingly not all the tumors show a similar susceptibility in terms of prognosis and some tumors showed an higher number of metabolisms impacting on survival (see Figure 29 and Figure 32). Low grade gliomas and kidney cancer are the tumors that shown an higher number of metabolisms correlating with survival. Summaries of the top, by frequency, tumors which survival is affected by metabolism and summary of the top, by frequency, metabolisms affecting survival in different tumors, are shown in Table 7 and Table 8. In Figure 30 and Figure 33 the pvalues of the correlation between metabolisms and OS and RFS respectively, are shown. In Figure 31 and Figure 34 the pvalues of the correlation between a subset of metabolisms and OS and RFS respectively, are shown. Consistently with these results, similar data were recently obtained, with a different approach on a smaller number tumors (Gaude and Frezza 2016).

CDEs availability in tumor types is very heterogeneous, with few parameters present in almost the totality of analysed tumors i.e. gender (not present in gender specific tumors such as ovarian carcinoma and testicular germ cell tumors ) and many tumor specific, i.e. StGallen classification of breast cancer and Gleason score in prostate cancer. The complete list can be found in Table 1.

The result of the correlation between metabolism and a categorical variable, gender in this case is shown in Figure 2. Fatty acid metabolism clusters have a strong gender bias in low grade gliomas, with cluster 1 patients being prevalently females. In general the almost totality (80%) of tumor types have at least one CDEs associated

with metabolism. In Figure 35 the amount of CDEs correlating with metabolism in different tumor types can be appreciated. The histological type is the CDEs more frequently correlated with metabolism (Figure 36 and Figure 37) suggesting the presence of a tissue specific metabolic rewiring. Association strength between histological types and metabolism in the different tumor types can be seen in Figure 38 and Figure 39. Summaries of the top, by frequency, tumors where CDEs are affected by metabolism and summary of the top, by frequency, metabolisms affecting CDEs in different tumor types, are shown in Table 9.

CDE	Tumors (n)	CDE	Tumors (n)
gender	32	excess adrenal hormone history type-2	1
histological type	28	extravascular matrix patterns	1
primary therapy outcome success	28	fetoprotein outcome lower limit	1
pathologic t	27	her2 immunohistochemistry level result	1
pathologic n	26	histological type-2	1
pathologic m	25	human papillomavirus type	1
karnofsky performance score	17	human papillomavirus type-2	1
neoplasm histologic grade	14	hysterectomy performed type	1
clinical m	11	melanoma clark level value	1
clinical t	9	new gleason score	1
additional treatment completion success	6	nuclear grade iii iv	1
outcome			
clinical n	6	primary neoplasm focus type	1
histological type other	3	primary pattern	1
child pugh classification grade	2	prior systemic therapy type	1
fetoprotein outcome upper limit	2	progesterone receptor level cell percent	1
		category	
fibrosis ishak score	2	relative cancer type	1
food allergy types	2	secondary pattern	1
gleason score	2	StGallen2013	1
tumor type	2	tertiary pattern	1
vascular tumor cell type	2	therapeutic mitotane levels achieved	1
asbestos exposure type	1	tumor response cdus type	1
breast carcinoma immunohistochemistry pos	1	tumor shape pathologic clinical	1
cell score			
chemotherapy regimen type	1	weiss score	1
excess adrenal hormone history type	1		

Table 1: **List of analysed CDEs.** This table contains all analysed CDEs with their frequency across tumor types.

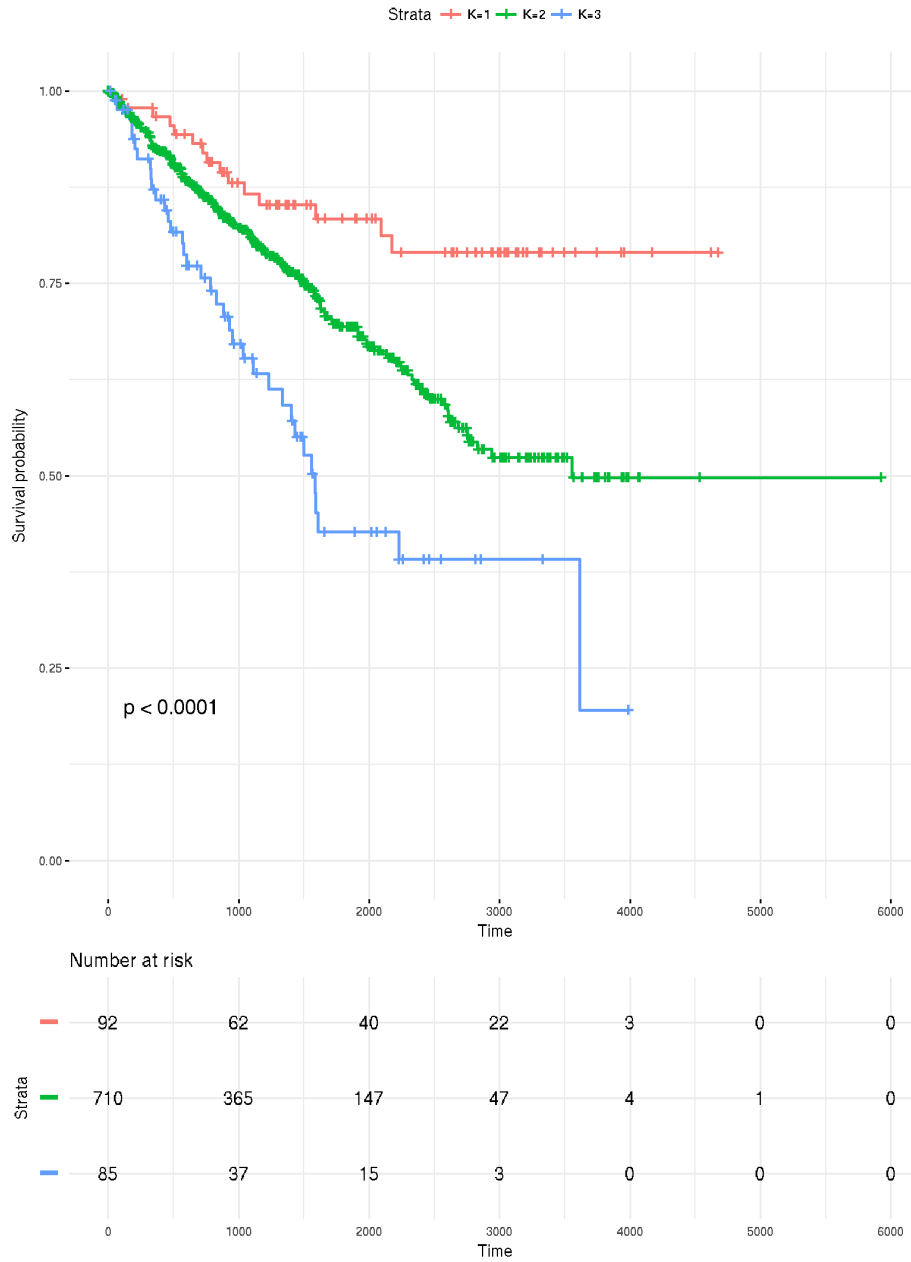


Figure 1: **Correlation between GMP metabolism and survival in kidney cancer.** The Kaplan-Meier plot showing the differences in survival between the three clusters obtained by GMP metabolism (upper panel) and the table with the number of patients at risk (lower panel).



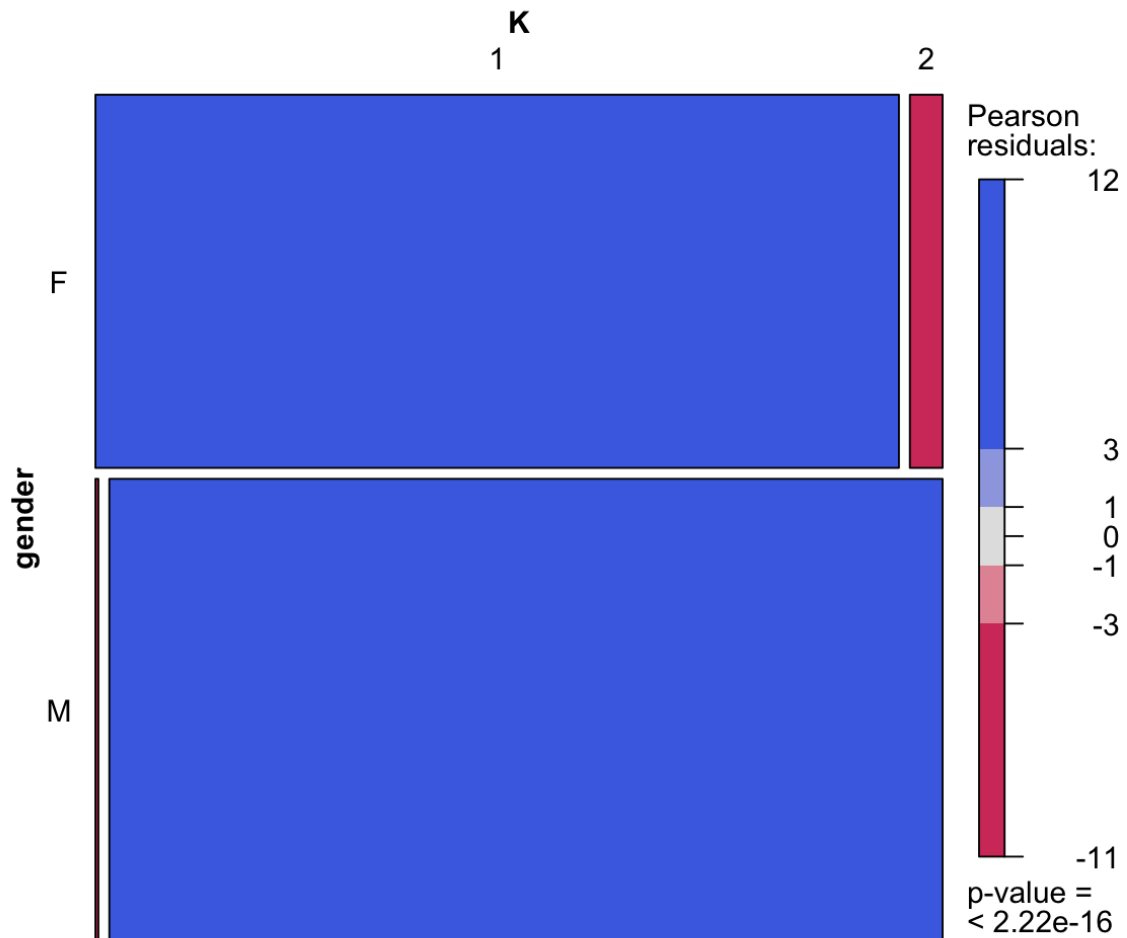


Figure 2: **Association between fatty acid metabolism and gender in low grade glioma.** The mosaic plot represents the abundance of female (F) and male (M) patients (rows) in the two clusters (K1,K2, columns) generated by fatty acid metabolism in low grade glioma. Pearson's positive and negative residuals are highlighted in blue and red respectively. It is striking the enrichment of females patients in K1 and the enrichment of male patients in K2. Pearson residuals are given by  $\frac{(observed - expected)}{\sqrt{expected}}$ .

### 5.1.2 Genomic Relevance Of Metabolic Clusters

For what concerns genomic characterization of metabolic clusters, the data analysed are the following:

1. Mutations, 3 different depth levels:
  - a. Protein coding gene level (PCGL, n=18833). Nucleotide variations occurring in the exons of coding genes.
  - b. Missense/Nonsense-mutations and Deletions/Insertions at protein level of coding genes (n=886274). Nucleotide variations inducing an alteration in the aminoacidic composition of a protein.
  - c. Nucleotide variations at genome level (n=1283947). Nucleotide variations occurring in any region comprised between the 5' and 3' UTRs of a coding gene.
2. CNA data at cytogenetic band level
3. Methylations, 2 different depth levels:
  - a. Whole genome level
  - b. Gene level
4. RPPA (Reverse Phase Protein Array) data of specific protein/protein post translational modification, abundance level. Protein arrays allow the simultaneous quantification of proteins and protein post translational modifications.
5. miRNAs expression level

All these genomic features are continuous numerical variable exception made for mutation data that were considered as Boolean variables, the mutation could be present or absent. For the first case non-parametric Mann-Whitney U or Kruskal-Wallis tests have been used, while Fisher or  $\chi^2$  tests have been used for mutation data.

#### 5.1.2.1 Mutation Data

Mutations were analysed at two main levels of detail: 1) protein coding gene level (PCGL) mutations, the less specific, and 2) at protein level and nucleotide level, more specific. The results of the correlation between PCGL

mutations and metabolisms returned a discrete number (33 out of 18833, 0.17%, across all cancer samples) of mutated genes associated with at least one metabolism at a Bonferroni corrected pvalue  $< 0.05$  ( $m=311 * 10^6$ ), however any information about the specific kind of mutation was lost since all the different kind of mutations (missense, nonsense, insertions, deletions, frameshift, duplication and repeat expansion) were reduced to a single binary variable. Among the more frequently metabolism-correlated, mutated genes, we find TP53 (the most frequently correlated among different tumor types), EGFR, IDH1, PTEN, PI3KCA and MET. On the other hand protein and nucleotide levels results yielded a remarkably lower amount of significant correlated mutations (6 out of 886274 0.0007% for protein level and 8 out of 1283947 0.0006%), but they preserved the information about the type of mutation. Moreover for what concerns the latter results a high level of redundancy between protein and nucleotide level was observed since IDH1, TVP23C, RNF43, BRAF, NRAS and GTF2I genes were found in both analysis. More interestingly two genes (DNAJC18 and ZNF43) were found to have single deletions at intronic level. Since the diversity of mutational processes observed in human tumors ((Alexandrov et al. 2013) and (Chalmers et al. 2017)) normalization on mutational landscape of every tumor type will surely be done in the future. In Figure 3 it can be appreciated the result of the correlation between a protein level mutational and retinol metabolism in thyroid carcinoma, while more details about the mutational analysis can be found in Figure 40 - Figure 54 and Tables 10 - 11.

### 5.1.2.2 Copy Number Alteration Data

Chromosomal instability (CIN) has been extensively studied in cancer ((Bakhoun and Compton 2012) and (Giam and Rancati 2015) ) and a recent study showed that metabolic requirements of highly proliferating tumors fosters the selection of cancer cells harbouring amplification of glycolytic genes and other cancer linked metabolic enzymes (Graham et al. 2017). Copy number alterations, summarized at cytogenetic band level, are correlated with at least one metabolism in 32 out 38 (~84%) tumors. The tumors with the higher frequency of correlation are kidney, lung, breast and brain Figure 55. If we observe the correlations between CNAs and metabolism in different tumor types we can see a subgroup which is more frequently associated (Figure 56). Results of the correlations between copy number alteration (CNA) data and metabolism and are summarized in Table 13. The CNAs most frequently correlated with metabolism are found on chromosome 3, 1, 19 and 20. Details of the correlations between those CNAs and metabolism can be appreciated in Figure 57 and Figure 58.

Below is shown the difference in chr3p24 observed between the two clusters, created by glycerophospholipids

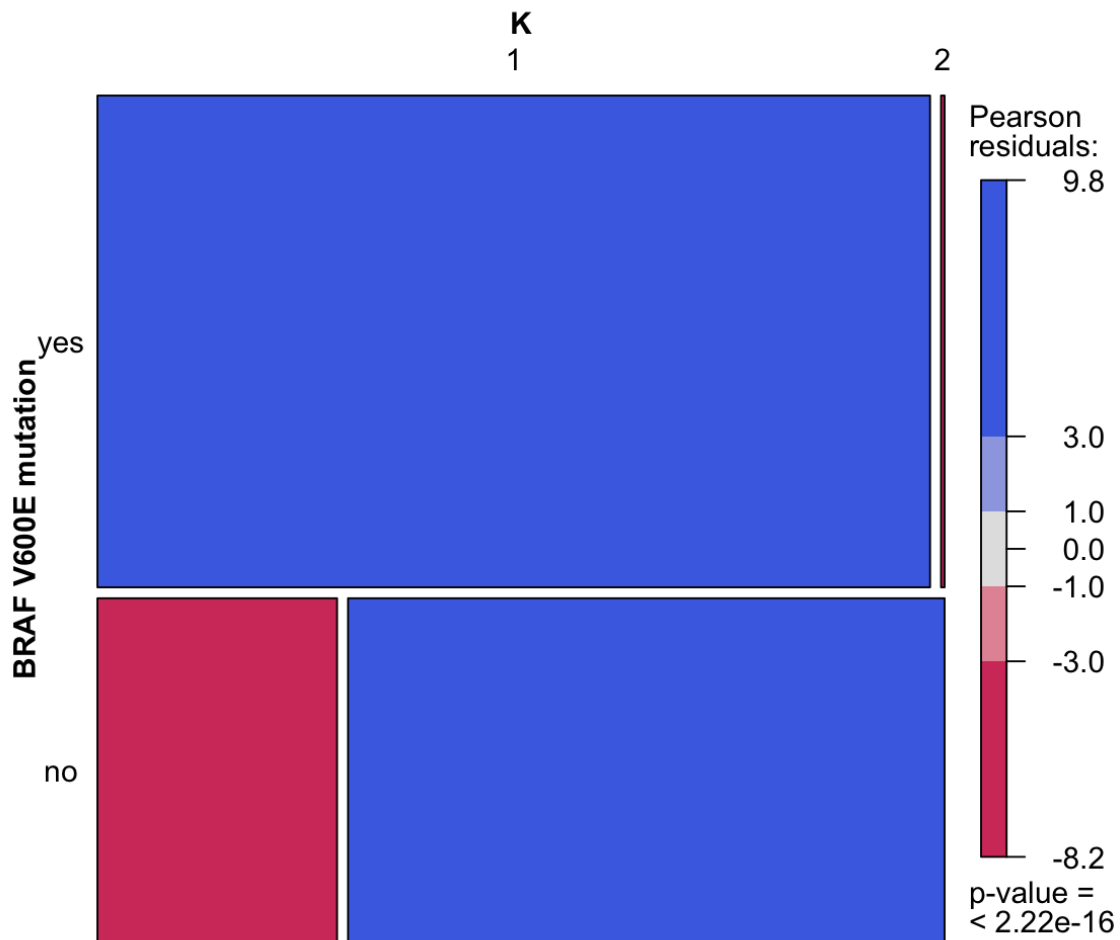


Figure 3: **Correlation between retinol metabolism and BRAF v600E in thyroid carcinoma.** The mosaic plot represents the distribution (presence: yes or absence: no; rows) of  $BRAF^{V600E}$  mutation in the two clusters (K1,K2, columns) generated by retinol metabolism in thyroid carcinoma. Pearson's positive and negative residuals are highlighted in blue and red respectively and they are calculated as follows:  $\frac{(observed - expected)}{\sqrt{expected}}$

metabolism, in kidney cancer cohort. The box plot shows the difference of a single cytogenetic band, while in Figure 15 and Figure 24 (in the detailed results section) examples of complete CNA landscape are shown.

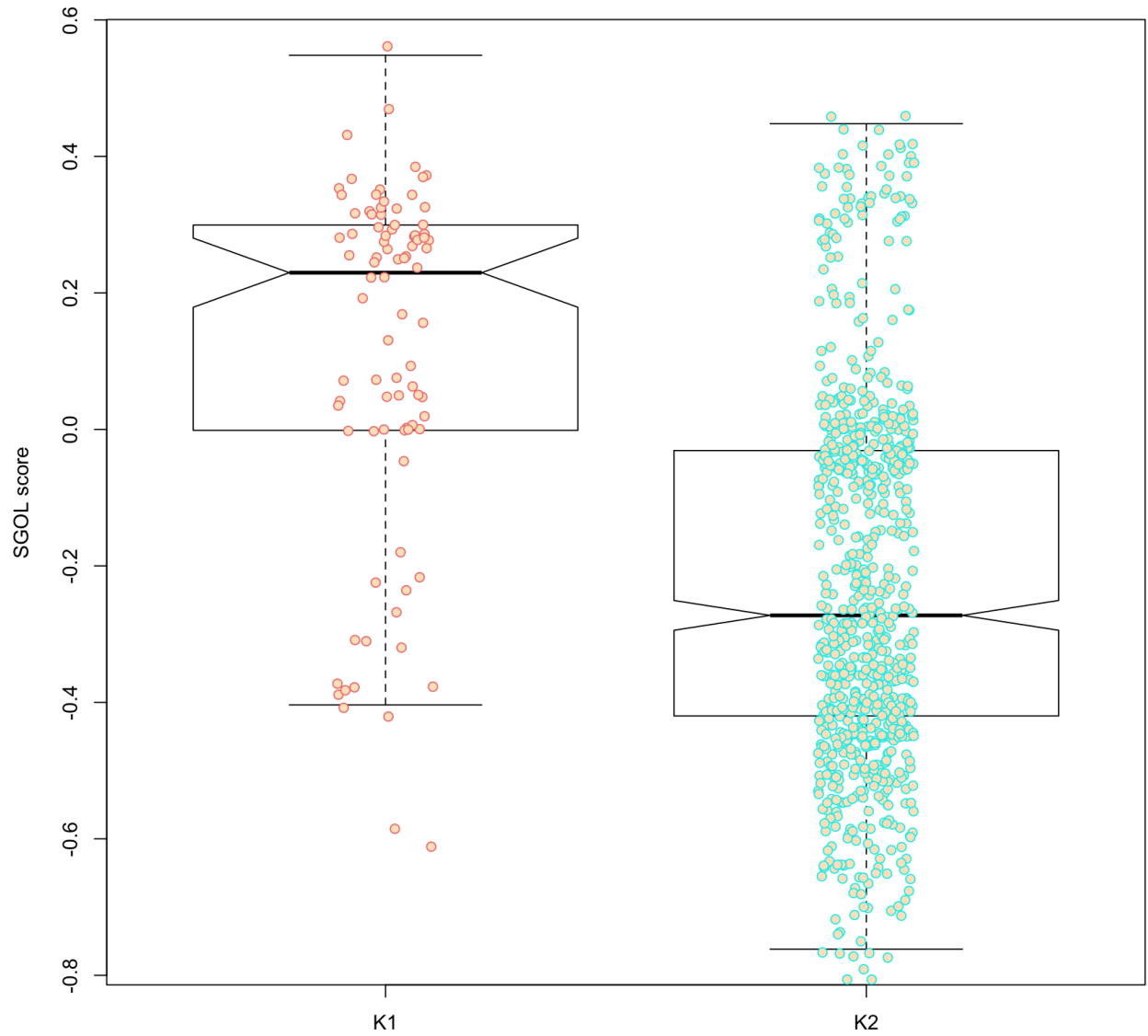


Figure 4: **Correlation between chr3p24 and glycerophospholipid in kidney cancer.** Box plot representing the difference in copy number of chr3p24 between the two clusters generated by glycerophospholipid metaolism in kidney cohort. The notch in the boxes indicates 95% confidence interval for the medians in each cluster.  $P_{\text{val}} < 2.16 * 10^{-16}$

### 5.1.2.3 Methylation Data

Methylation data were analysed at two different levels:

1. Whole genes level: for each sample in tumor cohort the median level of methylation of all genes was calculated and then tested for correlation with metabolic clusters, in order to generate a proxy for whole genome methylation level.
2. Single gene level: each gene  $\beta$  value has been correlated with metabolic clusters.

DNA methylation of CpG islands plays a key role in the epigenetic control of gene expression and correlates with cellular differentiation. It has been extensively studied in both pathological and physiological conditions ((Razin and Cedar 1991), (Phillips 2008), (Wagner et al. 2014) and (Moarii et al. 2015)).

#### 5.1.2.3.1 All Genes Methylation Level

The correlative analysis between metabolic pathways and methylation levels, highlighted an high number of tumor types, 26 out of 38 (68%), in which methylation correlates with at least one metabolic pathway. Moreover is it possible to see a subgroup of tumors, i.e. GBMLGG, KIPAN, LUNG and TGCT, which show a stronger correlation between broad methylation and metabolism. Summaries of the presence of correlation between metabolism and all genes methylation data are presented in Figure 60 - Figure 62 and Table 14. In Figure 5 a representative result of the correlation between all genes methylation level in lung cancer and glutamine metabolism is shown and in Table 2 details of pvalues from pairwise comparisons are shown. Glutamine has been reported to influence chromatin organization, cellular histone and DNA methylation levels by means of the production of glutamine-derived metabolite  $\alpha$ -KG (Ji Zhang, Pavlova, and Thompson 2017).

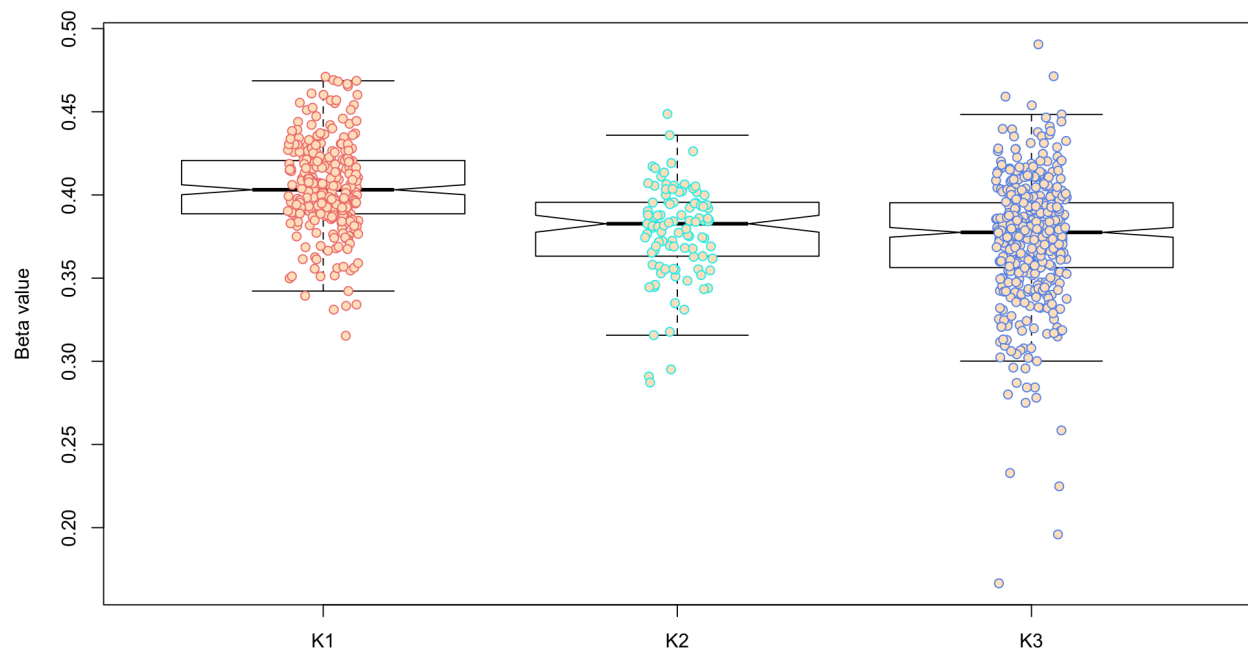


Figure 5: **Correlation between all genes methylation and glutamine metabolism in lung cancer.** Box plot representing the difference in all genes methylation levels (Beta value) between the three clusters generated by glutamine metabolism in kidney cancer. The notch in the boxes indicates 95% confidence interval for the difference between the medians. It is appreciable the difference in terms of higher methylation of cluster 1 compared to clusters 2 and 3, Kruskal-Wallis Pval  $< 2.16 * 10^{-16}$



- **method:** Mann-Whitney U test
- **data.name:** All genes methylation and K
- **p.value:**

	1	2
2	2.5e-14	NA
3	1.036e-33	0.1738

Table 2: **Pairwise Pvalues results** This table contains the Pvalues of pairwise comparisons of the analysis of all genes methylation in the three different cluster generated by glutamine metabolism in lung cohort

### 5.1.2.3.2 Gene Specific Methylation Level

The analysis at single gene level resulted in 34 out of 38 (89%) tumor samples, Table 15, which have the methylation of a gene significantly correlated with at least a metabolic pathway (results summarized in Figure 63 - Figure 67). Interestingly the analysis highlighted \* *LOC100130933*\* (SMIM6/ELN), a recently discovered gene (D. M. Anderson et al. 2016) involved in the regulation SERCA activity, to be the gene which methylation is more frequently correlated with metabolism, in different tumors. However its direct link to metabolism hasn't been investigated yet and this results could led to better investigate its expression in the different metabolic clusters. In Figure 6 correlation between *LOC100130933* methylation and glycosaminoglycan metabolism in bladder cancer is shown.

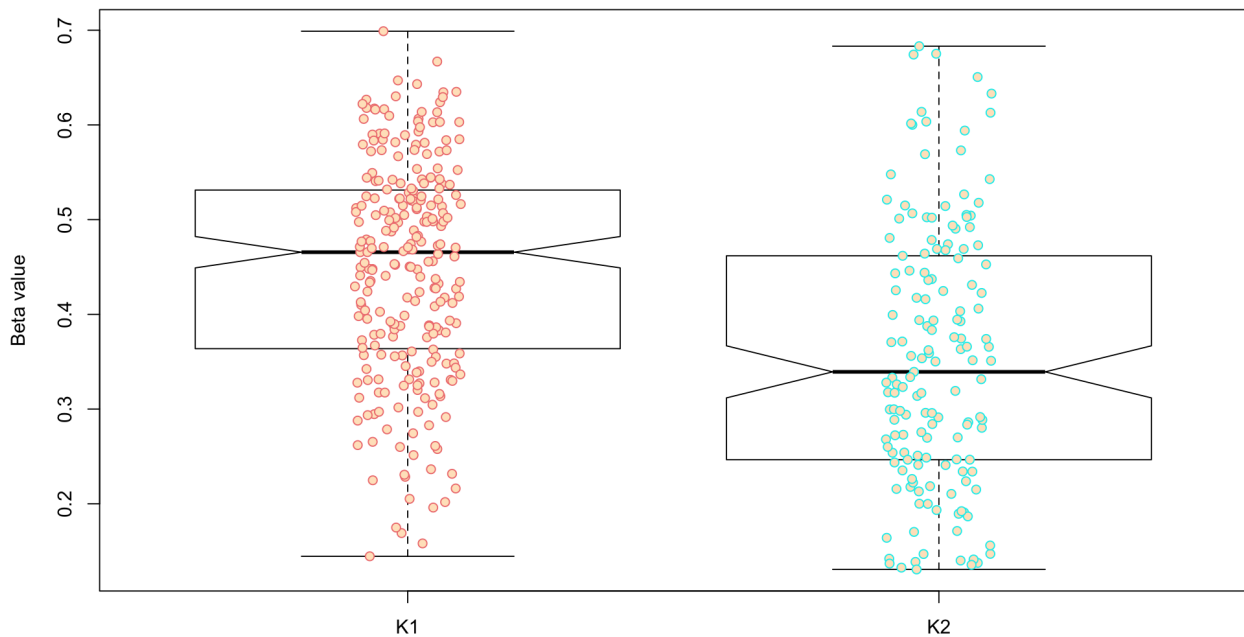


Figure 6: **Correlation between *LOC100130933* methylation and glycosaminoglycan metabolism in bladder cancer.** Box plot representing the difference in *LOC100130933* methylation level (Beta value) between the two clusters generated by glycosaminoglycan metabolism in bladder cancer. The notch in the boxes indicates 95% confidence interval for the medians in each cluster.  $P_{\text{val}} = 5.5 * 10^{-10}$

#### 5.1.2.4 RPPA Data

TCGA Reverse Phase Protein Array data represents one of the widest collections of cancer proteomic data. These data are not only relevant for what concerns protein abundance, but they also provide quantification of some of the most relevant protein post translational modifications. The identification of the top differentially present post translational modifications between the clusters helps, in turn, an easier and more direct molecular characterization at a cell signalling level. Results of the correlation between RPPA data and the different metabolisms are summarized in Figure 68 - Figure 72 and Table 16. Figure 7 shows the representative correlation between sterols metabolism and GSK3 phosphorylation in serine 9 in sarcoma. GSK3A/B are highly conserved kinases that have roles in a number of signal transduction pathways regulating cell growth, differentiation, and development. GSK3B activity is regulated negatively by the phosphorylation of serine 9 ((Stambolic and Woodgett 1994), (G. Cai et al. 2007) and (Giovannini et al. 2013)). In Table 3 details of pvalues from pairwise comparisons are shown.

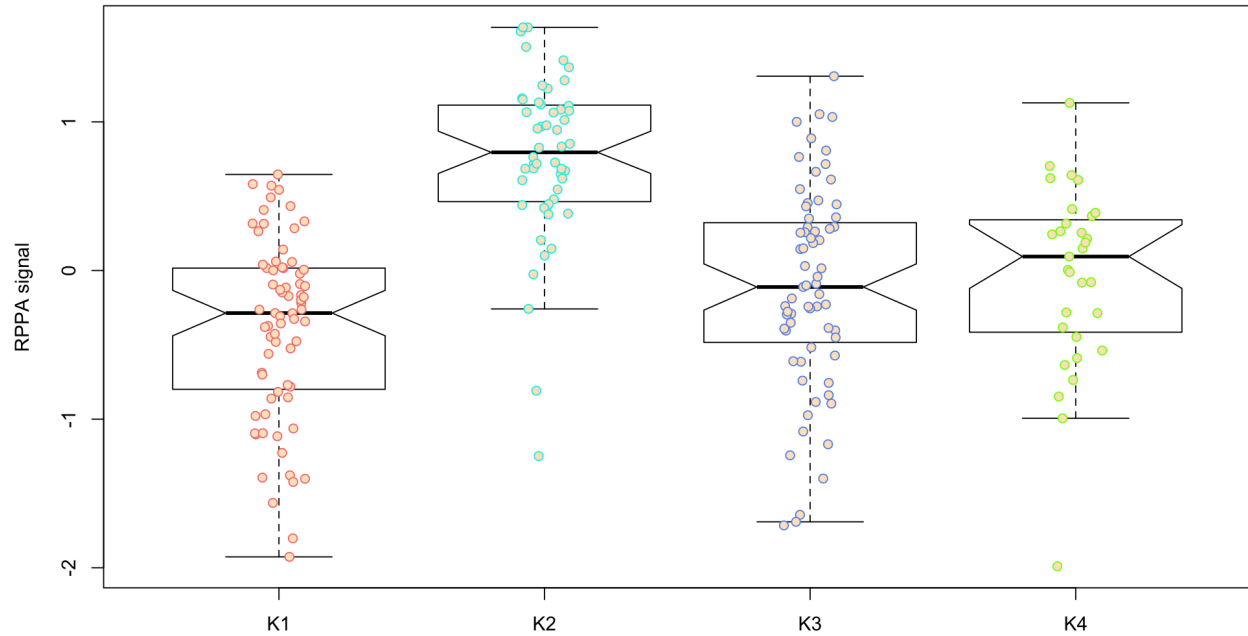


Figure 7: **Correlation between *GSK3* Serine 9 phosphorylation and sterol metabolism in sarcoma.** Box plot representing the difference in *GSK pS9* level (RPPA signal) between the four clusters generated by sterol metabolism in bladder cancer. The notch in the boxes indicates 95% confidence interval for the medians. Kruskal-Wallis Pval  $< 2.16 * 10^{-16}$

- **method:** Mann-Whitney U test
- **data.name:** GSK3\_pS9 and K
- **p.value:**

	1	2	3
2	3.857e-16	NA	NA
3	0.02794	4.487e-11	NA
4	0.009004	3.551e-08	0.5719

Table 3: **Pairwise Pvalues results** This table contains the Pvalues of pairwise comparisons of the analysis of *GSK pS9* in the four different clusters generated by sterol metabolism in sarcoma cohort

#### 5.1.2.5 miRNA Data

miRNAs are responsible for gene silencing and post translational regulation of gene expression. Almost 80% of analysed tumors showed at least one correlation between a miRNA and a metabolism (Table 17). Among the most frequently metabolism-correlated miRNAs, *hsa-mir-375* is the top of the list. Its high rank have been anticipated mainly due to the fact that on one hand side its transcriptional regulation is controlled by FOXO2, a transcription factor that regulates the expression of genes important for glucose sensing in pancreatic beta-cells and glucose homeostasis and on the other hand its direct involvement in the regulation of the genes regulating insulin secretion ((Latreille et al. 2015), (Deiuliis 2016) and (Vienberg et al. 2017)). In Figure 73 - Figure 75 heatmaps of the number of tumors, metabolisms and miRNAs are shown. In Figure 76 -Figure 77 details of the metabolism correlating with *hsa-mir-375* are shown, while in Figure 8 the detail of the differential expression of *hsa-mir-375* in the clusters generated by inositol phosphate metabolism in oesophageal and stomach adenocarcinoma is shown and in Table 4 detail of pvalues from pairwise comparisons are shown.

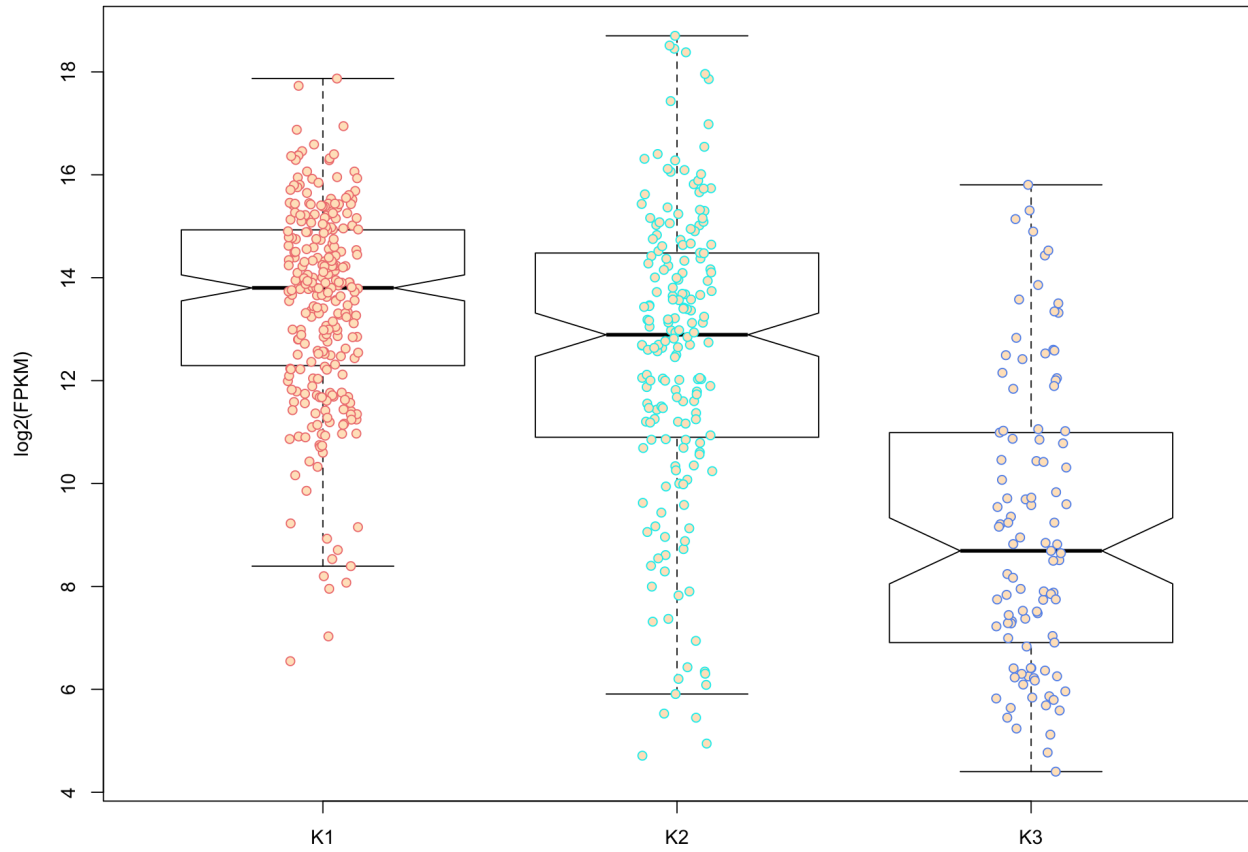


Figure 8: **Correlation between *hsa-mir-375* and inositol phosphate metabolism in oesophageal and stomach adenocarcinoma.** Box plot representing the difference in *hsa-mir-375* expression levels ( $\log_2(\text{FPKM})$ ) between the three clusters generated by inositol phosphate metabolism in bladder cancer. The notch in the boxes indicates 95% confidence interval for the medians. Kruskal-Wallis Pval  $< 2.16 \times 10^{-16}$

- **method:** Wilcoxon rank sum test
- **data.name:** hsa-mir-375 and K
- **p.value:**

	1	2
<b>2</b>	0.0001366	NA
<b>3</b>	1.363e-30	8.538e-17

Table 4: **Pairwise Pvalues results** This table contains the Pvalues of pairwise comparisons of the analysis of *hsa-mir-375* expression levels in the three different clusters generated by inositol phosphate in oesophageal and stomach adenocarcinoma cohort



## 5.2 Context Specific Results

The previously shown data can highlight common metabolism-feature correlation among different tumors, however, in order to obtain more biologically relevant informations, is it possible to focus on a specific metabolism in a specific tumor (context specific results). Below are presented a few examples of the observations that can be drawn from context specific results analysing the unsaturated fatty acid metabolism in LGG and hexose metabolism in BRCA.

### 5.2.1 Unsaturated Fatty Acids Metabolism In LGG

Below are summarized results of the impact of the unsaturated fatty acids metabolism gene set, **GO:0033559**, on LGG. This gene set contains genes which expression is linked to the chemical reactions and pathways involving an unsaturated fatty acid, any fatty acid containing one or more double bonds between carbon atoms.

The expression profile of unsaturated fatty acid metabolism genes divides the LGG sample cohort in two distinct clusters (K1 and K2). The differences in gene expression can be appreciated in Figure 9.

GSEA analysis performed on differentially expressed genes between the two clusters reveals a reduction in, among the others:

- A. broad activation of inflammatory pathways (IL6-JAK-STAT3, INFNA, INFNB, TNF $\alpha$ -NFKB, IL2-STAT5)
- B. epithelial to mesenchimal transition
- C. angiogenesis and hypoxia
- D. glycolysis
- E. proliferation

in K1 samples. Figure 10

From a clinical perspective the patients of the two clusters behave differently in terms of survival, with K1 patients having a better prognosis compare to K2 ones Figure 11. As expected differences between histological subtypes are present, interestingly differences in grade and in primary therapy outcome are also observed (Figure 12).

For what concerns differences in the mutation landscape of the two sample cohorts, we can observe an higher frequency of IDH1 and CIC mutations in K1 and an higher frequency of PTEN and EGFR1 mutations in K2 (Figure 13). An in depth analysis of mutation data highlights IDH1-R132H and EGFR-G598V as the most differentially present mutations in K1 and K2 respectively.

In Figure 14 the CNA landscape of the two cluster is shown. It is striking the difference between K1 (upper plot) and K2 (lower plot) in terms of CNA abundance. K1 is characterized by a severe loss in CHR1p and CHR19q while K2 by a severe loss of CHR9p12-p22 and a gain in CHR7p11 (Figure 15).

Methylation analysis Figure 16 clearly shows a hypermethylation in K1 cohort, characterized by higher frequency of mutated IDH1. This results is consistent with literature data (Turcan et al. 2012), and with grading score (Figure 12, panel C) result since grade depends on tumor differentiation level (higher grade, less differentiation) and hypomethylation usually correlates with differentiation level ((L. Shen et al. 1998), (Soares et al. 1999) and (C. H. Lin et al. 2001), reviewed in (Ehrlich 2002)).

Some of the differences observed in protein abundance between the two cohorts could be explained by CNA data, i.e. the higher abundance of EGFR (located in CHR7p11) and lower abundance of CDKN1B (located in CHR9p21) in K2 samples. Clusters characterization by means of RPPA data is particularly interesting for what concerns post translational modifications. We can see that K2 samples are characterized by a higher A-Raf S299, PKC\_α S657 and HER3 Y1289 phosphorylation. On the other hand K1 samples are characterized by a higher YB-1 S102, EGFR Y1173 and HER2 Y1248 phosphorylation. Interestingly K1 samples show a decrease in GAPDH abundance Figure 17.

Finally the differentially expressed miRNAs are shown in Figure 18. In Table 5 the data of the ten most up and down regulated miRNAs are shown. Interestingly *has-mir-105* are found to be upregulated in K1, the cluster that shows better prognosis has been previously described as having prognostic value in hepatocellular carcinoma (Ma et al. 2017). Consistently with GSEA results, *has-mir-10a*, which involvement in EMT has been recently reported (Y. Liu et al. 2017) is found as the most downregulated one. Moreover *has-mir-196a*, is consistently downregulated in the better prognosis, K1, cluster as suggested by literature (Y. S. Lee et al., n.d.) in pancreatic cancer. At last *has-mir-155* which downregulation correlates with enhance apoptosis and anoikis (cell death triggered by loss of adhesion, ((F.-q. Zhu et al. 2016) and (Rajasekhar et al. 2017)) is found to be downregulated as well. Even more puzzling is the interpretation of higher expression, in K1 samples, of *has-mir-767*, which has

been reported as oncomir (miRNA promoting tumorigenesis) in tandem with CYLD expression (K. Zhang and Guo 2018), which is slightly but significantly (pvalue  $< 2.2e-16$  ) downregulated in K1 (median expression 10.19) respect to K2 (median expression 10.38).

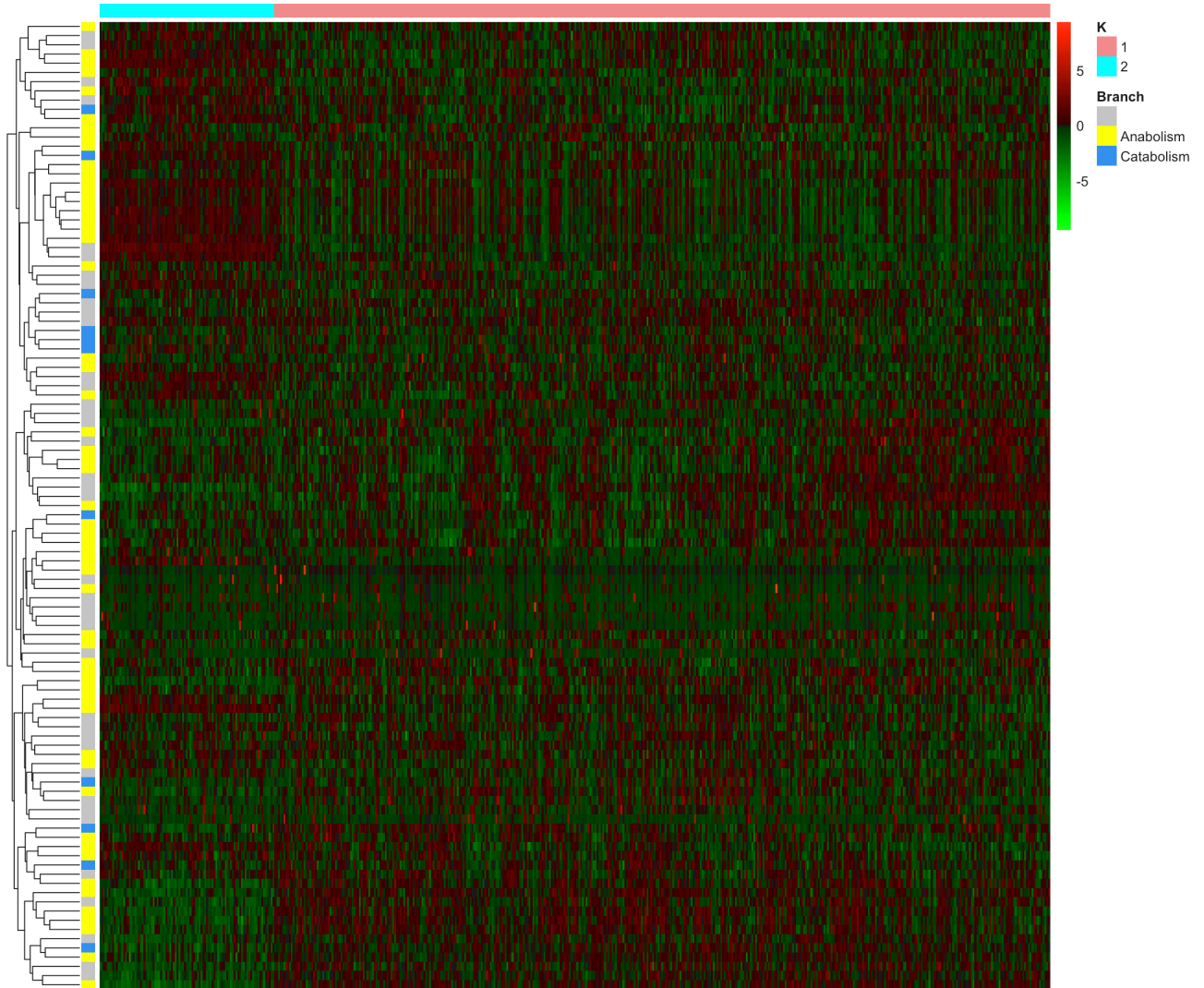


Figure 9: **Differentially expressed metabolic genes in LGG.** Heatmap representing the expression of the genes of the unsaturated fatty acid metabolism (rows) in LGG cohort. The colour bar above heatmap columns (samples) represents samples from the two clusters. The samples are been subdivided as resulted from PAM algorithm. The colour bar on the rows of the heatmap highlights genes belonging to anabolism (yellow) and catabolism (blue). Colours ranging from red (upregulated) to green (downregulated), represent genes expression Z scores.

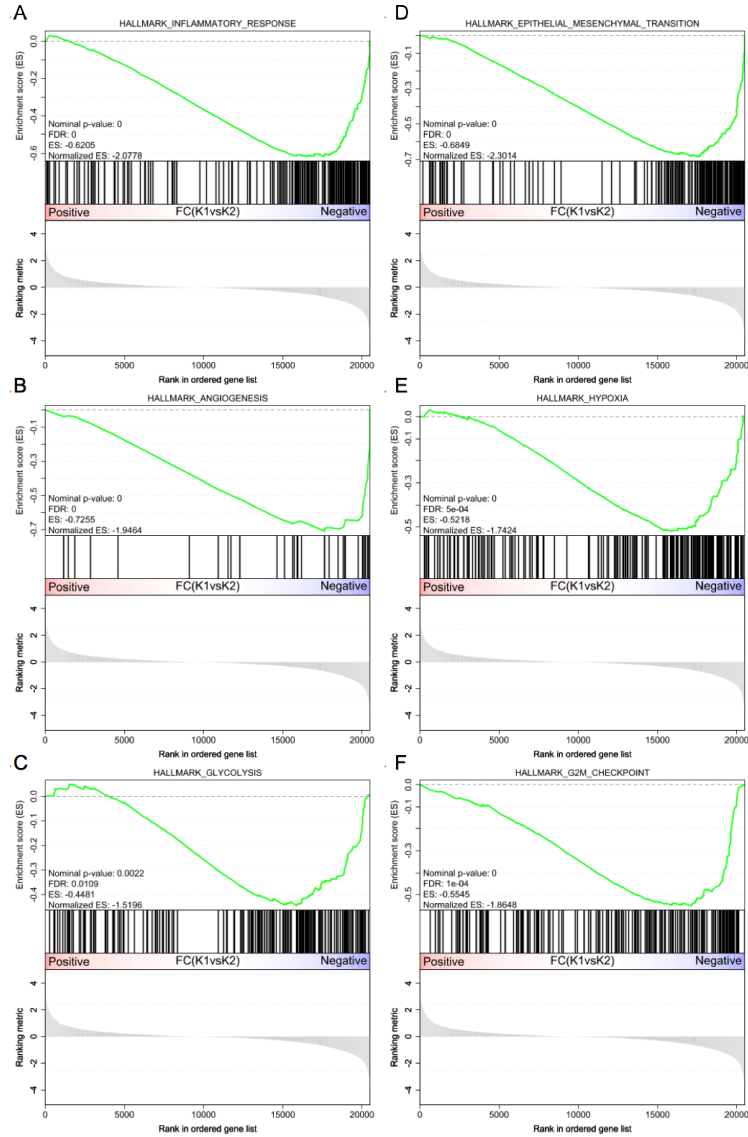


Figure 10: **GSEA results from K1 and K2 differentially expressed genes.** GSEA plots of the most differentially enriched gene sets between the two cohorts. A) Inflammatory response, B) Angiogenesis, C) Glycolysis, D) EMT, E) Hypoxia and F) G2M checkpoint. In each panel the following informations are shown: upper panel, green line representing enrichment statistic, middle panel, black horizontal lines indicating the position of gene set genes respect to the expression profile (lower plot, genes ordered by ranking metric,  $\log_2(\text{Fc})$ ). ES = enrichment score.

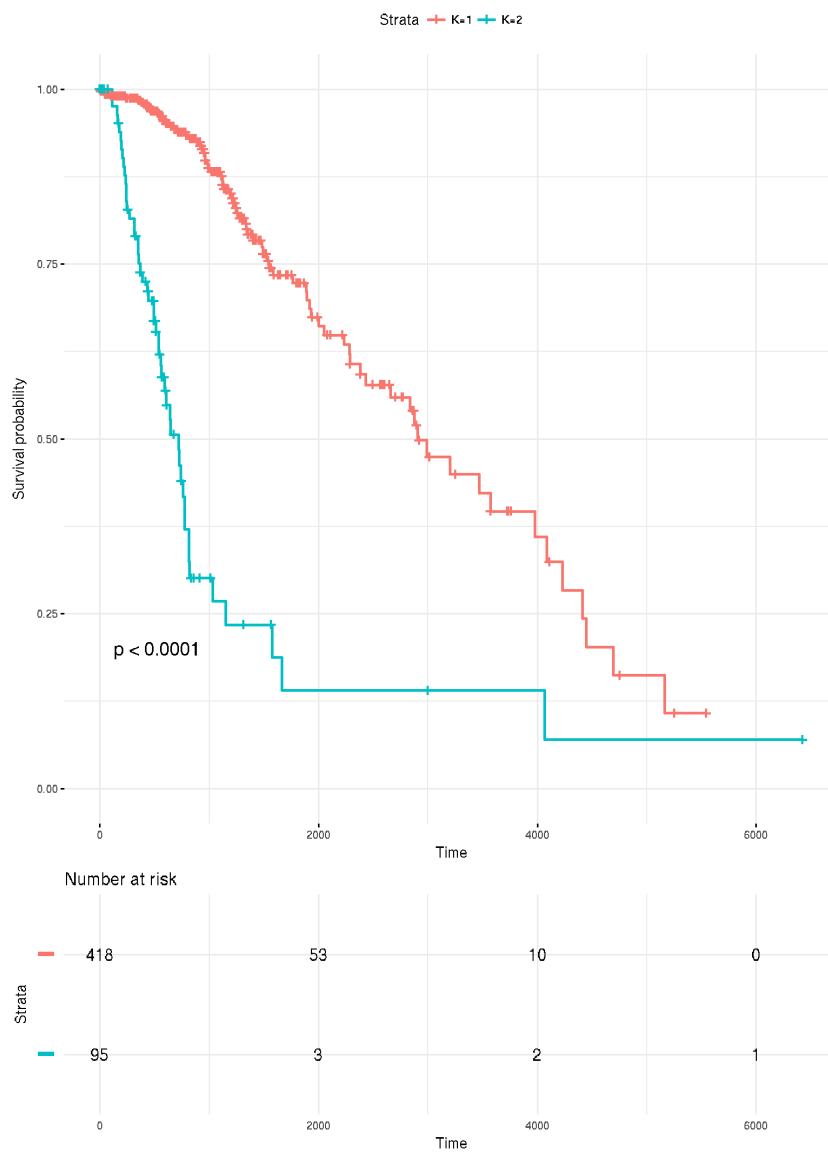


Figure 11: **Kaplan-Meier plot of the two clusters in LGG.** This plot shows the difference in survival between the two clusters generated by the unsaturated fatty acid metabolism in LGG. Cluster 1 (K1) shows a significantly better prognosis compared to cluster 2 (K2) samples.

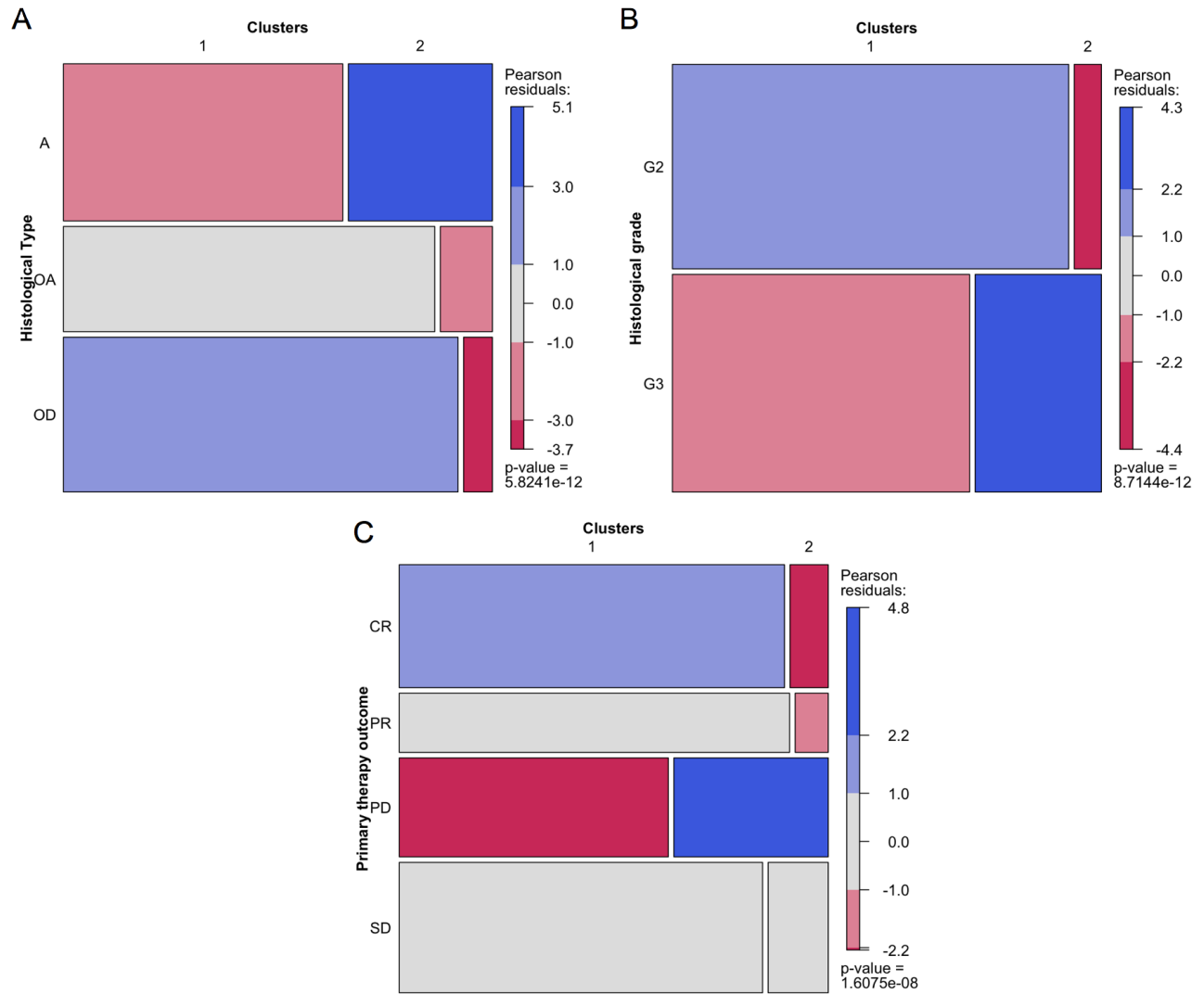


Figure 12: **Difference in CDEs distribution between LGG clusters.** Mosaic plot of three different clinical elements in LGG cohort. A) Histological type (A = Astrocytoma, OA = Oligoastrocytoma and OD = Oligodendrocytoma), B) Histological grade (G2 = Grade 2 and G3 = grade 3) and C) Primary therapy outcome (CR = complete response, PR = partial response, PD = progressive disease, SD = stable disease). Pearson residuals are given by  $\frac{(observed - expected)}{\sqrt{expected}}$ .

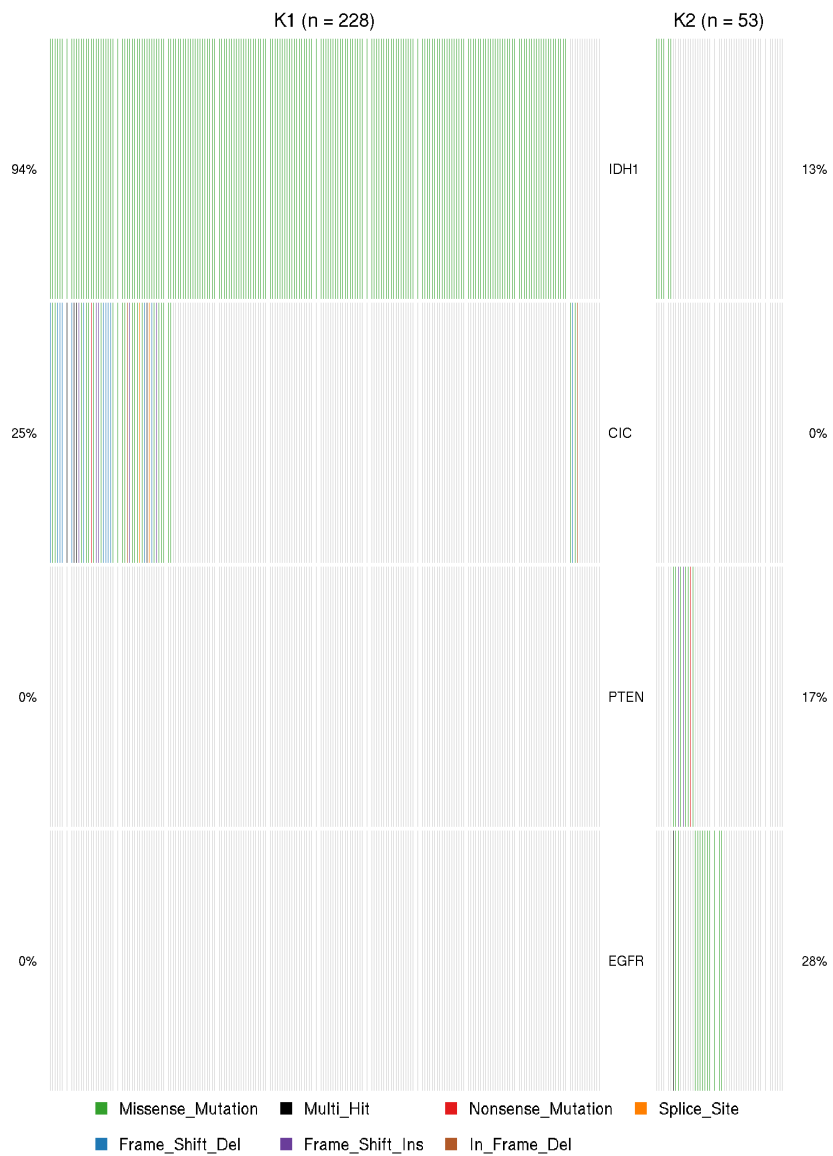


Figure 13: **Difference in mutations distribution between the two clusters in LGG cohort.** Oncoplot representing the different mutation occurrence and typology between the clusters generated by unsaturated fatty acid metabolism in LGG.



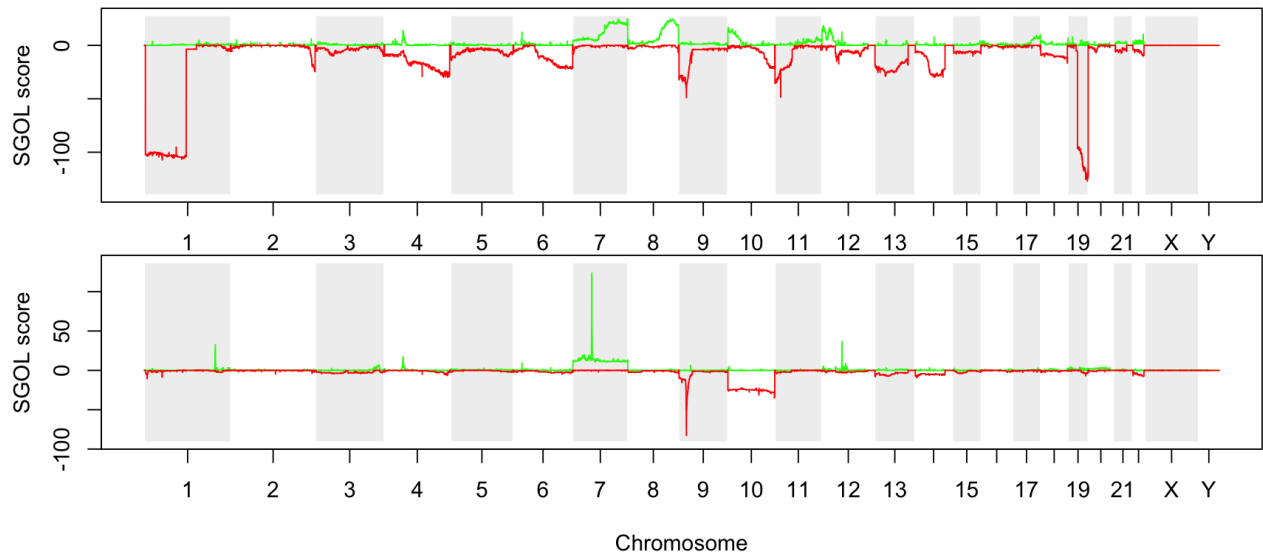


Figure 14: **CNAs landscape in LGG clusters.** SGOL plot of the different regions of amplification (green) and deletion (red) in K1 (upper panel) and K2 (lower panel). On the y-axis SGOL scores are shown, while on the x-axis the chromosomes are shown.



Figure 15: **Regions of copy number gains or losses at the cytogenetic band level.** Heatmap representing different chromosomal regions (rows), showing differential chromosomal cytogenetic band abundance in the samples of the two clusters (columns). Colour bar on the rows represents chromosomes arms and column colours bar identify samples from cluster 1 and cluster 2. Colours ranging from red (gain) to blue (loss) represent Z scores of SGOL scores between the two clusters.

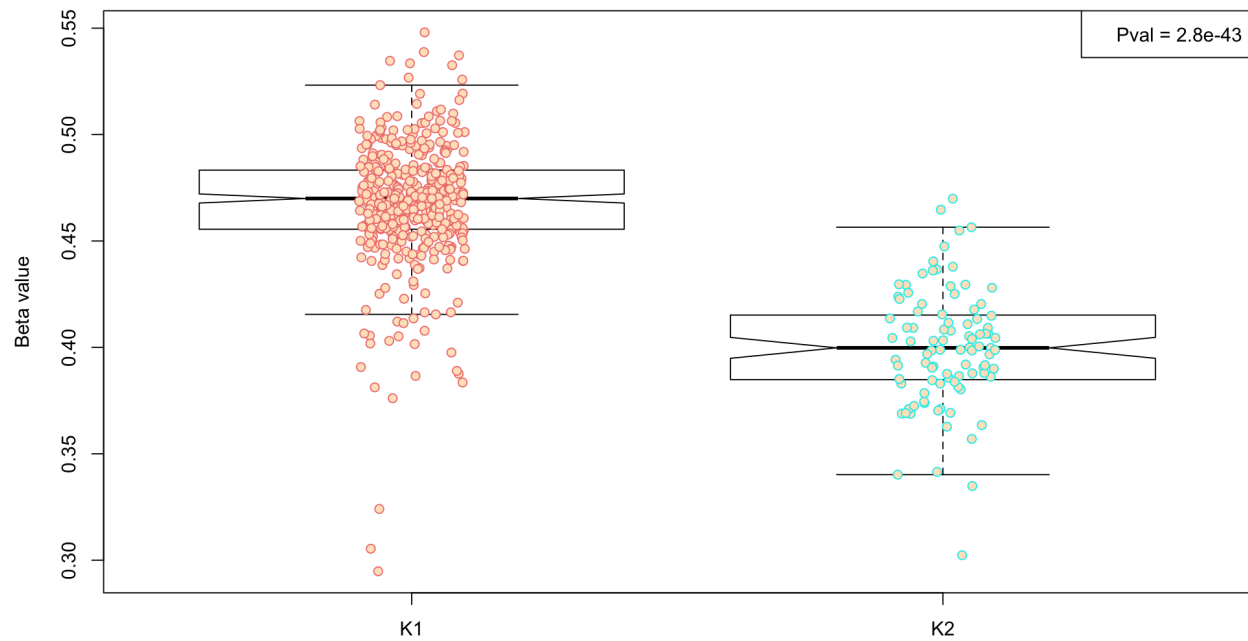


Figure 16: **Difference in all genes methylation levels between the two clusters.** Box plot representing the difference in all genes methylation level between cluster 1 and cluster 2 samples. The notch in the boxes indicates 95% confidence interval for the medians.



Figure 17: **Difference in protein abundance level between samples between the two clusters.** Heatmap representing differentially abundance in proteins level (rows) between cluster 1 and cluster 2 samples (columns). Colours ranging from red (upregulated) to green (downregulated), represent proteins signal Z scores.

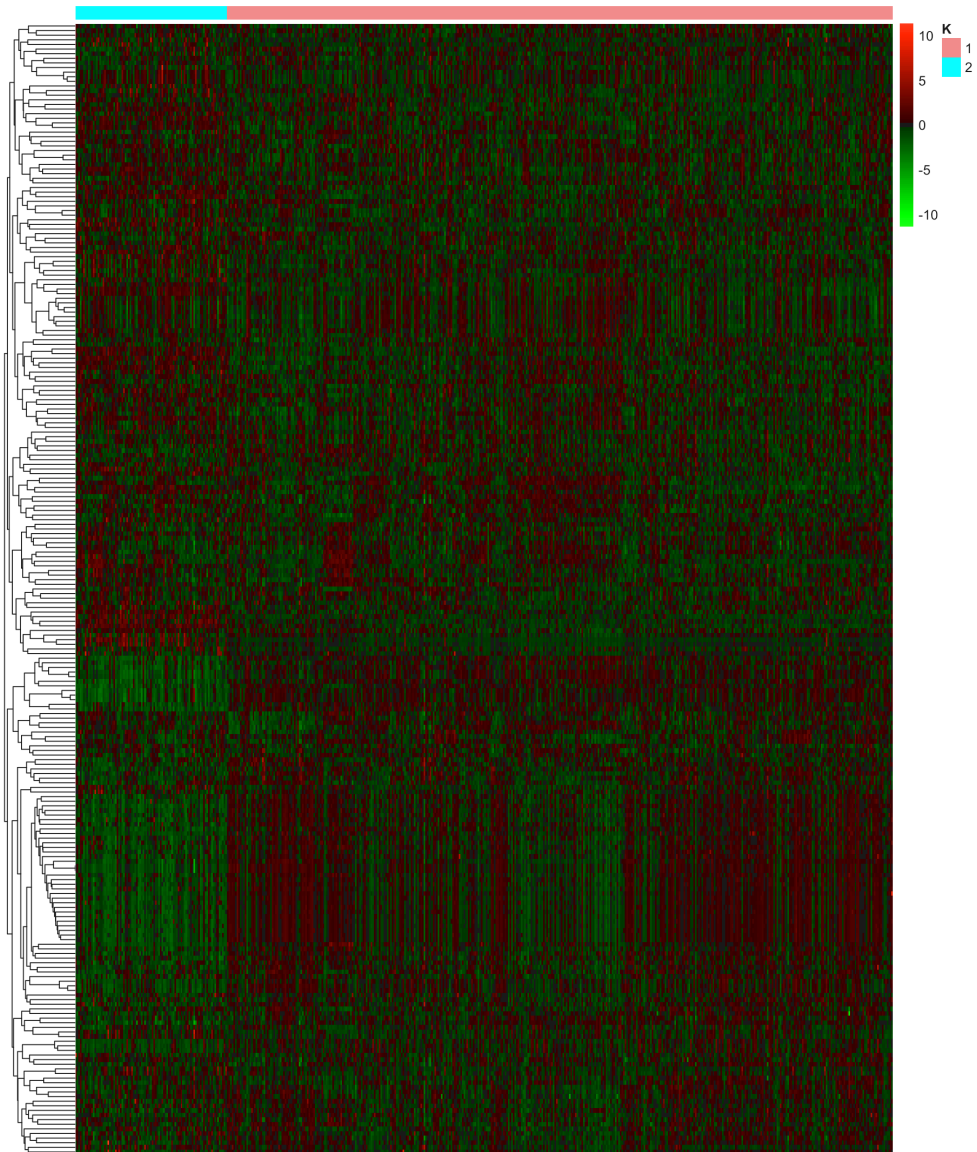


Figure 18: **Difference in miRNAs expression levels between the two clusters.** Heatmap of differentially expressed miRNAs (rows), at Bonferroni corrected pvalue  $< 0.05$ , between the two clusters samples (columns). The colour bar above heatmap columns represents samples from the two clusters. Colours ranging from red (upregulated) to green (downregulated), represent miRNAs expression Z scores.

	logFC	AveExpr	t	P.Value	adj.P.Val	B
<b>hsa-mir-767</b>	1.431	5.318	10.58	8.778e-24	2.24e-22	41.81
<b>hsa-mir-105-2</b>	1.292	4.974	9.883	3.431e-21	7.325e-20	35.89
<b>hsa-mir-105-1</b>	1.262	4.959	9.437	1.359e-19	2.787e-18	32.24
<b>hsa-mir-182</b>	1.204	11.84	5.906	6.408e-09	3.989e-08	8.028
<b>hsa-mir-183</b>	1.164	10.04	5.459	7.503e-08	4.242e-07	5.633
<b>hsa-mir-96</b>	1.106	2.905	6.381	3.951e-10	2.973e-09	10.75
<b>hsa-mir-1296</b>	1.07	4.97	8.896	1.009e-17	1.552e-16	27.98
<b>hsa-mir-139</b>	1.02	7.987	6.237	9.391e-10	6.505e-09	9.903
<b>hsa-mir-219-2</b>	1.018	11.33	6.135	1.707e-09	1.144e-08	9.32
<b>hsa-mir-128-1</b>	0.9753	9.516	7.006	7.789e-12	6.904e-11	14.6
<b>hsa-mir-221</b>	-1.661	5.355	-10.81	1.127e-24	3.413e-23	43.85
<b>hsa-mir-34a</b>	-1.814	5.849	-15.55	6.333e-45	1.104e-42	90.29
<b>hsa-mir-204</b>	-1.893	5.512	-9.104	1.956e-18	3.41e-17	29.6
<b>hsa-mir-10b</b>	-1.963	11.81	-5.643	2.776e-08	1.659e-07	6.6
<b>hsa-mir-21</b>	-2.053	13.59	-13.15	3.288e-34	2.023e-32	65.69
<b>hsa-mir-148a</b>	-2.46	9.753	-17.23	9.431e-53	2.466e-50	108.3
<b>hsa-mir-155</b>	-2.506	4.287	-22.78	8.776e-80	9.18e-77	170.5
<b>hsa-mir-196a-1</b>	-2.879	0.7985	-19.36	5.212e-63	2.726e-60	131.9
<b>hsa-mir-196b</b>	-3.267	2.49	-13.86	2.669e-37	2.327e-35	72.78
<b>hsa-mir-10a</b>	-3.605	6.293	-18.11	5.56e-57	1.939e-54	118

Table 5: **Summary of the differentially expressed miRNAs in LGG.** This table represents the 20 most differentially expressed miRNAs between K1 and K2 cohorts, sorted by expression fold change. *logFC* is log2 fold change of miRNA expression in K1 vs K2, *AveExpr* average miRNA log2 expression, *t* is the moderated t-statistic, *P.Value* is the raw pvalue of differential expression, *adj.P.Value* is BH adjusted (FDR) adjusted p-value and B\* is log-odds that the miRNA is differentially expressed.

### 5.2.2 Hexose Metabolism in BRCA

Below are summarized results of the impact of the hexose metabolism gene set, **GO:0019318**, on BRCA. This gene set contains genes whose expression is linked to the chemical reactions and pathways involving a hexose, any monosaccharide with a chain of six carbon atoms.

The expression profile of hexose metabolism genes divides the BRCA sample cohort in two distinct clusters (K1 and K2). The differences in gene expression can be appreciated in Figure 19.

GSE analysis performed on differentially expressed genes between the two clusters reveals a reduction in, among the others:

- A. estrogen response genes
- B. epithelial to mesenchymal transition
- C. KRAS signalling
- D. adipogenesis

and an increase in:

- E. proliferation
- F. MYC target genes
- G. MTORC1 signalling
- H. oxidative phosphorylation
- I. glycolysis

in K1 samples. Figure 20

From a clinical perspective the patients belonging to cluster 1 (K1) have a worse prognosis in respect to those of cluster 2 (K2) Figure 21. As expected differences between histological subtypes are present, interestingly differences in pathologic T (tumor score) and in StGallen classification are also observed (Figure 22).

For what concerns differences in the mutation landscape of the two sample cohorts, we can observe a higher frequency of TP53 (mainly R175H and R248W) in K1 samples and an higher frequency of CDH1 mutations in K2 (Figure 23). Both TP53 observed mutations have been reported to impair DNA binding, by the introduction of a large hydrophobic side chain (R248W mutation) or by disrupting the zinc binding domain, as it happens in R175H, the most frequent hotspot mutant (Olivier, Hollstein, and Hainaut 2010). CDH1 mutations are cancer predisposing mutations in hereditary diffuse gastric cancer (HDGC) (Hansford et al. 2015). Consistently with literature, a higher CDH1 mutation rate is observed in invasive lobular carcinoma (ILC) (Dossus and Benusiglio 2015).

In Figure 24 the CNA landscape of the two clusters is shown. Although there are no striking qualitative differences in regions of gain and losses between K1 (upper plot) and K2 (lower plot) samples, exception made for chr3q gain in K1 samples, quantitative differences can be appreciated. Notably K1 shows higher grade of gains and losses (see graphs y-axes range and Figure 25) such as chr8q gains and chr8p and chr13q losses.

Methylation analysis Figure 26 shown no difference in all genes methylation between the two.

More striking differences arise from RPPA data shown in Figure 27. K1 samples have a clearly higher proliferating signature i.e. higher phosphorylation of CDKs (cycle dependent kinases), higher amount of cyclines (CCNE1/2, CCNB1) and cell cycle checkpoint genes (CHECK1/2), consistently with GSEA results on proliferation, and a reduction in apoptotic genes such as BCL2, ANXA7. Moreover K1 samples show a hyper activation of NFKB, whose implication in breast cancer metastasis has been extensively described ((Helbig et al. 2003) and (M. A. Huber et al. 2004)).

The miRNA analysis identifies the oncomir, has-mir-210 as the most upregulated one, consistently with the literature (Camps et al. 2008). On the other hand the tumor suppressor hsa-mir-139 (Krishnan et al. 2013) is the most downregulated.

Bottom line the results obtained by the subdivision of the BRCA cohort on expression profile of hexose metabolic genes recapitulates known features of breast cancer tumorigenesis.



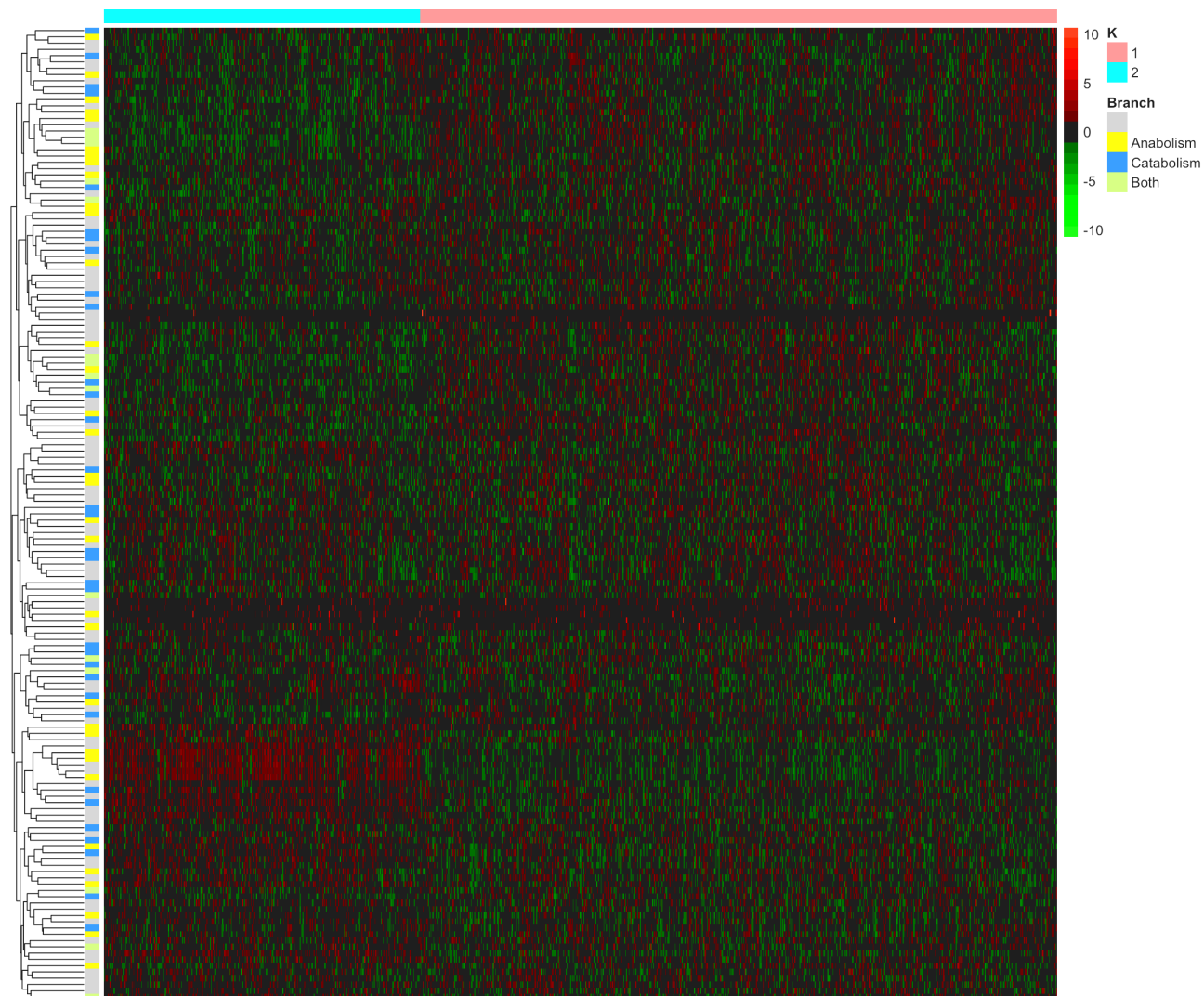


Figure 19: **Differentially expressed metabolic genes in BRCA.** Heatmap representing the expression of the genes of the hexose metabolism (rows) in BRCA cohort. The colour bar above heatmap columns (samples) represents samples from the two clusters. The samples have been subdivided as resulted from PAM algorithm. The colour bar on the rows of the heatmap highlights genes belonging to anabolism (yellow), catabolism (blue) and enzymes catalysing reversible reactions (green). Colours ranging from red (upregulated) to green (downregulated), represent genes expression Z scores.

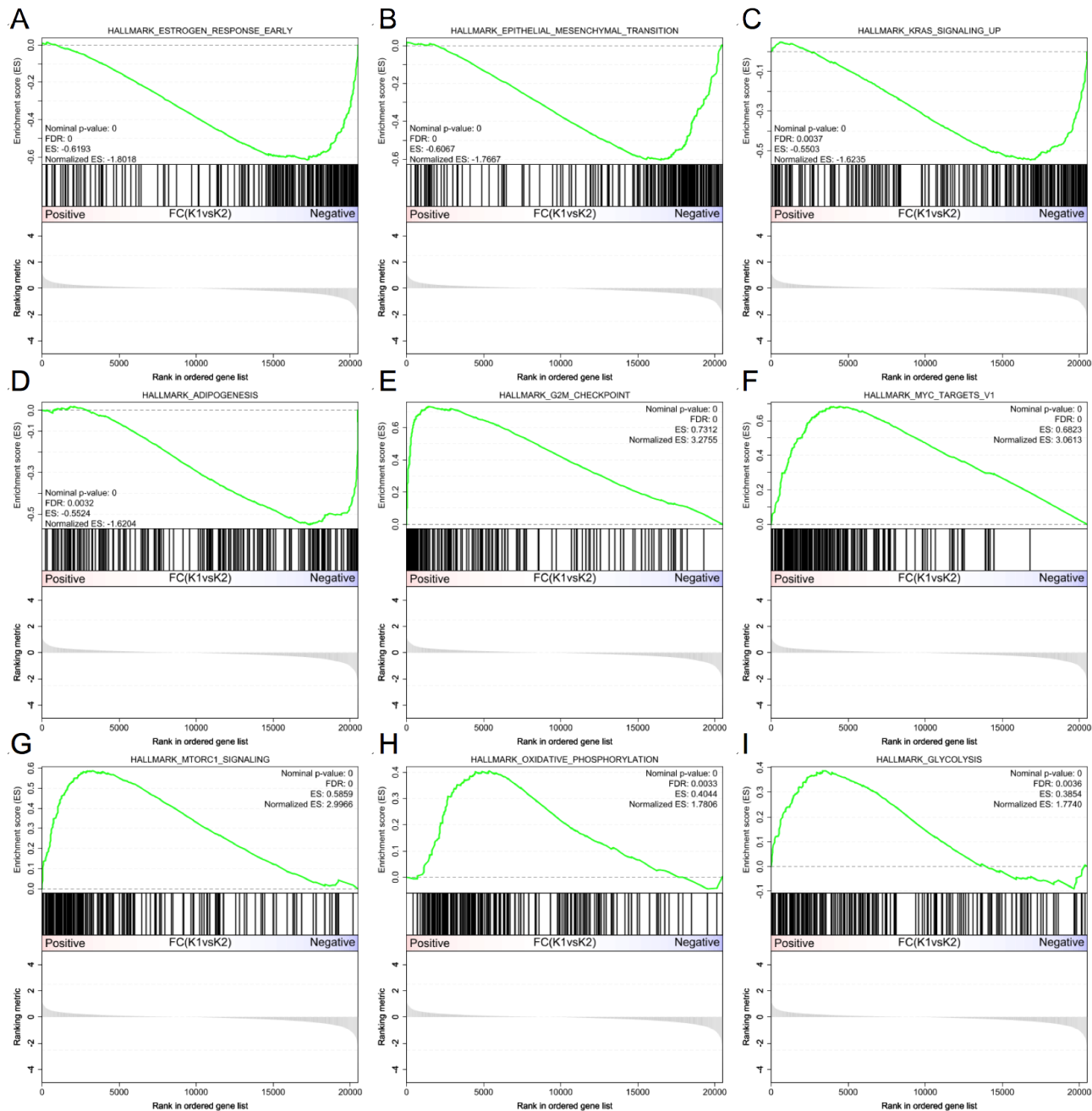


Figure 20: **GSEA results from K1 and K2 differentially expressed genes.** GSEA plots of the most differentially enriched gene sets between the two cohorts. A) Estrogen response, B) EMT, C) KRAS signalling, D) Adipogenesis, E) G2M checkpoint, F) MYC targets, G) MTORC1 signalling, H) Oxydative phosphorylation and I) Glycolysis . In each panel the following informations are shown: upper panel, green line representing enrichment statistic, middle panel, black horizontal lines indicating the position of gene set genes respect to the expression profile (lower plot, genes ordered by ranking metric,  $\log_2(Fc)$ ). ES = enrichment score.

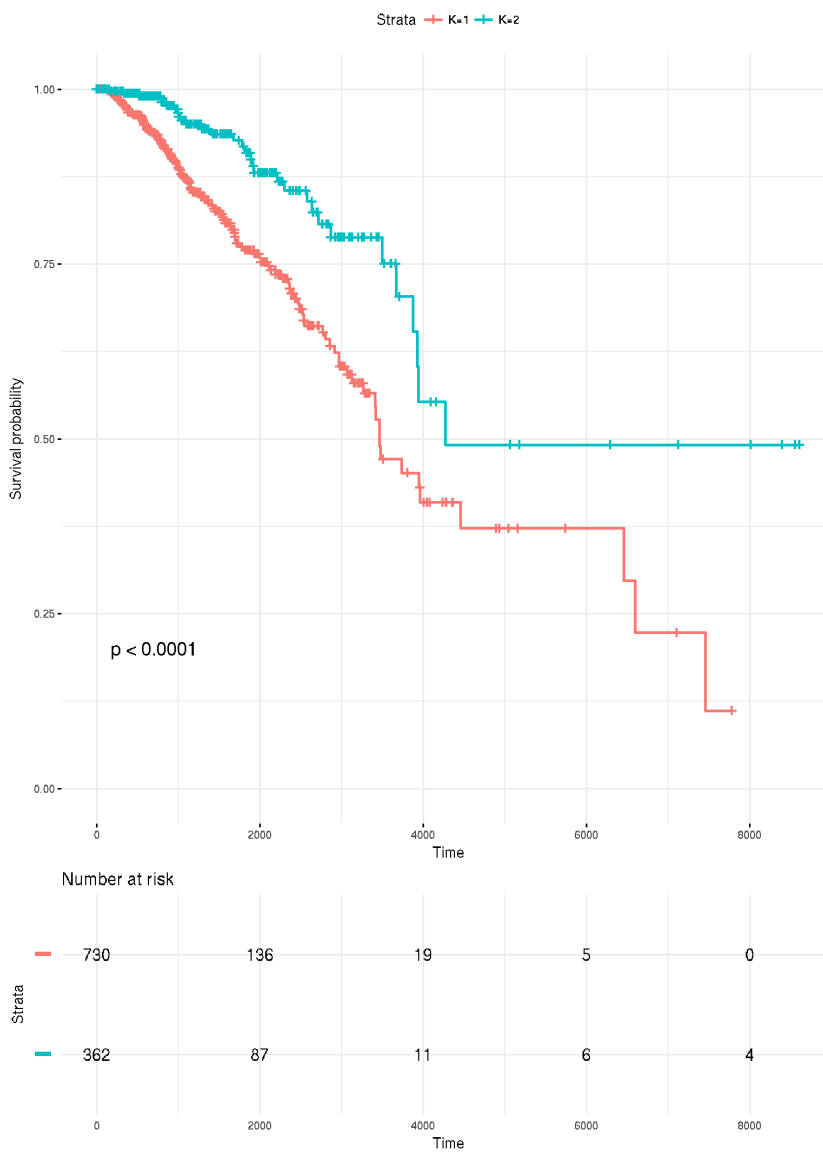


Figure 21: **Kaplan-Meier plot of the two clusters in BRCA** This plot shows the difference in survival between the two clusters generated by the hexose metabolism in BRCA Cluster 1 (K1) show a significantly worse prognosis compared to cluster 2 (K2) samples.

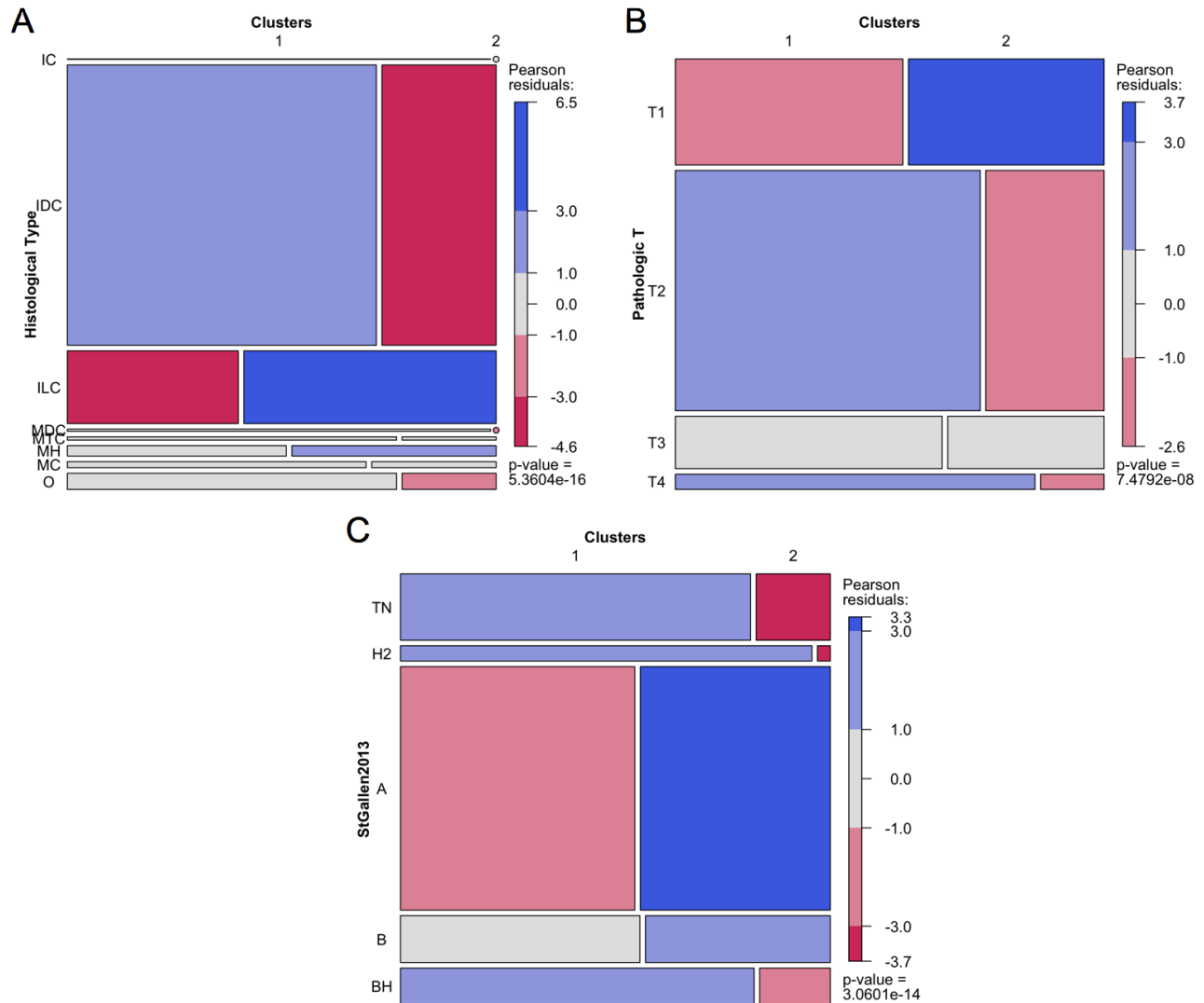


Figure 22: **Difference in CDEs distribution between LGG clusters.** Mosaic plot of three different clinical elements in LGG cohort. A) Histological type (IC = Infiltrating carcinoma, IDC = Infiltrating ductal carcinoma, ILB = Infiltrating lobular carcinoma, MDC = Medullary carcinomam, MTC =Metaplastic carcinoma, MHC = Mixed histology, MC = Mucinous carcinoma and O = Other), B) Pathologic T (from T1 to T4, grade of tumor burder) and C) StGallen classification (TN = triple negative, H2 = HER2 +, A = Luminal A, B = Luminal B and BH = Luminal B HER2 +). Pearson residuals are given by  $\frac{(observed - expected)}{\sqrt{expected}}$ .

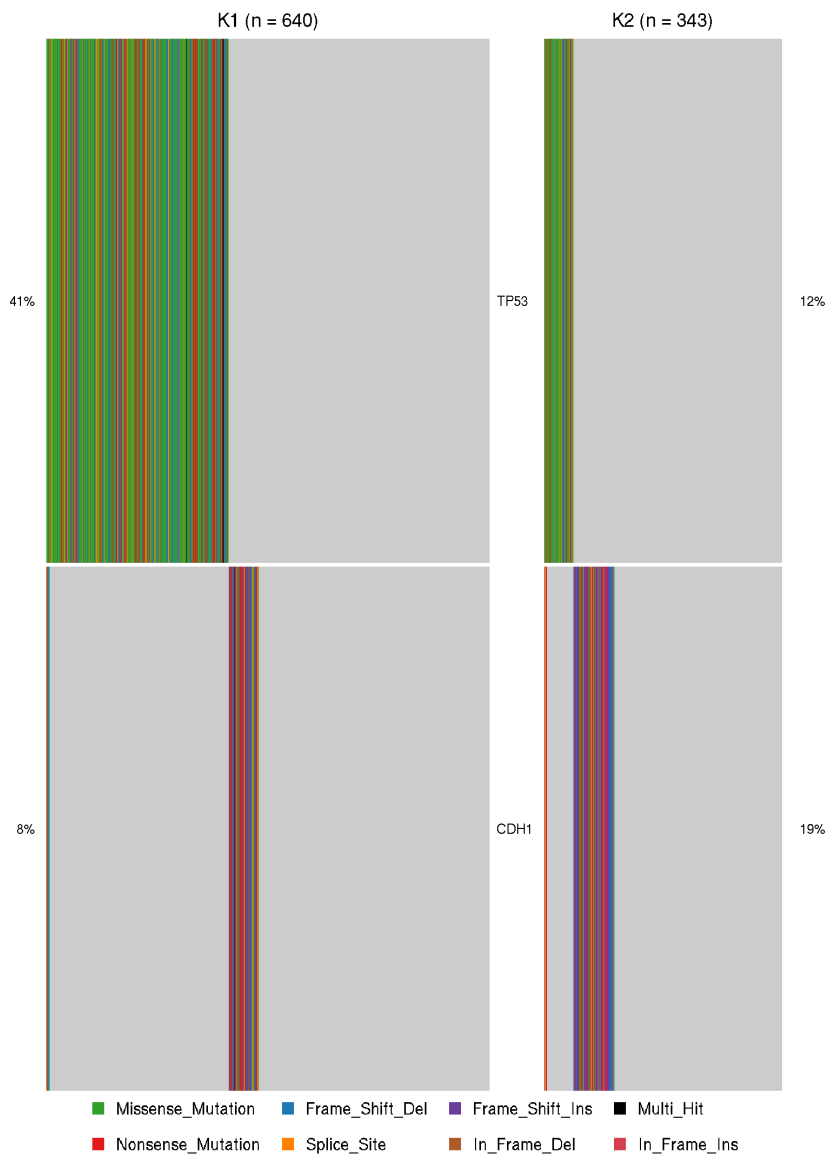


Figure 23: **Difference in mutations distribution between the two clusters in BRCA cohort.** Oncoplot representing the different mutations occurrence and typology between the clusters generated by hexose metabolism in BRCA.

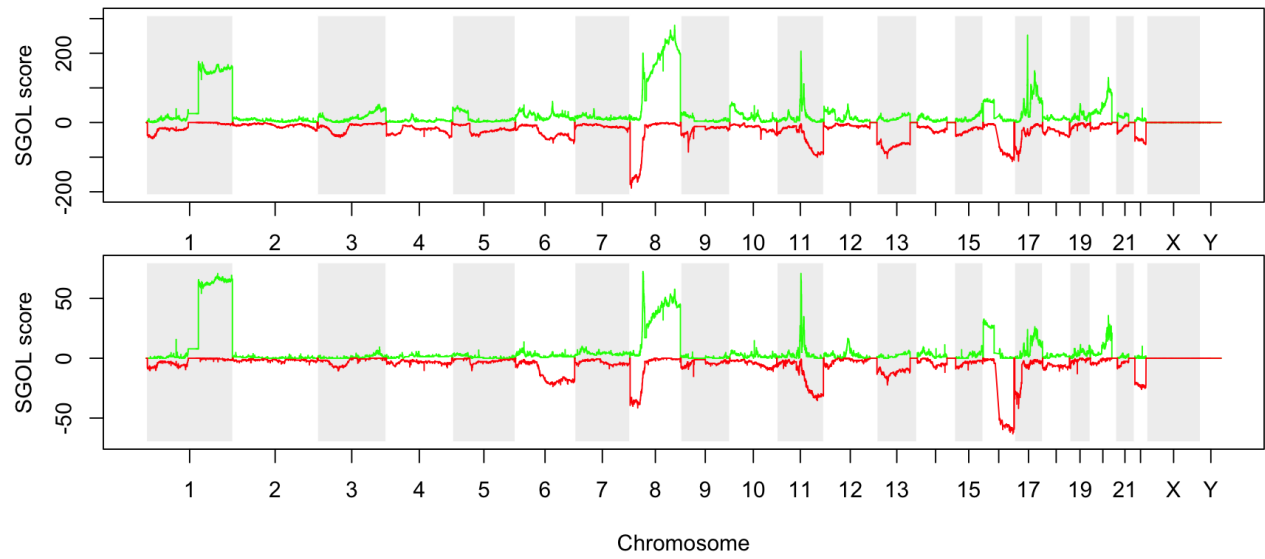


Figure 24: **CNAs landscape in BRCA clusters.** SGOL plot of the different regions of amplification (green) and deletion (red) in K1 (upper panel) and K2 (lower panel). On the y-axis SGOL scores are shown, while on the x-axis the chromosomes are shown.

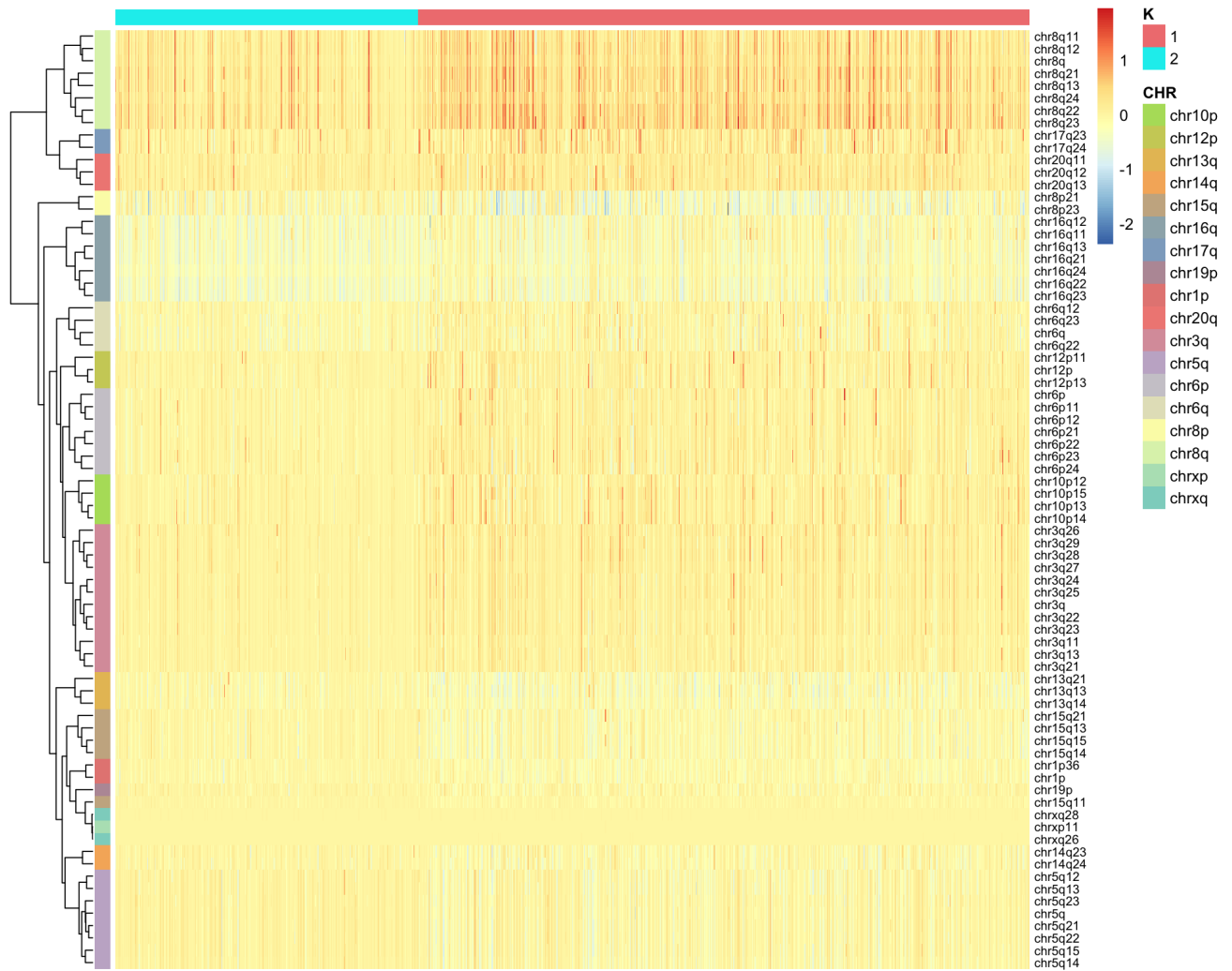


Figure 25: **Regions of copy number gains or losses at the cytogenetic band level.** Heatmap representing different chromosomal regions (rows), showing differential abundance in the samples of the two clusters (columns). Colour bar on the rows represents chromosomes arms and column colours bar identify samples from cluster 1 and cluster 2. Colours ranging from red (gain) to blue (loss) represent Z scores of SGOL scores between the two clusters.

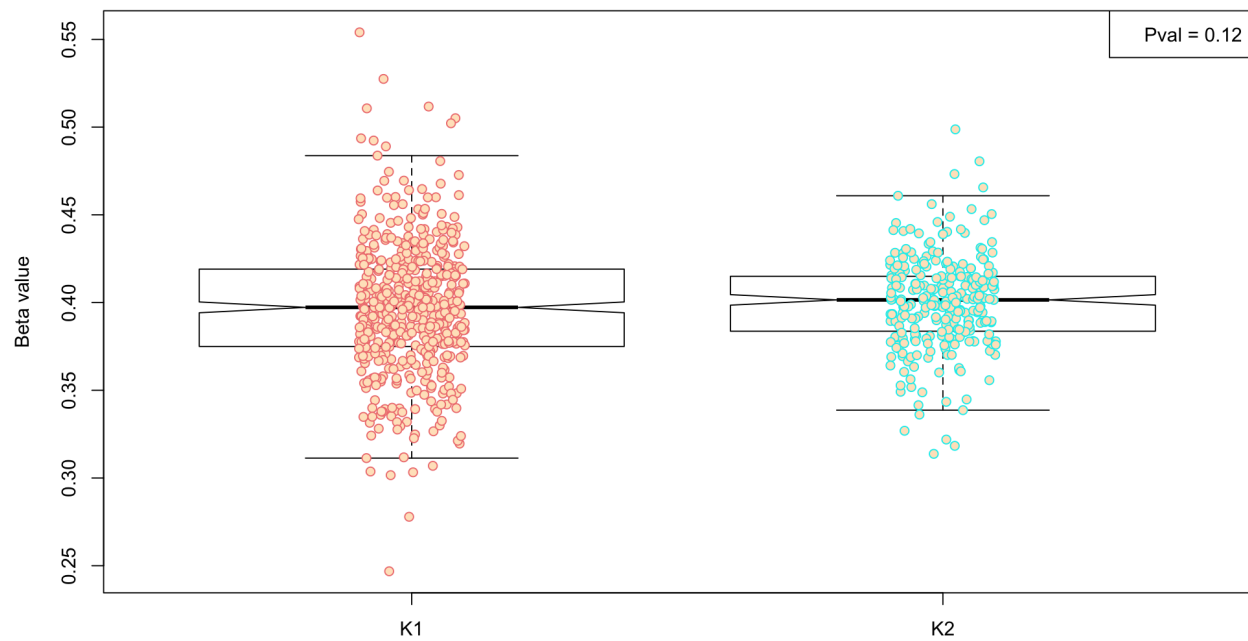


Figure 26: **Difference in all genes methylation levels between the two clusters.** Box plot representing the difference in all genes methylation level between cluster 1 and cluster 2 samples.





Figure 27: **Difference in protein abundance level between samples between the two clusters.** Heatmap representing differentially abundance in proteins level (rows) between cluster 1 and cluster 2 samples (columns).

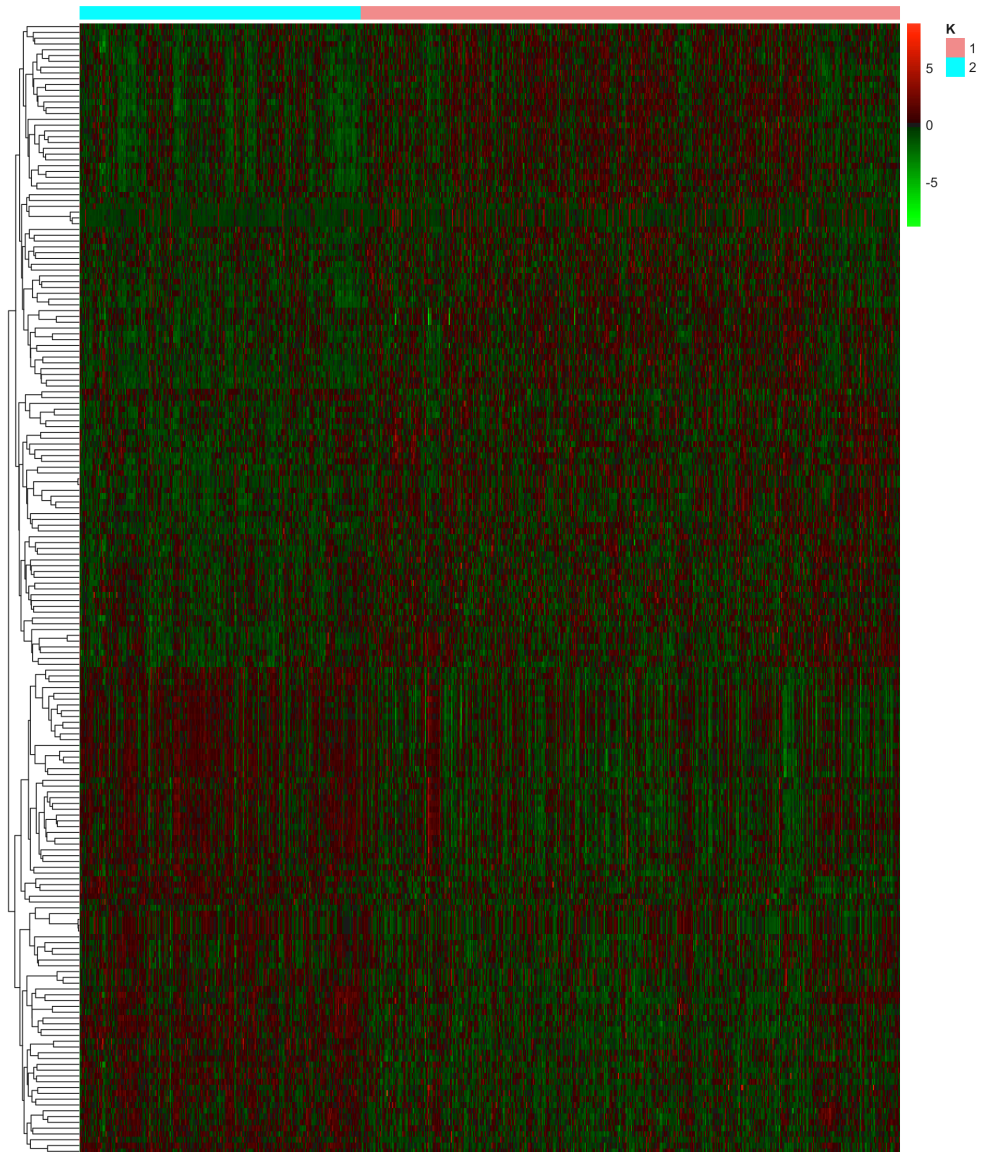


Figure 28: **Difference in miRNAs expression levels between the two clusters.** Heatmap of differentially expressed miRNAs (rows), at Bonferroni corrected pvalue  $< 0.05$ , between the two clusters samples (columns). The colour bar above heatmap columns represents samples from the two clusters. Colours ranging from red (upregulated) to green (downregulated), represent miRNAs expression Z scores

	logFC	AveExpr	t	P.Value	adj.P.Val	B
<b>hsa-mir-210</b>	1.676	8.21	10.82	1.867e-25	6.511e-23	45.87
<b>hsa-mir-1269</b>	1.046	2.062	4.681	3.39e-06	2.749e-05	2.23
<b>hsa-mir-9-2</b>	0.9182	8.735	5.477	5.904e-08	7.817e-07	6.142
<b>hsa-mir-9-1</b>	0.9089	8.739	5.428	7.684e-08	9.923e-07	5.886
<b>hsa-mir-105-2</b>	0.8454	1.094	5.482	5.735e-08	7.691e-07	6.17
<b>hsa-mir-1307</b>	0.8222	10.28	11.28	2.209e-27	1.156e-24	50.28
<b>hsa-mir-105-1</b>	0.8095	1.12	5.255	1.926e-07	2.166e-06	4.996
<b>hsa-mir-301a</b>	0.8019	3.637	7.729	3.47e-14	2.135e-12	20.19
<b>hsa-mir-3200</b>	0.7999	2.036	8.432	1.728e-16	1.643e-14	25.42
<b>hsa-mir-3677</b>	0.7648	3.337	7.261	9.634e-13	4.381e-11	16.92
<b>hsa-mir-30a</b>	-0.631	15.42	-6.113	1.575e-09	2.745e-08	9.67
<b>hsa-mir-150</b>	-0.6448	9.007	-4.637	4.174e-06	3.234e-05	2.03
<b>hsa-mir-34c</b>	-0.6894	3.415	-8.266	6.282e-16	5.476e-14	24.15
<b>hsa-mir-195</b>	-0.6997	5.369	-10.66	8.276e-25	2.164e-22	44.39
<b>hsa-mir-483</b>	-0.7035	2.289	-7.38	4.191e-13	1.993e-11	17.74
<b>hsa-mir-99a</b>	-0.8019	9.191	-9.048	1.228e-18	1.605e-16	30.31
<b>hsa-let-7c</b>	-0.8031	10.92	-9.723	3.971e-21	8.307e-19	35.99
<b>hsa-mir-1247</b>	-0.8209	4.206	-6.324	4.381e-10	9.165e-09	10.92
<b>hsa-mir-204</b>	-0.8668	1.674	-9.648	7.637e-21	1.331e-18	35.34
<b>hsa-mir-139</b>	-0.9843	5.361	-13.39	8.146e-37	8.521e-34	71.88

Table 6: **Summary of differentially expressed miRNAs.** The table represents the most differentially expressed miRNAs between K1 and K2 clusters *logFC* is log2 fold change of miRNA expression in K1 vs K2, *AveExpr* average miRNA log2 expression, *t* is the moderated t-statistic, *P.Value* is the raw Pvalue of differential expression, *adj.P.Value* is BH adjusted (FDR) adjusted Pvalue and *B* is log-odds that the miRNA is differentially expressed.

## 6 Discussion

This study provides a quantitative description of the metabolic heterogeneities found in different human tumor types. By means of the subdivision of cancer patients based on the metabolic gene expression profile and subsequent correlative tests, it has been possible to provide a quantitative clinical and genomic description linked to metabolic differences.

For sure, direct clustering on metabolomic data (Hakimi et al. 2016) would reduce the uncertainty generated by the impact of post-translational modification on protein function, in particular for what concerns metabolic enzymes. However the availability of metabolomic data doesn't allow the analysis of vast cohort of tumors such as the one presented in this study. Besides, nowadays transcriptome profile of tumor biopsies is more frequently adopted in clinical settings, moreover, the correlative analysis between metabolites levels and the expression levels of the genes responsible for their synthesis (data from X. Tang et al. (2014)) showed a concordance above 75% between enzymes and their reaction products (data not shown). Another advantage of using transcriptomic data is the possibility to perform flux balance analysis (FBA, Blazier and Papin (2012) and Bordbar et al. (2014)). Despite the mere correlative nature of the results, so far, and in turn the lack of any causal link between a metabolic profile and a given feature, this study provides data that can provide hypotheses that can be further experimentally validated and, in some cases, can lead to the generation of new precision therapies.

Indeed, nowadays, the development of an effective metabolic therapy will require defining the stage of tumor progression in which each pathway provides its benefit. Key step would be the establishment of therapeutic windows that target tumor metabolism considering that normal proliferating cells share similar metabolic requirements and adaptations. Moreover therapeutics targets are nominated from simple experimental models, like culture cells. It will be essential to define their context-specific roles in biologically accurate models of tumor initiation and progression. We think that this study may represent a small step toward the identification of tumors metabolic heterogeneity.

## 7 Supplementary Figures And Tables

## 7.1 Clinical Data

### 7.1.1 Overall Survival

- **Tumors\_freq**: 39.5 % (15 out of 38 tumors)
- **Tumors\_top**:

	Metabolisms (n)
<b>GBMLGG</b>	305
<b>LGG</b>	145
<b>KIRC</b>	137
<b>KIPAN</b>	79
<b>UVM</b>	53

- **Metabolisms\_freq**: 75 % (324 out of 432 metabolisms)
- **Metabolisms\_top**:

	Tumors (n)
<b>GO NUCLEOBASE CONTAINING SMALL MOLECULE METABOLIC PROCESS</b>	7
<b>GO CELLULAR MODIFIED AMINO ACID METABOLIC PROCESS</b>	6
<b>GO FATTY ACID METABOLIC PROCESS</b>	6
<b>GO GMP METABOLIC PROCESS</b>	6
<b>GO MULTICELLULAR ORGANISMAL MACROMOLECULE METABOLIC PROCESS</b>	6

Table 7: **Summary of OS results.** *Tumors\_freq/Metabolism\_freq* is the number of tumors/metabolisms where correlation between OS and any metabolism/tumor was observed. *Tumors\_top/Metabolism\_top* are the top five tumors/metabolisms sorted by number of metabolisms/tumors correlating with OS.

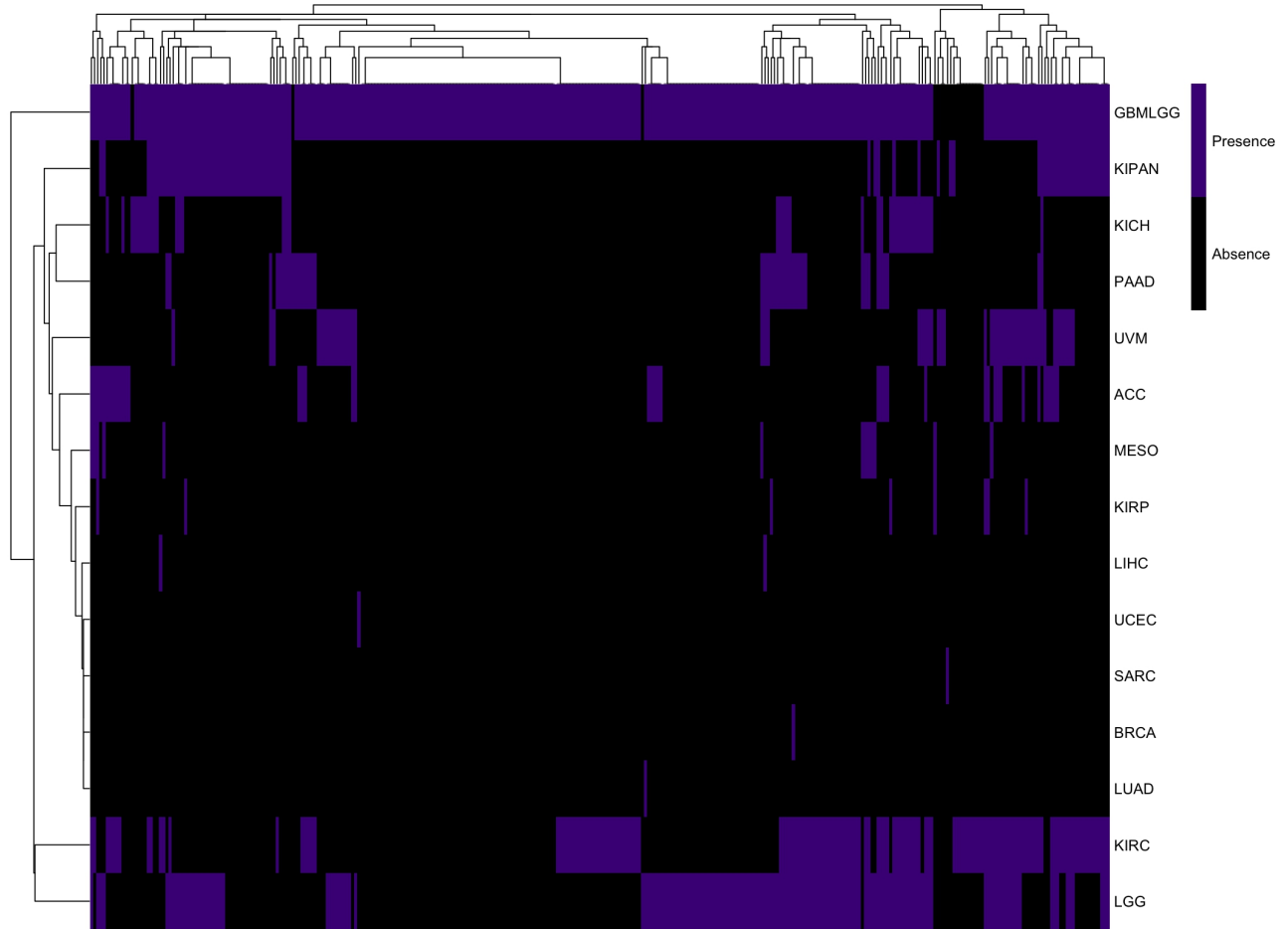


Figure 29: **Heatmap of the presence of correlation between OS and metabolism across different tumor types.** This heatmap summarizes the presence (dark violet) or absence (black) of correlation between OS and specific metabolisms (columns) in tumors cohort (rows). Statistical significance was set as a global Bonferroni corrected  $p$ val  $<.05$ .



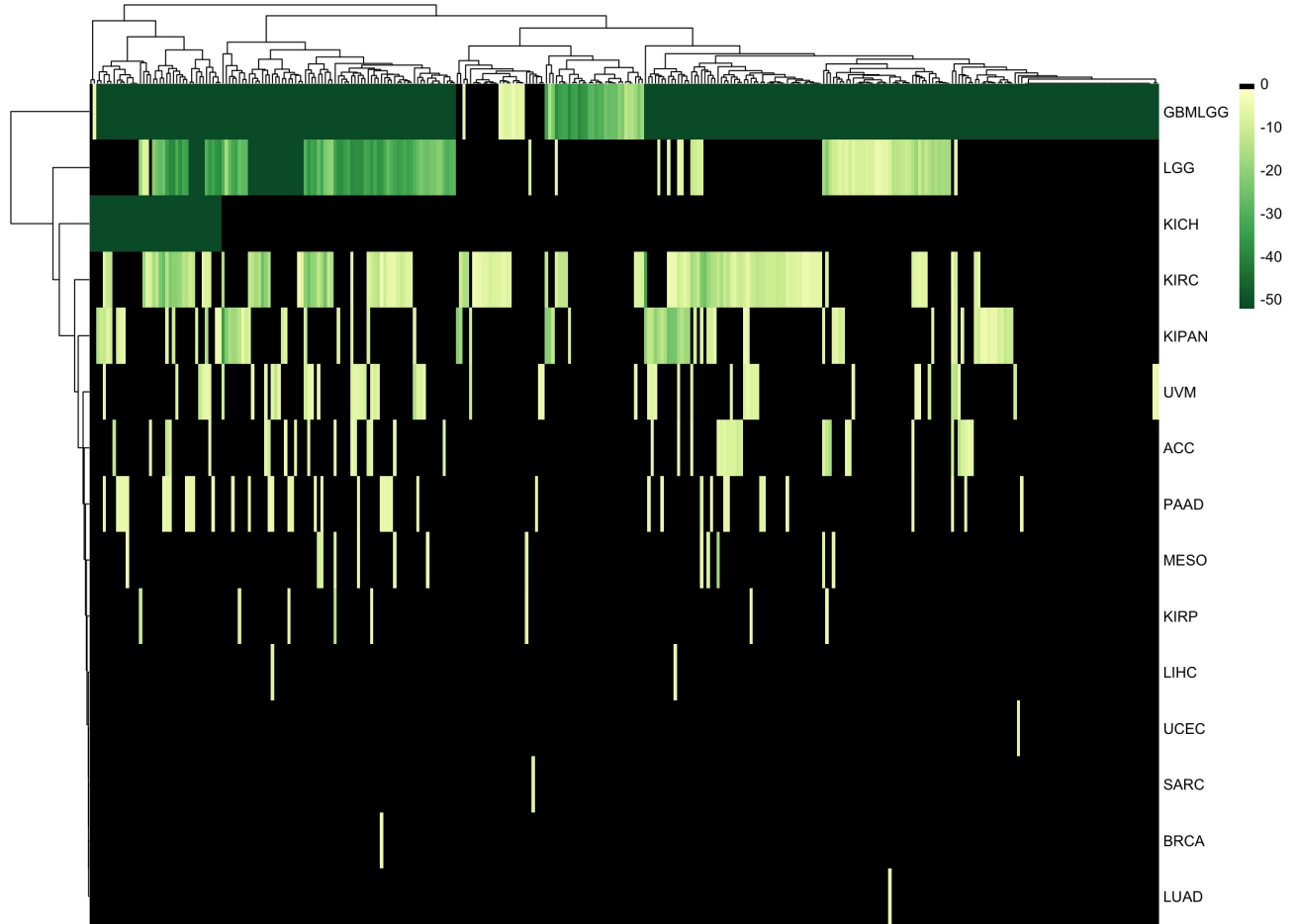


Figure 30: **Heatmap of the association strength between OS and metabolisms.** This heatmap represents the correlation of OS, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected pvalues; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

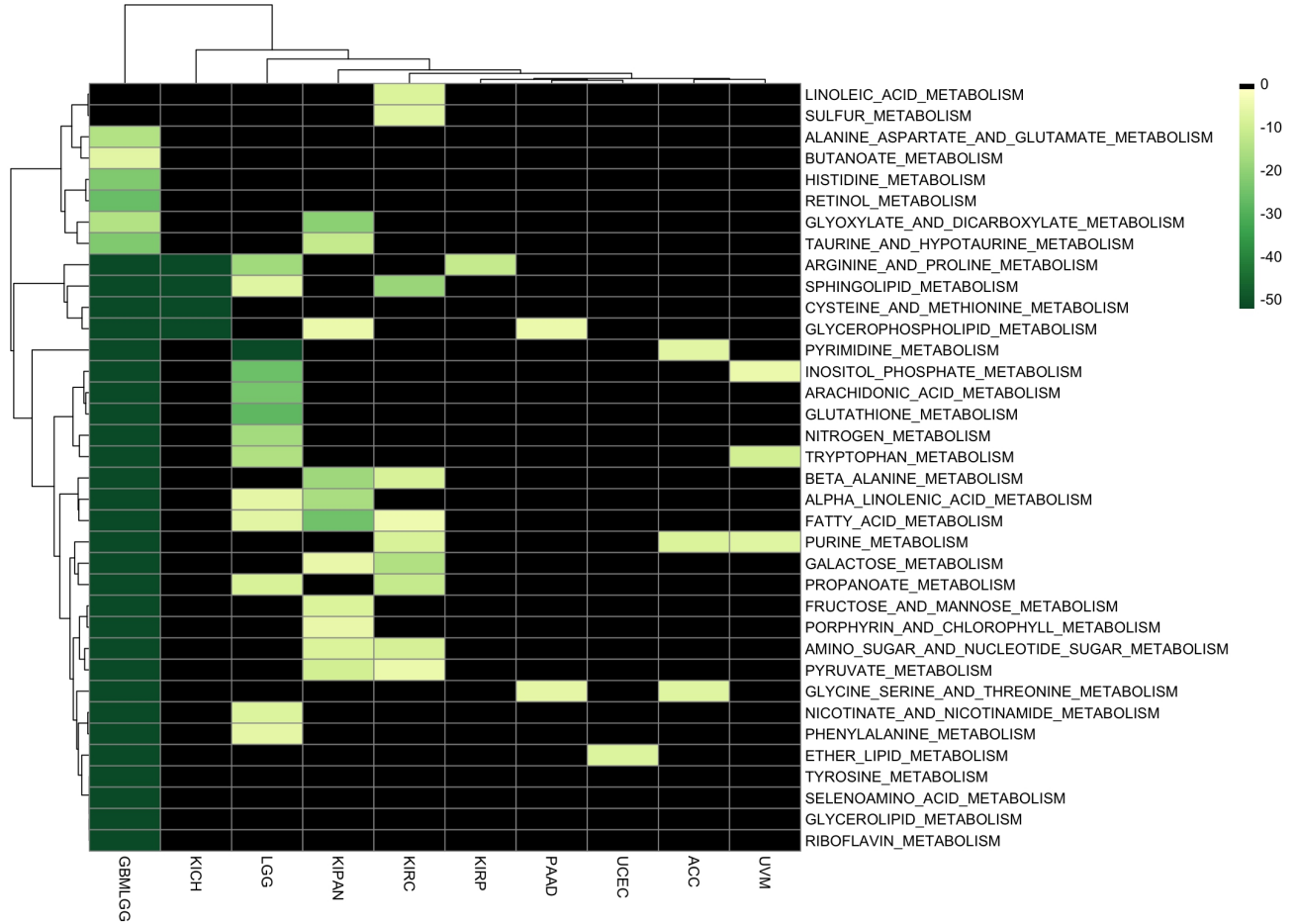


Figure 31: **Heatmap of the association strength between OS and a subset of metabolisms.** This heatmap represents the correlation of OS, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

### 7.1.2 Relapse Free Survival

- **Tumors\_freq**: 18.9 % (7 out of 37 tumors)
- **Tumors\_top**:

	Metabolisms (n)
<b>LGG</b>	102
<b>PAAD</b>	71
<b>ACC</b>	59
<b>LUNG</b>	6
<b>UVM</b>	3

- **Metabolisms\_freq**: 39.4 % (170 out of 432 metabolisms)
- **Metabolisms\_top**:

	Tumors (n)
<b>GO ALCOHOL METABOLIC PROCESS</b>	3
<b>GO ALPHA AMINO ACID METABOLIC PROCESS</b>	3
<b>GO CARBOHYDRATE DERIVATIVE METABOLIC PROCESS</b>	3
<b>GO CELLULAR LIPID METABOLIC PROCESS</b>	3
<b>GO COENZYME METABOLIC PROCESS</b>	3

Table 8: **Summary of RFS results.**  $Tumors\_freq/Metabolism\_freq$  is the number of tumors/metabolisms where correlation between RFS and any metabolism/tumor was observed.  $Tumors\_top/Metabolism\_top$  are the top five tumors/metabolisms sorted by number of metabolisms/tumors correlating with RFS.

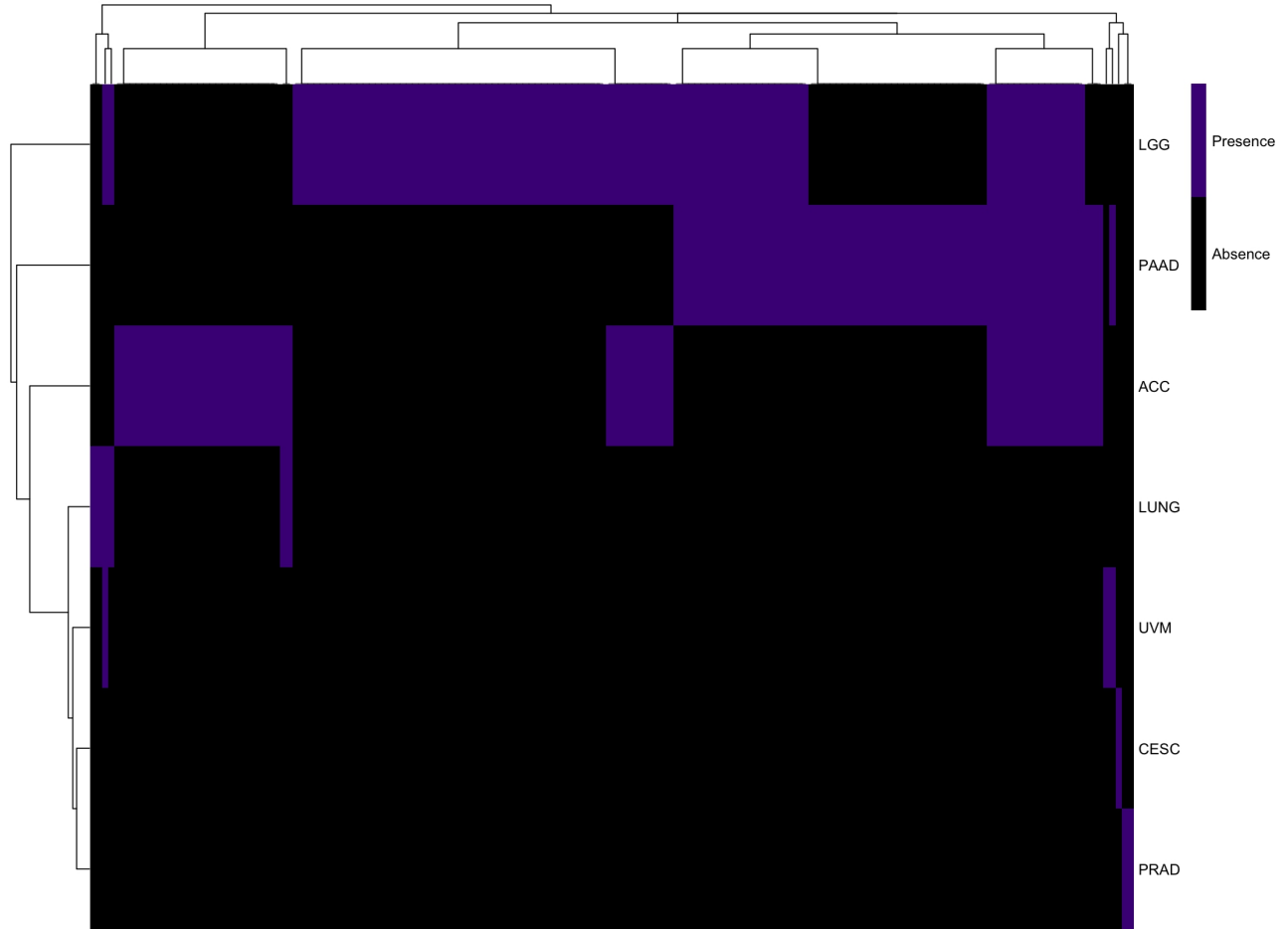


Figure 32: **Heatmap of the presence of correlation between RFS and metabolism across different tumor types.** This heatmap summarizes the presence (dark violet) or absence (black) of correlation between RFS and specific metabolisms (columns) in tumors cohort (rows). Statistical significance was set as a global Bonferroni corrected  $pval < .05$ .

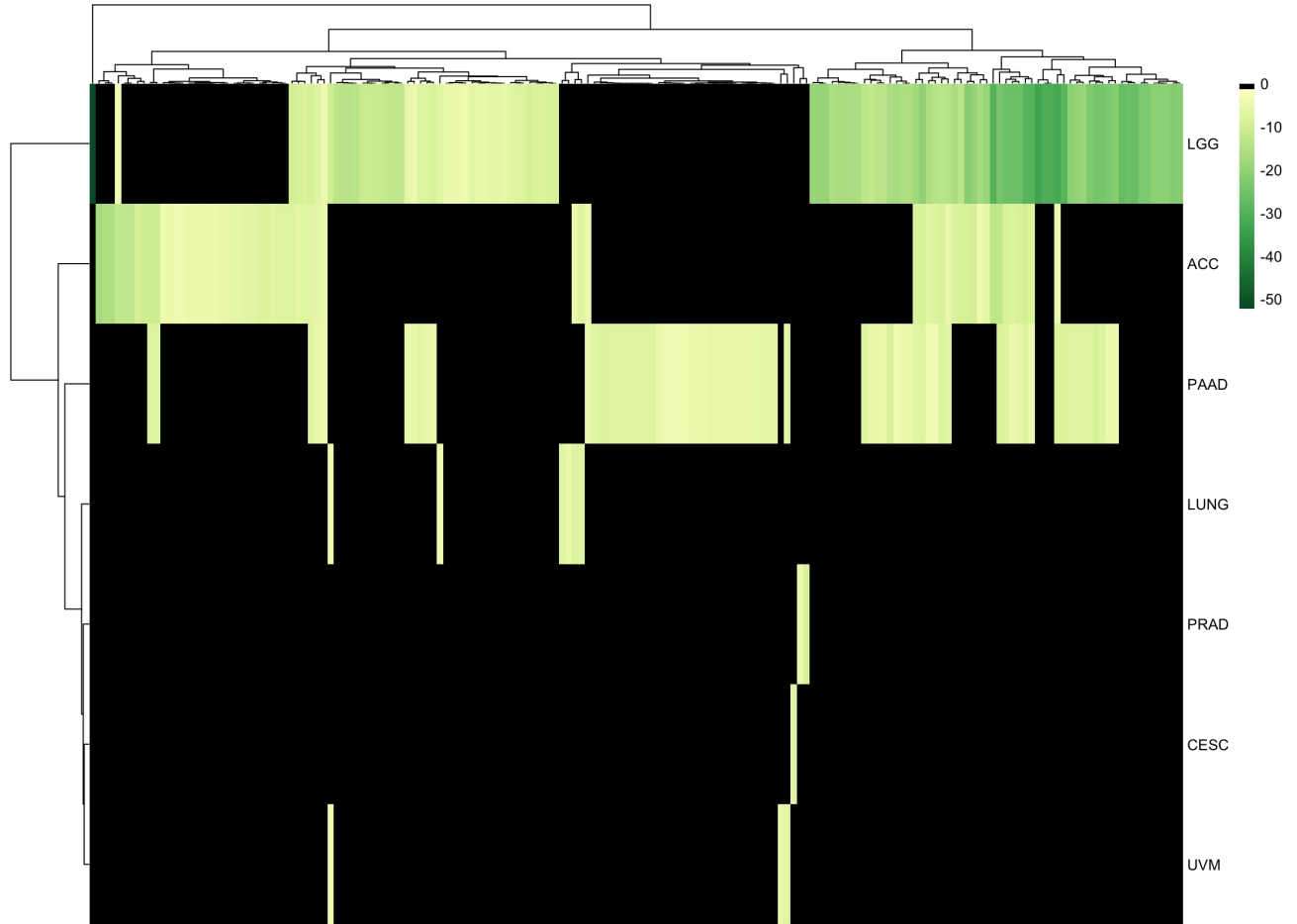


Figure 33: **Heatmap of the association strength between RFS and metabolisms.** This heatmap represents the correlation of RFS, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

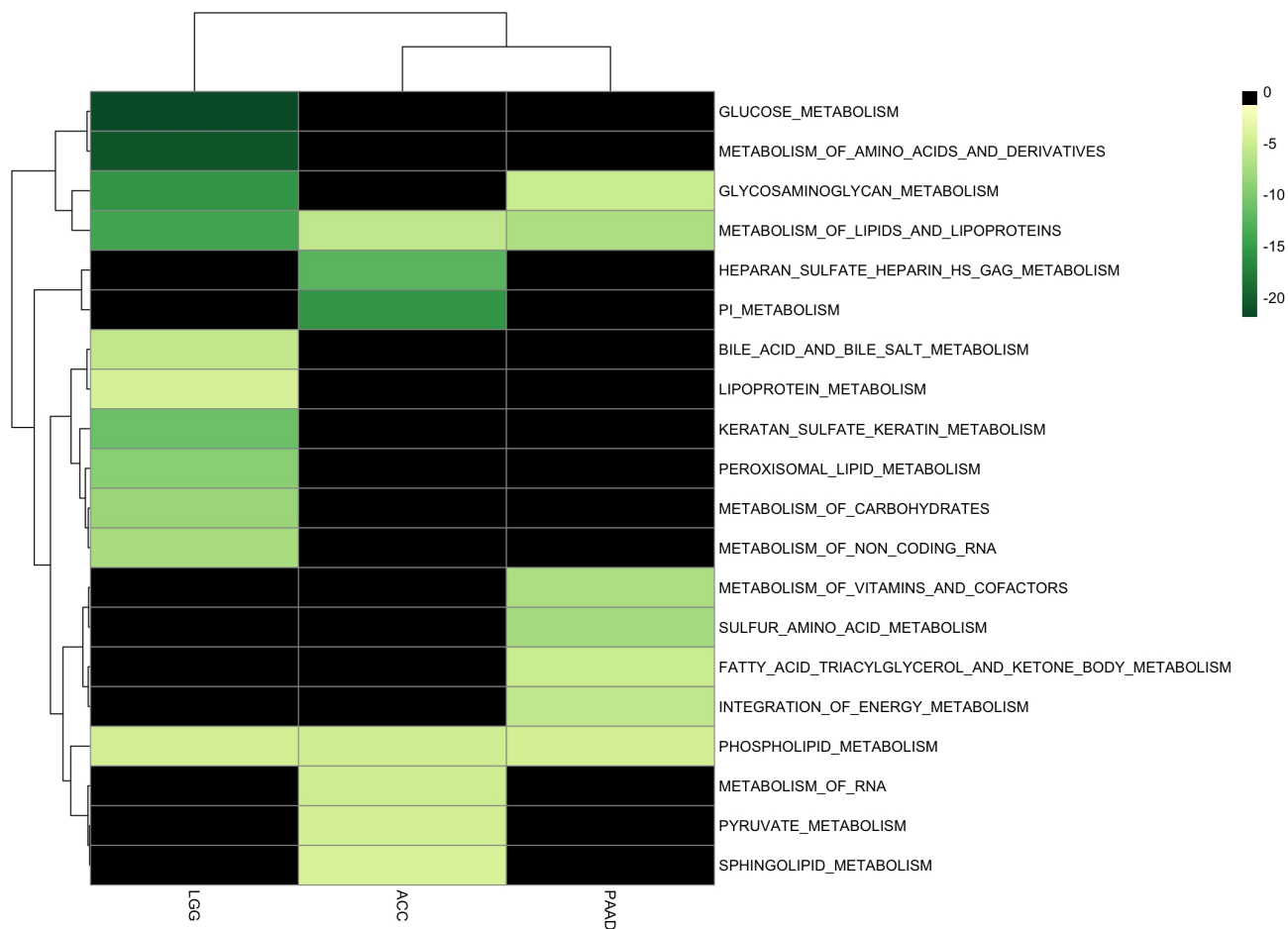


Figure 34: **Heatmap of the association strength between RFS and a subset of metabolisms.** This heatmap represents the correlation of RFS, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

### 7.1.3 Clinical Data Elements

- **Tumors\_freq:** 84.2 % (32 out of 38 tumors)
- **Tumors\_top:**

	Metabolisms (n)
<b>KIPAN</b>	342
<b>GBMLGG</b>	324
<b>LUNG</b>	323
<b>ESCA</b>	307
<b>BRCA</b>	299

- **Metabolisms\_freq:** 79.9 % (345 out of 432 metabolisms)
- **Metabolisms\_top:**

	Tumors (n)
<b>HALLMARK FATTY ACID METABOLISM</b>	21
<b>GO AMINOGLYCAN METABOLIC PROCESS</b>	19
<b>GO AMMONIUM ION METABOLIC PROCESS</b>	19
<b>GO AMINE METABOLIC PROCESS</b>	18
<b>GO MEMBRANE LIPID METABOLIC PROCESS</b>	18

- **histological type:**

Tumor	Metabolisms (n)	Metabolism	Tumors (n)
KIPAN	342	GO AMINOGLYCAN METABOLIC PROCESS	15
GBMLGG	323	GO SMALL MOLECULE METABOLIC PROCESS	15
LUNG	323	GO AMINE METABOLIC PROCESS	14
ESCA	307	GO CELLULAR LIPID METABOLIC PROCESS	14
SARC	248	GO MEMBRANE LIPID METABOLIC PROCESS	14

Table 9: **Summary of CDEs results.** *Tumors\_freq/Metabolism\_freq* is the number of tumors/metabolisms in which at least one correlation between any CDE and any metabolism/tumor is found. *Tumors\_top/Metabolism\_top* are the top five tumors/metabolisms sorted by number of metabolisms/tumors correlating with at least one CDE. The last table represent the top five tumors/metabolism where the specific CDE has been observed.



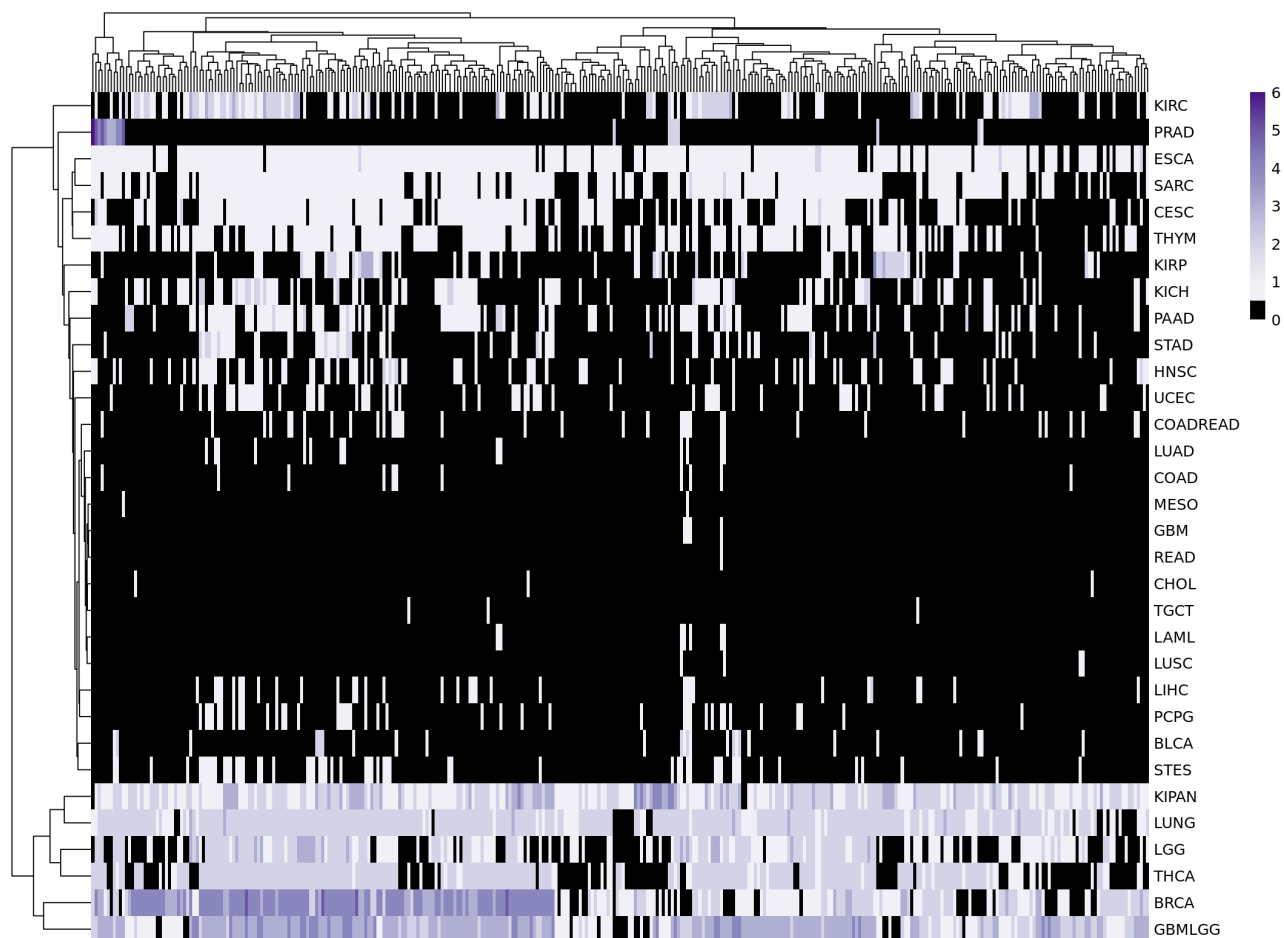


Figure 35: **Heatmap of the amount of CDEs correlated with metabolism across different tumor types.** This heatmap represents the number of CDEs correlating with a specific metabolism (rows) in tumors cohort (columns). The colour scale, from white to dark violet, represent the amount of CDEs; black means no observation found at a global Bonferroni corrected  $p$ val  $< .05$ .

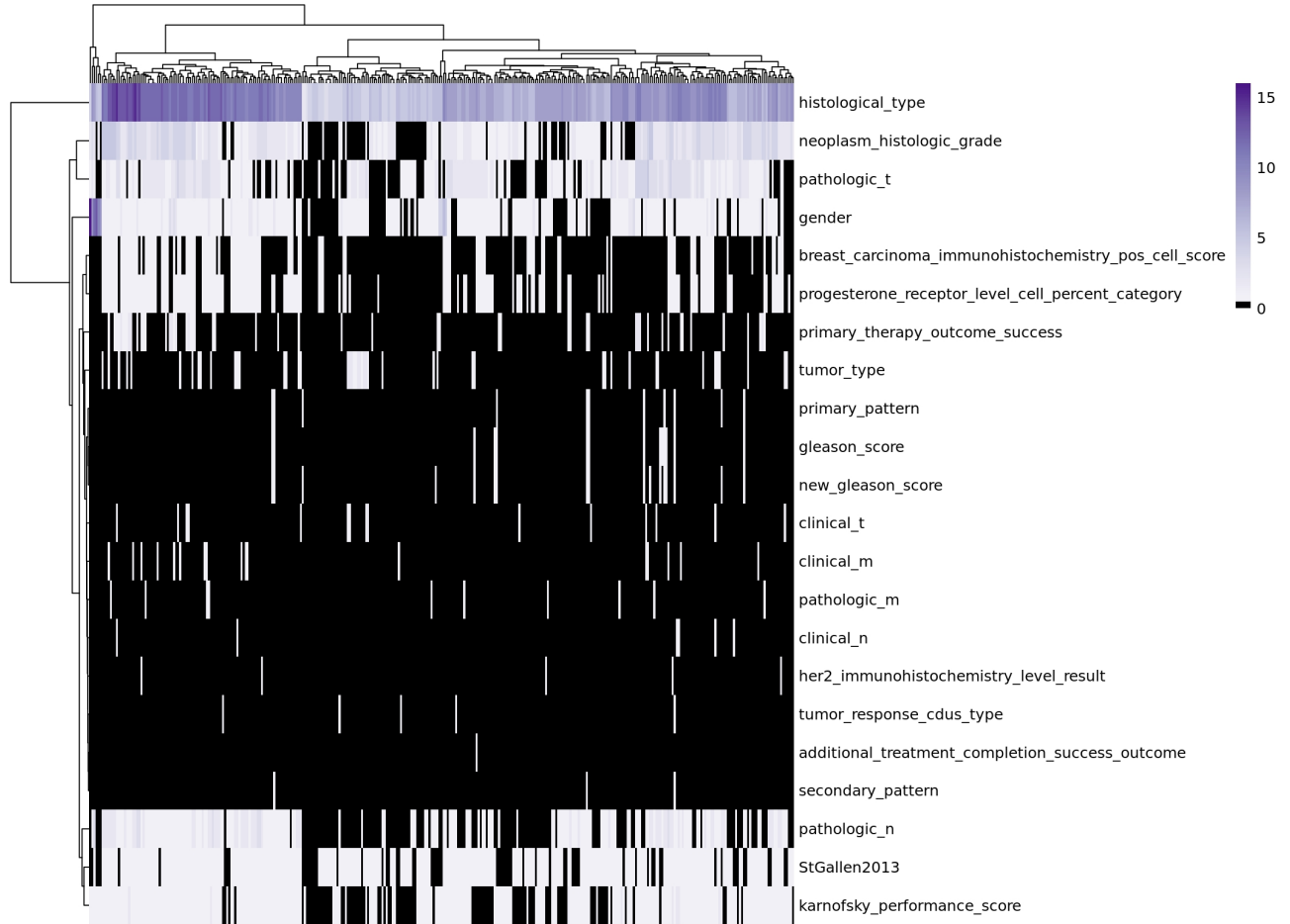


Figure 36: **Heatmap of the frequency of association between CDEs and metabolisms.** This heatmap represents the number of tumors in which a correlation between a CDE (columns) and a metabolism (rows) has been observed. The colour scale, from white to dark violet, represent the number of tumors; black means no observation found at a global Bonferroni corrected pval <.05.

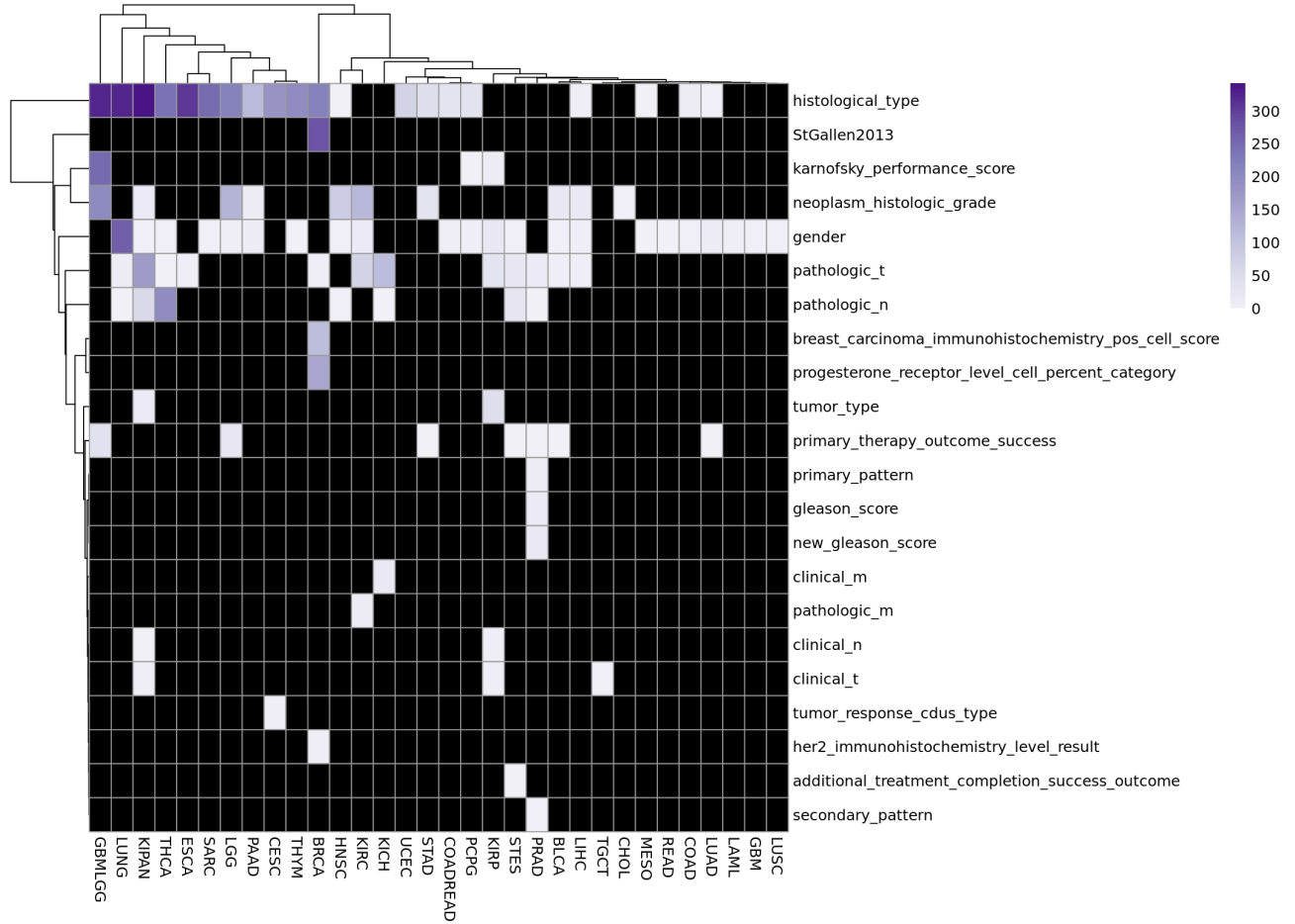


Figure 37: **Heatmap of the frequency of metabolisms.** This heatmap represents the number of metabolisms that correlates with a specific CDE (rows) in different tumor types (columns). The colour scale, from white to dark violet, represent the number of tumors; black means no observation found at a global Bonferroni corrected  $pval < .05$ .

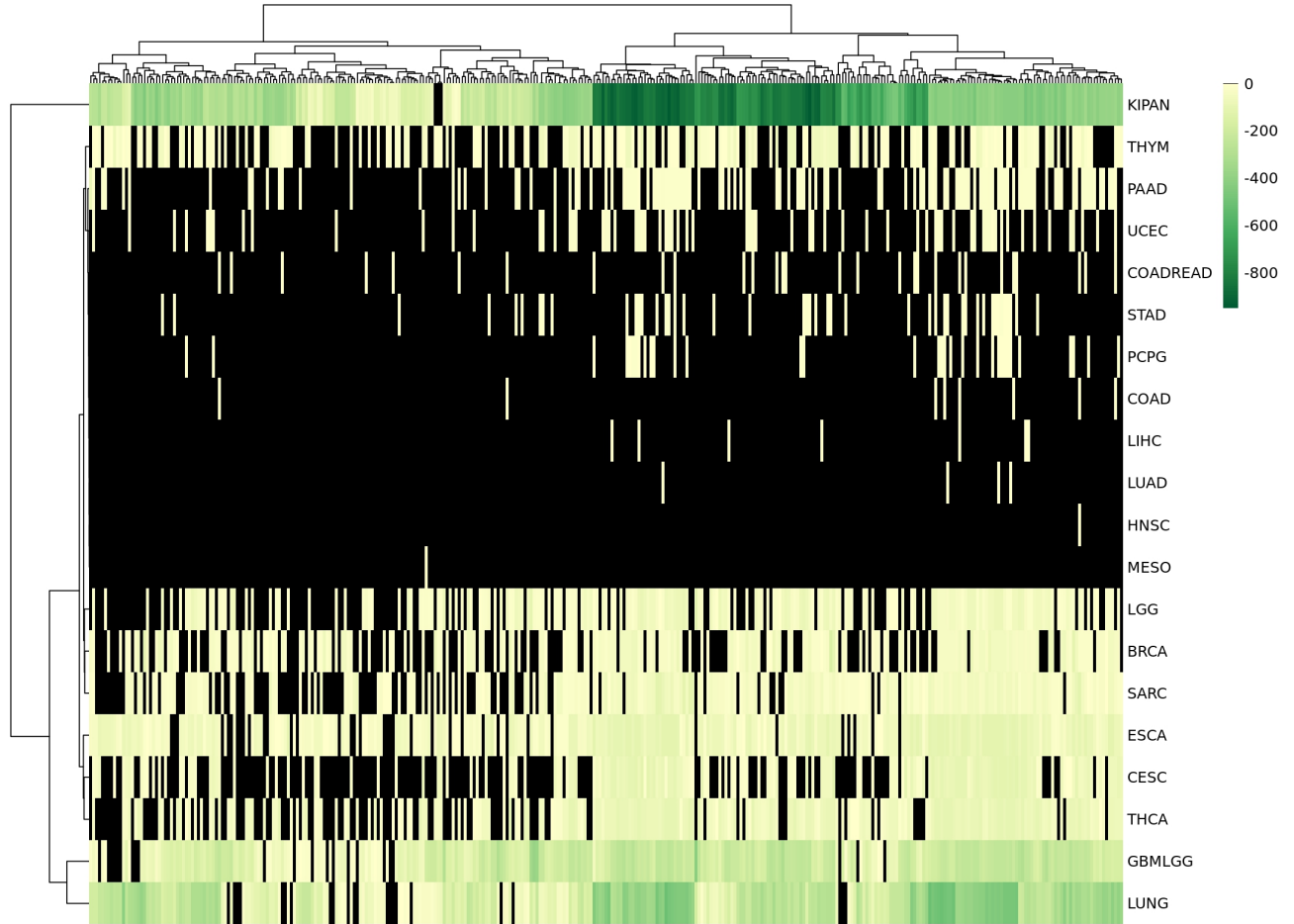


Figure 38: **Heatmap of the association strength between a specific CDE and metabolisms.** This heatmap represents the correlation of a specific CDE, *Histological type*, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

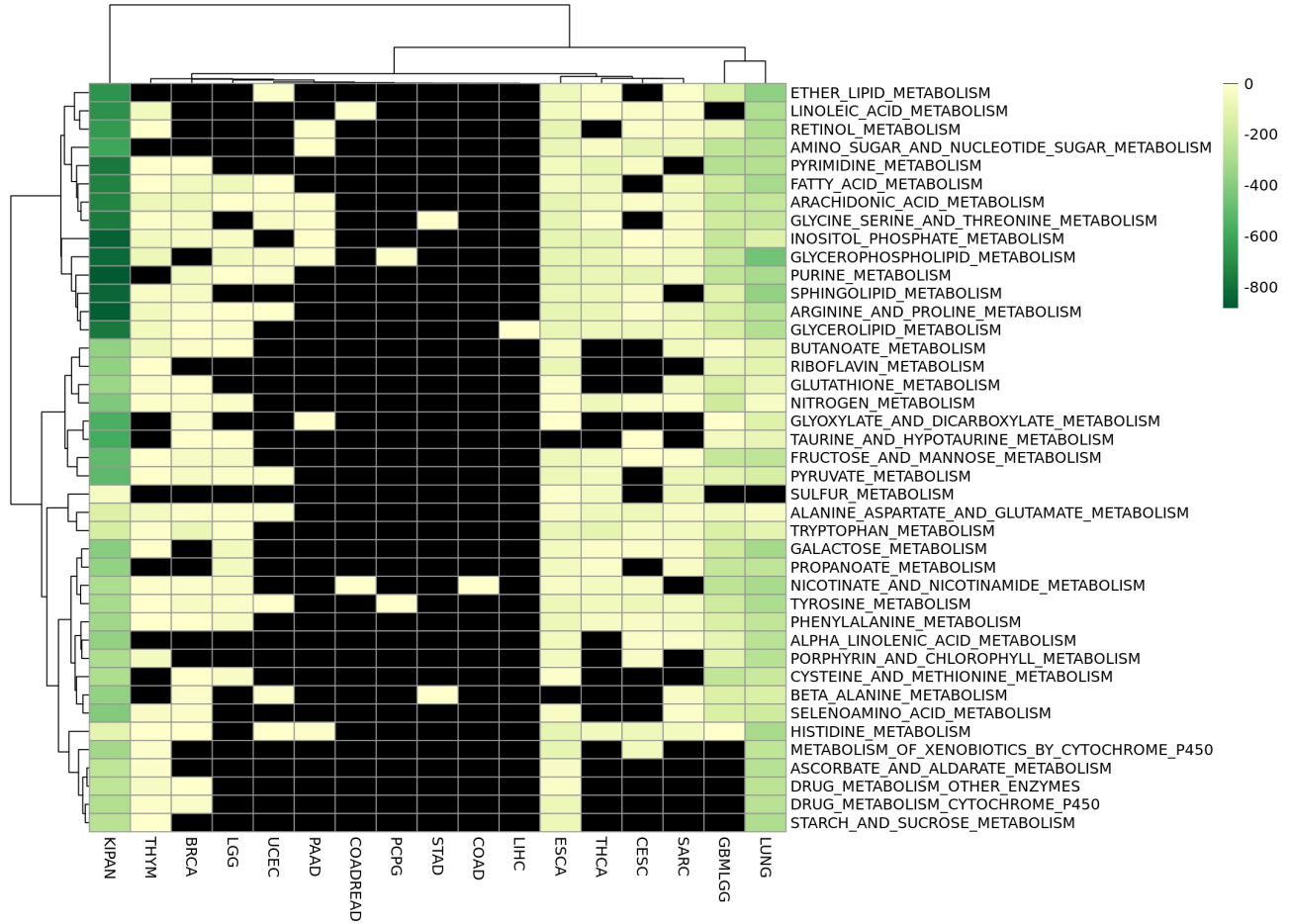


Figure 39: **Heatmap of the association strength between a specific CDE and a subset of metabolisms.** This heatmap represents the correlation of a specific CDE, *Histological type*, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

## 7.2 Genomic Data

## 7.2.1 Mutation Data

### 7.2.1.1 Protein Coding Gene Level Mutations

- **Tumors\_freq:** 39.5 % (15 out of 38 tumors)
- **Tumors\_top:**

	Metabolisms (n)
<b>GBMLGG</b>	259
<b>THCA</b>	212
<b>LUNG</b>	203
<b>BRCA</b>	192
<b>KIPAN</b>	184

- **Metabolisms\_freq:** 78.7 % (340 out of 432 metabolisms)
- **Metabolisms\_top:**

	Tumors (n)
<b>GO CELLULAR LIPID METABOLIC PROCESS</b>	9
<b>KEGG TYROSINE METABOLISM</b>	9
<b>GO CARBOHYDRATE DERIVATIVE METABOLIC PROCESS</b>	8
<b>GO LIPID METABOLIC PROCESS</b>	8
<b>GO SMALL MOLECULE METABOLIC PROCESS</b>	8

- **TP53:**

Tumor	Metabolisms (n)	Metabolism	Tumors (n)
BRCA	185	GO ACYL COA METABOLIC PROCESS	4
LUNG	108	GO ALCOHOL METABOLIC PROCESS	4
KIPAN	90	GO ETHANOLAMINE CONTAINING COMPOUND METABOLIC PROCESS	4
GBMLGG	38	GO LIPID METABOLIC PROCESS	4
LGG	38	GO ORGANIC HYDROXY COMPOUND METABOLIC PROCESS	4

Tumor	Metabolisms (n)	Metabolism	Tumors (n)
-------	-----------------	------------	------------

Table 10: **Summary of PCGL mutations results.** *Tumors\_freq/Metabolism\_freq* is the number of tumors/metabolisms in which at least one correlation between any PCGL mutation and any metabolism/tumor is found. *Tumors\_top/Metabolism\_top* are the top five tumors/metabolisms sorted by number of metabolisms/tumors correlating with at least one PCGL mutation. The last table represent the top five tumors/metabolism in which the specific gene has been observed.



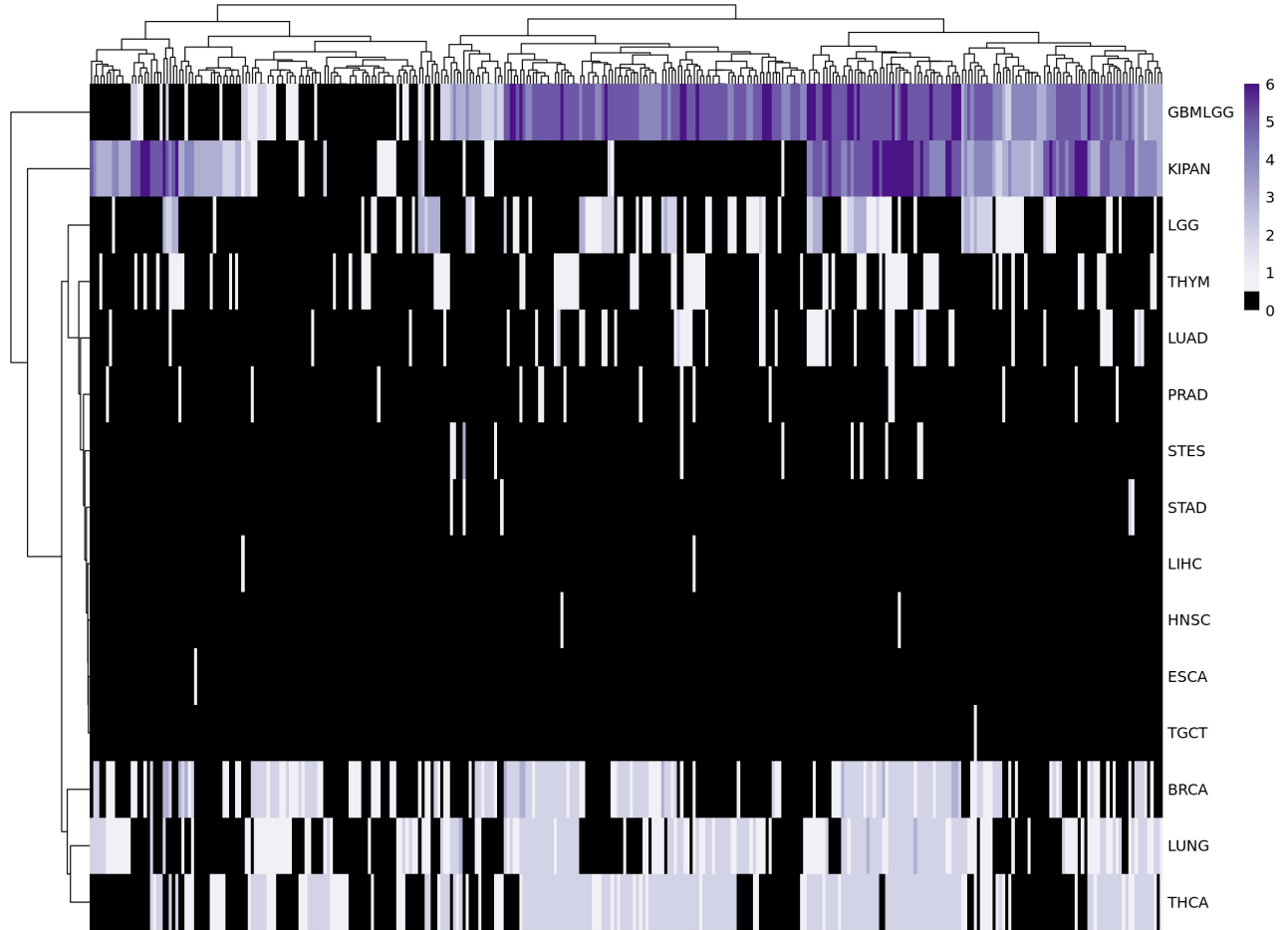


Figure 40: **Heatmap of the amount of PCGL mutations correlated with metabolism across different tumor types.** This heatmap represents the number of PCPGL mutations correlating with a specific metabolism (columns) in tumors cohort (rows). The colour scale, from white to dark violet, represents the amount of PCGL mutations; black means no observation found at a global Bonferroni corrected pval <.05.

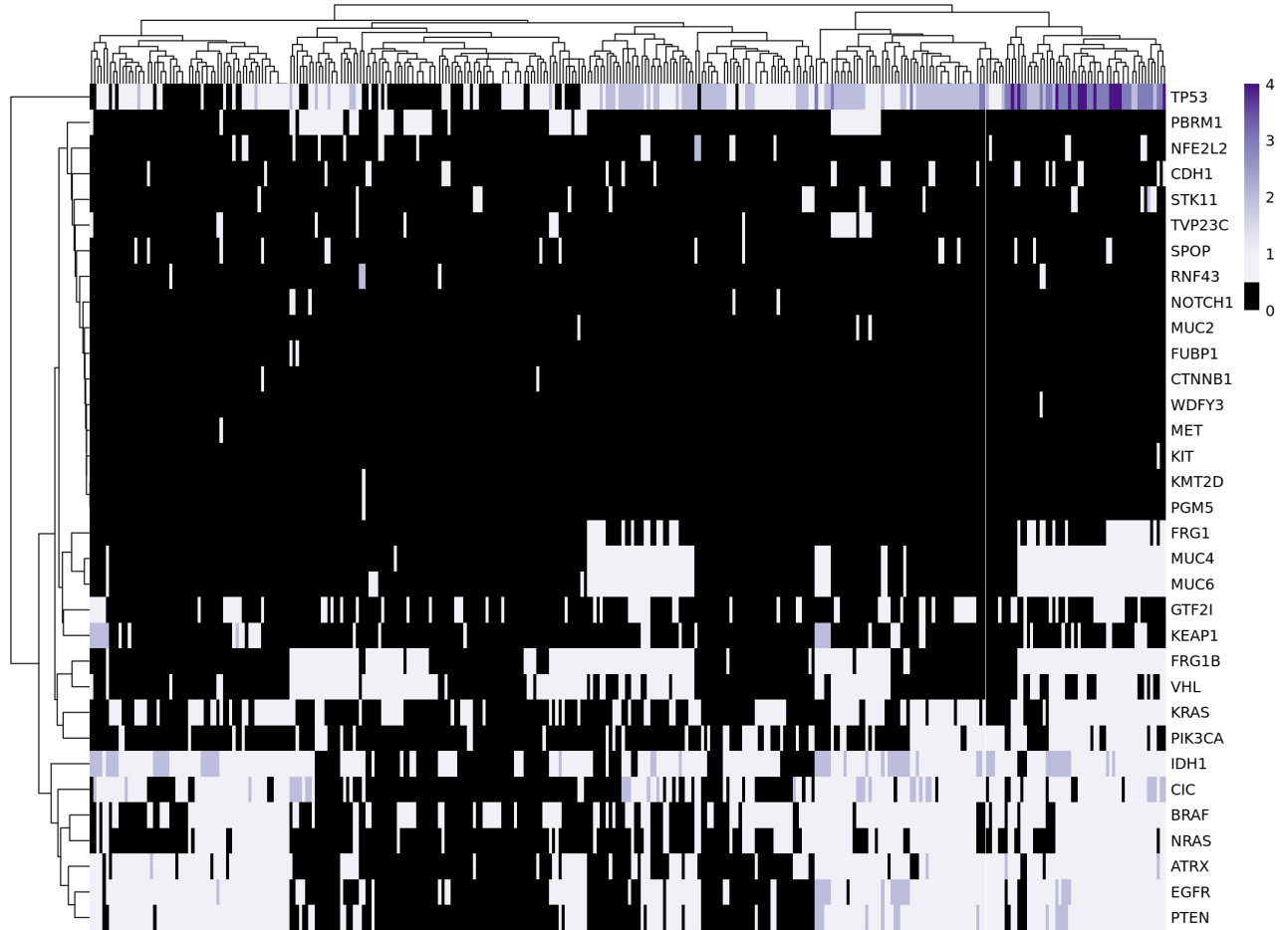


Figure 41: **Heatmap of the frequency of association between PCGL mutations and metabolisms.** This heatmap represents the number of tumors in which a correlation between PCPGL mutations (rows) and a metabolism (columns) has been observed. The colour scale, from white to dark violet, represent the number of tumors; black means no observation found at a global Bonferroni corrected  $p$ val  $< .05$ .

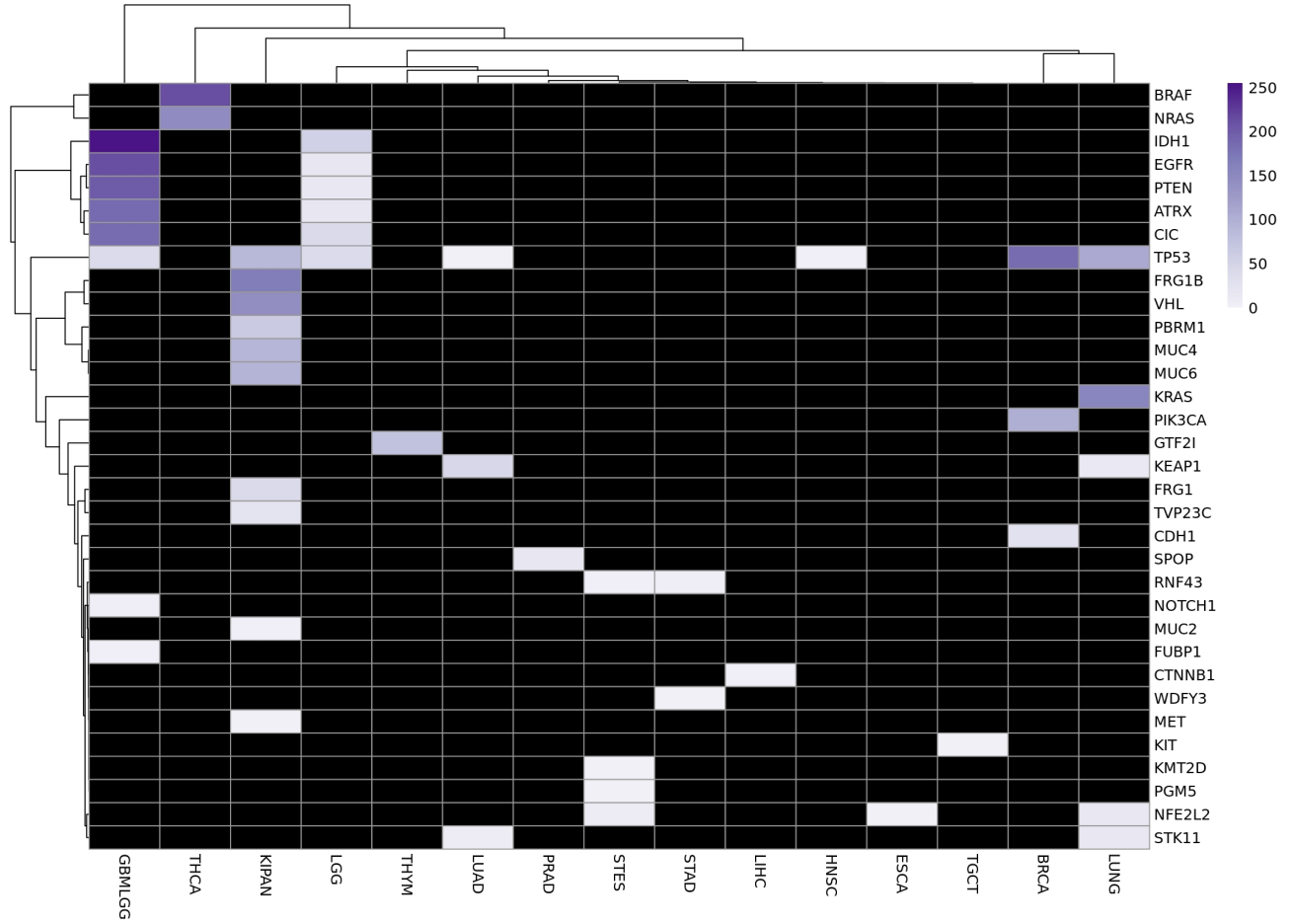


Figure 42: **Heatmap of the frequency of metabolisms.** This heatmap represents the number of metabolisms that correlates with a specific PCGL mutation (rows) in different tumor types (columns). The colour scale, from white to dark violet, represent the number of tumors; black means no observation found at a global Bonferroni corrected pval < .05.

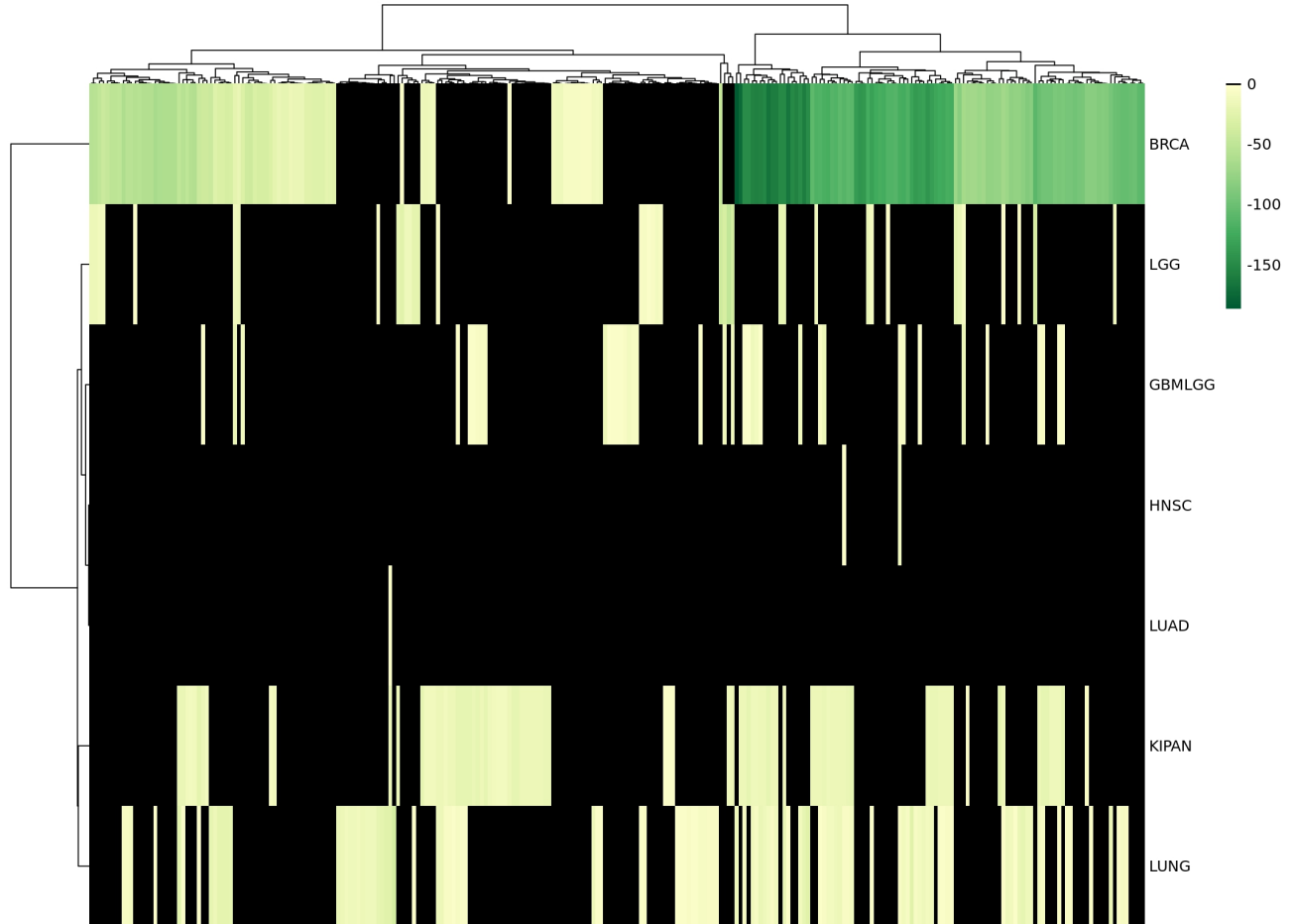


Figure 43: **Heatmap of the association strength between a specific PCGL mutation and metabolisms.** This heatmap represents the correlation of a specific PCGL mutation in *TP53*, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

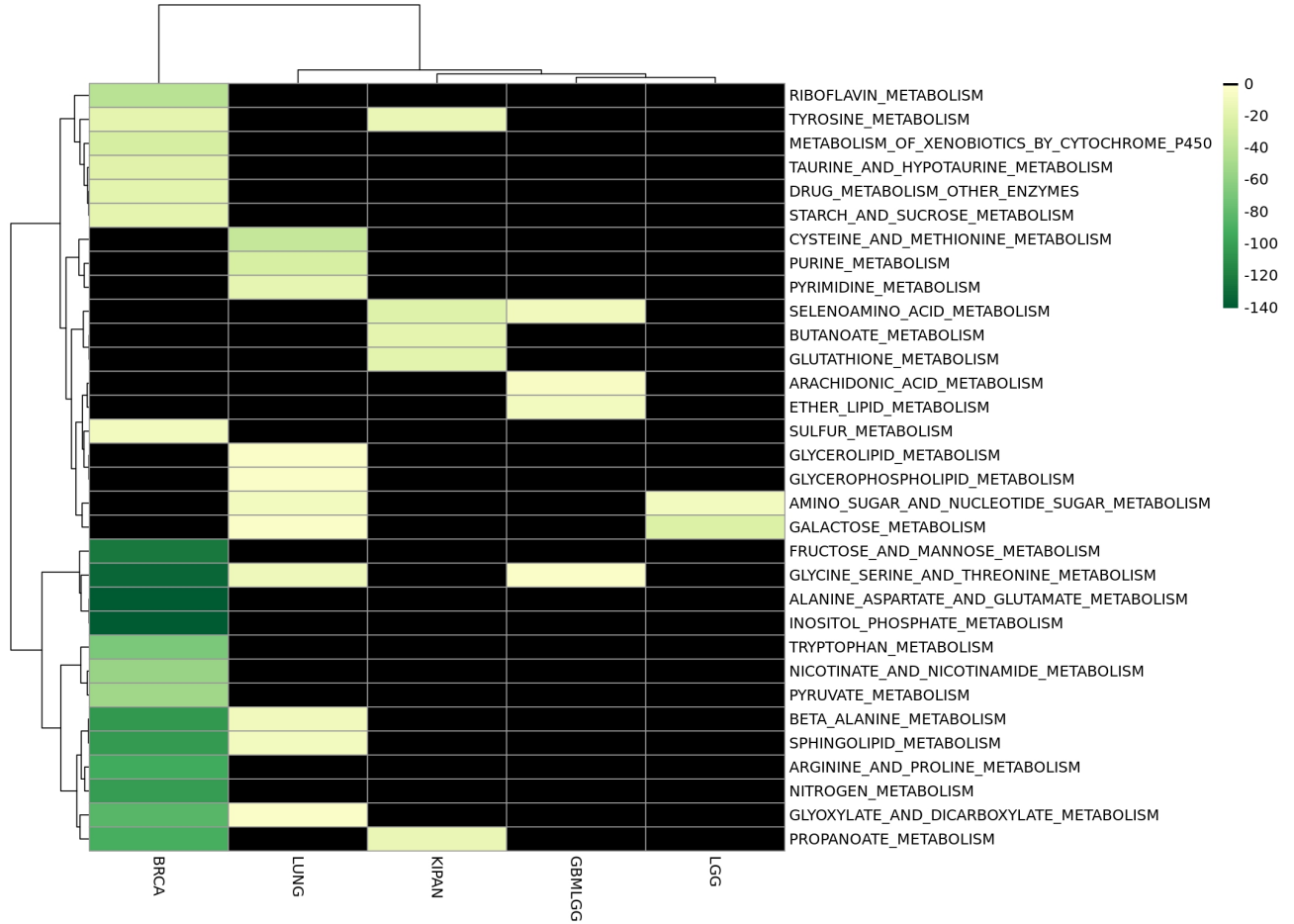


Figure 44: **Heatmap of the association strength between a specific PCGL mutation and a subset of metabolisms.** This heatmap represents the correlation of a specific PCGL mutation in *TP53*, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

### 7.2.1.2 Protein Level Mutations

- **Tumors\_freq:** 18.4 % (7 out of 38 tumors)
- **Tumors\_top:**

	Metabolisms (n)
<b>GBMLGG</b>	252
<b>THCA</b>	216
<b>THYM</b>	72
<b>LGG</b>	36
<b>KIPAN</b>	24

- **Metabolisms\_freq:** 69.9 % (302 out of 432 metabolisms)
- **Metabolisms\_top:**

	Tumors (n)
<b>GO CARBOHYDRATE DERIVATIVE METABOLIC PROCESS</b>	4
<b>GO CELLULAR LIPID METABOLIC PROCESS</b>	4
<b>GO CELLULAR MODIFIED AMINO ACID METABOLIC PROCESS</b>	4
<b>GO GLUCOSE METABOLIC PROCESS</b>	4
<b>GO HEXOSE METABOLIC PROCESS</b>	4

- **IDH1 p.R132H:**

Tumor	Metabolisms (n)	Metabolism	Tumors (n)
GBMLGG	252	GO ACETYL COA METABOLIC PROCESS	2
LGG	36	GO ALDITOL PHOSPHATE METABOLIC PROCESS	2

Table 11: **Summary of protein level mutations results.** *Tumors\_freq/Metabolism\_freq* is the number of tumors/metabolisms in which at least one correlation between any protein level mutation and any metabolism/tumor is found. *Tumors\_top/Metabolism\_top* are the top five tumors/metabolisms sorted by number of metabolisms/tumors correlating with at least one protein level mutation. The last table represent the top five tumors/metabolism in which the

specific protein level mutation has been observed.

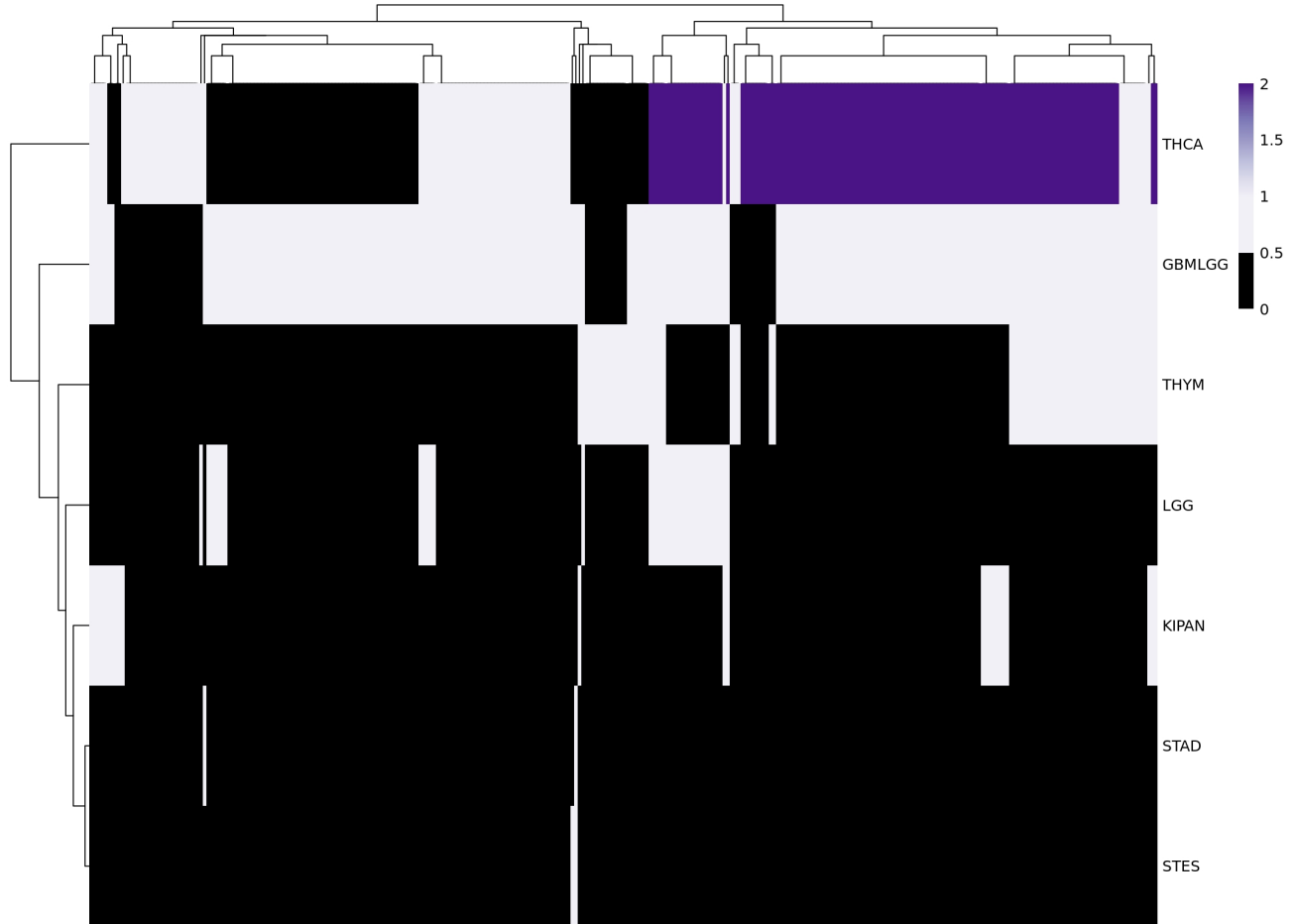


Figure 45: **Heatmap of the amount of protein level mutations correlated with metabolism across different tumor types.** This heatmap represents the number of protein level mutations correlating with a specific metabolism (columns) in tumors cohort (rows). The colour scale, from white to dark violet, represents the amount of protein level mutations; black means no observation found at a global Bonferroni corrected pval  $<.05$ .





Figure 46: **Heatmap of the frequency of association between protein level mutations and metabolisms.** This heatmap represents the number of tumors in which a correlation between protein level mutation (rows) and a metabolism (columns) has been observed. The colour scale, from white to dark violet, represent the number of tumors; black means no observation found at a global Bonferroni corrected pval  $<.05$ .

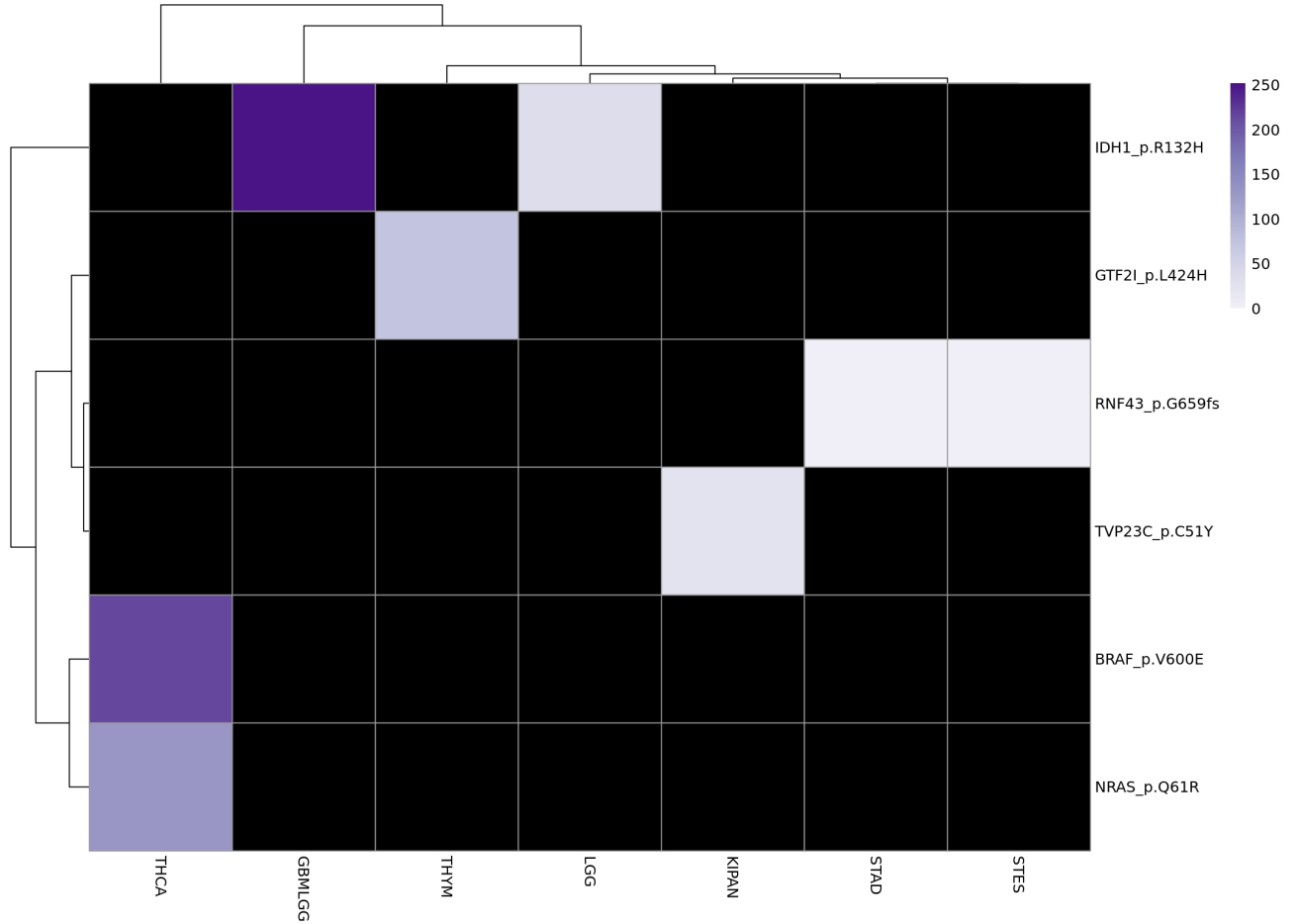


Figure 47: **Heatmap of the frequency of metabolisms.** This heatmap represents the number of metabolisms that correlates with a specific protein level mutation (rows) in different tumor types (columns). The colour scale, from white to dark violet, represents the number of tumors; black means no observation found at a global Bonferroni corrected  $pval < .05$ .

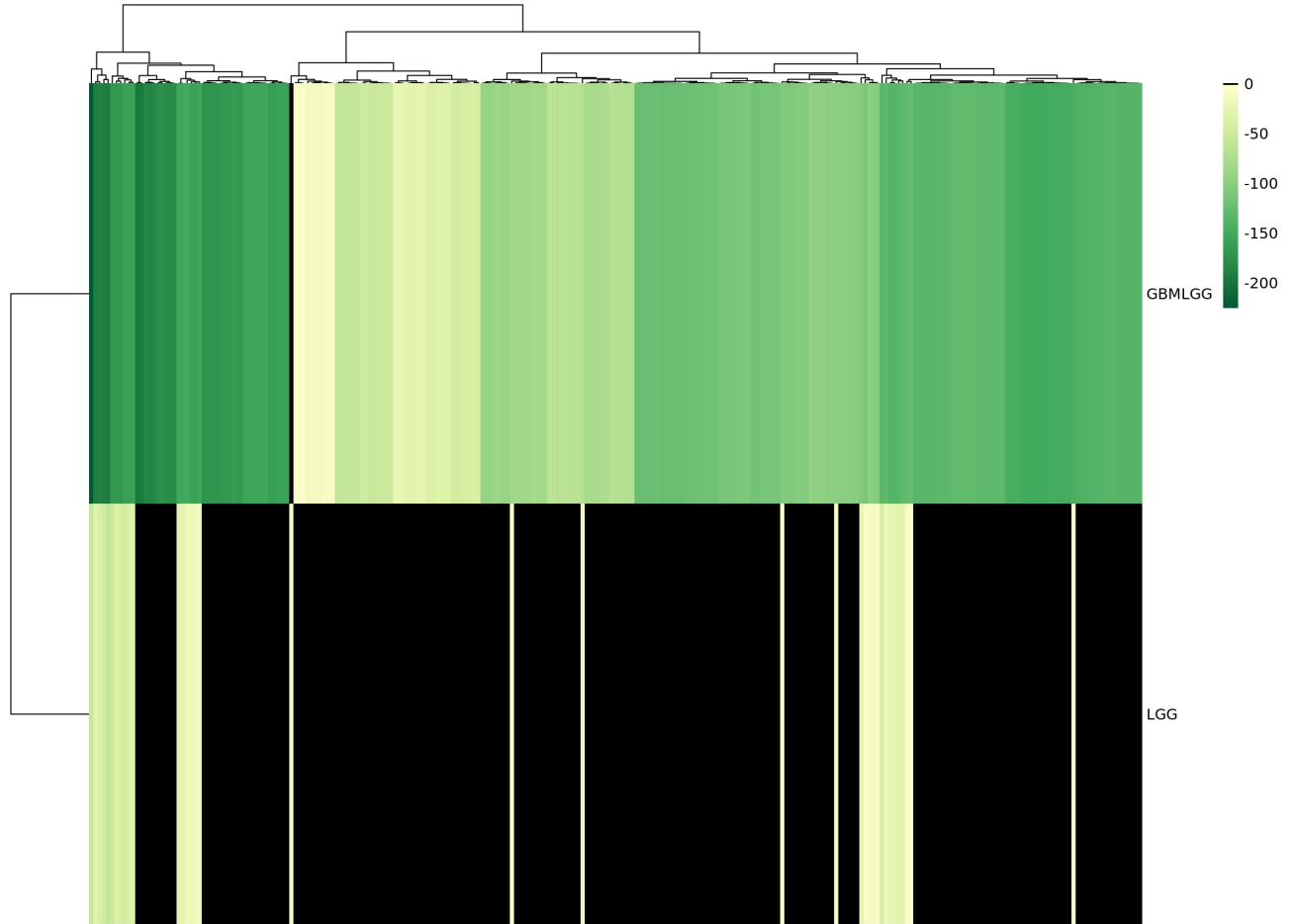


Figure 48: **Heatmap of the association strength between a specific protein level mutation and metabolisms.** This heatmap represents the correlation of a specific protein level mutation in *IDH1\_p.R132H*, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected pvalues; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

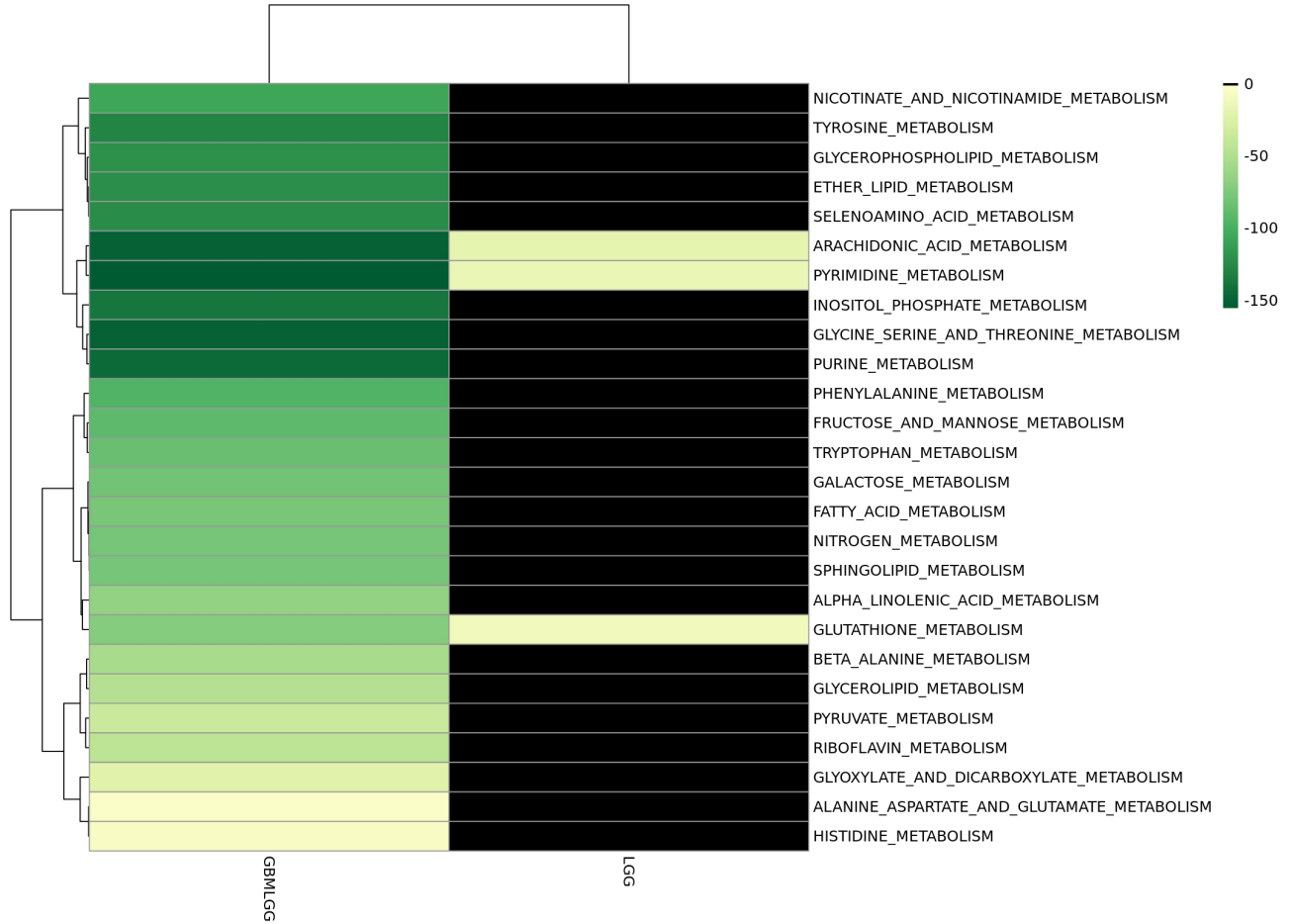


Figure 49: **Heatmap of the association strength between a specific protein level mutation and a subset of metabolisms.** This heatmap represents the correlation of a specific protein level mutation in *IDH1\_p.R132H*, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

### 7.2.1.3 Genomic Specific Nucleotide Variations

- **Tumors\_freq:** 18.4 % (7 out of 38 tumors)
- **Tumors\_top:**

	Metabolisms (n)
<b>GBMLGG</b>	251
<b>THCA</b>	216
<b>THYM</b>	74
<b>LGG</b>	35
<b>KIPAN</b>	23

- **Metabolisms\_freq:** 70.1 % (303 out of 432 metabolisms)
- **Metabolisms\_top:**

	Tumors (n)
<b>GO CARBOHYDRATE DERIVATIVE METABOLIC PROCESS</b>	4
<b>GO CELLULAR LIPID METABOLIC PROCESS</b>	4
<b>GO CELLULAR MODIFIED AMINO ACID METABOLIC PROCESS</b>	4
<b>GO GLUCOSE METABOLIC PROCESS</b>	4
<b>GO HEXOSE METABOLIC PROCESS</b>	4

- **IDH1 37 2 209113112 209113112 + Missense Mutation SNP C C T rs121913500:**

Tumor	Metabolisms (n)	Metabolism	Tumors (n)
GBMLGG	251	GO ACETYL COA METABOLIC PROCESS	2
LGG	35	GO ALDITOL PHOSPHATE METABOLIC PROCESS	2

Table 12: **Summary of gene specific nucleotide variation results.** *Tumors\_freq/Metabolism\_freq* is the number of tumors/metabolisms in which at least one correlation between any gene specific nucleotide variation and any metabolism/tumor is found. *Tumors\_top/Metabolism\_top* are the top five tumors/metabolisms sorted by number of metabolisms/tumors correlating with at least one gene specific nucleotide variation. The last table represent the top five

tumors/metabolism in which the specific gene specific nucleotide variation has been observed. The name of the mutation contains the following informations: 1) gene name, 2) genome assembly version, 3) chromosome number, 4) start position, 5) end position, 6) strand, 7) variant classification, 8) variant type, 9) reference allele, 8) tumor allele, 9) aberrant allele and 10) dbSNP RS id.

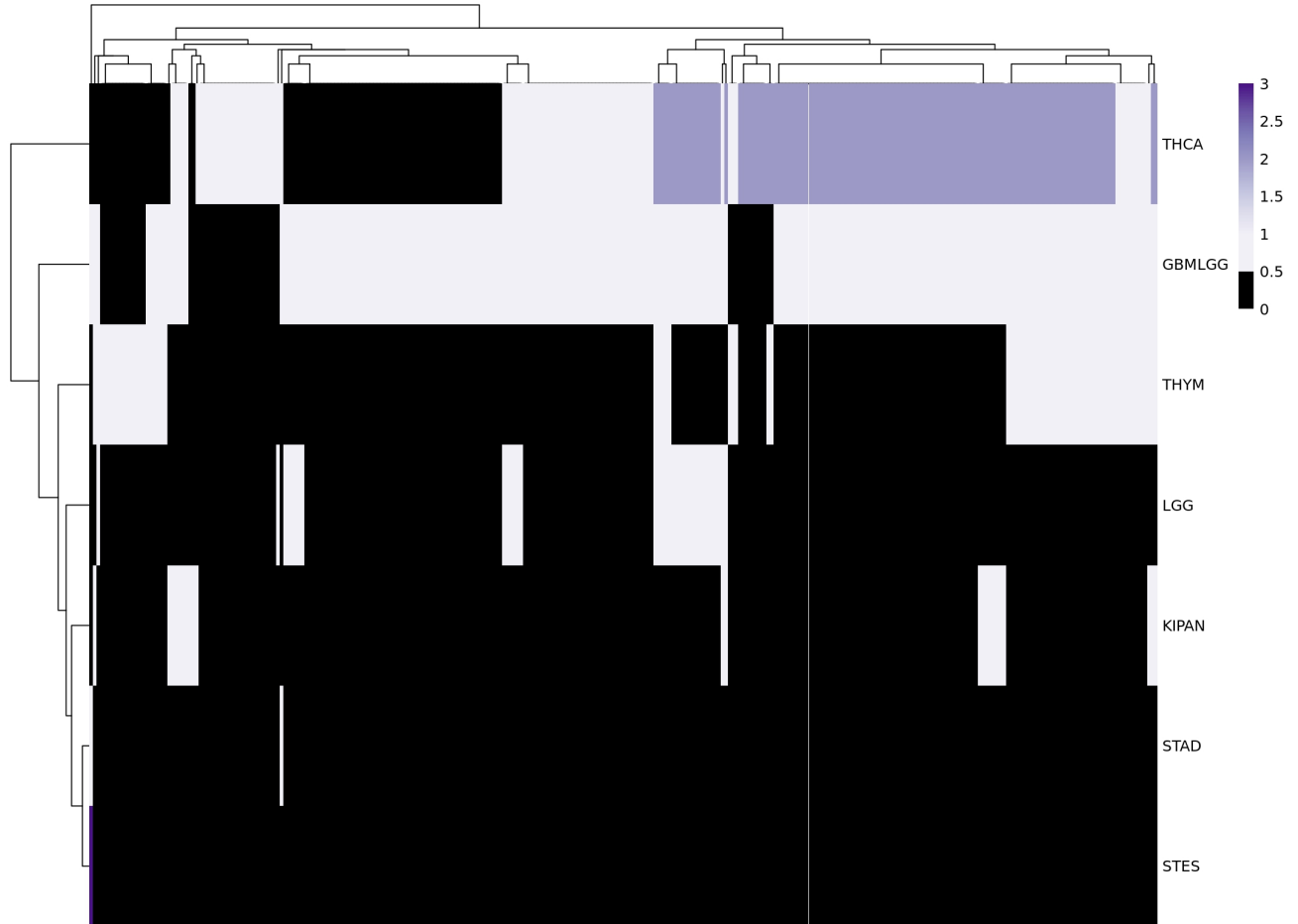


Figure 50: **Heatmap of the amount of gene specific nucleotide variation correlated with metabolism across different tumor types.** This heatmap represents the number of gene specific nucleotide variations correlating with a specific metabolism (columns) in tumors cohort (rows). The colour scale, from white to dark violet, represents the amount of gene specific nucleotide variations; black means no observation found at a global Bonferroni corrected  $p$ val  $<.05$ .



Figure 51: **Heatmap of the frequency of association between gene specific nucleotide variations and metabolisms.** This heatmap represents the number of tumors in which a correlation between gene specific nucleotide variation (rows) and a metabolism (columns) has been observed. The colour scale, from white to dark violet, represents the number of tumors; black means no observation found at a global Bonferroni corrected pval <.05.



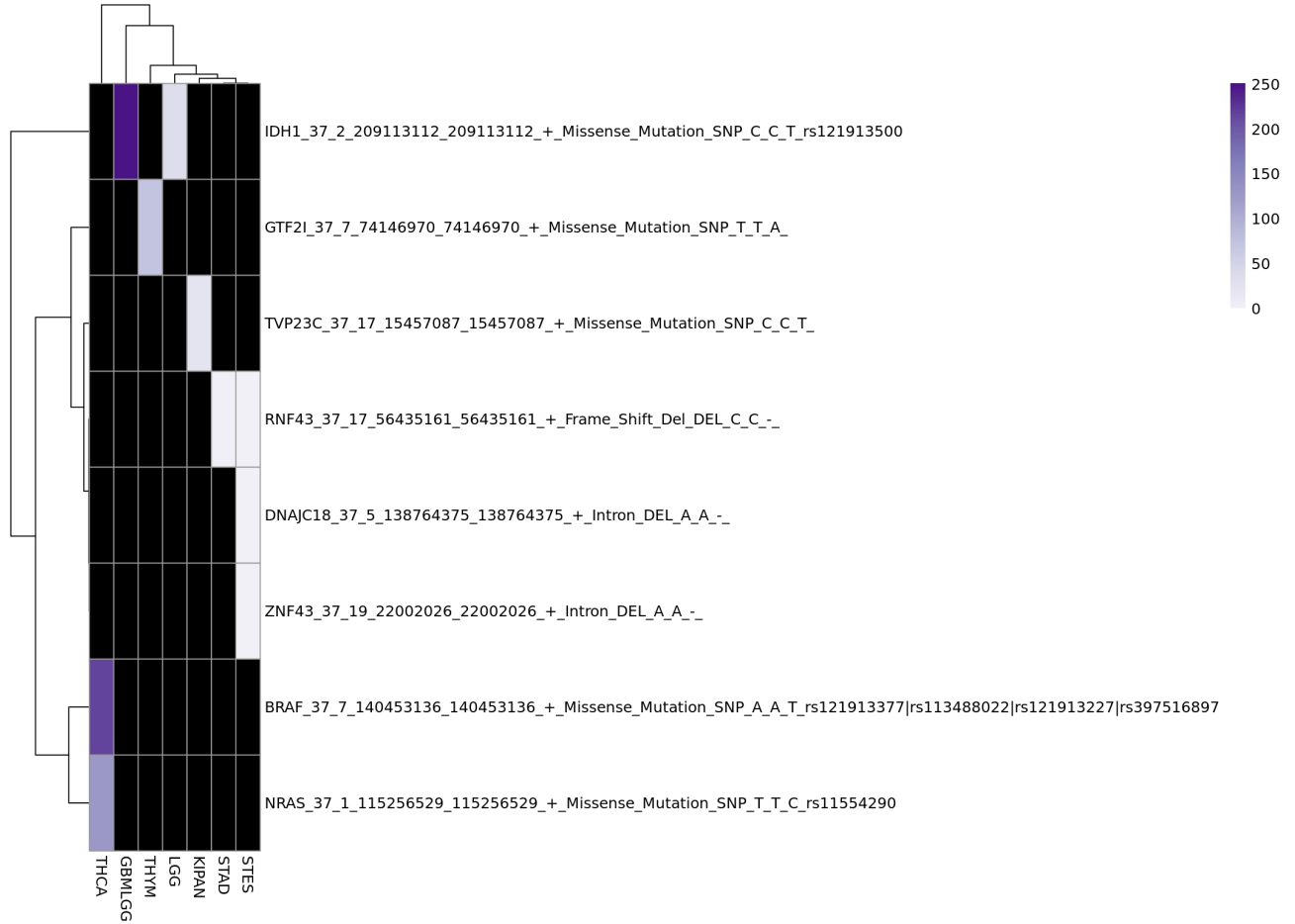


Figure 52: **Heatmap of the frequency of metabolisms.** This heatmap represents the number of metabolisms that correlates with a specific gene specific nucleotide variation (rows) in different tumor types (columns). The colour scale, from white to dark violet, represents the number of tumors; black means no observation found at a global Bonferroni corrected pval < .05.

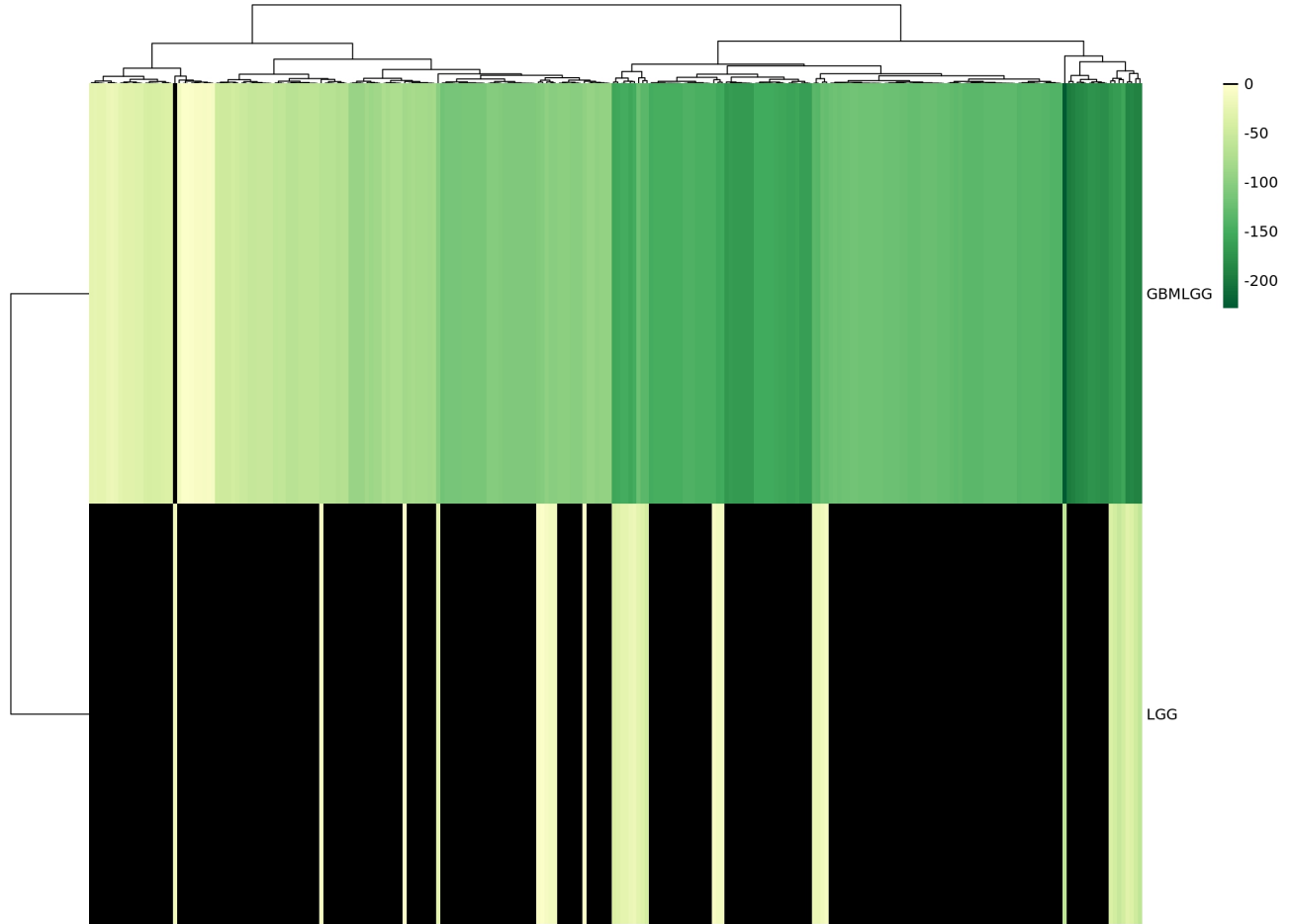


Figure 53: **Heatmap of the association strength between a specific gene specific nucleotide variation and metabolisms.** This heatmap represents the correlation of a specific gene specific nucleotide variation in \*IDH1\_37\_2\_209113112\_209113112+\_Missense\_Mutation\_SNP\_C\_C\_T\_rs121913500\*, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected pvalues; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

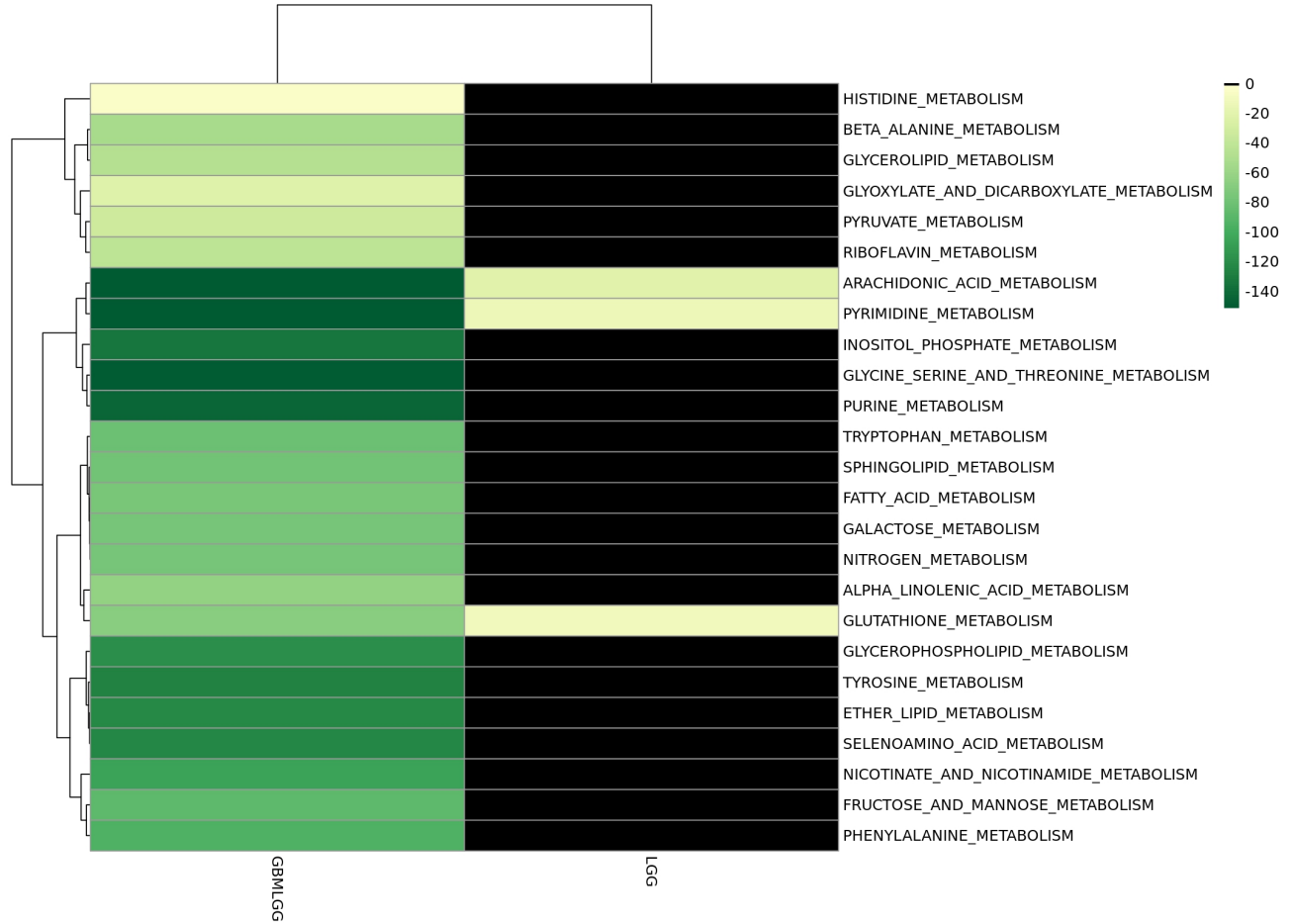


Figure 54: **Heatmap of the association strength between a specific gene specific nucleotide variation and a subset of metabolisms.** This heatmap represents the correlation of a specific gene specific nucleotide variation in *\*IDH1\_37\_2\_209113112\_209113112+\_Missense\_Mutation\_SNP\_C\_C\_T\_rs121913500\**, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected pvalues; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

## 7.2.2 Copy Number Alteration Data

- **Tumors\_freq:** 84.2 % (32 out of 38 tumors)

- **Tumors\_top:**

	Metabolisms (n)
<b>KIPAN</b>	343
<b>LUNG</b>	336
<b>GBMLGG</b>	330
<b>BRCA</b>	304
<b>KIRC</b>	281

- **Metabolisms\_freq:** 79.9 % (345 out of 432 metabolisms)

- **Metabolisms\_top:**

	Tumors (n)
<b>GO AMMONIUM ION METABOLIC PROCESS</b>	25
<b>GO ORGANIC HYDROXY COMPOUND METABOLIC PROCESS</b>	25
<b>GO ALPHA AMINO ACID METABOLIC PROCESS</b>	24
<b>GO CARBOHYDRATE METABOLIC PROCESS</b>	24
<b>GO CELLULAR AMINO ACID METABOLIC PROCESS</b>	24

- **chr3p:**

Tumor	Metabolisms (n)	Metabolism	Tumors (n)
KIPAN	329	GO ALPHA AMINO ACID METABOLIC PROCESS	12
LUNG	300	GO AMINE METABOLIC PROCESS	12
KIRC	246	GO MONOCARBOXYLIC ACID METABOLIC PROCESS	12
STES	154	GO CARBOHYDRATE METABOLIC PROCESS	11
ESCA	146	GO DICARBOXYLIC ACID METABOLIC PROCESS	11

Tumor	Metabolisms (n)	Metabolism	Tumors (n)
-------	-----------------	------------	------------

Table 13: **Summary of CNA results.** *Tumors\_freq/Metabolism\_freq* is the number of tumors/metabolisms in which at least one correlation between any CNA and any metabolism/tumor is found. *Tumors\_top/Metabolism\_top* are the top five tumors/metabolisms sorted by number of metabolisms/tumors correlating with at least one CNA. The last table represents the top five tumors/metabolisms in which the specific CNA has been observed.

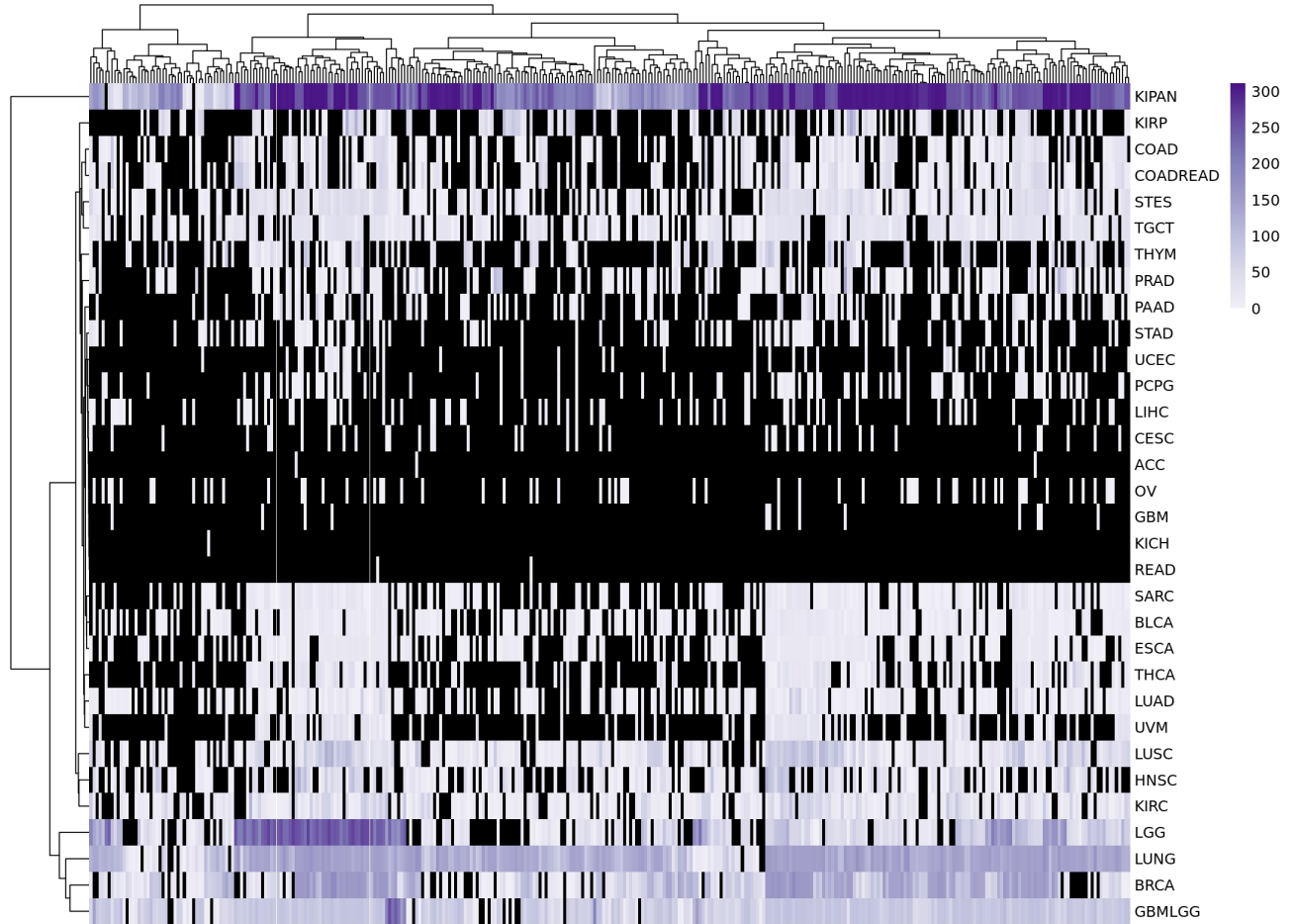


Figure 55: **Heatmap of the amount of CNAs correlated with metabolism across different tumor types.** This heatmap represents the number of CNAs correlating with a specific metabolism (columns) in tumors cohort (rows). The colour scale, from white to dark violet, represents the amount of CNAs; black means no observation found at a global Bonferroni corrected  $pval < .05$ .

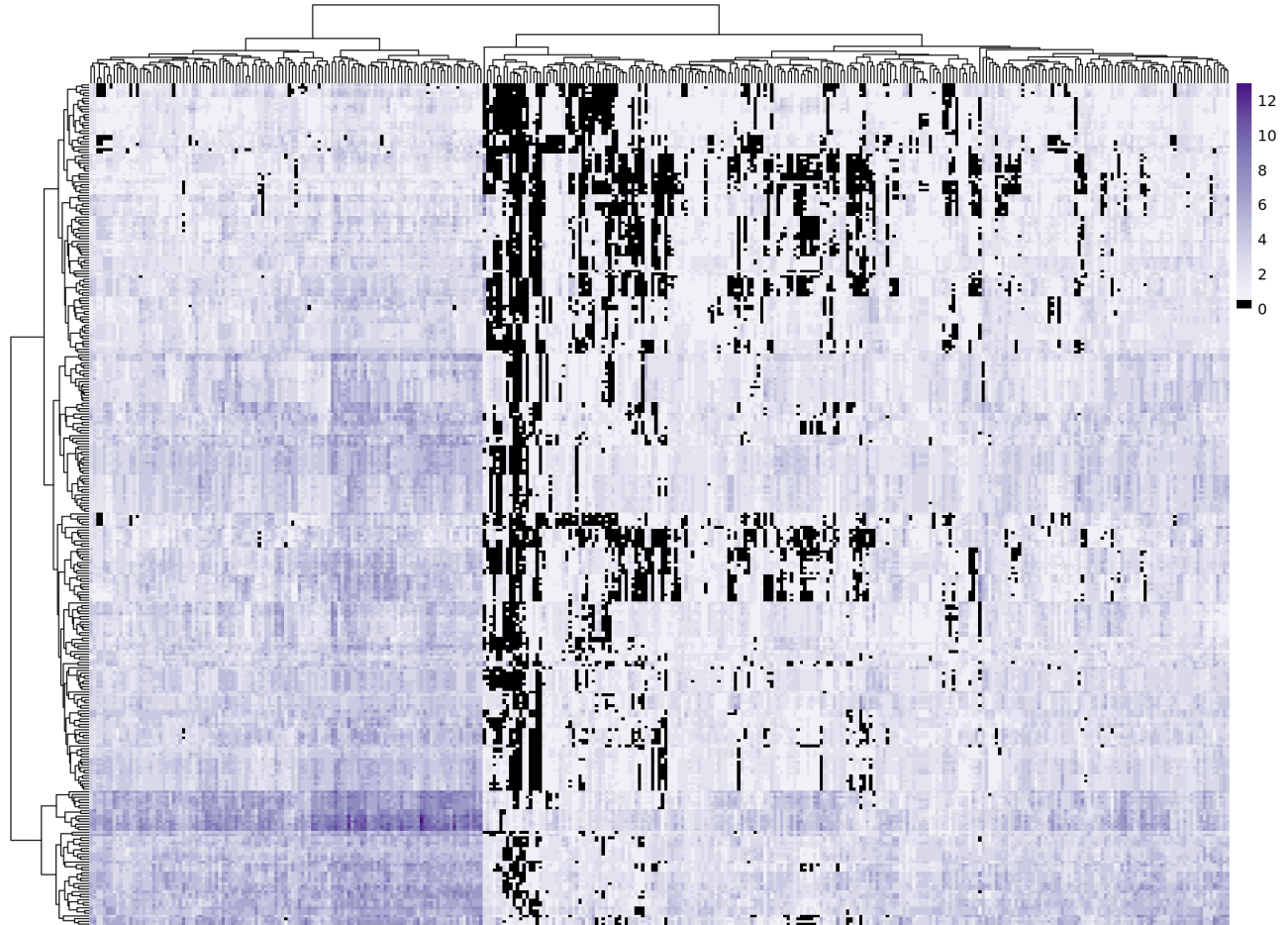


Figure 56: **Heatmap of the frequency of association between CNAs and metabolisms.** This heatmap represents the number of tumors in which a correlation between CNA (columns) and a metabolism (rows) has been observed. The colour scale, from white to dark violet, represent the number of tumors; black means no observation found at a global Bonferroni corrected  $pval < .05$ .

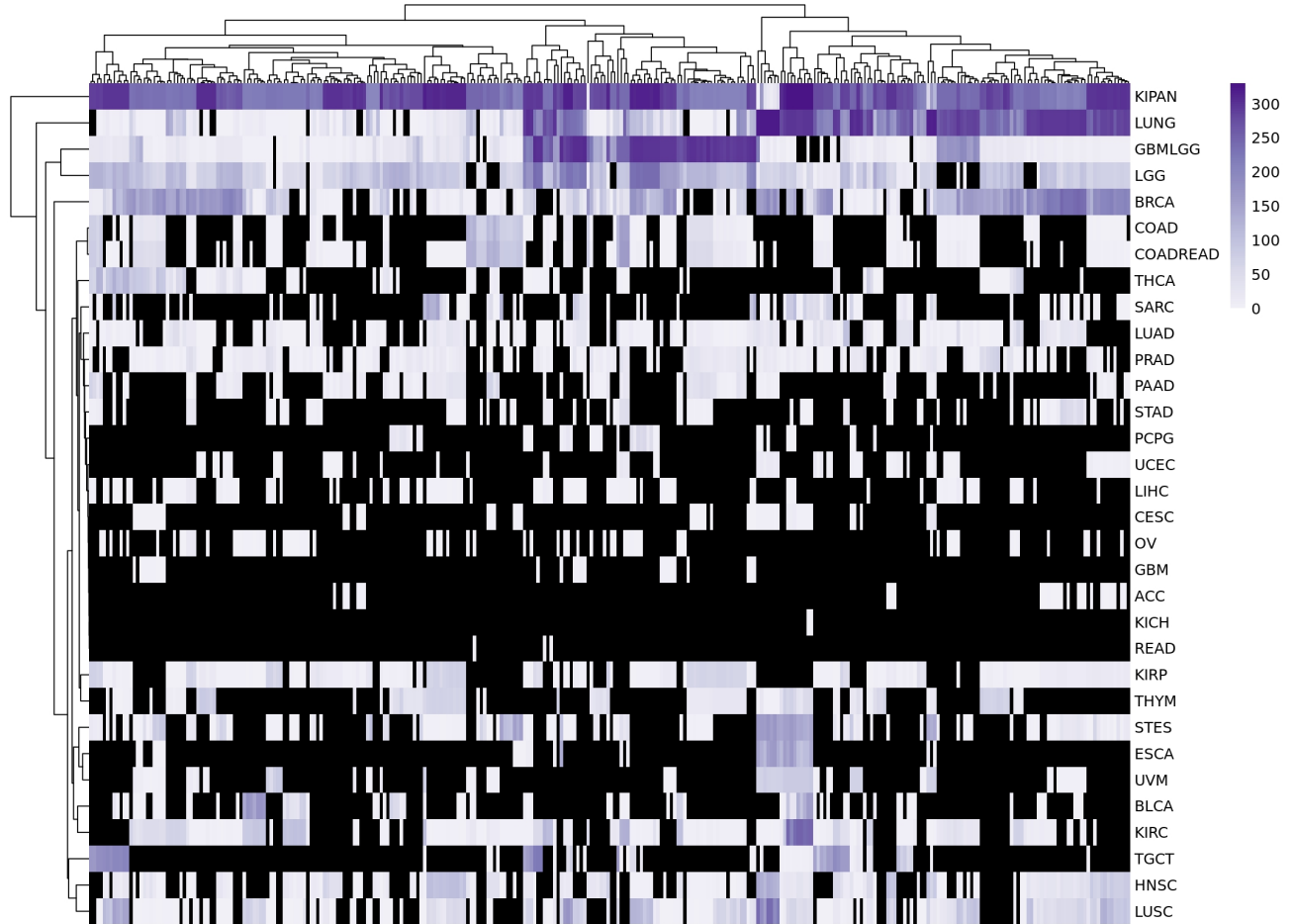


Figure 57: **Heatmap of the frequency of metabolisms.** This heatmap represents the number of metabolisms that correlates with a specific CNA (columns) in different tumor types (rows). The colour scale, from white to dark violet, represents the number of tumors; black means no observation found at a global Bonferroni corrected pval <.05.



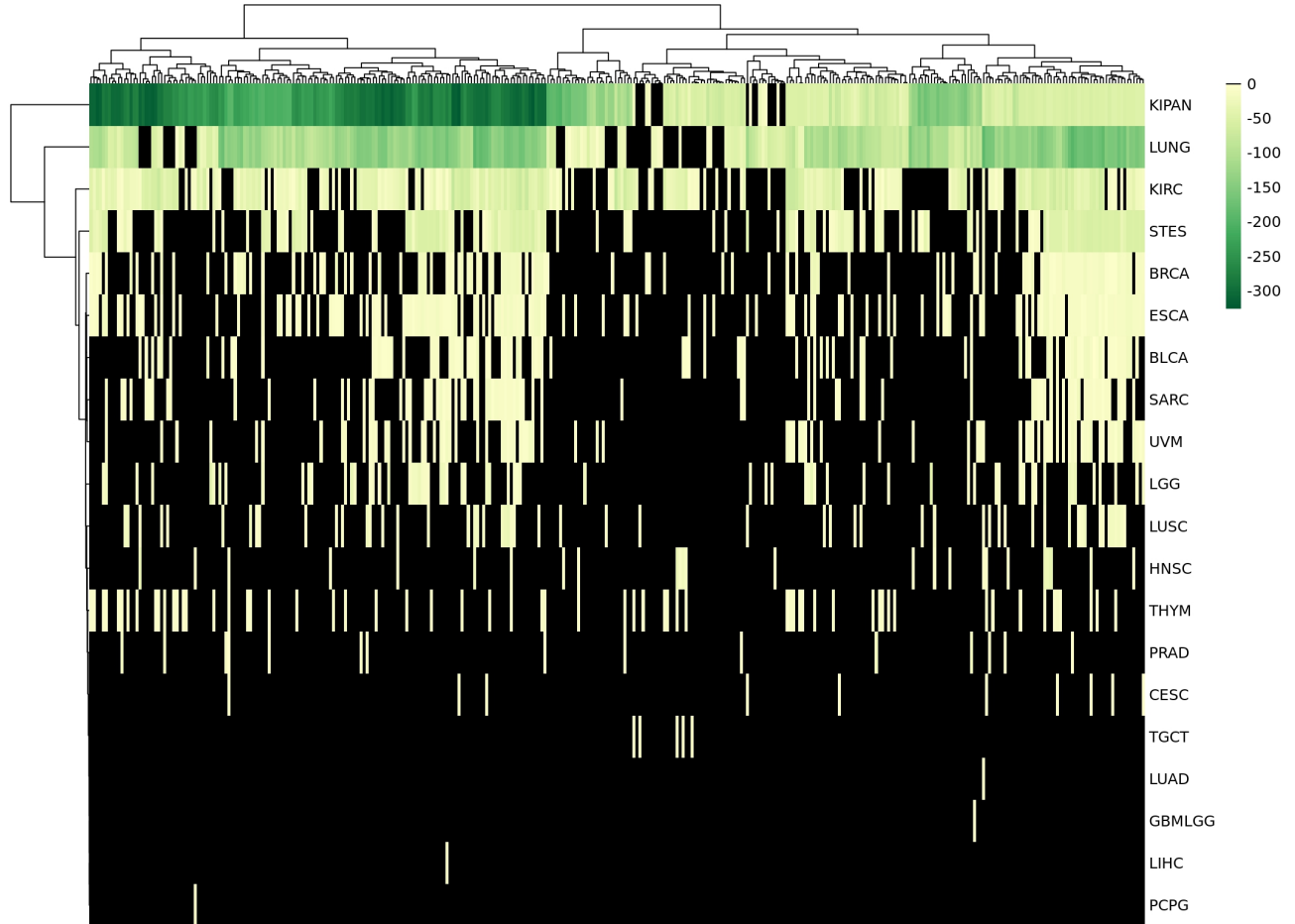


Figure 58: **Heatmap of the association strength between a specific CNA and metabolisms.** This heatmap represents the correlation of a specific CNA of *chr3p*, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected pvalues; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

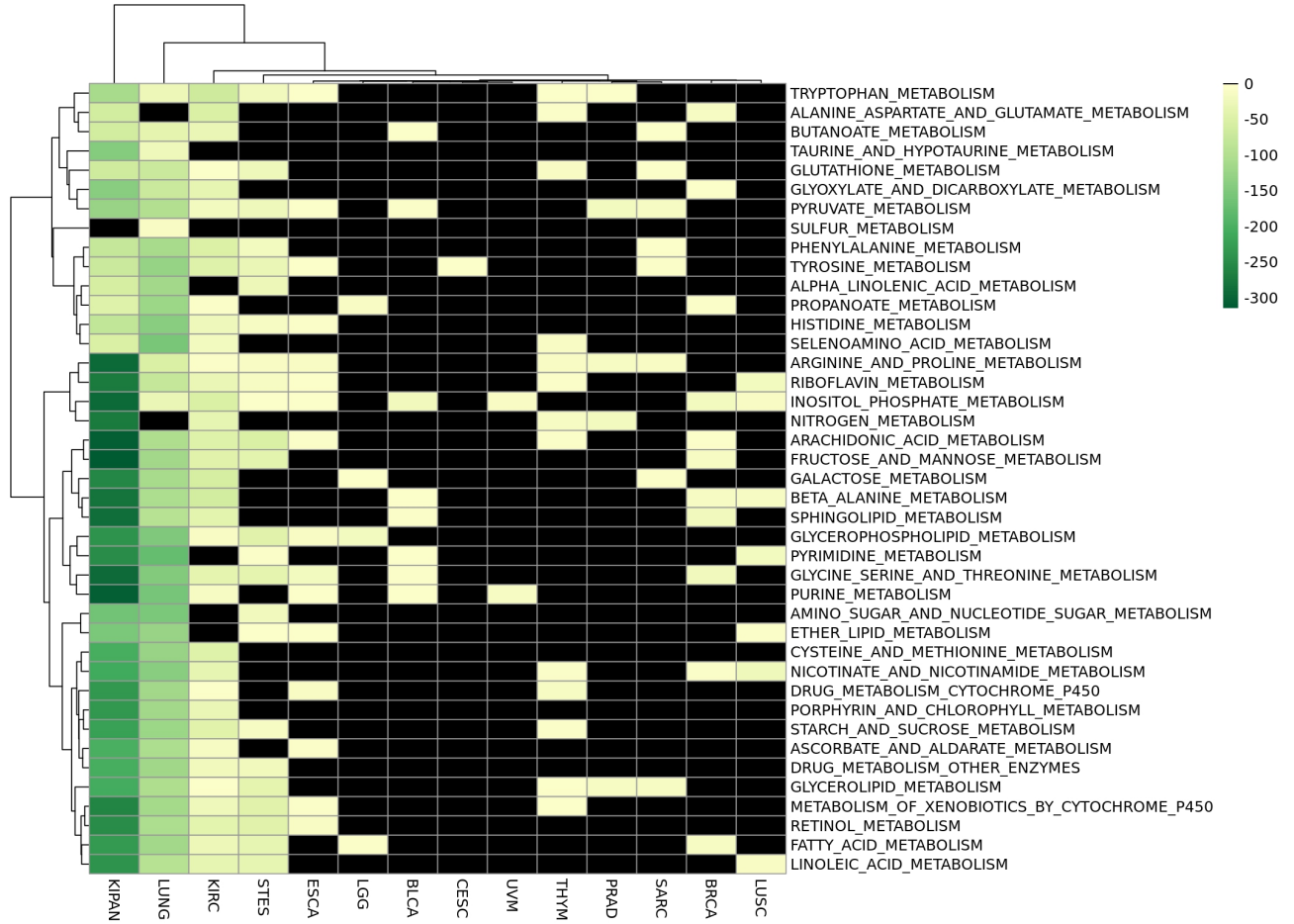


Figure 59: **Heatmap of the association strength between a specific CNA and a subset of metabolisms.** This heatmap represents the correlation of a specific CNA of *chr3p*, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected pvalues; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

## 7.2.3 Methylation Data

### 7.2.3.1 All Genes Methylation Level

- **Tumors\_freq:** 68.4 % (26 out of 38 tumors)
- **Tumors\_top:**

	Metabolisms (n)
<b>TGCT</b>	329
<b>LUNG</b>	296
<b>GBMLGG</b>	295
<b>KIPAN</b>	247
<b>LGG</b>	229

- **Metabolisms\_freq:** 79.9 % (345 out of 432 metabolisms)
- **Metabolisms\_top:**

	Tumors (n)
<b>GO MONOCARBOXYLIC ACID METABOLIC PROCESS</b>	19
<b>GO COFACTOR METABOLIC PROCESS</b>	17
<b>GO MONOSACCHARIDE METABOLIC PROCESS</b>	17
<b>GO SULFUR COMPOUND METABOLIC PROCESS</b>	17
<b>REACTOME METABOLISM OF LIPIDS AND LIPOPROTEINS</b>	17

Table 14: **Summary of whole genome methylation results.** *Tumors\_freq/Metabolism\_freq* is the number of tumors/metabolisms where the correlation between whole genome methylation and any metabolism/tumor was observed *Tumors\_top/Metabolism\_top* are the top five tumors/metabolisms sorted by occurrence of metabolisms/tumors correlating with whole genome methylation.

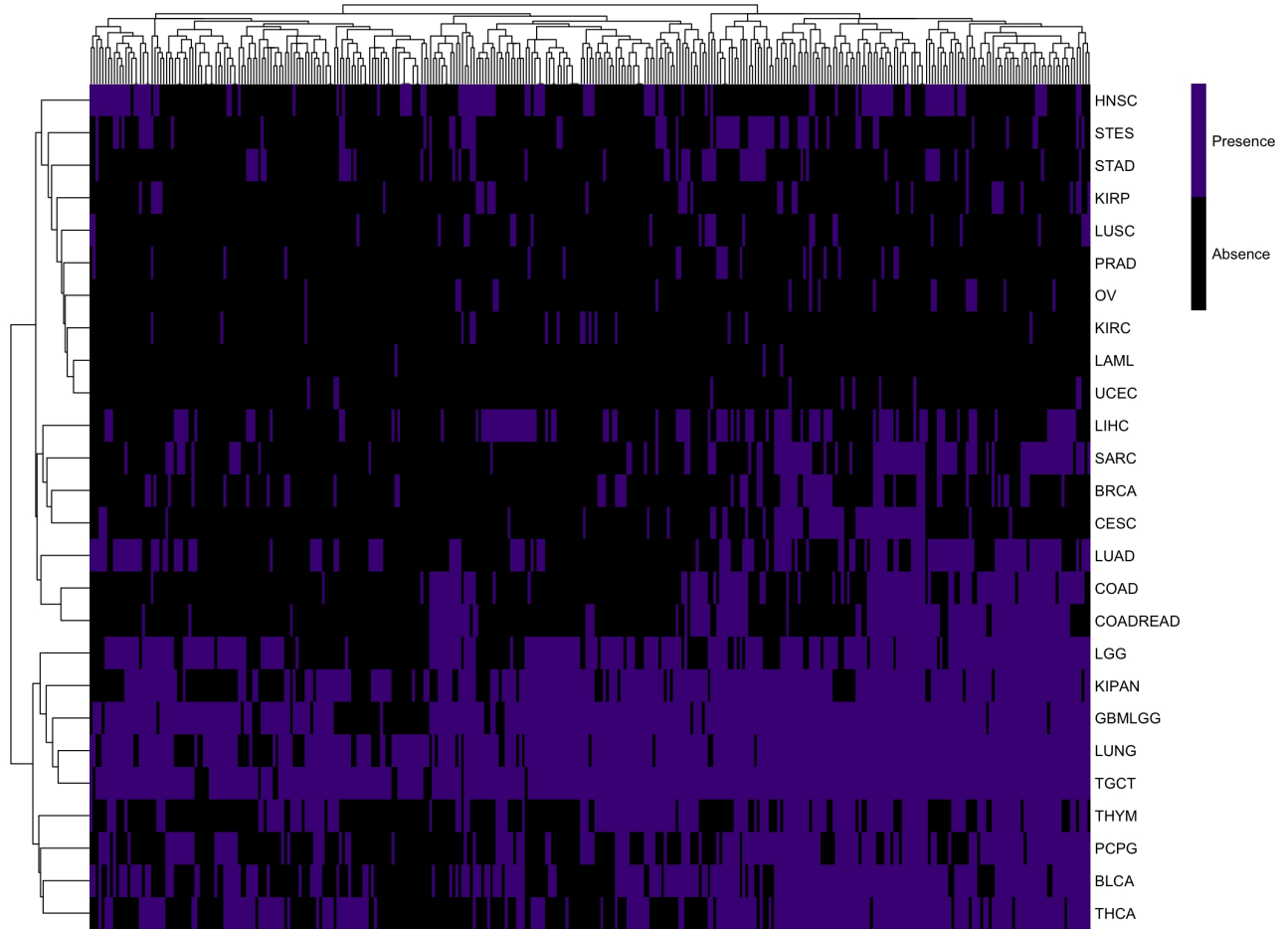


Figure 60: **Heatmap of the presence of the correlation between whole genome methylation and metabolism across different tumor types.** This heatmap represents the presence of the correlation between whole genome methylation with a given metabolism (columns) in tumors cohort (rows). The colour scale, black for absence and dark violet for presence, represents the presence of association between whole genome methylation and metabolism in every tumor types, at a global Bonferroni corrected  $p$ val  $<.05$ .

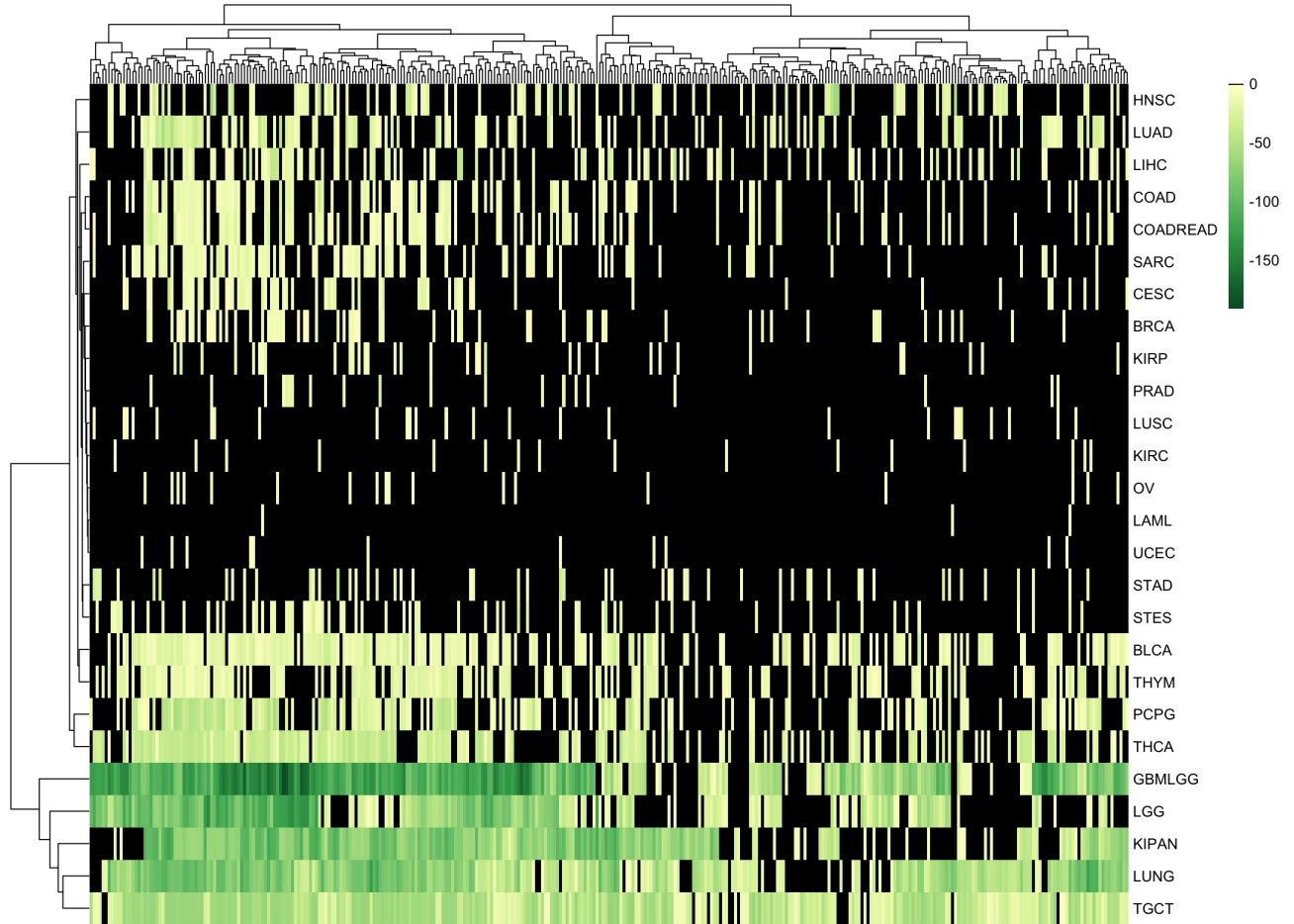


Figure 61: **Heatmap of the association strength between whole genome methylation and metabolisms across different tumor types.** This heatmap represents the correlation of whole genome methylation, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

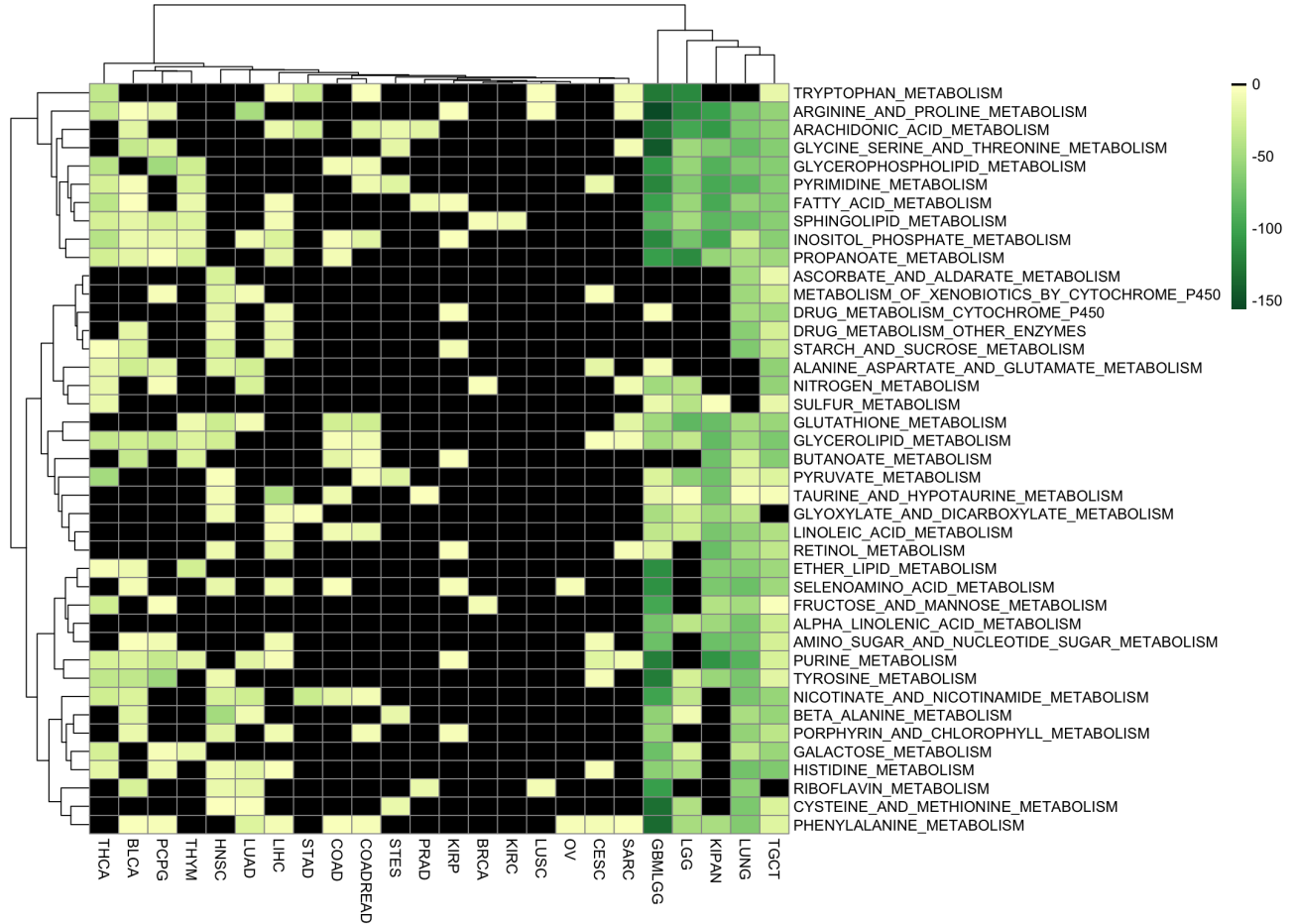


Figure 62: **Heatmap of the association strength between whole genome methylation and a subset of metabolisms across different tumor types.** This heatmap represents the correlation of whole genome methylation, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

### 7.2.3.2 Gene Specific Methylation Level

- **Tumors\_freq:** 89.5 % (34 out of 38 tumors)
- **Tumors\_top:**

	Metabolisms (n)
<b>KIPAN</b>	345
<b>TGCT</b>	342
<b>LUNG</b>	341
<b>BRCA</b>	340
<b>GBMLGG</b>	340

- **Metabolisms\_freq:** 79.9 % (345 out of 432 metabolisms)
- **Metabolisms\_top:**

	Tumors (n)
<b>GO AMINOGLYCAN METABOLIC PROCESS</b>	30
<b>GO GLUTATHIONE METABOLIC PROCESS</b>	30
<b>GO GLYCOPROTEIN METABOLIC PROCESS</b>	30
<b>GO LIPID METABOLIC PROCESS</b>	30
<b>GO NEUTRAL LIPID METABOLIC PROCESS</b>	30

- **LOC100130933:**

Tumor	Metabolisms (n)	Metabolism	Tumors (n)
KIPAN	316	GO FATTY ACID METABOLIC PROCESS	19
LUNG	312	REACTOME GLYCOSAMINOGLYCAN METABOLISM	19
GBMLGG	275	GO GLYCEROPHOSPHOLIPID METABOLIC PROCESS	18
ESCA	262	GO MONOCARBOXYLIC ACID METABOLIC PROCESS	18
BLCA	258	GO ORGANIC ACID METABOLIC PROCESS	18

Table 15: **Summary of gene methylation level results.** *Tumors\_freq/Metabolism\_freq* is the number of tumors/metabolisms in which at least one correlation between any gene methylation level and any metabolism/tumor is found. *Tumors\_top/Metabolism\_top* are the top five tumors/metabolisms sorted by number of metabolisms/tumors correlating with at least one gene methylation level. The last table represents the top five tumors/metabolisms in which the specific gene methylation level has been observed.



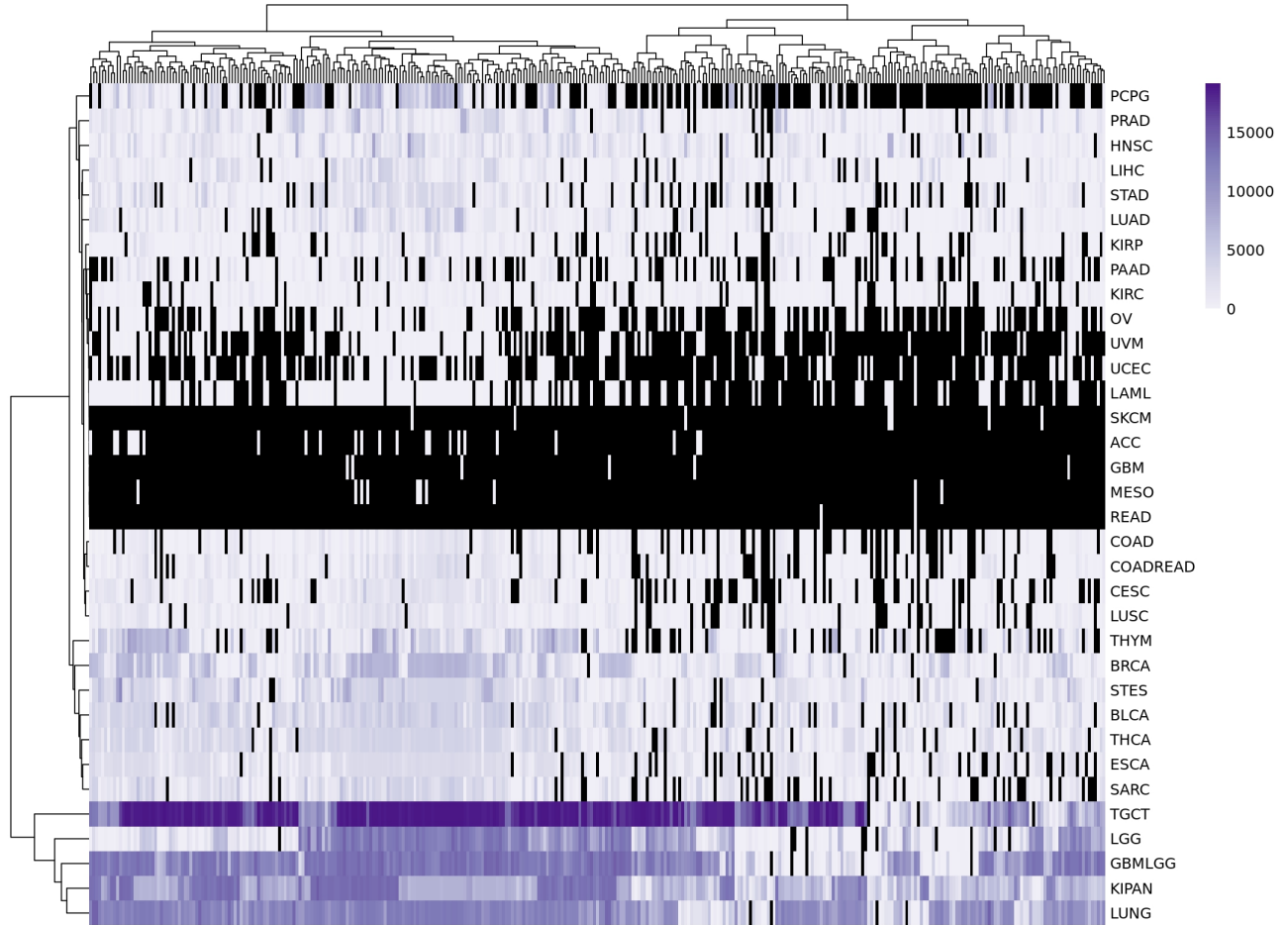


Figure 63: **Heatmap of the amount of gene methylation levels correlated with metabolism across different tumor types.** This heatmap represents the number of CpG sites correlating with a specific metabolism (columns) in tumors cohort (rows). The colour scale, from white to dark violet, represents the amount of gene methylation levels; black means no observation found at a global Bonferroni corrected  $p$ val  $<.05$ .

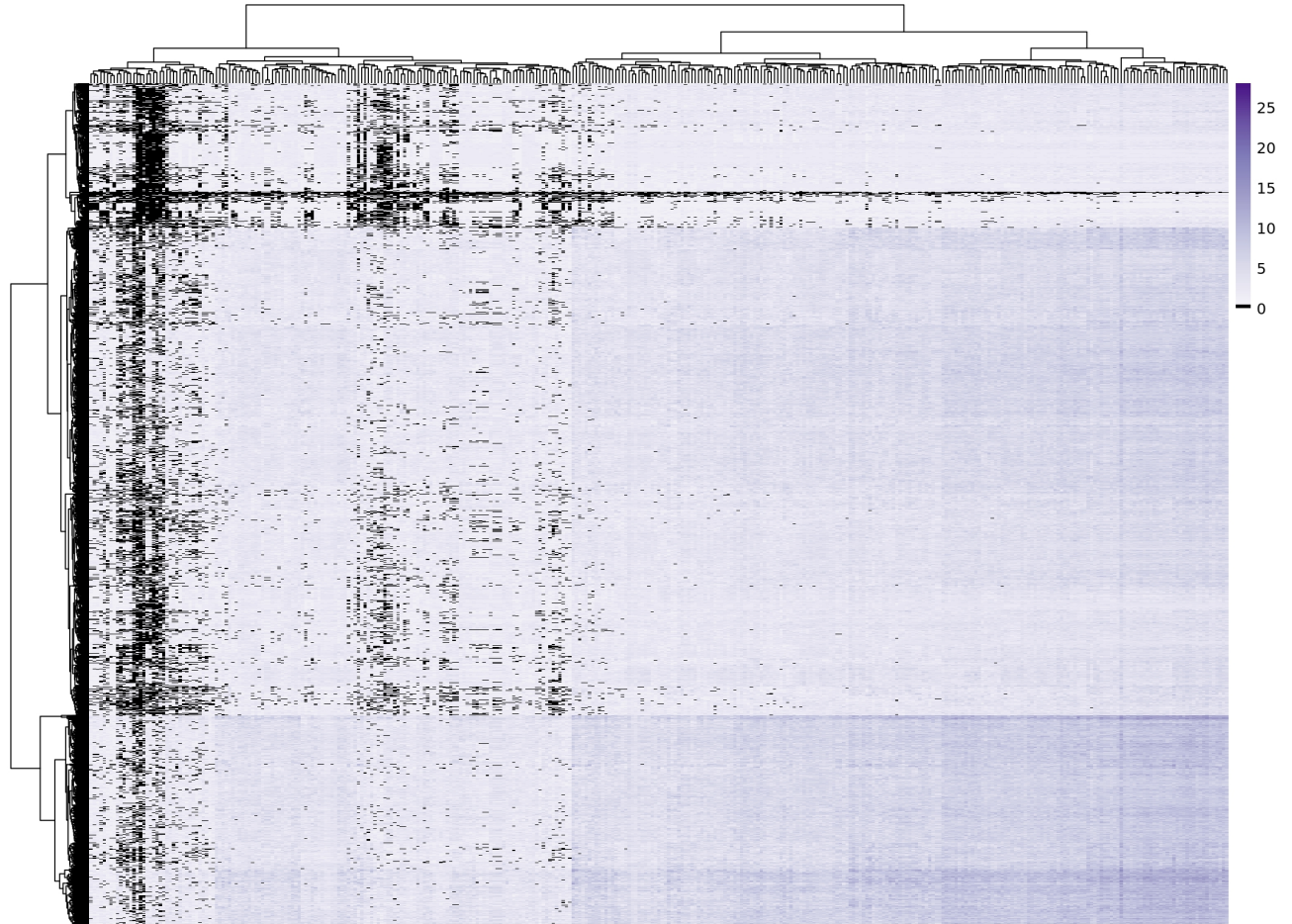


Figure 64: **Heatmap of the frequency of association between single gene methylation levels and metabolisms.** This heatmap represents the number of tumors in which a correlation between CpG sites (rows) and a metabolisms (columns) has been observed. The colour scale, from white to dark violet, represents the number of tumors; black means no observation found at a global Bonferroni corrected  $p$ val  $<.05$ .

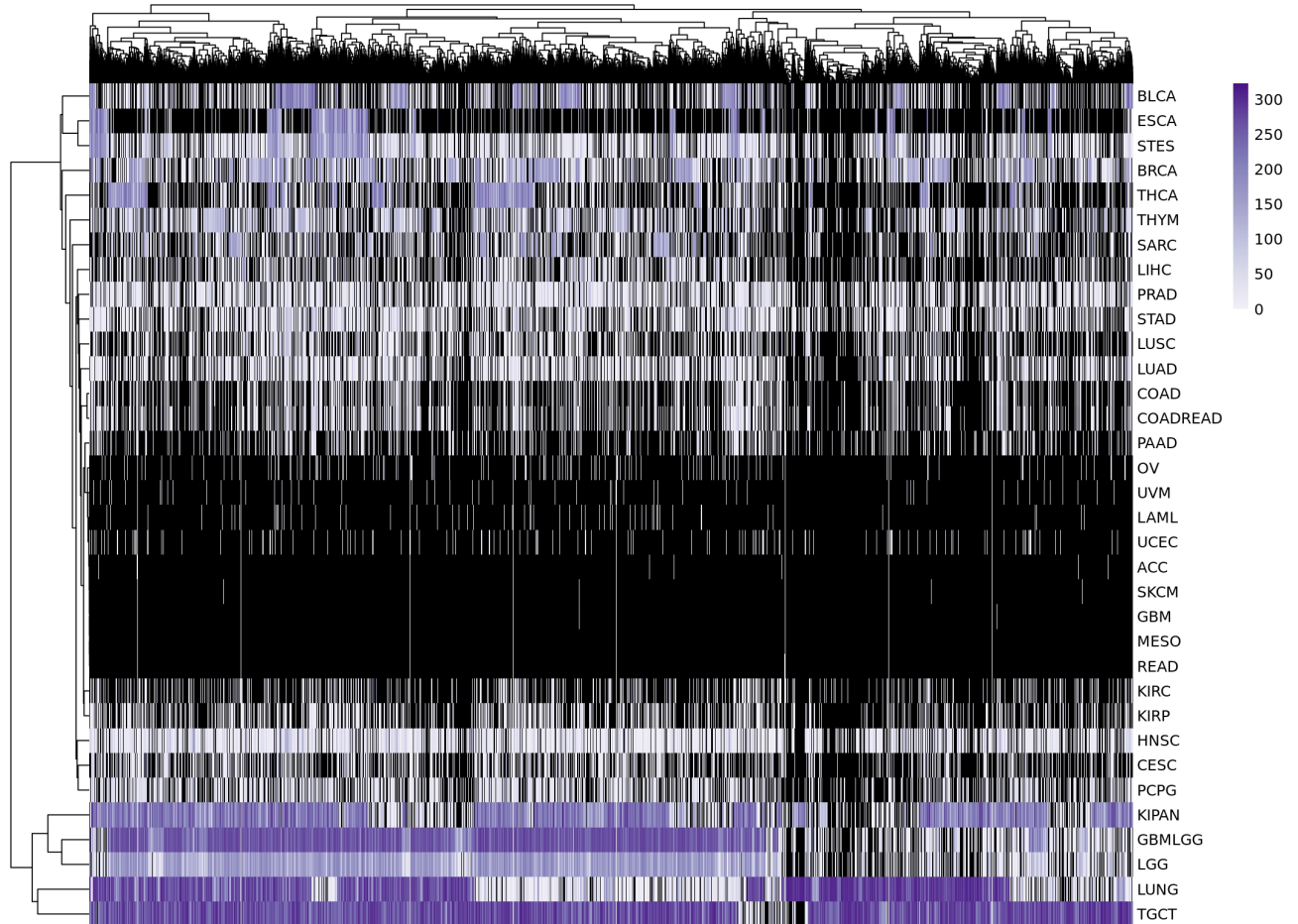


Figure 65: **Heatmap of the frequency of metabolisms correlating with gene specific methylation levels.** This heatmap represents the number of metabolisms that correlates with gene specific methylation levels (columns) in different tumor types (rows). The colour scale, from white to dark violet, represents the number of tumors; black means no observation found at a global Bonferroni corrected pval <.05.

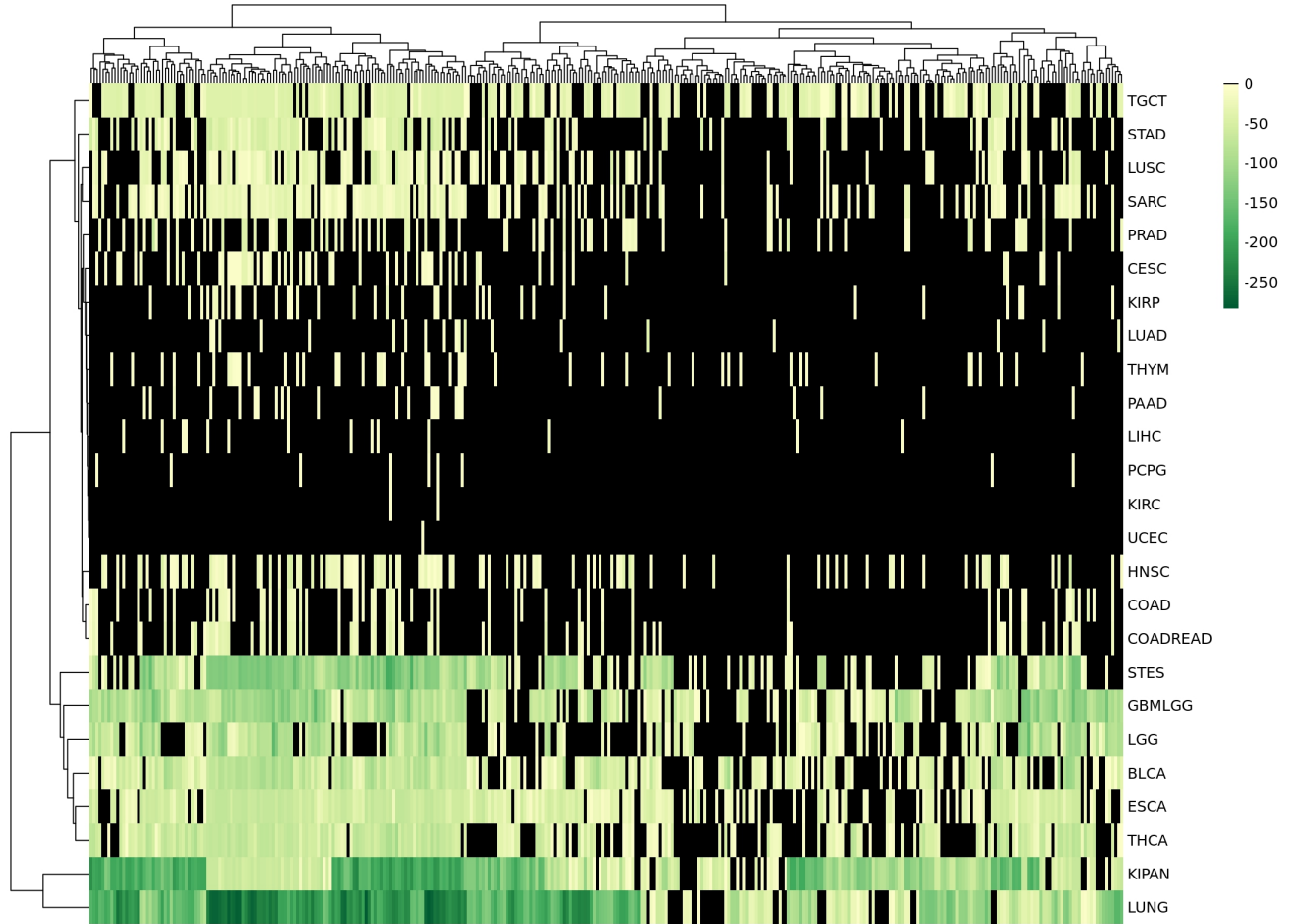


Figure 66: **Heatmap of the association strength between a specific gene methylation level and metabolisms.** This heatmap represents the correlation of the specific *LOC100130933* methylation level, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected pvalues; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

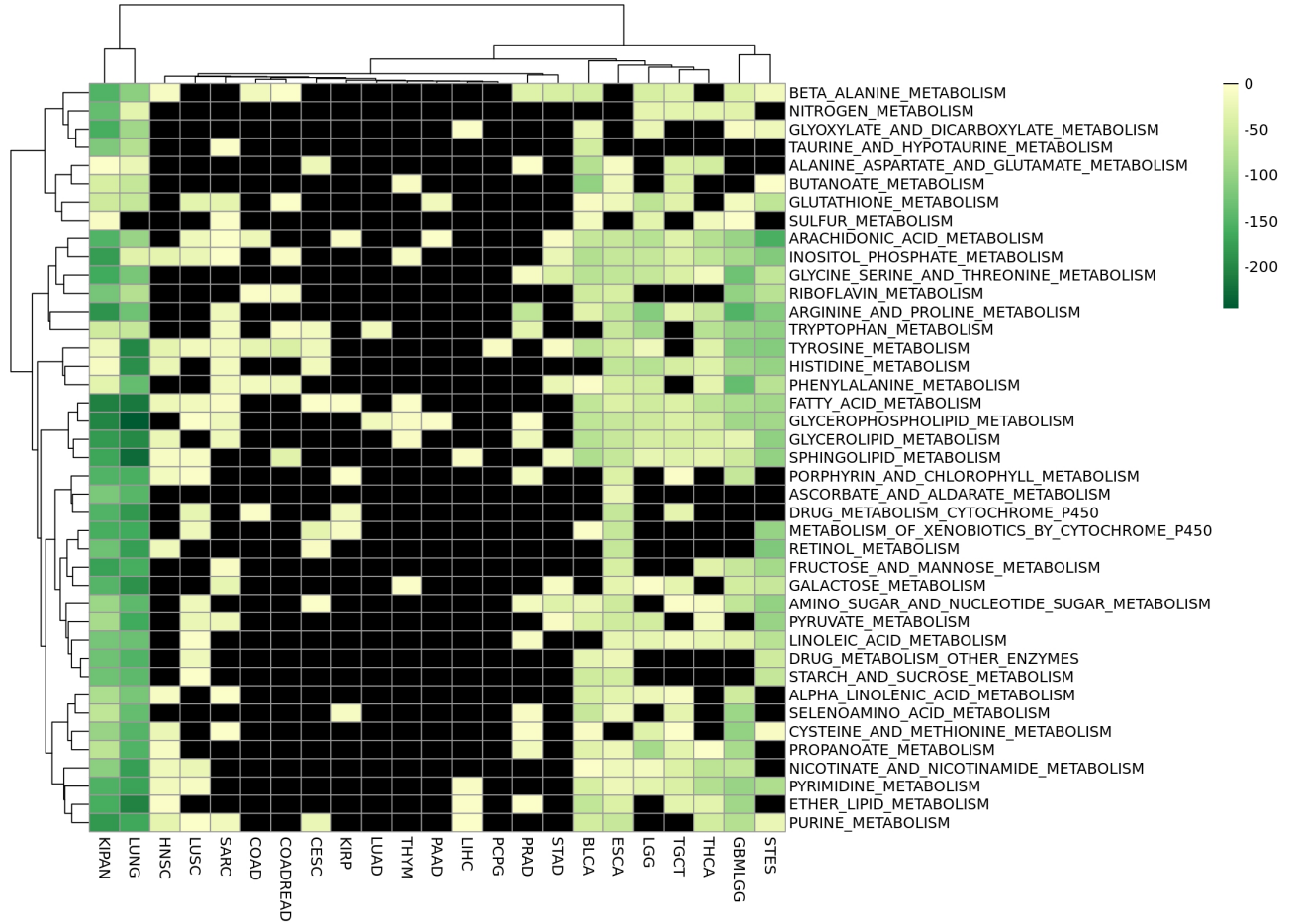


Figure 67: **Heatmap of the association strength between a specific gene methylation level and a subset of metabolisms.** This heatmap represents the correlation of the specific *LOC100130933* methylation level, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

## 7.2.4 Reverse Phase Protein Array Data

- **Tumors\_freq:** 68.4 % (26 out of 38 tumors)

- **Tumors\_top:**

	Metabolisms (n)
<b>TGCT</b>	336
<b>BRCA</b>	322
<b>LGG</b>	317
<b>GBMLGG</b>	313
<b>KIRC</b>	310

- **Metabolisms\_freq:** 79.9 % (345 out of 432 metabolisms)

- **Metabolisms\_top:**

	Tumors (n)
<b>GO GLYCOPROTEIN METABOLIC PROCESS</b>	23
<b>GO AROMATIC AMINO ACID FAMILY METABOLIC PROCESS</b>	22
<b>GO CELLULAR AMINO ACID METABOLIC PROCESS</b>	22
<b>GO FATTY ACID METABOLIC PROCESS</b>	22
<b>GO STEROID METABOLIC PROCESS</b>	22

- **AR|AR:**

Tumor	Metabolisms (n)	Metabolism	Tumors (n)
BRCA	236	GO MONOCARBOXYLIC ACID METABOLIC PROCESS	9
LGG	205	GO PHOSPHOLIPID METABOLIC PROCESS	9
GBMLGG	180	GO CARBOHYDRATE METABOLIC PROCESS	8
KIRC	149	GO CELLULAR HORMONE METABOLIC PROCESS	8
LIHC	135	GO CELLULAR LIPID METABOLIC PROCESS	8

Table 16: **Summary of protein abundance results.** *Tumors\_freq/Metabolism\_freq* is the number of tumors/metabolisms in which at least one correlation between any protein production level and any metabolism/tumor is found. *Tumors\_top/Metabolism\_top* are the top five tumors/metabolisms sorted by number of metabolisms/tumors correlating with at least one protein production level. The last table represents the top five tumors/metabolisms in which the specific protein production level has been observed.

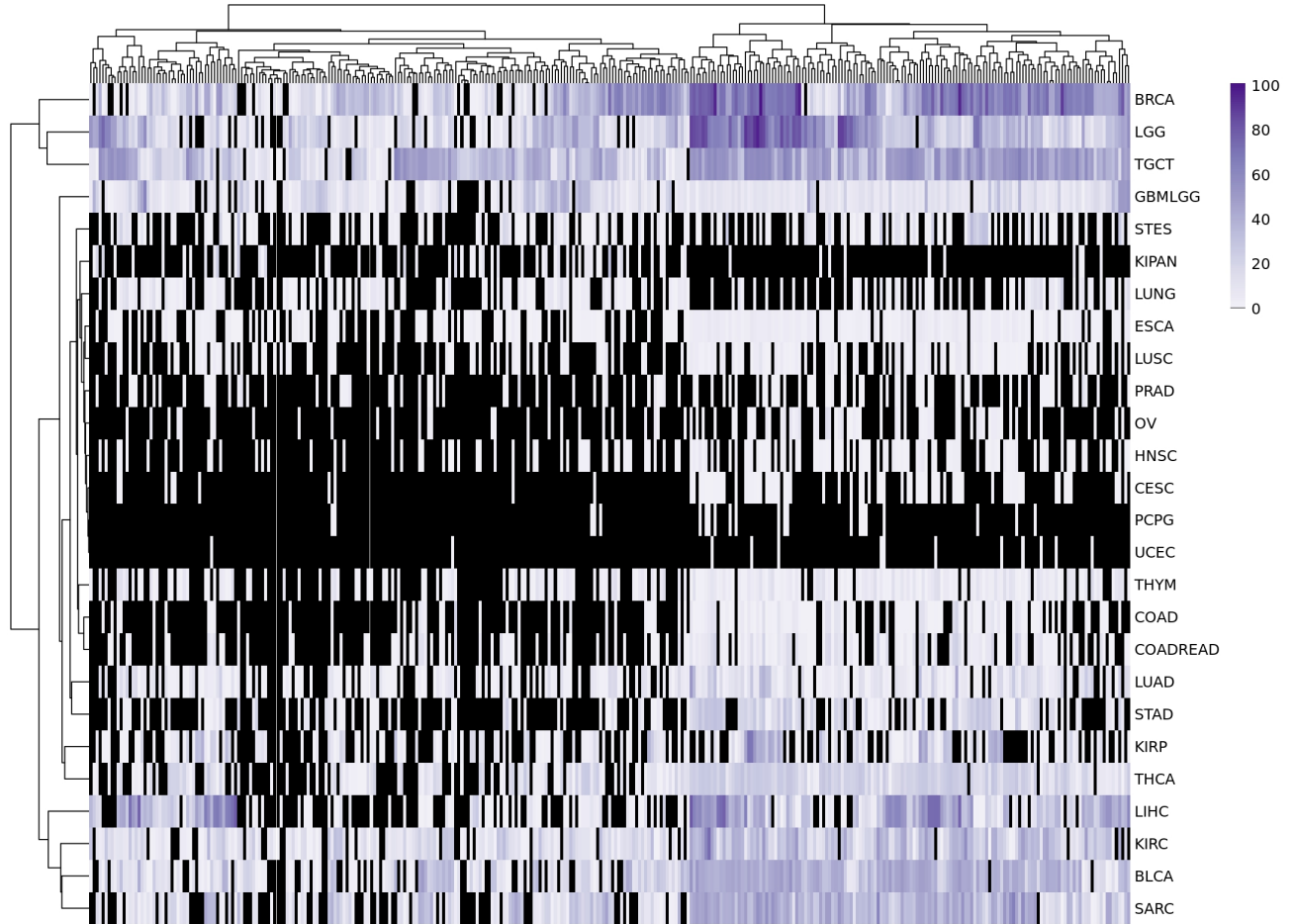


Figure 68: **Heatmap of the amount of protein production levels correlated with metabolism across different tumor types.** This heatmap represents the number of protein production levels correlating with a specific metabolisms (columns) in tumors cohort (rows). The colour scale, from white to dark violet, represents the amount of protein production levels; black means no observation found at a global Bonferroni corrected pval  $<.05$ .



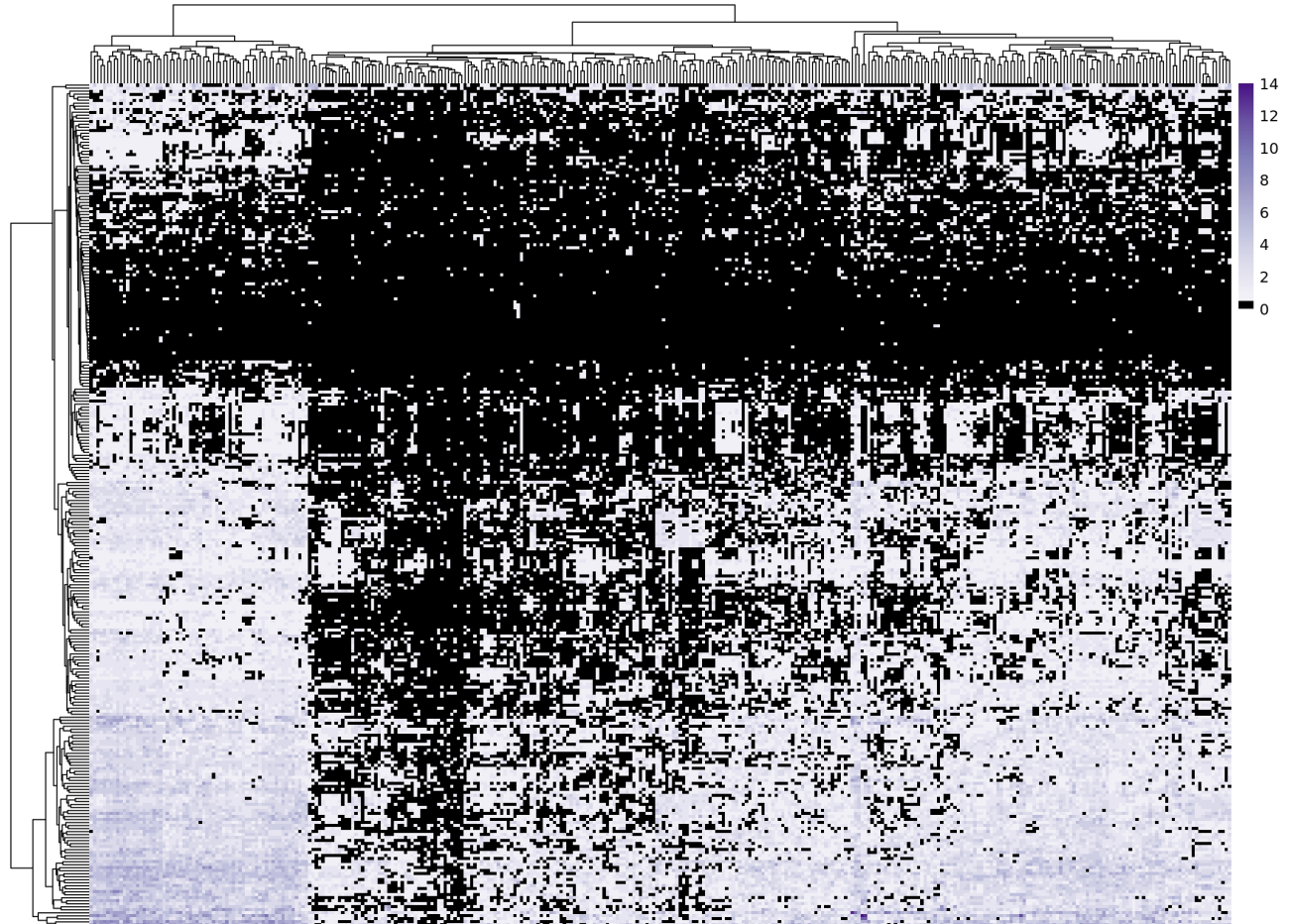


Figure 69: **Heatmap of the frequency of association between protein production levels and metabolisms.** This heatmap represents the number of tumors in which a correlation between protein production level (columns) and a metabolism (rows) has been observed. The colour scale, from white to dark violet, represents the number of tumors; black means no observation found at a global Bonferroni corrected pval <.05.

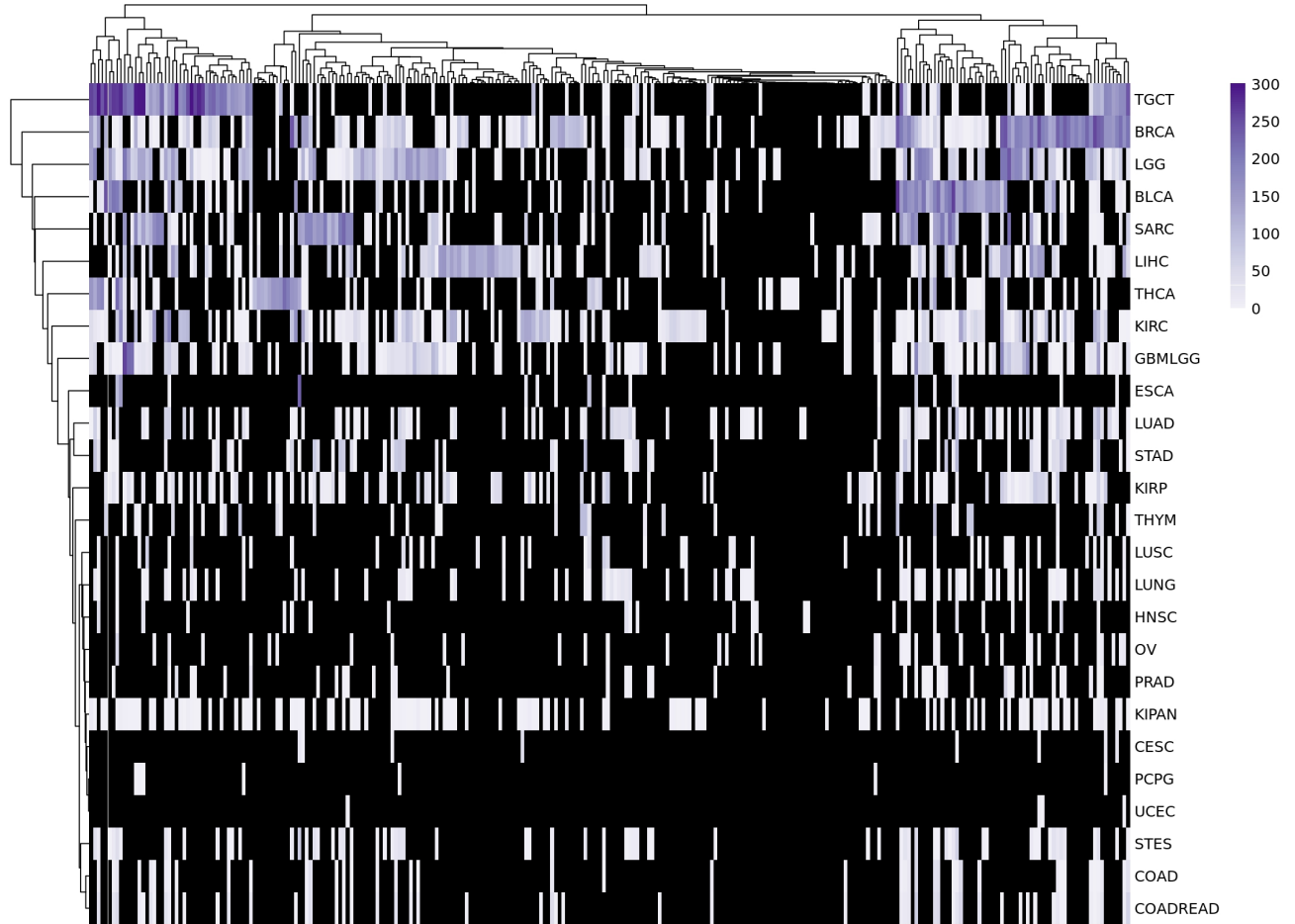


Figure 70: **Heatmap of the frequency of metabolisms.** This heatmap represents the number of metabolisms that correlates with a specific protein production level (columns) in different tumor types (rows). The colour scale, from white to dark violet, represents the number of tumors; black means no observation found at a global Bonferroni corrected pval < .05.

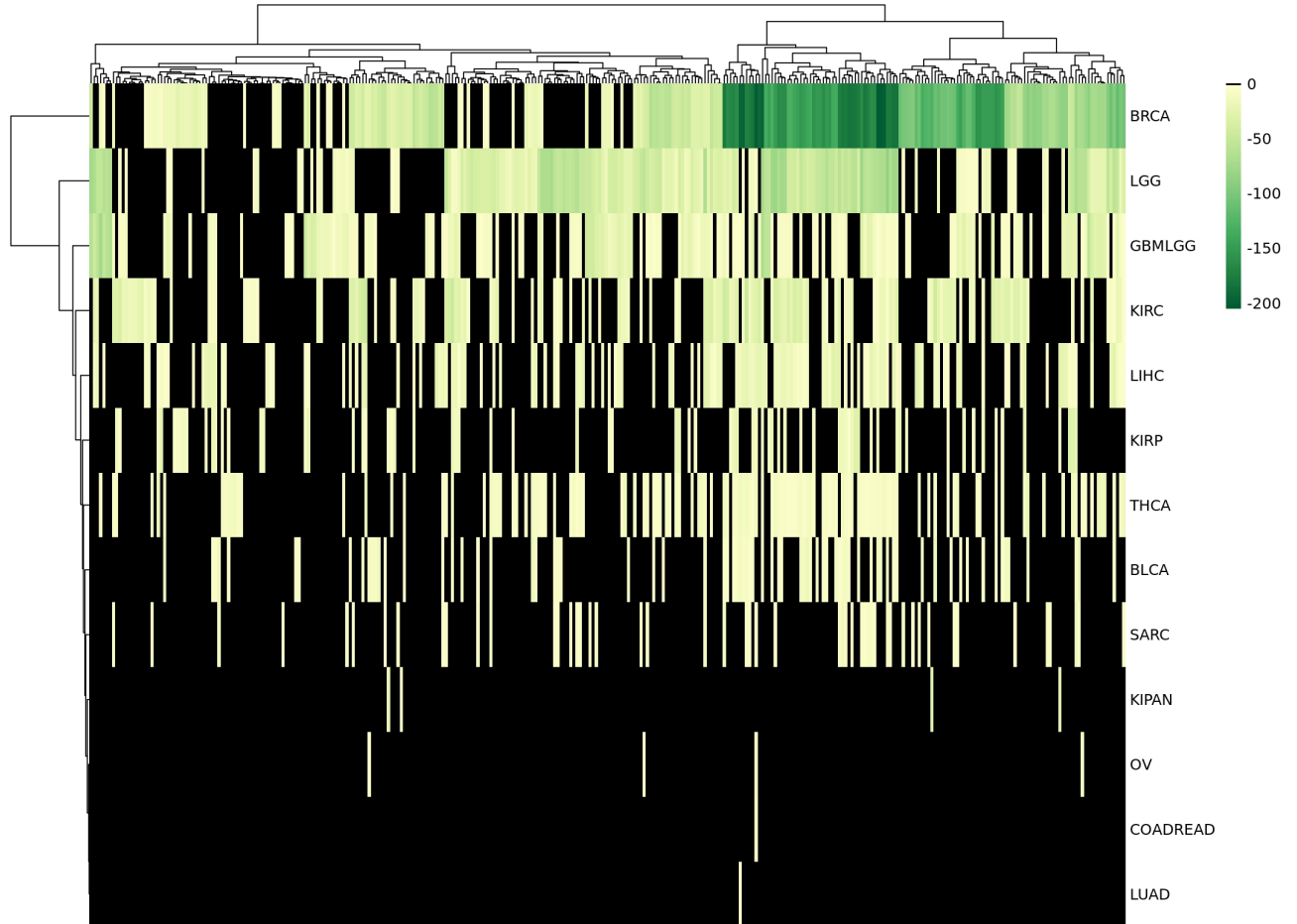


Figure 71: **Heatmap of the association strength between a specific protein production level and metabolisms.** This heatmap represents the correlation of the specific *AR* abundance level, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

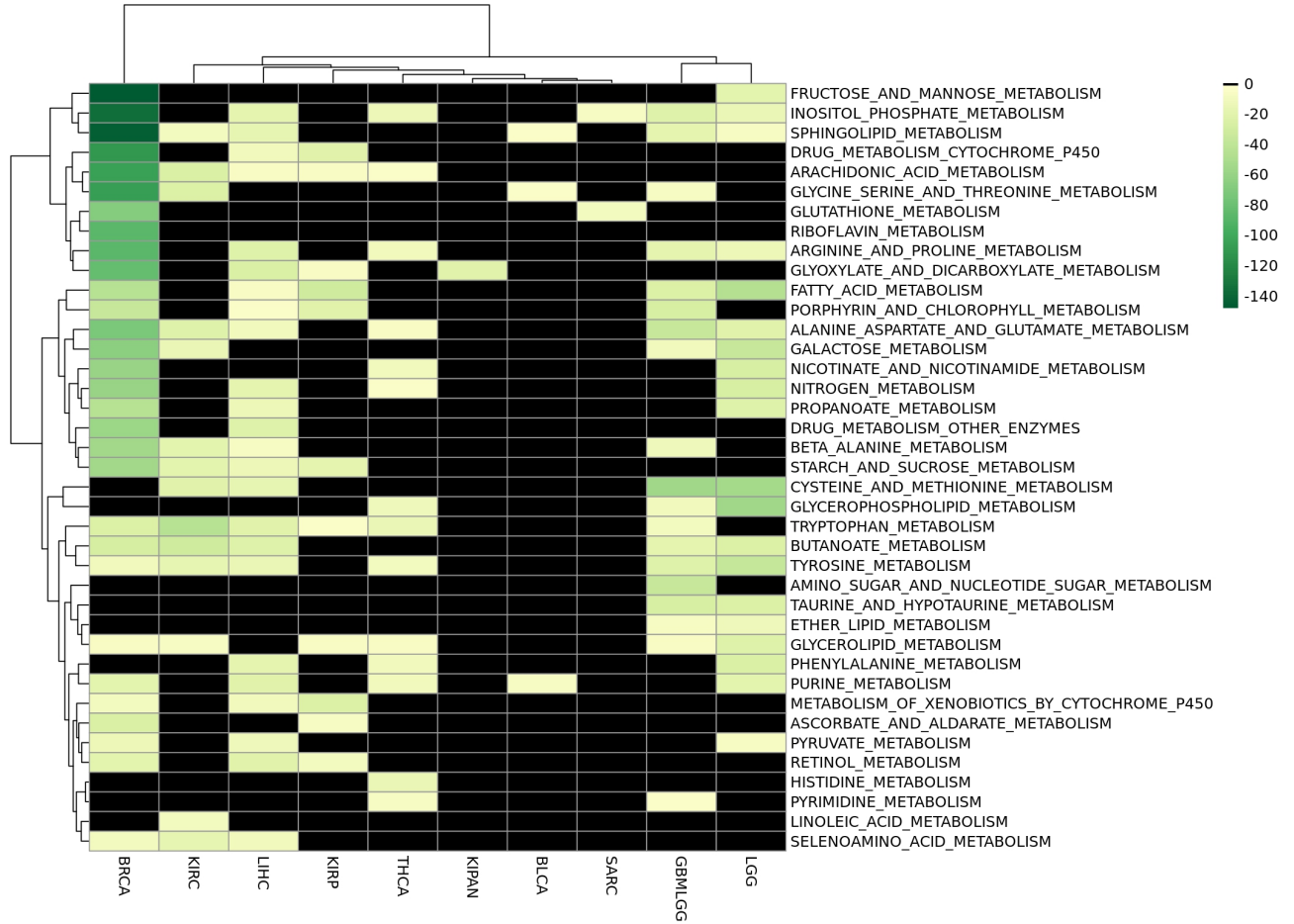


Figure 72: **Heatmap of the association strength between a specific protein production level and a subset of metabolisms.** This heatmap represents the correlation of the specific *AR* abundance level, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

## 7.2.5 miRNA Expression Data

- **Tumors\_freq:** 78.9 % (30 out of 38 tumors)
- **Tumors\_top:**

	Metabolisms (n)
<b>KIPAN</b>	339
<b>LUNG</b>	328
<b>LGG</b>	307
<b>GBMLGG</b>	304
<b>BRCA</b>	298

- **Metabolisms\_freq:** 79.9 % (345 out of 432 metabolisms)
- **Metabolisms\_top:**

	Tumors (n)
<b>GO AMINOGLYCAN METABOLIC PROCESS</b>	26
<b>GO AMINE METABOLIC PROCESS</b>	25
<b>GO MUCOPOLYSACCHARIDE METABOLIC PROCESS</b>	24
<b>GO ORGANIC ACID METABOLIC PROCESS</b>	24
<b>GO PHENOL CONTAINING COMPOUND METABOLIC PROCESS</b>	24

- **hsa-mir-375:**

Tumor	Metabolisms (n)	Metabolism	Tumors (n)
LUNG	294	GO GLYCOPROTEIN METABOLIC PROCESS	10
TGCT	219	GO INOSITOL PHOSPHATE METABOLIC PROCESS	10
STES	185	GO PYRUVATE METABOLIC PROCESS	10
ESCA	178	GO ALCOHOL METABOLIC PROCESS	9
THCA	165	GO ORGANIC ACID METABOLIC PROCESS	9

Table 17: **Summary of mirna expression level results.** *Tumors\_freq/Metabolism\_freq* is the number of tumors/metabolisms in which at least one correlation between any mirna expression level and any metabolism/tumor is

found. *Tumors\_top/Metabolism\_top* are the top five tumors/metabolisms sorted by number of metabolisms/tumors correlating with at least one mirna expression level. The last table represents the top five tumors/metabolisms in which the specific mirna expression level has been observed.

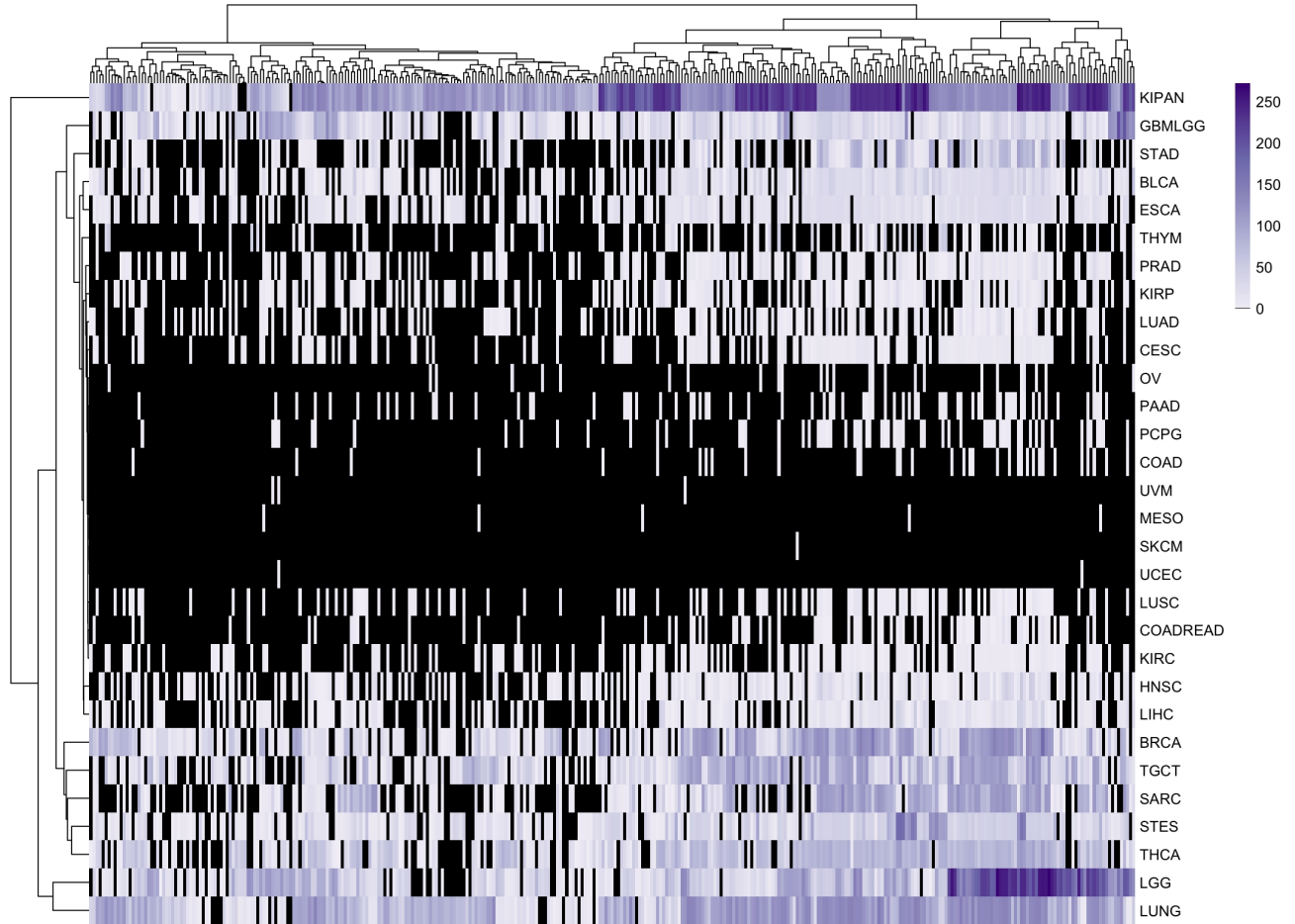


Figure 73: **Heatmap of the amount of mirna expression levels correlated with metabolism across different tumor types.** This heatmap represents the number of mirna expression levels correlating with a specific metabolism (columns) in tumors cohort (rows). The colour scale, from white to dark violet, represents the amount of mirna expression levels; black means no observation found at a global Bonferroni corrected pval  $<.05$ .



Figure 74: **Heatmap of the frequency of association between mirna expression levels and metabolisms.** This heatmap represents the number of tumors in which a correlation between mirna expression level (columns) and a metabolism (rows) has been observed. The colour scale, from white to dark violet, represents the number of tumors; black means no observation found at a global Bonferroni corrected pval < .05.



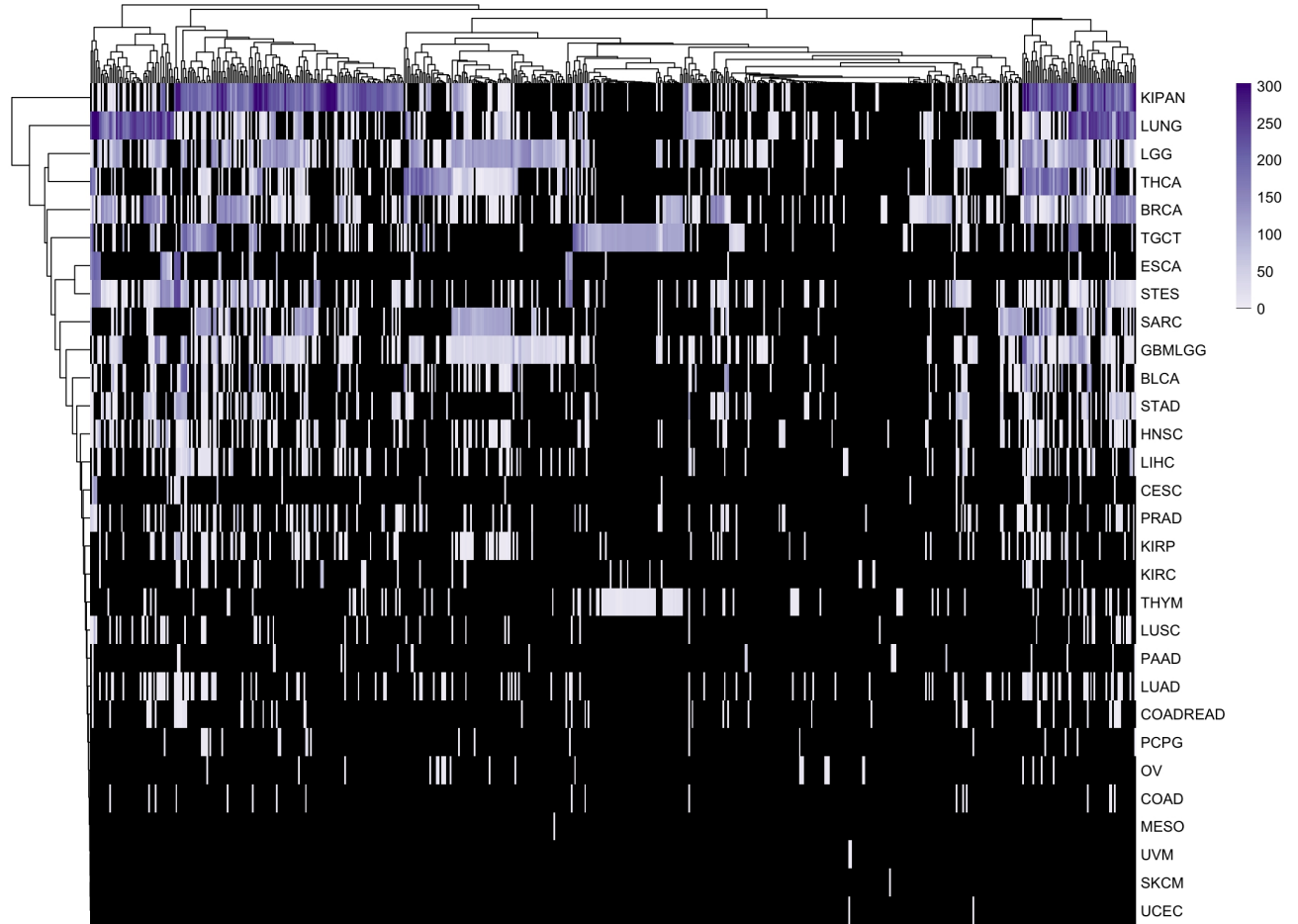


Figure 75: **Heatmap of the frequency of metabolisms associated to differentially expressed miRNAs across different tumor types.** This heatmap represents the number of metabolisms that correlates with a specific mirna expression level (columns) in different tumor types (rows). The colour scale, from white to dark violet, represents the number of tumors; black means no observation found at a global Bonferroni corrected pval <.05.

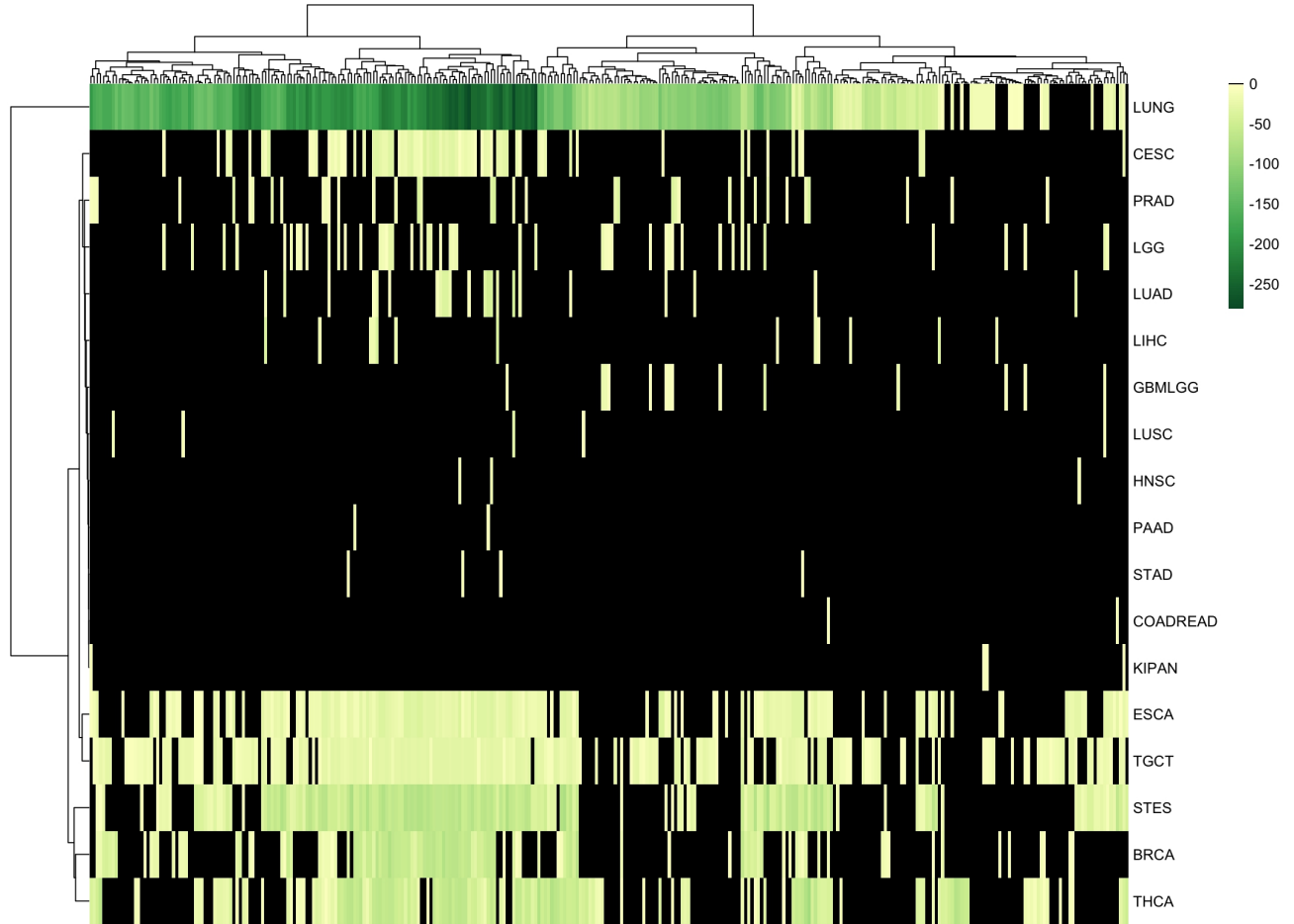


Figure 76: **Heatmap of the association strength between a specific mirna expression level and metabolisms.** This heatmap represents the correlation of the specific *hsa-mir-514-1*: expression level, with different metabolisms (columns) in different tumors (rows). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected pvalues; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

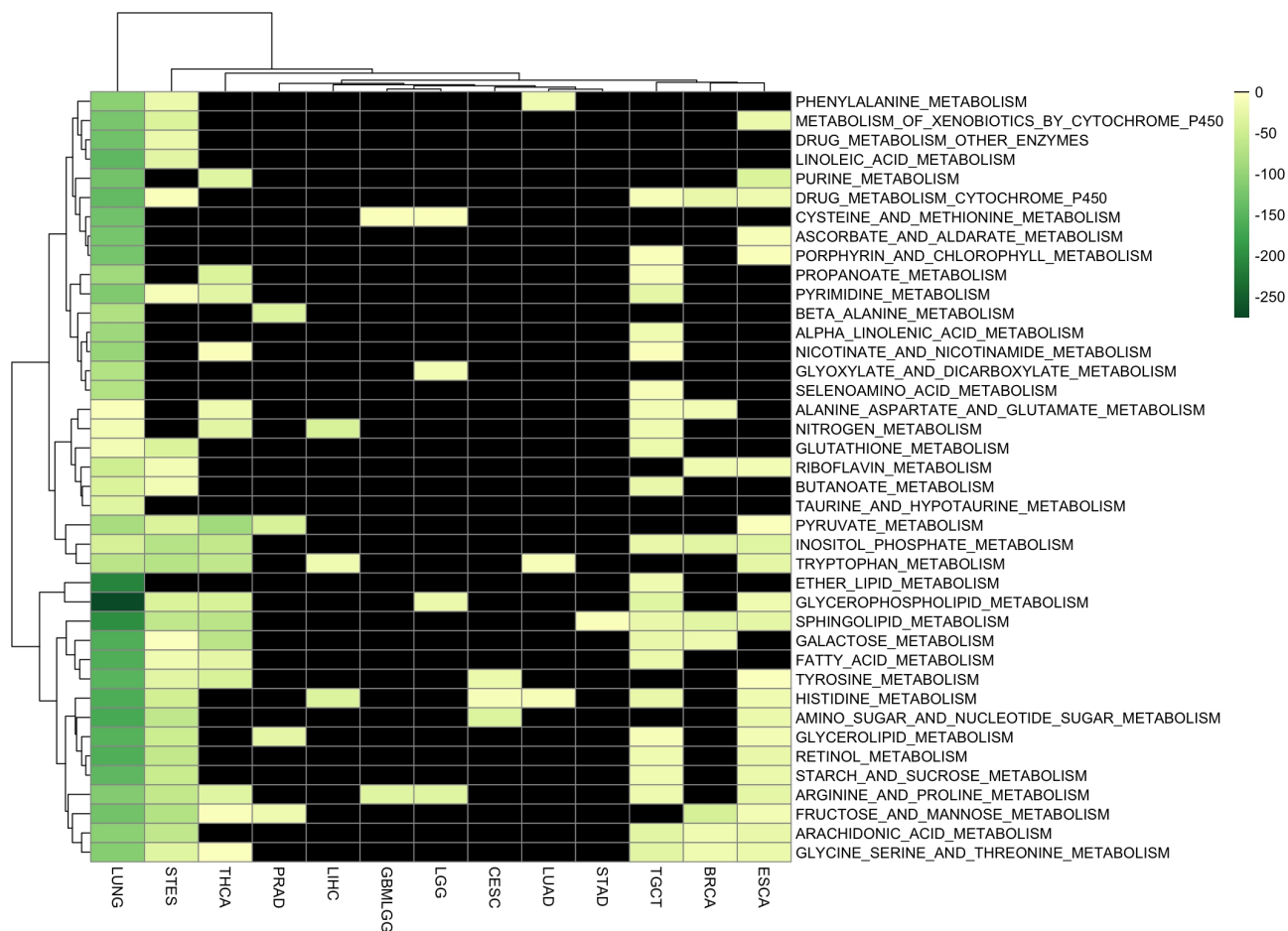


Figure 77: **Heatmap of the association strength between a specific mirna expression level and a subset of metabolisms.** This heatmap represents the correlation of the specific *hsa-mir-514-1*: expression level, with different metabolisms (rows) in different tumors (columns). The colour scale, from light to dark green, represents the  $\log_2$  of the globally Bonferroni-corrected p-values; black means lack of statistical significance ( $\log_2(pval_{Bonf}) > -1.3$ ).

## 8 Publications List

- Fusella F, Seclì L, Busso E, Krepelova A, **Moiso E**, Rocca S, Conti L, Annaratone L, Rubinetto C, Mello-Grand M, Singh V, Chiorino G, Silengo L, Altruda F, Turco E, Morotti A, Oliviero S, Castellano I, Cavallo F, Provero P, Tarone G, Brancaccio M.  
*The IKK/NF- $\kappa$ B signaling pathway requires Morgana to drive breast cancer metastasis.*  
Nat Commun. 2017 Nov 21;8(1):1636. doi: 10.1038/s41467-017-01829-1
- Salaroglio IC, Panada E, **Moiso E**, Buondonno I, Provero P, Rubinstein M, Kopecka J, Riganti C.  
*PERK induces resistance to cell death elicited by endoplasmic reticulum stress and chemotherapy.*  
Mol Cancer. 2017 May 12;16(1):91. doi: 10.1186/s12943-017-0657-0.
- Grasso S, Chapelle J, Salemme V, Aramu S, Russo I, Vitale N, Verdun di Cantogno L, Dallaglio K, Castellano I, Amici A, Centonze G, Sharma N, Lunardi S, Cabodi S, Cavallo F, Lamolinara A, Stramucci L, **Moiso E**, Provero P, Albini A, Sapino A, Staaf J, Di Fiore PP, Bertalot G, Pece S, Tosoni D, Confalonieri S, Iezzi M, Di Stefano P, Turco E, Defilippi P.  
*The scaffold protein p140Cap limits ERBB2-mediated breast cancer progression interfering with Rac GTPase-controlled circuitries.*  
Nat Commun. 2017 Mar 16;8:14797. doi: 10.1038/ncomms14797.
- **Moiso E**, Accardo M, Tamagnone L.  
*Experimental Approaches for Studying Semaphorin Signals in Tumor Growth and Metastasis in Mouse Models.*  
Methods Mol Biol. 2017;1493:467-484.
- Perico ME, Grasso S, Brunelli M, Martignoni G, Munari E, **Moiso E**, Fracasso G, Cestari T, Naim HY, Bronte V, Colombatti M, Ramarli D.  
*Prostate-specific membrane antigen (PSMA) assembles a macromolecular complex regulating growth and survival of prostate cancer cells “in vitro” and correlating with progression “in vivo”.*  
Oncotarget. 2016 Nov 8;7(45):74189-74202. doi: 10.18632/oncotarget.12404.

## 9 References

- Alexandrov, Ludmil B., Serena Nik-Zainal, David C. Wedge, Samuel A.J.R. Aparicio, Sam Behjati, Andrew V. Biankin, Graham R. Bignell, et al. 2013. “Signatures of mutational processes in human cancer.” *Nature* 500 (7463): 415–21. doi:10.1038/nature12477.
- Anderson, Douglas M, Catherine A Makarewich, Kelly M Anderson, John M Shelton, Svetlana Bezprozvannaya, Rhonda Bassel-Duby, and Eric N Olson. 2016. “Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides.” *Science Signaling* 9 (457). American Association for the Advancement of Science: ra119. doi:10.1126/scisignal.aaj1460.
- Arbin, Norazam, Nur Suhailayani Suhaimi, Nurul Zafirah Mokhtar, and Zalinda Othman. 2015. “Comparative Analysis between K-Means and K-Medoids for Statistical Clustering.” In *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (Aims)*, 117–21. IEEE. doi:10.1109/AIMS.2015.82.
- Bakhoun, Samuel F, and Duane A Compton. 2012. “Chromosomal instability and cancer: a complex relationship with therapeutic potential.” *The Journal of Clinical Investigation* 122 (4): 1138–43. doi:10.1172/JCI59954.
- Berg, Jeremy M, John L Tymoczko, and Lubert Stryer. 2002. “The Glycolytic Pathway Is Tightly Controlled.” W H Freeman. <https://www.ncbi.nlm.nih.gov/books/NBK22395/>.
- Blazier, Anna S., and Jason A. Papin. 2012. “Integration of expression data in genome-scale metabolic network reconstructions.” *Frontiers in Physiology* 3 (August). Frontiers: 299. doi:10.3389/fphys.2012.00299.
- Bordbar, Aarash, Jonathan M. Monk, Zachary A. King, and Bernhard O. Palsson. 2014. “Constraint-based models predict metabolic and associated cellular functions.” *Nature Reviews Genetics* 15 (2). Nature Publishing Group: 107–20. doi:10.1038/nrg3643.
- Cai, Guoqing, Jian Wang, Xiaoyan Xin, Zunji Ke, and Jia Luo. 2007. *International journal of oncology*. Vol. 31. 3. University of Crete, Faculty of Medicine, Laboratory of Clinical Virology. <https://www.spandidos-publications.com/ijo/31/3/657>.
- Camps, Carme, Francesca M Buffa, Stefano Colella, John Moore, Christos Sotiriou, Helen Sheldon, Adrian L Harris, Jonathan M Gleadly, and Jiannis Ragoussis. 2008. “hsa-miR-210 Is induced by hypoxia and is an independent prognostic factor in breast cancer.” *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 14 (5). American Association for Cancer Research: 1340–8. doi:10.1158/1078-

0432.CCR-07-1755.

Cantor, Jason R, and David M Sabatini. 2012. "Cancer cell metabolism: one hallmark, many faces." *Cancer Discovery* 2 (10). American Association for Cancer Research: 881–98. doi:10.1158/2159-8290.CD-12-0345.

Chalmers, Zachary R., Caitlin F. Connelly, David Fabrizio, Laurie Gay, Siraj M. Ali, Riley Ennis, Alexa Schrock, et al. 2017. "Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden." *Genome Medicine* 9 (1). doi:10.1186/s13073-017-0424-2.

Colombo, Sergio L, Miriam Palacios-Callender, Nanci Frakich, Joel De Leon, Christoph A Schmitt, Leanne Boorn, Nicola Davis, and Salvador Moncada. 2010. "Anaphase-promoting complex/cyclosome-Cdh1 coordinates glycolysis and glutaminolysis with transition to S phase in human T lymphocytes." *Proceedings of the National Academy of Sciences of the United States of America* 107 (44). National Academy of Sciences: 18868–73. doi:10.1073/pnas.1012362107.

Cox, D. R. 1972. "Regression Models and Life-Tables." WileyRoyal Statistical Society. doi:10.2307/2985181.

Curi, R, P Newsholme, and E A Newsholme. 1988. "Metabolism of pyruvate by isolated rat mesenteric lymphocytes, lymphocyte mitochondria and isolated mouse macrophages." *The Biochemical Journal* 250 (2): 383–88. doi:10.1042/bj2500383.

Dang, L., K. Yen, and E. C. Attar. 2016. "IDH mutations in cancer and progress toward development of targeted therapeutics." *Annals of Oncology* 27 (4): 599–608. doi:10.1093/annonc/mdw013.

Daróczi, Gergely, and Roman Tsegelskyi. 2017. *Pander: An R 'Pandoc' Writer*. <http://CRAN.R-project.org/package=pander>.

Deiuliis, J A. 2016. "MicroRNAs as regulators of metabolic disease: pathophysiologic significance and emerging role as biomarkers and therapeutics." *International Journal of Obesity (2005)* 40 (1). Nature Publishing Group: 88–101. doi:10.1038/ijo.2015.170.

Dossus, Laure, and Patrick R Benusiglio. 2015. "Lobular breast cancer: incidence and genetic and non-genetic risk factors." *Breast Cancer Research : BCR* 17 (March). BioMed Central: 37. doi:10.1186/s13058-015-0546-7.

Dowle, Matt, and Arun Srinivasan. 2017. *Data.table: Extension of 'Data.frame'*. <http://CRAN.R-project.org/>



package=data.table.

Eagle, H. 1955. "Nutrition needs of mammalian cells in tissue culture." *Science (New York, N.Y.)* 122 (3168): 501–14. <http://www.ncbi.nlm.nih.gov/pubmed/13255879>.

Ehrlich, Melanie. 2002. "DNA methylation in cancer: too much, but also too little." *Oncogene* 21 (35). Nature Publishing Group: 5400–5413. doi:10.1038/sj.onc.1205651.

Fisher, R. A. 1922. "On the Interpretation of  $\chi^2$  from Contingency Tables, and the Calculation of P." *Journal of the Royal Statistical Society* 85 (1). WileyRoyal Statistical Society: 87. doi:10.2307/2340521.

Gaude, Edoardo, and Christian Frezza. 2016. "Tissue-specific and convergent metabolic transformation of cancer correlates with metastatic potential and patient survival." *Nature Communications* 7. doi:10.1038/ncomms13041.

Gentleman, R., V. Carey, W. Huber, and F. Hahne. n.d. *Genefilter: Genefilter: Methods for Filtering Genes from High-Throughput Experiments*.

Giam, Maybelline, and Giulia Rancati. 2015. "Aneuploidy and chromosomal instability in cancer: A jackpot to chaos." doi:10.1186/s13008-015-0009-7.

Giovannini, Catia, Michele Baglioni, Marco Baron Toaldo, Cristiano Ventrucchi, Stefania D'Adamo, Mario Cipone, Pasquale Chieco, et al. 2013. "Notch3 inhibition enhances sorafenib cytotoxic efficacy by promoting GSK3 $\beta$  phosphorylation and p21 down-regulation in hepatocellular carcinoma." *Oncotarget* 4 (10). Impact Journals: 1618–31. doi:10.18632/oncotarget.1221.

Graham, Nicholas A, Aspram Minasyan, Anastasia Lomova, Ashley Cass, Nikolas G Balanis, Michael Friedman, Shawna Chan, et al. 2017. "Recurrent patterns of DNA copy number alterations in tumors reflect metabolic selection pressures." *Molecular Systems Biology* 13 (2): 914. doi:10.15252/msb.20167159.

Hakimi, A. Ari, Ed Reznik, Chung-Han Lee, Chad J. Creighton, A. Rose Brannon, Augustin Luna, B. Arman Aksoy, et al. 2016. "An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma." *Cancer Cell* 29 (1). Elsevier: 104–16. doi:10.1016/j.ccell.2015.12.004.

Hanahan, Douglas, and Robert A Weinberg. 2011. "Review Hallmarks of Cancer : The Next Generation." *Cell*

144 (5): 646–74. doi:10.1016/j.cell.2011.02.013.

Hansford, Samantha, Pardeep Kaurah, Hector Li-Chang, Michelle Woo, Janine Senz, Hugo Pinheiro, Kasmintan A. Schrader, et al. 2015. “Hereditary Diffuse Gastric Cancer Syndrome.” *JAMA Oncology* 1 (1). American Medical Association: 23. doi:10.1001/jamaoncol.2014.168.

Haq, Rizwan, Jonathan Shoag, Pedro Andreu-Perez, Satoru Yokoyama, Hannah Edelman, Glenn C Rowe, Dennie T Frederick, et al. 2013. “Oncogenic BRAF regulates oxidative metabolism via PGC1 $\alpha$  and MITF.” *Cancer Cell* 23 (3). Elsevier: 302–15. doi:10.1016/j.ccr.2013.02.003.

Helbig, Gregory, Kent W Christopherson, Poornima Bhat-Nakshatri, Suresh Kumar, Hiromitsu Kishimoto, Kathy D Miller, Hal E Broxmeyer, and Harikrishna Nakshatri. 2003. “NF-kappaB promotes breast cancer cell migration and metastasis by inducing the expression of the chemokine receptor CXCR4.” *The Journal of Biological Chemistry* 278 (24). American Society for Biochemistry; Molecular Biology: 21631–8. doi:10.1074/jbc.M300609200.

Hennig, Christian. 2015. *Fpc: Flexible Procedures for Clustering*. <http://CRAN.R-project.org/package=fpc>.

Heyland, Daren, John Muscedere, Paul E. Wischmeyer, Deborah Cook, Gwynne Jones, Martin Albert, Gunnar Elke, Mette M. Berger, and Andrew G. Day. 2013. “A Randomized Trial of Glutamine and Antioxidants in Critically Ill Patients.” *New England Journal of Medicine* 368 (16). Massachusetts Medical Society: 1489–97. doi:10.1056/NEJMoa1212722.

Hsu, Peggy P, and David M Sabatini. 2008. “Cancer cell metabolism: Warburg and beyond.” *Cell* 134 (5). Elsevier: 703–7. doi:10.1016/j.cell.2008.08.021.

Huber, Margit A, Ninel Azoitei, Bernd Baumann, Stefan Grünert, Andreas Sommer, Hubert Pehamberger, Norbert Kraut, Hartmut Beug, and Thomas Wirth. 2004. “NF-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression.” *The Journal of Clinical Investigation* 114 (4). American Society for Clinical Investigation: 569–81. doi:10.1172/JCI21358.

Jr, Frank E Harrell, with contributions from Charles Dupont, and many others. 2017. *Hmisc: Harrell*

*Miscellaneous.* <http://CRAN.R-project.org/package=Hmisc>.

Kaufman, L, and P J Rousseeuw. 1987. “Clustering by means of medoids.”

Koenker, Roger, and Pin Ng. 2017. *SparseM: Sparse Linear Algebra*. <http://CRAN.R-project.org/package=SparseM>.

Krishnan, Keerthana, Anita L Steptoe, Hilary C Martin, Diwakar R Pattabiraman, Katia Nones, Nic Waddell, Mythily Mariasegaram, et al. 2013. “miR-139-5p is a regulator of metastatic pathways in breast cancer.” *RNA (New York, N.Y.)* 19 (12). Cold Spring Harbor Laboratory Press: 1767–80. doi:10.1261/rna.042143.113.

Kruskal, William H., and W. Allen Wallis. 1952. “Use of Ranks in One-Criterion Variance Analysis.” *Journal of the American Statistical Association* 47 (260): 583–621. doi:10.1080/01621459.1952.10483441.

Kvamme, E, and G Svenneby. 1960. “Effect of anaerobiosis and addition of keto acids on glutamine utilization by Ehrlich ascites-tumor cells.” *Biochimica et Biophysica Acta* 42 (July): 187–8. <http://www.ncbi.nlm.nih.gov/pubmed/13755527>.

Latreille, Mathieu, Karolin Herrmanns, Neil Renwick, Thomas Tuschl, Maciej T. Malecki, Mark I. McCarthy, Katharine R. Owen, Thomas Rüllicke, and Markus Stoffel. 2015. “miR-375 gene dosage in pancreatic  $\beta$ -cells: implications for regulation of  $\beta$ -cell mass and biomarker development.” *Journal of Molecular Medicine* 93 (10). Springer Berlin Heidelberg: 1159–69. doi:10.1007/s00109-015-1296-9.

Lee, Yoon Suk, Haeryoung Kim, Hyoung Woo Kim, Jong-Chan Lee, Kyu-Hyun Paik, Jingu Kang, Jaihwan Kim, et al. n.d. “High Expression of MicroRNA-196a Indicates Poor Prognosis in Resected Pancreatic Neuroendocrine Tumor.” doi:10.1097/MD.0000000000002224.

Liberzon, A., A. Subramanian, R. Pinchback, H. Thorvaldsdottir, P. Tamayo, and J. P. Mesirov. 2011. “Molecular signatures database (MSigDB) 3.0.” *Bioinformatics* 27 (12). Oxford University Press: 1739–40. doi:10.1093/bioinformatics/btr260.

Lin, C H, S Y Hsieh, I S Sheen, W C Lee, T C Chen, W C Shyu, and Y F Liaw. 2001. “Genome-wide hypomethylation in hepatocellular carcinogenesis.” *Cancer Research* 61 (10): 4238–43. <http://www.ncbi.nlm.nih>.

gov/pubmed/11358850.

Liu, Yankun, Yingnan Zhang, Haidong Wu, Yufeng Li, Yi Zhang, Min Liu, Xin Li, and Hua Tang. 2017. “miR-10a suppresses colorectal cancer metastasis by modulating the epithelial-to-mesenchymal transition and anoikis.” *Cell Death & Disease* 8 (4). Nature Publishing Group: e2739–e2739. doi:10.1038/cddis.2017.61.

Lucas, Antoine. 2014. *Amap: Another Multidimensional Analysis Package*. <http://CRAN.R-project.org/package=amap>.

Ma, Yu-Shui, Ting-Miao Wu, Zhong-Wei Lv, Gai-Xia Lu, Xian-Ling Cong, Ru-Ting Xie, Hui-Qiong Yang, et al. 2017. “High expression of miR-105-1 positively correlates with clinical prognosis of hepatocellular carcinoma by targeting oncogene NCOA1.” *Oncotarget* 8 (7). Impact Journals: 11896–11905. doi:10.18632/oncotarget.14435.

Mann, H. B., and D. R. Whitney. 1947. “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other.” *The Annals of Mathematical Statistics* 18 (1): 50–60. doi:10.1214/aoms/1177730491.

Matés, José M, Cristina Pérez-Gómez, Ignacio Núñez de Castro, Maite Asenjo, and Javier Márquez. 2002. “Glutamine and its relationship with intracellular redox status, oxidative stress and cell proliferation/death.” *The International Journal of Biochemistry & Cell Biology* 34 (5). Pergamon: 439–58. doi:10.1016/S1357-2725(01)00143-1.

Moarii, Matahi, Valentina Boeva, Jean-Philippe Vert, and Fabien Reyat. 2015. “Changes in correlation between promoter methylation and gene expression in cancer.” *BMC Genomics* 16 (1). BioMed Central: 873. doi:10.1186/s12864-015-1994-2.

Mootha, Vamsi K, Cecilia M Lindgren, Karl-Fredrik Eriksson, Aravind Subramanian, Smita Sihag, Joseph Lehar, Pere Puigserver, et al. 2003. “PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.” *Nature Genetics* 34 (3). Nature Publishing Group: 267–73. doi:10.1038/ng1180.

Olivier, Magali, Monica Hollstein, and Pierre Hainaut. 2010. “TP53 mutations in human cancers: origins, consequences, and clinical use.” *Cold Spring Harbor Perspectives in Biology* 2 (1). Cold Spring Harbor Laboratory Press: a001008. doi:10.1101/cshperspect.a001008.

Pearson, Karl. 1900. “X. <i>On the criterion that a given system of deviations from the probable in the case of a

- correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling</i>.” *Philosophical Magazine Series 5* 50 (302). Taylor & Francis Group: 157–75. doi:10.1080/14786440009463897.
- Phillips, T. 2008. “The role of methylation in gene expression.” *Nature Education* 1: 1.
- R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Rajasekhar, Megha, Anton M. Olsson, Kathryn J.A. Steel, Mirella Georgouli, Ushan Ranasinghe, Christine Brender Read, Klaus S. Frederiksen, and Leonie S. Taams. 2017. “MicroRNA-155 contributes to enhanced resistance to apoptosis in monocytes from patients with rheumatoid arthritis.” *Journal of Autoimmunity* 79 (May). Academic Press: 53–62. doi:10.1016/J.JAUT.2017.01.002.
- Razin, A, and H Cedar. 1991. “DNA methylation and gene expression.” *Microbiological Reviews* 55 (3): 451–8. doi:10.1002/wsbm.64.
- Ritchie, Matthew E, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. 2015. “limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.” *Nucleic Acids Research* 43 (7): e47.
- RStudio Team. 2015. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>.
- Sarkar, Deepayan. 2008. *Lattice: Multivariate Data Visualization with R*. New York: Springer. <http://lmdvr.r-forge.r-project.org>.
- Seshan, Venkatraman E., and Adam Olshen. n.d. *DNACopy: DNA Copy Number Data Analysis*.
- Shen, L, J Fang, D Qiu, T Zhang, J Yang, S Chen, and S Xiao. 1998. “Correlation between DNA methylation and pathological changes in human hepatocellular carcinoma.” *Hepato-Gastroenterology* 45 (23): 1753–9. <http://www.ncbi.nlm.nih.gov/pubmed/9840141>.
- Soares, J, A E Pinto, C V Cunha, S André, I Barão, J M Sousa, and M Cravo. 1999. “Global DNA hypomethylation in breast carcinoma: correlation with prognostic factors and tumor progression.” *Cancer* 85 (1):

112–8. <http://www.ncbi.nlm.nih.gov/pubmed/9921982>.

Stambolic, V, and J R Woodgett. 1994. “Mitogen inactivation of glycogen synthase kinase-3 beta in intact cells via serine 9 phosphorylation.” *The Biochemical Journal* 303 ( Pt 3 (3)). Portland Press Limited: 701–4. doi:10.1042/BJ3030701.

Subramanian, Aravind, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, et al. 2005. “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (43). National Academy of Sciences: 15545–50. doi:10.1073/pnas.0506580102.

Tang, Xiaohu, Chao-Chieh Lin, Ivan Spasojevic, Edwin S Iversen, Jen-Tsan Chi, and Jeffrey R Marks. 2014. “A joint analysis of metabolomics and genetics of breast cancer.” *Breast Cancer Research* 16 (4). BioMed Central: 415. doi:10.1186/s13058-014-0415-9.

Terry M. Therneau, and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Therneau, Terry M. 2015. *A Package for Survival Analysis in S*. <https://CRAN.R-project.org/package=survival>.

Thiele, Ines, Neil Swainston, Ronan M.T. Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K. Aurich, Hulda Haraldsdottir, et al. 2013. “A community-driven global reconstruction of human metabolism.” *Nature Biotechnology* 31 (5): 419–25. doi:10.1038/nbt.2488.

Turcan, Sevin, Daniel Rohle, Anuj Goenka, Logan A Walsh, Fang Fang, Emrullah Yilmaz, Carl Campos, et al. 2012. “IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype.” *Nature* 483 (7390). NIH Public Access: 479–83. doi:10.1038/nature10866.

Vander Heiden, Matthew G., and Ralph J. DeBerardinis. 2017. “Understanding the Intersections between Metabolism and Cancer Biology.” *Cell* 168 (4). Elsevier: 657–69. doi:10.1016/j.cell.2016.12.039.

Vander Heiden, Matthew G., Lewis C. Cantley, and Craig B. Thompson. 2009. “No Title” 324 (5930). American Association for the Advancement of Science. doi:10.1126/science.1160809.

Vienberg, S., J. Geiger, S. Madsen, and L. T. Dalgaard. 2017. “MicroRNAs in metabolism.” *Acta Physiologica*

219 (2): 346–61. doi:10.1111/apha.12681.

Wagner, James R, Stephan Busche, Bing Ge, Tony Kwan, Tomi Pastinen, and Mathieu Blanchette. 2014. “The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts.” *Genome Biology* 15 (2). BioMed Central: R37. doi:10.1186/gb-2014-15-2-r37.

Warburg, O. 1956. “Injuring of Respiration the Origin of Cancer Cells.” *Science* 123 (3191): 309–14. doi:10.1126/science.123.3191.309.

Warburg, Otto, Erwin Negelein, and Karl Posener. 1924. “Versuche an Überlebendem Carcinomgewebe.” *Klinische Wochenschrift* 3 (24): 1062–4. doi:10.1007/BF01736087.

Zhang, J., and B. Feng. n.d. *CghMCR: Find Chromosome Regions Showing Common Gains/Losses*.

Zhang, Ji, Natalya N Pavlova, and Craig B Thompson. 2017. “Cancer cell metabolism: the essential role of the nonessential amino acid, glutamine.” *The EMBO Journal* 36 (10): 1302–15. doi:10.15252/emj.201696151.

Zhang, Jianhua. n.d. *CNTools: Convert Segment Data into a Region by Sample Matrix to Allow for Other High Level Computational Analyses*.

Zhang, Kejin, and Ling Guo. 2018. “MiR-767 promoted cell proliferation in human melanoma by suppressing CYLD expression.” *Gene* 641 (January). Elsevier: 272–78. doi:10.1016/J.GENE.2017.10.055.

Zhu, Fu-qiang, Li Zeng, Na Tang, Ya-ping Tang, Bo-ping Zhou, Fang-fang Li, Wei-gang Wu, Xiao-bing Zeng, and Shu-song Peng. 2016. “MicroRNA-155 Downregulation Promotes Cell Cycle Arrest and Apoptosis in Diffuse Large B-Cell Lymphoma.” *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics* 24 (6): 415–27. doi:10.3727/096504016X14685034103473.