Glucose metabolism transcriptome clustering identifies subsets of resectable lung adenocarcinoma with different prognosis.

Enzo Alifano, Mathilde Prieto, MD, Marco Alifano, MD PhD

PII: S2666-2736(24)00170-0
DOI: https://doi.org/10.1016/j.xjon.2024.06.010
Reference: XJON 1094

To appear in: JTCVS Open

Received Date: 18 April 2024
Revised Date: 6 June 2024
Accepted Date: 17 June 2024

Please cite this article as: Alifano E, Prieto M, Alifano M, Glucose metabolism transcriptome clustering identifies subsets of resectable lung adenocarcinoma with different prognosis., JTCVS Open (2024), doi: https://doi.org/10.1016/j.xjon.2024.06.010.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Copyright © 2024 The Authors. Published by Elsevier Inc. on behalf of The American Association for Thoracic Surgery
Glucose metabolism transcriptome clustering identifies subsets of resectable lung adenocarcinoma with different prognosis.

Enzo Alifano¹, Mathilde Prieto MD ¹, Marco Alifano MD PhD ¹

¹: Thoracic surgery department, Cochin hospital, APHP.CUP, Paris-University, France.

Short Title: Glucose metabolism gene clustering.

Disclosure statement: The authors have no conflict of interest in relation with the content of the manuscript.

Author contributions: Conceptualization: EA, MP, MA; Data curation: MP; Formal analysis: EA, MA; Investigation: EA, MP, MA; Methodology: EA, MA; Writing - original draft: MA, EA; Supervision: MA. Writing assistance: None

Funding: None

Corresponding author:

Pr Marco Alifano
Thoracic Surgery Department – Cochin Hospital
APHP Centre Université de Paris C
27 Rue de Faubourg Saint Jacques – 75014 Paris – France
e-mail: marco.alifano@aphp.fr

Article word count: 4274 words

The Institutional Review Board (IRB) did not approve this study, because of its nature (retrospective analysis of a publicly available database. Patient written consent for the publication of the study was not received, as they were anonymous.
Glossary of abbreviations

TCGA= The Cancer Genome Atlas; PCA= Principal Component Analysis; ALDOA= Aldolase, Fructose-Bisphosphate A, CS= Citrate Synthase, ENO1= Enolase 1, FH= Fumarate Hydratase, GAPDH= Glyceraldehyde-3-Phosphate Dehydrogenase, GPI= Glucose-6-Phosphate Isomerase, HK2= Hexokinase 2, IDH3A= Isocitrate Dehydrogenase (NAD(+)) 3 Catalytic Subunit Alpha, LDHA= Lactate Dehydrogenase A, MDH2=, Malate Dehydrogenase 2 MT-ATP6= Mitochondrially Encoded ATP Synthase Membrane Subunit 6, MT-ATP8= Mitochondrially Encoded ATP Synthase Membrane Subunit 8, MT-CO1= Mitochondrially Encoded Cytochrome C Oxidase I, MT-CO2= Mitochondrially Encoded Cytochrome C Oxidase II, MT-CO3=, Mitochondrially Encoded Cytochrome C Oxidase III MT-CYB= Mitochondrially Encoded Cytochrome B, MT-1= Metallothionein 1, MT-2= Metallothionein 2, MT-3= Metallothionein 3, OGDH= Oxoglutarate Dehydrogenase, PFKP= Phosphofructokinase, Platelet, PGAM1= Phosphoglycerate Mutase 1, PGK1= Phosphoglycerate Kinase 1, PKM= Pyruvate Kinase M, SDHA= Succinate Dehydrogenase Complex Flavoprotein Subunit A, SLC16A1= Solute Carrier Family 16 Member 1, SLC16A3= Solute Carrier Family 16 Member 3, SLC2A1= Solute Carrier Family 2 Member 1, SLC2A2= Solute Carrier Family 2 Member 2, SLC2A3= Solute Carrier Family 2 Member 3, SLC2A4= Solute Carrier Family 2 Member 4, SUCLG1= Succinate-CoA Ligase GDP/ADP-Forming Subunit Alpha, TOMM20= Translocase Of Outer Mitochondrial Membrane 20, TPI1= Triosephosphate Isomerase 1.
Central Picture

Summary of the ten steps of glycolysis and interaction with TCA cycle and OXPHOS
Central message

Using a clustering gene transcriptoma approach, we showed the existence of different energy metabolic phenotypes in lung adenocarcinoma and demonstrated that different clusters had different prognosis.
Perspective statements

Most of the energy production of cancer cells would originate from aerobic glycolysis. Concurrent tricarboxylic acid cycle and oxidative phosphorylation is a possible feature, but the clinical significance of co-existence of both pathways is unknown. By a clustering approach, we showed the existence of different metabolic phenotypes and demonstrated that different clusters had different prognosis.
Abstract

Objectives: Reprogramming of energy metabolism is a well-established hallmark of cancer, with aerobic glycolysis classically considered a prominent feature. We investigate the heterogeneity in glucose metabolism pathways within resectable primary lung adenocarcinoma and its clinical significance.

Methods: Using The Cancer Genome Atlas (TCGA) data, RNA expressions were extracted from 489 primary lung adenocarcinoma samples. Prognostic impact of glycolytic (MCT4), aerobic (MCT1), and mitochondrial (TOMM20) markers was assessed using Kaplan-Meier analysis. Clustering of 35 genes involved in glucose metabolism was performed using the k-means method. The clusters were then analyzed for associations with demographic, clinical, and pathologic variables. Overall survival was assessed using the Kaplan-Meier estimator. Multivariate analysis was performed to assess the independent prognostic value of cluster membership.

Results: Classical statistical approach showed that higher expression of MCT4 was associated with a significantly worse prognosis. Increased expression of TOMM20 was associated with a non-significant trend toward better prognosis, and increased expression of MCT1 was associated with a better outcome. Clustering identified three major metabolic phenotypes, dominantly hypometabolic, dominantly oxidative and dominantly mixed oxidative/glycolytic with significantly different pathologic stage distribution and prognosis; mixed oxidative/glycolytic was associated with worse survival. Cluster membership was independently associated with survival.

Conclusion: This study demonstrates the existence of distinct glucose metabolism clusters in resectable lung adenocarcinoma, providing valuable prognostic information. The findings highlight the potential relevance of considering metabolic profiles when designing strategies.
for reprogramming energy metabolism. Further studies are warranted to validate these findings in different cancer types and populations.

Key words: Glucose metabolism; gene clustering; lung cancer; survival.
Introduction

Reprogramming of energy metabolism has been recognized as a hallmark of cancer cells: increased energy demands are required to fuel cell growth and division\(^1\). In their landmark 2011 paper, Hanahan and Weinberg highlighted how aerobic glycolysis (the so-called Warburg effect) is largely favored over normal glucose metabolism in cancer cells: even under aerobic conditions, most of the energy production would come from cytoplasmic glycolysis, without the transfer of pyruvate to mitochondria for the tricarboxylic acid cycle and oxidative phosphorylation\(^2\).

This mechanism, originally described by Otto Warbourg in 1930 and confirmed in various settings in the following years, is rather counterintuitive because the energetic efficiency of glycolysis (in terms of ATP production) is about 18 times lower compared to the usual glucose metabolism of non-cancerous cells\(^3\). To compensate for the lower energetic efficiency and to enable enhanced glycolysis, cancer cells require more glucose to enter the cytosol, which is enabled by increased expression of glucose transporters, mainly GLUT1\(^4\) \(^5\). In this model, cancer can be viewed as a metabolic parasite whose cells compete for energy substrates with host cells in the rest of the body and in the tumor microenvironment \(^6\). The ability of tumors to take up glucose is clinically demonstrated using 18FDG positron emission tomography, and a large body of literature has accumulated highlighting the negative prognostic character of increased glucose uptake\(^7\).
It should be emphasized that aerobic conditions are not necessarily present inside tumors whose neovascularization favors reduced oxygen availability, especially in the depth of the lesion; in this case, only glycolytic fueling allows cancer cells to survive and proliferate. In addition, glycolytic function has been shown to be associated with activated oncogenes such as RAS or MYC and mutated tumour suppressors such as TP53.

Furthermore, glycolytic waste in the tumor stroma contributes to a decrease in local pH, which in turn leads to a decrease in the efficacy of anti-tumor immune cells in the tumor microenvironment.

Glycolysis is a sequence of ten enzymatic reactions and is thought to have three key regulatory steps, namely the reactions catalyzed by hexokinase, phosphofructokinase and pyruvate kinase (first, third and tenth steps, respectively). These regulatory steps are essentially irreversible and have large negative $\Delta G$ values. A large number of publications have been devoted to the impact of these enzymes on cancer progression and outcome, but most of the available literature focuses on a single enzyme, its expression and relationships with different oncogenic pathways. However, it has been highlighted how several glycolytic intermediates, which are also generated in non-key steps and whose levels would be increased in the event of enhanced glycolysis and enhanced expression of non-key enzymes, can be diverted to enter various biosynthetic pathways, including those that generate nucleosides and amino acids, which in turn facilitate the biosynthesis of macromolecules required for cell division. Thus, up-regulation of the whole glycolytic pathway would be a real advantage for cancer cells,
mitigating the lower energy efficiency compared to normal glucose metabolism, but clinical
evidence of increased aggressiveness in the case of current over-activation of the whole
glycolytic pathway is currently poor.

Cancer cells do not necessarily rely on glycolysis, and Hanahan and Weinberg highlighted that
some tumors have been found to contain two subpopulations of cancer cells that differ in their
energy production pathways: one subpopulation would consist of cells that rely on the Warburg
effect and produce lactate, whereas the second subpopulation would preferentially import and
use the lactate produced by their neighbors as their main energy source, converting lactate to
pyruvate for use in the tricarboxylic acid cycle and oxidative phosphorylation\(^2\). In this model,
the role of tricarboxylic acid carriers is crucial, allowing lactic acid to be secreted or imported:
the carriers responsible for this exchange, MCT4 and MCT1, have been proposed to be more
frequently associated with glycolytic or OXPHOS function in cancer cells, respectively\(^{15 \text{ 16 17}}\).

However, the clinical significance of the coexistence of both pathways, has been poorly
assessed.

Classifying tumor metabolism as oxidative, glycolytic, or both (if possible) on the basis of
analysis of single or groups of proteins or gene expressions is not trivial. Indeed, it is not clear
how one should classify the metabolism when – for example – some “glycolytic” genes show
high RNA expression and others a low RNA expression. To our knowledge, there is no
universally accepted heuristic for doing so. In this analysis, we propose to learn the
classification heuristic on a publicly available dataset using a clustering approach.
Thus, using the RNA expressions available thanks to The Cancer Genome Atlas (TCGA) project and performing a clustering approach, we aimed to: 1. assess the coexistence of different glucose metabolism pathways in a clinical model of resectable primary lung adenocarcinoma; 2. evaluate whether different clusters are differentially represented among stages, sex and age categories; 3. assess the prognostic significance of different glucose metabolism pathway clusters.

Material and Methods:

The R2 platform (‘R2: Genomics Analysis and Visualization Platform (http://r2.amc.nl)’) was used to extract demographic, pathologic, RNA expression and overall survival data from the TCGA Lung Adenocarcinoma database (Mixed Lung Adenocarcinoma (2022-v32) - tega - 589 - tpm - gencode36). We took into account data on patients whose samples were represented by the primary tumor and for whom a minimal follow-up beyond intervention was available (n=489). The Institutional Review Board (IRB) did not approve this study, because of its nature (retrospective analysis of a publicly available database. Patient written consent for the publication of the study was not received, as they were anonymous.

First, we assessed the prognostic impact of RNA expression of a marker of glycolytic activity (MCT4), a marker of aerobic activity (MCT1) and a marker of mitochondrial function (TOMM20) by Kaplan-Meier estimates of overall survival. We used the median, lower quartile and upper quartile as cut-offs. Log-rank was used for comparison.

Secondly, we used the platform’s “grabber” function to extract demographic and available clinical and pathological data. Together, we extracted RNA expression levels (normalized by log2) of 35 genes, a number considered compatible with the subsequent clustering as
representing approximately 50% more than the square number of observations. Thus, we extracted the RNA expression of genes of the three main glucose transporters (SLC2A1, SLC2A3, SLC2A4, whose proteins are known as GLUT1, GLUT3 and GLUT4), the ten genes of glycolysis (HK2, GPI, PFKP, ALDOA, TPI1, GAPDH, PGK1, PGAM1, ENO1, PKM), the eight genes of tricarboxylic cycle (CS, ACO2, IDH3A, OGDH, SUCLG1, SDHA, FH, MDH2), and all the ten mitochondrial genes involved in oxidative phosphorylation (MT-ATP6, MT-ATP8, MT-CO1, MT-CO2, MT-CO3, MT-CYB, MT-ND1, MT-ND2, MT-ND3, MT-ND4) as well as of non-mitochondrial gene of oxidative phosphorylation, TOMM20 belonging to the TOMM complex (translocase of the outer mitochondrial membrane). Finally, to take into account the possible utilization of lactate produced by glycolytic cancer cells by oxidative ones, we extracted the RNA expression of two tricarboxylic acid carrier genes, SLC16A3 and SLC16A1 (whose proteins are known as MCT1 and MCT4), and of LDH, which reversibly catalyze the pyruvate-lactate conversion.

For some of the enzymes different isoforms may exist; we selected, after careful literature review, the isoform more involved in glucose metabolism abnormalities in lung adenocarcinoma, whenever known, or, more generally, in cancer. For example, for hexokinase, we selected the isoform 2 (HK2).

Data were retrieved in the txt format and converted to Excel. The Excel file was imported into Python and Python scripts were written to perform data analysis and clustering.

Clustering was performed using the k-means approach. Kaplan-Meier estimates were used to assess overall survival and curves were compared by log-rank. Finally, Cox regression analysis was used to assess the independent character of prognostic factors. In the Cox model we included cluster, age, sex, and pathologic stage. Significance was accepted at p<0.05. Principal Component Analysis (PCA) analyses were performed to discuss the results qualitatively and
display the relevant clusters. Used dataset and all the Python scripts are available upon request.

IRB was not required due to the nature of the data analyzed.

**Results:**

Records of 489 patients who met our criteria were collected and available for analysis. All specimens were obtained at surgical resection for primary lung adenocarcinoma. The distribution of pathologic stages was I: 269, II: 116, III: 78, IV: 25.

*Prognostic impact of markers of glycolytic or oxidative metabolism and of mitochondrial activity.*

Using median or 75th percentile but not the 25th percentile as cut-off, higher expression of SLC16A3 (MCT4, marker of glycolytic metabolism) was associated with a significantly worse prognosis. Increased expression of SLC16A1 (MCT1, a marker of non-glycolytic metabolism) was associated with a better outcome when using the 75th percentile as cut-off, whereas the difference was not significant using the median or 25th percentile. Finally, increased expression of TOMM20 (a marker of oxidative phosphorylation), is associated with a trend toward better prognosis without reaching significance, regardless of the cut-off chosen (Figure 1).

**Clustering approach**

We performed a PCA on the gene expression variables to assist visualization of clusters. The cumulative explained variances for each principal component are shown in figure 2A. This result supports a two-dimensional analysis of the clusters, as the first two components explain more than half of the variance in the dataset. Subsequent components explain far less variance. These two components correlate with each gene as shown in figure 2B.
We identified four clusters. On these two axes, the clusters can be visualized with a color mapping in figure 2C. Table 1 shows the mean values and standard deviation of each analyzed gene expression as well as the means or proportions of available clinical variables in each cluster.

In a univariate Kaplan Meier analysis, the assigned cluster is associated with overall survival (Cluster 1 vs Cluster 2, p < 0.05; Cluster 0 vs Cluster 2, p = 0.01; Cluster 1 vs Cluster 0, p = 0.24). Cluster 3 is excluded from this analysis due to the small number of datapoints, and therefore large confidence intervals. This can be seen in figure 2D.

In a univariate Cox regression model (table 2, A), clusters 0 and 1 are significantly associated with better prognosis as compared to cluster 2, in the direction that is predicted by the Kaplan-Meier estimates.

When age, sex and stage were included in the regression (table 2, B), Cluster 1 maintained significance, whereas cluster 0 lost significance, but a trend in the same direction was maintained.

**Discussion**

In this study, we provide evidence that, in lung adenocarcinoma, genes of the entire glucose metabolism pathway (including glucose transporters, tricarboxylic acid transporters, glycolysis, tricarboxylic acid cycle and oxidative phosphorylation) can be clustered, resulting in a strong prognostic discrimination. The prognostic significance of clusters is maintained in multivariate analysis underlining the independent prognostic significance.
This clustering approach has the advantage of not requiring labeled data; as discussed above, the labeling itself (i.e. defining a tumor as merely glycolytic or not) may be hardly achievable. The expression of SLC16A3 (included in our analysis) has been reported as a general marker of glycolysis, because its product, MCT4 is a tricarboxylic acid carrier responsible for lactate export from glycolytic producing cells. By a conventional statistical approach, using median as cut-off, higher expression of SLC16A3, was associated with a significantly worse prognosis, as compared to tumors with lower expression of MCT4 which could be considered as non-glycolytic, although they cannot be directly qualified as relying on oxidative metabolism. We also show by a conventional approach that increased expression of TOMM20, classically considered as a marker of oxidative phosphorylation, is associated with trend toward better prognosis without reaching significance (regardless from the cut-off chosen). Yet, increased expression of SLC16A1 (considered as marker of non-glycolytic metabolism) was associated with a better outcome when the 75th percentile was used as a cut-off, whereas the difference was not significant at median. The product of SLC16A, MCT1, is a tricarboxylic acid carrier that uptake lactate from the interstitium (where it is rejected by glycolytic cells) into the non-glycolytic cells where it can be transformed into pyruvate by LDHA. The fate of pyruvate can be dual as it can proceed towards the tricarboxylic acid cycle and oxidative phosphorylation or even enter in the gluconeogenesis pathway to produce glucose, as in the Cori cycle. Using a conventional approach on TCGA data, one could conclude that glycolytic tumors have a worse prognosis as compared to non-glycolytic tumors and that tumors...
expressing a specific marker of oxidative phosphorylation show a non-significant trend toward better prognosis.

However, it should be underlined that most of the variance in gene expressions can be explained by a two component PCA. In figure 2, the left-right axis is correlated with mitochondrial genes and therefore explains the aerobic character of the phenotype; the up-down axis is correlated to anaerobic genes, but also with gene expressions related to lactate re-uptake and lactate to pyruvate transformations. The fact that these two variables explain most of the variance in the dataset is consistent with basic research: these lung adenocarcinoma phenotypes are explained by a theoretical “oxidative/mitochondrial variable”, and a theoretical “anaerobic variable”. This second variable correlates with lactate and pyruvate metabolism, as predicted by the two components model.

Using our approach, we were able to identify three main clusters, including 171, 165, and 130 patients, respectively, and a smaller one with 23 individuals. The latter was mainly composed of younger individuals, with lower average tobacco consumption, and more frequent early stage disease: expression of both oxidative and glycolytic genes was very low and few events were observed in the first years after surgery. The other three clusters were represented by a cluster (No. 1) expressing low levels of both glycolytic and OXPHOS genes, a cluster mainly relying on OXPHOS metabolism (No. 0), and a cluster (No. 2) with increased levels of expression of glycolytic genes, but also expressing genes responsible for lactate re-uptake and transformation to pyruvate at a higher level as compared to the clusters 0 and 1, and expressing...
also mitochondrial genes at a level on average superior to the hypometabolic cluster, but similar
to the cluster relying mainly on OXPHOS. The outcome was different among the three clusters,
with the best survival in the cluster (No. 1) with low expression of both oxidative and glycolytic
genes, followed by the cluster with dominant OXPHOS metabolism (No. 0), and finally by the
cluster (No.2) with glycolytic metabolism associated with an OXPHOS one. Notably, this
prognostic significance applies to the entire population, but similar trends are observed within
the different stages of the disease. In other words, our findings support the concept that inside
a tumor harboring a glycolytic component the presence of a non-glycolytic one conjure to
increase aggressiveness which is reflected by worsening prognosis.

Thus, the results of our approach seem to be consistent with some knowledge from basic science
and show that clustering could allow the identification of metabolic phenotypes associated with
different prognosis in clinical setting\textsuperscript{20, 21, 22}. This could be potentially relevant on clinical
grounds, and could support the concept that reprogramming tumor metabolism should be a
major research goal in the coming years\textsuperscript{2, 23, 24}. In the era of personalized medicine, allocating
patients to a specific group of metabolic profiles would be crucial in the design of such studies
on energy metabolism reprogramming strategies\textsuperscript{25}.

Previous studies based on TCGA data have shown the enormous potential of this project to
reveal multiple metabolic pathway dysregulation in different cancers\textsuperscript{26}. In this study we
focused on a specific cancer type and demonstrated the existence of different clusters with a
strong impact on survival: we believe that the main strength of our study is the relative
simplicity of the concepts. There are two main limitations of our study. The first is due to the
nature of the samples: the RNA was extracted from bulk tumors and therefore contained
variable amounts of host RNA along with the tumor RNA, which is likely to represent the
dominant part of the RNA: despite an unavoidable limitation, the bulk approach has the
advantage of taking into account tumor heterogeneity, which cannot be easily taken into account
when using other approaches, such as single cells RNA analysis. The other limitation is that the
results of our method should be considered as dataset specific. Unfortunately, to date, publicly
available transcriptome datasets (for example LUSC-CN or LUSC-KR \(^{27,28}\)) contain a limited
number of patients which precludes a clustering approach.

Overall, patterns and correlations derived from the TCGA dataset we used are specific to the
cancer type studied in this paper, primary lung adenocarcinoma. The behavior (including impact
on survival) in other histologic types of lung cancer and, more generally, in other cancers and
relative comparisons deserve further studies.

References

1. Icard P, Fournel L, Wu Z, Alifano M, Lincet H. Interconnection between Metabolism


12. Flier JS, Mueckler MM, Usher P, Lodish HF. Elevated levels of glucose transport and transporter messenger RNA are induced by ras or src oncogenes. Science. 1987;235(4795):1492-1495. doi:10.1126/science.3103217


<table>
<thead>
<tr>
<th></th>
<th>Cluster 0</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (years)</td>
<td>64.69±10.31</td>
<td>67.38±8.94</td>
<td>64.7±10.1</td>
<td>57.96±10.96</td>
</tr>
<tr>
<td>Percentage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>56%</td>
<td>58%</td>
<td>50%</td>
<td>43%</td>
</tr>
<tr>
<td>Stage I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>57%</td>
<td>65%</td>
<td>42%</td>
<td>74%</td>
</tr>
<tr>
<td>Stage II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>22%</td>
<td>29%</td>
<td>17%</td>
</tr>
<tr>
<td>Stage III</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18%</td>
<td>8%</td>
<td>21%</td>
<td>9%</td>
</tr>
<tr>
<td>Stage IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4%</td>
<td>4%</td>
<td>8%</td>
<td>0%</td>
</tr>
<tr>
<td>Pack/years</td>
<td>42,52±28.34</td>
<td>38,8±25.59</td>
<td>45,35±29.11</td>
<td>30,67±17.17</td>
</tr>
<tr>
<td>ACO2</td>
<td>4.02±0.56</td>
<td>4.29±0.68</td>
<td>4.47±0.53</td>
<td>3.4±0.52</td>
</tr>
<tr>
<td>ALDOA</td>
<td>2.68±0.75</td>
<td>2.58±0.59</td>
<td>3.49±0.72</td>
<td>2.46±0.99</td>
</tr>
<tr>
<td>CS</td>
<td>6.09±0.5</td>
<td>6.12±0.52</td>
<td>6.56±0.52</td>
<td>5.18±0.66</td>
</tr>
<tr>
<td>ENO1</td>
<td>9.87±0.64</td>
<td>9.74±0.73</td>
<td>10.64±0.6</td>
<td>8.82±1.24</td>
</tr>
<tr>
<td>FH</td>
<td>6.42±0.51</td>
<td>6.54±0.59</td>
<td>6.74±0.5</td>
<td>5.78±0.57</td>
</tr>
<tr>
<td>GAPDH</td>
<td>10.18±0.66</td>
<td>9.95±0.72</td>
<td>11.32±0.6</td>
<td>9.25±1.29</td>
</tr>
<tr>
<td>GPI</td>
<td>6.26±0.58</td>
<td>6.04±0.54</td>
<td>6.96±0.6</td>
<td>5.23±0.7</td>
</tr>
<tr>
<td>HK2</td>
<td>4.88±1.03</td>
<td>4.56±1.08</td>
<td>5.95±0.99</td>
<td>4.96±1.36</td>
</tr>
<tr>
<td>IDH3A</td>
<td>2.72±0.59</td>
<td>3.04±0.46</td>
<td>3.26±0.52</td>
<td>3.05±0.5</td>
</tr>
<tr>
<td>LDHA</td>
<td>8.31±0.62</td>
<td>8.34±0.57</td>
<td>9.31±0.65</td>
<td>8.27±0.68</td>
</tr>
<tr>
<td>MDH2</td>
<td>7.32±0.57</td>
<td>7.31±0.63</td>
<td>7.76±0.51</td>
<td>6.33±0.7</td>
</tr>
<tr>
<td>MT-ATP6</td>
<td>13.89±0.48</td>
<td>12.69±0.56</td>
<td>12.83±0.69</td>
<td>9.36±1.95</td>
</tr>
<tr>
<td>MT-ATP8</td>
<td>11.69±0.76</td>
<td>10.72±0.74</td>
<td>10.89±0.73</td>
<td>7.04±1.78</td>
</tr>
<tr>
<td>MT-CO1</td>
<td>14.82±0.49</td>
<td>13.73±0.48</td>
<td>13.71±0.69</td>
<td>11.63±1.25</td>
</tr>
<tr>
<td>MT-CO2</td>
<td>14.71±0.49</td>
<td>13.47±0.5</td>
<td>13.65±0.66</td>
<td>11.06±1.2</td>
</tr>
<tr>
<td>MT-CO3</td>
<td>14.79±0.52</td>
<td>13.53±0.5</td>
<td>13.67±0.71</td>
<td>11.56±1.13</td>
</tr>
<tr>
<td>MT-CYB</td>
<td>13.84±0.52</td>
<td>12.64±0.53</td>
<td>12.59±0.71</td>
<td>9.76±1.9</td>
</tr>
<tr>
<td>MT-ND1</td>
<td>13.73±0.57</td>
<td>12.55±0.58</td>
<td>12.51±0.83</td>
<td>9.04±1.96</td>
</tr>
<tr>
<td>MT-ND2</td>
<td>13.78±0.58</td>
<td>12.48±0.57</td>
<td>12.64±0.71</td>
<td>8.58±2.28</td>
</tr>
<tr>
<td>MT-ND3</td>
<td>13.53±0.56</td>
<td>12.63±0.56</td>
<td>12.63±0.62</td>
<td>9.96±1.12</td>
</tr>
<tr>
<td>MT-ND4</td>
<td>14.63±0.47</td>
<td>13.39±0.53</td>
<td>13.52±0.71</td>
<td>9.86±2.11</td>
</tr>
<tr>
<td>OGDH</td>
<td>6.45±0.62</td>
<td>6.33±0.59</td>
<td>6.68±0.57</td>
<td>5.47±0.75</td>
</tr>
<tr>
<td>PFKP</td>
<td>5.32±0.99</td>
<td>4.81±1</td>
<td>6.36±0.96</td>
<td>4.66±0.94</td>
</tr>
<tr>
<td>PGAM1</td>
<td>5.22±0.57</td>
<td>5.41±0.46</td>
<td>6.22±0.52</td>
<td>5.14±0.68</td>
</tr>
<tr>
<td>PGK1</td>
<td>7.91±0.7</td>
<td>8.12±0.59</td>
<td>9.09±0.63</td>
<td>8.19±0.96</td>
</tr>
<tr>
<td>PKM</td>
<td>8.59±0.59</td>
<td>8.59±0.65</td>
<td>9.47±0.5</td>
<td>7.66±0.87</td>
</tr>
<tr>
<td>SDHA</td>
<td>1.81±0.48</td>
<td>1.69±0.43</td>
<td>1.88±0.5</td>
<td>2.29±0.93</td>
</tr>
<tr>
<td>SLC16A1</td>
<td>2.84±1.08</td>
<td>2.85±0.99</td>
<td>4.37±1.44</td>
<td>3.82±0.94</td>
</tr>
<tr>
<td>SLC16A3</td>
<td>5.13±1.02</td>
<td>4.88±0.87</td>
<td>6.21±0.92</td>
<td>4.42±1.2</td>
</tr>
<tr>
<td>SLC2A1</td>
<td>5.32±1.21</td>
<td>4.73±0.97</td>
<td>7.17±1.06</td>
<td>5.24±1.28</td>
</tr>
<tr>
<td>SLC2A3</td>
<td>4.36±1.1</td>
<td>4.4±1.12</td>
<td>5.34±1.21</td>
<td>4.72±1.22</td>
</tr>
<tr>
<td>SLC2A4</td>
<td>0.72±0.48</td>
<td>0.65±0.37</td>
<td>0.62±0.52</td>
<td>0.47±0.32</td>
</tr>
<tr>
<td>SUCGL1</td>
<td>5.48±0.42</td>
<td>5.68±0.53</td>
<td>5.78±0.47</td>
<td>5.18±0.75</td>
</tr>
<tr>
<td>TOMM20</td>
<td>7.57±0.61</td>
<td>8.01±0.65</td>
<td>7.96±0.51</td>
<td>7.63±0.35</td>
</tr>
<tr>
<td>TPI1</td>
<td>8.97±0.51</td>
<td>8.86±0.69</td>
<td>9.88±0.53</td>
<td>8.15±1.24</td>
</tr>
</tbody>
</table>
Table 1: Clinical characteristics of each cluster and mean RNA expression levels. (ACO2=Aconitase 2, ALDOA=Aldolase, Fructose-Bisphosphate A, CS=Citrate Synthase, ENO1=Enolase 1, FH=Fumarate Hydratase, GAPDH=Glyceraldehyde-3-Phosphate Dehydrogenase, GPI=Glucose-6-Phosphate Isomerase, HK2=Hexokinase 2, IDH3A=Isocitrate Dehydrogenase (NAD(+) 3 Catalytic Subunit Alpha, LDHA=Lactate Dehydrogenase A, MDH2=, Malate Dehydrogenase 2 MT-ATP6= Mitochondrially Encoded ATP Synthase Membrane Subunit 6, MT-ATP8= Mitochondrially Encoded ATP Synthase Membrane Subunit 8, MT-CO1=Mitochondrially Encoded Cytochrome C Oxidase I, MT-CO2= Mitochondrially Encoded Cytochrome C Oxidase II, MT-CO3=, Mitochondrially Encoded Cytochrome C Oxidase III MT-CYB= Mitochondrially Encoded Cytochrome B, MT-ND1=Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1 MT-ND2= Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 2 MT-ND3= Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1, MT-ND4 = Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4, OGDH=Oxoglutarate Dehydrogenase, PFKP=Phosphofructokinase, Platelet, PGAM1=Phosphoglycerate Mutase 1, PGK1=Phosphoglycerate Kinase 1, PKM=Pyruvate Kinase M, SDHA=Succinate Dehydrogenase Complex Flavoprotein Subunit A, SLC16A1=Solute Carrier Family 16 Member 1, SLC16A3= Solute Carrier Family 16 Member 3, SLC2A1=Solute Carrier Family 2 Member 1, SLC2A3= Solute Carrier Family 2 Member 3, SLC2A4= Solute Carrier Family 2 Member 4, SUCLG1=Succinate-CoA Ligase GDP/ADP-Forming Subunit Alpha, TOMM20= Translocase Of Outer Mitochondrial Membrane 20, TPI1=Triosephosphate Isomerase 1.}
<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>coef lower 95%</th>
<th>coef upper 95%</th>
<th>exp(coef) lower 95%</th>
<th>exp(coef) upper 95%</th>
<th>exp to z</th>
<th>p</th>
<th>log2(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 0</td>
<td>-0.54</td>
<td>0.58</td>
<td>0.21</td>
<td>-0.95</td>
<td>-0.14</td>
<td>0.39</td>
<td>0.87</td>
<td>0.00</td>
<td>2.63</td>
<td>0.01</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>-0.87</td>
<td>0.42</td>
<td>0.25</td>
<td>-1.36</td>
<td>-0.37</td>
<td>0.26</td>
<td>0.69</td>
<td>0.00</td>
<td>3.44</td>
<td>&lt;0.005</td>
</tr>
</tbody>
</table>

Table 2: Univariate (A) and multivariate Cox regression (B) for each cluster. Cluster 2 and Stage I are the references.
Figure Legends:

Figure 1: Prognostic impact of MCT4, MCT1, TOMM20 using median, 75<sup>th</sup> and 25<sup>th</sup> percentile as cut-offs (MCT4: Monocarboxylate Transporter 4, MCT1: Monocarboxylate Transporter 1, TOMM20: Translocase Of Outer Mitochondrial Membrane 20): Kaplan Meier estimates of overall survival. Shadows indicate 95% Confidence Intervals.

Figure 2: A: Explained variation proportion on gene expression us for clustering; B: Correlation between gene expression and PCA; C: Representation of the four clusters: blue= cluster 0, orange= cluster 1, green= cluster 2, red= cluster 3. D: Kaplan Meier estimates of overall survival by Clusters. Shadows indicate 95% Confidence Intervals.