APPENDIX

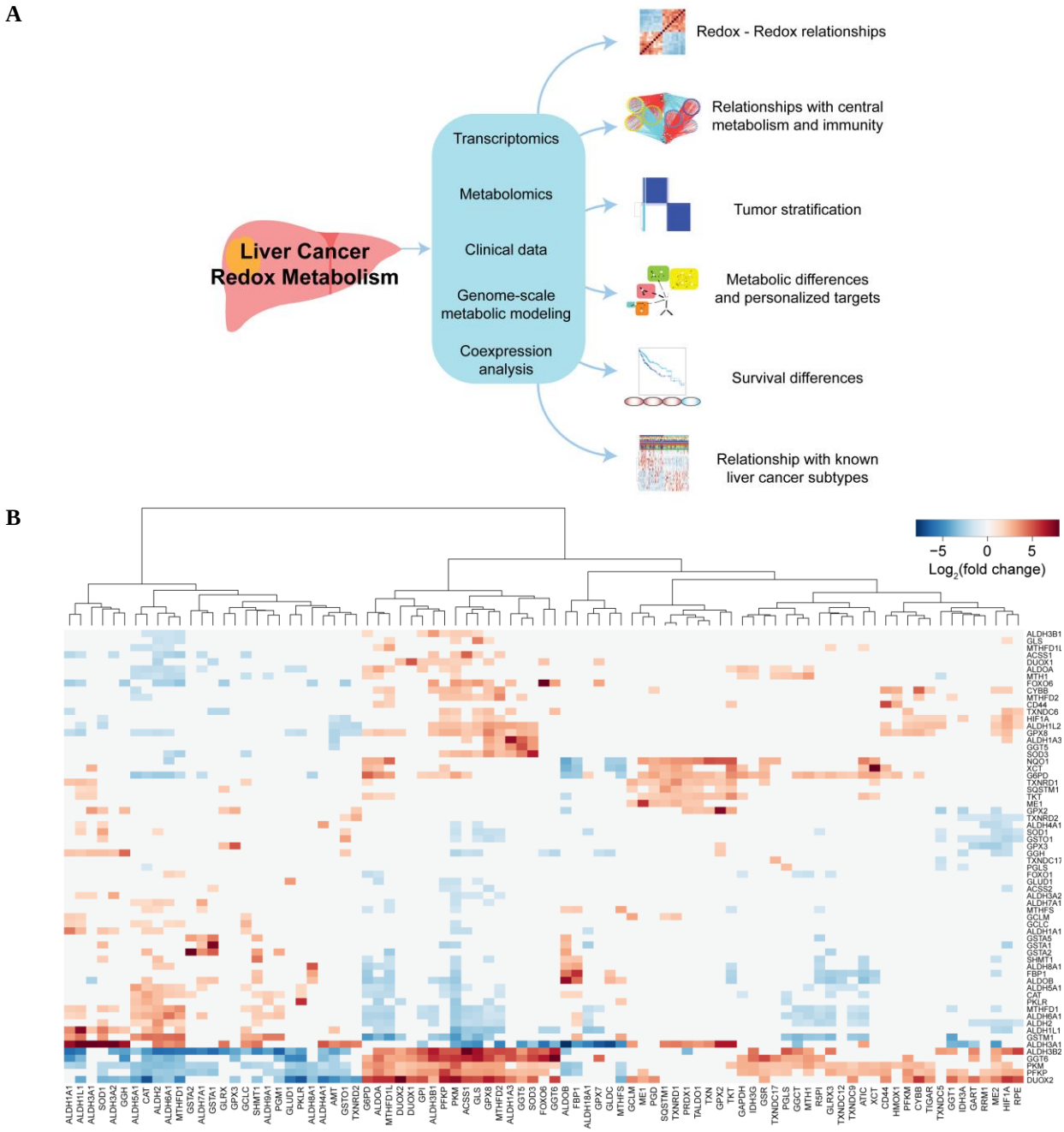
Characterization of heterogeneous redox responses in hepatocellular carcinoma patients using network analysis

Rui Benfeitas, Gholamreza Bidkhori, Bani Mukhopadhyay, Martina Klevstig, Muhammad Arif, Cheng Zhang, Sunjae Lee, Resat Cinar, Jens Nielsen, Mathias Uhlen, Jan Boren, George Kunos, Adil Mardinoglu

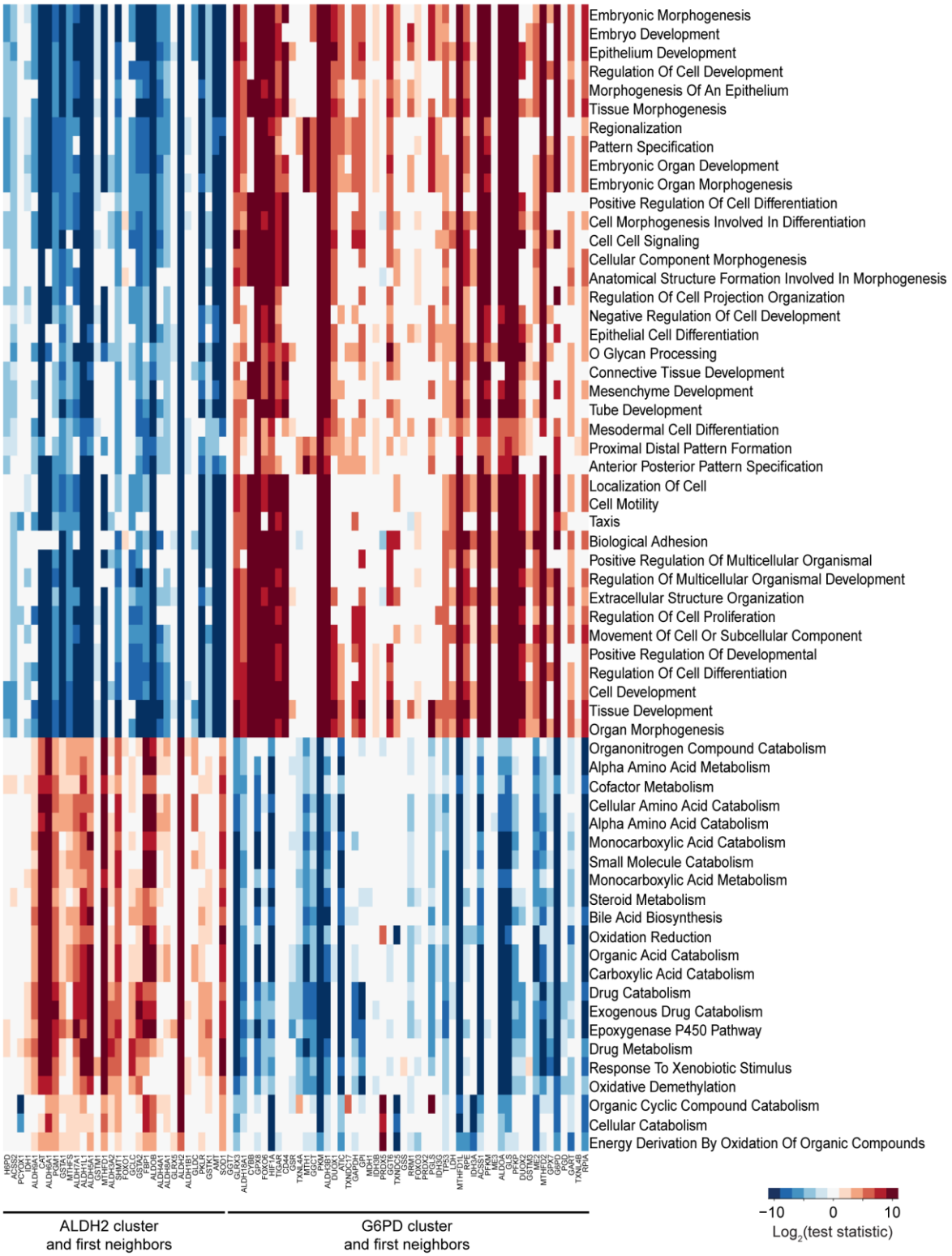
Adil Mardinoglu
Email: adilm@scilifelab.se

This PDF file includes:

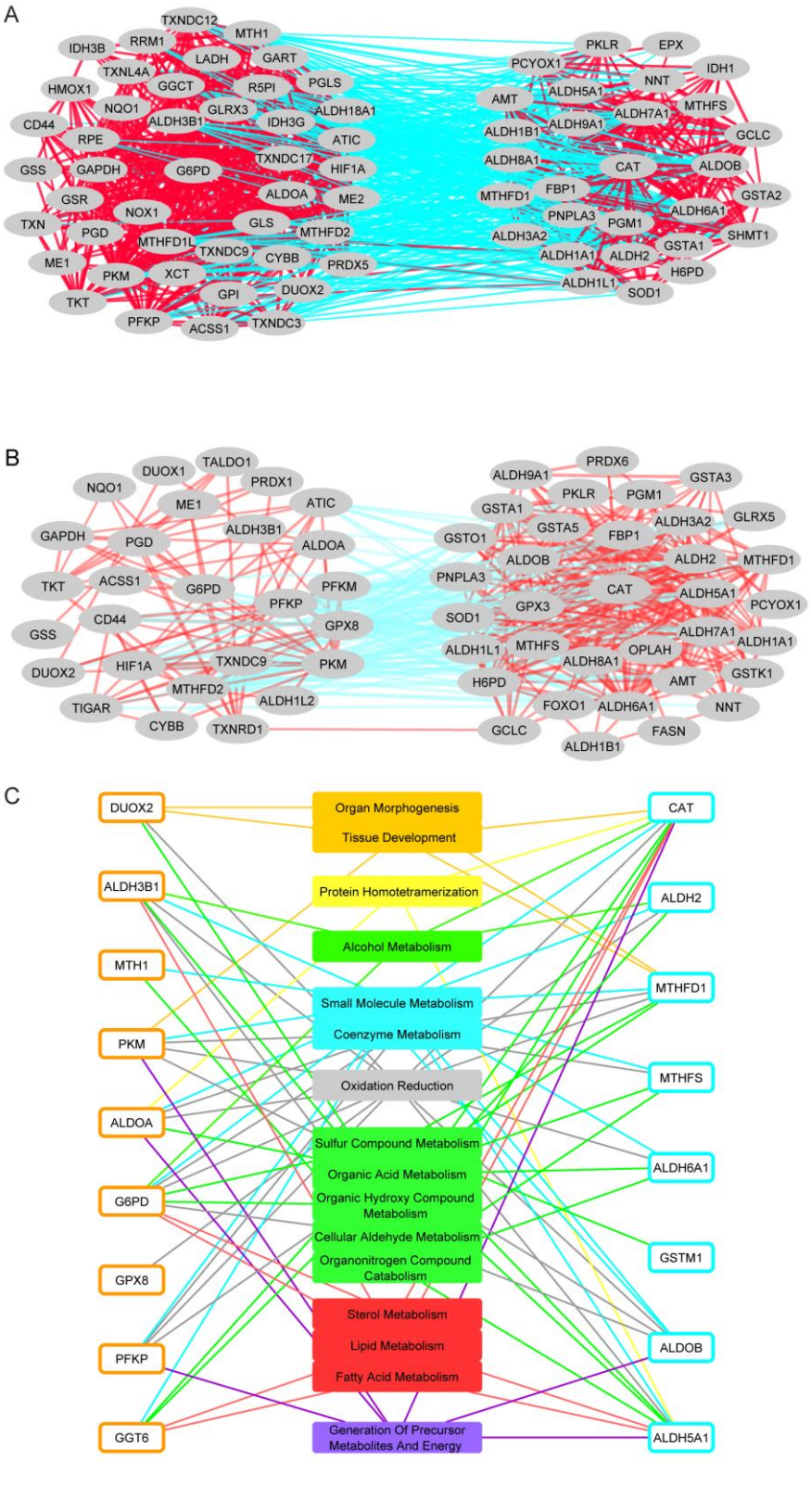
Appendix Figures S1 – S9
Doc. S1



Appendix Figure S1 – A. A systematic analysis of HCC integrating different omics highlighted mechanistic differences between redox genes and relationships with other processes, and permitted tumor stratification and identification of subtype-specific features. B. Similar patterns of differentially expressed genes are observed upon comparison of subjects displaying high vs low expression of redox genes (columns). Hierarchical clustering of columns for all genes displaying significantly differentially expressed ($Q < 0.05$) and $\log_2$-fold changes >1. Columns and rows displaying fewer than 5 statistically significant observations are not presented. Note that differential expression with respect to the top 10 correlated genes (Fig 1 inset) in the ALDH2 and G6PD clusters grouped the genes in each closely together and distant from one another.

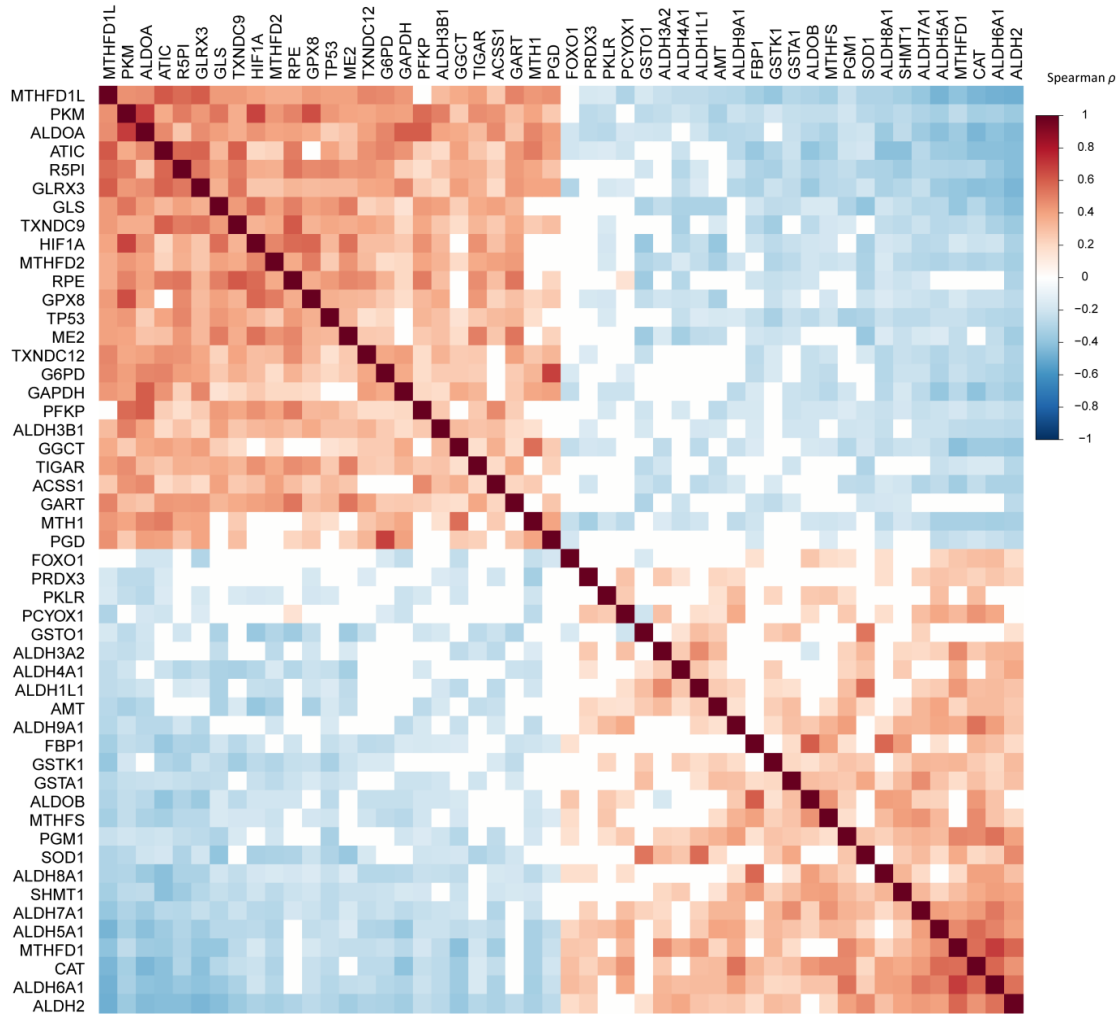


Appendix Figure S2– GSEA performed in tumors displaying high vs low expression of genes and the first neighbors of the ALDH2 (left) and G6PD (right) clusters (Fig 4). Biological processes (rows) that were significantly (Q < 0.05) enriched in more than 50% of the columns are displayed.

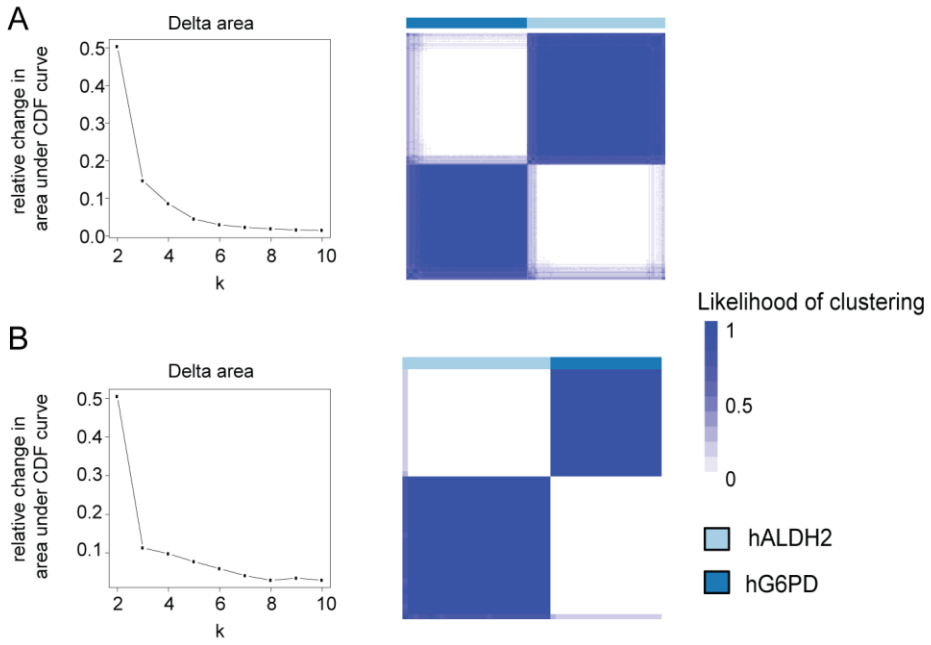


Appendix Figure S3- G6PD and CAT comprise two highly correlated gene groups with inter-group negative correlations and with opposing cellular functions. A. Redox genes that were highly correlated with CAT and G6PD are shown (absolute Spearman $\rho > 0.3$, $Q < 10^{-6}$), indicating the first neighbors of the two genes. Edges show negative

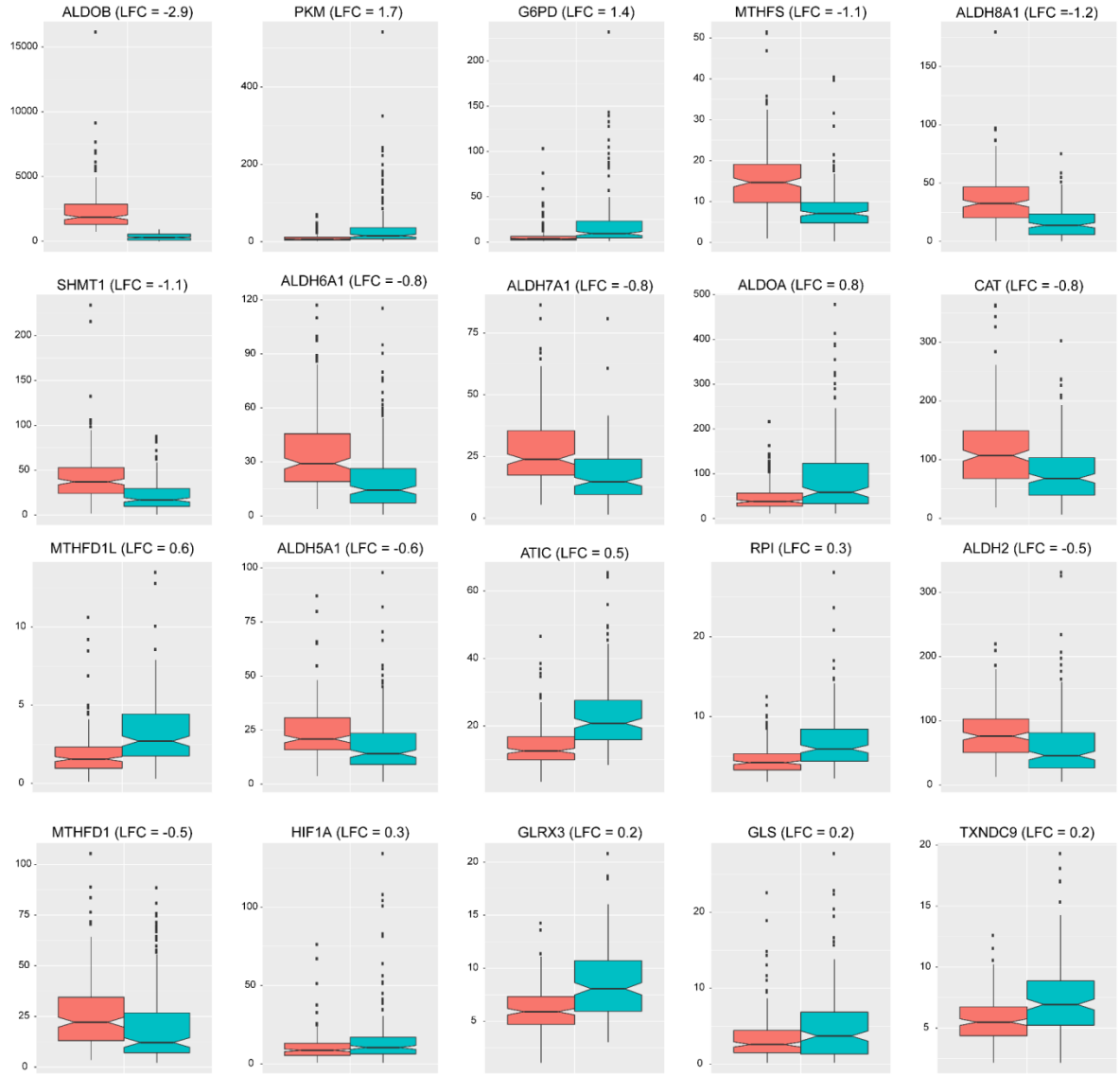
(blue) and positive (red) correlations, with thicker lines indicating stronger correlations (highest absolute Spearman ρ ≈ 0.8). The full set of correlations is displayed in Dataset 3. B. Using transcriptomics data from an independent dataset (Chaisaingmongkol et al, 2017), we identified co-expression with CAT and G6PD (C, all absolute Spearman ρ > 0.4), similar to our observations using transcriptomic data. Only statistically significant correlations (Q < 0.01) are presented. C. Redox genes that were differentially expressed with respect to both CAT and G6PD expression were determined (Q < 0.05), and their associated and significantly enriched processes (Q < 0.01) are indicated. Only processes involving at least one redox gene are displayed. The full enrichment analysis is shown in Dataset 8.



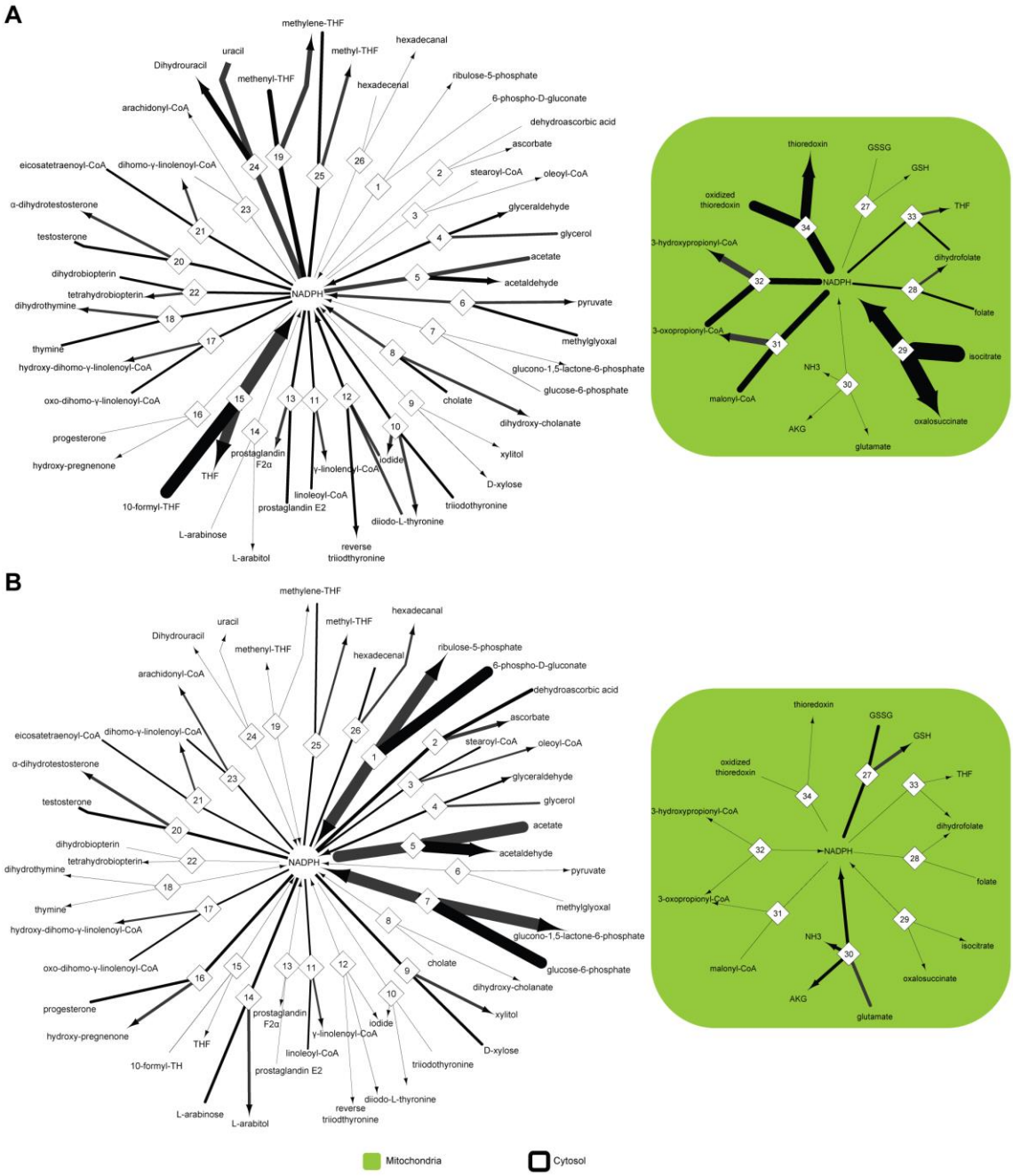
Appendix Figure S4– Co-expression analysis of the top 25 co-expressed genes (all displayed Spearman's correlations Q < 0.01).



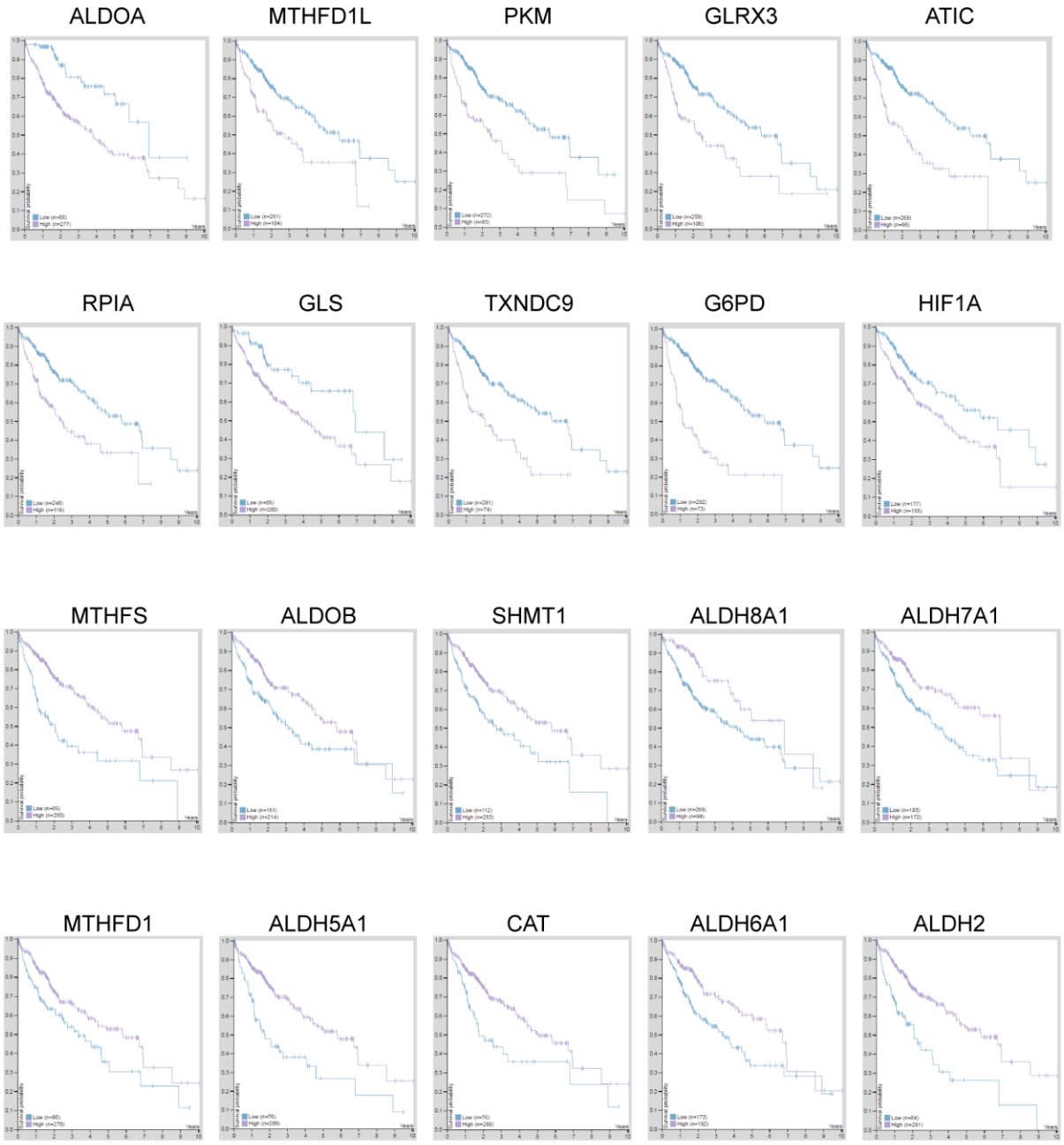
Appendix Figure S5– Consensus clustering using redox gene expression from two independent datasets clusters tumors into 2 major clusters. Relative change in area under the cumulative distribution function (CDF) and heat maps of the consensus clustering for 2 clusters (*k = 2*), using data from (Lee et al, 2004; Lee et al, 2006) (A) and (Chaisaingmongkol et al, 2017) (B).



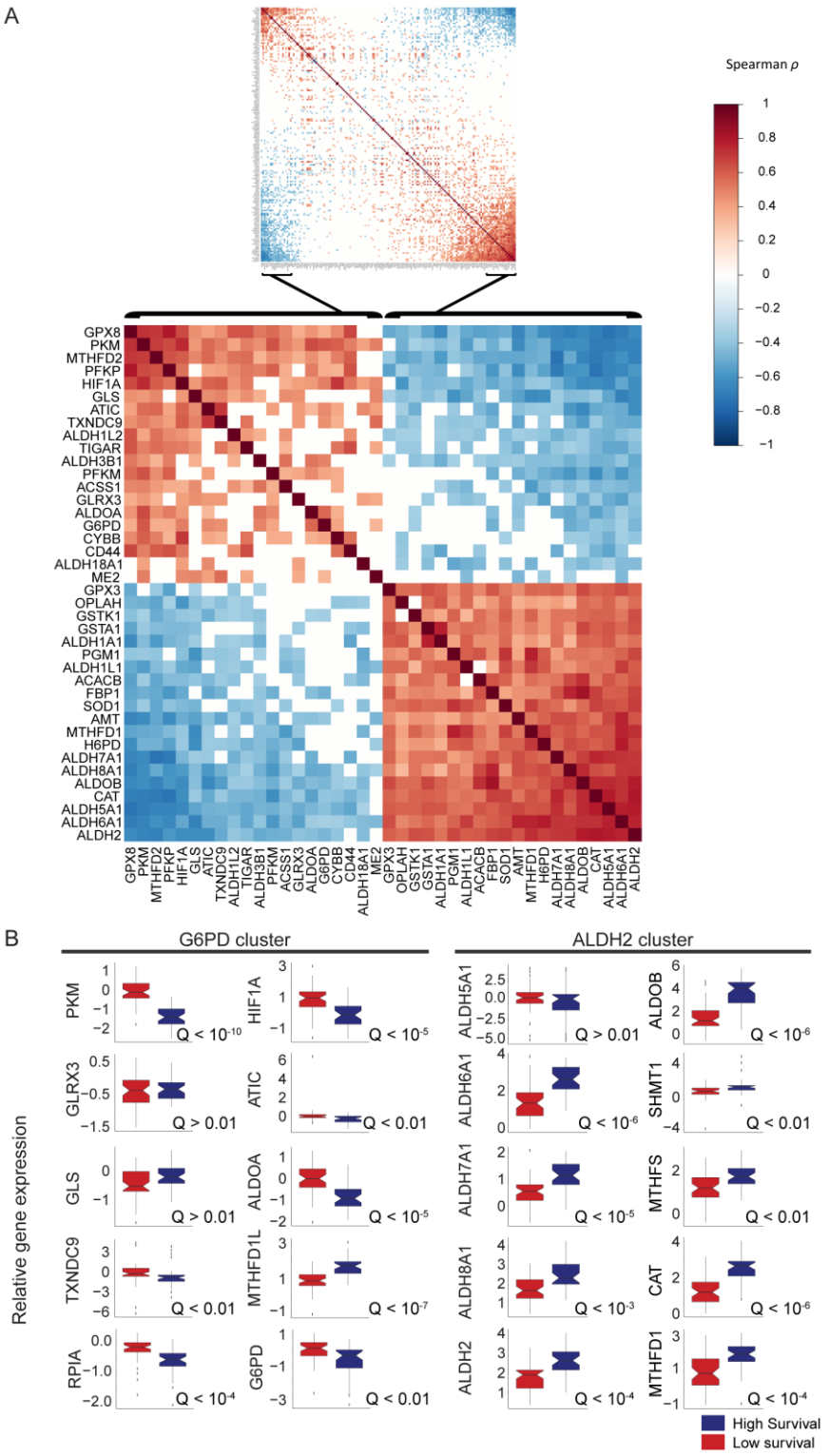
Appendix Figure S6–Redox genes are differentially expressed between tumor clusters, but the expression levels of some are markedly more different than others. Gene expression in the hALDH2 (red) and hG6PD (blue) clusters are displayed, together with log fold changes. Genes were sorted according to absolute log fold change, and the top row comprises the best stratifying genes. ALDOB, PKM, G6PD, ALDH8A1, and MTHFS were highly differentially expressed ($Q < 10^{-18}$, DESeq2). Note that the expression of *MTHFS* (log fold change = -1.1, $Q < 10^{-33}$) was clearly distinct between groups and was thus a better stratifying gene than *ALDH8A1* (log fold change = -1.2, $Q < 10^{-17}$). For all genes, $Q < 0.005$, with exception of *GLS* with $Q > 0.05$.



Appendix Figure S7- hALDH2 and hG6PD tumors display substantial differences in NADPH metabolism. GEMs for hALDH2 (A) and hG6PD (B) tumors were integrated with gene expression data for all tumors in a group. Mitochondrial (green box) and cytosolic processes are represented. Numbers indicate enzyme-catalyzed reactions or transport reactions (Dataset 11), e.g., glucose 6 phosphate dehydrogenase catalyzes reaction 7. Only reactions displaying fluxes >1 in at least one of the models are shown. For bidirectional reactions (e.g., transport reactions), fluxes were used to select the direction of that reaction. Thin arrows indicate reactions with low or null fluxes (A: 1-3, 7, 9, 14, 16, 26, 27, 30; B: 6, 8, 10, 12, 13, 15, 18, 19, 22, 24, 28, 29, 31-34).



Appendix Figure S8- Kaplan-Meier survival plots for the best separation for ALDH2 and G6PD cluster genes, retrieved from the Human Pathology Atlas (Uhlen et al., 2017). Detailed statistics for individual genes are present in Dataset 16.



Appendix Figure S9– Analysis of data from two separate cohorts reproduces observations in transcriptomic data. A. Using transcriptomics data from an independent cohort (Chaisaingmongkol et al, 2017), we identified antagonistic gene expression similar to our observations using transcriptomic data (main text). Only statistically significant correlations (Q < 0.01) are presented. B. Observations of another dataset were consistent between our observed patterns of gene expression and patient survival groups (Lee et al, 2004; Lee et al, 2006). Gene expression was

compared in high vs low survival groups, defined previously (Lee et al, 2004; Lee et al, 2006) as *Cluster B* and *A*, respectively, and $Q < 0.01$ was considered statistically significant (Mann-Whitney U test). Gene expression is reported with respect to matched non-tumor samples and considering samples with Cy-3 labeled cDNA (Lee et al, 2004; Lee et al, 2006).

DOC S1 - G6PD AND CAT COMPRISE TWO HIGHLY CORRELATED ANTAGONISTIC GENE CLUSTERS WITH OPPOSING CELLULAR FUNCTIONS

Our observations in the main text indicate that the heterogeneity in the redox metabolism in HCC tumors comprises two antagonistic clusters, respectively including *CAT* and *G6PD*. These two genes are of particular interest not only because they are potential prognostic markers (Table 1) but also due to their roles in redox metabolism. CAT is an important antioxidant with catalytic activity that does not require NADPH, whereas G6PD is one of the main sources of NADPH that may be used to maintain the antioxidant activity of peroxidases. Both proteins are often imbalanced in cancer cells (Benfeitas et al, 2017; Glorieux et al, 2011; Ray et al, 2000; Skrzydlewska et al, 2005). We therefore analyzed the expression patterns of these two genes in different tumors and the prognostic indications of their co-expressed genes.

Correlation analysis indicated that CAT and G6PD expression was associated with the expression of most genes in the two clusters (Appendix Figure S3A, absolute Spearman $\rho > 0.4$ and $Q < 2\times10^{-13}$, compare with Fig 1B). CAT and G6PD were positively co-expressed with 34 and 49 redox genes, respectively. Their co-expressed genes were highly correlated with genes within each of the two groups but negatively correlated with most genes in the other group. These observations were validated in an independent cohort (Appendix Figure S3B). Several potential markers of favorable prognosis among genes positively co-expressed with CAT and markers of unfavorable prognosis among genes positively were co-expressed with G6PD. Differential expression analysis showed that only 15 redox genes were significantly ($Q < 0.05$) and simultaneously differentially expressed in tumors with high vs low CAT and G6PD expression. GSTM1 and CAT-co-expressed genes (ALDOB, ALDH6A1, CAT, ALDH5A1, MTHFD1, and MTHFS) were positively co-expressed with CAT expression ($\log_2$-fold changes > 1.55, $Q < 0.05$) but negatively co-expressed with G6PD ($\log_2$-fold changes < -1.6, $Q < 0.005$). In turn, the GPX8, GGT6, DUOX2 and G6PD-co-expressed genes (ALDOA, ALDH3B1, MTH1, PFKP, PKM) were positively co-expressed with G6PD ($\log_2$-fold changes > 1.5, $Q < 0.05$) but negatively co-expressed with CAT ($\log_2$-fold changes <-1.6, $Q < 0.015$). GSEA and reporter metabolite analysis with respect to tumors displaying high vs low CAT or G6PD expression displayed similar observations to those above (Fig 3). Importantly, no functional process was simultaneously up- or downregulated with respect to CAT and G6PD expression. We also identified several processes involving redox genes in both clusters (Appendix Figure S3C), including those related to the metabolism of organic compounds, lipids, cofactors, coenzymes and small molecules, production of precursor metabolites and energy, as well as tissue development. Only organ morphogenesis and tissue development were upregulated in high G6PD and low CAT-expressing tumors. All other terms, including oxidation-reduction, showed the opposite trend.