Robust and independent metabolic transcriptional components in 34,494 patient-derived samples and cell lines: a resource for translational cancer research

V.C. Leeuwenburgh[1,2,#], C.G. Urzúa-Traslaviña[1,#], A. Bhattacharya[1], M.T.C. Walvoort[2], M. Jalving[1], S. de Jong[1], R.S.N. Fehrmann[1]

[1] Department of Medical Oncology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

[2] Department of Chemical Biology, Stratingh Institute for Chemistry, University of Groningen, the Netherlands

[#] Contributed equally

Corresponding author:

Rudolf S.N. Fehrmann, MD, PhD

Department of Medical Oncology, University Medical Center Groningen

P.O. Box 30.001, 9700 RB Groningen, the Netherlands

Tel. +31503612821, Fax +31503614862

E-mail: r.s.n.fehrmann@umcg.nl

SUMMARY

Patient-derived bulk expression profiles of cancers can provide insight into transcriptional changes that underlie reprogrammed metabolism in cancer. However, these bulk profiles represent the average expression pattern of all heterogeneous tumor and non-tumor cells present in the biopsy. Therefore, subtle transcriptional footprints of metabolic processes can be concealed by other biological processes and experimental artifacts. We therefore performed consensus Independent Component Analyses (c-ICA) with 34,494 bulk expression profiles of patient-derived tumor biopsies, non-cancer tissues and cell lines. c-ICA enabled us to create a transcriptional metabolic landscape in which many robust metabolic transcriptional components (mTCs) and their activation score in individual samples were defined. Here we demonstrate that this metabolic landscape can be used to explore associations between metabolic processes and drug sensitivities, patient outcomes, and the composition of the immune tumor microenvironment. The metabolic landscape can be explored at http://www.themetaboliclandscapeofcancer.com.

KEYWORDS

Metabolism, Immunometabolism, Transcriptome Analysis, Pan-Cancer, Independent Component Analysis

INTRODUCTION

Reprogrammed energy metabolism is a hallmark of cancer (Hanahan and Weinberg, 2011). Metabolic reprogramming supports the survival, proliferation, and maintenance of cancer cells by ensuring sufficient biosynthetic capacity, redox potential, and energy (Cantor and Sabatini, 2012; Pavlova and Thompson, 2016; Vazquez et al., 2016). Additionally, metabolic reprogramming enables tumor cells to adapt to challenging microenvironmental conditions, such as hypoxia and low nutrient availability, and ultimately become resistant to cancer treatment (Huang et al., 2014; Viale and Draetta, 2016). Moreover, metabolic reprogramming of cancer cells influences the composition and function of immune cells within the tumor microenvironment, which affects their sensitivity to immunotherapies (Herbel et al., 2016; Le Bourgeois et al., 2018; Liu et al., 2018; Quail and Joyce, 2013).

Metabolic dependencies have been successfully exploited to treat cancer, as illustrated by the efficacy of antifolate drugs such as methotrexate (Walling, 2006). Increased knowledge about cancer cell metabolism has resulted in novel therapeutic targets, including glutaminase or mutant forms of IDH1/2, that are currently being evaluated in pre-clinical models and phase I/II clinical trials (Luengo et al., 2017). However, adverse effects or a lack of effectiveness have hampered clinical development of most metabolic therapies, partly due to the complex involvement of the tumor micro-environment (Luengo et al., 2017). Many currently registered targeted drugs were developed based on insights into the molecular mechanisms of cancer based on model systems of human cancer (Ghaffari et al., 2015). However, most of these cell line-based models do not incorporate important features of the human cancer of interest, in particular the immune cells

present in the tumor microenvironment (Domcke et al., 2013; Hynds et al., 2018; Jiang et al., 2016; Klijn et al., 2015; Vincent and Postovit, 2017).

Evidence is emerging that predominantly transcriptional changes underlie the reprogrammed metabolism in cancer cells (Desvergne et al., 2006; Peng et al., 2018b; Ward and Thompson, 2012). The interplay between metabolism and regulation of gene expression may also explain the metabolic plasticity of cancer cells (Martin-Martin et al., 2018). The interaction is bidirectional: changes in gene expression can influence metabolite levels, and metabolic changes can result in altered gene expression (Martin-Martin et al., 2018). The availability of large numbers of gene expression profiles from a broad spectrum of cancer types in the public domain provides a unique opportunity to study cancer metabolism at the transcriptomic level in patient tumors. Since these expression profiles are derived from complex biopsies, they contain both tumor cells and immune cells present in the tumor microenvironment. Recent advances have made it possible to infer the composition and functionality of the immune tumor microenvironment from the average gene expression profiles of complex biopsies, thus enabling interrogation of the interaction between metabolic processes of tumor cells and the immune tumor microenvironment (Newman et al., 2015). This additional layer of information might help to identify novel metabolic targets and to design more effective metabolically-targeted therapies.

In the present study we analyzed 34,494 gene expression profiles obtained from the Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA), the Cancer Cell Line Encyclopedia (CCLE) and the Genomics of Drug Sensitivity in Cancer portal (GDSC) (Barret et al., 2013; Barretina et al., 2012; Yang et al., 2013). We created a metabolic landscape in which we defined cross-dataset robust metabolic transcriptional components (mTCs) and calculated the activity scores

4

for each mTC in individual samples. We also demonstrated how this metabolic landscape

(www.themetaboliclandscapeofcancer.com) can be used to explore associations between

metabolic processes and drug sensitivities, patient outcomes and the composition of the immune

tumor microenvironment.

RESULTS

A subset of transcriptional components is associated with metabolic processes

Previously, we collected gene expression data from four databases: the Gene Expression Omnibus

(GEO dataset, n = 21,592), The Cancer Genome Atlas (TCGA dataset, n = 10,817), the Cancer Cell

Line Encyclopedia (CCLE dataset, n = 1,067), and the Genomics of Drug Sensitivity in Cancer (GDSC

dataset, n = 1,018) (**Figure 1A**), totaling 34,494 samples (Bhattacharya et al., 2020). Of this total,

28,200 expression profiles originated from patient-derived complex tissue cancer biopsies, 4,209

from complex tissue biopsies of non-cancerous tissue, and 2,085 from cell lines. The samples in

these four databases represent 89 cancer tissue types and subtypes and 19 non-cancerous tissue

types. For GEO and CCLE data sets, the expression profiles were generated with Affymetrix HG-

U133 Plus 2.0. Expression profiles within the GDSC dataset were generated with Affymetrix

Human Genome U219, and TCGA profiles originated from RNA sequencing.

Gene expression profiling measures the net expression level of individual genes, thus

reflecting the integrated activity of underlying regulatory factors, including experimental, genetic,

and non-genetic factors. To gain insight into the number and nature of these regulatory factors

and their effects on gene expression levels, we previously performed consensus-independent

component analysis (c-ICA) on the total of 34,494 samples from each of the abovementioned

5

datasets (Bhattacharya et al., 2020). The c-ICA resulted in transcriptional components (TCs) in which each gene has a specific weight. This weight describes how strongly and in which direction the underlying transcriptional regulatory factor influences the expression level of that gene. The analysis yielded 855, 1383, 466, and 467 TCs for GEO, TCGA, CCLE, and GDSC datasets, respectively (**Figure 1A**).

Gene set enrichment analysis (GSEA) with 608 gene sets that describe metabolic processes was performed to identify TCs associated with metabolic processes. These gene sets were selected from the gene set collections Biocarta (n = 7), the Kyoto Encyclopedia of Genes and Genomes (KEGG, n = 64), the Gene Ontology Consortium (GO, n = 508), and Reactome (n = 29) within the Molecular Signatures DataBase (MSigDB, v6.1; for the systematic selection strategy see **Methods**). We performed consensus clustering on the GSEA results to reduce potential biological redundancy in the gene set definitions (**Figure S1A** and **B**). This resulted in 50 clusters that each described a different metabolic theme (**Table S1**). In total, 132 (GEO), 151 (TCGA), 136 (CCLE), and 137 (GDSC) TCs were identified that showed significant enrichment for metabolic processes (**Figure 1A** and **B**; see **Methods** for the systematic selection strategy). In the present study, this subset of TCs is referred to as metabolic TCs (mTCs). These mTCs constitute the metabolic transcriptional footprints present in our broad set of samples: patient-derived samples, cancer cell-line samples, and non-cancer samples.

Metabolic TCs are robust across different datasets and platforms

Pair-wise comparison between mTCs showed that at least 90% of mTCs identified in one dataset correlated strongly ($|r_s| \geq 0.5$) with at least one mTC identified in the other three datasets (**Figure**

6

1C and S1C). Focusing specifically on the mTCs identified in the two patient-derived datasets, we observed that 113/132 (85.6%) of the GEO mTCs and 127/151 (84.1%) of the TCGA mTCs had a strong and significant pair-wise correlation ($|r_s| \geq 0.5$ with P value < 0.05; Figure 1D). For 19/132 GEO mTCs, the highest pair-wise correlation with an mTC from the TCGA dataset was ≤ 0.5. However, for 13 out of these 19 mTCs, still a very high enrichment (Z-transformed P value > 5) for at least one metabolic gene set was observed (Figure 1E). Furthermore, ten of the 19 GEO mTCs captured a metabolic process, which was active in a single tumor type that was present only in the GEO dataset. For example, GEO mTC 119 is active in a subset of myeloma samples, which were not represented in the TCGA dataset (Figure S1D). These results indicate that most of the identified mTCs were robust and independent of dataset-specific or platform-specific characteristics.

Metabolic TCs identified new genes potentially involved in metabolic processes

To identify clusters of genes with similar weight patterns across the mTCs, we clustered the genes based on their weight in the mTCs. We did this only on the subset of genes with an absolute weight > 3 in at least one mTC (Figure 2A and 2D). In most clusters, the member genes had a high weight in only one mTC, indicating that these genes predominately participate in one metabolic process.

As mentioned in the previous paragraph, we observed many concordant mTCs between different datasets. Two strongly correlated mTCs, GEO mTC 54 and TCGA mTC 127 ($r_s = 0.77$), which both showed enrichment for glycolysis and the metabolic process of ADP were analyzed in depth (Figure 2B and 2C, Table S1). GEO mTC 54 contained 262 genes with an absolute weight >

7

3, of which 155 (59.1%) also had absolute weight > 3 in TCGA mTC 127. Both mTCs contained multiple top-ranked genes that are known targets of the HIF-1 complex and genes previously found to be part of a hypoxic signature (Benita et al., 2009; Ye et al., 2018). Several genes ranking at the top of both GEO mTC 54 and TCGA mTC 127 (e.g., *FAM162A*, *C4orf3*, *C4orf47*, and *ANKRD37*) are currently not a member of any of the 608 metabolic gene sets. However, these data suggest that these four genes are involved in glycolysis and are possibly hypoxia related. Indeed, several studies have suggested that at least *FAM162A* and *ANKRD37* are regulated by the transcription factor HIF-1α (Copple et al., 2012; Sørensen et al., 2015).

As a second example, we investigated two highly correlated mTCs, GEO mTC 11 and TCGA mTC 141 (r_s = 0.67), which showed enrichment for mitochondrial metabolic processes such as oxidative phosphorylation and the TCA cycle (**Figure 2E** and **2F, Table S1**). GEO mTC 11 contained 427 genes with an absolute weight > 3, of which 270 (63.2%) also had an absolute weight > 3 in TCGA mTC 141. In these two mTCs, *C6orf136* and *IMMT* are top-ranked genes which are currently not assigned to any of the 608 metabolic gene sets. *C6orf136* and *IMMT* were both previously identified in proteome profiles of functional mitochondria (Lefort et al., 2009). These results suggest that mTCs could be used to assign metabolic functions to genes currently not members of known gene sets describing metabolic processes.

Clustering activity scores of mTCs reveal multiple metabolic subtypes

In addition to mTCs, consensus ICA also provides a "mixing matrix" in which each column corresponds to an mTC and each row corresponds to a sample. The mixing matrix contains weights that are interpreted as measurements of the activity of the corresponding mTCs in an

individual sample; we refer to these as activity scores. By clustering these activity scores, we could investigate the heterogeneity of the metabolic transcriptome in a broad range of cancer subtypes.

For each of the four datasets, samples were hierarchically clustered based on their mTC activity scores (**Figure 3A**, **3B** and **S2A**, **S2B**). We selected the cutoff heights of the resulting dendrograms so that every cluster – referred to as metabolic subtype – contained at least 50 samples (**Figure S2C** and **S2D**). This clustering analysis divided the 21,592 GEO samples into 67 metabolic subtypes with a median of 276 samples per subtype (range 54-1252), and the 10,817 TCGA samples into 58 metabolic subtypes with a median of 167 samples per subtype (range 52-536). For an overview of the metabolic subtypes and their sample composition, see **Figure S3, S4** and **Table S2.** Three types of patterns emerged.

The first pattern consisted of several metabolic subtypes that mostly contained samples from a single cancer type. For example, 102/133 (76.7%) of thyroid cancer samples in the GEO dataset fell into a metabolic subtype that was characterized by a high activity score of GEO mTC 64 (**Table S2**). Similarly, 446/509 (87.6%) of thyroid cancer samples in the TCGA dataset fell into a metabolic subtype that was characterized by a high activity score of TCGA mTC 87 (**Table S2**). In line with the biology of thyroid tissue, both GEO mTC 64 and TCGA mTC 87 were highly enriched for thyroid hormone metabolism (**Table S1**).

The second pattern consisted of metabolic subtypes that contained samples from multiple tumor types. For example, GEO metabolic subtype 22 contained samples from 25 tumor types, including 42 ovarian cancer samples (22% of all ovarian cancer), 33 synovial sarcoma samples (97% of all synovial sarcoma), and 15 Ewing's sarcoma samples (58% of all Ewing's sarcoma; **Table S2**). GEO mTC 111 had the highest absolute median activity score in GEO metabolic subtype 22

9

(**Table S2**). This mTC showed enrichment for the metabolism of nicotinamide adenine dinucleotide phosphate (NADP) and genes involved in the activation of an innate immune response (**Table S1**).

The third pattern consisted of several tumor types that were not characterized by a few dominant metabolic subtypes, but the samples were divided across multiple metabolic subtypes. For example, the 3,512 breast cancer samples in the GEO dataset were divided across 33 metabolic subtypes (**Figure 3C**). These metabolic subtypes did not follow the breast cancer classification based on ER and HER2 receptor status (**Figure 3C** and **Table S2**). In line with this observation in the GEO dataset, the 1,100 breast cancer samples in the TCGA dataset were also scattered across 29 metabolic subtypes.

These results show that classification of samples based on metabolic subtype yields different patterns than current classification systems, such as histotype or receptor status in breast cancer.

Metabolic subtypes show prognostic value

We then investigated the clinical relevance of the metabolic subtypes. We had previously collected distant relapse-free survival (DRFS) data for 1,207 breast cancer samples (Bense et al., 2017). In the present study the breast cancer samples in the GEO dataset were divided across 33 of the 67 metabolic subtypes. Of these 33 subtypes, eight contained > 50 breast cancer samples with data available for DRFS: subtypes 15, 16, 20, 31, 32, 33, 34, and 35. We found that breast cancer samples assigned to metabolic subtypes 16 and 33 showed the best and worst DRFS, respectively (P value <0.05, Log-Rank test; **Figure 3E**). Distributions of standard prognostic factors

within these eight metabolic subtypes are presented in **Table S3**. Samples belonging to metabolic subtype 34 had the second-best DRFS. GEO mTC 50, which was enriched for genes involved in oxidative phosphorylation, had lower median activity in metabolic subtype 34 in comparison with other metabolic subtypes (P value <0.05, Mann-Whitney U test; **Figure S2E**). This indicates that lower expression of genes involved in oxidative phosphorylation was associated with a better DRFS.

Melanoma samples in the TCGA dataset were divided into 17 metabolic subtypes. Overall survival (OS) data was available for 437 samples. Metabolic subtypes 5 and 8 contained >20 samples with available OS data. Melanoma samples that were part of metabolic subtype 5 had better OS data than metabolic subtype 8 (P value <0.05, Log-Rank test; **Figure 3F**). In comparison to metabolic subtype 8, subtype 5 had a higher median activity score of TCGA mTC 7 (P value <0.05, Mann-Whitney U test; **Figure S2F**). TCGA mTC 7 was enriched for genes involved in the regulation of the metabolic processes of glycoproteins and phospholipids, cell-cell adhesion, and the activation of lymphocytes and leukocytes. Metabolic subtype 5 contained 82.9% melanoma samples from regional lymph nodes and no primary tumor samples (**Table S3**). In addition, subtype 5 (except for melanoma samples) contained samples from primary tumors of 27 other cancer types (**Table S2**). In contrast, subtype 8 contained 98% melanoma samples: 44.6% samples from regional lymph nodes and 28.0% primary tumor samples (**Table S2** and **S3**). This difference in sample origin and the associated difference in the activity of metabolic processes may explain the difference in OS data between the melanoma samples in metabolic subtypes 5 and 8. These results show that metabolic subtypes are associated with disease outcomes, and may therefore be relevant for clinical practice.

The activity of mTCs is associated with drug sensitivity

The CCLE and GDSC databases contain the sensitivities of cell lines to a large panel of drugs expressed as IC_{50} values. We observed clear associations between the activity scores of mTCs and the IC_{50} values of 270 drugs (**Table S4**).

For example, in the GDSC dataset the activity score of GDSC mTC 3 was associated with the IC_{50} values of nutlin-3a ($|r_s|$= 0.42; **Figure 4A**). Nutlin-3a targets the p53 pathway through inhibition of MDM2. In line with this, GDSC mTC 3 showed strong enrichment for genes involved in the p53 pathway with *MDM2* ranked as the second gene (**Table S1**). GDSC mTC 3 was strongly correlated with CCLE mTC 4 ($|r_s|$= 0.84), GEO mTC 57 ($|r_s|$= 0.79), and TCGA mTC 110 ($|r_s|$= 0.74) (**Figure 4B**), suggesting that this mTC was captured in cell line datasets as well as in the two patient-derived datasets. Cell lines with wildtype *TP53* had a higher activity score of GDSC mTC 3 (**Figure 4C**). Cell lines with wildtype *TP53* had a higher activity score of CCLE mTC 4 as well (**Figure 4D**).

In another example, the activity score of GDSC mTC 18 was found to be associated with the IC50 values of 142 drugs ($|r_s|$ range 0.20 – 0.44; **Figure 5A**). An increase in activity score of GDSC mTC 18 in a sample was associated with a higher IC_{50} value for 135 of these drugs, including the widely used antimetabolite 5-fluorouracil ($|r_s|$ = 0.41). GDSC mTC 18 was strongly correlated with CCLE mTC 28 ($|r_s|$= 0.84), GEO mTC 35 ($|r_s|$= 0.59), and TCGA mTC 58 ($|r_s|$= 0.55), indicating that this mTC is also captured in both cell line datasets and the two patient-derived datasets. All four of these highly correlated mTCs were enriched for genes involved in the metabolism of glutathione, cellular ketones and xenobiotics, and in drug detoxification (**Table S1**). Specifically,

genes in the aldo-keto reductase family were among the top-ranked genes in these mTCs. For both GEO mTC 35 and TCGA mTC 58, we observed high activity scores especially in non-small cell lung carcinoma samples (NSCLC; **Figure S5A** and **S5B**). In NSCLC, previous studies have reported a role for the glutathione system and drug detoxification/metabolism processes, and for the aldo-keto reductase family of genes in acquired drug resistance to platinum-based drugs (Chang et al., 2019; Tian et al., 2016; Yang et al., 2006; Zhu et al., 2018). In contrast, we observed that an increased activity score of GDSC mTC 18 was associated with a decrease in IC_{50} value (i.e., increased sensitivity) for only seven of the 142 drugs ($|r_s|$ range 0.20-0.41; **Figure 5A**). The drug with the highest negative correlation was tanespimycin (17-AAG), an Hsp90 inhibitor ($|r_s| = 0.41$). A direct link between the functions of glutathione and Hsp90 in oxidative stress has been suggested, as well as a relationship between tanespimycin sensitivity and *NQO1* expression, a gene coding for an enzyme reducing quinones to hydroquinones that is involved in detoxification pathways (Gaspar et al., 2009; Kim et al., 2015). In line with these findings, we found that the *NQO1* gene is present near the top of GDSC mTC 18.

As a final example, in the GDSC dataset in our study the activity score of GDSC mTC 108 was associated with the IC_{50} values of MEK inhibitor trametinib and histone deacetylase inhibitor vorinostat ($|r_s| = 0.48$ and $|r_s| = 0.46$, respectively; **Figure 5B** and **Table S4**). An increase in the activity score of GDSC mTC 108 in a sample was associated with a lower IC_{50} value of trametinib. In contrast, an increased activity score of GDSC mTC 108 was associated with a higher IC_{50} value of vorinostat. GDSC mTC 108 was found to be enriched for genes involved in the regulation of the biosynthetic process of cytokines and negative regulation of the MAPK cascade (**Table S1**). GDSC mTC 108 was strongly correlated with CCLE mTC 97 ($|r_s| = 0.61$). Consistent with the observation

13

for GDSC mTC 108, in the CCLE dataset we found that increased activity of CCLE mTC 97 was associated with a lower IC_{50} value (i.e., increased sensitivity) to the MEK inhibitor PD-0325901 ($|r_s|$ = 0.24) and a higher IC_{50} value (i.e., increased resistance) to the histone deacetylase inhibitor panobinostat ($|r_s|$ = 0.43). This contrasting sensitivity for MEK and histone deacetylase inhibition is in line with data from a study that used *BRAF*-mutated melanoma cell lines. The authors showed that cell lines with acquired resistance to MEK inhibitors subsequently became sensitive to treatment with the histone deacetylase inhibitor vorinostat (Wang et al., 2018). They concluded that the MEK-inhibitor resistance mechanism results from the activation (or reactivation) of MAPK cascades (Wagle et al., 2014). These findings are in line with our observation that both GDSC mTC 108 and CCLE mTC 97 were enriched for genes involved in the negative regulation of the MAPK cascade. These examples demonstrate that metabolic processes are associated with drug sensitivity.

The activity of mTCs are associated with the immune composition of the tumor microenvironment

We determined the association between the activity of mTCs and the immune composition of the tumor microenvironment (Table S5; see Methods for details). The immune composition for all samples in the GEO and TCGA dataset was determined by inferring fractions of 22 immune cell types using the CIBERSORT algorithm (Chen et al., 2018). We observed that the mTCs that were correlated with immune cell fractions could be divided into two groups. The first group included mTCs that were only identified in the patient-derived datasets. The second group contained mTCs that were identified in both the patient-derived and the cell line datasets.

14

For example, in the first group of mTCs that were correlated with immune cell fractions, the activity score of GEO mTC 123 was associated with CD8+ T-cell ($|r_s|$ = 0.40), γδ T-cell ($|r_s|$ = 0.36), activated CD4 memory T-cell ($|r_s|$ = 0.34), and regulatory T-cell ($|r_s|$ = 0.32) fractions (**Figure 5C**). GEO mTC 123 was correlated with TCGA mTC 34 ($|r_s|$= 0.33). In line with this, the activity score of TCGA mTC 34 was also associated with CD8+ T-cell fractions ($|r_s|$ = 0.58). Both GEO mTC 123 and TCGA mTC 34 showed enrichment for genes involved in immunological processes such as leukocyte activation and cytokine metabolism, as well as in metabolic processes such as phosphatidylinositol and phospholipid metabolism (**Table S1**). For both GEO mTC 123 and TCGA mTC 34, no strong correlation with the mTCs identified in the CCLE and GDSC cell line datasets was found. This suggests that GEO mTC 123 and TCGA mTC 34 both originated from non-cancerous cells in the immune tumor microenvironment and that their activity scores correlate especially with the presence of CD8+ T-cells.

GEO mTC 14 is illustrative of the second group of mTCs that were correlated with immune cell fractions. The activity scores of GEO mTC 14 were correlated with the fractions of M1 macrophages ($|r_s|$= 0.65) and M2 macrophages ($|r_s|$= 0.59; **Figure 5D**). GEO mTC 14 was correlated with TCGA mTC 70 ($|r_s|$= 0.44) and with CCLE mTC 118 ($|r_s|$= 0.21), and GDSC mTC 33 ($|r_s|$= 0.33). All four mTCs were enriched for genes involved in the metabolism of extracellular macromolecules (**Tables S1**). Genes coding for several types of collagens were among the top-ranked in these mTCs. This is in line with previous reports indicating that macrophages can function as collagen-producing cells in the tumor microenvironment (Schnoor et al., 2008; Vaage and Harlos, 1991). GEO mTC 14 and TCGA mTC 70 showed a high activity score in subsets of breast cancers, lung cancers, and sarcomas (**Figure S5C** and **S5D**). A negative activity score of GEO mTC

14 and TCGA mTC 70 was observed in a subset of hematological cancers, as well as in hematological cancer cell lines in both GDSC and CCLE mTCs. These mTCs are present in both patient data sets and cell-line datasets, which indicates that these metabolic processes are likely intrinsic to cancer cells and are associated with the fraction of macrophages present in the tumor microenvironment.

By correlating inferred immune cell fractions of samples with the activity scores of mTCs in samples, the relationship between metabolic gene expression and the various components of the immune tumor microenvironment could be assessed.

DISCUSSION

In the present study, we used consensus-Independent Component Analysis (c-ICA) in combination with Gene Set Enrichment Analysis (GSEA) to identify a broad set of robust metabolic Transcriptional Components (mTCs). With these mTCs, the transcriptional metabolic landscape was defined in patient-derived cancer tissue, cancer cell lines, and non-cancer samples. We also showed how this metabolic landscape can be used to explore associations between metabolic processes and drug sensitivities, patient outcomes, and the composition of the immune tumor microenvironment.

We used the wealth of publicly available pan-cancer transcriptomic data to study human metabolism on a large scale. Previous work used either single-cell sequencing or bulk cell transcriptomic profiles to study metabolism in specific cancer types (Hakimi et al., 2016; Xiao et al., 2019), or pan-cancer, but based on a single platform (Cubuk et al., 2018; Peng et al., 2018a; Rosario et al., 2018). Our present study differs from this previous work in two essential aspects.

16

Firstly, we used c-ICA to segregate the average expression patterns of complex biopsies into statistically independent components (Bhattacharya et al., 2020; Biton et al., 2014; Kong et al., 2008). Metabolic processes relevant to cancer biology can be challenging to detect when their transcriptional footprints (TFs) are subtle and concealed by more pronounced TFs from other biological processes and experimental artifacts. Using c-ICA, both the pronounced and more subtle TFs of metabolic processes can be identified. Moreover, c-ICA provided us with information on the activities of these metabolic processes. This enabled us to determine the correlation between the activity of pronounced and subtle metabolic processes with drug sensitivities, patient outcomes, and the composition of the immune tumor microenvironment. Secondly, the present study is the most extensive transcriptional analysis of metabolism and the first that integrated patient-derived data from GEO and TCGA with cell-line data from CCLE and GDSC. The samples in these four datasets were obtained from a multitude of independently constructed, publicly available cohorts, and the expression profiles were generated using different technologies (microarray or RNA-sequencing). This integrated dataset enabled us to demonstrate that most of the identified mTCs were robust and independent from dataset-specific and platform-specific characteristics. The observed overlap, or lack of overlap, between patient-derived and cell line-derived mTCs can help researchers understand how metabolic genes and pathways that are identified in cell lines can be translated to a patient tissue-context and vice versa. Furthermore, we hypothesize that metabolic processes that are identified only in patient-derived samples and not in cell line samples are more likely to originate from cells in the tumor microenvironment. These microenvironment-specific metabolic processes will not be captured

17

by mTCs in cell line datasets. This is because bulk expression profiles of cancer cell line samples do not harbor transcriptional footprints associated with non-tumor cells.

The metabolic landscape enabled us to classify samples based on the transcriptional activity of metabolic processes, resulting in metabolic subtypes. However, this metabolic classification was often not in alignment with current classification systems based on aspects such as histotype. For breast cancer and melanoma, we demonstrated that metabolic subtypes were associated with disease outcomes, emphasizing the relevance of metabolic pathway-based classification in cancer. The heterogeneity (metabolic and otherwise) within and between cancer types is well recognized, and alternative subtyping based on metabolite profiling and the metabolic transcriptome have been proposed before (Reznik et al., 2018; Rosario et al., 2018; Tang et al., 2014). More specifically, clinically significant, metabolism-based classifications have been proposed in breast cancer (Cappelletti et al., 2017; Serrano-Carbajal et al., 2020; Wang et al., 2019) and melanoma (Fischer et al., 2018; Wind et al., 2018). The most active mTCs in a clinically relevant metabolic subtype could thus be used to generate new hypotheses for treatment targets. Additionally, the association between the activity of mTCs and drug sensitivity could help to design therapeutic strategies, personalized or otherwise.

Metabolic heterogeneity and plasticity are not limited to cancer cells but are also applicable to the immune cells present in the tumor micro-environment. Immune cells undergo metabolic changes when activated, and their metabolic status can overlap with the metabolic state of cancer cells (Andrejeva and Rathmell, 2017). For example, the Warburg effect is classically seen as an example of a metabolic transformation in cancer cells. However, it is also observed in activated T cells (Bantug et al., 2018; Patel and Powell, 2017; Wang and Green, 2012). In the

context of metabolism, this complex interplay between cancer cells and immune cells present in the micro-environment gives a new dimension to the use of drugs that target metabolic processes (O'Sullivan et al., 2019; Patel et al., 2019). For instance, inhibiting glutamine metabolism has been shown to inhibit tumor growth and increase the sensitivity of triple-negative breast cancers to immune checkpoint blockade (Oh et al., 2020), and reducing oxidative stress has been shown to prevent the generation of tumor-associated macrophages (Zhang et al., 2013). Furthermore, modulating metabolism in T-cells from glycolytic towards an OXPHOS-weighted profile has been shown to improve CAR T cell immunotherapy (Fraietta et al., 2018; O'Sullivan and Pearce, 2015; Sukumar et al., 2017). Our transcriptional metabolic landscape can contribute to knowledge on immunometabolism and, combined with the association of mTCs with drug sensitivity, also contribute to the formulation of new hypotheses on how to metabolically engage the tumor and its immune microenvironment, thus improving the response to immunotherapy.

Further research to gain an even more comprehensive understanding of metabolism in patient-derived cancer samples should ideally integrate genomics, transcriptomics, proteomics, and metabolomics to capture the complexity of metabolic processes within cancer cells (Buescher and Driggers, 2016). Recent initiatives to this end are the Recon1, Edinburgh Human Metabolic Network (EHMN), and Human1 projects (Brunk et al., 2018; Ma et al., 2007; Robinson et al., 2020). However, challenges for these initiatives lie in the limited set of samples for which these high-dimensional multi-omics features are available, and the use of predominantly cell line samples. Paired datasets on a large scale are needed to unleash the full potential of such an integrated approach.

In the present study we demonstrated several examples of how metabolic processes were associated with the composition of the immune tumor microenvironment, drug sensitivity, and patient outcomes. To facilitate the use of our transcriptional metabolic landscape, we have provided access to all data via a web portal (www.themetaboliclandscapeofcancer.com). In this portal, users can explore genes, metabolic processes, and tissue types of interest. We invite researchers and clinicians to use this portal as a guide to the metabolic transcriptome in cancer, or as a starting point for further research into cancer metabolism.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

R.S.N.F, V.C.L., C.G.U, and A.B collected and compiled the data. Data analyses were performed by R.S.N.F., V.C.L., and C.G.U. All authors contributed to the data interpretation, writing of the manuscript, and the final decision to submit the manuscript.

DECLARATION OF INTERESTS

All authors declare no competing interests.

FIGURE TITLES AND LEGENDS

Figure 1 – Identification of metabolic transcriptional components (mTCs). mTCs are highly concordant across different datasets and platforms. **(A)** Workflow for identification of mTCs. Consensus-Independent Component Analysis (c-ICA) on 34,494 samples from four datasets was used to identify transcriptional components (TCs). Subsequent systematic selection of TCs enriched for metabolic processes resulted in in 132, 151, 136, and 137 mTCs for the GEO, TCGA, CCLE, and GDSC datasets, respectively. **(B)** Hierarchically clustered heatmaps showing the enrichment of the 608 metabolic gene sets for the mTCs identified in GEO, TCGA, CCLE, and GDSC datasets. **(C)** Citrus plots showing the absolute Spearman correlations of mTCs identified in one dataset with mTCs identified in the other three datasets. Colored lines indicate $|r_s| > 0.5$. The fraction and percentage of mTCs of a dataset having at least one $|r_s| > 0.5$ are depicted for each dataset. **(D)** Scatter plot showing the correlations between all GEO mTCs and TCGA mTCs (x-axis), versus their Z-transformed P values (y-axis). Colored dots show the correlations > 0.5 having a P value < 0.05. **(E)** Scatter plot illustrating the highest absolute correlation coefficient $|r_s|$ of a GEO mTC to a TCGA mTC (x-axis), versus the highest enrichment score for a metabolic gene set found for that GEO mTC (y-axis).

Figure 2 - For GEO and TCGA datasets, genes with an absolute weight > 3 in at least one mTC were clustered hierarchically. **(A)** Heatmap of hierarchically clustered genes with an absolute weight > 3 in GEO mTCs. Sets of top genes in GEO mTC 54 and GEO mTC 11, as highlighted in the

21

text, are enlarged in black boxes. The blue and red colors in the heatmap designate the sign and magnitude of each gene weight, as given in the legend. However, interpreting the sign of a gene weight only has value when the activity score of the respective mTC in a sample is taken into account. This activity score, too, can be either positive or negative. Only the product of a gene weight and activity score can therefore be interpreted, and the sign of this product will give an indication of a gene from a mTC being "more active" or "less active" in a sample than the average. **(B-C)** Top genes in GEO mTC 54 and TCGA mTC 127. Text colored white shows genes that are a member of the 608 metabolic gene sets. Lines signify genes that are present in the top of both GEO and TCGA mTCs. **(D-E)** Top genes in GEO mTC 11 and TCGA mTC 141. **(F)** Heatmap of hierarchically clustered genes with an absolute weight > 3 in TCGA mTCs. Sets of top genes in TCGA mTC 127 and TCGA mTC 141, as highlighted in the text, are enlarged in black boxes.

Figure 3 – Clustering activity scores of mTCs reveal multiple metabolic subtypes, which show prognostic value

(A) 21,592 GEO samples were hierarchically clustered based on mTC activity scores and divided into 67 metabolic subtypes. **(B)** 10,817 TCGA samples were hierarchically clustered based on mTC activity scores and divided into 58 metabolic subtypes. **(C)** Metabolic landscape of the subset of breast tissue samples in the GEO dataset. **(D)** Metabolic landscape of the subset of melanoma samples in the TCGA dataset. **(E)** Retrospective distant relapse-free survival of breast cancer patients in the GEO dataset. Patient-derived samples were stratified per metabolic subtype. Kaplan Meier curves are shown with a confidence interval of 0.95. **(F)** Retrospective overall

survival of melanoma patients in the TCGA dataset. Patient-derived samples were stratified per metabolic subtype. Kaplan Meier curves are plotted with a confidence interval of 0.95.

Figure 4 – Activity of GDSC mTC 3 is highly correlated with the IC50 value of Nutlin-3a and is active in cell lines with wildtype *TP53*. (A) Spearman correlations between drug IC50 values and the activity of GDSC mTC 3 **(B)** Scatter plot showing the association between the IC50 value of Nutlin-3a and activity of GDSC mTC 3. **(C)** Highest pair-wise correlations between GDSC mTC 3 and mTCs from GEO, TCGA and CCLE datasets. **(D)** Box plot of activity of GDSC mTC 3 across cell lines, colored for their *TP53* mutation status. **(E)** Box plot of activity of CCLE mTC 4 across cell lines, colored for their *TP53* mutation status.

Figure 5 – Associations between mTCs, drug sensitivities, and the composition of the immune tumor microenvironment for selected examples. (A-B) Spearman correlations between drug IC50 values and the activity of GDSC mTC 18 and mTC 108. **(C-D)** Spearman correlations between CIBERSORT estimated immune cell fractions and the activity of GEO mTC 123 and mTC 14.

METHODS

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Rudolf S.N. Fehrmann (r.s.n.fehrmann@umcg.nl).

Materials availability

N.A.

Data and code availability

Data can be explored at http://themetaboliclandscapeofcancer.com. Code is available at github.com/MetabolicLandscape/

Method details

Data acquisition, data preprocessing and c-ICA

A detailed description of the methods are available in the **Supplementary Note**. The methods for data acquisition, preprocessing of the four datasets GEO, TCGA, GDSC, and CCLE and c-ICA was published previously (Bhattacharya et al., 2020). These methods are summarized below.

ICA was performed on a preprocessed and whitened dataset using the FastICA algorithm, resulting in the extraction of estimated sources (ESs) and a mixing matrix. The number of principal components which captured 90% of the variance seen in the whitened dataset was chosen as the number of ESs to extract. We performed 25 ICA runs with different random initialization weight factors to assess the robustness of the ESs and exclude ICA results derived from convergence at local solutions. ESs extracted from these runs were clustered together if the absolute value of the Pearson correlation between them was > 0.9. We considered a consensus TC robust when clustering included individual TCs from > 50% of the runs. The consensus TCs, in combination with the original input expression profiles, were used to obtain the consensus mixing matrix with the individual activity scores of the consensus TCs in each sample via matrix inversion.

Identification of TCs enriched for metabolic processes

Gene sets defining metabolic process were selected from gene set collections obtained from the Molecular Signatures Database (MSigDb version 6.1): BioCarta, Gene Ontology – Biological Process (GO-BP), Gene Ontology – Molecular Function (GO-MF), KEGG, and Reactome (See **Supplemental Information** for details on selection process). To identify transcriptional components enriched for metabolic processes, gene set enrichment analysis (GSEA) was performed using the selected metabolic gene sets. Enrichment of each metabolic gene set was tested according to the two-sample Welch's t-test for unequal variance between the metabolic set of genes, and the Welch's t statistic was transformed to a Z-score to allow comparison between gene sets.

To reduce the redundancy in gene sets from different gene set collections, consensus clustering was performed set-wise on the GSEA data for the GEO, TCGA, CCLE, and GDSC datasets. Consensus clustering was performed using the ConsensusClusterPlus-package (v1.51.1) within R, using the default hierarchical clustering algorithm and Pearson correlation distance, a maximum number of clusters (maxK) of 150, 2000 resamplings (reps), with 80% row and 80% column resampling (pFeature and pItem, respectively). The optimal number of clusters (k) was determined as the k at which the relative change in area under the CDF curve was minimized (<0.01). This resulted in a k of 50 clusters (**Figure S1**).

The 50 clusters of gene sets were subsequently used to select transcriptional components based on their enrichment for metabolic processes. Per gene set cluster, the three TCs with the highest absolute enrichment score for any gene set in that cluster were selected. We also selected

25

the three TCs with the highest absolute mean enrichment score for all gene sets in that cluster. These selected TCs were then referred to as metabolic Transcriptional Components (mTCs).

Clustering of Metabolic Transcriptional Components, Genes and Samples

Hierarchical clustering of mTCs, genes, and samples was performed using ward.D2 as the method and 1-cor(data) as the distance function. Heatmaps were created using R's *gplots* package (v3.0.1).

CIBERSORT

Relative and absolute immune fractions for 22 immune cell types were estimated for all samples in GEO and TCGA datasets using the CIBERSORT algorithm running with default parameters, 1000 permutations, and selecting 'nosumto1' as output. This output was then associated with the activity of the mTCs in samples, through Spearman correlation.

Statistical Analyses

Univariate DRFS on breast cancer samples from GEO and univariate OS analyses on melanoma samples from TCGA were performed using a Cox regression model through *survminer* (v0.4.3) and *survival* (v2.43-3) packages in R. Confidence intervals were set at 0.95. Significance was tested through the Log Rank test. Scripts are available at github.com/MetabolicLandscape/. Pearson correlations were performed in R using the cor.test()-function from the *stats* package (v.3.5.1). Spearman correlations and the corresponding exact P values were calculated using the *pspearman*-package (v0.3-0) in R, with a t-distribution as an approximation.

SUPPLEMENTAL INFORMATION TITLES AND LEGENDS

Figure S1 – Related to Figure 1;

(A) Consensus clustering gene set enrichment scores of all TCs in the GEO dataset. Consensus matrix for a k of 50 gene set clusters. **(B)** Consensus clustering gene set enrichment scores of all TCs in the GEO dataset. Relative change in area under the consensus CDF curve with increasing k.

(C) Histograms quantifying the gene-based pair-wise correlations between mTCs from each dataset with an $|r_s| > 0.5$ and a P value < 0.05

(D) The activity of GEO mTC 119 in multiple myeloma, other cancer and non-cancer samples in the GEO dataset. **(E)** Only very low correlations were found between GEO mTC 119 and mTCs from either TCGA, CCLE or GDSC datasets.

Figure S2 – Related to Figure 3;

(A) Metabolic landscape for CCLE samples. The 1,067 samples were hierarchically clustered and divided into 38 metabolic subtypes. **(B)** Metabolic landscape for GDSC samples. The 1,018 samples were hierarchically clustered and divided into 36 clusters metabolic subtypes. Grey labels designate tissue types that are present in other datasets, but are not present in the given dataset;

(C) Hierarchical clustering of samples based on activity scores of mTCs in GEO and TCGA datasets in order to define metabolic subtypes. The plot shows the minimum sample size of a cluster depending on the chosen cutoff height of the dendrogram resulting from hierarchical clustering. The heights at which the minimum cluster size reaches 50 is given for both GEO and TCGA datasets. **(D)** Hierarchical clustering of samples based on activity scores of mTCs in CCLE and GDSC datasets in order to define metabolic subtypes. The plot shows the minimum sample size of a

cluster depending on the chosen cutoff height of the dendrogram resulting from hierarchical clustering. The heights at which the minimum cluster size reaches 50 is given for both CCLE and GDSC datasets;

(E) Activity of GEO mTC 50 in breast cancer samples with survival data, grouped by metabolic subtype. The selected mTC was enriched for oxidative phosphorylation. **(F)** Activity of TCGA mTC 7 in melanoma samples with survival data, grouped by metabolic subtype. The selected mTC was enriched for genes involved in the regulation of the metabolic processes of glycoproteins and phospholipids, cell-cell adhesion, and the activation of lymphocytes and leukocytes.

Figure S3 – Related to Figure 3; Pie graphs depicting the tissue type composition of the 67 metabolic subtypes defined for the GEO dataset.

Figure S4 – Related to Figure 3; Pie graphs depicting the tissue type composition of the 58 metabolic subtypes defined for the TCGA dataset.

Figure S5 - Related to Figures 4 and 5; (A) Activity of GEO mTC 35 in samples, grouped per tissue type. Tissue types mentioned in the text are highlighted on the x-axis. **(B)** Activity of TCGA mTC 58 in samples, grouped per tissue type. Tissue types mentioned in the text are highlighted on the x-axis. **(C)** Activity of GEO mTC 14 in samples, grouped per tissue type. Tissue types with a higher median activity highlighted in the text are given a red axis label, tissue types with a lower median activity highlighted in the text are given a blue axis label. **(D)** Activity of TCGA mTC 70 in samples, grouped per tissue type. Tissue types with a higher median activity highlighted in the text are

28

given a red axis label, tissue types with a lower median activity highlighted in the text are given a blue axis label.

Table S1 – Related to Figure 1; This file contains five sheets: (1) List of 608 metabolic gene sets, selected in a systematic way, divided into 50 gene set clusters. These gene set clusters resulted from consensus clustering the gene set enrichment scores of TCs. (2-5) Gene set enrichment scores for the 608 metabolic gene sets for all GEO, TCGA, CCLE and GDSC mTCs.

Table S2 – Related to Figure 3; This file contains four sheets: (1-2) Tissue type composition of metabolic subtypes as defined for the GEO and TCGA datasets. (3-4) The mean activity score of all GEO and TCGA mTCs in samples belonging to the metabolic subtypes as defined in the GEO an TCGA datasets, respectively.

Table S3 – Related to Figure 3; Clinicopathological parameters of breast cancer tissue samples with DRFS survival data from the GEO dataset, and of melanoma samples with OS data from the TCGA dataset, stratified per metabolic subtype.

Table S4 – Related to Figure 5; Spearman correlations between drug IC50 values and the activity of CCLE and GDSC mTCs in cell lines.

Table S5 – Related to Figure 5; Spearman correlations between CIBERSORT estimated immune cell fractions and the activity of GEO and TCGA mTCs in samples.

REFERENCES

Andrejeva, G., and Rathmell, J.C. (2017). Similarities and distinctions of cancer and immune metabolism in inflammation and tumors. Cell Metab *26*, 49-70.

Bantug, G.R., Galluzzi, L., Kroemer, G., and Hess, C. (2018). The spectrum of T cell metabolism in health and disease. Nat Rev Immunol *18*, 19-34.

Barret, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets - update. Nucleic Acids Res *41*, D991-D995.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, V., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603-607.

Benita, Y., Kikuchi, H., Smith, A.D., Zhang, M.Q., Chung, D.C., and Xavier, R.J. (2009). An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. Nucleic Acids Res *37*, 4587-4602.

Bense, R.D., Sotiriou, C., Piccart-Gebhart, M.J., Haanen, J.B.A.G., van Vugt, M.A.T.M., de Vries, E.G.E., Schröder, C.P., and Fehrmann, R.S.N. (2017). Relevance of tumor-infiltrating immune cell composition and functionality for disease outcome in breast cancer. J Natl Cancer Inst *109*, djw192.

Bhattacharya, A., Bense, R.D., Urzúa-Traslaviña, C.G., de Vries, E.G.E., van Vugt, M., and Fehrmann, R.S.N. (2020). Transcriptional effects of copy number alterations in a large set of human cancers. Nat Commun *11*, 715.

Biton, A., Bernard-Pierrot, I., Lou, Y., Krucker, C., Chapeaublanc, E., Rubio-Pérez, C., López-Bigas, N., Kamoun, A., Neuzillet, Y., Gestraud, P., et al. (2014). Independent component analysis uncovers the landscape of the bladder tumor transcriptome and reveals insights into luminal and basal subtypes. Cell Rep *9*, 1235-1245.

Brunk, E., Sahoo, S., Zielinski, D.C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Gonzalez, G.A.P., Aurich, M.K., et al. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. Nat Biotechnol *36*, 272-281.

Buescher, J.M., and Driggers, E.M. (2016). Integration of omics: more than the sum of its parts. Cancer Metab *4*, 4.

Cantor, J.R., and Sabatini, D.M. (2012). Cancer cell metabolism: one hallmark, many faces. Cancer Discov *2*, 881-898.

Cappelletti, V., Iorio, E., Miodini, P., Silvestri, M., Dugo, M., and Daidone, M.G. (2017). Metabolic footprints and molecular subtypes in breast cancer. Dis Markers *2017*, 7687851.

Chang, W.M., Chang, Y.C., Yang, Y.C., Lin, S.K., Chang, P.M., and Hsiao, M. (2019). AKR1C1 controls cisplatin-resistance in head and neck squamous cell carcinoma through cross-talk with the STAT1/3 signaling pathway. J Exp Clin Cancer Res *38*, 245.

Chen, B., Khodadoust, M.S., Liu, C.L., Newman, A.M., and Alizadeh, A.A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. Methods Mol Biol *1711*, 243-259.

Copple, B.L., Bai, S., Burgoon, L.D., and Moon, J.O. (2012). Hypoxia-inducible Factor-1α regulates expression of genes in hypoxic hepatic stellate cells important for collagen deposition and angiogenesis. Liver int. *31*, 230-244.

Cubuk, C., Hidalgo, M.R., Amadoz, A., Pujana, M.A., Mateo, F., Herranz, C., Carbonell-Caballero, J., and Dopazo, J. (2018). Gene expression integration into pathway modules reveals a pan-cancer metabolic landscape. Cancer Res *78*, 6059-6072.

Desvergne, B., Michalik, L., and Wahli, W. (2006). Transcriptional regulation of metabolism. Physiol Rev *86*, 465-514.

Domcke, S., Sinha, R., Levine, D.A., Sander, C., and Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nat Commun *4*, 2126.

Fischer, G.M., Vashisht Gopal, Y.N., McQuade, J.L., Peng, W., DeBerardinis, R.J., and Davies, M.A. (2018). Metabolic strategies of melanoma cells: Mechanisms, interactions with the tumor microenvironment, and therapeutic implications. Pigment Cell Melanoma Res *31*, 11-30.

Fraietta, J.A., Lacey, S.F., Orlando, E.J., Pruteanu-Malinici, I., Gohil, M., Lundh, S., Boesteanu, A.C., Wang, Y., O'Connor, R.S., Hwang, W.T., et al. (2018). Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. Nat Med *24*, 563-571.

Gaspar, N., Sharp, S.Y., Pacey, S., Jones, C., Walton, M., Vassal, G., Eccles, S., Pearson, A., and Workman, P. (2009). Acquired resistance to 17-allylamino-17-demethoxygeldanamycin (17-AAG, tanespimycin) in glioblastoma cells. Cancer Res *69*, 1966-1975.

Ghaffari, P., Mardinoglu, A., and Nielsen, J. (2015). Cancer metabolism: a modeling perspective. Front Physiol *6*, 382.

Hakimi, A.A., Reznik, E., Lee, C.H., Creighton, C.J., Brannon, A.R., Luna, A., Aksoy, B.A., Liu, E.M., Shen, R., Lee, W., et al. (2016). An integrated metabolic atlas of clear cell renal cell carcinoma. Cancer Cell *29*, 104-116.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Herbel, C., Patsoukis, N., Bardhan, K., Seth, P., Weaver, J.D., and Boussiotis, V.A. (2016). Clinical significance of T cell metabolic reprogramming in cancer. Clin Transl Med *5*, 29.

Huang, M., Shen, A.J., Ding, J., and Geng, M.Y. (2014). Molecularly targeted cancer therapy: some lessons from the past decade. Trends Pharmacol Sci *35*, 41-50.

Hynds, R.E., Vladimirou, E., and Janes, S.M. (2018). The secret lives of cancer cell lines. Dis Model Mech *11*, 1-5.

Jiang, G.L., Zhang, S.J., Yazdanparast, A., Li, M., Pawar, A.V., Liu, Y.L., Inavolu, S.M., and Cheng, L.J. (2016). Comprehensive comparison of molecular portraits between cell lines and tumors in breast cancer. BMC Genomics *17*, 525.

Kim, Y.S., Seo, H.W., and Jung, G. (2015). Reactive oxygen species promote heat shock protein 90-mediated HBV capsid assembly. Biochem Biophys Res Commun *457*, 328-333.

Klijn, C., Durinck, S., Stawiski, E.W., Haverty, P.M., Jiang, Z.S., Liu, H.B., Degenhardt, J., Mayba, O., Gnad, F., Liu, J.F., et al. (2015). A comprehensive transcriptional portrait of human cancer cell lines. Nat Biotechnol *33*, 306-312.

Kong, W., Vanderburg, C.R., Gunshin, H., Rogers, J.T., and Huang, X. (2008). A review of independent component analysis application to microarray gene expression data. Biotechniques *45*, 501-520.

Le Bourgeois, T., Strauss, L., Aksoylar, H.-I., Daneshmandi, S., Seth, P., Patsoukis, N., and Boussiotis, V.A. (2018). Targeting T cell metabolism for improvement of cancer immunotherapy. Front Oncol *8*, 237.

Lefort, N., Yi, Z., Bowen, B., Glancy, B., De Filippis, E.A., Mapes, R., Hwang, H., Flynn, C.R., Willis, W.T., Civitarese, A., et al. (2009). Proteome profile of functional mitochondria from human skeletal muscle using one-dimensional gel electrophoresis and HPLC-ESI-MS/MS. J Proteomics *72*, 1046-1060.

Liu, X., Mo, W., Ye, J., Li, L.Y., Zhang, Y.P., Hsueh, E.C., Hoft, D.F., and Peng, G.Y. (2018). Regulatory T cells trigger effector T cell DNA damage and senescence caused by metabolic competition. Nat Commun *9,* 249.

Luengo, A., Gui, D.Y., and Vander Heiden, M.G. (2017). Targeting metabolism for cancer therapy. Cell Chem Biol *24*, 1161-1180.

Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., and Goryanin, I. (2007). The Edinburgh human metabolic network reconstruction and its functional analysis. Mol Syst Biol *3*, 135.

Martín-Martín, N., Carracedo, A., and Torrano, V. (2018). Metabolism and transcription in cancer: merging two classic tales. Front Cell Dev Biol *5*, 119.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat Methods *12*, 453-457.

O'Sullivan, D., and Pearce, E.L. (2015). Targeting T cell metabolism for therapy. Trends Immunol *36*, 71-80.

O'Sullivan, D., Sanin, D.E., Pearce, E.J., and Pearce, E.L. (2019). Metabolic interventions in the immune response to cancer. Nat Rev Immunol *19*, 324-335.

Oh, M.H., Sun, I.H., Zhao, L., Leone, R.D., Sun, I.M., Xu, W., Collins, S.L., Tam, A.J., Blosser, R.L., Patel, C.H., et al. (2020). Targeting glutamine metabolism enhances tumor specific immunity by modulating suppressive myeloid cells. J Clin Invest *130*, 3865-3884.

Patel, C.H., Leone, R.D., Horton, M.R., and Powell, J.D. (2019). Targeting metabolism to regulate immune responses in autoimmunity and cancer. Nat Rev Drug Discov *18*, 669-688.

Patel, C.H., and Powell, J.D. (2017). Targeting T cell metabolism to regulate T cell activation, differentiation and function in disease. Curr Opin Immunol *46*, 82-88.

Pavlova, N.N., and Thompson, C.B. (2016). The Emerging Hallmarks of Cancer Metabolism. Cell Metab *23*, 27-47.

Peng, X., Chen, Z., Farshidfar, F., Xu, X., Lorenzi, P.L., Wang, Y., Cheng, F., Tan, L., Mojumdar, K., Du, D., et al. (2018). Molecular characterization and clinical relevance of metabolic expression subtypes in human cancers. Cell Rep *23*, 255-269.

Quail, D.F., and Joyce, J.A. (2013). Microenvironmental regulation of tumor progression and metastasis. Nat Med *19*, 1423-1437.

Reznik, E., Luna, A., Aksoy, B.A., Liu, E.M., La, K., Ostrovnaya, I., Creighton, C.J., Hakimi, A.A., and Sander, C. (2018). A landscape of metabolic variation across tumor types. Cell Syst *6*, 301-313 e303.

Robinson, J.L., Kocabaş, P., Wang, H., Cholley, P.E., Cook, D., Nilsson, A., Anton, M., Ferreira, R., Domenzain, I., Billa, V., et al. (2020). An atlas of human metabolism. Sci Signal *13*, eeaz1482.

Rosario, S.R., Long, M.D., Affronti, H.C., Rowsam, A.M., Eng, K.H., and Smiraglia, D.J. (2018). Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. Nat Commun *9*, 5330.

Schnoor, M., Cullen, P., Lorkowski, J., Stolle, K., Robenek, H., Troyer, D., Rauterberg, J., and Lorkowski, S. (2008). Production of type VI collagen by human macrophages: a new dimension in macrophage functional heterogeneity. J Immunol *180*, 5707-5719.

Serrano-Carbajal, E.A., Espinal-Enríquez, J., and Hernández-Lemus, E. (2020). Targeting metabolic deregulation landscapes in breast cancer subtypes. Front Oncol *10*, 97.

Sørensen, B.S., Knudsen, A., Wittrup, C.F., Nielsen, S., Aggerholm-Pedersen, N., Busk, M., Horsman, M., Høyer, M., Bouchelouche, P.N., Overgaard, J., et al. (2015). The usability of a 15-gene hypoxia classifier as a universal hypoxia profile in various cancer cell types. Radiother Oncol *116*, 346-351.

Sukumar, M., Kishton, R.J., and Restifo, N.P. (2017). Metabolic reprograming of anti-tumor immunity. Curr Opin Immunol *46*, 14-22.

Tang, X., Lin, C.C., Spasojevic, I., Iversen, E.S., Chi, J.T., and Marks, J.R. (2014). A joint analysis of metabolomics and genetics of breast cancer. Breast Cancer Res *16*, 415.

Tian, H., Li, X., Jiang, W., Lv, C., Sun, W., Huang, C., and Chen, R. (2016). High expression of AKR1C1 is associated with proliferation and migration of small-cell lung cancer cells. Lung Cancer (Auckl) *7*, 53-61.

Vaage, J., and Harlos, J.P. (1991). Collagen production by macrophages in tumour encapsulation and dormancy. Br J Cancer *63*, 758-762.

Vazquez, A., Kamphorst, J.J., Markert, E., Schug, Z.T., Tardito, S., and Gottlieb, E. (2016). Cancer metabolism at a glance. J Cell Sci *129*, 3367-3373.

Viale, A., and Draetta, G.F. (2016). Metabolic features of cancer treatment resistance. Recent Results Cancer Res *207*, 135-156.

Vincent, K.M., and Postovit, L.M. (2017). Investigating the utility of human melanoma cell lines as tumour models. Oncotarget *8*, 10498-10509.

Wagle, N., Van Allen, E.M., Treacy, D.J., Frederick, D.T., Cooper, Z.A., Taylor-Weiner, A., Rosenberg, M., Goetz, E.M., Sullivan, R.J., Farlow, D.N., et al. (2014). MAP kinase pathway alterations in BRAF-mutant melanoma patients with acquired resistance to combined RAF/MEK inhibition. Cancer Discov *4*, 61-68.

Walling, J. (2006). From methotrexate to pemetrexed and beyond. A review of the pharmacodynamic and clinical properties of antifolates. Invest New Drug *24*, 37-77.

Wang, L., Leite de Oliveira, R., Huijberts, S., Bosdriesz, E., Pencheva, N., Brunen, D., Bosma, A., Song, J.Y., Zevenhoven, J., Los-de Vries, G.T., et al. (2018). An acquired vulnerability of drug-resistant melanoma with therapeutic potential. Cell *173*, 1413-1425 e1414.

Wang, R., and Green, D.R. (2012). Metabolic checkpoints in activated T cells. Nat Immunol *13*, 907-915.

Wang, R., Zhao, H., Zhang, X., Zhao, X., Song, Z., and Ouyang, J. (2019). Metabolic discrimination of breast cancer subtypes at the single-cell level by multiple microextraction coupled with mass spectrometry. Anal Chem *91*, 3667-3674.

Ward, P.S., and Thompson, C.B. (2012). Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. Cancer Cell *21*, 298-308.

Wind, T., Jalving, M., de Haan, J.J., De Vries, E.G.E., van Vugt, M.A.T.M., Reijngoud, D.J., van Rijn, R.S., Haanen, J.B.A.G., Blank, C.U., Hospers, G.A.P., et al. (2018). A large pooled analysis refines gene expression-based molecular subclasses in cutaneous melanoma. Oncoimmunology *8*, 1558664.

Xiao, Z., Dai, Z., and Locasale, J.W. (2019). Metabolic landscape of the tumor microenvironment at single cell resolution. Nat Commun *10*, 3763.

Yang, P., Ebbert, J.O., Sun, Z., and Weinshilboum, R.M. (2006). Role of the glutathione metabolic pathway in lung cancer treatment and prognosis: a review. J Clin Oncol *24*, 1761-1769.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res *41*, D955-961.

Ye, I.C., Fertig, E.J., DiGiacomo, J.W., Considine, M., Godet, I., and Gilkes, D.M. (2018). Molecular portrait of hypoxia in breast cancer: a prognostic signature and novel HIF-regulated genes. Mol Cancer Res *16*, 1889-1901.

Zhang, Y., Choksi, S., Chen, K., Pobezinskaya, Y., Linnoila, I., and Liu, Z.G. (2013). ROS play a critical role in the differentiation of alternatively activated macrophages and the occurrence of tumor-associated macrophages. Cell Res *23*, 898-914.

Zhu, H., Chang, L.L., Yan, F.J., Hu, Y., Zeng, C.M., Zhou, T.Y., Yuan, T., Ying, M.D., Cao, J., He, Q.J., et al. (2018). AKR1C1 activates STAT3 to promote the metastasis of non-small cell lung cancer. Theranostics *8*, 676-692.
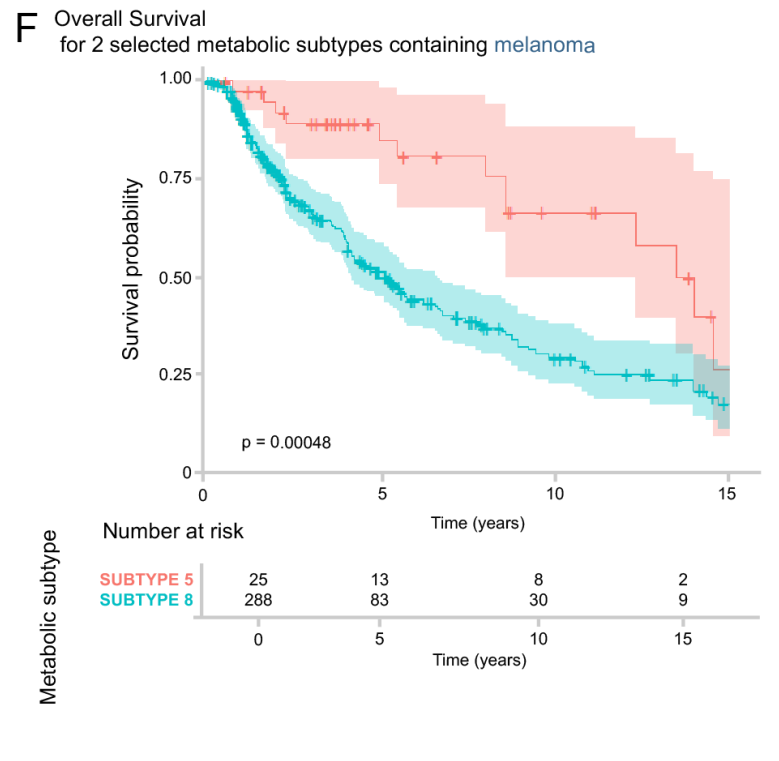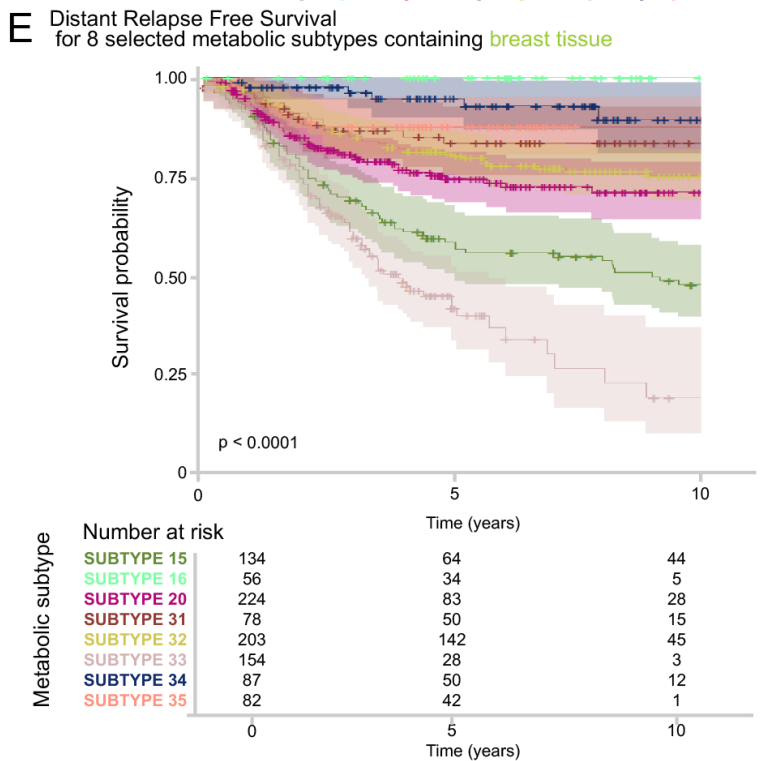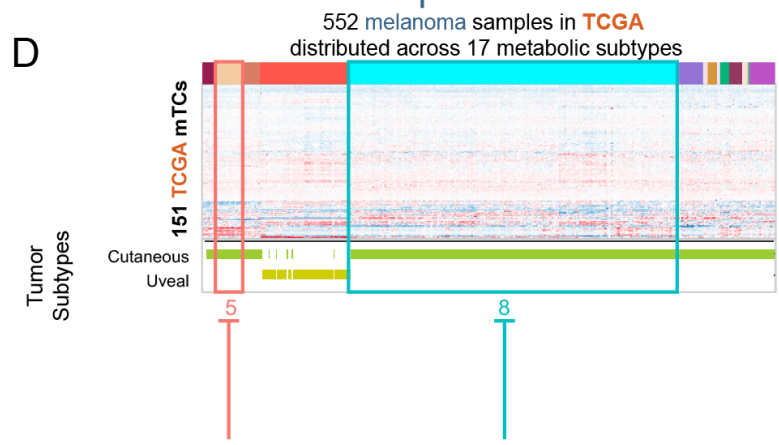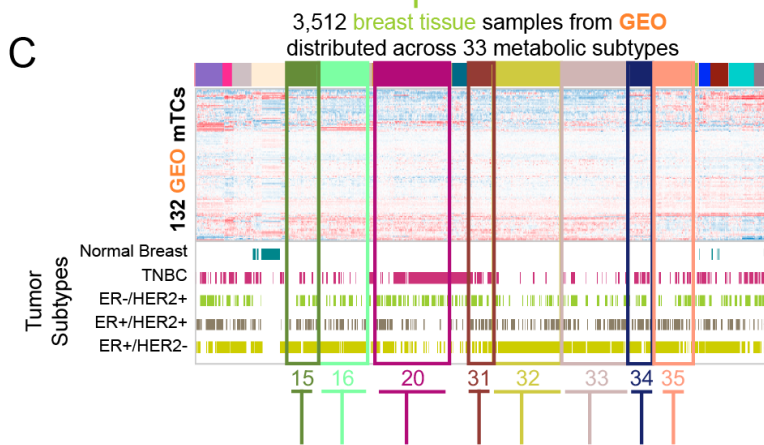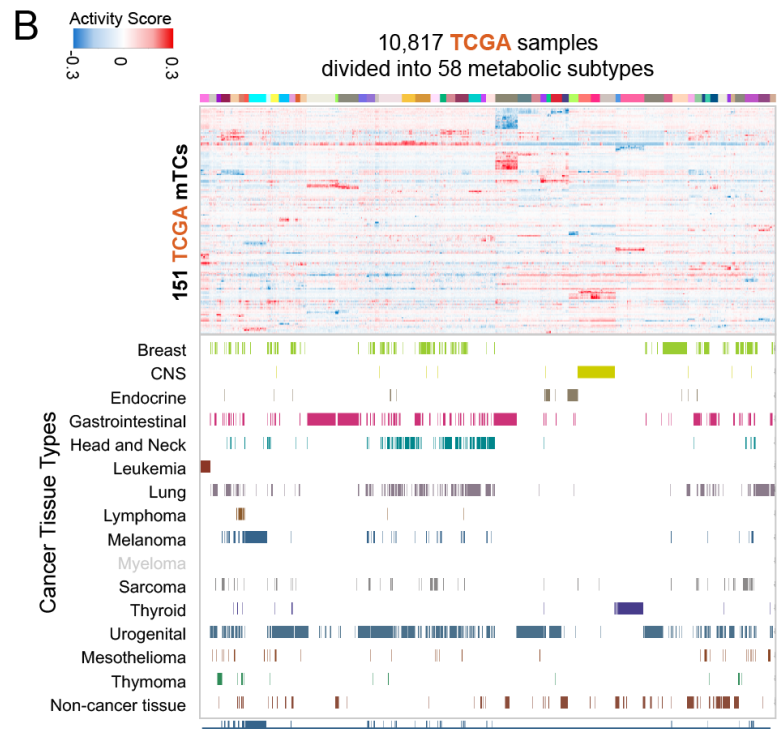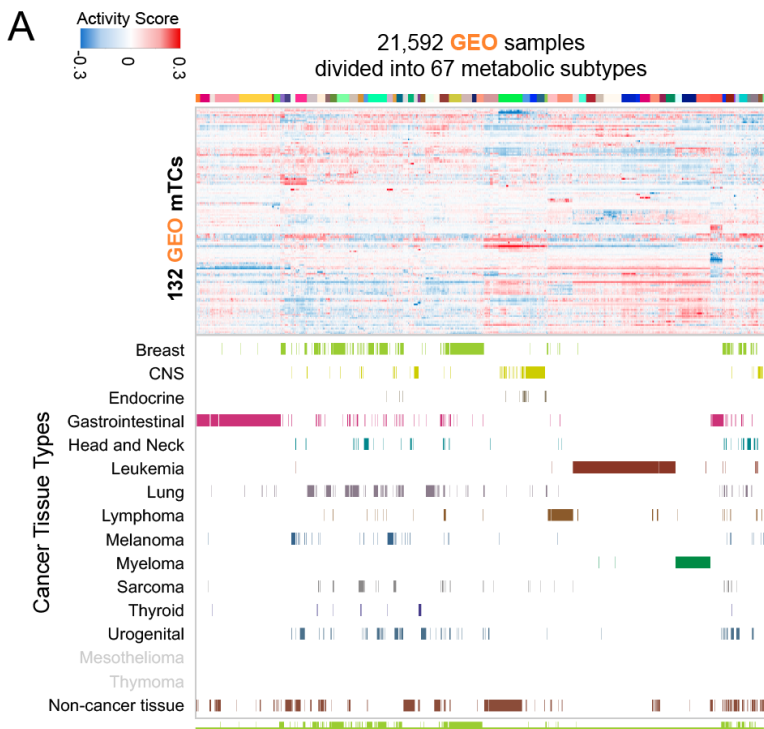
Figure 1 – **Identification of metabolic transcriptional components (mTCs). mTCs are highly concordant across different datasets and platforms**. **(A)** Workflow for identification of mTCs. Consensus-Independent Component Analysis (c-ICA) on 34,494 samples from four datasets was used to identify transcriptional components (TCs). Subsequent systematic selection of TCs enriched for metabolic processes resulted in in 132, 151, 136, and 137 mTCs for the GEO, TCGA, CCLE, and GDSC datasets, respectively. **(B)** Hierarchically clustered heatmaps showing the enrichment of the 608 metabolic gene sets for the mTCs identified in GEO, TCGA, CCLE, and GDSC datasets. **(C)** Citrus plots showing the absolute Spearman correlations of mTCs identified in one dataset with mTCs identified in the other three datasets. Colored lines indicate $|r_s| > 0.5$. The fraction and percentage of mTCs of a dataset having at least one $|r_s| > 0.5$ are depicted for each dataset. **(D)** Scatter plot showing the correlations between all GEO mTCs and TCGA mTCs (x-axis), versus their Z-transformed P values (y-axis). Colored dots show the correlations > 0.5 having a P value < 0.05. **(E)** Scatter plot illustrating the highest absolute correlation coefficient $|r_s|$ of a GEO mTC to a TCGA mTC (x-axis), versus the highest enrichment score for a metabolic gene set found for that GEO mTC (y-axis).

Figure 2 - **For GEO and TCGA datasets, genes with an absolute weight > 3 in at least one mTC were clustered hierarchically**. **(A)** Hierarchically clustered heatmap of genes with an absolute weight > 3 in GEO mTCs. Sets of top genes in GEO mTC 54 and GEO mTC 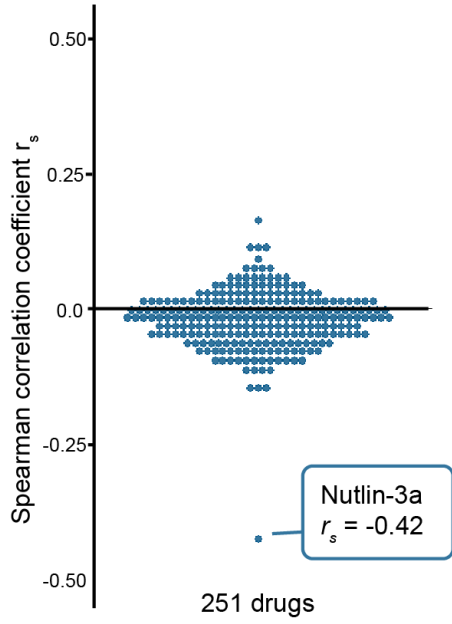11, as highlighted in the text, are enlarged in black boxes. The blue and red colors in the heatmap designate the sign and magnitude of each gene weight, as given in the legend. However, interpreting the sign of a gene weight only has value when the activity score of the respective mTC in a sample is taken into account. This activity score, too, can be either positive or negative. Only the product of a gene weight and activity score can therefore be interpreted, and the sign of this product will give an indication of a gene from a mTC being "more active" or "less active" in a sample than the average. **(B-C)** Top genes in GEO mTC 54 and TCGA mTC 127. Text colored white shows genes that are a member of the 608 metabolic gene sets. Lines signify genes that are present in the top of both GEO and TCGA mTCs. **(D-E)** Top genes in GEO mTC 11 and TCGA mTC 141. **(F)** Hierarchically clustered heatmap of genes with an absolute weight > 3 in TCGA mTCs. Sets of top genes in TCGA mTC 127 and TCGA mTC 141, as highlighted in the text, are enlarged in black boxes.
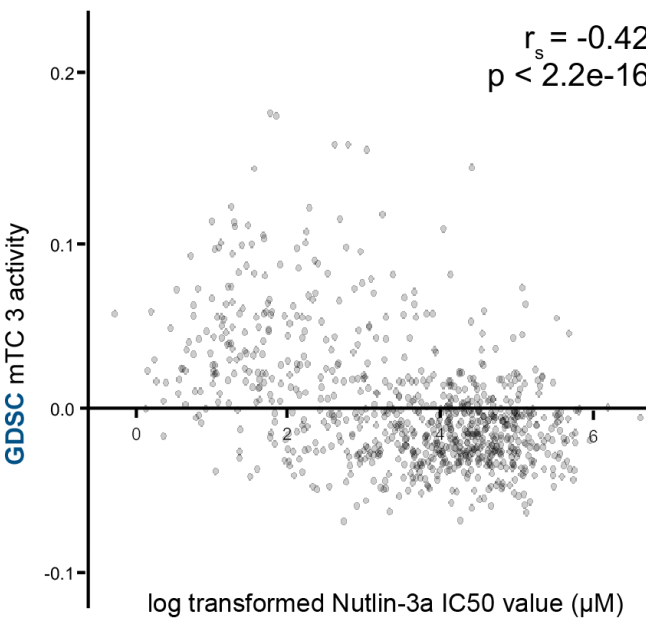
Figure 3 – Clustering activity scores of mTCs reveal multiple metabolic subtypes, which show prognostic value

(A) 21,592 GEO samples were hierarchically clustered based on mTC activity scores and divided into 67 metabolic subtypes. **(B)** 10,817 TCGA samples were hierarchically clustered based on mTC activity scores and divided into 58 metabolic subtypes. **(C)** Metabolic landscape of the subset of breast tissue samples in the GEO dataset. **(D)** Metabolic landscape of the subset of melanoma samples in the TCGA dataset. **(E)** Retrospective distant relapse-free survival of breast cancer patients in the GEO dataset. Patient-derived samples were stratified per metabolic subtype. Kaplan Meier curves are shown with a confidence interval of 0.95. **(F)** Retrospective overall survival of melanoma patients in the TCGA dataset. Patient-derived samples were stratified per metabolic subtype. Kaplan Meier curves are plotted with a confidence interval of 0.95.
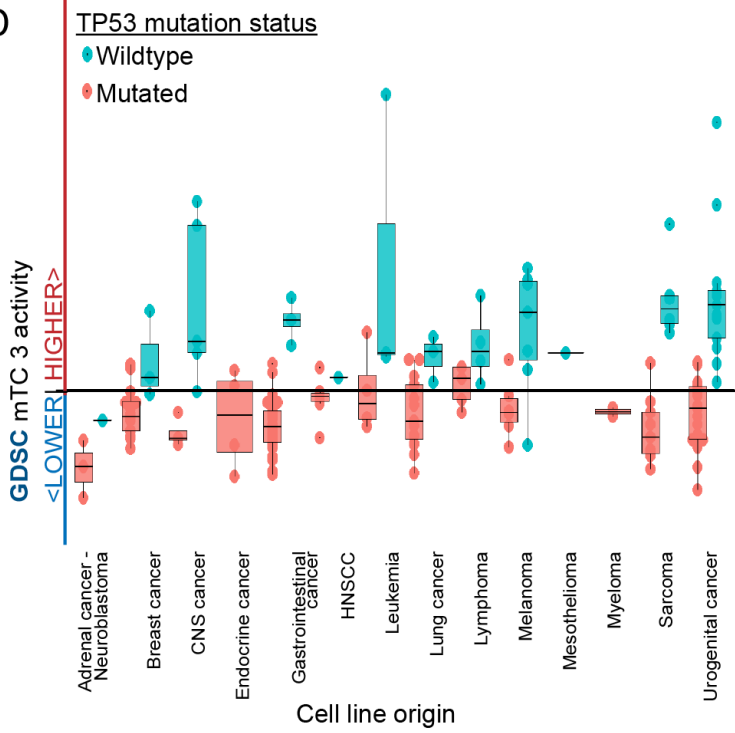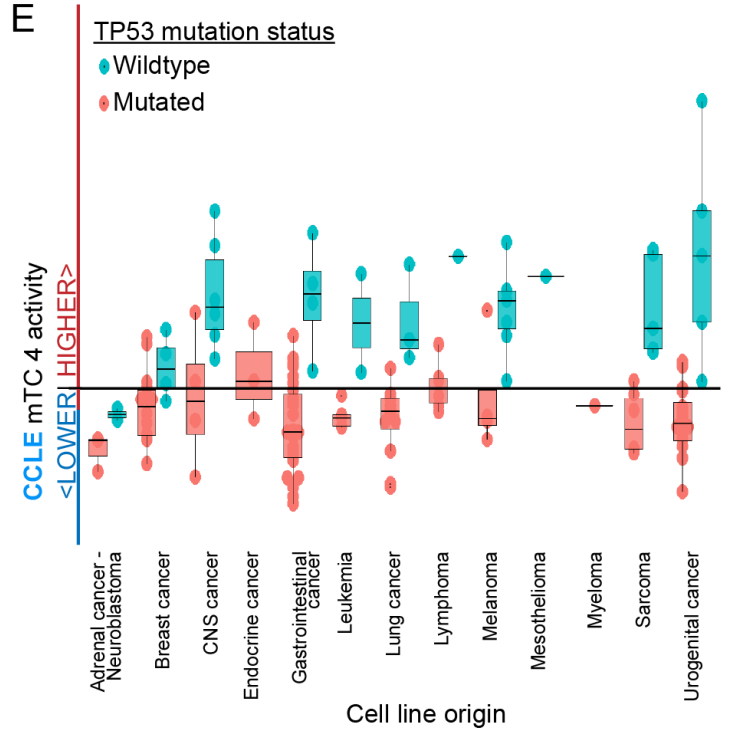
Figure 4 – Activity of GDSC mTC 3 is highly correlated with the IC50 value of Nutlin-3a and is active in cell lines with wildtype *TP53*. (A) Spearman correlations between drug IC50 values and the activity of GDSC mTC 3 **(B)** Scatter plot showing the association between the IC50 value of Nutlin-3a and activity of GDSC mTC 3. **(C)** Highest pair-wise correlations between GDSC mTC 3 and mTCs from GEO, TCGA and CCLE datasets. **(D)** Box plot of activity of GDSC mTC 3 across cell lines, colored for their *TP53* mutation status. **(E)** Box plot of activity of CCLE mTC 4 across cell lines, colored for their *TP53* mutation status.
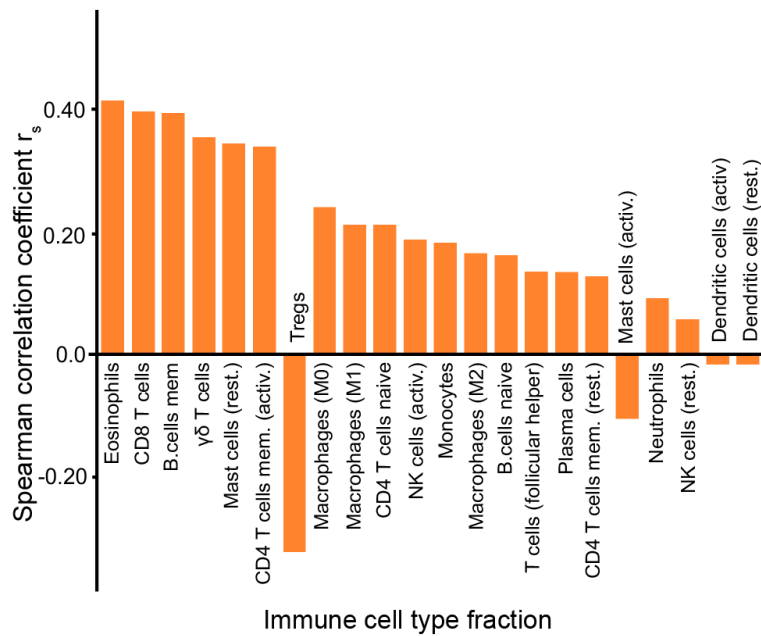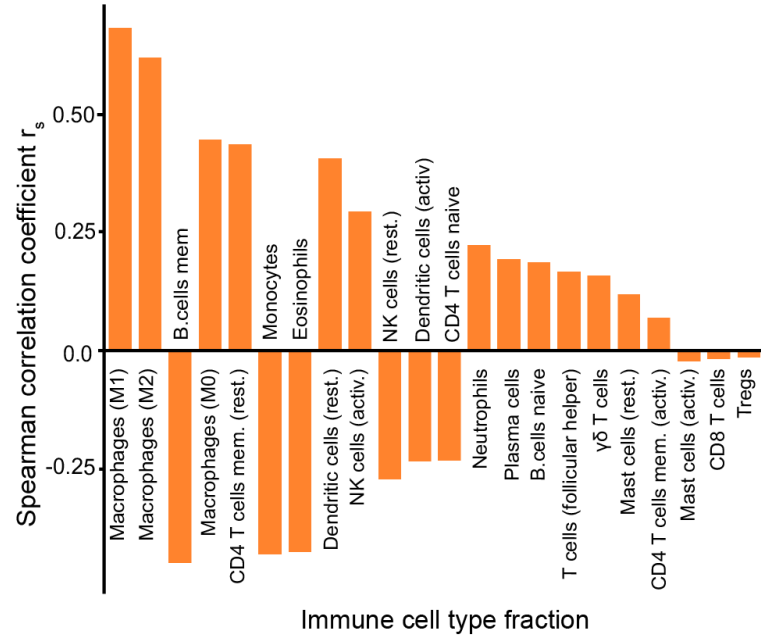
A **GDSC** mTC 18
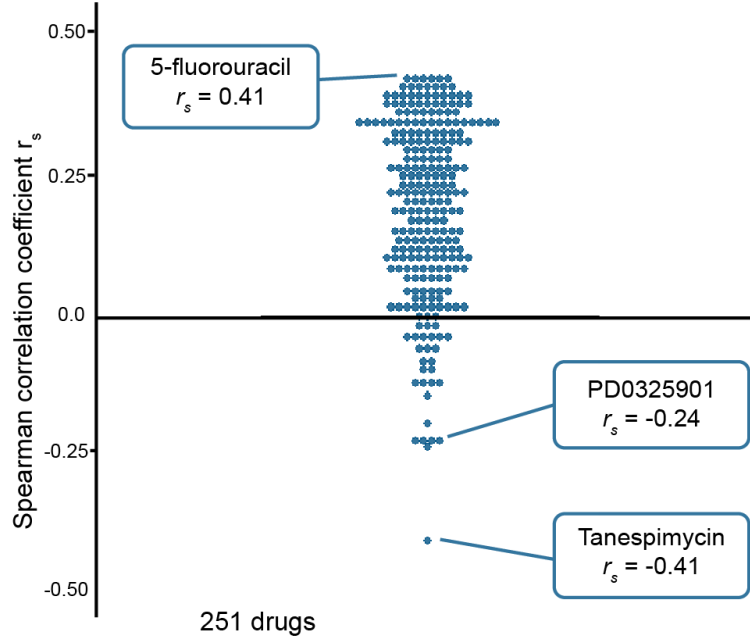
B **GDSC** mTC 108
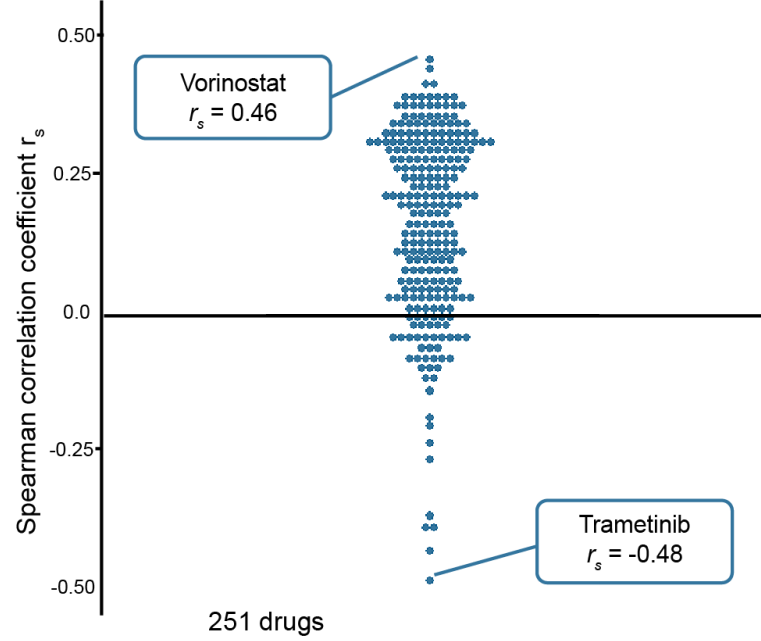
C **GEO** mTC 123

D **GEO** mTC 14

Figure 5 – Associations between mTCs, drug sensitivities, and the composition of the immune tumor microenvironment for selected examples. (A-B) Spearman correlations between drug IC50 values and the activity of GDSC mTC 18 and mTC 108. **(C-D)** Spearman correlations between CIBERSORT estimated immune cell fractions and the activity of GEO mTC 123 and mTC 14.