Supplementary Information for

# Metabolic landscape of the tumor microenvironment at single cell

resolution

Xiao et al.

# **Supplementary Methods**

#### Comparison of t-SNE results using relative mutual information

The relative mutual information metric was used to evaluate whether two clustering results are statistically similar or different. Specifically, we use  $C = \{C_1, ..., C_k\}$  to represent the t-SNE clustering results generated by using the expression profile of metabolic genes, randomly selected set of genes or the complete set of genes. For each clustering pattern the probability distribution  $P_C$  is defined as:

$$P_c(i) = \frac{n_i}{N} \qquad (1)$$

Where  $n_i$  is the number of samples in the cluster  $C_i$  and N is the total number of samples in dataset. For a pair of clustering results  $P_{CI} = (p_1 \dots p_n)$  and  $P_{C2} = (q_1 \dots q_n)$ , the mutual information of each pair of clusters I(p, q) is defined as:

$$I(p,q) = R(i,j) \log \frac{R(i,j)}{P_{C_1}(i)P_{C_2}(j)}$$
(2)

Where R(i, j) is the joint probability distribution defined as:

$$R(i,j) = \frac{|c_{1,i} \cap c_{2,j}|}{N}$$
(3)

In which  $C_{1,i}$  and  $C_{2,j}$  are the *i*-th and *j*-th cluster of  $C_1$  and  $C_2$ , respectively. Let  $n_i$  and  $m_j$  denote the numbers of samples in  $C_{1,i}$  and  $C_{2,j}$ . The mutual information between  $P_{C1}$  and  $P_{C2}$  is then defined as the sum of all I(p,q),

$$I(P_{C_1}, P_{C_2}) = \sum I(p, q) = \sum_{i,j} R(i, j) \log \frac{R(i, j)}{P_{C_1}(i) P_{C_2}(j)} = \sum_{i,j} \frac{|C_{1,i} \cap C_{2,j}|}{N} \log \frac{N \cdot |C_{1,i} \cap C_{2,j}|}{n_i m_j}$$
(4)

The relative mutual information of  $P_{C1}$  and  $P_{C2}$  is defined as:

$$I'(P_{C_1}, P_{C_2}) = \frac{I(P_{C_1}, P_{C_2})}{\max\{I(P_{C_1}, P_{C_2})\}} \quad (5)$$

Where  $Max\{I(P_{C1}, P_{C2})\}$  is the maximal possible mutual information between  $P_{C1}$  and  $P_{C2}$ . A value of  $I'(P_{C1}, P_{C2})$  close to 1 means nearly perfect consistency between  $C_1$  and  $C_2$ .



- 3 -

## **Supplementary Figure 1 (continued)**



#### **Supplementary Figure 1. (Related to Figure 1)**

(a) Distributions of numbers of expressed genes (i.e. genes with non-zero expression value)before and after imputation in different cell types from the melanoma dataset. Box-plot elements:box limits, 25 and 75 percentiles; center line, median; whiskers, minimal and maximal values.

(b) Same as in (a) but for the squama cell carcinoma of head and neck (HNSCC) dataset.

(c) Distributions of drop-out rates (i.e. fraction of zero gene expression values) in different cell

types before and after imputation in the melanoma dataset.

(d) Same as in (c) but for the HNSCC dataset.

(e) Scatter plots comparing average metabolic gene expression levels before and after imputation in different cell types from the melanoma dataset.

(f) Same as in (e) but for the HNSCC dataset.



Supplementary Figure 2. (Related to Figure 1)

(a) Distributions of Spearman's rank correlation coefficients of metabolic gene expression
between 500 randomly selected pairs of cells from the same tumor (i.e. intra-tumor) or from
different tumors (i.e. inter-tumor) in the melanoma dataset. Box-plot elements: box limits, 25 and
75 percentiles; center line, median; whiskers, minimal and maximal values.

(b) Same as in (a) but for the HNSCC dataset.



Supplementary Figure 3. (Related to Figure 1)

(a) t-SNE distribution of cells based on expression values of 1566 randomly selected genes in malignant cells from the melanoma dataset. The color of each dot indicates the tumor from which the cell was derived.

- (b) same as in (a) but for non-malignant cells from the melanoma dataset.
- (c) same as in (a) but for malignant cells from the HNSCC dataset.
- (d) same as in (a) but for non-malignant cells from the HNSCC dataset.
- (e) same as in (a) but based on expression values of the complete gene set.
- (f) same as in (b) but based on expression values of the complete gene set.
- (g) same as in (c) but based on expression values of the complete gene set.
- (h) same as in (d) but based on expression values of the complete gene set.



**Supplementary Figure 4. (Related to Figure 2)** 

(a) Distributions of relative gene expression levels in cell types from the melanoma dataset after data normalization using relative log expression (RLE), trimmed mean of M-values (TMM), upper quartile or deconvolution method. Box-plot elements: box limits, 25 and 75 percentiles; center line, median; whiskers, 1.5 × interquartile range (IQR); points, outliers.
(b) Same as in (a) but for the HNSCC dataset.





## **Supplementary Figure 5. (Related to Figure 2)**

(a) Distributions of average OXPHOS (left) and glycolysis (right) gene expression levels in cell

types in the melanoma dataset.

(b) Same as in (a) but for the HNSCC dataset.

P-values in each graph were computed using one-way ANOVA. Box-plot elements: box limits,

25 and 75 percentiles; center line, median; whiskers, 1.5 × interquartile range (IQR); points,

outliers.



С

Number of each category of metabolic pathways up-regulated in at least on cell type in the melanoma dataset



Amino acid metabolism Carbohydrate metabolism Glycan biosynthesis metabolism Metabolism of cofactors and vitamins Metabolism of other amino acids Xenobiotics biodegradation and metabolism

Nucleotide metabolism Energy metabolism

Metabolism of terpenoids and polyketides

b

Number of each category of metabolic pathways up-regulated in at least on cell type in the HNSCC dataset



Melanoma				HNSCC												
Amino acid metabolism	0.42	0.43	0.25	0.52	0.24	0.12	0.67	Amino acid metabolism	0.45	0.25	1.00	0.11	0.40	0.04	1.00	0.44
Biosynthesis of other secondary metabolites	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Biosynthesis of other secondary metabolites	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Carbohydrate metabolism	1.00	0.32	1.00	0.07	0.13	1.00	1.00	Carbohydrate metabolism	0.33	1.00	0.78	0.67	0.73	0.07	0.66	1.00
Energy metabolism	0.12	1.00	1.00	0.67	0.95	1.00	0.23	Energy metabolism	0.38	1.00	1.00	1.00	0.69	0.84	1.00	0.15
Glycan biosynthesis metabolism	1.00	0.43	0.67	0.77	0.99	0.52	0.25	Glycan biosynthesis metabolism	0.29	1.00	0.75	0.34	0.86	1.00	0.63	1.00
Lipid metabolism	0.48	0.95	0.32	0.42	0.97	0.58	0.32	Lipid metabolism	0.93	1.00	0.14	0.74	0.56	0.99	0.26	0.18
Metabolism of cofactors and vitamins	1.00	0.70	0.51	0.95	0.11	1.00	1.00	Metabolism of cofactors and vitamins	0.78	0.23	0.61	0.92	0.60	0.28	0.48	1.00
Metabolism of other amino acids	1.00	0.73	1.00	0.77	0.85	1.00	1.00	Metabolism of other amino acids	1.00	1.00	0.40	0.91	0.18	0.63	1.00	1.00
Metabolism of terpenoids and polyketides	1.00	0.27	1.00	1.00	0.62	1.00	1.00	Metabolism of terpenoids and polyketides	1.00	1.00	1.00	1.00	1.00	0.75	1.00	1.00
Nucleotide metabolism	1.00	0.48	1.00	1.00	0.38	1.00	1.00	Nucleotide metabolism	1.00	1.00	1.00	0.61	1.00	0.55	1.00	1.00
Xenobiotics biodegradation and metabolism	1.00	1.00	1.00	1.00	0.67	1.00	1.00	Xenobiotics biodegradation and metabolism	1.00	1.00	1.00	0.31	0.69	0.41	1.00	1.00
	3 cell	CAF	nellial	nage	Inant	MK.	r cell		3 cell	drittic	nelial	blast	nage	inant .	Mast -	( cell
		Enou	Macron	Mar	-				00	Enou	FID.	Macro	Mar			
p	value	0.0	0 0 0						ovalue	e 📃		000		ļ.		
		1 2	ω 4 υ	167	890	0				-	4ως	50 00 ~	489	0		
0																
				f		4 - 1		g		<b>•</b> • • • •	4		41			
e Melanoma metabolic path ≩ ³	ways			f	1	lelano	oma n	on-metabolic pathways g	INSC	C nor	n-met	abolio	c path	ways		
Melanoma metabolic path	ways	I	1	f 4	5 N	leland	oma n	on-metabolic pathways g +	INSC	C nor	n-met	abolio	c path	iways		
Melanoma metabolic path	ways			f		leland	oma n	on-metabolic pathways	INSC	C nor	n-met	abolio	c path	iways	1	
Melanoma metabolic path the second s	ways		•	f		1elano	oma n	on-metabolic pathways		C nor	n-met		c path	iways		>
Melanoma metabolic path Metapolic bit mathematic bit metabolic bit metapolic bit metap	ways	» +	· · ·	f	N A A A A A A A A A A A A A A A A A A A		oma n	on-metabolic pathways	INSC	C nor	n-met		c path	iways	Masi	>
Melanoma metabolic path metapolic bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathward bathw	ways	* *	*	f	S N S S S S S S S S S S S S S S S S S S		boma n	on-metabolic pathways	INSC st.	C nor	n-met	abolic	c path	bendritic	Nash	>
h Melanoma metabolic path b the second secon	erootreff	8 2	×+	f	A A A A A A A A A A A A A A A A A A A	Ieland	naligraf	g g g g cut there on the there of the the there of the the there of the the there of the	INSC bailgrad	C nor	h aconses	abolic	c path	bendritic	Nan	
h B cell B c	ways	> <	×	f	Mal	leland	nt	on-metabolic pathways	INSC Macr	C nor	age	abolic	Fi	brob	Noot Noot	> 
h b b b b c c c d b c d b c d c d d c d d d d	troometic T ce	> = = 0.5 p=3.6	58 e-8	f	Mal	leland رمة	nt R=0.45 0=4.9e-	on-metabolic pathways	INSC haingroot	C nor	age	e abolic	Fi	brob	Notion Notion Iasts	> 0.004 0.97
h B cell B cell	T Ce	ell R=0.5 p=3.6		f	Mal	leland رمة	nt R=0.45 p=4.9e-	on-metabolic pathways	INSC b hellocor Macr	C nor	age 0.44 6.9e-5	abolic Performente Supportente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente Performente	Fi	brob	No <sup>sh</sup>	<b>5</b> 0.004 0.97
h <sup>3</sup> <sup>4</sup> <sup>4</sup> <sup>4</sup> <sup>4</sup> <sup>4</sup> <sup>4</sup> <sup>4</sup> <sup>4</sup>	ways	<b>bil</b> R=0.5 p=3.6	58 e-8	f 3 2 2- 2- 1-	Mal	leland رچه iigna	https://www.communications.com/ https://www.communications.com/ ntt R=0.45 =4.9e-	on-metabolic pathways	HNSC S UNIT	C nor	age 0.44 6.9e-5	abolic Provine <sup>nd</sup>	Fi	opertures	Hasts R=C p=	≥ 0.004 0.97
h B cell b b b b b b b b b b b b b b b b b b	T ce	ell R=0.5 p=3.6	58 e-8 3	f 3 2 2 2 1 0 0 0	Mal	igna	oma n , 1000000000000000000000000000000000000	on-metabolic pathways	hinscr heiner Macr	C nor	age 0.44 6.9e-5	abolic Crauneed Crauneed Crauneed	Fi	brob	lasts p=	> 0.004 0.97

**Supplementary Figure 6. (Related to Figure 2)** 

(a) Numbers of different categories of metabolic pathways up-regulated in at least one cell type in the melanoma dataset.

(b) Same as in (a) but for the HNSCC dataset.

(c) One-sided Fisher's exact test p-values for enrichment of categories of pathways in pathways up-regulated in different cell types in the melanoma dataset.

- (d) Same as in (c) but for the HNSCC dataset.
- (e) Distributions of metabolic pathway activity scores in cell types in the melanoma dataset.
- (f) Distributions of non-metabolic pathway activity scores in cell types in the melanoma dataset.

(g) Same as in (f) but for the HNSCC dataset.

(h) Scatter plots comparing metabolic pathway activities between the melanoma and HNSCC datasets for cell types shared by the two datasets. The Spearman correlation coefficient is shown on the right of each graph



**Supplementary Figure 7. (Related to Figure 2)** 

Distributions of Pearson correlation coefficients between TCGA bulk tumors and bulk tumors reconstructed using the single cell dataset. The horizontal line indicating the mean value.



**Supplementary Figure 8. (Related to Figure 2)** 

(a) Violin plots and box plots comparing the distributions of metabolic pathway activities in single malignant cells and reconstructed bulk tumors from the melanoma (left) and HNSCC (right) datasets. The standard variation (SD) values of metabolic pathway activities in each group are shown on the bottom. P-values were computed using one-sided Wilcoxon's rank-sum test. Box-plot elements: box limits, 25 and 75 percentiles; center line, median; whiskers,  $1.5 \times$  interquartile range (IQR); points, outliers.

(b) Quantile-quantile (Q-Q) plots showing the discrepancy between single malignant cells and reconstructed bulk tumors in distribution of metabolic pathway activities for the melanoma (left) and HNSCC (right) datasets. P-values were computed using one-sided Kolmogorov–Smirnov (K-S) test.



#### **Supplementary Figure 9. (Related to Figure 3)**

(a) Explained variance of principal components (PCs) from principal component analysis (PCA) of metabolic gene expression levels in malignant cells from different tumors in the melanoma dataset. Top PCs accounting for 80% of the variance are highlighted in red.(b) Same as in (a) but for the HNSCC dataset.

- 15 -



**Supplementary Figure 10. (Related to Figure 3)** 

(a) Metabolic pathways enriched in genes with highest coefficient of variation (CV) in

melanoma dataset.

(b) Same as in (a) but for genes with highest standard deviation (SD).

(c) Scatter plot comparing mean gene expression and CV in malignant cells from the melanoma

tumor MEL7 as a representative example.

(d) Same as in (c) but for SD.

- 16 -

(e) Same as in (c) but for PCA score.

(f) Swarm plots comparing distributions of gene expression levels in OXPHOS and steroid hormone biosynthesis. Red lines indicate mean values.



**Supplementary Figure 11. (Related to Figure 3)** 

(a) Pairwise Pearson's correlation coefficients of gene expression levels between oxidative phosphorylation (OXPHOS), glycolysis, and response to hypoxia in malignant cells from the melanoma dataset. Color bars on the left and top show the standard deviations of expression levels of the genes.

(b) Same as in (a) but for HNSCC dataset.



#### Supplementary Figure 12. (Related to Figure 3)

- (a) Top3 GO terms enriched in genes up-regulated in cells with lowest expression levels of glycolysis, OXPHOS and hypoxia pathways for each patient in the melanoma dataset.
- (b) Same as in (a) but for HNSCC dataset.



## Supplementary Figure 13. (Related to Figure 4)

(a) Explained variance of principal components (PCs) from principal component analysis (PCA) of metabolic gene expression levels in different types of non-malignant cells from the melanoma dataset. Top PCs accounting for 80% of the variance are highlighted in red.

(b) Same as (a) but for HNSCC dataset.



## Supplementary Figure 14. (Related to Figure 5)

(a) Distributions of average expression of all OXPHOS genes (left) and the leading-edge genes enriched in the GSEA analysis (right) in CD4+ and CD8+ T cells in the melanoma dataset.

(b) Distributions of average gene expression levels for pathways differentially expressed between CD4+ and CD8+ T cells in the HNSCC dataset.

(c) Same as in (b) but for pathways differentially expressed in Th and Treg cells in the melanoma dataset.

(d) Same as in (c) but for the HNSCC dataset.

(e) same as in (d) but for pathways differentially expressed in CAF cells and Myofib cells in the HNSCC dataset.

P-values shown in each graph were computed using one-sided Wilcoxon's rank-sum test for comparison of gene expression levels between the cell subtypes. Box-plot elements: box limits, 25 and 75 percentiles; center line, median; whiskers,  $1.5 \times$  interquartile range (IQR); points, outliers.

Data set		Metabolic gene set <i>vs</i> Complete gene set	Metabolic gene set <i>vs</i> Random gene set
N 1	Malignant cells	1	1
Melanoma	Non-malignant cells	0.96	0.96
HNSCC	Malignant cells	1	1
	Non-malignant cells	1	1

**Supplementary Table 1.** Relative mutual information between t-SNE results generated using different gene sets