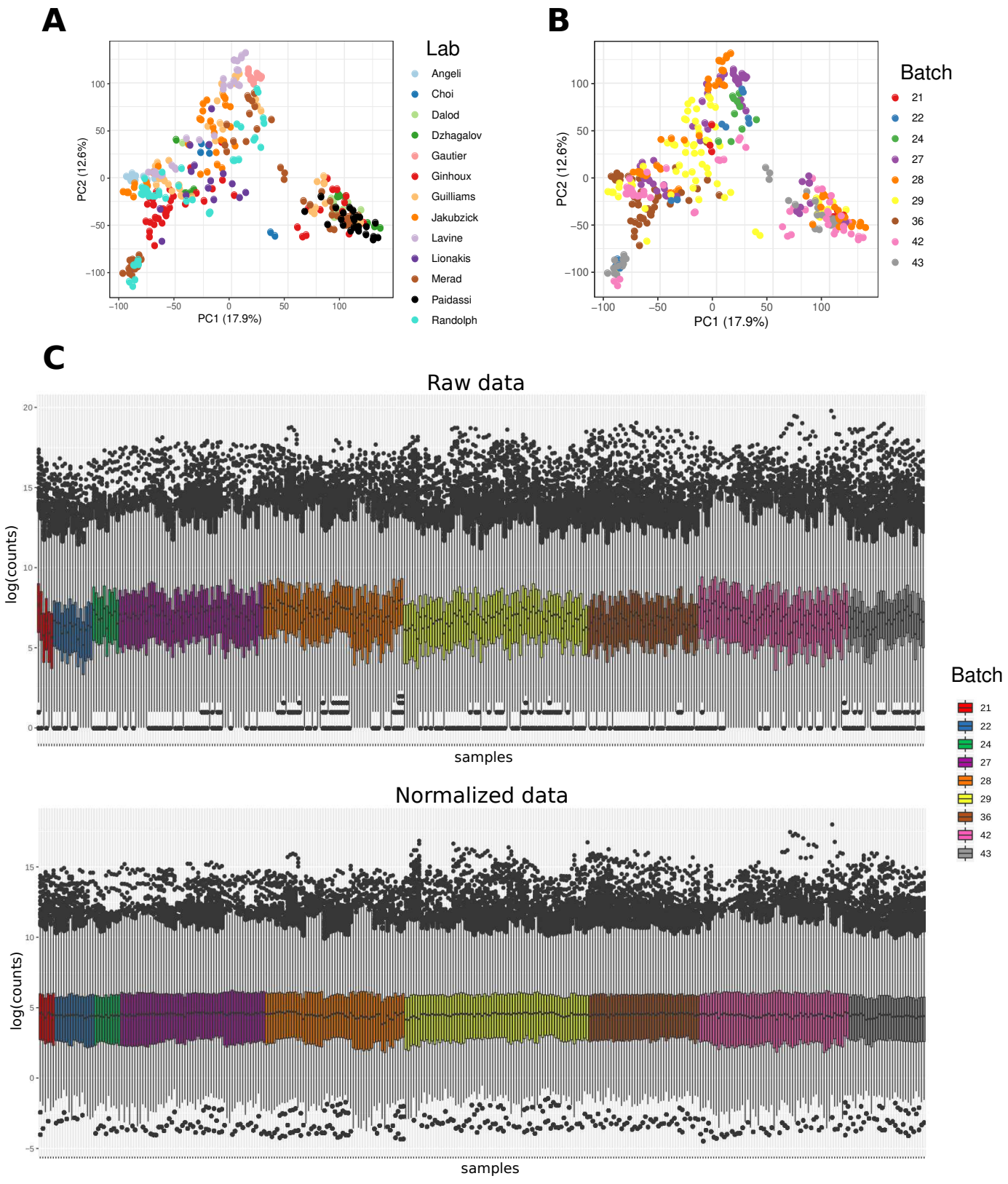
Cell Reports, Volume *42*

Supplemental information

Network analysis of large-scale ImmGen and Tabula Muris datasets highlights metabolic diversity of tissue mononuclear phagocytes

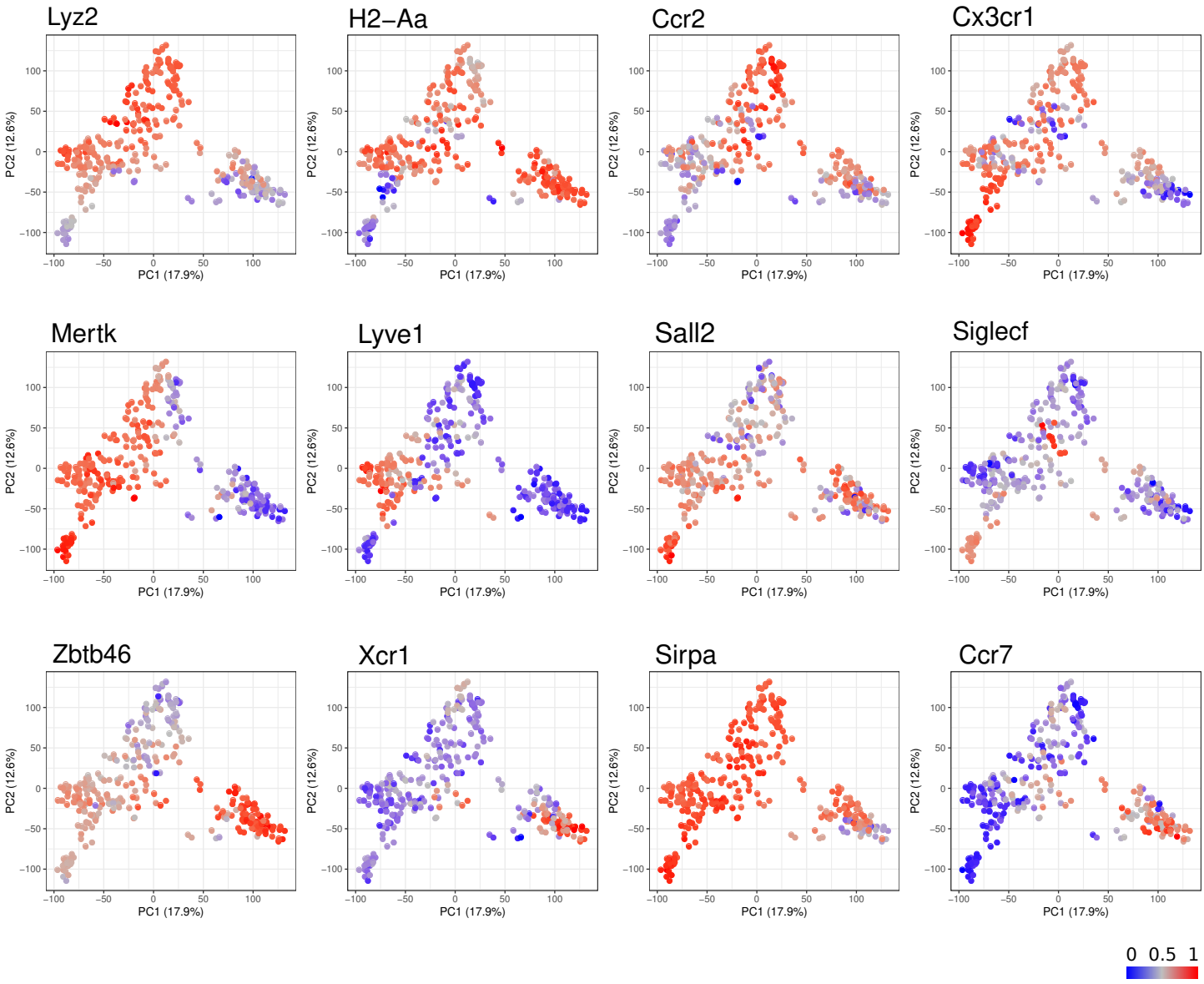
Anastasiia Gainullina, Denis A. Mogilenko, Li-Hao Huang, Helena Todorov, Vipin Narang, Ki-Wook Kim, Lim Sheau Yng, Andrew Kent, Baosen Jia, Kumba Seddu, Karen Krchma, Jun Wu, Karine Crozat, Elena Tomasello, Regine Dress, Peter See, Charlotte Scott, Sophie Gibbings, Geetika Bajpai, Jigar V. Desai, Barbara Maier, Sébastien This, Peter Wang, Stephanie Vargas Aguilar, Lucie Poupel, Sébastien Dussaud, Tyng-An Zhou, Veronique Angeli, J. Magarian Blander, Kyunghee Choi, Marc Dalod, Ivan Dzhagalov, Emmanuel L. Gautier, Claudia Jakubzick, Kory Lavine, Michail S. Lionakis, Helena Paidassi, Michael H. Sieweke, Florent Ginhoux, Martin Guilliams, Christophe Benoist, Miriam Merad, Gwendalyn J. Randolph, Alexey Sergushichev, Maxim N. Artyomov, and ImmGen Consortium

Supplementary figure 1



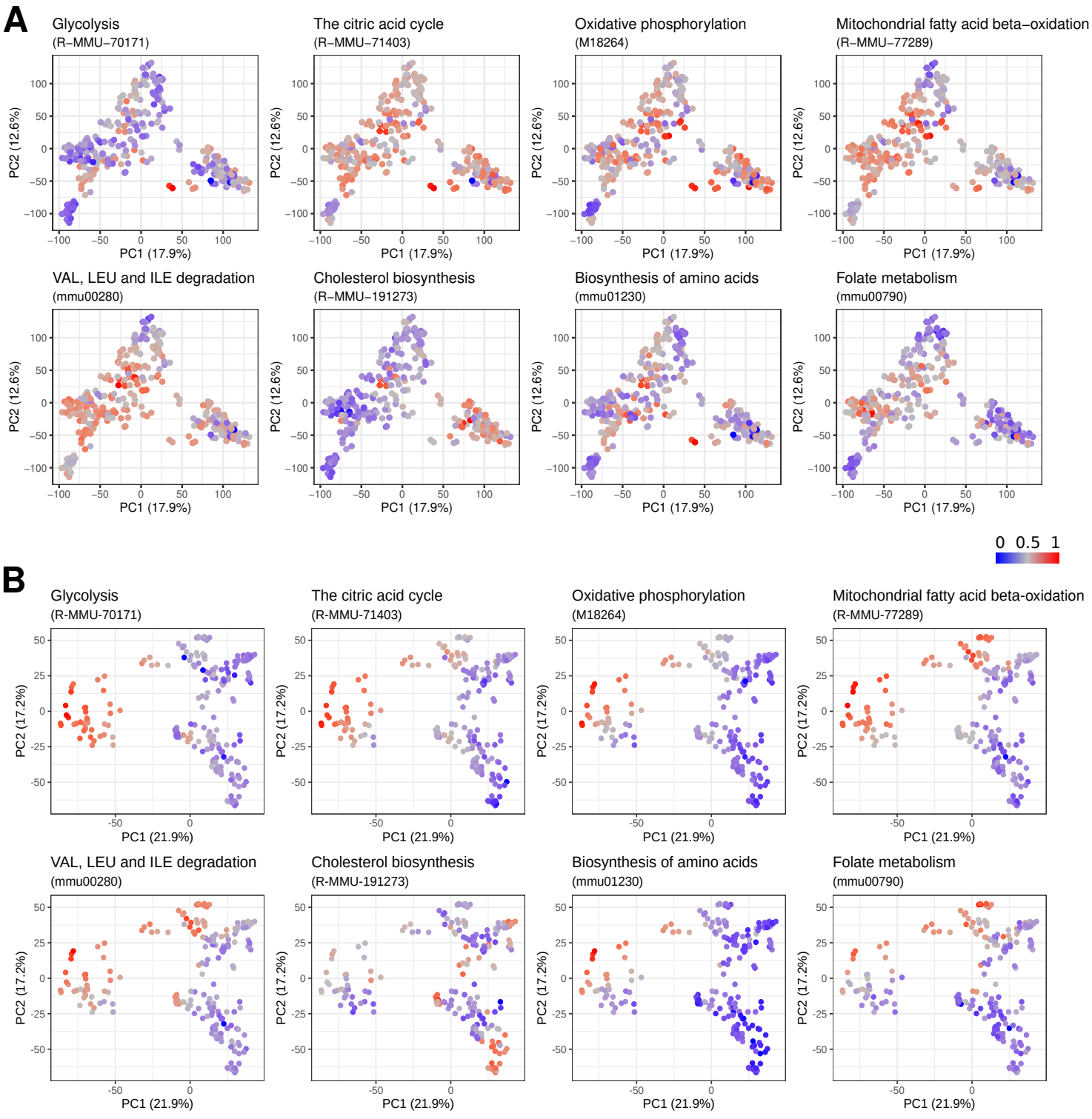
Supplementary figure 1. ImmGen MNP OS dataset sequencing characteristics. Principal component analysis (PCA) of ImmGen MNP OS dataset based on 12,000 most expressed genes across all samples colored by lab of samples sorting (a) and batch of samples sequencing (b). Axes are the first two principal components (PCs). c, Boxplots of either raw either normalized counts of 12,000 most expressed genes across all samples. Related to **Figure 1**.

Supplementary figure 2



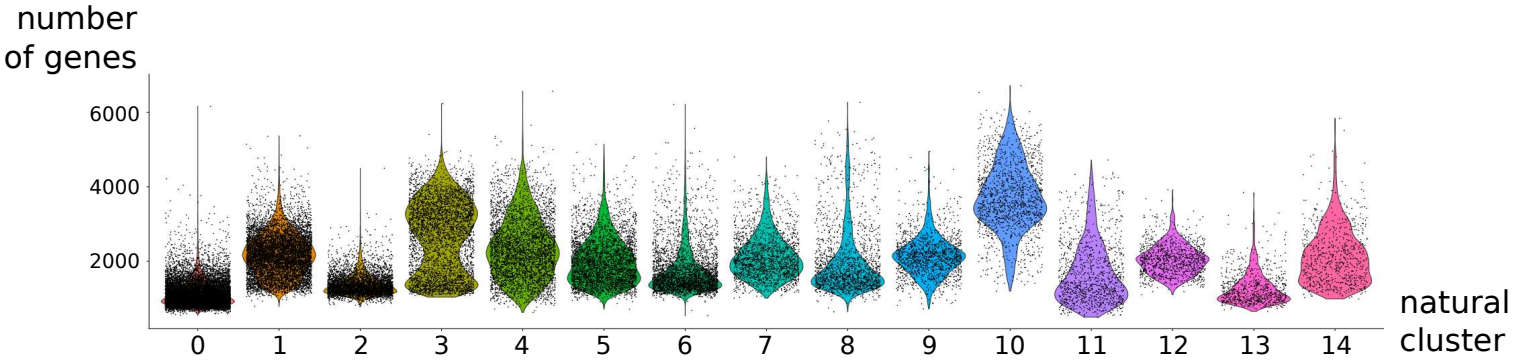
Supplementary figure 2. ImmGen MNP OS dataset cell types annotation. Principal component analysis (PCA) of ImmGen MNP OS dataset based on 12,000 most expressed genes across all samples colored by cell specific markers expression (from lowest as blue to highest as red). Axes are the first two principal components (PCs). Related to **Figure 1**.

Supplementary figure 3



Supplementary figure 3. ImmGen MNP OS and Phase1 datasets metabolic characteristics. a, Principal component analysis (PCA) based on 12,000 most expressed genes across all samples colored by intensity of gene expression (from lowest as blue to highest as red) of several KEGG and Reactome canonical pathways for ImmGen MNP OS (a) and ImmGen MNP P1 (b) datasets. Related to **Figure 1**.

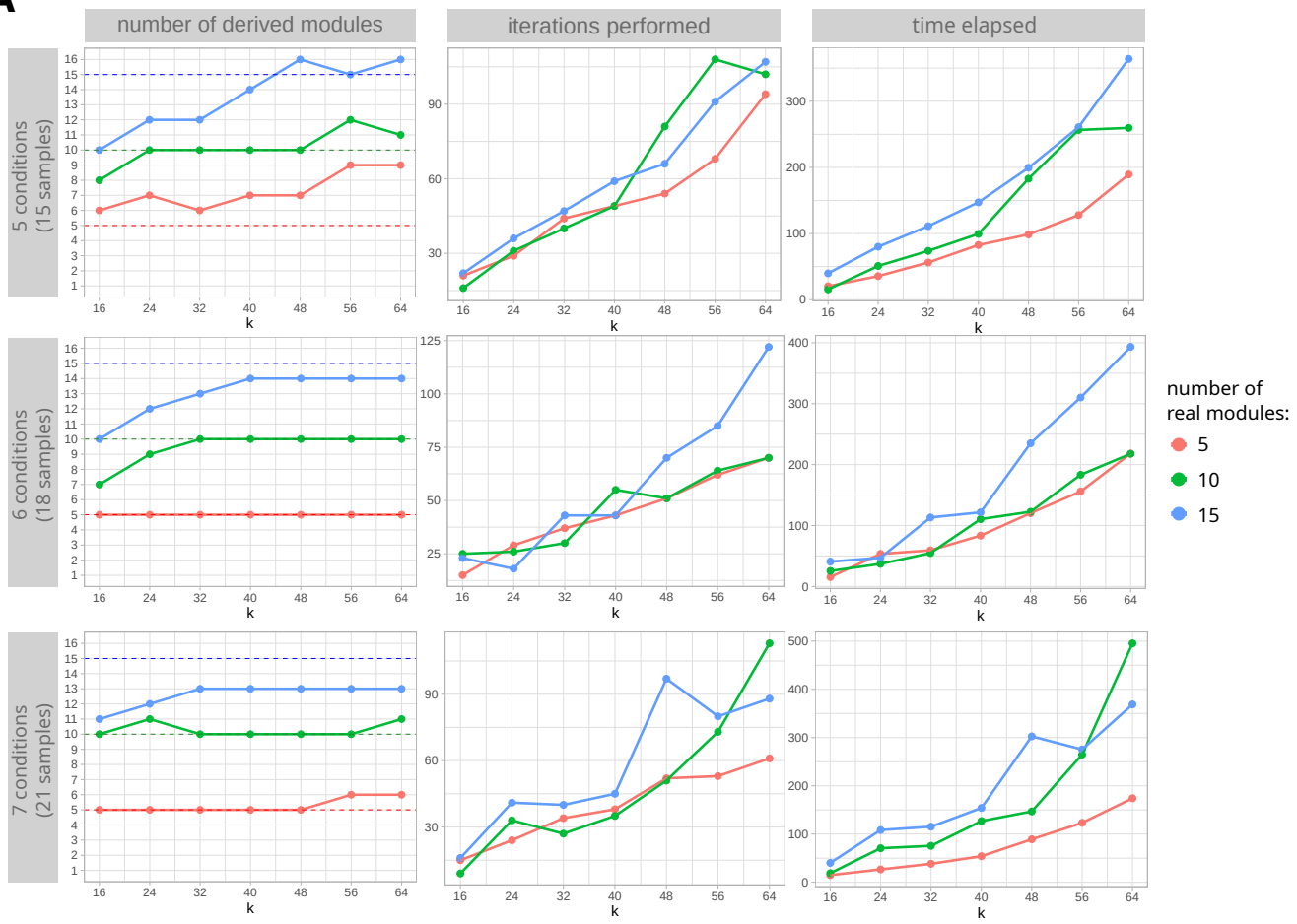
Supplementary figure 4



Supplementary figure 4. Myeloid Tabula Muris Senis dataset sequencing characteristics. Violin plot for the number of genes in the cells of each natural cluster of mTMS dataset. Related to **Figure 2**.

Supplementary figure 5

A

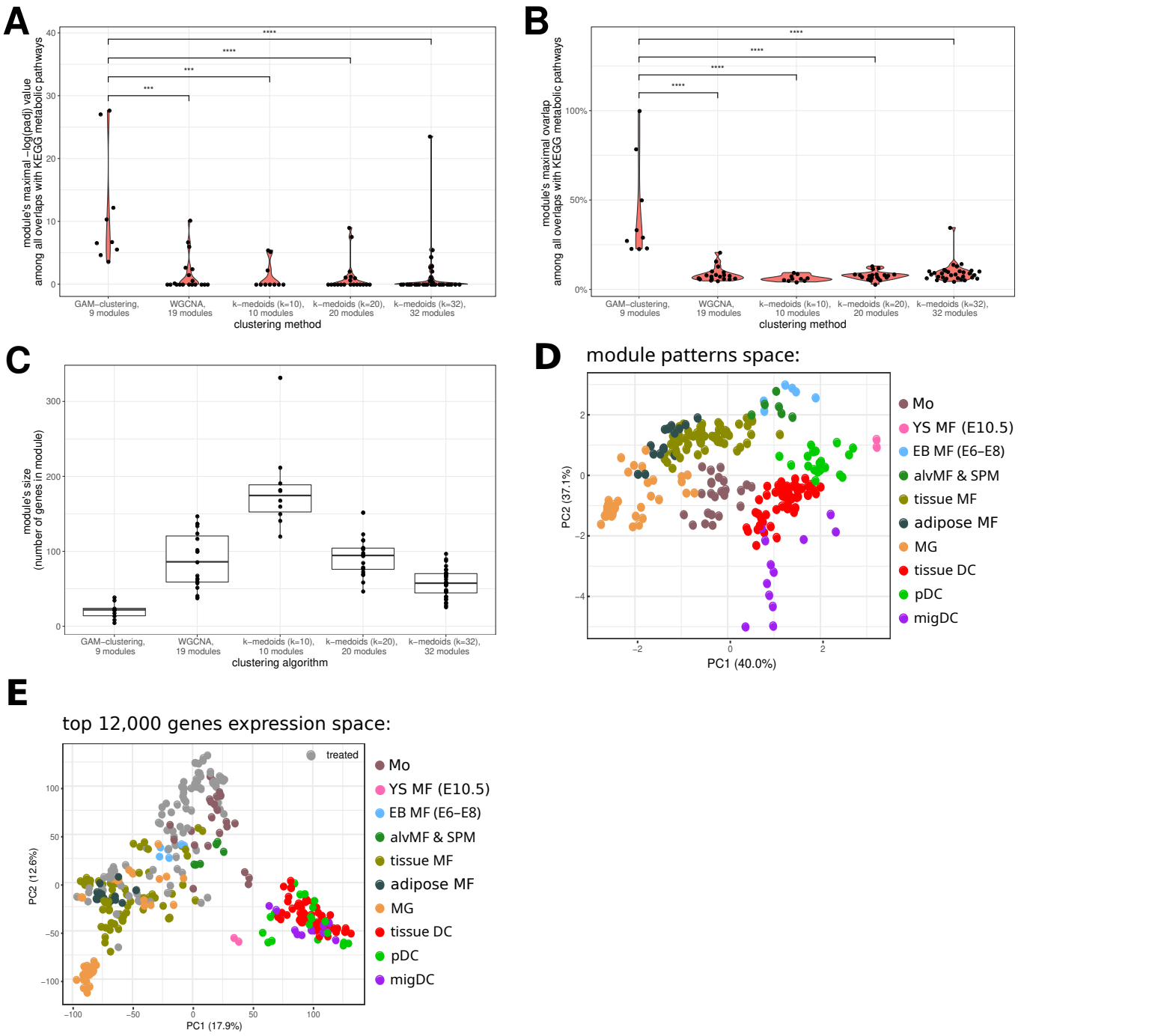


B

base	0.3	0.4	0.5	0.6
number of modules	3	9	10	16
mean number of module edges	24.3	20.9	33.9	24.6
mean correlation of module edges	0.79	0.74	0.67	0.64
number of unique annotating pathways	8	25	36	36
number of pathways related to one module only	8	17	25	21
mean percent of genes with negative score in module	17.4%	12.5%	18.7%	19.2%
mean percent of genes with negative correlation in module	0%	0%	0.4%	1%
mean percent of genes with correlation less than 1-*base* in module	17.4%	9.4%	15.5%	12.1%

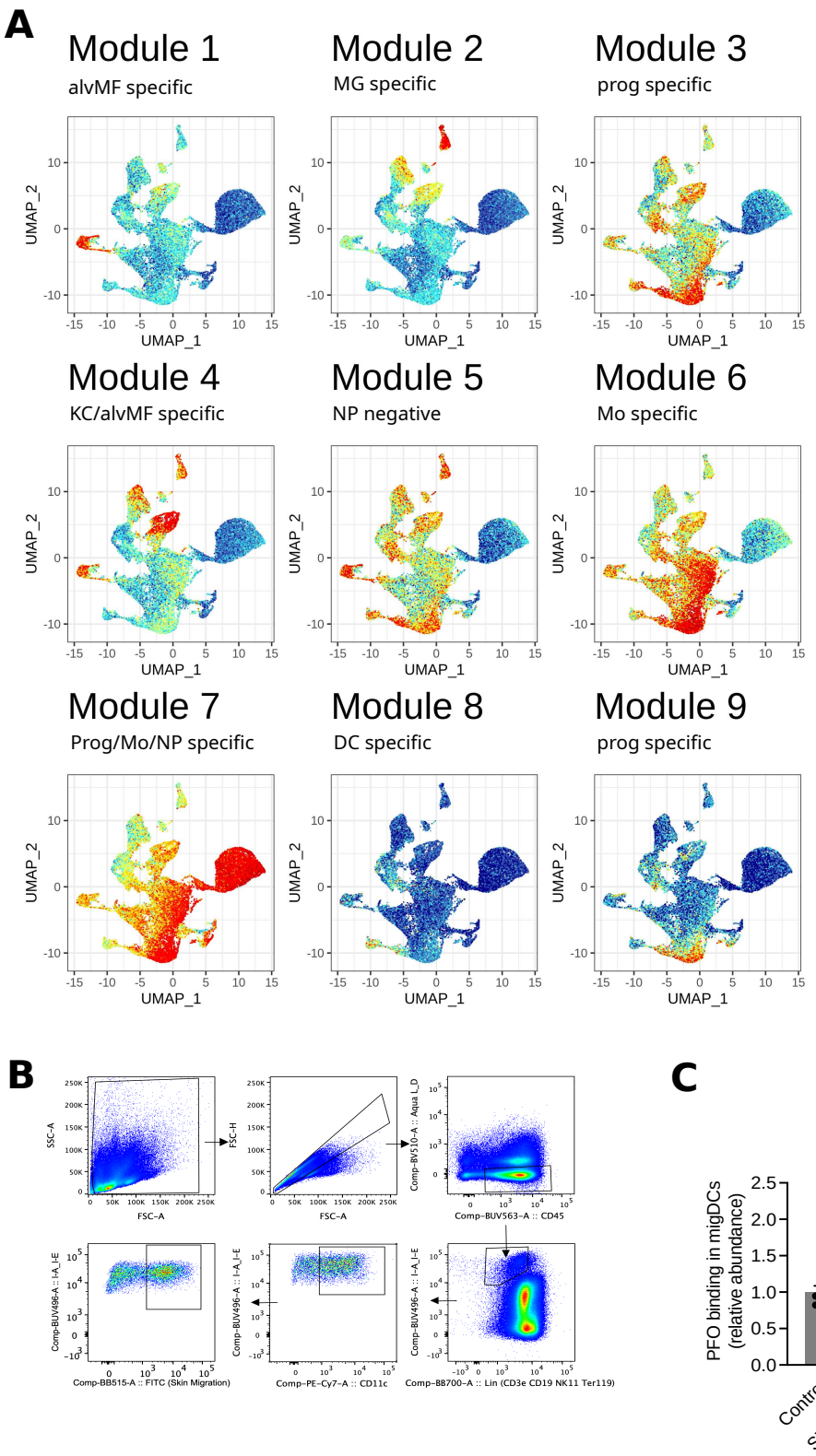
Supplementary figure 5. Analysis of and parameter values influence to characteristics of final modules performed on model (a) and ImmGen MNP OS (b) data. a, Model data imitate experiment with complex design (15, 18 or 21 samples; shown by row splitting) where several modules (5, 10 or 15; shown by colored dashed line in the first column) are active each in a particular subset of samples. All combinations of these data were analysed by the GAM-clustering method with values equal to 16, 24, 32, 40, 48, 56 or 64 and the following output features were calculated: number of final modules found by method (first column), number of iterations performed (second column) and time elapsed during the analysis (third column). b, Comparative study of the results obtained after GAM-clustering analysis of the ImmGen MNP OS data (with 32 initial clusters) with different values (0.3, 0.4, 0.5 or 0.6). Optimal value for the parameter (framed) was determined by the calculation of various characteristics of the output modules. Related to **Figure 3** and to **STAR Method "Statistical analysis of the GAM-clustering method"**.

Supplementary figure 6



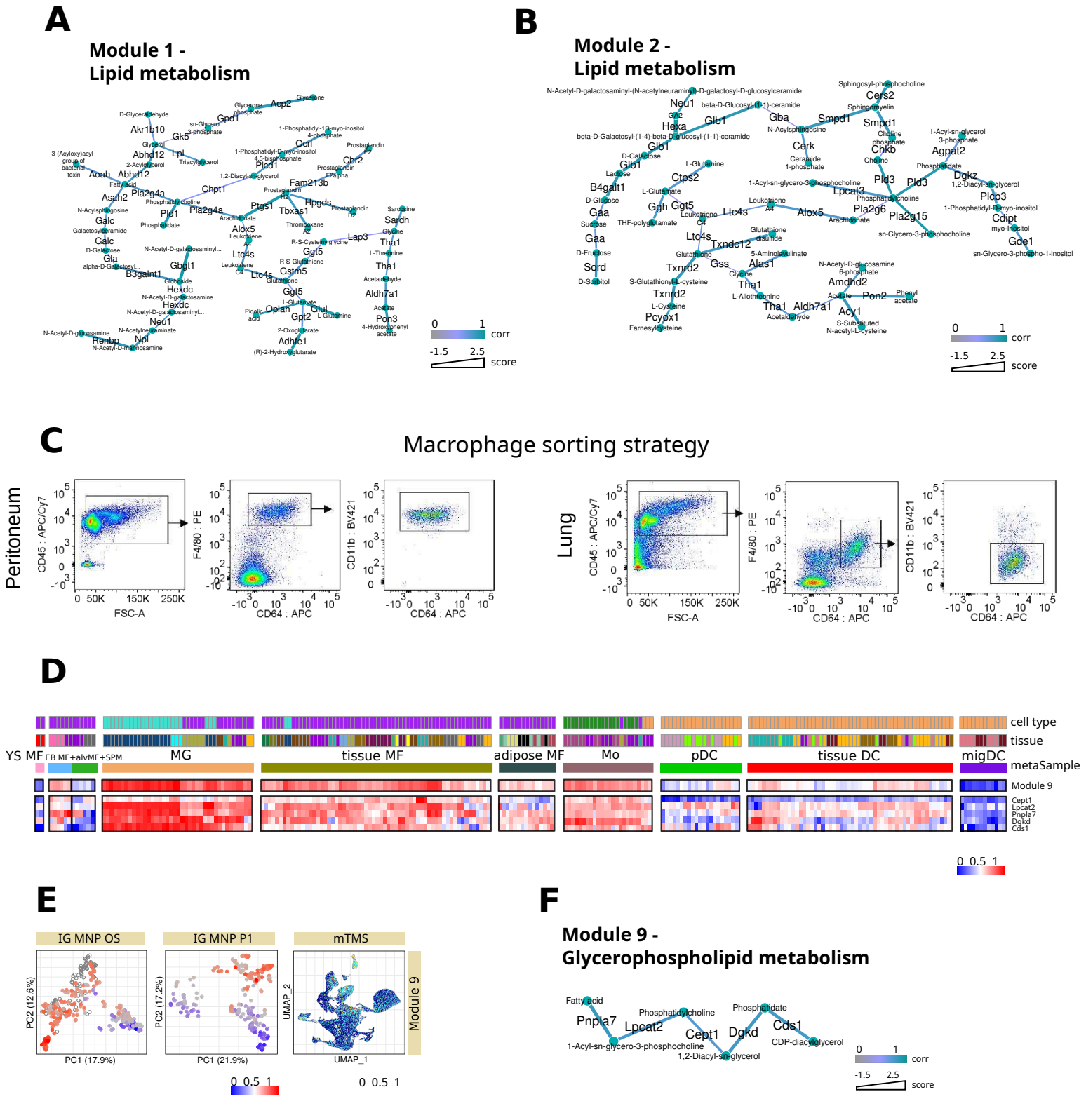
Supplementary figure 6. Comparison of GAM-clustering method with commonly used unbiased clustering approaches. a, b, Used approaches: WGCNA (weighted gene co-expression network analysis) and k-medoids (with Pearson correlation distance as a distance measure). For each clustering method, obtained clusters (modules) were examined against eighty KEGG murine metabolic pathways by hypergeometric test (common and descriptive pathways corresponding to KEGG's "Global and overview maps" were not considered in order to increase the specificity of annotation, pathways with less than 10 constituting genes were also excluded). Modules' overlap with the individual KEGG metabolic pathways was also analysed in terms of p-value (p-adjusted in case of multiple comparison) (a) and percentage of module's covered genes (b). c, Modules's sizes. *** – p-value ≤ 0.001; **** – p-value ≤ 0.0001. d, e, Principal component analysis (PCA) of ImmGen MNP OS dataset in module (d) and transcriptional (e) spaces across all samples colored on the basis of its belonging to a particular metasample (metasample names are given based on the major cell type in the current metasample: Mo – monocyte, MF – macrophage, DC – dendritic cell, YS MF – yolk sac macrophage, EB MF – embryoid body macrophage, alvMF – alveolar macrophage, SPM – small peritoneal macrophage, MG – microglia, pDC – plasmacytoid dendritic cell, migDC – migratory dendritic cell). Related to **Figure 3** and to **STAR Method "GAM-clustering"**.

Supplementary figure 7



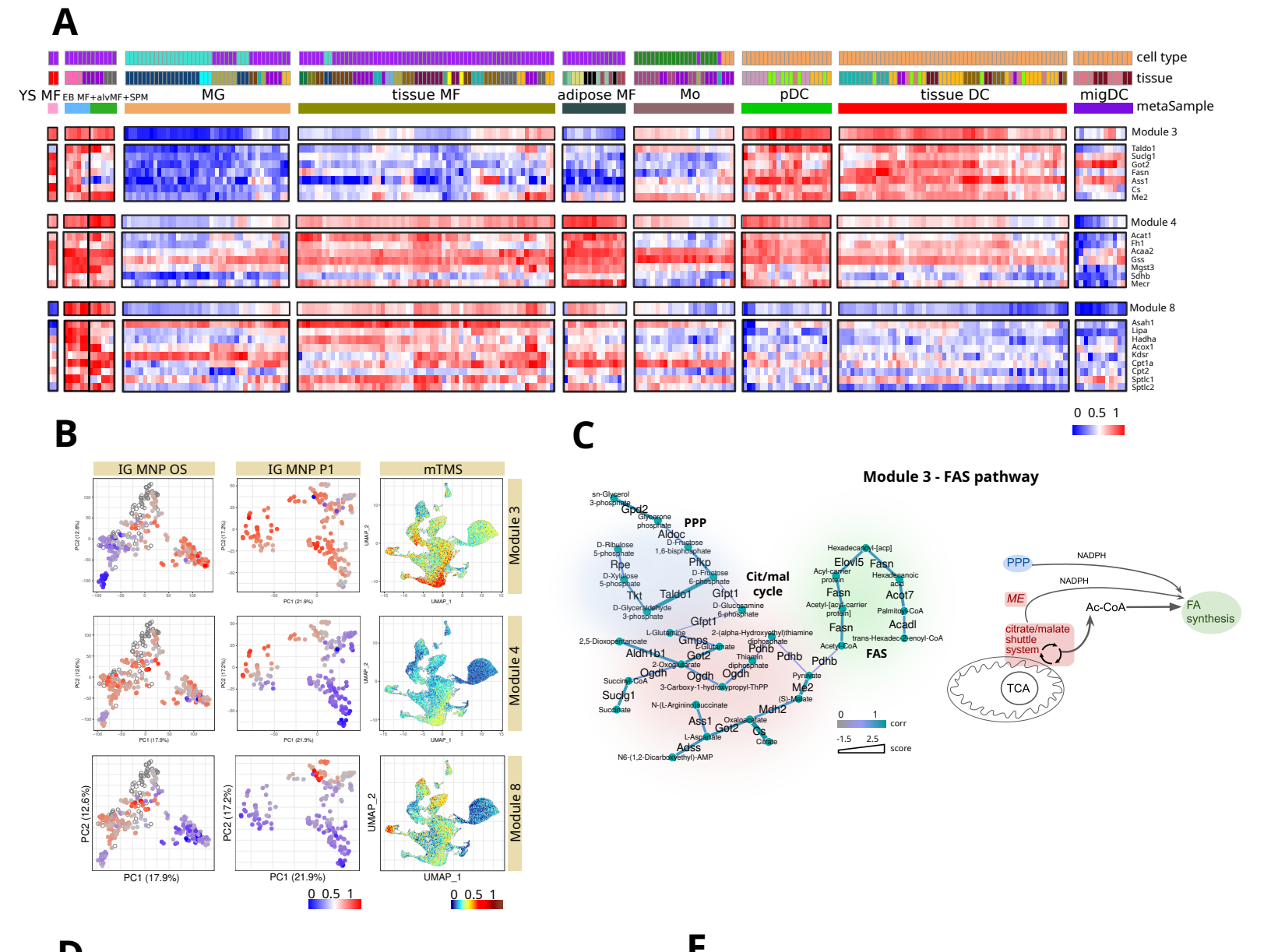
Supplementary figure 7. Modules derived from GAM-clustering analysis of mTMS dataset. a, UMAP plots colored by intensity of gene expression (from lowest as blue to highest as red) of modules derived from GAM-clustering analysis of mTMS dataset; prog – progenitor, MF – macrophage, alvMF – alveolar macrophage, MG – microglia, KC – Kupffer cell, Mo – monocyte, DC – dendritic cell, NP – neutrophil. b, Sorting strategy for dendritic cells. c, Membrane cholesterol levels of migDCs isolated from draining lymph nodes; n = 5 mice in each group; statistical analysis by unpaired two-tail t-test. Related to **Figure 4** and to **STAR Method "GAM-clustering"**.

Supplementary figure 8



Supplementary figure 8. Metabolic modules associated with lipid metabolism. Edges of modules are attributed with color according to correlation of its enzyme's gene expression to this particular module pattern and with thickness according to its score (a,b,f). c, Sorting strategy for macrophages. d, Heatmaps of module patterns along with the expression of some of its genes (from lowest as blue to highest as red). e, Enrichment of modules genes expression (from lowest as blue to highest as red, transparent dots correspond to treated samples) across all three analysed datasets. Related to **Figure 4**.

Supplementary figure 9



Supplementary figure 9. Metabolic modules associated with fatty acid synthesis in cytosol (b) and mitochondria (c) and sphingolipid synthesis (e). a, Heatmaps of module patterns along with the expression of some of its genes (from lowest as blue to highest as red). b, Enrichment of modules genes expression (from lowest as blue to highest as red, transparent dots correspond to treated samples) across all three analysed datasets. c,d,e, Metabolic modules per se and corresponding schematic diagrams. Edges of modules are attributed with color according to correlation of its enzyme's gene expression to this particular module pattern and with thickness according to its score. Related to **Figure 4**.