

IOBR Tutorial

Dongqiang Zeng

2023-08-24

Contents

IOBR introduction	5
0.1 Introduction	6
0.2 License	6
0.3 Publishment	7
0.4 Reporting bugs	7
1 How to install IOBR	9
1.1 Installing Dependency Packages	9
1.2 Install IOBR package	9
1.3 The main pipeline of IOBR	10
1.4 Main Functions	10
2 RNA Data preprocessing	15
2.1 Loading packages	15
2.2 Downloading data for example	15
2.3 Gene Annotation	16
2.4 Identifying outlier samples	18
2.5 PCA analysis of molecular subtypes	20
2.6 Batch effect correction	21
2.7 References	24
3 Tumor ecosystem analysis	25
3.1 Determine TME subtype of gastric cancer using TMEclassifier	25
3.2 DEG analysis	27
3.3 GSEA analysis of DEGs	28
3.4 DEG analysis	32
3.5 Identifying signatures associated with TME clusters	35
4 Signatures and relevant phenotypes	45
4.1 Loading packages	45
4.2 Downloading data for example	45
4.3 Signature score estimation	47
4.4 Identifying features associated with survival	47
4.5 Visulization using heatmap	49

4.6	Focus on target signatures	50
4.7	Survival analysis	53
4.8	Batch correlation analysis	56
4.9	Visulization of correlations	59
5	TME deconvolution	61
5.1	Loading packages	61
5.2	Downloading data for example	61
5.3	Available Methods to Decode TME Contexture	63
5.4	Method 1: CIBERSORT	64
5.5	Method 2: EPIC	65
5.6	Method 3: MCPcounter	66
5.7	Method 4: xCELL	66
5.8	Method 5: ESTIMATE	67
5.9	Method 6: TIMER	68
5.10	Method 7: quanTIseq	68
5.11	Method 8: IPS	70
5.12	Combination of above deconvolution results	71
6	Cross-references	73
6.1	Chapters and sub-chapters	73
6.2	Captioned figures and tables	73
7	Footnotes and citations	77
7.1	Footnotes	77
7.2	Citations	77
8	Blocks	79
8.1	Equations	79
8.2	Theorems and proofs	79
8.3	Callout blocks	80
9	Sharing your book	81
9.1	Publishing	81
9.2	404 pages	81
9.3	Metadata for sharing	81
10	References	83
10.1	TME deconvolution	83
10.2	TME Signatures	84
10.3	Data sets	85
10.4	Others	86

IOBR introduction

Preface

IOBR R package workflow

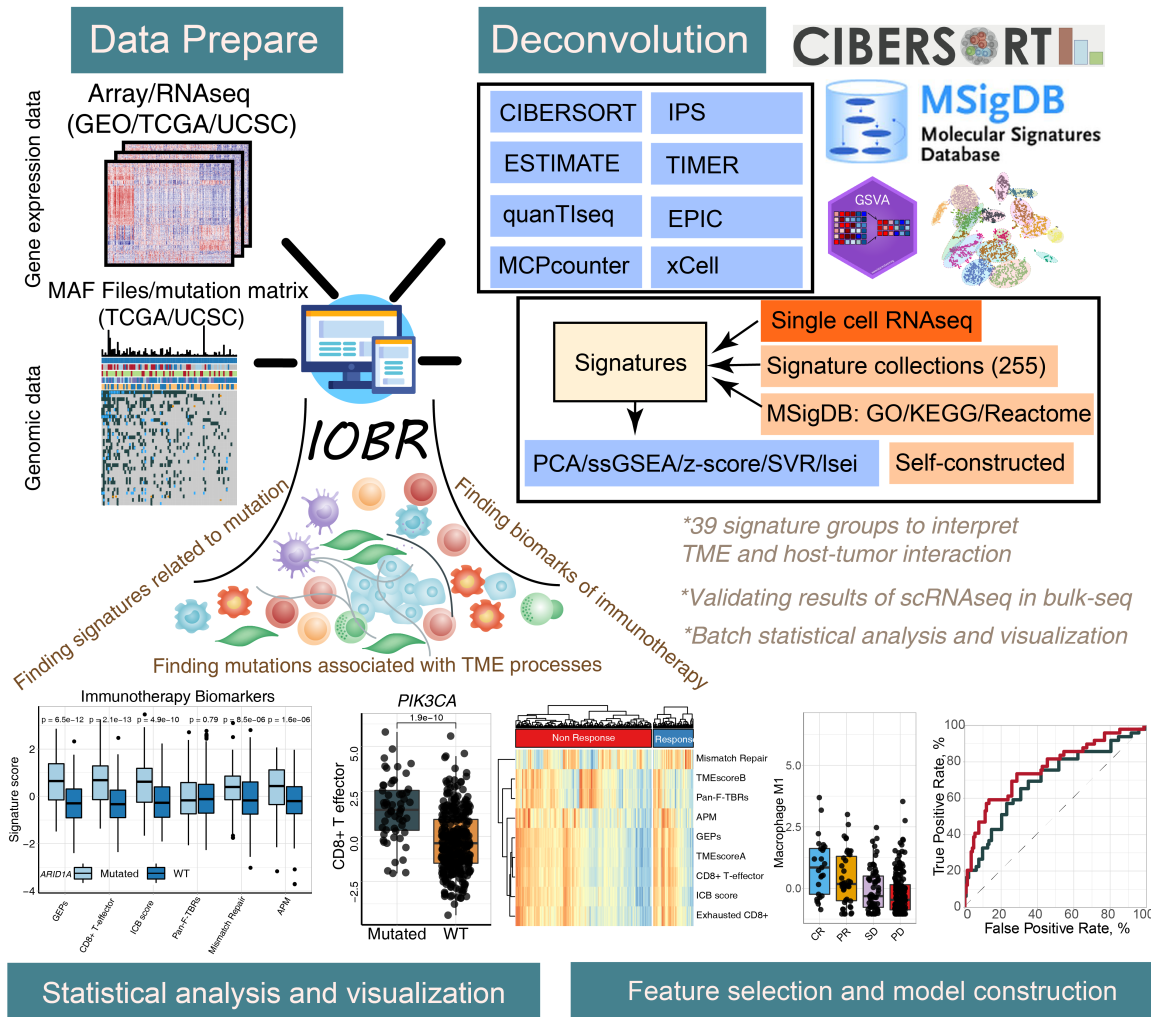


Figure 1: The workflow of IOBR

0.1 Introduction

IOBR is design for Immuno-Oncology Biological Research. Recent advance in next-generation sequencing has triggered the rapidly accumulating publicly available multi-omics data. The application of integrated omics to exploring robust signatures for clinical translation is increasingly highlighted in immuno-oncology but raises computational and biological challenges. This vignette aims to demonstrate how to utilize the package named IOBR to perform multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures for clinical translation.

This R package integrates 8 published methodologies for decoding tumor microenvironment (TME) contexture: CIBERSORT, TIMER, xCell, MCPcounter, ESITMATE, EPIC, IPS, quanTIseq. Moreover, 255 published signature gene sets were collected by IOBR, involving tumor microenvironment, tumor metabolism, m6A, exosomes, microsatellite instability, and tertiary lymphoid structure. Run the function `signature_collection_citation` to obtain the source papers, and the function `signature_collection` returns the detail signature genes of all given signatures. Subsequently, IOBR adopts three computational methods to calculate the signature score, comprising PCA, z-score, and ssGSEA. To note, IOBR collected and employed multiple approaches for variable transition, visualization, batch survival analysis, feature selection, and statistical analysis. Batch analysis and visualization of corresponding results are supported. The details of how IOBR works are described below.

0.2 License

IOBR is released under the GPL v3.0 license. See LICENSE for details. The code contained in this book is simultaneously available under the GPL license; this means that you are free to use it in your packages, as long as you cite the source. The online version of this book is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

0.3 Publishment

Zeng D, Ye Z, Shen R, Yu G, Wu J, Xiong Y,..., Liao W (2021) **IOBR**: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. *Frontiers in Immunology*. 12:687975. doi: 10.3389/fimmu.2021.687975

0.4 Reporting bugs

Please report bugs to the Github issues page

E-mail any questions to dongqiangzeng0808@gmail.com

Chapter 1

How to install IOBR

1.1 Installing Dependency Packages

It is essential that you have R 3.6.3 or above already installed on your computer or server. IOBR is a pipeline that utilizes many other R packages that are currently available from CRAN, Bioconductor and GitHub.

```
if (!requireNamespace("BiocManager", quietly = TRUE)) install.packages("BiocManager")
depends<-c('tibble', 'survival', 'survminer', 'limma', "DESeq2", "devtools", 'limSolve', '
          "devtools", "tidyHeatmap", "caret", "glmnet", "ppcor", "timeROC", "pracma", "
          "FactoMineR", "WGCNA", "patchwork", 'ggplot2', "biomaRt", 'ggpubr')
for(i in 1:length(depends)){
  depen<-depends[i]
  if (!requireNamespace(depen, quietly = TRUE)) BiocManager::install(depen, update = FA
}
```

1.2 Install IOBR package

When the dependent environments are built, users are able to install IOBR from github by typing the following code into your R session:

```
if (!requireNamespace("IOBR", quietly = TRUE)) devtools::install_github("IOBR/IOBR")

## Warning: package 'tidyHeatmap' was built under R version 4.2.3

library(IOBR)

## Warning: package 'tibble' was built under R version 4.2.3

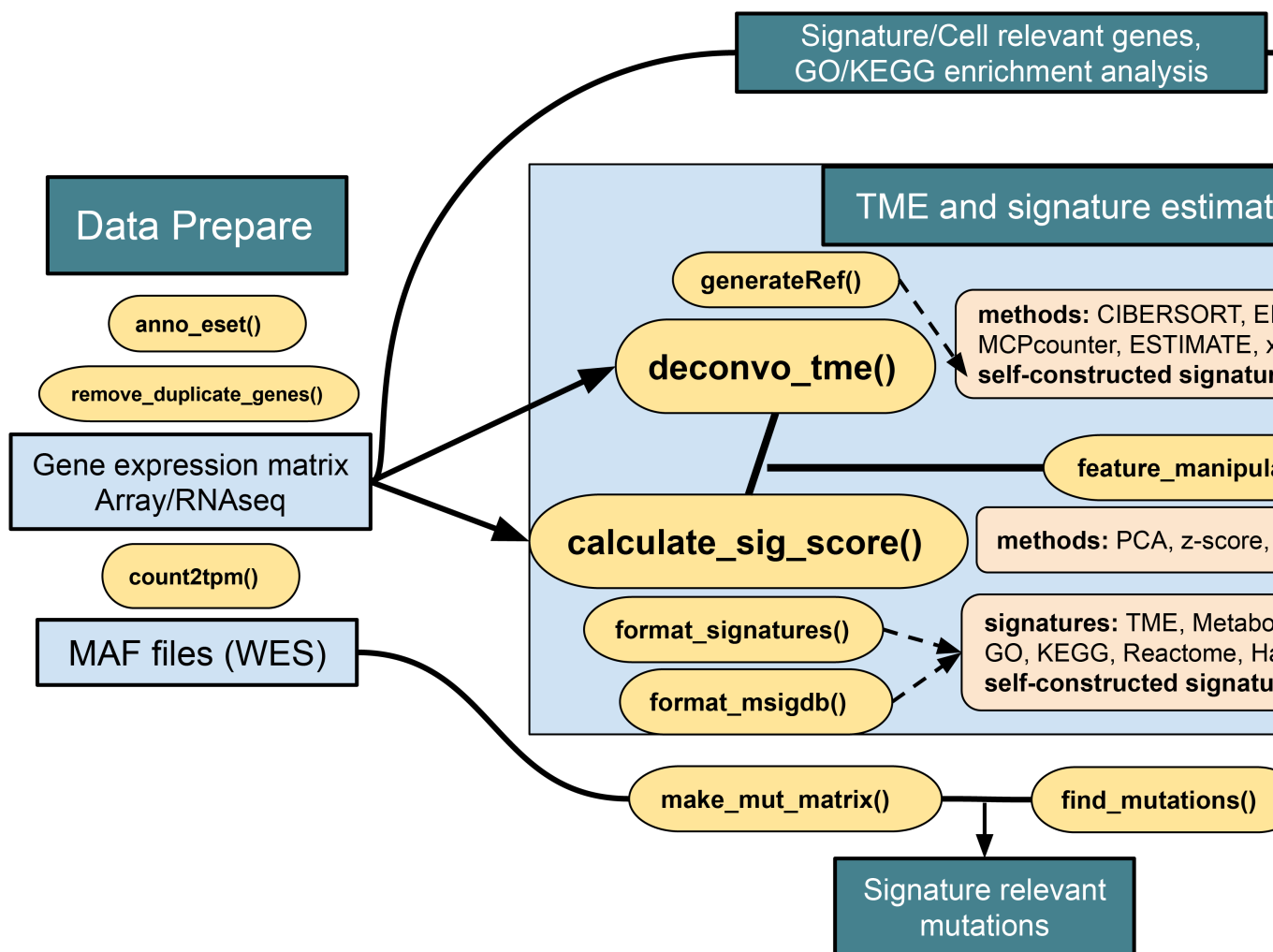
## Warning: package 'dplyr' was built under R version 4.2.3

## Warning: package 'ggplot2' was built under R version 4.2.3
```

1.3 The main pipeline of IOBR

1.4 Main Functions

- **Data Preparation: data annotation and transformation**
 - `count2tpm()`: transform count data of RNA sequencing into TPM data.
 - `anno_eset()`: annotate the normalized genes expression matrix, including RNAseq and array (Affymetrix or Illumina).
 - `remove_duplicate_genes()`: remove the genes annotated with the duplicated symbol after normalization and retain only the symbol with highest expression level.
- **TME Deconvolution Module: integrate multiple algorithms to decode immune contexture**
 - `deconvo_tme()`: decode the TME infiltration with different deconvolution methodologies, based on bulk RNAseq, microarray or single cell RNAseq data.
 - `generateRef()`: generate a novel gene reference matrix for a specific feature such as infiltrating cell, through the SVR and lsei algorithm.
- **Signature Module: calculate signature scores, estimate phenotype related signatures and corresponding genes, and evaluate signatures generated from**



IOBR (Immuno-Oncology Bioinformatics)

Figure 1.1: The main pipeline of IOBR

single-cell RNA sequencing data

- `calculate_sig_score()`: estimate the interested signatures enrolled in IOBR R package, which involves TME-associated, tumor-metabolism, and tumor-intrinsic signatures.
- `feature_manipulation()`: manipulate features including the cell fraction and signatures generated from multi-omics data for latter analysis and model construction. Remove missing values, outliers and variables without significant variance.
- `format_signatures()`: generate the object of `calculate_sig_score()` function, by inputting a data frame with signatures as column names of corresponding gene sets, and return a list contain the signature information for calculating multiple signature scores.
- `format_msigdb()`: transform the signature gene sets data with gmt format, which is not included in the signature collection and might be downloaded in the MSigDB website, into the object of `calculate_sig_score()` function.
- **Batch Analysis and Visualization: batch survival analysis and batch correlation analysis and other batch statistical analyses**
 - * `batch_surv`: batch survival analysis of multiple continuous variables including varied signature scores.
 - * `subgroup_survival`: batch survival analysis of multiple categorized variables with different number of subgroups.
 - * `batch_cor()`: batch analysis of correlation between two continuous variables using Pearson correlation coefficient or Spearman's rank correlation coefficient .
 - * `batch_wilcoxon()`: conduct batch wilcoxon analyses of binary variables.
 - * `batch_pcc()`: batch analyses of Partial Correlation coefficient(PCC) between continuous variables and minimize the interference derived from confounding factors.
 - * `iobr_cor_plot()`: visualization of batch correlation analysis of signatures

from ‘sig_group’. Visualize the correlation between signature or phenotype with expression of gene sets in target signature is also supported.

- * `cell_bar_plot()`: batch visualization of TME cell fraction, supporting input of deconvolution results from ‘CIBERSORT’, ‘EPIC’ and ‘quanTIseq’ methodologies to further compare the TME cell distributions within one sample or among different samples.

- **Signature Associated Mutation Module: identify and analyze mutations relevant to targeted signatures**

- `make_mut_matrix()`: transform the mutation data with MAF format(contain the columns of gene ID and the corresponding gene alterations which including SNP, indel and frameshift) into a mutation matrix in a suitable manner for further investigating signature relevant mutations.
- `find_mutations()`: identify mutations associated with a distinct phenotype or signature.

- **Model Construction Module: feature selection and fast model construct to predict clinical phenotype**

- `BinomialModel()`: select features and construct a model to predict a binary phenotype.
- `PrognosticMode()`: select features and construct a model to predict clinical survival outcome.

Chapter 2

RNA Data preprocessing

2.1 Loading packages

Load the IOBR package in your R session after the installation is complete:

```
library(IOBR)
library(tidyverse)
library(clusterProfiler)
```

2.2 Downloading data for example

Obtaining data set from GEO Gastric cancer: GSE62254 using GEOquery R package.

```
if (!requireNamespace("GEOquery", quietly = TRUE)) BiocManager::install("GEOquery")
library("GEOquery")
# NOTE: This process may take a few minutes which depends on the internet connection
eset_geo<-getGEO(GEO = "GSE62254", getGPL = F, destdir = "./")
eset    <-eset_geo[[1]]
eset    <-exprs(eset)
eset[1:5,1:5]
```

```
##          GSM1523727 GSM1523728 GSM1523729 GSM1523744 GSM1523745
## 1007_s_at  3.2176645  3.0624323  3.0279131   2.921683   2.8456013
## 1053_at   2.4050109  2.4394879  2.2442708   2.345916   2.4328582
## 117_at    1.4933412  1.8067380  1.5959665   1.839822   1.8326058
## 121_at    2.1965561  2.2812181  2.1865556   2.258599   2.1874363
## 1255_g_at 0.8698382  0.9502466  0.8125414   1.012860   0.9441993
```

2.3 Gene Annotation

Annotation of genes in the expression matrix and removal of duplicate genes.

```
# Load the annotation file `anno_hug133plus2` in IOBR.
head(anno_hug133plus2)
```

```
## # A tibble: 6 x 2
##   probe_id symbol
##   <fct>      <fct>
## 1 1007_s_at MIR4640
## 2 1053_at   RFC2
## 3 117_at    HSPA6
## 4 121_at    PAX8
## 5 1255_g_at GUCA1A
## 6 1294_at   MIR5193
```

```
# Load the annotation file `anno_grch38` in IOBR.
head(anno_grch38)
```

```
##          id eff_length      gc entrez  symbol chr   start      end
## 1 ENSG000000000003      4536 0.3992504   7105  TSPAN6  X 100627109 100639991
## 2 ENSG000000000005      1476 0.4241192  64102   TNMD   X 100584802 100599885
## 3 ENSG000000000419      9276 0.4252911   8813   DPM1  20  50934867  50958555
## 4 ENSG000000000457      6883 0.4117391  57147  SCYL3   1 169849631 169894267
```



```
## 5 ENSG000000000460      5970 0.4298157  55732 Clorf112    1 169662007 169854080
## 6 ENSG000000000938      3382 0.5644589   2268      FGR      1  27612064  27635277
##   strand      biotype
## 1      -1 protein_coding
## 2       1 protein_coding
## 3      -1 protein_coding
## 4      -1 protein_coding
## 5       1 protein_coding
## 6      -1 protein_coding
##
## 1                                     tetraspanin 6 [Source:HGNC]
## 2                                     tenomodulin [Source:HGNC]
## 3 dolichyl-phosphate mannosyltransferase polypeptide 1, catalytic subunit [Source:HGNC]
## 4                                     SCY1-like, kinase-like 3 [Source:HGNC]
## 5                                     chromosome 1 open reading frame 112 [Source:HGNC]
## 6                                     FGR proto-oncogene, Src family tyrosine kinase [Source:HGNC]
```

Load the annotation file `anno_gc_vm32` in IOBR for mouse RNAseq data

```
head(anno_gc_vm32)
```

```
##           id eff_length      gc symbol      mgi_id      gene_type
## 1 ENSMUSG000000000001      3262 0.4350092  Gnai3    MGI:95773 protein_coding
## 2 ENSMUSG000000000003       902 0.3481153   Pbsn  MGI:1860484 protein_coding
## 3 ENSMUSG000000000028      3506 0.4962921  Cdc45  MGI:1338073 protein_coding
## 4 ENSMUSG000000000031      2625 0.5588571   H19    MGI:95891      lncRNA
## 5 ENSMUSG000000000037      6397 0.4377052  Scml2  MGI:1340042 protein_coding
## 6 ENSMUSG000000000049      1594 0.5050188  Apoh    MGI:88058 protein_coding
##           start      end transcript_id  ont
## 1 108014596 108053462          <NA> <NA>
## 2  76881507  76897229          <NA> <NA>
## 3 18599197 18630737          <NA> <NA>
```

```
## 4 142129262 142131886      <NA> <NA>
## 5 159865521 160041209      <NA> <NA>
## 6 108234180 108305222      <NA> <NA>
```

2.3.1 For Array data: HGU133PLUS-2 (Affymetrix)

Conduct gene annotation using `anno_hug133plus2` file; If identical gene symbols exist

```
eset<-anno_eset(eset      = eset,
                annotation = anno_hug133plus2,
                symbol     = "symbol",
                probe      = "probe_id",
                method      = "mean")
eset[1:5, 1:3]
```

```
##          GSM1523727 GSM1523728 GSM1523729
## SH3KBP1      4.327974  4.316195  4.351425
## RPL41        4.246149  4.246808  4.257940
## EEF1A1       4.293762  4.291038  4.262199
## COX2         4.250288  4.283714  4.270508
## LOC101928826 4.219303  4.219670  4.213252
```

2.3.2 For RNAseq data

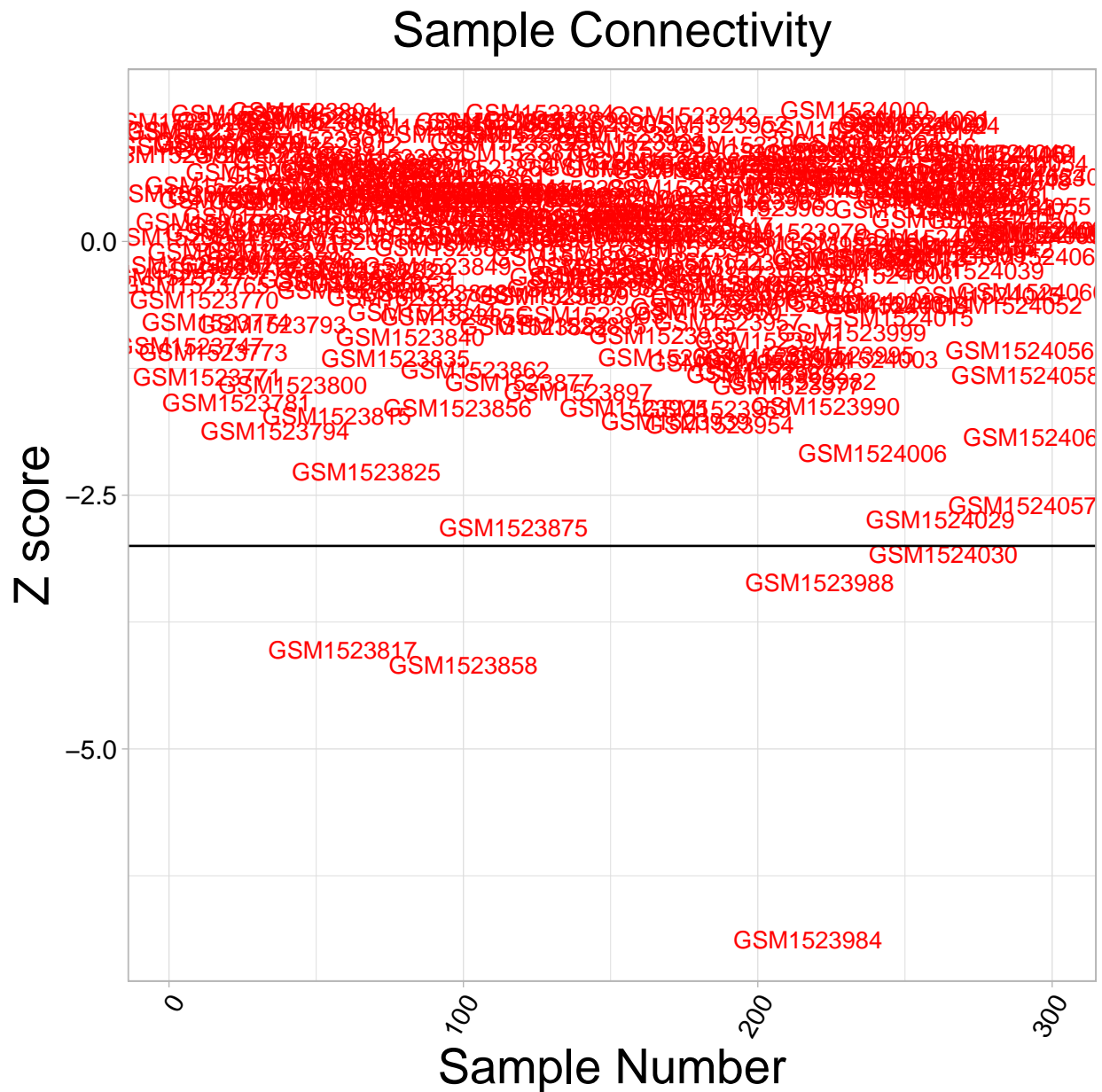
Download RNAseq data using UCSCXenaTools

Transform gene expression matrix into TPM format, and conduct subsequent annotation.

2.4 Identifying outlier samples

Take ACRG microarray data for example

```
# source("E:/18-Github/Organization/IOBR/R/find_outlier_samples.R")
res <- find_outlier_samples(eset = eset, project = "ACRG", show_plot = TRUE)
```



```
## [1] "GSM1523817" "GSM1523858" "GSM1523984" "GSM1523988" "GSM1524030"
```

```
eset1 <- eset[, !colnames(eset)%in%res]
```

```

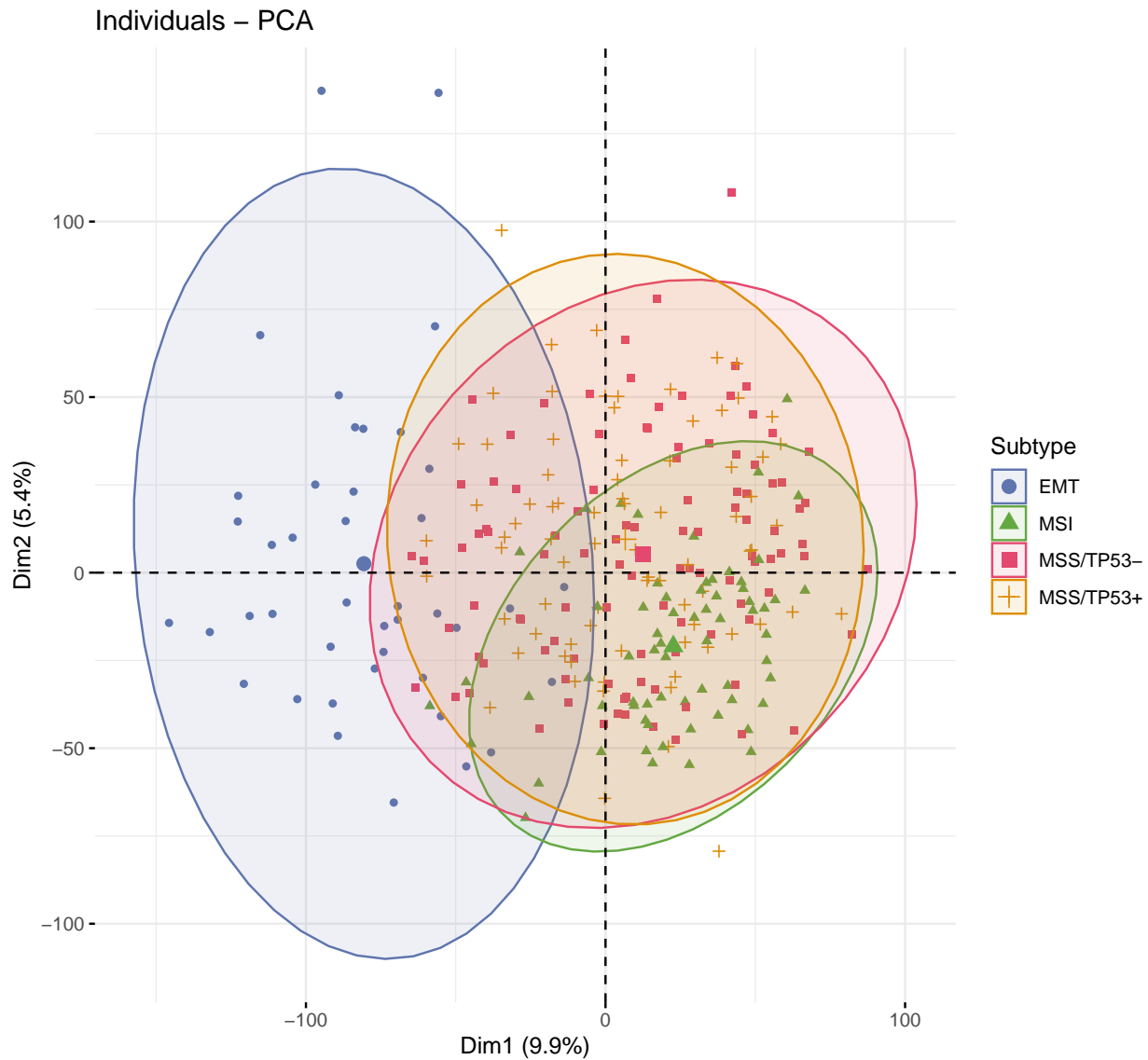
data("pdata_acrg")
res<- iobr_pca(data      = eset1,
               is.matrix = TRUE,
               scale     = TRUE,
               is.log    = FALSE,
               pdata     = pdata_acrg,
               id_pdata  = "ID",
               group     = "Subtype",
               geom.ind  = "point",
               cols      = "normal",
               palette   = "jama",
               repel     = FALSE,
               ncp       = 5,
               axes      = c(1, 2),
               addEllipses = TRUE)

```

```

##
##      CIN      EBV      EMT      GS      MSI MSS/TP53- MSS/TP53+
##      0        0      42      0      68      106      79
## [1] ">>-- colors for PCA: #5f75ae" ">>-- colors for PCA: #64a841"
## [3] ">>-- colors for PCA: #e5486e" ">>-- colors for PCA: #de8e06"
res

```



2.6 Batch effect correction

Obtaining another data set from GEO Gastric cancer: GSE57303 using GEOquery R package.

```
# NOTE: This process may take a few minutes which depends on the internet connection s
eset_geo<-getGEO(GEO      = "GSE57303", getGPL = F, destdir = "./")
eset2    <-eset_geo[[1]]
eset2    <-exprs(eset2)
eset2[1:5,1:5]
```

##	GSM1379261	GSM1379262	GSM1379263	GSM1379264	GSM1379265
## 1007_s_at	8.34746	9.67994	8.62643	8.59301	8.63046
## 1053_at	5.07972	4.46377	5.29685	5.78983	4.33359
## 117_at	5.65558	4.48732	4.21615	5.47984	5.20816
## 121_at	5.95123	7.09056	6.19903	5.89872	5.91323
## 1255_g_at	1.66923	1.98758	1.73083	1.56687	1.63332

Annotation of genes in the expression matrix and removal of duplicate genes.

```

eset2<-anno_eset(eset      = eset2,
                  annotation = anno_hug133plus2,
                  symbol    = "symbol",
                  probe     = "probe_id",
                  method    = "mean")

eset2[1:5, 1:5]

```

##	GSM1379261	GSM1379262	GSM1379263	GSM1379264	GSM1379265
## ND4	13.1695	13.1804	13.0600	12.4544	13.0457
## ATP6	13.1433	13.0814	13.0502	12.4831	13.1168
## SH3KBP1	12.9390	13.1620	12.9773	12.8745	13.1169
## COX2	13.0184	13.0489	12.8621	12.7489	12.9732
## RPL41	13.0201	12.6034	12.7929	13.0153	12.9404

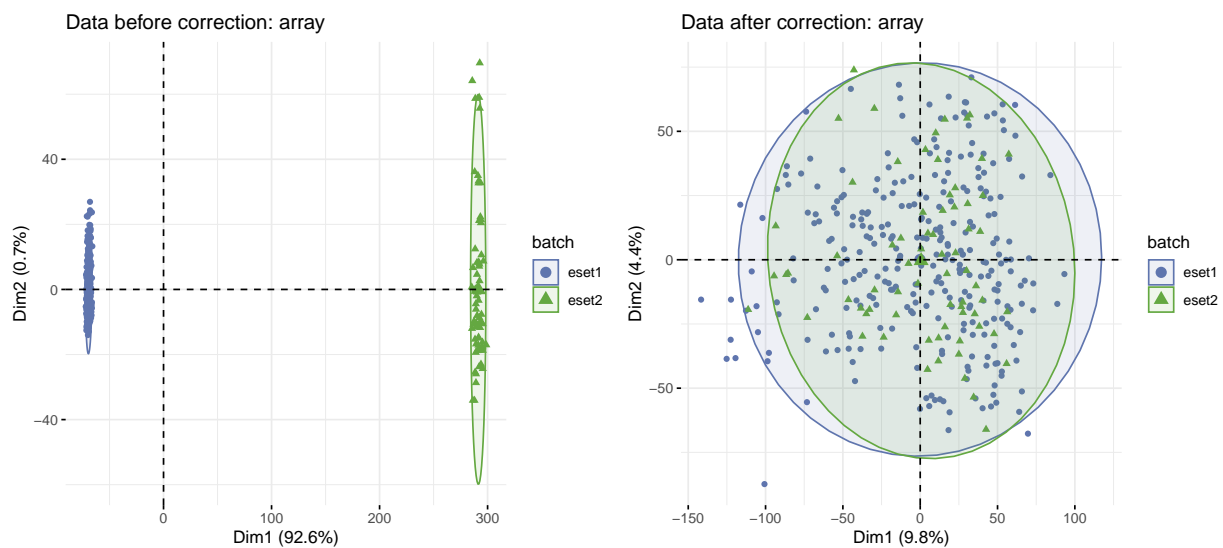
```

eset_com <- remove_batcheffect( eset1      = eset1,
                                eset2      = eset2,
                                eset3      = NULL,
                                id_type    = "symbol",
                                data_type  = "array",
                                cols       = "normal",
                                palette    = "jama",
                                log2       = TRUE,
                                check_eset = TRUE,

```

```
adjust_eset = TRUE,
repel       = FALSE,
path        = "result")
```

```
##
## eset1 eset2
##   295    70
## [1] ">>-- colors for PCA: #5f75ae" ">>-- colors for PCA: #64a841"
##
## eset1 eset2
##   295    70
## [1] ">>-- colors for PCA: #5f75ae" ">>-- colors for PCA: #64a841"
```



```
dim(eset_com)
```

```
## [1] 21752  365
```

-RNAseq count, combat-seq

2.7 References

Yuqing Zhang and others, ComBat-seq: batch effect adjustment for RNA-seq count data, NAR Genomics and Bioinformatics, Volume 2, Issue 3, September 2020, lqaa078, <https://doi.org/10.1093/nargab/lqaa078>

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, 28(6), 882-883.

Chapter 3

Tumor ecosystem analysis

3.1 Determine TME subtype of gastric cancer using TMEclassifier

TME R TMEclassifier

```
library(TMEclassifier)
tme <- tme_classifier(eset = eset1, scale = TRUE)

## Step-1: Expression data preprocessing...
## Step-2: TME deconvolution...
## Step-3: Predicting TME phenotypes...
## [23:36:59] WARNING: amalgamation/./src/learner.cc:1040:
##   If you are loading a serialized model (like pickle in Python, RDS in R) generated by
##   older XGBoost, please export the model by calling `Booster.save_model` from that version
##   first, then load it back in current version. See:
##
##   https://xgboost.readthedocs.io/en/latest/tutorials/saving_model.html
##
##   for more details about differences between saving model and serializing.
```

```
##
```

```
## [23:36:59] WARNING: amalgamation/./src/learner.cc:749: Found JSON model saved before
```

```
## >>>--- DONE!
```

```
table(tme$TMEcluster)
```

```
##
```

```
##  IA  IE  IS
```

```
## 106  94  95
```

```
head(tme)
```

```
##           ID           IE           IS           IA TMEcluster
## 1 GSM1523727 0.21034596 0.1089890 0.68066503           IA
## 2 GSM1523728 0.01008306 0.1120120 0.87790490           IA
## 3 GSM1523729 0.85603729 0.1119151 0.03204757           IE
## 4 GSM1523744 0.05645817 0.0770771 0.86646473           IA
## 5 GSM1523745 0.05785239 0.8128097 0.12933787           IS
## 6 GSM1523746 0.55697476 0.3827613 0.06026396           IE
```

```
table(tme$TMEcluster)
```

```
##
```

```
##  IA  IE  IS
```

```
## 106  94  95
```

```
head(tme)
```

```
##           ID           IE           IS           IA TMEcluster
## 1 GSM1523727 0.21034596 0.1089890 0.68066503           IA
## 2 GSM1523728 0.01008306 0.1120120 0.87790490           IA
## 3 GSM1523729 0.85603729 0.1119151 0.03204757           IE
## 4 GSM1523744 0.05645817 0.0770771 0.86646473           IA
## 5 GSM1523745 0.05785239 0.8128097 0.12933787           IS
## 6 GSM1523746 0.55697476 0.3827613 0.06026396           IE
```

3.2 DEG analysis

IA IE

```
pdata <- tme[!tme$TMEcluster=="IS", ]
deg <- iobr_deg(eset      = eset,
               annoation  = NULL,
               pdata      = pdata,
               group_id    = "TMEcluster",
               pdata_id    = "ID",
               array       = TRUE,
               method      = "limma",
               contrast    = c("deg_group", "IA", "IE"),
               path        = NULL,
               padj_cutoff = 0.01,
               logfc_cutoff = 0.5)
```

```
##
## Attaching package: 'limma'

## The following object is masked from 'package:BiocGenerics':
##
##      plotMA

## group1 = IA

## group2 = IE

## # A tibble: 6 x 11
##   symbol log2FoldChange AveExpr      t    pvalue      padj      B sigORnot label
##   <chr>          <dbl>   <dbl> <dbl>    <dbl>    <dbl> <dbl> <chr>    <chr>
## 1 TMEM100        -0.755    1.83 -14.0 1.54e-31 3.35e-27 60.9 Down_regul~ Both
## 2 ABCA8          -0.923    1.89 -12.9 3.18e-28 3.46e-24 53.4 Down_regul~ Both
## 3 HHIP           -0.625    1.74 -12.4 1.52e-26 1.10e-22 49.6 Down_regul~ Both
```

```
## 4 ADH1B          -0.901    1.84 -12.2 4.71e-26 2.09e-22 48.5 Down_regul~ Both
## 5 LMNB2           0.285    2.26  12.2 4.80e-26 2.09e-22 48.5 NOT          Sign~
## 6 FCER1A          -0.552    1.57 -12.0 2.34e-25 7.27e-22 46.9 Down_regul~ Both
## # i 2 more variables: IA <dbl>, IE <dbl>
```

3.3 GSEA analysis of DEGs

IOBR signature collection

```
head(deg)
```

```
## # A tibble: 6 x 11
##   symbol log2FoldChange AveExpr      t  pvalue      padj      B sigORnot label
##   <chr>          <dbl>   <dbl> <dbl>    <dbl>    <dbl> <dbl> <chr>      <chr>
## 1 TMEM100        -0.755    1.83 -14.0 1.54e-31 3.35e-27 60.9 Down_regul~ Both
## 2 ABCA8          -0.923    1.89 -12.9 3.18e-28 3.46e-24 53.4 Down_regul~ Both
## 3 HHIP           -0.625    1.74 -12.4 1.52e-26 1.10e-22 49.6 Down_regul~ Both
## 4 ADH1B          -0.901    1.84 -12.2 4.71e-26 2.09e-22 48.5 Down_regul~ Both
## 5 LMNB2           0.285    2.26  12.2 4.80e-26 2.09e-22 48.5 NOT          Sign~
## 6 FCER1A          -0.552    1.57 -12.0 2.34e-25 7.27e-22 46.9 Down_regul~ Both
## # i 2 more variables: IA <dbl>, IE <dbl>
```

```
sig_list <- signature_collection[c("TMEscoreB_CIR", "TMEscoreA_CIR", "DNA_replication",
                                   "Pan_F_TBRs", "TGFb.myCAF", "Ferroptosis", "TLS_Natur
sig_list
```

```
## $TMEscoreB_CIR
##   [1] "DCN"          "SEPP1"          "ACTA2"          "SPARCL1"        "BEX3"
##   [6] "MYLK"         "AKR1C1"         "TIMP2"          "MXRA7"          "C11orf96"
##  [11] "CAV1"         "PDGFRA"         "FHL1"           "MGP"            "EID1"
##  [16] "LOC101930400" "DST"            "GREM1"          "FERMT2"         "TNC"
##  [21] "CYBRD1"       "LTBP1"          "ACTG2"          "TMEM47"         "SERPINE2"
```

##	[26]	"ANTXR2"	"GNG11"	"TAGLN"	"GSTA4"	"PKIG"
##	[31]	"MAOA"	"PTRF"	"FAM3B"	"PBX1"	"WLS"
##	[36]	"SELM"	"SVIL"	"MYH11"	"AGT"	"SPON1"
##	[41]	"TGFB1I1"	"PDLIM3"	"PDK4"	"SYNP02"	"MSRB3"
##	[46]	"PROS1"	"EDNRA"	"AKAP12"	"PSD3"	"TNS1"
##	[51]	"JAM3"	"PDZRN3"	"DDR2"	"HMGCS2"	"SGCE"
##	[56]	"MRVI1"	"WFDC1"	"FBLN1"	"FM05"	"MAOB"
##	[61]	"AMOTL1"	"AKT3"	"CNRIP1"	"CPE"	"MAP1B"
##	[66]	"RBP1"	"GNAI1"	"FOXF2"	"SORBS2"	"ZCCHC24"
##	[71]	"ZNF704"	"ARMCX1"	"DIXDC1"	"SSTR1"	"THRB"
##	[76]	"C3orf70"	"PKIB"	"CNN1"	"SYTL5"	"DACT1"
##	[81]	"SYNP0"	"GAS1"	"DPYSL3"	"CCDC80"	"TSPYL5"
##	[86]	"DCHS1"	"SOBP"	"AOC3"	"NDN"	"FGF7P3"
##	[91]	"SMAD9"	"MCC"	"CLMP"	"MYL9"	"RBP4"
##	[96]	"PLN"	"SPOCK1"	"COL14A1"	"CRYAB"	"SRPX"
##	[101]	"EML1"	"RERG"	"PPP1R3C"	"LOC100506718"	"CH25H"
##	[106]	"HSPB8"	"PID1"	"TTC28"	"STON1"	"ABCG2"
##	[111]	"ZSCAN18"	"SCIN"	"C14orf132"	"TMEM55A"	"WASF3"
##	[116]	"PAPLN"	"COLEC12"	"ACKR1"	"TMEM150C"	"RAI2"
##	[121]	"TSPAN7"	"MRGPRF"	"ABCA8"	"CHIC1"	"NBEA"
##	[126]	"FAM13C"	"SETBP1"	"LDOC1"	"TMEM100"	"LOC101930349"
##	[131]	"PRICKLE2"	"TSPAN18"	"FABP4"	"ARHGEF26"	"ERICH5"
##	[136]	"MYOCD"	"BEX2"	"PPP1R14A"	"FGF13"	"RUNX1T1"
##	[141]	"MAGI2-AS3"	"LINC01279"	"REEP1"	"PLAC9"	"MYEF2"
##	[146]	"PRKD1"	"RGN"	"CLDN11"	"ANK2"	"ESRRG"
##	[151]	"SYNC"	"ZNF667-AS1"	"FGF7"	"SFRP1"	"HMCN1"
##	[156]	"TCEAL7"	"OGN"	"MAGI2"	"MIR100HG"	"FILIP1"
##	[161]	"LOC100507334"	"ANKRD6"	"PLEKHH2"	"ZNF542P"	"ARMCX4"
##	[166]	"NOV"	"DCLK1"	"ARHGAP28"	"C2orf40"	"TRHDE"
##	[171]	"EPHA7"	"SCRG1"	"ZNF677"	"ZFPM2"	"PEG3"

```

## [176] "SERP2"          "ZNF415"          "MAMDC2"          "RBM24"          "MEOX2"
##
## $TMEscoreA_CIR
##  [1] "HLA-DPB1"          "UBD"             "LOC100509457"    "WARS"
##  [5] "TAP1"              "HLA-DMA"         "TRIM22"          "PSAT1"
##  [9] "CXCL10"            "SOCS3"           "CXCL9"           "PBK"
## [13] "CCL4"              "CCL5"            "BCL2A1"          "TRBC1"
## [17] "ID01"              "NFE2L3"          "CCL3L3"          "DTL"
## [21] "MMP9"              "SLC2A3"          "ZNF367"          "RCC1"
## [25] "STIL"              "TRAC"            "HELLS"           "GZMB"
## [29] "RTEL1-TNFRSF6B"    "CXCL11"          "GBP5"            "CD2"
## [33] "CDCA2"             "CDT1"            "TNFAIP2"          "TYMP"
## [37] "MICB"              "SLC2A14"         "GZMK"            "CD8A"
## [41] "CENPH"             "MND1"            "BATF2"           "BRIP1"
## [45] "E2F7"              "KIF18A"          "AIM2"            "ETV7"
## [49] "ITK"               "GNLY"            "GPR171"          "WDHD1"
## [53] "GBP4"              "MB21D1"          "NLRP3"           "MCEMP1"
## [57] "POLR3G"            "NLRC3"           "KLRC2"           "CLEC5A"
## [61] "ARHGAP11A"         "GPR84"           "IFNG"            "ZBED2"
##
## $DNA_replication
##  [1] "RNASEH2A" "POLD3"  "DNA2"   "FEN1"   "POLA2"  "RNASEH1"
##  [7] "RPA4"      "LIG1"   "MCM2"   "MCM3"   "MCM4"   "MCM5"
## [13] "MCM6"      "MCM7"   "PCNA"   "POLE3"  "POLA1"  "POLD1"
## [19] "POLD2"     "POLE"   "POLE2"  "PRIM1"  "PRIM2"  "POLE4"
## [25] "POLD4"     "RFC1"   "RFC2"   "RFC3"   "RFC4"   "RFC5"
## [31] "RPA1"      "RPA2"   "RPA3"   "SSBP1"  "RNASEH2B" "RNASEH2C"
##
## $Base_excision_repair
##  [1] "PARP2" "PARP3" "POLD3" "PARP1" "PARP4" "FEN1"  "SMUG1" "NEIL2" "APEX2"

```

```

## [10] "POLL" "HMGB1" "APEX1" "LIG1" "LIG3" "MPG" "MUTYH" "NTHL1" "OGG1"
## [19] "PCNA" "POLE3" "POLB" "POLD1" "POLD2" "POLE" "POLE2" "NEIL3" "POLE4"
## [28] "POLD4" "UNG" "XRCC1" "NEIL1" "MBD4"
##
## $Pan_F_TBRs
## [1] "ACTA2" "ACTG2" "ADAM12" "ADAM19" "CNN1" "COL4A1"
## [7] "CTGF" "CTPS1" "FAM101B" "FSTL3" "HSPB1" "IGFBP3"
## [13] "PXDC1" "SEMA7A" "SH3PXD2A" "TAGLN" "TGFB1" "TNS1"
## [19] "TPM1"
##
## $TGFB.myCAF
## [1] "CST1" "LAMP5" "LOXL1" "EDNRA" "TGFB1" "TGFB3" "TNN"
## [8] "CST2" "HES4" "COL10A1" "ELN" "THBS4" "NKD2" "OLFM2"
## [15] "COL6A3" "LRRRC17" "COL3A1" "THY1" "HTRA3" "TMEM204" "11-Sep"
## [22] "COMP" "TNFAIP6" "ID4" "GGT5" "INAFM1" "CILP" "OLFML2B"
##
## $Ferroptosis
## [1] "ACSL4" "AKR1C1-3" "ALOXs" "ATP5G3" "CARS"
## [6] "CBS" "CD44v" "CHAC1" "CISD1" "CS"
## [11] "DPP4" "FANCD2" "GCLC/GCLM" "GLS2" "GPX4"
## [16] "GSS" "HMGCR" "HSPB1/5" "KOD" "LPCAT3"
## [21] "MT1G" "NCOA4" "NFE2L2" "PTGS2" "RPL8"
## [26] "SAT1" "SLC7A11" "SQS" "TFRC" "TP53"
## [31] "TTC35/EMC2" "MESH1"
##
## $TLS_Nature
## [1] "CD79B" "CD1D" "CCR6" "LAT" "SKAP1" "CETP" "EIF1AY" "RBP5"
## [9] "PTGDS"
##
## $Glycolysis

```

```
## [1] "ACSS1" "ACSS2" "ADH1A" "ADH1B" "ADH1C" "ADH4" "ADH5"
## [8] "ADH6" "ADH7" "ADPGK" "AKR1A1" "ALDH1A3" "ALDH1B1" "ALDH2"
## [15] "ALDH3A1" "ALDH3A2" "ALDH3B1" "ALDH3B2" "ALDH7A1" "ALDH9A1" "ALDOA"
## [22] "ALDOB" "ALDOC" "BPGM" "DLAT" "DLD" "ENO1" "ENO2"
## [29] "ENO3" "FBP1" "FBP2" "G6PC" "G6PC2" "GALM" "GAPDH"
## [36] "GAPDHS" "GCK" "GPI" "HK1" "HK2" "HK3" "HKDC1"
## [43] "LDHA" "LDHAL6A" "LDHAL6B" "LDHB" "LDHC" "PANK1" "PCK1"
## [50] "PCK2" "PDHA1" "PDHA2" "PDHB" "PFKFB1" "PFKFB2" "PFKFB3"
## [57] "PFKFB4" "PFKL" "PFKM" "PFKP" "PGAM1" "PGAM2" "PGAM4"
## [64] "PGK1" "PGK2" "PGM1" "PGM2" "PKLR" "PKM" "SLC2A2"
## [71] "TPI1"
```

3.4 DEG analysis

```
find_markers_in_bulk TME
```

```
library(Seurat)
res <- find_markers_in_bulk(pdata      = tme,
                           eset        = eset,
                           group       = "TMEcluster",
                           nfeatures   = 2000,
                           top_n       = 20,
                           thresh.use  = 0.15,
                           only.pos    = TRUE,
                           min.pct     = 0.10)
```

```
##
## IA IE IS
## 106 94 95
## # A tibble: 56 x 7
```



```
## # Groups:   cluster [3]
```

```
##      p_val avg_log2FC pct.1 pct.2 p_val_adj cluster gene
##      <dbl>      <dbl> <dbl> <dbl>      <dbl> <fct>   <chr>
##  1 4.66e-19      0.208     1     1 1.01e-14 IA     IFNG
##  2 7.41e-18      0.170     1     1 1.61e-13 IA     CXCL10
##  3 1.21e-15      0.177     1     1 2.62e-11 IA     GZMB
##  4 1.34e-15      0.247     1     1 2.91e-11 IA     CXCL11
##  5 5.33e-15      0.166     1     1 1.16e-10 IA     CXCL9
##  6 6.17e-15      0.158     1     1 1.34e-10 IA     POLR3G
##  7 3.06e-14      0.214     1     1 6.66e-10 IA     IDO1
##  8 1.95e-13      0.179     1     1 4.24e- 9 IA     GBP4
##  9 4.17e-13      0.150     1     1 9.07e- 9 IA     ZBED2
## 10 3.74e-11      0.153     1     1 8.13e- 7 IA     GNLY
## # i 46 more rows
```

```
top15 <- res$top_markers %>% dplyr:: group_by(cluster) %>% dplyr::top_n(15, avg_log2F
top15$gene
```

```
##  [1] "IFNG"          "CXCL10"        "GZMB"          "CXCL11"
##  [5] "CXCL9"         "POLR3G"        "IDO1"          "GBP4"
##  [9] "GNLY"         "PLEKHS1"       "KLRC2"         "VSNL1"
## [13] "AIM2"         "SLC01B3"       "COL11A1"       "TMEM100"
## [17] "ADH1B"        "ABCA8"         "MAMDC2"        "SCN7A"
## [21] "C7"          "C2orf40"       "LIPF"          "PGA4"
## [25] "SCGB2A1"      "FUT9"         "GKN1"          "GKN2"
## [29] "OGN"         "GIF"          "IL1A"          "EREG"
## [33] "PPBP"        "IL11"         "CXCL6"         "PI15"
## [37] "TNFRSF11B"    "PROK2"        "CLEC5A"        "MAGEA10-MAGEA5"
## [41] "MAGEA4"       "MAGEA12"      "MAGEA6"        "MAGEA2B"
## [45] "REG1B"
```

Seurat DoHeatmap

```
#
cols <- c('#2692a4', '#fc0d3a', '#ffbe0b')
p1 <- DoHeatmap(res$sce, top15$gene, group.colors = cols )+
  scale_fill_gradientn(colours = rev(colorRampPalette(RColorBrewer::brewer.pal(11,"RdBu
```

TME

```
input <- combine_pd_eset(eset = eset, pdata = tme, fea = top15$gene, scale = T)

p2 <- sig_box(input, variable = "TMEcluster", signature = "IFNG", jitter = TRUE,
  cols = cols, show_pvalue = TRUE, size_of_pvalue = 4)
```

```
## # A tibble: 3 x 8
```

	.y.	group1	group2	p	p.adj	p.format	p.signif	method
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>	<chr>
## 1	signature	IA	IE	7.08e-16	2.10e-15	7.1e-16	****	Wilcoxon
## 2	signature	IA	IS	5.12e-13	1 e-12	5.1e-13	****	Wilcoxon
## 3	signature	IE	IS	9.36e- 2	9.4 e- 2	0.094	ns	Wilcoxon

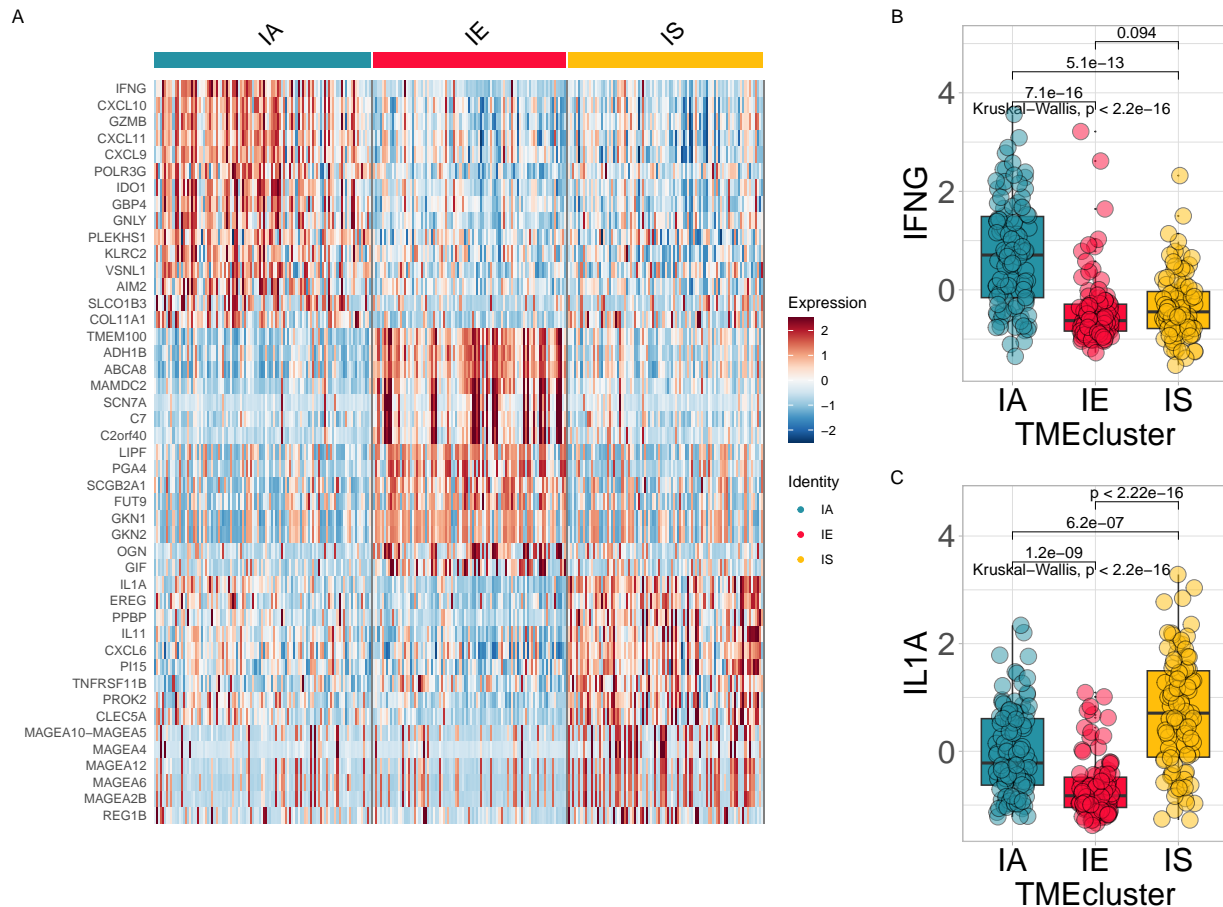
```
p3 <- sig_box(input, variable = "TMEcluster", signature = "IL1A",
  jitter = TRUE, cols = cols, show_pvalue = TRUE, size_of_pvalue = 4)
```

```
## # A tibble: 3 x 8
```

	.y.	group1	group2	p	p.adj	p.format	p.signif	method
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>	<chr>
## 1	signature	IA	IE	1.17e- 9	2.3 e- 9	1.2e-09	****	Wilcoxon
## 2	signature	IA	IS	6.19e- 7	6.20e- 7	6.2e-07	****	Wilcoxon
## 3	signature	IE	IS	4.51e-19	1.4 e-18	< 2e-16	****	Wilcoxon

```
if (!requireNamespace("patchwork", quietly = TRUE)) install.packages("patchwork")
library(patchwork)
p <- (p1|p2/p3) + plot_layout(widths = c(2.3,1))
```

```
p + plot_annotation(tag_levels = 'A')
```



3.5 Identifying signatures associated with TME clusters

Calculate TME associated signatures-(through PCA method).

```
sig_tme <- calculate_sig_score(pdata = NULL,
                              eset = eset,
                              signature = signature_collection,
                              method = "pca",
                              mini_gene_count = 2)

sig_tme <- t(column_to_rownames(sig_tme, var = "ID"))
```

```
sig_tme[1:5, 1:3]
```

```
##          GSM1523727 GSM1523728 GSM1523729
## CD_8_T_effector -2.5513794  0.7789141 -2.1770675
## DDR            -0.8747614  0.7425162 -1.3272054
## APM            1.1098368  2.1988688 -0.9516419
## Immune_Checkpoint -2.3701787  0.9455120 -1.4844104
## CellCycle_Reg   0.1063358  0.7583302 -0.3649795
```

TMEcluster

```
res <- find_markers_in_bulk(pdata = tme, eset = sig_tme, group = "TMEcluster", nfeatures
```

```
##
##  IA  IE  IS
## 106  94  95
## # A tibble: 60 x 7
## # Groups:   cluster [3]
##      p_val avg_log2FC pct.1 pct.2 p_val_adj cluster gene
##      <dbl>      <dbl> <dbl> <dbl>      <dbl> <fct>   <chr>
##  1 2.73e-18      17.8  0.792 0.365  6.98e-16 IA      Base-excision-repair
##  2 8.18e-18       3.64  0.698 0.323  2.09e-15 IA      IFNG-signature-Ayers-et-al
##  3 9.03e-18       8.94  0.84  0.381  2.31e-15 IA      Mismatch-Repair
##  4 9.73e-18       4.17  0.821 0.402  2.49e-15 IA      Th2-cells-Bindea-et-al
##  5 1.20e-16       3.97  0.821 0.381  3.08e-14 IA      Folate-One-Carbon-Metaboli~
##  6 1.14e-15       4.36  0.811 0.354  2.91e-13 IA      Homologous-recombination
##  7 2.13e-14       3.09  0.83  0.397  5.45e-12 IA      Purine-Biosynthesis
##  8 3.61e-14       4.20  0.67  0.28  9.25e-12 IA      CD-8-T-effector
##  9 1.49e-13       2.70  0.708 0.344  3.82e-11 IA      Th1-cells-Bindea-et-al
## 10 7.03e-13       2.61  0.792 0.429  1.80e-10 IA      TIP-Release-of-cancer-cell~
## # i 50 more rows
```

```
top15 <- res$top_markers %>% dplyr:: group_by(cluster) %>% dplyr::top_n(15, avg_log2F
```

```
p1 <- DoHeatmap(res$sce, top15$gene, group.colors = cols)+
  scale_fill_gradientn(colours = rev(colorRampPalette(RColorBrewer::brewer.pal(11,"RdBu
```

```
top15$gene <- gsub(top15$gene, pattern = "\\-", replacement = "\\_")
input <- combine_pd_eset(eset = sig_tme, pdata = tme, feas = top15$gene, scale = T)

p2 <- sig_box(input, variable = "TMEcluster", signature = "IFNG_signature_Ayers_et_al",
  cols = cols, show_pvalue = TRUE, size_of_pvalue = 4, size_of_font = 6)
```

```
## # A tibble: 3 x 8
```

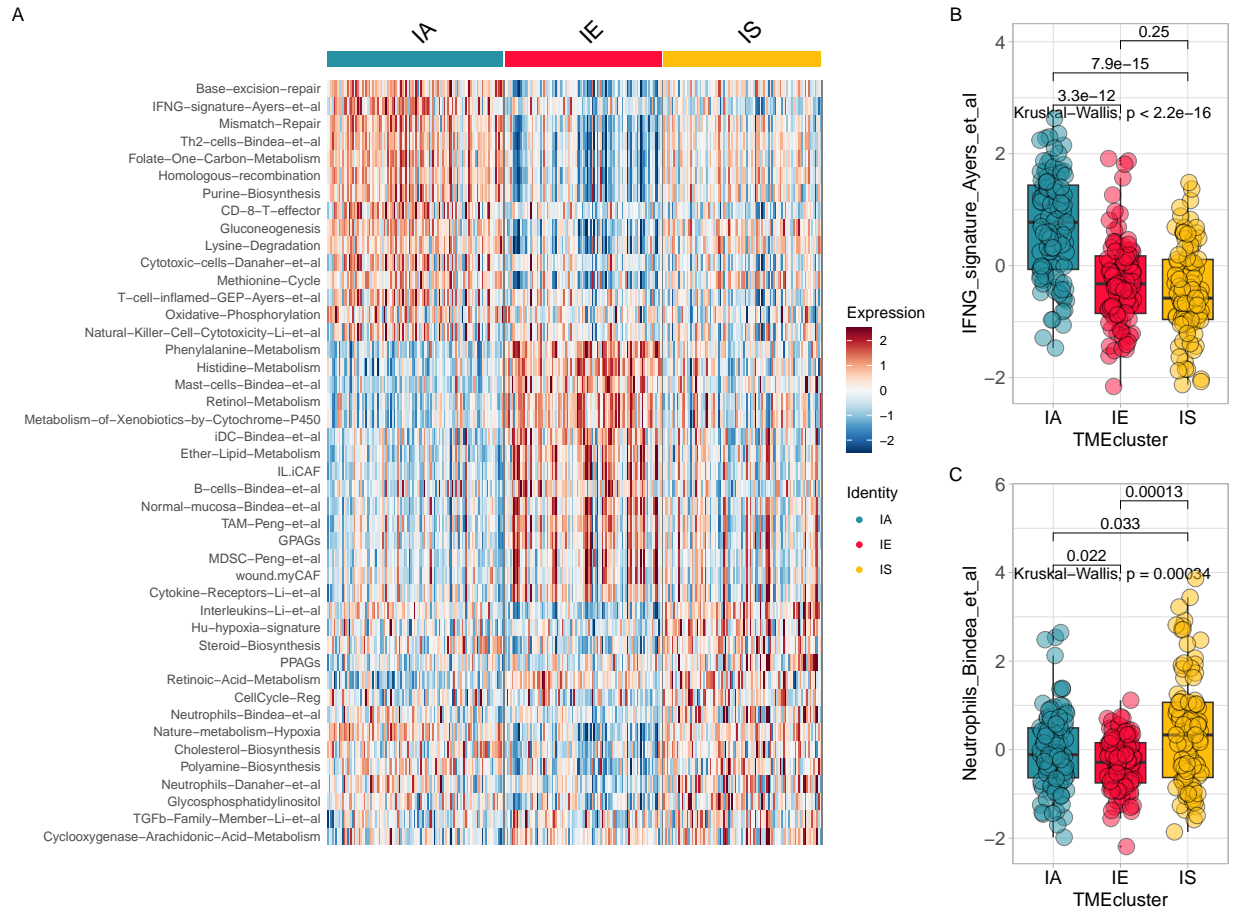
```
##   .y.      group1 group2      p    p.adj p.format p.signif method
##   <chr>    <chr> <chr>    <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 signature IA    IE    3.34e-12 6.70e-12 3.3e-12 ****    Wilcoxon
## 2 signature IA    IS    7.93e-15 2.40e-14 7.9e-15 ****    Wilcoxon
## 3 signature IE    IS    2.54e- 1 2.5 e- 1 0.25     ns      Wilcoxon
```

```
p3 <- sig_box(input, variable = "TMEcluster", signature = "Neutrophils_Bindea_et_al",
  jitter = TRUE, cols = cols, show_pvalue = TRUE, size_of_pvalue = 4, size_
```

```
## # A tibble: 3 x 8
```

```
##   .y.      group1 group2      p    p.adj p.format p.signif method
##   <chr>    <chr> <chr>    <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 signature IA    IE    0.0222   0.044   0.02217 *      Wilcoxon
## 2 signature IA    IS    0.0333   0.044   0.03326 *      Wilcoxon
## 3 signature IE    IS    0.000133 0.0004 0.00013 ***     Wilcoxon
```

```
p <- (p1|p2/p3) + plot_layout(widths = c(2.3,1))
p + plot_annotation(tag_levels = 'A')
```



```
library(survminer)
```

```
##
```

```
## Attaching package: 'survminer'
```

```
## The following object is masked from 'package:survival':
```

```
##
```

```
## myeloma
```

```
data(pdata_acrg, package = "IOBR")
```

```
input <- merge(pdata_acrg, input, by = "ID")
```

```
p1<-surv_group(input_pdata      = input,
                target_group     = "TMEcluster",
                ID                = "ID",
                reference_group   = "High",
```

```
project      = "ACRG",  
cols         = cols,  
time         = "OS_time",  
status       = "OS_status",  
time_type    = "month",  
save_path    = "result")
```

```
## >>> Dataset's survival follow up time is range between 1 to 105.7 months
```

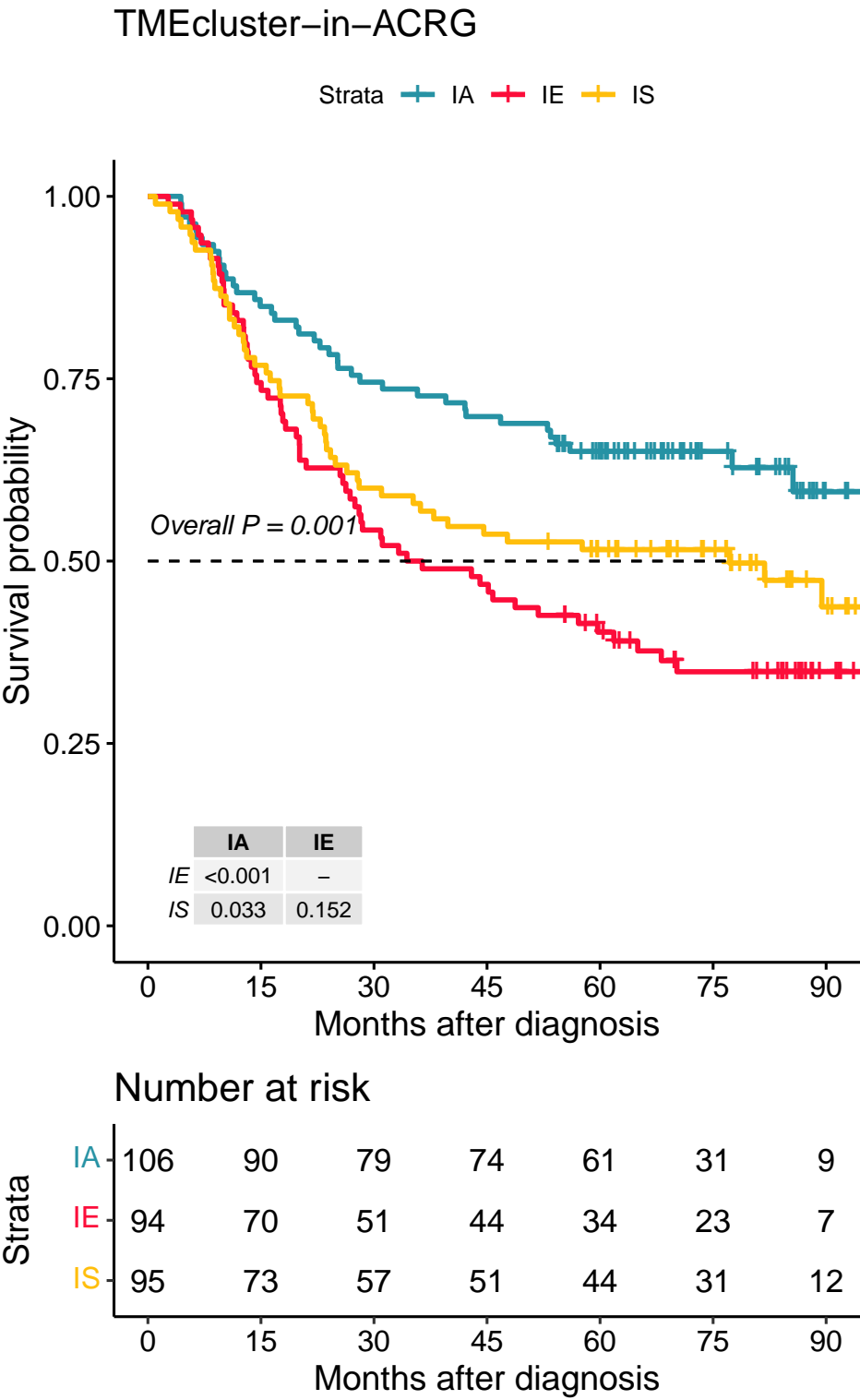
```
##  IA  IE  IS
```

```
## 106 94 95
```

```
## 1069495
```

```
## Maximum of follow up time is 105.7 months; and will be divided into 6 sections;
```

```
p1
```



```
p1<- percent_bar_plot(input, x = "TMEcluster" , y = "Subtype", palette = "jama")
```

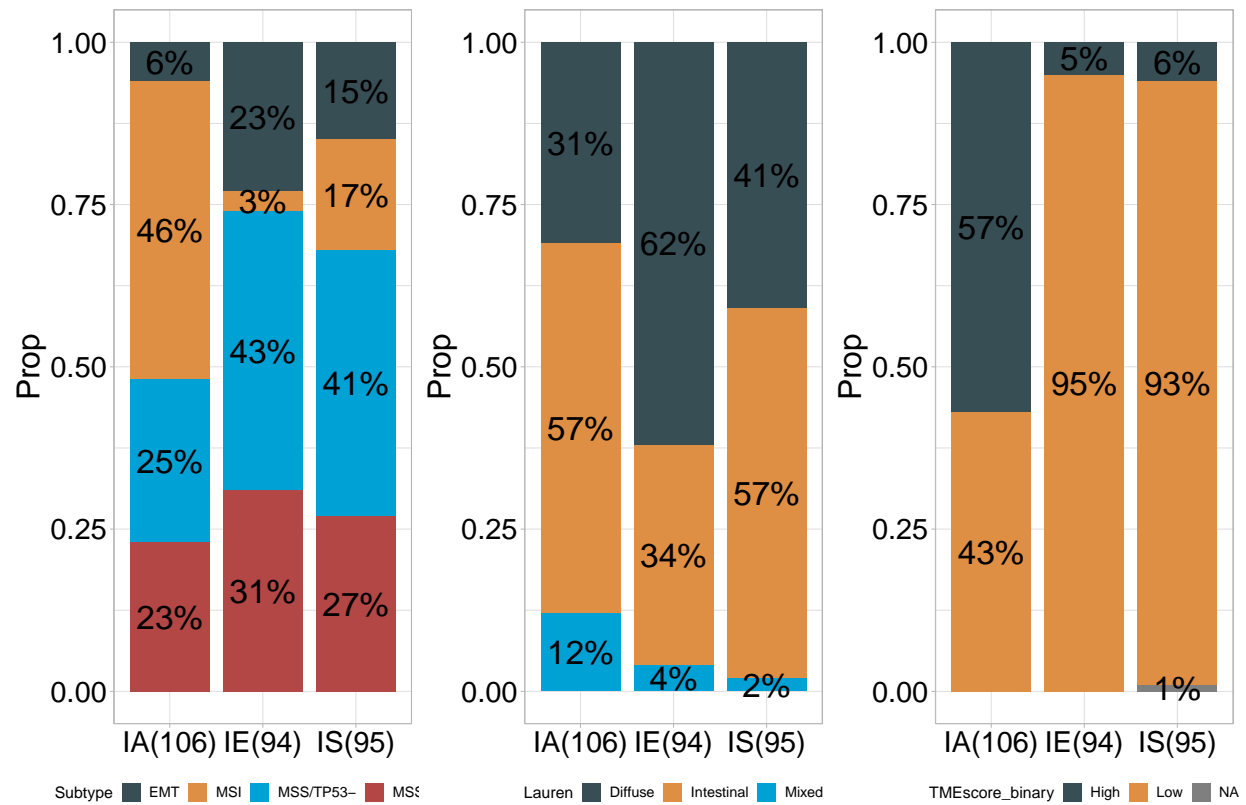
A tibble: 12 x 5


```
## # Groups:   TMEcluster [3]
##   TMEcluster Subtype      Freq  Prop count
##   <chr>      <fct>      <dbl> <dbl> <dbl>
##  1 IA        EMT          6  0.06  106
##  2 IA        MSI          49  0.46  106
##  3 IA        MSS/TP53-    27  0.25  106
##  4 IA        MSS/TP53+    24  0.23  106
##  5 IE        EMT          22  0.23   94
##  6 IE        MSI          3  0.03   94
##  7 IE        MSS/TP53-    40  0.43   94
##  8 IE        MSS/TP53+    29  0.31   94
##  9 IS        EMT          14  0.15   95
## 10 IS        MSI          16  0.17   95
## 11 IS        MSS/TP53-    39  0.41   95
## 12 IS        MSS/TP53+    26  0.27   95
## [1] "'#374E55FF', '#DF8F44FF', '#00A1D5FF', '#B24745FF', '#79AF97FF', '#6A6599FF', '#
p2<- percent_bar_plot(input, x = "TMEcluster" , y = "Lauren", palette = "jama")
```

```
## # A tibble: 9 x 5
## # Groups:   TMEcluster [3]
##   TMEcluster Lauren      Freq  Prop count
##   <chr>      <fct>      <dbl> <dbl> <dbl>
##  1 IA        Diffuse      33  0.31  106
##  2 IA        Intestinal    60  0.57  106
##  3 IA        Mixed        13  0.12  106
##  4 IE        Diffuse      58  0.62   94
##  5 IE        Intestinal    32  0.34   94
##  6 IE        Mixed         4  0.04   94
##  7 IS        Diffuse      39  0.41   95
##  8 IS        Intestinal    54  0.57   95
```

```
## 9 IS          Mixed          2 0.02    95
## [1] "'#374E55FF', '#DF8F44FF', '#00A1D5FF', '#B24745FF', '#79AF97FF', '#6A6599FF', '#
p3<- percent_bar_plot(input, x = "TMEcluster" , y = "TMEscore_binary", palette = "jama"
```

```
## # A tibble: 7 x 5
## # Groups:   TMEcluster [3]
##   TMEcluster TMEscore_binary Freq Prop count
##   <chr>      <fct>          <dbl> <dbl> <dbl>
## 1 IA        High           60 0.57  106
## 2 IA        Low            46 0.43  106
## 3 IE        High            5 0.05   94
## 4 IE        Low            89 0.95   94
## 5 IS        High            6 0.06   95
## 6 IS        Low            88 0.93   95
## 7 IS        <NA>            1 0.01   95
## [1] "'#374E55FF', '#DF8F44FF', '#00A1D5FF', '#B24745FF', '#79AF97FF', '#6A6599FF', '#
p1|p2|p3
```



Chapter 4

Signatures and relevant phenotypes

4.1 Loading packages

Load the IOBR package in your R session after the installation is complete:

```
library(IOBR)
library(survminer)
library(tidyverse)
```

4.2 Downloading data for example

Obtaining data set from GEO Gastric cancer: GSE62254 using GEOquery R package.

```
if (!requireNamespace("GEOquery", quietly = TRUE)) BiocManager::install("GEOquery")
library("GEOquery")
# NOTE: This process may take a few minutes which depends on the internet connection
eset_geo <- getGEO(GEO = "GSE62254", getGPL = F, destdir = "./")
eset <- eset_geo[[1]]
eset <- exprs(eset)
eset[1:5, 1:5]
```

```
##          GSM1523727 GSM1523728 GSM1523729 GSM1523744 GSM1523745
## 1007_s_at  3.2176645  3.0624323  3.0279131   2.921683  2.8456013
## 1053_at   2.4050109  2.4394879  2.2442708   2.345916  2.4328582
## 117_at    1.4933412  1.8067380  1.5959665   1.839822  1.8326058
## 121_at    2.1965561  2.2812181  2.1865556   2.258599  2.1874363
## 1255_g_at 0.8698382  0.9502466  0.8125414   1.012860  0.9441993
```

Annotation of genes in the expression matrix and removal of duplicate genes.

```
# Load the annotation file `anno_hug133plus2` in IOBR.
```

```
head(anno_hug133plus2)
```

```
## # A tibble: 6 x 2
##   probe_id symbol
##   <fct>      <fct>
## 1 1007_s_at MIR4640
## 2 1053_at   RFC2
## 3 117_at    HSPA6
## 4 121_at    PAX8
## 5 1255_g_at GUCA1A
## 6 1294_at   MIR5193
```

```
# Conduct gene annotation using `anno_hug133plus2` file; If identical gene symbols exist
```

```
eset<-anno_eset(eset      = eset,
                annotation = anno_hug133plus2,
                symbol     = "symbol",
                probe      = "probe_id",
                method     = "mean")
eset[1:5, 1:3]
```

```
##          GSM1523727 GSM1523728 GSM1523729
## SH3KBP1      4.327974  4.316195  4.351425
```

## RPL41	4.246149	4.246808	4.257940
## EEF1A1	4.293762	4.291038	4.262199
## COX2	4.250288	4.283714	4.270508
## LOC101928826	4.219303	4.219670	4.213252

4.3 Signature score estimation

```
sig_tme<-calculate_sig_score(pdata      = NULL,
                             eset       = eset,
                             signature  = signature_collection,
                             method     = "pca",
                             mini_gene_count = 2)
sig_tme <- t(column_to_rownames(sig_tme, var = "ID"))
sig_tme[1:5, 1:3]
```

##	GSM1523727	GSM1523728	GSM1523729
## CD_8_T_effector	-2.5513794	0.7789141	-2.1770675
## DDR	-0.8747614	0.7425162	-1.3272054
## APM	1.1098368	2.1988688	-0.9516419
## Immune_Checkpoint	-2.3701787	0.9455120	-1.4844104
## CellCycle_Reg	0.1063358	0.7583302	-0.3649795

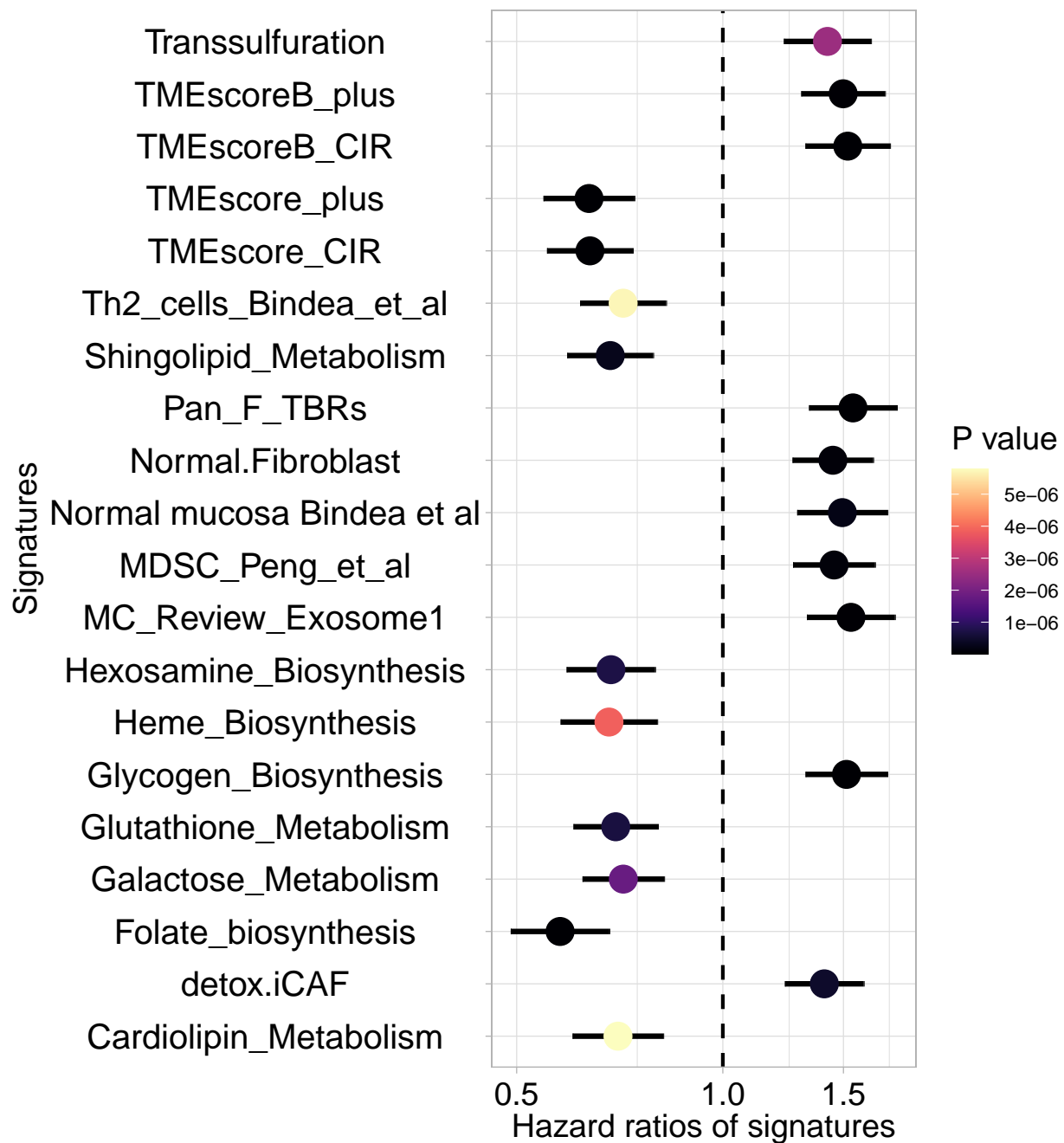
4.4 Identifying features associated with survival

```
data("pdata_acrg")
input <- combine_pd_eset(eset = sig_tme, pdata = pdata_acrg, scale = T)
res<- batch_surv(pdata      = input,
                 time       = "OS_time",
                 status     = "OS_status",
                 variable = colnames(input)[69:ncol(input)])
```

```
head(res)
```

```
## # A tibble: 6 x 5
##   ID                P    HR CI_low_0.95 CI_up_0.95
##   <chr>            <dbl> <dbl>      <dbl>      <dbl>
## 1 Folate_biosynthesis 1.00e-10 0.579      0.490      0.683
## 2 TMEscore_CIR        1.32e- 9 0.640      0.554      0.739
## 3 Glycogen_Biosynthesis 3.24e- 9 1.52      1.32      1.74
## 4 Pan_F_TBRS         6.33e- 9 1.55      1.34      1.80
## 5 TMEscoreB_CIR       7.17e- 9 1.52      1.32      1.75
## 6 TMEscore_plus       8.08e- 9 0.638      0.547      0.743
```

```
res<- res[nchar(res$ID)<=28, ]
p1<- sig_forest(res, signature = "ID", n = 20)
```

4.5 Visulization using heatmap

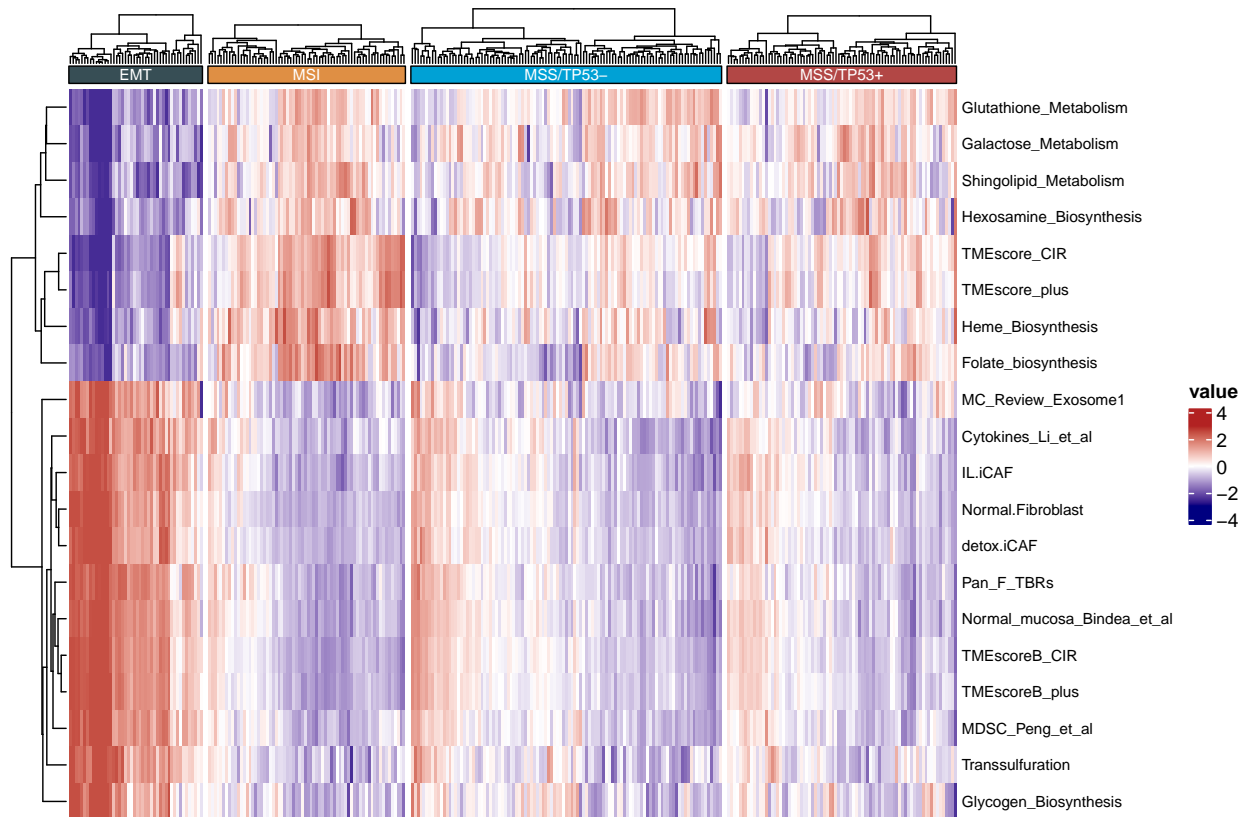
Signatures IOBR sig_heatmap

```
p2 <- sig_heatmap(input = input,
                  features = res$ID[1:20],
```

```

group          = "Subtype",
palette_group  = "jama",
palette        = 6)

```



4.6 Focus on target signatures

```

p1 <- sig_box(data          = input,
signature      = "Glycogen_Biosynthesis",
variable       = "Subtype",
jitter        = TRUE,
cols           = NULL,
palette        = "jama",
show_pvalue    = TRUE,

```

```

size_of_pvalue = 5,
hjust          = 1,
angle_x_text   = 60,
size_of_font   = 8)

```

```
## # A tibble: 6 x 8
```

	.y.	group1	group2	p	p.adj	p.format	p.signif	method
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>	<chr>
## 1	signature	EMT	MSI	5.39e-15	3.20e-14	5.4e-15	****	Wilcoxon
## 2	signature	EMT	MSS/TP53-	5.53e-13	2.8 e-12	5.5e-13	****	Wilcoxon
## 3	signature	EMT	MSS/TP53+	1.90e-12	7.6 e-12	1.9e-12	****	Wilcoxon
## 4	signature	MSI	MSS/TP53-	1.14e- 3	3.4 e- 3	0.0011	**	Wilcoxon
## 5	signature	MSI	MSS/TP53+	7.05e- 3	1.4 e- 2	0.0071	**	Wilcoxon
## 6	signature	MSS/TP53-	MSS/TP53+	7.16e- 1	7.2 e- 1	0.7161	ns	Wilcoxon

```

p2 <- sig_box(data      = input,
               signature  = "Pan_F_TBRs",
               variable   = "Subtype",
               jitter     = TRUE,
               cols       = NULL,
               palette    = "jama",
               show_pvalue = TRUE,
               angle_x_text = 60,
               hjust      = 1,
               size_of_pvalue = 5,
               size_of_font  = 8)

```

```
## # A tibble: 6 x 8
```

	.y.	group1	group2	p	p.adj	p.format	p.signif	method
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>	<chr>
## 1	signature	EMT	MSI	7.98e-17	3.20e-16	<2e-16	****	Wilcoxon

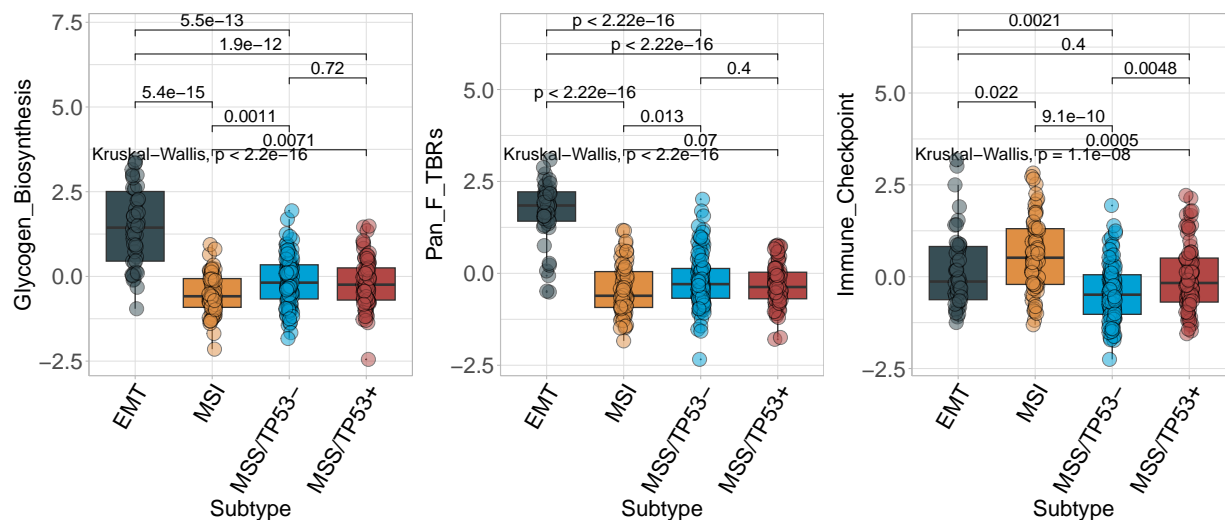
```
## 2 signature EMT      MSS/TP53- 1.70e-17 1 e-16 <2e-16 **** Wilcoxon
## 3 signature EMT      MSS/TP53+ 2.57e-17 1.3 e-16 <2e-16 **** Wilcoxon
## 4 signature MSI      MSS/TP53- 1.32e- 2 4 e- 2 0.013 * Wilcoxon
## 5 signature MSI      MSS/TP53+ 6.99e- 2 1.4 e- 1 0.070 ns Wilcoxon
## 6 signature MSS/TP53- MSS/TP53+ 4.02e- 1 4 e- 1 0.402 ns Wilcoxon
```

```
p3 <- sig_box(data      = input,
               signature  = "Immune_Checkpoint",
               variable   = "Subtype",
               jitter     = TRUE,
               cols       = NULL,
               palette     = "jama",
               show_pvalue = TRUE,
               angle_x_text = 60,
               hjust      = 1,
               size_of_pvalue = 5,
               size_of_font  = 8)
```

```
## # A tibble: 6 x 8
```

```
##   .y.      group1    group2      p      p.adj p.format p.signif method
##   <chr>    <chr>    <chr>    <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 signature EMT      MSI      2.20e- 2 0.044      0.0220 *      Wilcoxon
## 2 signature EMT      MSS/TP53- 2.11e- 3 0.0085      0.0021 **     Wilcoxon
## 3 signature EMT      MSS/TP53+ 4.03e- 1 0.4        0.4026 ns     Wilcoxon
## 4 signature MSI      MSS/TP53- 9.13e-10 0.0000000055 9.1e-10 ****   Wilcoxon
## 5 signature MSI      MSS/TP53+ 5.03e- 4 0.0025      0.0005 ***    Wilcoxon
## 6 signature MSS/TP53- MSS/TP53+ 4.82e- 3 0.014      0.0048 **     Wilcoxon
```

```
p1|p2|p3
```



4.7 Survival analysis

Signature

```
res <- sig_surv_plot(input_pdata = input,
                     signature    = "Glycogen_Biosynthesis",
                     cols         = NULL,
                     palette      = "jco",
                     project      = "ACRG",
                     time         = "OS_time",
                     status       = "OS_status",
                     time_type    = "month",
                     save_path    = "result")
```

##	ID	time	status	Glycogen_Biosynthesis	group3	group2	bestcutoff
## 1	GSM1523727	88.73	0	-0.3612213	Middle	Low	Low
## 2	GSM1523728	88.23	0	-0.6926726	Low	Low	Low
## 3	GSM1523729	88.23	0	-0.9388531	Low	Low	Low
## 4	GSM1523744	105.70	0	-1.1825136	Low	Low	Low
## 5	GSM1523745	105.53	0	-0.3034304	Middle	Low	Low


```
fig.type = "pdf",
width     = 5,
height    = 5.2)
```

```
## [1] ">>>-- Range of Time: "
```

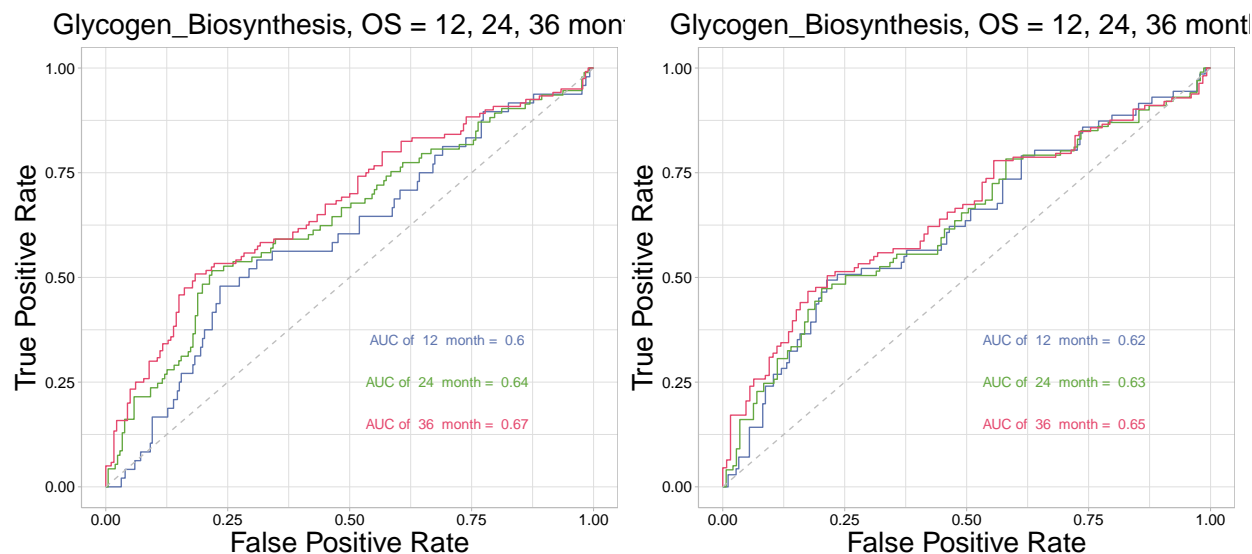
```
## [1] 1.0 105.7
```

```
p2<- roc_time(input      = input,
               vars       = "Glycogen_Biosynthesis",
               time       = "RFS_time",
               status     = "RFS_status",
               time_point = c(12, 24, 36),
               time_type  = "month",
               palette     = "jama",
               cols       = "normal",
               seed        = 1234,
               show_col    = FALSE,
               path        = "result",
               main        = "OS",
               index       = 1,
               fig.type    = "pdf",
               width       = 5,
               height      = 5.2)
```

```
## [1] ">>>-- Range of Time: "
```

```
## [1] 0.10 100.87
```

```
p1|p2
```



4.8 Batch correlation analysis

signature signatures

```
res <- batch_cor(data = input, target = "Glycogen_Biosynthesis", feature = colnames(input))
head(res)
```

```
## # A tibble: 6 x 6
```

##	sig_names	p.value	statistic	p.adj	log10pvalue	stars
##	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<fct>
## 1	TMEscoreB_CIR	8.89e-42	0.678	2.27e-39	41.1	****
## 2	Glycine__Serine_and_Threonine_M~	7.49e-40	-0.666	9.54e-38	39.1	****
## 3	Ether_Lipid_Metabolism	3.84e-39	0.662	3.27e-37	38.4	****
## 4	MDSC_Peng_et_al	1.13e-38	0.659	7.21e-37	37.9	****
## 5	Glycerophospholipid_Metabolism	8.72e-38	-0.653	4.44e-36	37.1	****
## 6	TIP_Release_of_cancer_cell_anti~	2.32e-37	-0.650	9.86e-36	36.6	****

```
p1<- get_cor(eset = sig_tme, pdata = pdata_acrg, var1 = "Glycogen_Biosynthesis", var2 =
```

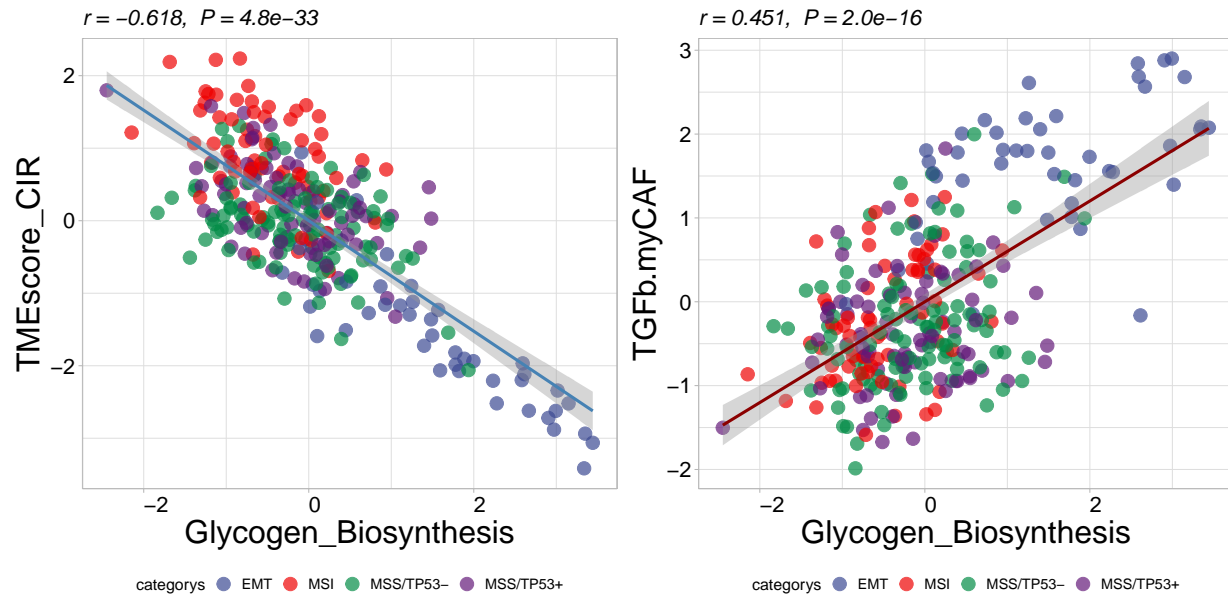
```
##
```

```
## Spearman's rank correlation rho
```

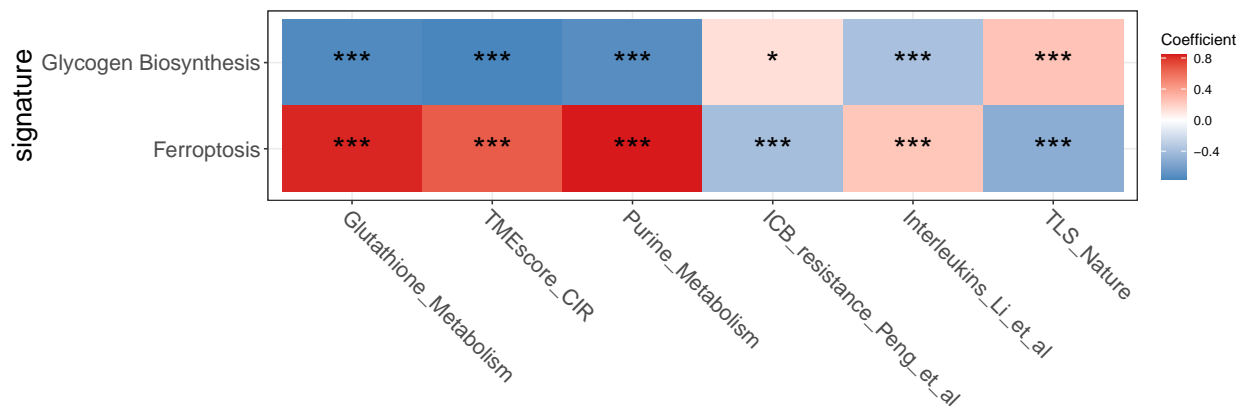


```
##
## data:  data[, var1] and data[, var2]
## S = 7282858, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##          rho
## -0.6184309
##
## [1] ">>>--- The exact p value is: 4.78971420439895e-33"
##      EMT      MSI MSS/TP53- MSS/TP53+
##      46      68      107      79
p2<- get_cor(eset = sig_tme, pdata = pdata_acrg, var1 = "Glycogen_Biosynthesis", var2 =

##
## Spearman's rank correlation rho
##
## data:  data[, var1] and data[, var2]
## S = 2471758, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##          rho
## 0.4507143
##
## [1] ">>>--- The exact p value is: 2.04505761057615e-16"
##      EMT      MSI MSS/TP53- MSS/TP53+
##      46      68      107      79
p1|p2
```



```
feas1 <- c("Glycogen_Biosynthesis", "Ferroptosis")
feas2 <- c("Glutathione_Metabolism", "TMEscore_CIR", "Purine_Metabolism", "ICB_resistance")
p <- get_cor_matrix(data = input,
                     feas1 = feas2,
                     feas2 = feas1,
                     method = "pearson",
                     font.size.star = 8,
                     font.size = 15,
                     fill_by_cor = FALSE,
                     round.num = 1)
```

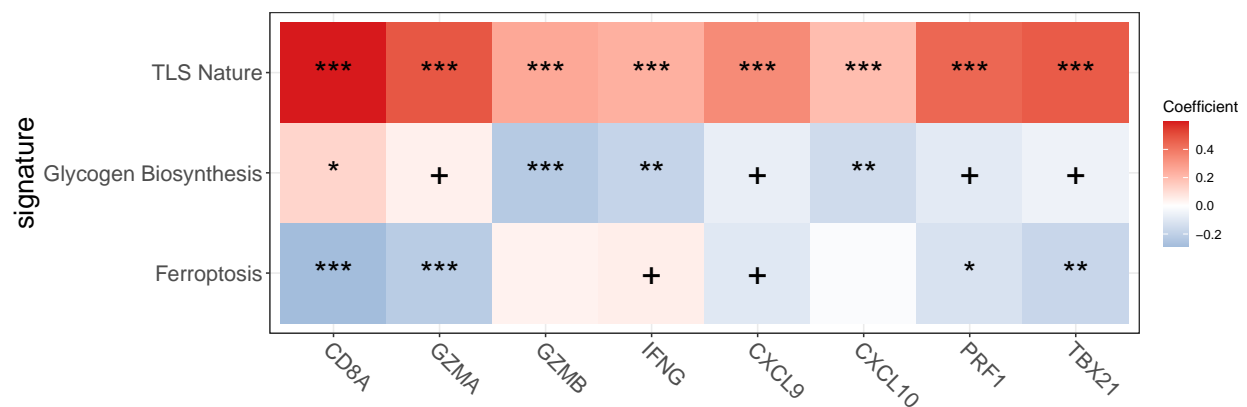


4.9 Visulization of correlations

```
input2 <- combine_pd_eset(eset = eset, pdata = input[, c("ID", "Glycogen_Biosynthesis",
feas1 <- c("Glycogen_Biosynthesis", "TLS_Nature", "Ferroptosis")
feas2 <- signature_collection$CD_8_T_effector
feas2
```

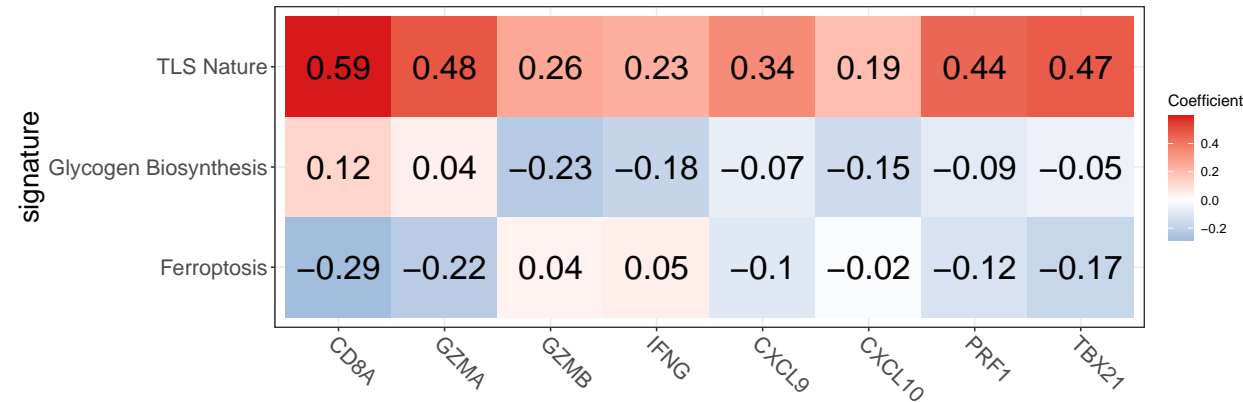
```
## [1] "CD8A" "GZMA" "GZMB" "IFNG" "CXCL9" "CXCL10" "PRF1" "TBX21"
```

```
p <- get_cor_matrix(data      = input2,
                    feas1     = feas2,
                    feas2     = feas1,
                    method    = "pearson",
                    scale     = T,
                    font.size.star = 8,
                    font.size  = 15,
                    fill_by_cor = FALSE,
                    round.num  = 1)
```



```
p <- get_cor_matrix(data      = input2,
                    feas1     = feas2,
                    feas2     = feas1,
                    method    = "pearson",
                    scale     = T,
```

```
font.size.star = 8,  
font.size      = 15,  
fill_by_cor    = TRUE,  
round.num      = 2)
```



Chapter 5

TME deconvolution

5.1 Loading packages

Load the IOBR package in your R session after the installation is complete:

```
library(IOBR)
library(survminer)
library(tidyverse)
```

5.2 Downloading data for example

Obtaining data set from GEO Gastric cancer: GSE62254 using GEOquery R package.

```
if (!requireNamespace("GEOquery", quietly = TRUE)) BiocManager::install("GEOquery")
library("GEOquery")
# NOTE: This process may take a few minutes which depends on the internet connection speed
eset_geo<-getGEO(GEO = "GSE62254", getGPL = F, destdir = "./")
eset      <-eset_geo[[1]]
eset      <-exprs(eset)
eset[1:5,1:5]
```

```
##          GSM1523727 GSM1523728 GSM1523729 GSM1523744 GSM1523745
## 1007_s_at  3.2176645  3.0624323  3.0279131   2.921683  2.8456013
## 1053_at   2.4050109  2.4394879  2.2442708   2.345916  2.4328582
## 117_at    1.4933412  1.8067380  1.5959665   1.839822  1.8326058
## 121_at    2.1965561  2.2812181  2.1865556   2.258599  2.1874363
## 1255_g_at  0.8698382  0.9502466  0.8125414   1.012860  0.9441993
```

Annotation of genes in the expression matrix and removal of duplicate genes.

```
library(IOBR)
```

```
# Load the annotation file `anno_hug133plus2` in IOBR.
```

```
head(anno_hug133plus2)
```

```
## # A tibble: 6 x 2
##   probe_id symbol
##   <fct>      <fct>
## 1 1007_s_at MIR4640
## 2 1053_at   RFC2
## 3 117_at    HSPA6
## 4 121_at    PAX8
## 5 1255_g_at GUCA1A
## 6 1294_at   MIR5193
```

```
# Conduct gene annotation using `anno_hug133plus2` file; If identical gene symbols exist
```

```
eset<-anno_eset(eset      = eset,
                annotation = anno_hug133plus2,
                symbol     = "symbol",
                probe      = "probe_id",
                method     = "mean")
eset[1:5, 1:3]
```

```
##          GSM1523727 GSM1523728 GSM1523729
## SH3KBP1      4.327974   4.316195   4.351425
## RPL41        4.246149   4.246808   4.257940
## EEF1A1       4.293762   4.291038   4.262199
## COX2         4.250288   4.283714   4.270508
## LOC101928826 4.219303   4.219670   4.213252
```

5.3 Available Methods to Decode TME Contexture

```
tme_deconvolution_methods
```

```
##          MCPcounter          EPIC          xCell          CIBERSORT
##      "mcpcounter"      "epic"      "xcell"      "cibersort"
## CIBERSORT Absolute          IPS          ESTIMATE          SVR
##      "cibersort_abs"      "ips"      "estimate"      "svr"
##          lsei          TIMER          quanTIseq
##      "lsei"      "timer"      "quantiseq"
```

```
# Return available parameter options of deconvolution methods
```

The input data is a matrix subseted from ESET of ACRG cohort, with genes in rows and samples in columns. The row name must be HGNC symbols and the column name must be sample names.

```
eset_acrg <- eset[, 1:50]
eset_acrg[1:5, 1:3]
```

```
##          GSM1523727 GSM1523728 GSM1523729
## SH3KBP1      4.327974   4.316195   4.351425
## RPL41        4.246149   4.246808   4.257940
## EEF1A1       4.293762   4.291038   4.262199
## COX2         4.250288   4.283714   4.270508
## LOC101928826 4.219303   4.219670   4.213252
```

Check detail parameters of the function

```
# help(deconvo_tme)
```

5.4 Method 1: CIBERSORT

```
cibersort<-deconvo_tme(eset = eset_acrg, method = "cibersort", arrays = TRUE, perm = 100
```

```
##
```

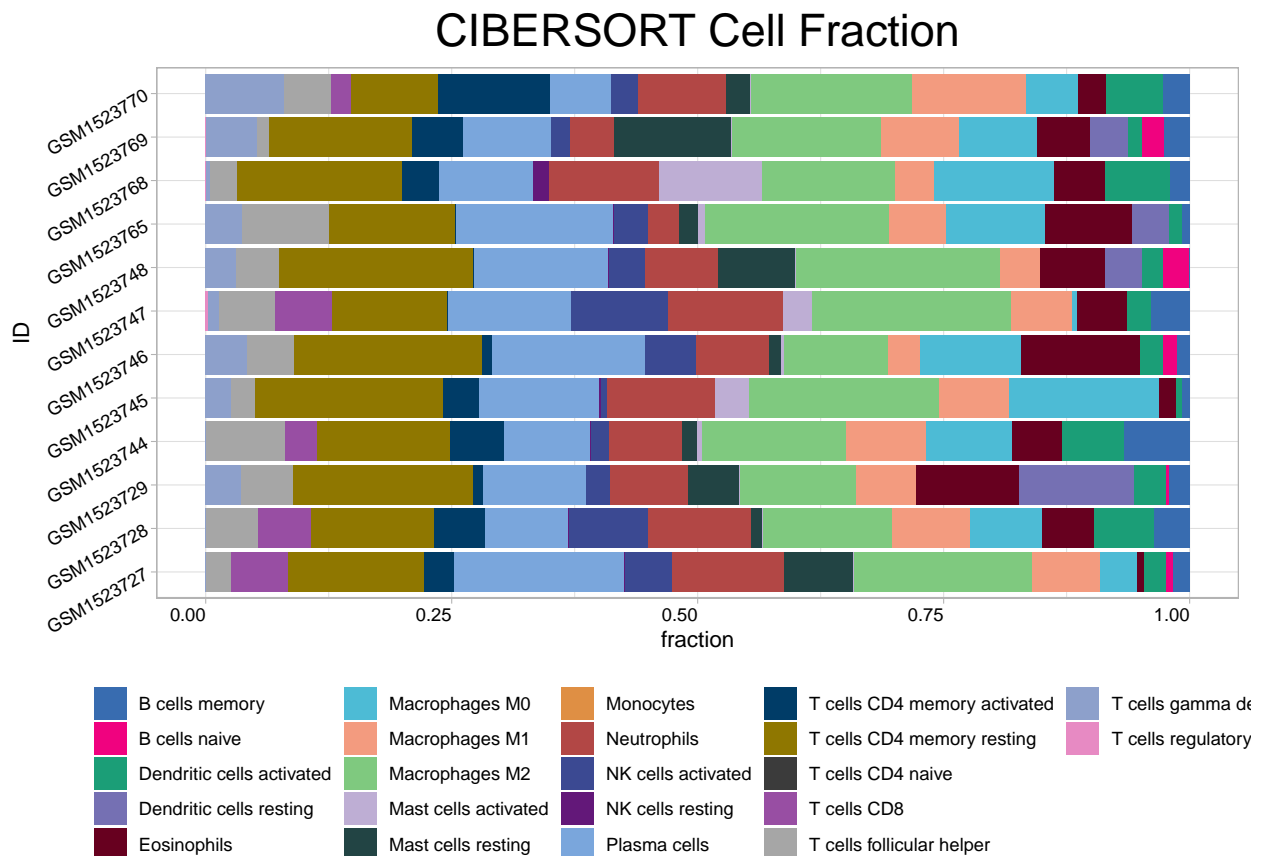
```
## >>> Running CIBERSORT
```

```
# head(cibersort)
```

```
res<-cell_bar_plot(input = cibersort[1:12,], title = "CIBERSORT Cell Fraction")
```

```
## There are seven categories you can choose: box, continue2, continue, random, heatmap,
```

```
## >>>=== Palette option for random: 1: palette1; 2: palette2; 3: palette3; 4: palette
```



5.5 Method 2: EPIC

```
# help(deconvo_epic)
```

```
epic<-deconvo_tme(eset = eset_acrg, method = "epic", arrays = TRUE)
```

```
##
```

```
## >>> Running EPIC
```

```
## Warning in IOBR::EPIC(bulk = eset, reference = ref, mRNA_cell = NULL, scaleExprs = TR
```

```
## GSM1523744; GSM1523746; GSM1523781; GSM1523786
```

```
## - check fit.gof for the convergeCode and convergeMessage
```

```
## Warning in IOBR::EPIC(bulk = eset, reference = ref, mRNA_cell = NULL, scaleExprs
```

```
## = TRUE): mRNA_cell value unknown for some cell types: CAFs, Endothelial - using
```

```
## the default value of 0.4 for these but this might bias the true cell proportions
```

```
## from all cell types.
```

```
head(epic)
```

```
## # A tibble: 6 x 9
```

```
##   ID          Bcells_EPIC CAFs_EPIC CD4_Tcells_EPIC CD8_Tcells_EPIC Endothelial_EPIC
```

```
##   <chr>          <dbl>      <dbl>          <dbl>          <dbl>          <dbl>
```

```
## 1 GSM152~      0.0292    0.00888          0.145          0.0756          0.0876
```

```
## 2 GSM152~      0.0293    0.0109          0.159          0.0745          0.0954
```

```
## 3 GSM152~      0.0308    0.0106          0.149          0.0732          0.0941
```

```
## 4 GSM152~      0.0273    0.0108          0.145          0.0704          0.0860
```

```
## 5 GSM152~      0.0280    0.0111          0.151          0.0707          0.0928
```

```
## 6 GSM152~      0.0320    0.00958          0.148          0.0716          0.0907
```

```
## # i 3 more variables: Macrophages_EPIC <dbl>, NKcells_EPIC <dbl>,
```

```
## #   otherCells_EPIC <dbl>
```

5.6 Method 3: MCPcounter

```
mcp<-deconvo_tme(eset = eset_acrg, method = "mcpcounter")
```

```
##
```

```
## >>> Running MCP-counter
```

```
head(mcp)
```

```
## # A tibble: 6 x 11
```

```
##   ID          T_cells_MCPcounter CD8_T_cells_MCPcounter Cytotoxic_lymphocytes_M~1
```

```
##   <chr>                <dbl>                <dbl>                <dbl>
```

```
## 1 GSM1523727          1.47                1.11                1.33
```

```
## 2 GSM1523728          1.53                1.05                1.60
```

```
## 3 GSM1523729          1.47                1.07                1.37
```

```
## 4 GSM1523744          1.46                1.02                1.44
```

```
## 5 GSM1523745          1.51                1.10                1.49
```

```
## 6 GSM1523746          1.51                0.992               1.40
```

```
## # i abbreviated name: 1: Cytotoxic_lymphocytes_MCPcounter
```

```
## # i 7 more variables: B_lineage_MCPcounter <dbl>, NK_cells_MCPcounter <dbl>,
```

```
## #   Monocytic_lineage_MCPcounter <dbl>,
```

```
## #   Myeloid_dendritic_cells_MCPcounter <dbl>, Neutrophils_MCPcounter <dbl>,
```

```
## #   Endothelial_cells_MCPcounter <dbl>, Fibroblasts_MCPcounter <dbl>
```

5.7 Method 4: xCELL

```
xcell<-deconvo_tme(eset = eset_acrg, method = "xcell", arrays = TRUE)
```

```
head(xcell)
```

```
## # A tibble: 6 x 68
```

```
##   ID          aDC_xCell Adipocytes_xCell Astrocytes_xCell `B-cells_xCell`
```

```
##   <chr>                <dbl>                <dbl>                <dbl>                <dbl>
```

```
## 1 GSM1523727 4.78e-19 0.0250 0 0
## 2 GSM1523728 9.41e- 2 0.00433 7.70e- 3 0
## 3 GSM1523729 1.02e- 1 0.0789 2.04e- 2 0
## 4 GSM1523744 7.88e- 2 0.0538 4.82e-18 0.0126
## 5 GSM1523745 9.02e- 2 0.0136 1.93e- 2 0
## 6 GSM1523746 3.40e- 2 0.0331 9.22e- 2 0
## # i 63 more variables: Basophils_xCell <dbl>,
## # `CD4+_memory_T-cells_xCell` <dbl>, `CD4+_naive_T-cells_xCell` <dbl>,
## # `CD4+_T-cells_xCell` <dbl>, `CD4+_Tcm_xCell` <dbl>, `CD4+_Tem_xCell` <dbl>,
## # `CD8+_naive_T-cells_xCell` <dbl>, `CD8+_T-cells_xCell` <dbl>,
## # `CD8+_Tcm_xCell` <dbl>, `CD8+_Tem_xCell` <dbl>, cDC_xCell <dbl>,
## # Chondrocytes_xCell <dbl>, `Class-switched_memory_B-cells_xCell` <dbl>,
## # CLP_xCell <dbl>, CMP_xCell <dbl>, DC_xCell <dbl>, ...
```

5.8 Method 5: ESTIMATE

```
estimate<-deconvo_tme(eset = eset_acrg, method = "estimate")
```

```
## [1] "Merged dataset includes 9940 genes (472 mismatched)."
```

```
## [1] "1 gene set: StromalSignature overlap= 136"
```

```
## [1] "2 gene set: ImmuneSignature overlap= 138"
```

```
head(estimate)
```

```
## # A tibble: 6 x 5
```

##	ID	StromalScore_estimate	ImmuneScore_estimate	ESTIMATEScore_estimate
##	<chr>	<dbl>	<dbl>	<dbl>
## 1	GSM1523727	-1250.	268.	-982.
## 2	GSM1523728	197.	1334.	1531.
## 3	GSM1523729	-111.	822.	711.
## 4	GSM1523744	-119.	662.	544.

```
## 5 GSM1523745          324.          1015.          1339.
## 6 GSM1523746         -594.          621.          27.0
## # i 1 more variable: TumorPurity_estimate <dbl>
```

5.9 Method 6: TIMER

```
timer<-deconvo_tme(eset = eset_acrg, method = "timer", group_list = rep("stad",dim(eset,
```

```
## [1] "Outlier genes: AGR2 B2M COL1A2 COL3A1 COX2 CYAT1 EEF1A1 EIF1 FTH1 GKN1 HUWE1 IGK
```

```
head(timer)
```

```
## # A tibble: 6 x 7
```

```
##   ID          B_cell_TIMER T_cell_CD4_TIMER T_cell_CD8_TIMER Neutrophil_TIMER
##   <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 GSM1523727      0.104          0.128          0.183          0.108
## 2 GSM1523728      0.103          0.130          0.192          0.118
## 3 GSM1523729      0.106          0.130          0.190          0.110
## 4 GSM1523744      0.101          0.126          0.187          0.111
## 5 GSM1523745      0.104          0.127          0.191          0.116
## 6 GSM1523746      0.105          0.129          0.192          0.111
```

```
## # i 2 more variables: Macrophage_TIMER <dbl>, DC_TIMER <dbl>
```

5.10 Method 7: quanTIseq

```
quantiseq<-deconvo_tme(eset = eset_acrg, tumor = TRUE, arrays = TRUE, scale_mrna = TRUE,
```

```
##
```

```
## Running quanTIseq deconvolution module
```

```
## Gene expression normalization and re-annotation (arrays: TRUE)
```

```
## Removing 17 genes with high expression in tumors
```

```
## Signature genes found in data set: 152/153 (99.35%)
```

```
## Mixture deconvolution (method: lsei)
```

```
## Deconvolution sucessful!
```

```
head(quantiseq)
```

```
## # A tibble: 6 x 12
```

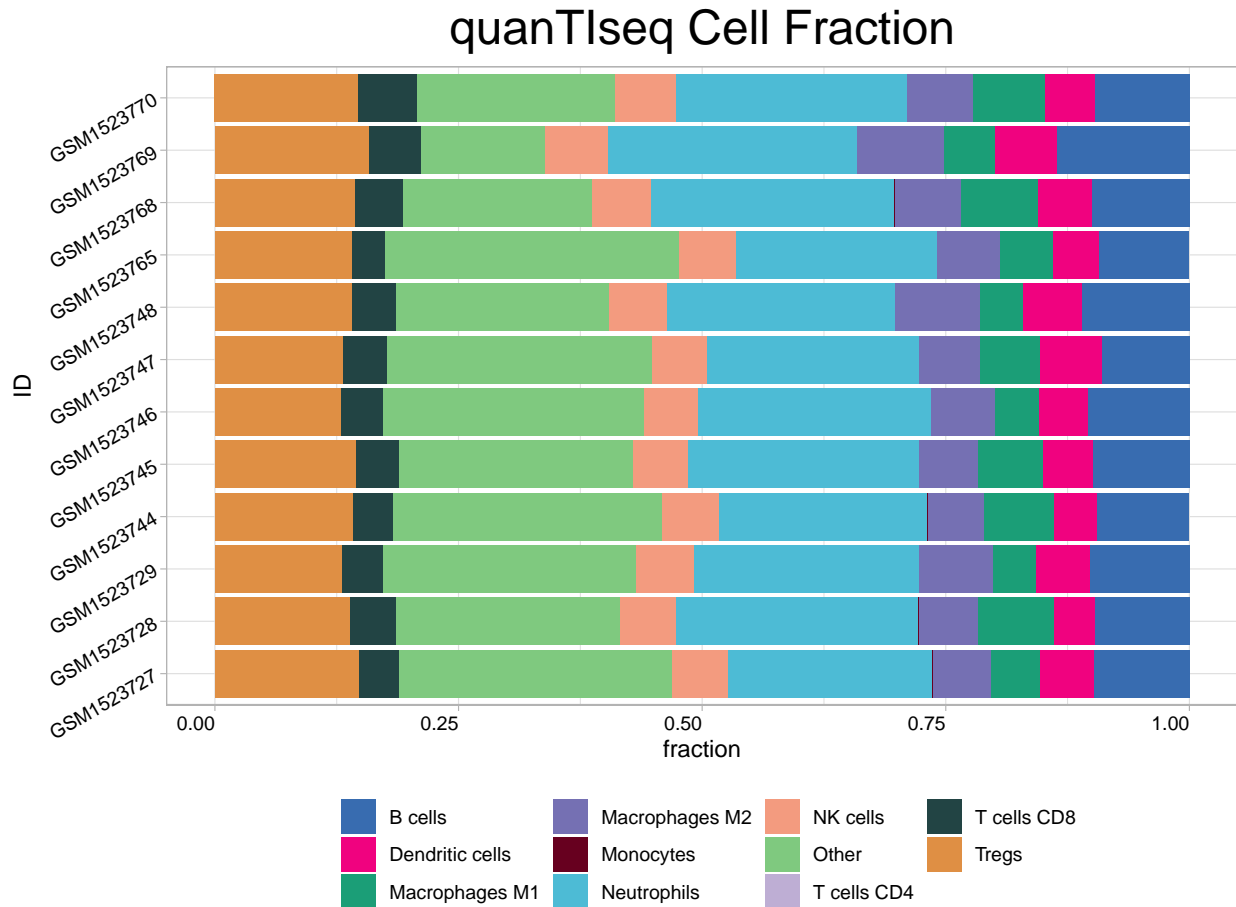
```
##   ID          B_cells_quantiseq Macrophages_M1_quantiseq Macrophages_M2_quantiseq
##   <chr>                <dbl>                <dbl>                <dbl>
## 1 GSM1523727          0.0983                0.0510                0.0598
## 2 GSM1523728          0.0967                0.0795                0.0607
## 3 GSM1523729          0.102                 0.0450                0.0758
## 4 GSM1523744          0.0954                0.0725                0.0579
## 5 GSM1523745          0.0991                0.0669                0.0613
## 6 GSM1523746          0.105                 0.0453                0.0662
```

```
## # i 8 more variables: Monocytes_quantiseq <dbl>, Neutrophils_quantiseq <dbl>,
## #   NK_cells_quantiseq <dbl>, T_cells_CD4_quantiseq <dbl>,
## #   T_cells_CD8_quantiseq <dbl>, Tregs_quantiseq <dbl>,
## #   Dendritic_cells_quantiseq <dbl>, Other_quantiseq <dbl>
```

```
res<-cell_bar_plot(input = quantiseq[1:12, ], title = "quanTIsseq Cell Fraction")
```

```
## There are seven categories you can choose: box, continue2, continue, random, heatmap,
```

```
## >>>>== Palette option for random: 1: palette1; 2: palette2; 3: palette3; 4: palette
```



5.11 Method 8: IPS

```
ips<-deconvo_tme(eset = eset_acrg, method = "ips", plot= FALSE)
head(ips)
```

```
## # A tibble: 6 x 7
```

##	ID	MHC_IPS	EC_IPS	SC_IPS	CP_IPS	AZ_IPS	IPS_IPS
##	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	GSM1523727	2.25	0.404	-0.192	0.220	2.68	9
## 2	GSM1523728	2.37	0.608	-0.578	-0.234	2.17	7
## 3	GSM1523729	2.10	0.480	-0.322	0.0993	2.36	8
## 4	GSM1523744	2.12	0.535	-0.333	0.0132	2.34	8
## 5	GSM1523745	1.91	0.559	-0.479	0.0880	2.08	7

```
## 6 GSM1523746      1.94  0.458 -0.346  0.261      2.31      8
```

5.12 Combination of above deconvolution results

```
tme_combine<-cibersort %>%
  inner_join(.,mcp,by      = "ID") %>%
  inner_join(.,xcell,by    = "ID") %>%
  inner_join(.,epic,by     = "ID") %>%
  inner_join(.,estimate,by = "ID") %>%
  inner_join(.,timer,by    = "ID") %>%
  inner_join(.,quantiseq,by = "ID") %>%
  inner_join(.,ips,by      = "ID")
dim(tme_combine)
```

```
## [1] 50 138
```

If you use this package in your work, please cite both our package and the method(s) you are using.

Licenses of the deconvolution methods | method | license | citation | |———|———|———|—|
 | CIBERSORT | free for non-commercial use only | Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., ... Alizadeh, A. A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*, 12(5), 453–457. <https://doi.org/10.1038/nmeth.3337> | | ESTIMATE | free (GPL2.0) | Vegesna R, Kim H, Torres-Garcia W, ..., Verhaak R. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Communications* 4, 2612. <http://doi.org/10.1038/ncomms3612> | | quanTIseq | free (BSD) | Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., ..., Sopper, S. (2019). Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome medicine*, 11(1), 34. <https://doi.org/10.1186/s13073-019-0638-6> | | TIMER | free (GPL 2.0) | Li, B., Severson, E., Pignon, J.-C., Zhao, H., Li, T., Novak, J., ... Liu, X. S. (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biology*, 17(1),

174. <https://doi.org/10.1186/s13059-016-1028-7> | | IPS | free (BSD) | P. Charoentong et al., Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Reports* 18, 248-262 (2017). <https://doi.org/10.1016/j.celrep.2016.12.019> | | MCPCounter | free (GPL 3.0) | Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., ... de Reyniès, A. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology*, 17(1), 218. <https://doi.org/10.1186/s13059-016-1070-5> | | xCell | free (GPL 3.0) | Aran, D., Hu, Z., & Butte, A. J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology*, 18(1), 220. <https://doi.org/10.1186/s13059-017-1349-1> | | EPIC | free for non-commercial use only (Academic License) | Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., & Gfeller, D. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *ELife*, 6, e26476. <https://doi.org/10.7554/eLife.26476> |

Licenses of the signature-estimation method | method | license | citation | |———|———|———| | GSVA | free (GPL (≥ 2)) | Hänzelmann S, Castelo R, Guinney J (2013). “GSVA: gene set variation analysis for microarray and RNA-Seq data.” *BMC Bioinformatics*, 14, 7. doi: 10.1186/1471-2105-14-7, <http://www.biomedcentral.com/1471-2105/14/7> |

Chapter 6

Cross-references

Cross-references make it easier for your readers to find and link to elements in your book.

6.1 Chapters and sub-chapters

There are two steps to cross-reference any heading:

1. Label the heading: `# Hello world {#nice-label}`.
 - Leave the label off if you like the automated heading generated based on your heading title: for example, `# Hello world = # Hello world {#hello-world}`.
 - To label an un-numbered heading, use: `# Hello world {-#nice-label}` or `{# Hello world .unnumbered}`.
2. Next, reference the labeled heading anywhere in the text using `\@ref(nice-label)`; for example, please see Chapter 6.
 - If you prefer text as the link instead of a numbered reference use: any text you want can go here.

6.2 Captioned figures and tables

Figures and tables *with captions* can also be cross-referenced from elsewhere in your book using `\@ref(fig:chunk-label)` and `\@ref(tab:chunk-label)`, respectively.

See Figure 6.1.

```
par(mar = c(4, 4, .1, .1))  
plot(pressure, type = 'b', pch = 19)
```

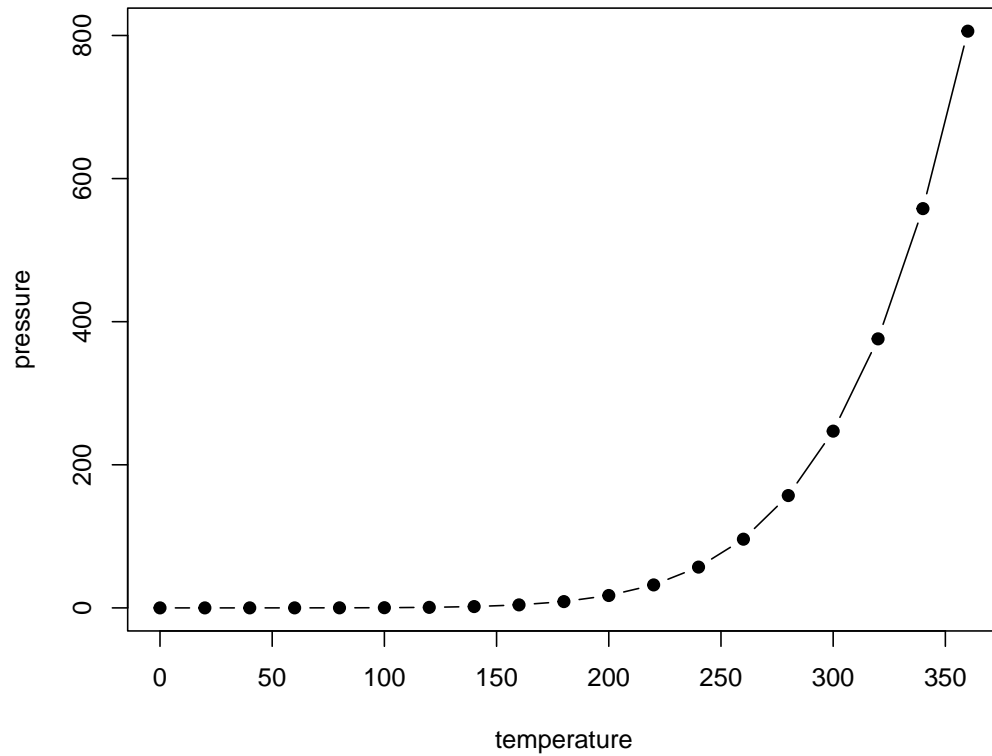


Figure 6.1: Here is a nice figure!

Don't miss Table 6.1.

```
knitr::kable(  
  head(pressure, 10), caption = 'Here is a nice table!',  
  booktabs = TRUE  
)
```

Table 6.1: Here is a nice table!

temperature	pressure
0	0.0002
20	0.0012
40	0.0060
60	0.0300
80	0.0900
100	0.2700
120	0.7500
140	1.8500
160	4.2000
180	8.8000

Chapter 7

Footnotes and citations

7.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one ¹.

7.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package (Xie, 2023) (check out the last code chunk in `index.Rmd` to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015) (this citation was added manually in an external file `book.bib`). Note that the `.bib` files need to be listed in the `index.Rmd` with the YAML `bibliography` key.

The RStudio Visual Markdown Editor can also make it easier to insert citations: <https://rstudio.github.io/visual-markdown-editing/#/citations>

¹This is a footnote.

Chapter 8

Blocks

8.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{8.1}$$

You may refer to using `\@ref{eq:binom}`, like see Equation (8.1).

8.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 8.1.

Theorem 8.1. *For a right triangle, if c denotes the length of the hypotenuse and a and b denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here <https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdo>

wn.html.

8.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: <https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html>

Chapter 9

Sharing your book

9.1 Publishing

HTML books can be published online, see: <https://bookdown.org/yihui/bookdown/publishing.html>

9.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

9.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `gitbook` uses the same social sharing data across all chapters in your book- all links

shared will look the same.

Specify your book's source repository on GitHub using the `edit` key under the configuration options in the `_output.yml` file, which allows users to suggest an edit by linking to a chapter's source file.

Read more about the features of this output format here:

<https://pkgs.rstudio.com/bookdown/reference/gitbook.html>

Or use:

```
?bookdown::gitbook
```

Chapter 10

References

If IOBR R package is utilized in your published research, please cite:

Zeng D, Ye Z, Shen R, Yu G, Wu J, Xiong Y,..., Liao W (2021) **IOBR**: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. *Frontiers in Immunology*. 12:687975. doi: 10.3389/fimmu.2021.687975

10.1 TME deconvolution

Please cite the following papers appropriately for TME deconvolution algorithm if used:

CIBERSORT: Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., ... Alizadeh, A. A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*, 12(5), 453–457. <https://doi.org/10.1038/nmeth.3337>

ESTIMATE: Vegesna R, Kim H, Torres-Garcia W, ..., Verhaak R.*(2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Communications* 4, 2612. <http://doi.org/10.1038/ncomms3612>

quanTIseq: Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., ..., Sopper, S.* (2019). Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome medicine*, 11(1), 34. <https://doi.org/10.1038/s41467-019-10000-0>

[//doi.org/10.1186/s13073-019-0638-6](https://doi.org/10.1186/s13073-019-0638-6)

TIMER: Li, B., Severson, E., Pignon, J.-C., Zhao, H., Li, T., Novak, J., ... Liu, X. S.* (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biology*, 17(1), 174.

IPS: P. Charoentong et al.*, Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Reports* 18, 248-262 (2017). <https://doi.org/10.1016/j.celrep.2016.12.019>

MCPCounter: Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., ... de Reyniès, A*. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology*, 17(1), 218. <https://doi.org/10.1186/s13059-016-1070-5>

xCell: Aran, D., Hu, Z., & Butte, A. J.* (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology*, 18(1), 220. <https://doi.org/10.1186/s13059-017-1349-1>

EPIC: Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., & Gfeller, D*. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *ELife*, 6, e26476. <https://doi.org/10.7554/eLife.26476>

10.2 TME Signatures

For signature score estimation, please cite corresponding literature below:

ssgsea: Barbie, D.A. et al (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*, 462(5):108-112.

gsva: Hänzelmann, S., Castelo, R. and Guinney, J. (2013). GSVA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*, 14(1):7.

zscore: Lee, E. et al (2008). Inferring pathway activity toward precise disease classification. *PLoS Comp Biol*, 4(11):e1000217.

10.3 Data sets

For the datasets enrolled in IOBR, please cite the data sources:

UCSCXena: Wang et al., et al (2019). The UCSCXenaTools R package: a toolkit for accessing genomics data from UCSC Xena platform, from cancer multi-omics to single-cell RNA-seq. *Journal of Open Source Software*, 4(40), 1627

TLSScore: Helmink BA, Reddy SM, Gao J, et al. B cells and tertiary lymphoid structures promote immunotherapy response. *Nature*. 2020 Jan;577(7791):549-555.

IMvigor210 immunotherapy cohort: Mariathasan S, Turley SJ, Nickles D, et al. TGF α attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*. 2018 Feb 22;554(7693):544-548. **HCP5:** Kulski, J.K. Long Noncoding RNA HCP5, a Hybrid HLA Class I Endogenous Retroviral Gene: Structure, Expression, and Disease Associations. *Cells* 2019, 8, 480.

HCP5: Li, Y., Jiang, T., Zhou, W. et al. Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat Commun* 11, 1000 (2020). **HCP5:** Sun J, Zhang Z, Bao S, et al Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer *Journal for ImmunoTherapy of Cancer* 2020;8:e000110.

LINC00657: Feng Q, Zhang H, Yao D, Chen WD, Wang YD. Emerging Role of Non-Coding RNAs in Esophageal Squamous Cell Carcinoma. *Int J Mol Sci*. 2019 Dec 30;21(1):258. doi: 10.3390/ijms21010258.

LINC00657: Qin X, Zhou M, Lv H, Mao X, Li X, Guo H, Li L, Xing H. Long noncoding RNA LINC00657 inhibits cervical cancer development by sponging miR-20a-5p and targeting RUNX3. *Cancer Lett*. 2020 Oct 28:S0304-3835(20)30578-4. doi: 10.1016/j.canlet.2020.10.044. **LINC00657:** Zhang XM, Wang J, Liu ZL, Liu H, Cheng YF, Wang T. LINC00657/miR-26a-5p/CKS2 ceRNA network promotes the growth of esophageal cancer cells via the MDM2/p53/Bcl2/Bax pathway. *Biosci Rep*. 2020;40(6):BSR20200525.

TCGA-STAD: Cancer Genome Atlas Research Network. Comprehensive molecular charac-

terization of gastric adenocarcinoma. *Nature*. 2014 Sep 11;513(7517):202-9. doi: 10.1038/nature13480. TCGA.STAD MAF data: <https://api.gdc.cancer.gov/data/c06465a3-50e7-46f7-b2dd-7bd654ca206b>

10.4 Others

1. Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., ... Alizadeh, A. A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*, 12(5), 453–457.
2. Vegesna R, Kim H, Torres-Garcia W, ..., Verhaak R.* (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Communications* 4, 2612.
3. Rieder, D., Hackl, H., ..., Sopper, S.* (2019). Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome medicine*, 11(1), 34.
4. Li, B., Severson, E., Pignon, J.-C., Zhao, H., Li, T., Novak, J., ... Liu, X. S.* (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biology*, 17(1), 174.
5. P. Charoentong et al.*, Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Reports* 18, 248-262 (2017).
6. Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., ... de Reyniès, A*. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology*, 17(1), 218.
7. Aran, D., Hu, Z., & Butte, A. J.* (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology*, 18(1), 220.
8. Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., & Gfeller, D*. (2017).

- Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *ELife*, 6, e26476.
9. Barbie, D.A. et al (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*, 462(5):108-112.
 10. Hänzelmann, S., Castelo, R. and Guinney, J. (2013). GSEA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*, 14(1):7.
 11. Lee, E. et al (2008). Inferring pathway activity toward precise disease classification. *PLoS Comp Biol*, 4(11):e1000217.
 12. Wang et al.,et al (2019). The UCSCXenaTools R package: a toolkit for accessing genomics data from UCSC Xena platform, from cancer multi-omics to single-cell RNA-seq. *Journal of Open Source Software*, 4(40), 1627
 13. Helmink BA, Reddy SM, Gao J, et al. B cells and tertiary lymphoid structures promote immunotherapy response. *Nature*. 2020 Jan;577(7791):549-555.
 14. Mariathasan S, Turley SJ, Nickles D, et al. TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*. 2018 Feb 22;554(7693):544-548.
 15. Kulski, J.K. Long Noncoding RNA HCP5, a Hybrid HLA Class I Endogenous Retroviral Gene: Structure, Expression, and Disease Associations. *Cells* 2019, 8, 480.
 16. Li, Y., Jiang, T., Zhou, W. et al. Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat Commun* 11, 1000 (2020).
 17. Sun J, Zhang Z, Bao S, et al Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer *Journal for ImmunoTherapy of Cancer* 2020;8:e000110.
 18. Feng Q, Zhang H, Yao D, Chen WD, Wang YD. Emerging Role of Non-Coding RNAs in Esophageal Squamous Cell Carcinoma. *Int J Mol Sci*. 2019 Dec 30;21(1):258. doi: 10.3390/ijms21010258.

19. Qin X, Zhou M, Lv H, Mao X, Li X, Guo H, Li L, Xing H. Long noncoding RNA LINC00657 inhibits cervical cancer development by sponging miR-20a-5p and targeting RUNX3. *Cancer Lett.* 2020 Oct
20. Zhang XM, Wang J, Liu ZL, Liu H, Cheng YF, Wang T. LINC00657/miR-26a-5p/CKS2 ceRNA network promotes the growth of esophageal cancer cells via the MDM2/p53/Bcl2/Bax pathway. *Biosci Rep.* 2020;40(6):BSR20200525.
21. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature.* 2014 Sep 11;513(7517):202-9. doi: 10.1038/nature13480.

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2023). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.35.