

# Package ‘ROKET’

May 7, 2026

**Type** Package

**Title** Optimal Transport-Based Kernel Regression

**Version** 1.0.0

**Date** 2025-03-04

**Description** Perform optimal transport on somatic point mutations and kernel regression hypothesis testing by integrating pathway level similarities at the gene level (Little et al. (2023) <[doi:10.1111/biom.13769](https://doi.org/10.1111/biom.13769)>). The software implements balanced and unbalanced optimal transport and omnibus tests with 'C++' across a set of tumor samples and allows for multi-threading to decrease computational runtime.

**Encoding** UTF-8

**Imports** Rcpp, smarter, ggplot2, survival, MiRKAT

**LinkingTo** Rcpp, RcppArmadillo

**RoxygenNote** 7.2.3

**License** GPL (>= 3)

**Suggests** rmarkdown, knitr, reshape2, ggdendro

**VignetteBuilder** knitr

**NeedsCompilation** yes

**Author** Paul Little [aut, cre]

**Maintainer** Paul Little <[pllittle321@gmail.com](mailto:pllittle321@gmail.com)>

**Repository** CRAN

**Date/Publication** 2025-03-06 17:00:02 UTC

## Contents

kernTEST . . . . .	2
kOT_sim_AGG . . . . .	3
kOT_sim_make . . . . .	3
kOT_sim_OT . . . . .	4
kOT_sim_REG . . . . .	4
run_myOT . . . . .	5
run_myOTs . . . . .	6

---

kernTEST	<i>kernTEST</i>
----------	-----------------

---

**Description**

kernTEST

**Usage**

```
kernTEST(
  RESI = NULL,
  KK,
  YY = NULL,
  XX = NULL,
  OMNI,
  nPERMS = 1e+05,
  ncores = 1
)
```

**Arguments**

RESI	A numeric vector of null model residuals names(RESI) must be set to maintain sample ordering for survival regression, otherwise set RESI to NULL.
KK	An array containing double-centered positive semi-definite kernel matrices. Refer to <code>MiRKAT::D2K()</code> for transforming distance matrices to kernel matrices. The <code>dimnames(KK)[[1]]</code> and <code>dimnames(KK)[[2]]</code> must match <code>names(RESI)</code> . Also set <code>dimnames(KK)[[3]]</code> to keep track of each kernel matrix.
YY	A numeric vector of continuous outcomes to be fitted in a linear model. Defaults to NULL for survival model.
XX	A numeric data matrix with first column for intercept, a column of ones.
OMNI	A matrix of zeros and ones. Each column corresponds to a distance matrix while each row corresponds to an omnibus test. Set <code>rownames(OMNI)</code> for labeling outputted p-values and <code>colnames(OMNI)</code> which should match <code>dimnames(KK)[[3]]</code> .
nPERMS	A positive integer to specify the number of permutation-based p-value calculation
ncores	A positive integer for the number of cores/threads to reduce computational runtime when running for loops

**Value**

A R list of p-values and omnibus p-values.

---

kOT_sim_AGG	<i>kOT_sim_AGG</i>
-------------	--------------------

---

**Description**

kOT\_sim\_AGG

**Usage**

kOT\_sim\_AGG(work\_dir)

**Arguments**

work\_dir            A full path to create "sim\_ROKET" and subdirectories

**Value**

Nothing. Png files are created within the simulation ROKET directory.

---

kOT_sim_make	<i>kOT_sim_make</i>
--------------	---------------------

---

**Description**

Generates simulation files

**Usage**

kOT\_sim\_make(work\_dir, NN = 200, nGENE = 500, nPATH = 25, RR = 200)

**Arguments**

work\_dir            A full path to create "sim\_ROKET" and subdirectories  
 NN                    A positive integer for sample size  
 nGENE                A positive integer for number of genes to simulate  
 nPATH                A positive integer for number of pathways to simulate  
 RR                    A positive integer for number of replicates to simulate

**Value**

Nothing. Rds files are created within the simulation ROKET directory.

---

kOT_sim_OT	<i>kOT_sim_OT</i>
------------	-------------------

---

**Description**

kOT\_sim\_OT

**Usage**

kOT\_sim\_OT(work\_dir, NN, nGENE, nPATH, SCEN, ncores = 1)

**Arguments**

work_dir	A full path to create "sim_ROKET" and subdirectories
NN	A positive integer for sample size
nGENE	A positive integer for number of genes to simulate
nPATH	A positive integer for number of pathways to simulate
SCEN	An integer taking values 1, 2, 3, or 4
ncores	A positive integer specifying the number of cores/threads to use for optimal transport calculations

**Value**

Nothing. Rds files are created within the simulation ROKET directory.

---

kOT_sim_REG	<i>kOT_sim_REG</i>
-------------	--------------------

---

**Description**

kOT\_sim\_REG

**Usage**

kOT\_sim\_REG(work\_dir, NN, nGENE, nPATH, SCEN, rr)

**Arguments**

work_dir	A full path to create "sim_ROKET" and subdirectories
NN	A positive integer for sample size
nGENE	A positive integer for number of genes to simulate
nPATH	A positive integer for number of pathways to simulate
SCEN	An integer taking values 1, 2, 3, or 4
rr	A positive integer indexing a replicate

**Value**

Nothing. A rds file is created within the simulation ROKET directory.

---

run_myOT	<i>run_myOT</i>
----------	-----------------

---

**Description**

Runs balanced or unbalanced optimal transport on two input vectors

**Usage**

```
run_myOT(
  XX,
  YY,
  COST,
  EPS,
  LAMBDA1,
  LAMBDA2 = NULL,
  balance = FALSE,
  conv = 1e-05,
  max_iter = 3000,
  verbose = TRUE,
  show_iter = 50
)
```

**Arguments**

XX	A numeric vector of positive masses
YY	A numeric vector of positive masses
COST	A numeric matrix of non-negative values representing the costs to transport masses between features of XX and YY. The rows of COST and features of XX need to be aligned. The columns of COST and features of YY need to be aligned.
EPS	A positive numeric value representing the tuning parameter for entropic regularization.
LAMBDA1	A non-negative numeric value representing the tuning parameter penalizing the distance between XX and the row sums of the optimal transport matrix.
LAMBDA2	A non-negative numeric value representing the tuning parameter penalizing the distance between YY and the column sums of the optimal transport matrix.
balance	Boolean set to TRUE to run balanced optimal transport regardless of LAMBDA1 and LAMBDA2. Otherwise run unbalanced optimal transport.
conv	A positive numeric value to determine algorithmic convergence. The default value is 1e-5.
max_iter	A positive integer denoting the maximum iterations to run the algorithm.

verbose	Boolean value to display verbose function output.
show_iter	A positive integer to display iteration details at multiples of show_iter but only if verbose = TRUE.

**Value**

A R list containing the optimal transport matrix and associated distance metric.

---

run_myOTs	<i>run_myOTs</i>
-----------	------------------

---

**Description**

run\_myOTs

**Usage**

```
run_myOTs(
  ZZ,
  COST,
  EPS,
  LAMBDA1,
  LAMBDA2 = NULL,
  balance,
  conv = 1e-05,
  max_iter = 3000,
  ncores = 1,
  verbose = TRUE,
  show_iter = 50
)
```

**Arguments**

ZZ	A numeric matrix of non-negative mass to transport. Rows correspond to features (e.g. genes) and columns correspond to samples or individuals. Each column must have strictly positive mass
COST	A numeric square matrix of non-negative values representing the non-negative costs to transport masses between pairs of features
EPS	A positive numeric value representing the tuning parameter for entropic regularization.
LAMBDA1	A non-negative numeric value representing the tuning parameter penalizing the distance between XX and the row sums of the optimal transport matrix.
LAMBDA2	A non-negative numeric value representing the tuning parameter penalizing the distance between YY and the column sums of the optimal transport matrix.
balance	Boolean set to TRUE to run balanced optimal transport regardless of LAMBDA1 and LAMBDA2. Otherwise run unbalanced optimal transport.

<code>conv</code>	A positive numeric value to determine algorithmic convergence. The default value is $1e-5$ .
<code>max_iter</code>	A positive integer denoting the maximum iterations to run the algorithm.
<code>ncores</code>	A positive integer for the number of cores/threads to reduce computational runtime when running for loops
<code>verbose</code>	Boolean value to display verbose function output.
<code>show_iter</code>	A positive integer to display iteration details at multiples of <code>show_iter</code> but only if <code>verbose = TRUE</code> .

**Value**

A R numeric matrix of pairwise distances.

# Index

kernTEST, [2](#)  
kOT\_sim\_AGG, [3](#)  
kOT\_sim\_make, [3](#)  
kOT\_sim\_OT, [4](#)  
kOT\_sim\_REG, [4](#)  
  
run\_myOT, [5](#)  
run\_myOTs, [6](#)