

# Package: RFlocalfdr.data (via r-universe)

August 27, 2024

**Title** Data for the Vignette and Examples in 'RFlocalfdr'

**Version** 0.0.3

**Description** Data for the vignette and examples in 'RFlocalfdr'.

Contains a dataset of 1103547 importance values, and the table of variables used in the random forest splits. The data is Chromosome 22 taken from Auton et al. (2015) [doi:10.1038/nature15393](https://doi.org/10.1038/nature15393). It also contains a 51 samples by 22283 genes data set taken from Spira et al. (2004) [doi:10.1165/rcmb.2004-0273OC](https://doi.org/10.1165/rcmb.2004-0273OC).

**License** GPL (>= 3)

**Encoding** UTF-8

**RoxygenNote** 7.2.3

**Depends** R (>= 2.10)

**LazyDataCompression** xz

**LazyData** true

**NeedsCompilation** no

**Author** Robert Dunne [aut, cre]  
(<https://orcid.org/0000-0003-1946-7279>)

**Maintainer** Robert Dunne <rob.dunne@csiro.au>

**Date/Publication** 2023-10-27 10:30:02 UTC

**Repository** <https://parsifal9.r-universe.dev>

**RemoteUrl** <https://github.com/cran/RFlocalfdr.data>

**RemoteRef** HEAD

**RemoteSha** c0801515c8c0a5c8a3956fc9f1da62a15d6484bd

## Contents

ch22 . . . . .	2
smoking . . . . .	3
<b>Index</b>	<b>4</b>

---

 ch22

*ch22 importance values*


---

### Description

A dataset containing 1103547 importance values, and a table of variables used in splits. Note that the importances have not been logged.

### Usage

```
ch22
```

### Format

A list

**imp** importances

**C** table of counts

### Source

A Global Reference for Human Genetic Variation, Auton et al., Nature, 2015, 526:7571 pp 68–74

### Examples

```
## Not run:
library(ranger)
system.time(fit.ranger.7 <- ranger(dependent.variable.name= "V1", data = aa2,
                                importance = "impurity",
                                num.threads=20,num.trees = 100000,
                                seed=123))

#Ranger result
#Call:
#ranger(dependent.variable.name = "V1", data = aa2, importance = "impurity",
#       num.threads = 20, num.trees = 1e+05, seed = 123)
#Type:                Classification
#Number of trees:     1e+05
#Sample size:         2504
#Number of independent variables: 1103547
#Mtry:                1050
#Target node size:    1
#Variable importance mode: impurity
#Splitrule:           gini
#OOB prediction error: 4.27 %
C <-count_variables(fit.ranger.7)
imp<-rf1$variable.importance

ch22<-list(imp,C)
names(ch22)<-c("imp", "C")
```

```
## End(Not run)

data(ch22)
```

---

smoking	<i>Effects of cigarette smoke on the human airway epithelial cell transcriptome</i>
---------	---

---

### Description

A dataset containing normalized transcript measurements for 51 subjects and 22283 transcripts. See Spira et al (2004). "Gene Expression Profiling of Human Lung Tissue from Smokers with Severe Emphysema", Am J Respir Cell Mol Biol.

### Usage

```
smoking
```

### Format

A list with `rma` (the transcript data) and `y` (the class labels):

**rma** 51 by 22283, log2 real values

**y** a character vector, "smoking" and "never-smoked" ...

### Source

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE994>

# Index

\* **datasets**

ch22, [2](#)

smoking, [3](#)

ch22, [2](#)

smoking, [3](#)