

# Package ‘preciseTADhub’

July 18, 2024

**Type** Package

**Title** Pre-trained random forest models obtained using preciseTAD

**Version** 1.13.0

**Description** An experimentdata package to supplement the preciseTAD package containing pre-trained models and the variable importances of each genomic annotation used to build the model parsed into list objects and available in ExperimentHub. In total, preciseTADhub provides access to n=84 random forest classification models optimized to predict TAD/chromatin loop boundary regions and stored as .RDS files. The value, n, comes from the fact that we considered l=2 cell lines {GM12878, K562}, g=2 ground truth boundaries {Arrowhead, Peakachu}, and c=21 autosomal chromosomes {CHR1, CHR2, ..., CHR22} (omitting CHR9). Furthermore, each object is itself a two-item list containing: (1) the model object, and (2) the variable importances for CTCF, RAD21, SMC3, and ZNF143 used to predict boundary regions. Each model is trained via a ‘‘holdout’’ strategy, in which data from chromosomes {CHR1, CHR2, ..., CHRi-1, CHRi+1, ..., CHR22} were used to build the model and the ith chromosome was reserved for testing. See <https://doi.org/10.1101/2020.09.03.282186> for more detail on the model building strategy.

**License** MIT + file LICENSE

**Depends** R (>= 4.1)

**Encoding** UTF-8

**LazyData** true

**Suggests** knitr, rmarkdown, markdown, BiocStyle, preciseTAD

**Imports** ExperimentHub

**VignetteBuilder** knitr

**biocViews** ExperimentData, PackageTypeData, ExperimentHub, Genome

**NeedsCompilation** no

**RoxygenNote** 7.1.1

**BugReports** <https://github.com/dozmorovlab/preciseTADhub/issues>

**URL** <https://github.com/dozmorovlab/preciseTADhub>

**git\_url** https://git.bioconductor.org/packages/preciseTADhub  
**git\_branch** devel  
**git\_last\_commit** 842246a  
**git\_last\_commit\_date** 2024-04-30  
**Repository** Bioconductor 3.20  
**Date/Publication** 2024-07-18  
**Author** Spiro Stilianoudakis [aut],  
 Mikhail Dozmorov [aut, cre]  
**Maintainer** Mikhail Dozmorov <mikhail.dozmorov@gmail.com>

## Contents

|                                 |          |
|---------------------------------|----------|
| preciseTADhub-package . . . . . | 2        |
| readEH . . . . .                | 3        |
| <b>Index</b>                    | <b>4</b> |

---

preciseTADhub-package *Pre-trained models obtained using preciseTAD as list objects.*

---

## Description

preciseTADhub is package that give users access to pre-trained random forest models that can be leveraged to predict TAD and/or chromatin loop boundaries using the preciseTAD R package. These data have been parsed into list objects and RDS files and are available in ExperimentHub.

## Details

See the vignette for examples of using these data in predicting precise boundary location at base-level resolution.

```
browseVignettes("preciseTADhub")
```

Details of how these data were created are in the scripts/ directory of the source package.

## Examples

```
## Not run:
library(ExperimentHub)
eh <- ExperimentHub()
myfiles <- query(eh, "preciseTADhub")
CHR1_GM12878_5kb_Arrowhead <- myfiles[[1]]

## End(Not run)
```

---

|        |  |
|--------|--|
| readEH | <i>A wrapper function for efficiently reading in user-specified random forest models generated by <code>preciseTAD::TADrandomForest</code>, built on cell-line specific CTCF, RAD21, SMC3, and ZNF143 ChIP-seq peak regions.</i> |
|--------|--|

---

## Description

A wrapper function for efficiently reading in user-specified random forest models generated by `preciseTAD::TADrandomForest`, built on cell-line specific CTCF, RAD21, SMC3, and ZNF143 ChIP-seq peak regions.

## Usage

```
readEH(chr, cl, gt, source)
```

## Arguments

|        |   |
|--------|---|
| chr    | Which chromosome was used as the holdout during the training process. That is, all other chromosomes were combined when building the random forest. |
| cl     | The cell line that was used (either "GM12878" or "K562")  |
| gt     | The ground-truth TAD or chromatin loop boundaries used to construct the binary response vector (either "Arrowhead" or "Peakachu").                  |
| source | The source of the files stored on ExperimentHub using <code>query(hub, "package_name")</code> .   |

## Value

A trained model object from `caret`

## Examples

```
# Suppose we want to read in the model that was built using CHR1-CHR21,  
# on GM12878, using Arrowhead defined TAD boundaries at 5kb resolution.
```

```
#Initialize ExperimentHub  
library(ExperimentHub)  
hub <- ExperimentHub()  
query(hub, "preciseTADhub")  
myfiles <- query(hub, "preciseTADhub")
```

```
CHR22_GM12878_5kb_Arrowhead <- readEH(chr = "CHR22",  
                                     cl = "GM12878",  
                                     gt = "Arrowhead",  
                                     source = myfiles)
```

# Index

## \* **utilities**

preciseTADhub-package, [2](#)

preciseTADhub (preciseTADhub-package), [2](#)

preciseTADhub-package, [2](#)

readEH, [3](#)