

# Package ‘xhmmScripts’

July 2, 2014

**Type** Package

**Title** XHMM R scripts

**Version** 1.1

**Date** 2014-05-21

**Author** Menachem Fromer

**Maintainer** Menachem Fromer <fromer@broadinstitute.org>

**Description** R scripts for plotting and assessing XHMM whole-exome-sequencing-based CNV calls. XHMM (eXome Hidden Markov Model) is a C++ software package (<http://atgu.mgh.harvard.edu/xhmm>) written to call copy number variation (CNV) from next-generation sequencing projects, where exome capture was used (or targeted sequencing, more generally). This R package enables the user to visualize both the PCA normalization performed by XHMM and the CNVs it has called.

**License** GPL-2

**Depends** R (>= 2.14.0)

**Imports** gplots, plotrix

**Suggests**

**URL** <http://atgu.mgh.harvard.edu/xhmm>

**BugReports** <http://groups.google.com/a/broadinstitute.org/group/xhmm-users>

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2014-06-15 07:38:03

**R topics documented:**

xhmmScripts-package . . . . .	2
calcSegmentIntervals . . . . .	3
listOfNulls . . . . .	3
loadNamedVectorNoHeaderMayNotExist . . . . .	4
loadTargetsToGenes . . . . .	4
loadXCNVcalls . . . . .	5
loadXHMMdata . . . . .	5
pedigreeDataToBinarySampleProperties . . . . .	6
phenotypeDataToBinarySampleProperties . . . . .	6
plotAllChromosomeValues . . . . .	7
plot_XHMM_genes . . . . .	7
plot_XHMM_region . . . . .	9
plot_XHMM_targets . . . . .	10
readNamedMatrix . . . . .	12
readPedigreeFile . . . . .	12
readPhenotypesFile . . . . .	13
scanVectorMayNotExist . . . . .	13
sourceDir . . . . .	14
targetsToChrBp1Bp2 . . . . .	14
targetsToSizes . . . . .	15
XHMM_plots . . . . .	15
<b>Index</b>	<b>17</b>

---

xhmmScripts-package	<i>Examine and plot XHMM whole-exome-sequencing-based CNV calls</i>
---------------------	---

---

**Description**

Examine and plot XHMM whole-exome-sequencing-based CNV calls.

XHMM (eXome Hidden Markov Model) is a C++ software package (<http://atgu.mgh.harvard.edu/xhmm>) written to call copy number variation (CNV) from next-generation sequencing projects, where exome capture was used (or targeted sequencing, more generally).

This R package enables the user to visualize both the PCA normalization performed by XHMM and the CNVs it has called.

**Details**

Package:	xhmmScripts
Type:	Package
Version:	1.0
Date:	2014-05-22
License:	GPL-2

A typical use case is outlined here:

[https://bitbucket.org/statgen/xhmm/src/master/sources/scripts/example\\_make\\_XHMM\\_plots.R](https://bitbucket.org/statgen/xhmm/src/master/sources/scripts/example_make_XHMM_plots.R)

### Author(s)

Menachem Fromer

Maintainer: Menachem Fromer <fromer@broadinstitute.org>

### References

<http://atgu.mgh.harvard.edu/xhmm/citations.shtml>

---

calcSegmentIntervals *Determine all of the indices where a transition in value occurs.*

---

### Description

Determine all of the vector indices where a transition in value occurs along the input vector.

### Usage

```
calcSegmentIntervals(valuesVectorList)
```

### Arguments

valuesVectorList

### Examples

```
ints = c(rep(1, 100), rep(2, 200), rep(3, 300))
breaks = calcSegmentIntervals(list(ints))
```

---

listOfNulls *Create an empty list object of given size.*

---

### Description

Create an empty list object of given size, with NULL at each position.

### Usage

```
listOfNulls(n)
```

### Arguments

n

**Examples**

```
lst = listOfNulls(17)
print(length(lst))
```

---

```
loadNamedVectorNoHeaderMayNotExist
```

*Load a vector with named rows.*

---

**Description**

Load a vector with named rows. If the file does not exist, warn the user and return an empty vector.

**Usage**

```
loadNamedVectorNoHeaderMayNotExist(fileName)
```

**Arguments**

```
fileName
```

**Examples**

```
## Not run: vec = loadNamedVectorNoHeaderMayNotExist("named_vector.txt")
```

---

```
loadTargetsToGenes
```

*Load the exome-target to gene mappings.*

---

**Description**

Load the exome-target to gene mappings.

**Usage**

```
loadTargetsToGenes(targToGenesFile, targRegexReplace =
list(c("\\.\\.\\.\"", "-")), geneDelim = "\\|",
targColumn = 1, geneColumn = 3, CHROMOSOMES_START_WITH_CHR = FALSE)
```

**Arguments**

```
targToGenesFile
```

```
targRegexReplace
```

```
geneDelim
```

```
targColumn
```

```
geneColumn
```

```
CHROMOSOMES_START_WITH_CHR
```

**Examples**

```
## Not run: targsToGenes = loadTargetsToGenes("targs_to_genes.txt")
```

---

loadXCNVcalls	<i>Load the XHMM CNV calls.</i>
---------------	---------------------------------

---

**Description**

Load the XHMM CNV calls.

**Usage**

```
loadXCNVcalls(xcnvFile)
```

**Arguments**

xcnvFile

**Examples**

```
## Not run: xcnv = loadXCNVcalls("run.xcnv")
```

---

loadXHMMdata	<i>Load all data files generated during an XHMM run.</i>
--------------	--

---

**Description**

Load all data files generated during an XHMM run.

**Usage**

```
loadXHMMdata(basename, EXCLUDE_LARGE_MATRICES = c())
```

**Arguments**

basename  
EXCLUDE\_LARGE\_MATRICES

**Examples**

```
## Not run: all_xhmm_data = loadXHMMdata("XHMM_PATH_AND_PREFIX")
```

pedigreeDataToBinarySampleProperties

*Convert a parsed Plink/Seq pedinfo file into a matrix of binary sample properties.*

---

**Description**

Convert a parsed Plink/Seq pedinfo file into a matrix of binary sample properties.

**Usage**

```
pedigreeDataToBinarySampleProperties(pedigreeData)
```

**Arguments**

pedigreeData

**Examples**

```
## Not run: pedigreeDataToBinarySampleProperties(readPedigreeFile("a.pedinfo"))
```

---

phenotypeDataToBinarySampleProperties

*Convert a parsed Plink/Seq phenotype file into a matrix of binary sample properties.*

---

**Description**

Convert a parsed Plink/Seq phenotype file into a matrix of binary sample properties.

**Usage**

```
phenotypeDataToBinarySampleProperties(phenotypes)
```

**Arguments**

phenotypes

**Examples**

```
## Not run: phenotypeDataToBinarySampleProperties(readPhenotypesFile("a.phe"))
```

---

`plotAllChromosomeValues`*Plot to a png file the values for the given intervals.*

---

**Description**

Plot to a png file the values for the given intervals, alternating color by chromosome.

**Usage**

```
plotAllChromosomeValues(chr, values, ylab, title,
outputName, specialColors = NULL)
```

**Arguments**

```
chr
values
ylab
title
outputName
specialColors
```

**Examples**

```
## Not run: plotAllChromosomeValues(c(rep(1, 200), rep(2, 100), rep(3, 50)),
c(sample(1:200), sample(1:100), sample(1:50)), "Values", "Genomic values", "./out")
## End(Not run)
```

---

`plot_XHMM_genes`*Plot the XHMM CNV spanning the input genes.*

---

**Description**

Plot the XHMM CNV spanning the input genes.

**Usage**

```
plot_XHMM_genes(BASE_OUTPUT_NAME, xhmm_data, allTargsToGenes,
binarySampleFeatures, SQ_THRESH, genes,
PLOT_ONLY_PNG = TRUE, PLOT_LINE_SEGMENTS = TRUE, DEFAULT_WIDTH = 14,
DEFAULT_HEIGHT = 7, UPDATE_FIG_HEIGHT_FOR_MARGINS = TRUE,
GENE_NAME_cex = 0.8, EXACT_GENE_MATCH = FALSE, PLOT_PIPELINE_TRANSITIONS
= TRUE, DIFFERENTIATE_SAMPLE_GROUPS = NULL, COLOR_DIFFERENTIAL_TARGETS =TRUE,
MARK_SAMPLES = NULL, MARK_SAMPLES_COLORS = DEFAULT_MARK_SAMPLES_COLORS,
MARK_SAMPLES_ONLY_IN_CALLS = FALSE, EXCLUDE_SAMPLES = NULL, MARK_INTERVALS = NULL)
```

**Arguments**`BASE_OUTPUT_NAME``xhmm_data``allTargsToGenes``binarySampleFeatures``SQ_THRESH``genes``PLOT_ONLY_PNG``PLOT_LINE_SEGMENTS``DEFAULT_WIDTH``DEFAULT_HEIGHT``UPDATE_FIG_HEIGHT_FOR_MARGINS``GENE_NAME_cex``EXACT_GENE_MATCH``PLOT_PIPELINE_TRANSITIONS``DIFFERENTIATE_SAMPLE_GROUPS``COLOR_DIFFERENTIAL_TARGETS``MARK_SAMPLES``MARK_SAMPLES_COLORS``MARK_SAMPLES_ONLY_IN_CALLS``EXCLUDE_SAMPLES``MARK_INTERVALS`**Examples**

```
## Not run: plot_XHMM_genes("./out", xhmm_data,  
allTargsToGenes, binarySampleFeatures, SQ_THRESH, genes)  
## End(Not run)
```



---

plot_XHMM_region	<i>Plot the XHMM CNV spanning the input region.</i>
------------------	---

---

### Description

Plot the XHMM CNV spanning the input region.

### Usage

```
plot_XHMM_region(BASE_OUTPUT_NAME, xhmm_data, allTargsToGenes,  
binarySampleFeatures, SQ_THRESH, chrom, start = NULL, stop = NULL,  
PLOT_ONLY_PNG = TRUE, PLOT_LINE_SEGMENTS = TRUE, DEFAULT_WIDTH = 14,  
DEFAULT_HEIGHT = 7,  
UPDATE_FIG_HEIGHT_FOR_MARGINS = TRUE, GENE_NAME_cex = 0.8,  
PLOT_PIPELINE_TRANSITIONS = TRUE, DIFFERENTIATE_SAMPLE_GROUPS = NULL,  
COLOR_DIFFERENTIAL_TARGETS = TRUE, MARK_SAMPLES = NULL,  
MARK_SAMPLES_COLORS = DEFAULT_MARK_SAMPLES_COLORS,  
MARK_SAMPLES_ONLY_IN_CALLS = FALSE, EXCLUDE_SAMPLES = NULL,  
MARK_INTERVALS = NULL, APPEND_REGION_NAME = TRUE)
```

### Arguments

BASE\_OUTPUT\_NAME

xhmm\_data

allTargsToGenes

binarySampleFeatures

SQ\_THRESH

chrom

start

stop

PLOT\_ONLY\_PNG

PLOT\_LINE\_SEGMENTS

DEFAULT\_WIDTH

DEFAULT\_HEIGHT

UPDATE\_FIG\_HEIGHT\_FOR\_MARGINS

GENE\_NAME\_cex

PLOT\_PIPELINE\_TRANSITIONS

DIFFERENTIATE\_SAMPLE\_GROUPS

COLOR\_DIFFERENTIAL\_TARGETS

MARK\_SAMPLES

MARK\_SAMPLES\_COLORS

MARK\_SAMPLES\_ONLY\_IN\_CALLS

EXCLUDE\_SAMPLES

MARK\_INTERVALS

APPEND\_REGION\_NAME

### Examples

```
## Not run: plot_XHMM_region("./out", xhmm_data, allTargsToGenes,  
binarySampleFeatures, SQ_THRESH, "X", 10000, 20000)  
## End(Not run)
```

---

plot\_XHMM\_targets      *Plot the XHMM CNV spanning the input target indices.*

---

### Description

Plot the XHMM CNV spanning the input target indices.

### Usage

```
plot_XHMM_targets(BASE_OUTPUT_NAME, xhmm_data, allTargsToGenes,  
binarySampleFeatures, SQ_THRESH, startTarg, stopTarg,  
PLOT_ONLY_PNG = TRUE, PLOT_LINE_SEGMENTS = TRUE, DEFAULT_WIDTH = 14,  
DEFAULT_HEIGHT = 7, UPDATE_FIG_HEIGHT_FOR_MARGINS = TRUE, GENE_NAME_cex  
= 0.8,  
PLOT_PIPELINE_TRANSITIONS = TRUE, DIFFERENTIATE_SAMPLE_GROUPS = NULL,  
COLOR_DIFFERENTIAL_TARGETS = TRUE, MARK_SAMPLES = NULL,  
MARK_SAMPLES_COLORS = DEFAULT_MARK_SAMPLES_COLORS,  
MARK_SAMPLES_ONLY_IN_CALLS = FALSE,  
EXCLUDE_SAMPLES = NULL, MARK_INTERVALS = NULL, APPEND_REGION_NAME = TRUE)
```

### Arguments

BASE\_OUTPUT\_NAME

xhmm\_data

```
allTargsToGenes

binarySampleFeatures

SQ_THRESH
startTarg
stopTarg
PLOT_ONLY_PNG
PLOT_LINE_SEGMENTS

DEFAULT_WIDTH
DEFAULT_HEIGHT
UPDATE_FIG_HEIGHT_FOR_MARGINS

GENE_NAME_cex
PLOT_PIPELINE_TRANSITIONS

DIFFERENTIATE_SAMPLE_GROUPS

COLOR_DIFFERENTIAL_TARGETS

MARK_SAMPLES
MARK_SAMPLES_COLORS

MARK_SAMPLES_ONLY_IN_CALLS

EXCLUDE_SAMPLES

MARK_INTERVALS
APPEND_REGION_NAME
```

### **Examples**

```
## Not run: plot_XHMM_targets("./out", xhmm_data, allTargsToGenes,
binarySampleFeatures, SQ_THRESH, 150, 180)
## End(Not run)
```

---

readNamedMatrix      *Quickly read a rectangular matrix that has row and column names.*

---

**Description**

Quickly read a rectangular matrix that has row and column names.

**Usage**

```
readNamedMatrix(matFile, what = double())
```

**Arguments**

matFile      If matFile ends in '.gz', then this file will be read as a gzip-ed text file.  
what

**Examples**

```
## Not run: readNamedMatrix("table.txt")
```

---

readPedigreeFile      *Read a Plink/Seq pedinfo file.*

---

**Description**

Read a Plink/Seq pedinfo file.

**Usage**

```
readPedigreeFile(file)
```

**Arguments**

file

**Examples**

```
## Not run: readPedigreeFile("a.pedinfo")
```

---

readPhenotypesFile     *Read a Plink/Seq phenotype file.*

---

**Description**

Read a Plink/Seq phenotype file.

**Usage**

readPhenotypesFile(file)

**Arguments**

file

**Examples**

```
## Not run: readPhenotypesFile("a.phe")
```

---

scanVectorMayNotExist     *Read in an unnamed vector.*

---

**Description**

Read in an unnamed vector. If the file does not exist, warn the user and return an empty vector.

**Usage**

scanVectorMayNotExist(fileName)

**Arguments**

fileName

**Examples**

```
## Not run: scanVectorMayNotExist("values.txt")
```

---

sourceDir	<i>Run source across the entire set of specified files in a given directory.</i>
-----------	--

---

**Description**

Run source across the entire set of specified files in a given directory.

**Usage**

```
sourceDir(path, trace = FALSE, recursive = FALSE, ...)
```

**Arguments**

path  
trace  
recursive  
...

**Examples**

```
## Not run: sourceDir("./source/R_functions")
```

---

targetsToChrBp1Bp2	<i>Parse a list of genomic intervals into their component chromosome, start, and stop.</i>
--------------------	--

---

**Description**

Parse a list of genomic intervals into their component chromosome, start, and stop.

**Usage**

```
targetsToChrBp1Bp2(targets, startStopDelim = "-")
```

**Arguments**

targets  
startStopDelim

**Examples**

```
targetsToChrBp1Bp2(c("1:23-29", "X:100-200", "Y:10-50"))
```

---

targetsToSizes	<i>Parse a list of genomic intervals into their corresponding span lengths.</i>
----------------	---

---

**Description**

Parse a list of genomic intervals into their corresponding span lengths.

**Usage**

```
targetsToSizes(targets)
```

**Arguments**

targets

**Examples**

```
targetsToSizes(c("1:23-29", "X:100-200", "Y:10-50"))
```

---

XHMM_plots	<i>Make a series of plots documenting the output of XHMM.</i>
------------	---

---

**Description**

Make a series of plots documenting the output of XHMM.

**Usage**

```
XHMM_plots(PLOT_PATH, JOB_PREFICES, JOB_TARGETS_TO_GENES = NULL,  
SAMPLE_FEATURES = NULL, SQ_THRESH = 60, NUM_ADD_TARGS = 2,  
PLOT_READ_DEPTHS = TRUE, PLOT_PC_CORRS = TRUE,  
PLOT_ALL_CNVS = TRUE, USE_XCNV_TO_PLOT = NULL,  
INCLUDE_PEDIGREE_SAMPLES = NULL, PLOT_ONLY_PNG = TRUE, LIMIT_MEMORY = FALSE)
```

**Arguments**

PLOT\_PATH

JOB\_PREFICES

JOB\_TARGETS\_TO\_GENES

SAMPLE\_FEATURES

SQ\_THRESH

NUM\_ADD\_TARGS

PLOT\_READ\_DEPTHS

PLOT\_PC\_CORRS

PLOT\_ALL\_CNVS

USE\_XCNV\_TO\_PLOT

INCLUDE\_PEDIGREE\_SAMPLES

PLOT\_ONLY\_PNG

LIMIT\_MEMORY

### **Examples**

```
## Not run: XHMM_plots(PLOT_PATH, JOB_PREFICES, JOB_TARGETS_TO_GENES, SAMPLE_FEATURES)
```



# Index

## \*Topic `\textasciitildekwd1`

- `calcSegmentIntervals`, 3
- `listOfNulls`, 3
- `loadNamedVectorNoHeaderMayNotExist`, 4
- `loadTargetsToGenes`, 4
- `loadXCNVcalls`, 5
- `loadXHMMdata`, 5
- `pedigreeDataToBinarySampleProperties`, 6
- `phenotypeDataToBinarySampleProperties`, 6
- `plot_XHMM_genes`, 7
- `plot_XHMM_region`, 9
- `plot_XHMM_targets`, 10
- `plotAllChromosomeValues`, 7
- `readNamedMatrix`, 12
- `readPedigreeFile`, 12
- `readPhenotypesFile`, 13
- `scanVectorMayNotExist`, 13
- `sourceDir`, 14
- `targetsToChrBp1Bp2`, 14
- `targetsToSizes`, 15
- `XHMM_plots`, 15

## \*Topic `\textasciitildekwd2`

- `calcSegmentIntervals`, 3
- `listOfNulls`, 3
- `loadNamedVectorNoHeaderMayNotExist`, 4
- `loadTargetsToGenes`, 4
- `loadXCNVcalls`, 5
- `loadXHMMdata`, 5
- `pedigreeDataToBinarySampleProperties`, 6
- `phenotypeDataToBinarySampleProperties`, 6
- `plot_XHMM_genes`, 7
- `plot_XHMM_region`, 9
- `plot_XHMM_targets`, 10

- `plotAllChromosomeValues`, 7
- `readNamedMatrix`, 12
- `readPedigreeFile`, 12
- `readPhenotypesFile`, 13
- `scanVectorMayNotExist`, 13
- `sourceDir`, 14
- `targetsToChrBp1Bp2`, 14
- `targetsToSizes`, 15
- `XHMM_plots`, 15

- `calcSegmentIntervals`, 3
- `listOfNulls`, 3
- `loadNamedVectorNoHeaderMayNotExist`, 4
- `loadTargetsToGenes`, 4
- `loadXCNVcalls`, 5
- `loadXHMMdata`, 5
- `pedigreeDataToBinarySampleProperties`, 6
- `phenotypeDataToBinarySampleProperties`, 6
- `plot_XHMM_genes`, 7
- `plot_XHMM_region`, 9
- `plot_XHMM_targets`, 10
- `plotAllChromosomeValues`, 7

- `readNamedMatrix`, 12
- `readPedigreeFile`, 12
- `readPhenotypesFile`, 13
- `scanVectorMayNotExist`, 13
- `sourceDir`, 14

- `targetsToChrBp1Bp2`, 14
- `targetsToSizes`, 15
- `XHMM_plots`, 15
- `xhmmScripts (xhmmScripts-package)`, 2
- `xhmmScripts-package`, 2