CopyhelpeR: Helper files for CopywriteR.

Thomas Kuilman

November 25, 2026

Department of Molecular Oncology
Netherlands Cancer Institute
The Netherlands

t.kuilman@nki.nl or thomaskuilman@yahoo.com

Contents

1 Usage 1

1 Usage

To get the full path to the helper files folder for the relevant reference genomes (hg18, hg19, hg38, mm9 or mm10), use the `loadHelperFile()` function:

```
> library(CopyhelpeR)
> getPathHelperFiles("hg19")
```

```
/tmp/RtmpuBtD4K/Rinst60e4cc9345f/CopyhelpeR/extdata/hg19_1kb
```

The helper files are provided as annotation data for the *CopywriteR* package, which uses off-target reads from targeted sequencing for copy number detection. The helper files contain pre-assembled 1kb bin GC-content and mappability files for the reference genomes hg18, hg19, hg38, mm9 and mm10. In addition, it contains a blacklist filter to remove regions that display CNV. Files in the returned folder are stored as GRanges objects from the *GenomicRanges* package.

The GC-content data were obtained using 'bedtools nuc' (http://bedtools.readthedocs.org/en/latest/content/tools/nuc.html). Only relevant columns were saved in the GC.mappa.grange GRanges object.

The mappability data were obtained by aligning all possible 51 base pair genomic fragments using BWA (http://bio-bwa.sourceforge.net/). The mappability of every fragment was binarized, and the mappability of a specific region is taken as the average mappability of all fragments that fall into this region.

The regions of copy number variation (CNV) were obtained from the 'wgEncodeDacMapabilityConsensusExcludable.bed.gz' and the 'wgEncodeDukeMapabilityRegionsExcludable.bed.gz' files from the ENCODE project (http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeMapability/). Blacklisted regions were merged using 'bedtools merge' (http://bedtools.readthedocs.org/en/latest/content/tools/merge.html).