

# Package ‘plotgardenerData’

December 10, 2024

**Title** Datasets and test data files for the plotgardener package

**Version** 1.12.0

**Description** This is a supplemental data package for the plotgardener package. Includes example datasets used in plotgardener vignettes and example raw data files. For details on how to use these datasets, see the plotgardener package vignettes.

**Depends** R (>= 4.1.0)

**Suggests** rmarkdown, knitr

**biocViews** ExperimentData, Homo\_sapiens\_Data, ExpressionData, Genome, ChIPSeqData, ENCODE

**License** MIT + file LICENSE

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.1.1

**VignetteBuilder** knitr

**URL** <https://github.com/PhanstiellLab/plotgardenerData>,  
<https://phanstiellab.github.io/plotgardener>

**BugReports** <https://github.com/PhanstiellLab/plotgardenerData/issues>

**git\_url** <https://git.bioconductor.org/packages/plotgardenerData>

**git\_branch** RELEASE\_3\_20

**git\_last\_commit** 5cb4d78

**git\_last\_commit\_date** 2024-10-29

**Repository** Bioconductor 3.20

**Date/Publication** 2024-12-10

**Author** Nicole Kramer [aut, cre] (<<https://orcid.org/0000-0001-9617-9671>>)

**Maintainer** Nicole Kramer <nekramer@live.unc.edu>

## Contents

COVID_NY_FL_tracking . . . . .	2
COVID_NY_FL_vaccines . . . . .	2
COVID_USA_cases . . . . .	3
GM12878_ChIP_CTCF_signal . . . . .	4

GM12878_ChIP_H3K27ac_signal . . . . .	4
GM12878_HiC_10kb . . . . .	5
hg19_insulin_GWAS . . . . .	6
IMR90_ChIP_CTCF_reads . . . . .	6
IMR90_ChIP_CTCF_signal . . . . .	7
IMR90_ChIP_H3K27ac_signal . . . . .	8
IMR90_DNAloops_pairs . . . . .	9
IMR90_HiC_10kb . . . . .	9

<b>Index</b>	<b>11</b>
--------------	-----------

---

COVID\_NY\_FL\_tracking *plotgardener example data for tracked COVID-19 cases in New York and Florida*

---

### Description

A timeline dataset tracking positive COVID-19 cases in New York and Florida from 2020-01-29 to 2021-03-07.

### Usage

```
data("COVID_NY_FL_tracking")
```

### Format

a dataframe with 3 columns

**date** The date of the case count.

**state** The state of the case count. Either "new york" or "florida".

**caseIncrease** The increase number of positive COVID-19 cases.

### Source

Data was downloaded from The COVID Tracking Project <https://covidtracking.com/>.

---

COVID\_NY\_FL\_vaccines *plotgardener example data for COVID-19 vaccinations in New York and Florida*

---

### Description

A dataset describing groups of COVID-19 vaccinations in New York and Florida.

### Usage

```
data("COVID_NY_FL_vaccines")
```

**Format**

a dataframe with 4 columns

**state** The state of the vaccinations. Either "new york" or "florida".

**vax\_group** Character value describing the 3 possibilities for vaccination status: "not", "partially", or "fully" vaccinated.

**value** Raw state population value in vaccination group.

**percent** State percentage in vaccination group.

**Source**

State population data and state COVID-19 vaccination data were downloaded from the John Hopkins Centers for Civic Impact COVID-19 GitHub repository "<https://github.com/govex/COVID-19/>".

---

COVID_USA_cases	<i>plotgardener example map data for COVID-19 cases in the United States</i>
-----------------	--

---

**Description**

A data frame of United States map data and COVID-19 cases as of 2021-03-07.

**Usage**

```
data("COVID_USA_cases")
```

**Format**

a dataframe with 7 columns

**state** The associated state in the United States.

**group** Numeric value describing a group for each state.

**long** Longitude value.

**lat** Latitude value.

**cases** The cumulative number of COVID-19 cases.

**population** Numeric value of total state population.

**cases\_100K** The cumulative number of COVID-19 cases, per 100000 individuals.

**Source**

COVID-19 case data was downloaded from The COVID Tracking Project <https://covidtracking.com/>. Data was turned into map data with [map\\_data](#).

---

GM12878\_ChIP\_CTCF\_signal

*plotgardener example GM12878 CTCF ChIP signal data*

---

### Description

A dataset listing read depths across the genome resulting from CTCF ChIP-seq in the GM12878 cell line. Genomic coordinates fall within the region chr21:28000000-30300000 according to the hg19 genome build.

### Usage

```
data("GM12878_ChIP_CTCF_signal")
```

### Format

a dataframe in BED format with a "score" column

**chrom** The name of the chromosome on which the genome feature exists.

**start** The starting position of the feature in the chromosome.

**end** The ending position of the feature in the chromosome.

**score** Score value of read depth.

### Source

Data from **Michael Snyder, Stanford** with accession number **ENCFF312KXX** was downloaded from the ENCODE portal <https://www.encodeproject.org/>.

### References

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6;489(7414):57-74. doi: 10.1038/nature11247. PMID: 22955616; PMCID: PMC3439153.

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, Cherry JM. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018 Jan 4;46(D1):D794-D801. doi: 10.1093/nar/gkx1081. PMID: 29126249; PMCID: PMC5753278.

---

GM12878\_ChIP\_H3K27ac\_signal

*plotgardener example GM12878 H3K27ac ChIP signal data*

---

### Description

A dataset listing read depths across the genome resulting from H3K27ac ChIP-seq in the GM12878 cell line. Genomic coordinates fall within the region chr21:28000000-30300000 according to the hg19 genome build.

**Usage**

```
data("GM12878_ChIP_H3K27ac_signal")
```

**Format**

a dataframe in BED format with a "score" column

**chrom** The name of the chromosome on which the genome feature exists.

**start** The starting position of the feature in the chromosome.

**end** The ending position of the feature in the chromosome.

**score** Score value of read depth.

**Source**

Data with reference epigenome identifier **E116** was downloaded from the NIH Roadmap Epigenomics Project <http://www.roadmapepigenomics.org/>.

**References**

Roadmap Epigenomics Consortium., Integrative analysis coordination., Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330 (2015). <https://doi.org/10.1038/nature14248>

---

GM12878\_HiC\_10kb

*plotgardener example GM12878 Hi-C data at 10 Kb resolution*

---

**Description**

A dataset containing interaction frequency matrix counts along genomic coordinates in the region chr21:28000000-30300000 according to the hg19 genome build. This data is from the GM12878 cell line.

**Usage**

```
data("GM12878_HiC_10kb")
```

**Format**

a 3-column data frame in sparse upper triangular format.

**References**

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014 Dec 18;159(7):1665-80. doi: 10.1016/j.cell.2014.11.021. Epub 2014 Dec 11. Erratum in: *Cell*. 2015 Jul 30;162(3):687-8. PMID: 25497547; PMCID: PMC5635824. ([PubMed](#))

---

hg19\_insulin\_GWAS      *plotgardener example insulin GWAS data*

---

### Description

A dataset representing GWAS data from a GWAS study of insulin response with coordinates based on the hg19 genome build.

### Usage

```
data("hg19_insulin_GWAS")
```

### Format

a dataframe with the following columns:

**chrom** The name of the chromosome of the SNP.

**pos** The basepair position of the SNP.

**p** The p-value of the SNP.

**snp** The rsID of the SNP.

**LD** A simulated linkage disequilibrium score for the SNP.

### Source

GWAS summary statistics were downloaded from LocusZoom <http://locuszoom.org/>.

### References

Prokopenko I, Poon W, Mägi R, Prasad B R, Salehi SA, Almgren P, Osmark P, Bouatia-Naji N, Wierup N, Fall T, Stančáková A, Barker A, Lagou V, Osmond C, Xie W, Lahti J, Jackson AU, Cheng YC, Liu J, O'Connell JR, Blomstedt PA, Fadista J, Alkayyali S, Dayeh T, Ahlqvist E, Taneera J, Lecoeur C, Kumar A, Hansson O, Hansson K, Voight BF, Kang HM, Levy-Marchal C, Vatín V, Palotie A, Syvänen AC, Mari A, Weedon MN, Loos RJ, Ong KK, Nilsson P, Isomaa B, Tuomi T, Wareham NJ, Stumvoll M, Widen E, Lakka TA, Langenberg C, Tönjes A, Rauramaa R, Kuusisto J, Frayling TM, Froguel P, Walker M, Eriksson JG, Ling C, Kovacs P, Ingelsson E, McCarthy MI, Shuldiner AR, Silver KD, Laakso M, Groop L, Lyssenko V. A central role for GRB10 in regulation of islet function in man. *PLoS Genet.* 2014 Apr 3;10(4):e1004235. doi: 10.1371/journal.pgen.1004235. PMID: 24699409; PMCID: PMC3974640.

---

IMR90\_ChIP\_CTCF\_reads      *plotgardener example CTCF read data*

---

### Description

A dataset listing aligned sequencing reads for CTCF in the IMR90 cell line as determined by ChIP-seq. Genomic coordinates fall within the region chr21:28000000-30300000 according to the hg19 genome build.

**Usage**

```
data("IMR90_ChIP_CTCF_reads")
```

**Format**

a dataframe in BED (ranges) format

**chrom** The name of the chromosome on which the genome feature exists.

**start** The starting position of the feature in the chromosome.

**end** The ending position of the feature in the chromosome.

**strand** An optional column defining the strand of the feature as either '+' or '-'.

**Source**

Data from **Michael Snyder, Stanford** with accession number **ENCFF847VPR** was downloaded from the ENCODE portal <https://www.encodeproject.org/>.

**References**

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6;489(7414):57-74. doi: 10.1038/nature11247. PMID: 22955616; PMCID: PMC3439153.

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymurov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, Cherry JM. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018 Jan 4;46(D1):D794-D801. doi: 10.1093/nar/gkx1081. PMID: 29126249; PMCID: PMC5753278.

---

IMR90\_ChIP\_CTCF\_signal

*plotgardener example IMR90 CTCF ChIP signal data*

---

**Description**

A dataset listing read depths across the genome resulting from CTCF ChIP-seq in the IMR90 cell line. Genomic coordinates fall within the region chr21:28000000-30300000 according to the hg19 genome build.

**Usage**

```
data("IMR90_ChIP_CTCF_signal")
```

**Format**

a dataframe in BED format with a "score" column

**chrom** The name of the chromosome on which the genome feature exists.

**start** The starting position of the feature in the chromosome.

**end** The ending position of the feature in the chromosome.

**score** Score value of read depth.

**Source**

Data from **Michael Snyder, Stanford** with accession number **ENCFF603PYX** was downloaded from the ENCODE portal <https://www.encodeproject.org/>.

**References**

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6;489(7414):57-74. doi: 10.1038/nature11247. PMID: 22955616; PMCID: PMC3439153.

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, Cherry JM. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018 Jan 4;46(D1):D794-D801. doi: 10.1093/nar/gkx1081. PMID: 29126249; PMCID: PMC5753278.

---

IMR90\_ChIP\_H3K27ac\_signal

*plotgardener example IMR90 H3K27ac ChIP signal data*

---

**Description**

A dataset listing read depths across the genome resulting from H3K27ac ChIP-seq in the IMR90 cell line. Genomic coordinates fall within the region chr21:28000000-30300000 according to the hg19 genome build.

**Usage**

```
data("IMR90_ChIP_H3K27ac_signal")
```

**Format**

a dataframe in BED format with a "score" column

**chrom** The name of the chromosome on which the genome feature exists.

**start** The starting position of the feature in the chromosome.

**end** The ending position of the feature in the chromosome.

**score** Score value of read depth.

**Source**

Data with reference epigenome identifier **E017** was downloaded from the NIH Roadmap Epigenomics Project <http://www.roadmapepigenomics.org/>.

**References**

Roadmap Epigenomics Consortium., Integrative analysis coordination., Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330 (2015). <https://doi.org/10.1038/nature14248>



---

IMR90\_DNAloops\_pairs *plotgardener example DNA loop pair data*

---

### Description

A dataset listing interaction data along genomic coordinates in the region chr21:28000000-30300000 according to the hg19 genome build. This data represents called DNA loops in the IMR90 cell line.

### Usage

```
data("IMR90_DNAloops_pairs")
```

### Format

a dataframe in BEDPE (paired ranges) format

**chrom1** The name of the chromosome on which the first end of the feature exists.

**start1** The starting position of the first end of the feature on chrom1.

**end1** The ending position of the first end of the feature on chrom1.

**chrom2** The name of the chromosome on which the second end of the feature exists.

**start2** The starting position of the second end of the feature on chrom2.

**end2** The ending position of the second end of the feature on chrom2.

### References

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014 Dec 18;159(7):1665-80. doi: 10.1016/j.cell.2014.11.021. Epub 2014 Dec 11. Erratum in: *Cell*. 2015 Jul 30;162(3):687-8. PMID: 25497547; PMCID: PMC5635824. ([PubMed](#))

---

IMR90\_HiC\_10kb *plotgardener example IMR90 Hi-C data at 10 Kb resolution*

---

### Description

A dataset containing interaction frequency matrix counts along genomic coordinates in the region chr21:28000000-30300000 according to the hg19 genome build. This data is from the IMR90 cell line.

### Usage

```
data("IMR90_HiC_10kb")
```

### Format

a 3-column data frame in sparse upper triangular format.

**References**

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014 Dec 18;159(7):1665-80. doi: 10.1016/j.cell.2014.11.021. Epub 2014 Dec 11. Erratum in: *Cell*. 2015 Jul 30;162(3):687-8. PMID: 25497547; PMCID: PMC5635824. ([PubMed](#))

# Index

## \* datasets

- COVID\_NY\_FL\_tracking, 2
- COVID\_NY\_FL\_vaccines, 2
- COVID\_USA\_cases, 3
- GM12878\_ChIP\_CTCF\_signal, 4
- GM12878\_ChIP\_H3K27ac\_signal, 4
- GM12878\_HiC\_10kb, 5
- hg19\_insulin\_GWAS, 6
- IMR90\_ChIP\_CTCF\_reads, 6
- IMR90\_ChIP\_CTCF\_signal, 7
- IMR90\_ChIP\_H3K27ac\_signal, 8
- IMR90\_DNAloops\_pairs, 9
- IMR90\_HiC\_10kb, 9

- COVID\_NY\_FL\_tracking, 2
- COVID\_NY\_FL\_vaccines, 2
- COVID\_USA\_cases, 3

- GM12878\_ChIP\_CTCF\_signal, 4
- GM12878\_ChIP\_H3K27ac\_signal, 4
- GM12878\_HiC\_10kb, 5

- hg19\_insulin\_GWAS, 6

- IMR90\_ChIP\_CTCF\_reads, 6
- IMR90\_ChIP\_CTCF\_signal, 7
- IMR90\_ChIP\_H3K27ac\_signal, 8
- IMR90\_DNAloops\_pairs, 9
- IMR90\_HiC\_10kb, 9

- map\_data, 3