Description of test data

This is a description file for test data sets provided in the GitHub repository of HTSvis https://github.com/boutroslab/HTSvis . HTSvis is a shiny app for exploratory analysis and visualization of arrayed high-throughput screens. Detailed information concerning the required data structure and instructions for data input can be found on the help page of HTSvis.

humanSGI.RData

This table contains extracted features from an image-based RNAi screen in human cells [1]. Features describing phenotypes of single cells were extracted using image analysis tools and averaged over all cells per well. The original table contained row-wise entries per well with 353 features (one column per feature). For reasons of simplicity and data size we selected 10 features for this example data set, e.g. cell count, shape or intensity measurements. The screen was performed in replicates with 85 384-well plates per replicate. Columns containing the well, plate and experiment allocation are named *well, plate* and *replicate*. No check has to be set as the default setting is for tables with an experiment column. An annotation column (named accordingly) with a per-well annotation of RNAi reagent IDs is present. Positive controls for cell viability of different strength are located in wells of every second row of column 23 starting from row B (B23 – P23). Negative controls are located in the alternating pattern of the positive controls in column 23 (C23 - 023). Wells of row A and column 24 were left empty. Note that each plate within this screen received a separate treatment with a second RNAi reagent for all sample wells.

topTable.txt

This is a result table from a cell-based RNAi screen in *Drosophila melanogaster* Kc_{167} cells with a single channel viability readout obtained upon analysis using the Bioconductor/R package cellHTS2 [1,2]. Accordingly, the *'cellHTS topTable'* check has to be set on the data input tab. The table

(topTable.txt) was generated following the cellHTS2 vignette which can be found at https://www.bioconductor.org/packages/release/bioc/html/cellHTS2.html.

The screen was conducted in 57 384-well plates per replicate with two replicates in total. Negative controls are located in well 'B01', positive controls in well 'B02' for all plates over the entire screen. Additional negative controls are located in wells A01 and A02. Tables in the cellHTS *tobTable* format have columns with data points and metrics per cannel and are termed 'Experiments/channels' columns here. Experiments/channels that can be chosen are: raw_r1_ch1, raw_r2_ch1, median_ch1 diff_ch1, raw/PlateMedian_r1_ch1, raw/PlateMedian_r2_ch1, normalized_r1_ch1, normalized_r2_ch1 and score. Those columns contain the raw and normalized measured values per well and further metrics (e.g. raw/plateMedian_r1_ch1) or the score (summarized over both replicates). Four different well annotation columns can be chosen by the user: wellAnno, finalWellAnno, GeneID, HFAID. The annotation was carried out using cellHTS2 and the wellAnno/finalWellAnno columns contain the annotation of sample/control wells (wellAnno) and of flagged wells (finalWellAnno). GeneID and HFAID are species- or project-specific annotations where the GeneID column contains CG-identifiers for the genes of *Drosophila melanogaster* and HFAID is a project specific annotation. Columns containing the well and plate identifiers are named *well* and *plate*.

topTable_dc.txt

This is a result table from a cell-based RNAi screen in *Drosophila melanogatser* cells with a dual channel reporter readout obtained upon analysis using the Bioconductor/R package cellHTS2 [1,2]. Accordingly, the *'cellHTS topTable'* check has to be set on the data input tab. The table (topTable_dc.txt) was generated following the cellHTS2 vignette for dual channel experiments, which can be found at https://www.bioconductor.org/packages/release/bioc/html/cellHTS2.html. This example data set contains 2 replicates with three 384-well plates per

replicate. During data acquisition two channels per well were measured, one channel signal (ch1) was used as a proxy for cell viability whilst the second channel (ch2) was used to capture a pathway specific reporter signal. Due to the assay type, controls are channel specific: in channel 1, there is one negative control, named geneA in well A01, and no positive controls. In the pathway-specic reporter channel 2 there are two different negative controls (geneA in well A01 and geneB in well A02), and two different positive controls (geneC in well B01 and geneD in well B02). Two annotation columns are available: wellAnno, finalWellAnno. The annotation was carried out using cellHTS2 and wellAnno/finalWellAnno contain the annotation of sample/control (wellAnno) wells and of flagged wells (finalWellAnno). The remaining columns are Experiments/channels and contain the raw and normalized measured values per well and further metrics (e.g. raw/plateMedian_r1_ch1) or the score (summarized over both replicates) and the user can select columns of interest.

96wellFACS.csv

This is table contains data from a high-throughput flow-cytometry experiment with two 96-well plates. We chose this example to demonstrate the flexibility of the application to be used with data obtained in different plate- and assay types. There is no experimental correlation between the two plates and the plates do not contain controls. As this table contains a single set of plates without a column with the experiment allocation, the *'single experiment'* check hast to be set on the data input tab.

References:

1. Laufer C, Fischer B, Billmann M, Huber W, Boutros M. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. Nature Methods. 2013;10:427–31.

2. Boutros M. Genome-Wide RNAi Analysis of Growth and Viability in Drosophila Cells. Science. 2004;303:832–5.

3. 14. Boutros M, Brás LP, Huber W. Analysis of cell-based RNAi screens. Genome Biol. 2006;7:R66.