Description of scripts used to run and evaluate SNACS

The script “**functions.R**” includes the functions that are used by the scripts that run SNACS and estimate its performance for the core data illustrated in the SNACS manuscript.

The script “**truthCall.R**” is used to make “truth calls” for the empiric multi-sample experiments. Within this script, the main function to make the calls is named “makeTruthCall.R”.

To run SNACS on the core multi-sample empiric experiment data (Experiments 5-7\_ illustrated in the SNACS manuscript:  
  
The script "**runSNACS.R**" runs SNACS on the empiric data from the core multi-sample experiments (Experiments 5-7) (see *Methods*, section 3.1 of the manuscript). The script first creates raw SNACS objects from input hdf5 files. Then, it runs SNACS on those objects to make hash calls. Next, it generates heatmaps of the SNP data based on the SNPs used to make the hash calls for all single cells; cells are annotated with both the hash data and SNACS calls. Next, it incorporates the optional doubletD algorithm to make SNACS + doubletD calls. Finally, it regenerates the heatmaps annotated with those calls. For these heatmaps, ot creates a legend of the colors used to depict the samples using the sampleColorLegend function in heatmap4 package. This provides a useful tool to visualize how well the SNACS and SNACS plus doubletD algorithms demultiplexed the data.

To estimate the performance of SNACS on the above empiric experimental data:

The script “**accuracy\_script.R**” estimates the accuracy with which SNACS makes calls (see *Methods,* section 3.2 of the manuscrupt). It uses the functions “getAnnoForAccuracy”, “getTruthCall”, and “createAnnotatedHeatmap” from the script “**functions.R**” and “makeTruthCall” from the script “**truthCall.R**”. This script compares the multi-sample experiments (Experiments 5-7) to the single-sample experiments (Experiments 1-4). First, it creates raw SNACS objects of the single-sample experiments from hdf5 files. Next, it filters the SNP data and imputes missing values in the SNP data, as done by SNACS with for the multi-sample experiments (*Results,* section 4.5) using the “filterData” and “imputeMissingMutations” functions from SNACS package. Then, it creates annotation tables containing both the SNACS calls and the genotypes used to make these calls for each single cell in a given multi-sample experiment. It uses the function “getAnnoForAccuracy” from the script “**functions.R**” to create the tables. Next, it creates tables of truth calls from the single-sample experiments using the“makeTruthCall” function from the “**truthCall.R**” script. It then adds the truth calls to the above cell annotation tables using the “getAnnoForAccuracy” function from the script “**functions.R**”. To visually evaluate the performance of SNACS, it also creates heatmaps based on the SNP data of both the multi-sample experiment annotated with the truth calls and of the constituent single-sample experiments. Lastly, the script computes the accuracy, sensitivity and specificity metrics for the multi-sample experiments.

To make “truth calls” for all cells:

The script “**truthCallForAllCells.R**” makes truth calls for all the cells in the raw data of the multi-sample Experiments 5-7. It creates annotation tables with truth calls for all cells in a given multi-sample experiment. It is used to estimate the performance of the four comparison methods detailed in the manuscript (HTOdemux, CiteFuse, scSplit, and the Robinson Method).

To generate simulated data:

The scripts “**run\_simulation.R**”, “**functions\_sim.R**”, “**setup\_simulation.R**”, “create\_sim.sh “, “**create\_sim.R**”, “run\_snacs\_sim.sh”, and “**run\_snacs\_sim.R**” are used to generate simulated data (see *Methods,* section 3.4 of the manuscript).

* The script “**run\_simulation.R**” generates simulated data for a single multi-sample experiment. It is used by the script “**create\_sim.R**” to simulate multiple multi-sample experiments.
* The script “**functions\_sim.R**” includess functions that are used by the simulation scripts.
* The script “**setup\_simulation.R**” creates initial data by drawing from the empiric single-sample experiments (Experiments 1-4) that are used to simulate multi-sample experiments.
* The script “create\_sim.sh” runs the script “**create\_sim.R**”. The latter script simulates the multi-sample experiments. In order to that, for each simulation, it runs the function “sim" from script “**run\_simulation.R**”. It uses the function “truthName2snacsCallName” from the script “**functions\_sim.R** “ to properly format the names of the truth calls.
* The script “run\_snacs\_sim.sh” runs the script “**run\_snacs\_sim.R**”. The latter script runs SNACS on the simulated data. It uses the functions “modifyCallName”, “createHeatmapForSim”, and “getColVarInfo” from the script “**functions\_sim.R**” to properly format the names of the truth calls, generate heatmaps based on the SNPs, and get the appropriate variables and their colors for cell-level annotation in the heatmaps.