

## Introduction to the six molecular features supported by Transcell

Six measurement types of data used in Transcell were downloaded from the figshare or DepMap (figshare (data description [here](#)): [mutation](#), [gene expression](#), [copy number variation](#), [gene effect](#), [drug sensitivity](#). DepMap: CCLE\_RPPA\_20181003, CCLE\_metabolomics\_20190502). All the datasets can be downloaded from the [compiled RData file](#). In the RData file:

#octad\_cell\_line\_meta: cell line meta information (including disease type, etc.)

#octad\_cell\_line\_features: feature description (including gene name, etc.)

#octad\_cell\_line\_matrix: all values

### Metabolite:

The method is detailed in the [publication](#) (Li et al. *Nature Medicine*, 2019). Note that the abundance of different metabolites cannot be compared given the nature of the LC-MS methods. Only for the same metabolite, the levels could be compared between different cell lines.

### Protein:

Data are generated based on the Reverse Phase Protein Array (RPPA) technology, including isoforms and different protein states. Antibodies are listed here in the [portal](#).

### Copy number variation (CNV):

Gene level copy number data, log2 transformed with a pseudo count of 1. This is generated by mapping genes onto the segment level calls.

### Gene effect score:

The DepMap released data containing CRISPR knockout screen results from the project Achilles. A lower score means a high gene dependency (DepMap used -1 as a threshold to define the dependency).

### Drug Sensitivity:

The primary PRISM Repurposing dataset contains the results of pooled-cell line chemical-perturbation viability screens for >4,500 compounds screened against >550 cell lines. The data used in Transcell are logfold change values which are relative to DMSO and corrected for experimental confounders using ComBat. The more negative a value is, the more effective the compound is in the corresponding cell line.

### Mutation:

We define 1: mutated, 0: wild, NA: unknown (NA for those cell lines which do not have mutation profiles). The following mutation types listed in Variant\_Classification were ignored: "3'UTR", "Intron", "5'UTR", "splice\_Site\_SNP", "Silent", "De\_novo\_Start\_OutOfFrame", "De\_novo\_Start\_InFrame", "Unknown", "Substitution - coding silent", "No detectable mRNA/protein".