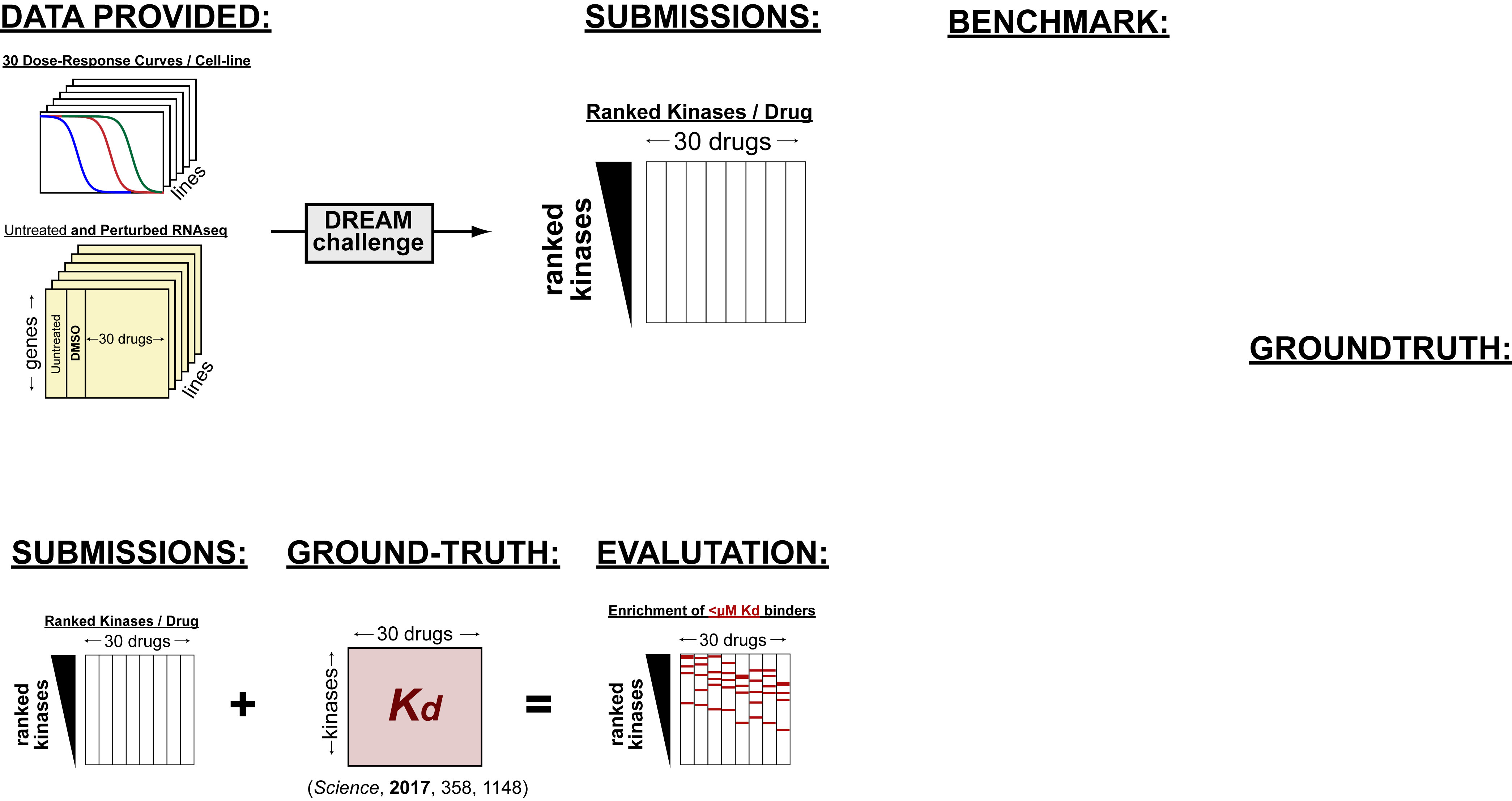
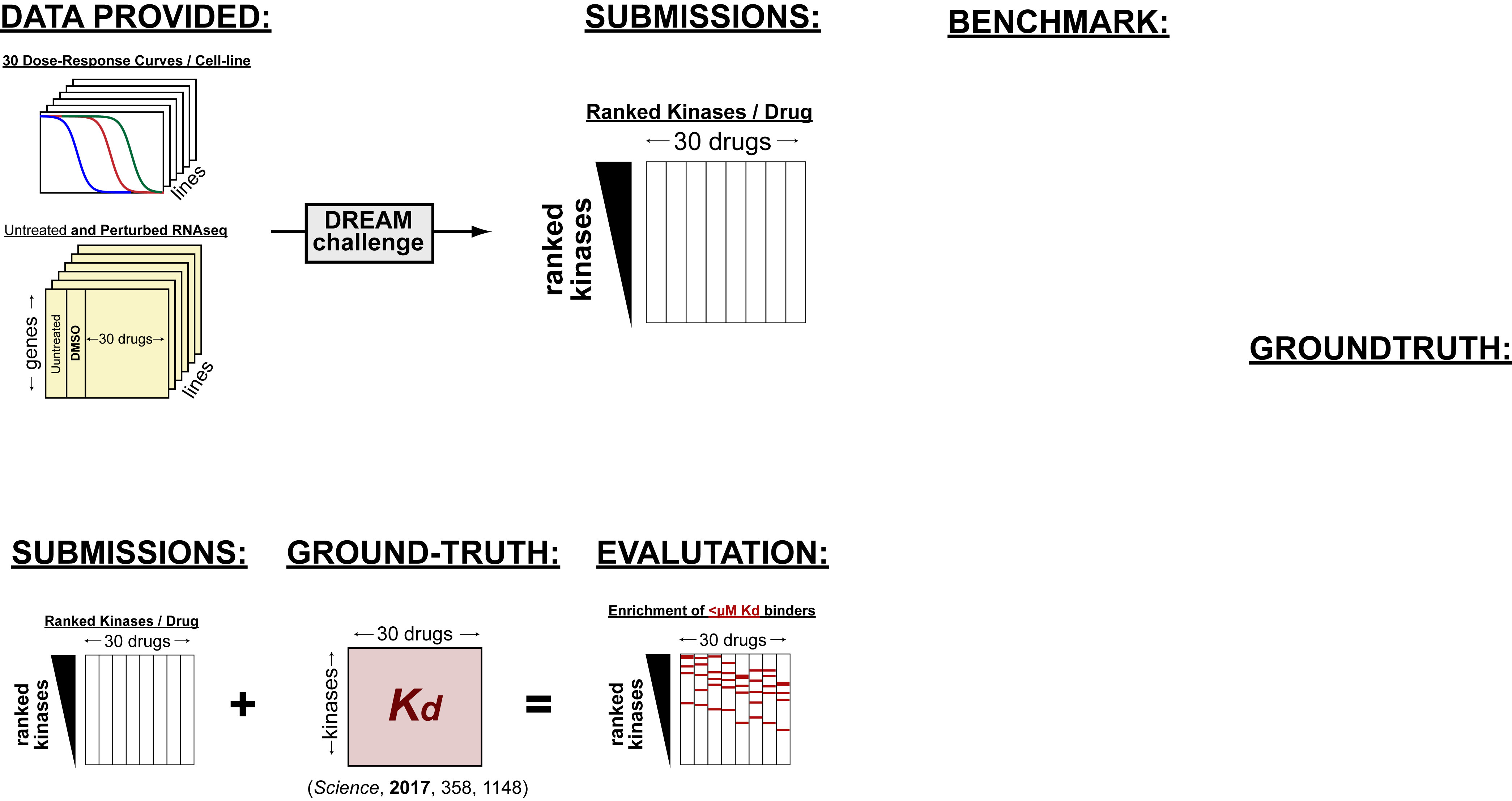
**CTD2 Pancancer Drug Activity Challenge**

CUIMC\_DREAMv1\_data.zip



The goal of the CTD2 Pancancer Drug Activity DREAM Challenge is to foster the development and benchmarking of algorithms to predict targets of chemotherapeutic compounds from post-treatment transcriptional data. The drug perturbational profiles on 11 cell lines and their dose-response curves for 32 chosen compounds with well-established targets will be provided to challenge participants, without revealing the identity of the drugs. These profiles will be removed from any public dataset and added back only after the challenge is completed. Transcriptional profiles for all the cell lines in which the compounds have been profiled have been provided to challenge participants, including the specific concentration at which the compound was titrated.



The package contains 2 metadata files, 22 data files, a README file that describes the data, and a COLUMNS file of descriptions of column headers shared by the 24 data files.

**Methods overview**

This dataset was developed in collaboration between Columbia University Irving Medical Centers (CUIMC)’s High Throughput Screening Center (HTS), Sulzberger Genome Center and the Califano Laboratory in the Department of Systems Biology. Briefly, HTS handled cell-culture, cell-perturbation experiments and RNA extraction; the Genome Center performed RNA sequencing and the Califano laboratory performed data normalization, quality control, benchmarking and scientific and statistical analysis.

**Compound titration curves**

To determine the 48h ED20 of each drug, cell lines were plated into 96-well tissue culture plates, in 100 μL total volume, and incubated at 37°C. After 16 hours the plates were removed from the incubator and compounds were transferred into assay wells (1 μL) in triplicate. Plates were then returned to the incubator. After 48 hours the assay plates were removed from the incubator and allowed to cool to room temperature prior to the addition of 100 μL of CellTiter-Glo (Promega Inc.) per well. The plates were then mechanically shaken for 5 minutes prior to readout on the EnVision Multi-Label Reader (Perkin Elmer Inc.) using the enhanced luminescence module. Relative cell viability was computed using matched DMSO control wells as reference. ED20 was estimated by fitting a four-parameter sigmoid model to the titration results.

**Perturbational profile generation**

Using the previously described plating and perturbation procedure we perturbed each cell-line with each drug at its 48h ED20 value (measured above) or its CMax concentration. In order to optimize the clinical translation potential of the perturbation databases, we used the CMax, defined as the maximum plasma concentration after the administration of the drug at the maximum tolerated dose in patients, (whenever available from published pharmacokinetic studies), as an upper bound for the perturbation studies (Table S1). The mRNA from these cells was isolated and profiled by PLATESeq (Nat. Commun. 2017, 8, 105) at 24h after each perturbation.

**Profile normalization**

RNASeq reads were mapped for each well to the human reference genome assembly 38 using the STAR aligner,57 version 2.5.2b. Individual plates counts files were then combined, normalized and corrected for batch effects. First, individual counts files were combined across genes and ERCC2 spike-in counts removed, yielding the raw counts file for each cell-line experiment. Second, raw counts were quantile normalized and variance stabilized based on the negative binomial distribution with the DESeq R system package.59 To account for plate-based batch effects (which are common with drug-perturbed transcriptomic data) normalized expression was batch corrected using ComBat.60

***metadata files***

* **CUIMC\_DREAMv1\_DrugLibrary.csv:** metadata file. pathway, target and clinical annotations per compound
* **drugTargets\_bindingconstants.csv:** metadata file. -log10(Kd) measurements of affinity for 32 drugs across 255 kinases as measured in Science 2017 358, 1148. This was used as the ground truth to evaluate Submission performance (see figure below)

***data files***

* dose-reponses folder:
  + **dose-responses/ASPC1\_DoseResponses.csv:** dose-response data.average viability for 32 compounds by 12 concentrations tested on ASPC1 cell-line
  + **dose-responses/DU145\_DoseResponses.csv:** average viability for 32 compounds by 12 concentrations tested on DU145 cell-line
  + **dose-responses/EFO21\_DoseResponses.csv:** dose-response data.average viability for 32 compounds by 12 concentrations tested on EFO21 cell-line
  + **dose-responses/H1793\_DoseResponses.csv:** dose-response data.average viability for 32 compounds by 12 concentrations tested on NCIH1793 cell-line
  + **dose-responses/HCC1143\_DoseResponses.csv:** dose-response data.average viability for 32 compounds by 12 concentrations tested on HCC1143 cell-line
  + **dose-responses/HF2597\_DoseResponses.csv:** dose-response data.average viability for 32 compounds by 12 concentrations tested on HF2597 cell-line
  + **dose-responses/HSTS\_DoseResponses.csv:** dose-response data.average viability for 32 compounds by 12 concentrations tested on HSTS cell-line
  + **dose-responses/KRJ1\_DoseResponses.csv:** dose-response data.average viability for 32 compounds by 12 concentrations tested on KRJ1 cell-line
  + **dose-responses/LNCAP\_DoseResponses.csv:** dose-response data.average viability for 32 compounds by 12 concentrations tested on LNCAP cell-line
  + **dose-responses/PANC1\_DoseResponses.csv:** dose-response data.average viability for 32 compounds by 12 concentrations tested on PANC1 cell-line
  + **dose-responses/U87\_DoseResponses.csv:** dose-response data.average viability for 32 compounds by 12 concentrations tested on U87MG cell-line
* rnaseq folder
  + **rnaseq/ASPC1\_RNAseq.csv:** rnaseq data.untreated, DMSO-treated and drug-DMSO treated RNAseq profile for ASPC1 cell-line across 32 drugs
  + **rnaseq/DU145\_ RNAseq.csv:** rnaseq data.untreated, DMSO-treated and drug-DMSO treated RNAseq profile for ADU145 cell-line across 32 drugs
  + **rnaseq/EFO21\_ RNAseq.csv:** rnaseq data.untreated, DMSO-treated and drug-DMSO treated RNAseq profile for EFO21 cell-line across 32 drugs
  + **rnaseq/H1793\_ RNAseq.csv:** rnaseq data.untreated, DMSO-treated and drug-DMSO treated RNAseq profile for NCIH1793 cell-line across 32 drugs
  + **rnaseq/HCC1143\_ RNAseq.csv:** rnaseq data.untreated, DMSO-treated and drug-DMSO treated RNAseq profile for HCC1143 cell-line across 32 drugs
  + **rnaseq/HF2597\_ RNAseq.csv:** rnaseq data.untreated, DMSO-treated and drug-DMSO treated RNAseq profile for HF2597 cell-line across 32 drugs
  + **rnaseq/HSTS\_ RNAseq.csv:** rnaseq data.untreated, DMSO-treated and drug-DMSO treated RNAseq profile for HSTS cell-line across 32 drugs
  + **rnaseq/KRJ1\_ RNAseq.csv:** rnaseq data.untreated, DMSO-treated and drug-DMSO treated RNAseq profile for KRJ1 cell-line across 32 drugs
  + **rnaseq/LNCAP\_ RNAseq.csv:** rnaseq data.untreated, DMSO-treated and drug-DMSO treated RNAseq profile for LNCAP cell-line across 32 drugs
  + **rnaseq/PANC1\_ RNAseq csv:** rnaseq data.untreated, DMSO-treated and drug-DMSO treated RNAseq profile for PANC1 cell-line across 32 drugs