

DRUG-seq U2OS MoA Box

[GitHub](#) Novartis DRUG-seq

[Zenodo](#) Novartis DRUG-seq U2OS MoABox Dataset

Authorship

Andrea Hadjikyriacou, Chian Yang, Martin Henault, Robin Ge, Leandra Mansur, Alicia Lindeman, Carsten Russ, Steffen Renner, Marc Hild, Jeremy Jenkins, Caroline Gubser-Keller, Jingyao Li, Daniel J. Ho, Marilisa Neri, Frederic D. Sigoillot, Robert Ihry

Background

DRUG-seq is an open-source high throughput transcriptomics platform. In order to facilitate the development of new computational analysis methods, that require large-scale unbiased transcriptomic data, we are making the DRUG-seq MoA Box dataset publicly available to drive innovation. Please reference the dataset using Ye et. al., 2018 [1], and the associated GitHub and Zenodo links provided. We look forward to hearing about the new methods development and hoping to see the future publications that follow!

We profiled a large set of small molecules ($N = 4,343$), including the previously released Mechanism of Action (MoA) Box (Canham et al., 2020 [2] and [Novartis MoA Box GitHub](#)) in the Osteosarcoma U2-OS cell line in a 4-dose response (0.01, 0.1, 1 and 10 microMolar) at one treatment timepoint (24.0hrs), using the high-throughput transcriptomics profiling method DRUG-seq (Ye et al., 2018 [1], Li et al., 2022 [3], and [Novartis DRUG-seq GitHub](#)). The profiles cover an additional 640 public compound structures part of the evolving Novartis MoA Box collection.

Here we share the following metadata and data:

- plate well-level gene UMI counts
- contrast-level for each compound, dose, timepoint and batch: gene fold change and associated statistics.
- reduced feature space derivatives of the UMI counts (PCA, UMAP) results (and associated pre-filtered and normalized count matrix data files) are provided before/after limma RemoveBatchEffect batch correction.

Dataset Content Statistics and processing information

1. Content description and metadata files

The dataset consists of 52 batches of 3 replicate 384-well plates (156 x 384-well plates), resulting in a total of $156 \times 384 = 59,904$ Next Generation Sequencing samples (uniquely identified by biosample_id/external_biosample_id columns in provided metadata) and associated transcriptome profiles. The subset of unused wells ($N = 3276$) was annotated with the filler well_type=EMPTY and water (H₂O) identifier (cmpd_sample_id: EC-27-RY89). There were 56,628 wells=samples effectively used in the data processing post UMI counts generation, representing 4,343 unique compounds tested in 4-dose response and triplicate (one replicate per plate across 3 replicate plates in a batch), except for a few positive controls (Homoharringtonine cmpd_sample_id: EA-18-FP00; BTdCPU cmpd_sample_id: SE-15-AV21; cmpd_sample_id: BD-11-DV28) shared between all batches and plates that were represented in triplicate per plate (9 wells per batch) at 10 microMolar and the Reference Control (RC) DMSO wells (cmpd_sample_id: CB-43-EP73) represented in column 23 (16 wells per plate). The positive controls were selected with the expectation they can have a large impact on cell proliferation and transcriptome profile.

a. Metadata files (3 files) for 59,904 DRUG-seq samples and MoABox public compounds annotation:

./DRUGseq_U2OS_MoABox_plate_wells_metadata_public.txt # contains plate and well-level metadata for the 59,904 DRUG-seq samples including plate and well identifiers, compound identifier (cmpd_sample_id), dose, timepoint. The cmpd_sample_id column allows mapping to compounds metadata in the next file.

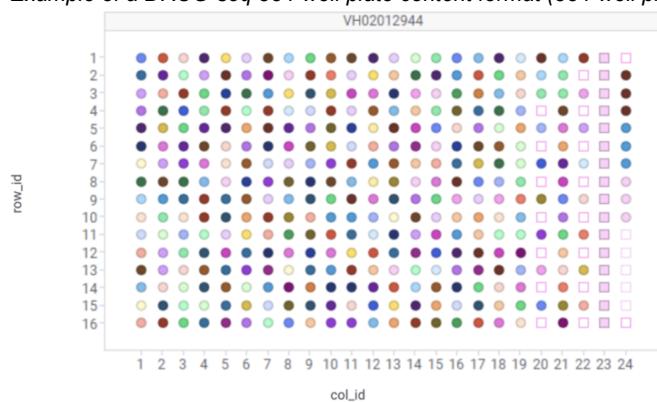
./MoABox_compounds_metadata.txt # contains 4,345 compounds (4,343 compounds + DMSO + Water) with annotation of structure (inchi_key and SMILES) and mechanism of action (MoA) when available.

./MoABox_compounds_target_pairs_public.txt # contains row-separated compounds (inchi_key) / target (gene, symbol) pairs annotation when available.

For reference:

- Water (H₂O; well_type = EMPTY): cmpd_sample_id: EC-27-RY89; inchi_key: 'XLYOFNOQVPJJNP-UHFFFAOYSA-N'; SMILES: 'O'
- Dimethyl Sulfoxide (DMSO; well_type=RC): cmpd_sample_id: CB-43-EP73; inchi_key: 'IAZDPXIOMUYVGZ-UHFFFAOYSA-N'; SMILES: 'CS(C)=O'
- Wells (NGS samples) are uniquely identified by metadata columns including 'biosample_id'/'external_biosample_id'
- Compound samples are uniquely identified by metadata columns including 'cmpd_sample_id'

Example of a DRUG-seq 384-well plate content format (384-well plate VH02012944)



- Tested samples (well_type = 'SA') are represented as circles
- The series of 16 Reference Control (well_type = 'RC') DMSO-treated wells are represented by filled squares in column 23
- Non-effective wells (no treatment) are represented by empty squares (well_type = EMPTY) and annotated as the water cmpd_sample_id.

- SA wells in column 24 consist of 3 replicates of 10 microMolar for each of 3 positive controls (BD-11-DV28, BTdCPU and Homoharringtonine).

b. Genes annotation files

./drugseq_ensembl_v98_annotation_and_entrez_mapping.RData # contains 3 items, drugseq_ensg_v98 has DRUG-seq data gene.ID to Ensembl (Version 98) ENSG gene identifiers mapping and two items have full join and inner join mapping to ENTREZ gene ID and associated gene information (drugseq_ensg_v98_entrez_mapping, drugseq_ensg_v98_entrez_mapping_notNULL)

2. Data processing and data files

The dataset was processed with the Novartis DRUG-seq data processing/analysis pipeline, following the method described in Li et al., 2022 [3] with the exception that after the TRUE NULL first step, two robust Reference Control (RC) DMSO wells were selected per plate, hence a total of 6 RC wells per batch, and used in fold change calculations final step. In order to generate the PCA and UMAP views, samples for contrasts defined as 'active' over TRUE NULL-defined background activity threshold (global threshold across 52 batches, 5-percentile of DMSO versus DMSO contrasts evaluated: >17 differentially expressed genes).

a. plate_well-level gene UMI counts are provided for all 59,904 wells in the file (well_type: SA, RC and EMPTY). Please note the dataset was saved with gzip compression to minimize the file size. We recommend loading the object in R and saving it locally as e.g. uncompressed RData file to allow faster loading for subsequent processing steps.

./Exp_gzip.RData

b. Sample groups and comparisons metadata files

./robust_RC_ReferenceControl_DMSO_wells.txt # set of Reference Control (RC) DMSO wells selected as robust after the TRUE NULL step1 and used in fold change calculations per batch

./comparisons_metadata.RData # contain 2 objects, comparison_group_member_info_public shows the samples groups setup to feed into the comparisons and comparisons_info_public shows metadata for the comparisons.

c. Fold Change calculations result files

./de.RData # data.frame aggregating all 17,731contrasts across 52 batches

3. PCA and UMAP dimensionality reduction and resulting files:

a. UMI count matrix filtering, before/after batch correction

./mat.filtered_before_RemoveBatchEffect.RData # UMI count matrix after filtering samples (samples for contrasts with a number of differentially expressed genes above the threshold of 17; TRUE NULL 95%ile threshold across batches and select RC control wells) and genes per the published method, before limma::RemoveBatchEffect was applied on plate_barcodes

./mat.filtered_after_RemoveBatchEffect.RData # UMI count matrix after filtering samples (samples for contrasts with a number of differentially expressed genes above the threshold of 17; TRUE NULL 95%ile threshold across batches and select RC control wells) and genes per the published method, after limma::RemoveBatchEffect was applied on plate_barcodes

b. PCA result files (before/after batch correction)

./pca_filtered_before_RemoveBatchEffect.RData

./pca_filtered_after_RemoveBatchEffect.RData

c. UMAP results files (2D and 3D, before and after RemoveBatchEffect)

./umap_2D_coordinates_filtered_before_and_after_RemoveBatchEffect.txt

./umap_3D_coordinates_filtered_before_and_after_RemoveBatchEffect.txt

Files content excerpts:

./DRUGseq_U2OS_MoABox_plate_wells_metadata_public.txt

| analysis_id | investigation_id | investigation_name | batch_id | plate_barcode | plate_index | well_id | well_index | col | row | well_type | cell_line_name | cell_line_ncn | concentration | unit | hours_post_treatment | biosample_id | external_biosample_id | cmpd_sample_id | plate_well |
|-------------|------------------|-----------------------|----------|---------------|------------------------------|------------|------------|-----|-------------|-----------|----------------|---------------|---------------|--------------|----------------------|---------------------|-----------------------|------------------------|------------|
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 18 | 2 | 6456_U-2_O5_MoA_batch18.rep1 | VH02011424 | CTACGACA | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 0.1 μ M | 24 | 1995238_QB-67-HQ74 | IA-00-RU8 | VH02011424_AACAAAGGTAC | |
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 20 | 1 | 6502_U-2_O5_MoA_batch20.rep1 | VH02010546 | GACCGTCA | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 10 μ M | 24 | 2025322_QB-22-HQ74 | IA-00-RU8 | VH02010546_AACAAAGGTAC | |
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 6 | 1 | 6570_U-2_O5_MoA_batch6.rep1 | VH02010542 | TGGTTGTC | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 0.1 μ M | 24 | 2025322_QB-22-HQ74 | IA-00-RU8 | VH02010542_AACAAAGGTAC | |
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 9 | 3 | 6579_U-2_O5_MoA_batch9.rep1 | VH02013398 | AGCTCTAC | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 0.01 μ M | 24 | 2089227 YE-51-01019 | TF-77-DV98 | VH02013398_AACAAAGGTAC | |
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 4 | 2 | 6568_U-2_O5_MoA_batch4.rep2 | VH02011379 | TATCCAC | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 10 μ M | 24 | 2070747 ED-31-K982 | DB-78-B12 | VH02011379_AACAAAGGTAC | |
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 3 | 2 | 6511_U-2_O5_MoA_batch3.rep2 | VH02010566 | CGGACAGC | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 1 μ M | 24 | 2040819 EE-53-M939 | DB-78-B12 | VH02010566_AACAAAGGTAC | |
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 17 | 2 | 6440_U-2_O5_MoA_batch17.rep2 | VH02011024 | TCGTGTC | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 0.01 μ M | 24 | 2002153 ED-73-QM91 | IA-00-RU8 | VH02011024_AACAAAGGTAC | |
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 26 | 2 | 6441_U-2_O5_MoA_batch26.rep2 | VH02011598 | CCTTAAAG | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 0.1 μ M | 24 | 2016847 GB-07-OL88 | LA-96-RT74 | VH02011598_AACAAAGGTAC | |
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 6 | 3 | 6576_U-2_O5_MoA_batch6.rep3 | VH02010998 | TCTCTACT | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 0.1 μ M | 24 | 2086917 SC-81-UP12 | LD-48-CY81 | VH02010998_AACAAAGGTAC | |
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 16 | 2 | 6479_U-2_O5_MoA_batch16.rep2 | VH02011034 | AGCGACCT | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 10 μ M | 24 | 2001546 OD-11-CJ68 | RE-21-BF58 | VH02011034_AACAAAGGTAC | |
| 24 | 2384 | DRUG-seq_CBT_U2OS_MoA | 49 | 2 | 6428_U-2_O5_MoA_batch49.rep2 | VH02013416 | AGCGAG | A01 | AACAAAGGTAC | 1 | 1 SA | U-2-05 | FH5-48QE | 0.01 μ M | 24 | 1975320 MA-60-DC69 | JB-63-NV02 | VH02013416_AACAAAGGTAC | |

./MoABox_compounds_metadata.txt

| cmpd_sample_id | inchi_key | smiles | cas_number | moa |
|----------------|------------------------------|--|--------------|---|
| EF-73-IP33 | AAAQFGUYHFJNH-SFHVFURJKSA-N | CCNC(=O)C[C@H]1N=C(c2ccc(C)c2)cc(O)c2c-... | 1260907-17-2 | BRD2/3/4-T BET family inhibitor |
| DE-81-OY65 | AAKILRGGTJIKAMG-UHFFFAOYSA-N | C#Cc1ccc(Nc2ncnc3cc(OCCOC)c(OCCOC)c2)c1 | 1429636-49-6 | Inhibitors of Signal Transduction Pathways:EGFR (HER... |
| IC-57-ZA36 | AAOYLOCWJSLLJU-UHFFFAOYSA-N | O=C(C(=O)c1cc(Br)c1)cc1Br)cc1O | 523-88-6 | Carboxylesterase Inhibitors |
| RA-81-JZ66 | AASYYGYJPZBFKU-UHFFFAOYSA-N | O=C(O)c1c1c(C(=O)O)n(Cc2cccc3cccc2)c2(F)c12 | 942191-15-3 | Known Chymase inhibitor |
| BD-60-ZK54 | AATHJLXXARFCZ-UHFFFAOYSA-N | O=c1ccc(c2cc[n+]([O-])c2)cc1)N1CCN(S(=O)(=O)... | 216957-20-9 | Inhibitors of Blood Coagulation Pathways:Coagulation... |
| EB-09-LQ45 | AATPYXMXFBKFO-UHFFFAOYSA-N | OC1cc(C(=O)c1c2cc[n+]([O-])c2)cc1)N2CCN(S(=O)(=O)... | 865305-30-2 | P2RX2 and P2RX3 Purinergic Receptor Antagonist |
| DB-32-FS68 | ABFMMCZFKIJUQG-UHFFFAOYSA-N | O=[N+](=O)[O-]c1cc(CNCP(=O)(O)c2nc(O)c)nc21 | 1230388-35-8 | AMPA Antagonists |
| MA-45-US84 | ABCOSOMRWYSQA0-UHFFFAOYSA-N | C1c1cc2(NC3CCCC3)ncn(-c3ncnc3)c2c1Cl | | SUV420H1/H2 inhibitor;kmt5b and kmt5c inhibitor |
| JD-48-EV39 | ABGYSGBNWQSGJD-UHFFFAOYSA-N | O=C(O)c1cccc2c(=O)c3cccc3oc12 | | |

./MoABox_compounds_target_pairs_public.txt

| inchi_key | moa | gene_symbol | entrez_gene_id | tax_id | tax_name |
|-----------------------------|----------------------------------|-------------|----------------|--------|--------------|
| RKUNBYITZUJHSG-UHFFFAOYSA-N | Muscarinic receptor antagonist | CHRM1 | 1128 | 9606 | Homo sapiens |
| RKUNBYITZUJHSG-UHFFFAOYSA-N | Muscarinic receptor antagonist | CHRM2 | 1129 | 9606 | Homo sapiens |
| RKUNBYITZUJHSG-UHFFFAOYSA-N | Muscarinic receptor antagonist | CHRM3 | 1131 | 9606 | Homo sapiens |
| RKUNBYITZUJHSG-UHFFFAOYSA-N | Muscarinic receptor antagonist | CHRM4 | 1132 | 9606 | Homo sapiens |
| RKUNBYITZUJHSG-UHFFFAOYSA-N | Muscarinic receptor antagonist | CHRM5 | 1133 | 9606 | Homo sapiens |
| PJDFLNOIAU1ZSL-UHFFFAOYSA-N | GABA Aminotransferase Inhibitors | ABAT | 18 | 9606 | Homo sapiens |
| GCWIIJLJLIHUNY-UHFFFAOYSA-N | Cathepsin inhibitor | CTSV | 1515 | 9606 | Homo sapiens |
| GCWIIJLJLIHUNY-UHFFFAOYSA-N | Cathepsin inhibitor | CTSZ | 1522 | 9606 | Homo sapiens |
| GCWIIJLJLIHUNY-UHFFFAOYSA-N | Cathepsin inhibitor | CTSF | 8722 | 9606 | Homo sapiens |

./robust_RC_ReferenceControl_DMSO_wells.txt

| batch_id | plate_barcode | well_id | logR_RS4_Down | logR_RS4_Up | N | comment | investigation_id | analysis_id | pipeline_run_key | pipeline_group_run_key | plate_index | well_index | external_biosample_id | inchi_key | smiles | cas_number | moa | cmpd_sample_id |
|----------|---------------|---------|---------------|-------------|----|--------------|------------------|-------------|------------------|------------------------|-------------|------------|-----------------------|-----------------------------|-----------------------------|------------|------------|----------------|
| 13 | VH02010536 | C23 | -4.517784 | -0.5809175 | 16 | 6_best_wells | logR_RS4_Down | 2384 | 24 | 27 | 27 | GGCGCTCA | CC-17-YP73 | IADZPKOMU1YVGZ-UHFFFAOYSA-N | CS(=O)=O | N/A | CB-43-EP73 | |
| 13 | VH02010536 | Q23 | -5.1325792 | -0.50760383 | 15 | 6_best_wells | logR_RS4_Down | 2384 | 24 | 27 | 27 | GGCGCTCA | IB-40-UE78 | IADZPKOMU1YVGZ-UHFFFAOYSA-N | CS(=O)=O | N/A | CB-43-EP73 | |
| 21 | VH02010538 | F23 | -1.5185139 | -1.26413337 | 16 | 6_best_wells | logR_RS4_Down | 2384 | 24 | 27 | 27 | AATGCCCT | TGTCATTGCG | OB-61-W89 | | N/A | CB-43-EP73 | |
| 21 | VH02010538 | Q23 | -1.7323938 | -1.31242988 | 11 | 6_best_wells | logR_RS4_Down | 2384 | 24 | 27 | 27 | AATGCCCT | TTGGTAACAG | CC-82-A887 | IADZPKOMU1YVGZ-UHFFFAOYSA-N | CS(=O)=O | N/A | CB-43-EP73 |
| 21 | VH02010540 | A23 | -1.7323938 | -0.75785185 | 11 | 6_best_wells | logR_RS4_Down | 2384 | 24 | 27 | 27 | AGTACTCC | TGACAGTAAC | IB-81-EU85 | IADZPKOMU1YVGZ-UHFFFAOYSA-N | CS(=O)=O | N/A | CB-43-EP73 |
| 21 | VH02010540 | L23 | -1.6276584 | -2.0985906 | 14 | 6_best_wells | logR_RS4_Down | 2384 | 24 | 27 | 27 | AGTACTCC | TTCACGGAAG | CE-06-0887 | IADZPKOMU1YVGZ-UHFFFAOYSA-N | CS(=O)=O | N/A | CB-43-EP73 |
| 51 | UAR1Y1H1A2 | r23 | -3.073745 | -1.3647031 | 11 | 6_best_wells | logR_RS4_Down | 2384 | 24 | 27 | 27 | AGTACTCC | TTACATTTTT | IB-82-YI751 | IADZPKOMU1YVGZ-UHFFFAOYSA-N | CS(=O)=O | N/A | CB-43-EP73 |

./drugseq_ensembl_v98_annotation_and_entrez_mapping.RData (two objects)

| ensembl_gene_id | geneID | chromosome | gene_name | entrez_gene_id |
|------------------|-------------------|------------|-----------|----------------|
| ENSG000000000003 | TSPAN6_grch38_X | grch38_X | TSPAN6 | 7105 |
| ENSG000000000005 | TNMD_grch38_X | grch38_X | TNMD | 64102 |
| ENSG000000000419 | DPM1_grch38_20 | grch38_20 | DPM1 | 8813 |
| ENSG000000000457 | SCYL3_grch38_1 | grch38_1 | SCYL3 | 57147 |
| ENSG000000000460 | C1orf112_grch38_1 | grch38_1 | C1orf112 | 55732 |
| ENSG000000000938 | FGR_grch38_1 | grch38_1 | FGR | 2268 |
| ENSG000000000971 | CFH_grch38_1 | grch38_1 | CFH | 3075 |
| ensembl_gene_id | geneID | chromosome | gene_name | entrez_gene_id |
| ENSG000000000003 | TSPAN6_grch38_X | grch38_X | TSPAN6 | 7105 |
| ENSG000000000005 | TNMD_grch38_X | grch38_X | TNMD | 64102 |
| ENSG000000000419 | DPM1_grch38_20 | grch38_20 | DPM1 | 8813 |
| ENSG000000000457 | SCYL3_grch38_1 | grch38_1 | SCYL3 | 57147 |
| ENSG000000000460 | C1orf112_grch38_1 | grch38_1 | C1orf112 | 55732 |
| ENSG000000000938 | FGR_grch38_1 | grch38_1 | FGR | 2268 |
| ENSG000000000971 | CFH_grch38_1 | grch38_1 | CFH | 3075 |

./Exp_gzip.RData

Structure of the Exp R object:

List of 52 batches named '1' to '52', each a list of 3 replicate plates (VH020...), each a list of 2 objects UMI.counts and Annotation
List of 52
\$ 13:List of 3
..\$ VH02010536:List of 2
...\$ UMI.counts: num [1:59594, 1:384] 16 4 0 0 0 0 0 7 28 68 ...
... ...- attr(*, "dimnames")=List of 2
...\$: chr [1:59594] "EDC3,grch38_15" "ARHGEF10L,grch38_1" "MTCO3P39,grch38_4" "RBMY2KP,grch38_Y" ...
...\$: chr [1:384] "VH02010536_AACACCTAGT" "VH02010536_AATTGCGATG" "VH02010536_TTGGTCAGTA" "VH02010536_GTTCAATTGCC" ...
... ..\$ Annotation:data.frame': 384 obs. of 18 variables:
...\$ batch_id : int [1:384] 13 13 13 13 13 13 13 13 13 13 ...
...\$ plate_barcode : chr [1:384] "VH02010536" "VH02010536" "VH02010536" ...
...\$ plate_index : chr [1:384] "GCGCTCTA" "GCGCTCTA" "GCGCTCTA" "GCGCTCTA" ...
...\$ well_id : chr [1:384] "E02" "E08" "L19" "K12" ...
...\$ plate_replicate : int [1:384] 1 1 1 1 1 1 1 1 1 1 ...
...\$ well_index : chr [1:384] "AACCGGCGTA" "ATAACGCCCTC" "GTTCCGGTGA" "CCTTGTTATTC" ...
...\$ col : int [1:384] 2 8 19 12 4 11 13 22 14 10 ...
...\$ row : int [1:384] 5 5 12 11 12 8 10 12 4 2 ...
...\$ biosample_id : int [1:384] 1997694 1997700 1997879 1997848 1997864 1997775 1997825 1997882 1997682 1997630 ...
...\$ external_biosample_id: chr [1:384] "CE-17-BF79" "GE-15-QX77" "ID-37-NE71" "WB-39-UD77" ...
...\$ cmpd_sample_id : chr [1:384] "FC-56-ZH16" "CA-39-ZJ18" "BA-44-WG29" "ED-32-JY57" ...
...\$ well_type : chr [1:384] "SA" "SA" "SA" "SA" ...
...\$ cell_line_name : chr [1:384] "U-2 OS" "U-2 OS" "U-2 OS" "U-2 OS" ...
...\$ cell_line_ncn : chr [1:384] "FH55-48QE" "FH55-48QE" "FH55-48QE" "FH55-48QE" ...
...\$ concentration : num [1:384] 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 ...
...\$ unit : chr [1:384] "uM" "uM" "uM" "uM" ...
...\$ hours_post_treatment : chr [1:384] "24.0" "24.0" "24.0" "24.0" ...
...\$ Sample : chr [1:384] "FC-56-ZH16_0.01uM_24.0hr_13" "CA-39-ZJ18_0.01uM_24.0hr_13" "BA-44-WG29_0.01uM_24.0hr_13" "ED-32-JY57_0.01uM_24.0hr_13" ...

./mat.filtered_before_RemoveBatchEffect.RData

| | VH02010986_AGTCTCGGCA | VH02010988_AGTCTCGGCA | VH02010984_AGTCTCGGCA |
|-----------------------------|-----------------------|-----------------------|-----------------------|
| 5_8S_rRNA,grch38_KI270733.1 | 0 | 1.530421 | 0 |
| 7SK,grch38_17 | 0 | 0 | 0 |
| A1BG,grch38_19 | 7.555737 | 7.156362 | 6.96272 |

./mat.filtered_after_RemoveBatchEffect.RData

| | VH02010986_AGTCTCGGCA | VH02010988_AGTCTCGGCA | VH02010984_AGTCTCGGCA |
|-----------------------------|-----------------------|-----------------------|-----------------------|
| 5_8S_rRNA,grch38_KI270733.1 | 0.106320567 | 1.77093519 | 0.4212733 |
| 7SK,grch38_17 | -0.000911934 | 0.04019551 | 0.1809353 |
| A1BG,grch38_19 | 7.465103895 | 7.24252875 | 7.1520947 |

./pca_filtered_before_RemoveBatchEffect.RData

| | PC1 | PC2 | PC3 |
|-----------------------|-----------|------------|----------|
| VH02010986_AGTCTCGGCA | -24.53069 | 17.706027 | 22.38801 |
| VH02010988_AGTCTCGGCA | -25.95049 | -10.333201 | 18.24073 |
| VH02010984_AGTCTCGGCA | -38.43094 | 6.044003 | 21.22492 |

./pca_filtered_after_RemoveBatchEffect.RData

| | PC1 | PC2 | PC3 |
|-----------------------|-----------|------------|-----------|
| VH02010986_AGTCTCGGCA | -4.825625 | 2.113132 | 8.2946427 |
| VH02010988_AGTCTCGGCA | 4.697556 | 11.392719 | 0.4948895 |
| VH02010984_AGTCTCGGCA | -3.65496 | -13.800878 | 28.827046 |

./umap_2D_coordinates_filtered_before_and_after_RemoveBatchEffect.txt***./umap_3D_coordinates_filtered_before_and_after_RemoveBatchEffect.txt***

| UMAP1 | UMAP2 | UMAP3 | UMAP4 | RE1 | RE2 | UMAP2_RBE | UMAP3_RBE | UMAP4_RBE | date | timepoint | N_deregulated | N_up | N_down | N_LNEU | plate_barcode | well_label | batch | batch_id | replicate | external_biosample_id | well_index | well_type | cell_line | n_m | n_pca | n_pca_RBE | n_neighbors | min_dist | investigation_id | analysis_id | pipeline_run_key | pipeline_group_run_key | key | comments | cmdp_sample_id |
|------------|----------|-----------|-----------|------------|------------|-----------|-----------|-----------|------------|-----------|---------------|------------|------------|---------|---------------|------------|------------|------------|-----------|-----------------------|------------|-----------|-----------|------|-------|-----------|-------------|----------|------------------|-------------|------------------|------------------------|-----|----------|----------------|
| 3.87014628 | 1.211468 | 10.416506 | 0.7090978 | -0.0036273 | -0.0036273 | 0.9364206 | 0.01uM | 24.0nM | 2018-08-08 | 10:00 | 8 | 450001 | VH02010536 | B06 | batch13 | 13 | 1 | AA-11-CM74 | C00TGTGTC | SA | FH05-48QE | 14 | 10 | 0.2 | 2384 | 24 | 27 | 27 | 27 | 04-19-2021 | | | | | |
| 3.3602279 | 1.158786 | 11.670748 | 0.8469313 | -0.0036272 | -0.0036272 | 15.396134 | 0.01uM | 24.0nM | 2018-08-08 | 10:00 | 9 | 330687 | VH02010536 | B06 | batch13 | 13 | 1 | AA-12-CM74 | A0CTGACAC | SA | FH05-48QE | 14 | 14 | 0.1 | 2384 | 24 | 27 | 27 | 27 | YC-11-U043 | | | | | |
| 3.3484716 | 1.159053 | 11.356178 | 0.6834297 | -2.9671168 | 9.618647 | 0.01uM | 24.0nM | 119 | 48 | 71 | 400087 | VH02010536 | F10 | batch13 | 13 | 1 | AA-22-CM74 | CAGCTCAT | SA | FH05-48QE | 14 | 14 | 0.1 | 2384 | 24 | 27 | 27 | 27 | AD-45-YB25 | | | | | | |
| 2.71636891 | 1.227864 | 10.526533 | 0.8474569 | -0.0023198 | -0.0023198 | 9.0002012 | 0.01uM | 24.0nM | 2018-08-08 | 10:00 | 44 | 21 | 23 | 520533 | VH02010536 | J14 | batch13 | 13 | 1 | AA-32-CM74 | C0TGACAGC | SA | FH05-48QE | 14 | 14 | 0.1 | 2384 | 24 | 27 | 27 | 27 | AB-53-AW00 | | | |
| 3.3403465 | 1.167066 | 11.652234 | 0.7830375 | -3.183781 | 10.388650 | 0.01uM | 24.0nM | 106 | 35 | 71 | 38201 | VH02010536 | C03 | batch13 | 13 | 1 | AB-19-UAT7 | A0TAGAGAC | SA | FH05-48QE | 14 | 14 | 0.1 | 2384 | 24 | 27 | 27 | 27 | JB-5A-VB5 | | | | | | |
| 3.3267707 | 3.393563 | 0.172495 | 0.6647021 | 10.7524462 | 3.982908 | 10.0uM | 24.0nM | 5303 | 206 | 1047 | 39471 | VH02010536 | J24 | batch13 | 13 | 1 | AB-20-GE70 | TOTOCATA | PC3 | FH05-48QE | 14 | 14 | 0.1 | 2384 | 24 | 27 | 27 | 27 | SE-25-VE21 | | | | | | |

./comparisons_metadata.RData (2 objects)

- comparison_group_member_info_public

| comparison_group_id | external_bio_sample_id | plate_barcode | plate_index | well_id | well_index | comparison_group_name | cmdp_sample_id | concentration | unit | hours_post_treatment | batch_id | plate_well |
|---------------------|------------------------|---------------|-------------|------------|-------------|----------------------------------|----------------|---------------|------------|----------------------|----------|---------------------------|
| 91966 | AA-00-HE00 | VH02012992 | AAATGCCCT | B06 | ACGTAGGCAC | OC-23-QX66_0.1uM_24.0_batchid42 | OC-23-QX66 | 0.1 | uM | | 24 | 42 VH02012992_ACCTAGGGAC |
| 80292 | AA-00-HE10 | VH02010996 | AGGTATTA | A05 | ACGCCCTTCGT | FF-61-PA27_0.1uM_24.0_batchid6 | FF-61-PA27 | 0.1 | uM | | 24 | 6 VH02010996_ACGCCCTTCGT |
| 94608 | AA-00-JY58 | VH02013428 | CGCTATG | N13 | CTCAAGAAC | AC-90-QA75_0.1uM_24.0_batchid46 | AC-90-QA75 | 0.1 | uM | | 24 | 46 VH02013428_CTCAAAGGACC |
| 85326 | AA-00-JY78 | VH02010986 | GGTGAACC | L20 | GTGAAATTGG | EB-66-TU32_1uM_24.0_batchid27 | EB-66-TU32 | 1 | uM | | 24 | 27 VH02010986_GTTGAATTGG |
| 87318 | AA-00-JY88 | VH02011372 | CAACAATG | F16 | GAATTACGGC | GB-57-XV52_0.01uM_24.0_batchid25 | GB-57-XV52 | 0.01 | uM | | 24 | 25 VH02011372_GAATTACGGC |
| | | | RRRRQ | AA-00-JY98 | VH02011104 | TCATCATA | F14 | ACATCCCGCA | Sn.FF.YPQH | 10uM_24.0 batchid29 | 29 | VH02011104_ACATCCCGCA |

- comparisons_info_public

| analysis | comparison_id | comparison_name | comparison_group_id | comparison_group_name | comparison_group_id2 | comparison_group_name2 | cmdp_sample_id | concentration | unit | hours_post_treatment | |
|----------|---------------|---|---------------------|-----------------------|----------------------|------------------------|----------------|---------------|------|----------------------|----|
| | 24 | 78519 SA RE-21-BF58_0.01uM_24.0_batchid13 vs RC CB-43-EP73_0uM_24.0_batchid13 | 78883 | RE-21-BF58_0.01uM | 78884 | CB-43-EP73_0uM_2 | RE-21-BF58 | 0.01 | uM | | 24 |
| | 24 | 78519 SA RE-21-BF58_0.01uM_24.0_batchid13 vs RC CB-43-EP73_0uM_24.0_batchid13 | 78883 | RE-21-BF58_0.01uM | 78884 | CB-43-EP73_0uM_2 | RE-21-BF58 | 0.01 | uM | | 24 |
| | 24 | 78519 SA RE-21-BF58_0.01uM_24.0_batchid13 vs RC CB-43-EP73_0uM_24.0_batchid13 | 78883 | RE-21-BF58_0.01uM | 78884 | CB-43-EP73_0uM_2 | RE-21-BF58 | 0.01 | uM | | 24 |
| | 24 | 78519 SA RE-21-BF58_0.01uM_24.0_batchid13 vs RC CB-43-EP73_0uM_24.0_batchid13 | 78883 | RE-21-BF58_0.01uM | 78884 | CB-43-EP73_0uM_2 | RE-21-BF58 | 0.01 | uM | | 24 |
| | 24 | 78519 SA RE-21-BF58_0.01uM_24.0_batchid13 vs RC CB-43-EP73_0uM_24.0_batchid13 | 78883 | RE-21-BF58_0.01uM | 78884 | CB-43-EP73_0uM_2 | RE-21-BF58 | 0.01 | uM | | 24 |
| | 24 | 78519 SA RE-21-BF58_0.01uM_24.0_batchid13 vs RC CB-43-EP73_0uM_24.0_batchid13 | 78883 | RE-21-BF58_0.01uM | 78884 | CB-43-EP73_0uM_2 | RE-21-BF58 | 0.01 | uM | | 24 |

(comparison_group_id1/2 map to comparison_group_id in comparison_group_member_info_public)

/de.RData

| Investigation | analysis | pipeline_group | method | normalization | comparison_name | comparison_group_id | comparison_group_name | comparison_group_id2 | comparison_group_name2 | cmdp_sample_id | concentration | unit | hours_post_treatment | batch | gene.ID | logFC | adj.P.Val | P.Value | AveExpr | t |
|---------------|----------|----------------|--------|---------------|-----------------|-----------------------------|---|----------------------|------------------------|----------------|---------------|------|----------------------|-------------|----------|----------|-------------|-------------|---------|---|
| | 2384 | 24 | 27 | 27 | limma | quantile=0.75;TMM;log2(CPM) | SA AF-80-QK39_0.01uM_24.0_batchid1 vs RC CB-43-EP73_0uM_24.0_batchid1 | AF-80-QK39 | 0.01 | uM | 24 | 1 | COK58_grch38.2 | 0.535137331 | 1.92e-07 | 2.17e-11 | 9.685192336 | 6.796986031 | | |
| | 2384 | 24 | 27 | 27 | limma | quantile=0.75;TMM;log2(CPM) | SA AF-80-QK39_0.01uM_24.0_batchid1 vs RC CB-43-EP73_0uM_24.0_batchid1 | AF-80-QK39 | 0.01 | uM | 24 | 1 | APRT_grch38.16 | 0.838632459 | 1.92e-07 | 2.63E-11 | 8.234166551 | 6.767676382 | | |
| | 2384 | 24 | 27 | 27 | limma | quantile=0.75;TMM;log2(CPM) | SA AF-80-QK39_0.01uM_24.0_batchid1 vs RC CB-43-EP73_0uM_24.0_batchid1 | AF-80-QK39 | 0.01 | uM | 24 | 1 | MCM4_grch38.8 | 1.658912297 | 3.70e-07 | 7.60E-11 | 5.34990408 | 6.60306915 | | |

References

[1] Ye C, Ho DJ, Neri M, Yang C, Kulkarni T, Randhawa R, Henault M, Mostacci N, Farmer P, Renner S, Ihry R, Mansur L, Keller CG, McAllister G, Hild M, Jenkins J, Kaykas A.

DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. Nat Commun. 2018 Oct 17;9(1):4307. PMID: 30333485; PMCID: PMC6192987.

DOI [10.1038/s41467-018-06500-x](https://doi.org/10.1038/s41467-018-06500-x)

[2] Canham SM, Wang Y, Cornett A, Auld DS, Baeschlin DK, Patoor M, Skaanderup PR, Honda A, Llamas L, Wendel G, Mapa FA, Aspasia P Jr, Labb  -Gigu  re N, Gamber GG, Palacios DS, Schuffenhauer A, Deng Z, Nigsch F, Frederiksen M, Bushell SM, Rothman D, Jain RK, Hemmerle H, Briner K, Porter JA, Tallarico JA, Jenkins JL. Systematic Chemogenetic Library Assembly. Cell Chem Biol. 2020 Sep 17;27(9):1124-1129. doi: 10.1016/j.chembiol.2020.07.004. Epub 2020 Jul 23. PMID: 32707038.

DOI [10.1016/j.chembiol.2020.07.004](https://doi.org/10.1016/j.chembiol.2020.07.004)

[3] Li J, Ho DJ, Henault M, Yang C, Neri M, Ge R, Renner S, Mansur L, Lindeman A, Kelly B, Tumkaya T, Ke X, Soler-Llavina G, Shanker G, Russ C, Hild M, Gubser Keller C, Jenkins JL, Worringer KA, Sigoillot FD, Ihry RJ. DRUG-seq Provides Unbiased Biological Activity Readouts for Neuroscience Drug Discovery. ACS Chem Biol. 2022 Jun 17;17(6):1401-1414. doi: 10.1021/acscchembio.1c00920. Epub 2022 May 4. PMID: 35508359; PMCID: PMC9207813.

DOI [10.1021/acscchembio.1c00920](https://doi.org/10.1021/acscchembio.1c00920)

GitHub links

GitHub Novartis MoaBox

GitHub Novartis DRUG-seq

GitHub Novartis DRUG-seq U2OS MoABox Dataset

Zenodo link:

[Zenodo Novartis DRUG-seq U2OS MoABox Dataset](#)