# - Supplement S1 -
# MERIDA: Data

## 1 Data acquisition

For our analyses, we use the Genomics of Drug Sensitivity in Cancer (GDSC) cancer cell line panel (Iorio *et al.*, 2016) to obtain information on the sensitivity of cancer cell lines to drugs as well as the molecular and genetic characteristics of the cell lines. To annotate the genetic data with pre-existing knowledge on sensitivity to the cancer drugs as well as oncogenicity information of specific alterations, we leverage the information from several databases: COSMIC (Tate *et al.*, 2019), CIViC (Griffith *et al.*, 2017), OncoKB (Chakravarty *et al.*, 2017), and the Cancer Genome Interpreter (CGI) (Tamborero *et al.*, 2018). In addition to that, we use the IntOGen cancer driver gene list (Gonzalez-Perez *et al.*, 2013) and a list by Sanchez-Vega et al. (Sanchez-Vega *et al.*, 2018) to reduce the number of genes and thereby focus mainly on cancer-associated ones.
In the following, we will describe how we obtained the data.

### 1.1 Cancer cell line data

We downloaded the pre-processed mutation (whole exome sequencing with Agilent SureSelectXT Human All Exon 50Mb bait set), copy number variation (Affymetrix SNP6.0 Array), gene expression (Affymetrix Human Genome U219 Array), and drug sensitivity data (GDSC1 compounds: Syto60 and resazurin assay, GDSC2 compounds: CellTiter-Glo assay) from the GDSC website. In particular, we use the drug data (IC50 values) from release 8.0, which are divided into two database subsets depending on the used assay type (GDSC1: Syto60 and resazurin assay, GDSC2: CellTiterGlo), the gene expression data from release 7.0, and the mutation and copy number data from release 6.1. Please refer to the respective documentation on the website to obtain information on how this data was pre-processed.

### 1.2 Annotation data

The data from the GDSC database is very high-dimensional since it contains genome-wide measurements on various omics-types. As a consequence, the search for influential aberrations becomes difficult using statistical learning, a problem that is known as the

curse of dimensionality in computer science. To counteract this issue, dimension reduction or feature selection can be applied to the data before application of a particular statistical learning method or as part of the method. We decided to do a literature-driven feature selection before we start the analyses with our method. To this end, we employ the IntOGen driver gene list from release 2016.5 for filtering the genes from the gene expression and mutation data. For filtering the genes from the copy number variation data, we use a list from the supplementary material of Sanchez-Vega et al. (Sanchez-Vega *et al.*, 2018). We annotate the alterations that are present in the remaining mutation and copy number data with oncogenicity information from CIViC (release 01-Oct-2018), OncoKB (version v1.16), and CGI (version 2018/01/17), and with sensitivity information from COSMIC (version v86), CIViC, OncoKB, and CGI.

# 2 Data processing

All the data described above is used to construct the input matrices and output vectors for the 41 investigated drugs (cf. Table 1). For each drug, we consider only cell lines that have full information status on all used data sets, i.e. drug response, gene expression, mutations, and copy number variations. For each drug, we only use the drug response data belonging to either the GDSC1 or GDSC2 database since we want to ensure that different assays do not disturb the analysis. If there is drug information in GDSC1 and GDSC2, we decided to use the newer GDSC2 database if there is a sufficient number of cell lines ($> 700$) available (cf. Table 1).
Note that in general, we focused on the analysis of mTOR-inhibitors. However, since there is an unfavourable ratio between sensitive and resistant cell lines, we additionally analyzed the four drugs CX-5461, NSC319726, Niraparib, and Talazoparib, which are the drugs with the highest number of sensitive cell lines in GDSC1 and GDSC2 (see Figure 1 and Figure 2).

## 2.1 Identifier mapping

To avoid inconsistencies between the used gene identifiers of the different data sources, we mapped the genes from the IntOGen cancer driver list, the genes from the copy number driver list by Sanchez-Vega et al. and all genes from CIViC, OncoKB, CGI and COSMIC to human official gene symbols. The genes from the GDSC data were not mapped since they were given as human official gene symbols.

## 2.2 Input matrix and output vector generation

For our ILP formulation, we need binarized input and output features. In the following, we will describe how we obtained them by using all of the data sets described before.

**Output vector** The output vector specifies for each cell line whether it is sensitive (1) or resistant (0) to a specific drug. To this end, the given logarithmized IC50 values

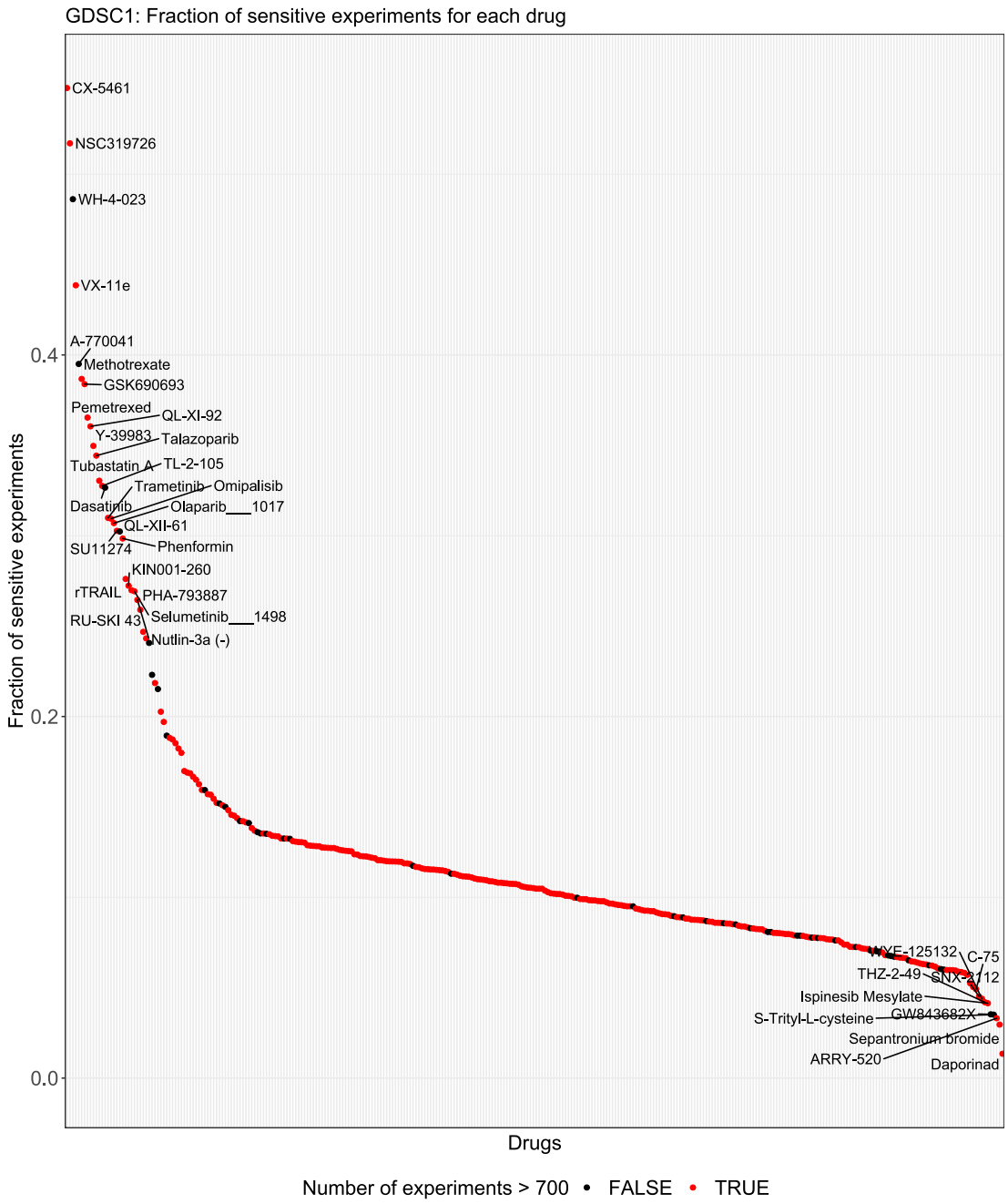


Figure 1: This figure depicts the fraction of sensitive cell lines for each drug using the data from GDSC1. Drugs that are tested on more than 700 cell lines are colored in red, otherwise the drug is colored in black.

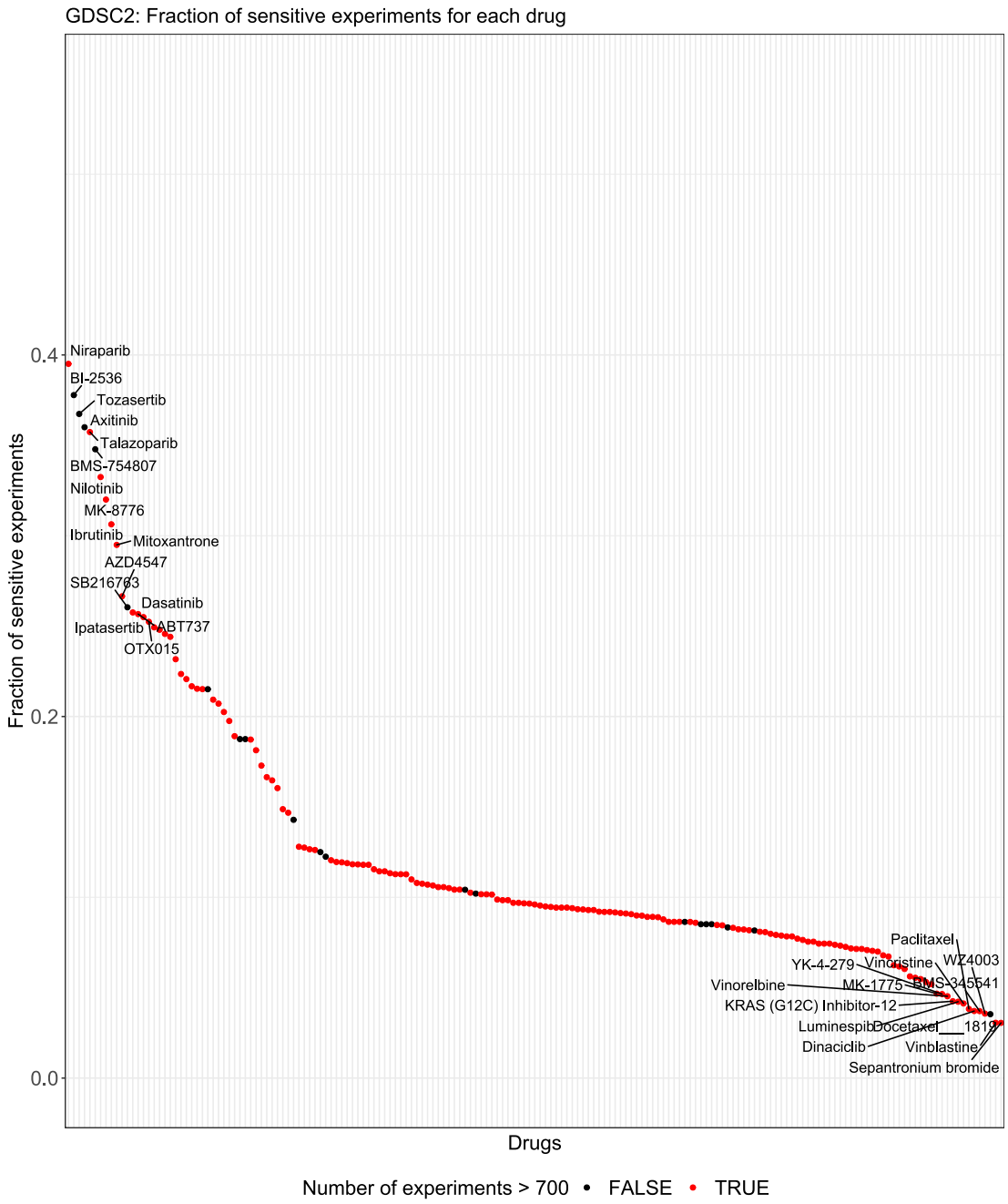


Figure 2: This figure depicts the fraction of sensitive cell lines for each drug using the data from GDSC2. Drugs that are tested on more than 700 cell lines are colored in red, otherwise the drug is colored in black.

from the GDSC database need to be binarized. This binarization was accomplished using the procedure described in Knijnenburg *et al.*, 2016 using a custom R-script. Briefly summarized, the distribution of resistant cell lines for each drug is estimated from the given IC50 values. Depending on the shape of this distribution, a drug-specific binarization threshold is determined. This threshold is then used as a dividing line between the sensitive and resistant cell lines. We used the parameter settings for the threshold identification as described in the supplementary documentation of Iorio *et al.*, 2016, which can be found on the GDSC website. The calculated thresholds for all 41 drugs are given in Table 2.
In addition to providing a possibility to divide the cell lines into sensitive and resistant ones, these thresholds can also be employed to retain some of the continuous information from the IC50 values in the binarized data. More specifically, the distance from a threshold is used as an importance measure for the cell lines in the final ILP formulation such that more weight is placed on cell lines further away from the threshold.

**Input matrices** Generally, the input matrices for the different drugs contain features for mutation, copy number variation, and gene expression data. The input matrix represents a collection of binary feature vectors. It specifies for each cell line and feature, whether the cell line is affected by a specific aberration (1) or not (0). Depending on whether there is a priori knowledge on biomarkers for drug sensitivity or resistance available, different input matrices can be constructed. We will now first describe how a matrix is constructed for drugs without available a priori knowledge (e.g. AZD8055, Omipalisib, Voxtalisib, NSC319726, and Niraparib). Afterwards, we will explain how we can calculate matrices for the other drugs (e.g. Rapamycin, Temsirolimus, Dactolisib, CX-5461, and Talazoparib). The number of available cell lines and used features for all 41 the drugs can be found in Table 3.

*Gene expression features* If there are replicate measurements for a particular cell line, we keep only the first measurement given. Then, we filter the genes of the gene expression data using the IntOGen cancer driver gene list. For each gene from the list, we define two feature vectors: one feature vector that indicates whether a gene is up-regulated in a sample (1) or not (0) and one feature vector that indicates whether a gene is down-regulated in a sample (1) or not (0). We binarized the gene expression values through calculation of a gene-wise z-score and then taking the upper and lower 5% of samples of the assumed distribution (z-score threshold +/- 1.68). Note that we used only the training data to obtain the distribution parameters for the z-score and re-use these to binarize the test data.

*Mutation features* The mutation data is given as specific point mutations of genes within the cell lines. However, using all single point mutations as features in a matrix would lead to a sparse, high dimensional matrix, which poses a problem to statistical learning methods. Therefore, we decided to filter the genes and then annotate the remaining mutations in order to define composite features that represent this annotation. In particular, we use four composite features per gene: oncogenic gain-of-function, onco-

genic loss-of-function, neutral, and status unknown.
We filter the genes from the mutation data using the IntOGen cancer driver gene list. Then we use the oncogenicity information from CIViC, oncoKB, and CGI to annotate the alterations of the filtered gene list. In particular, we annotate each alteration with one of the four states: oncogenic gain-of-function, oncogenic loss-of-function, neutral, and status unknown. Then, for each gene, all alterations with the same state are merged into one composite feature. This composite feature is 1 iff the respective feature for any of the merged alterations was 1. Otherwise, the composite feature is 0.
Since we consider frameshift and truncating mutations as more severe than missense mutations, we additionally consider these as extra composite features if they are not already included in the respective gain-of-function or loss-of-function composite feature for a gene.

*Copy number features* Similar to the gene expression and the mutation data, we first filter the genes from the copy number data using the list by Sanchez-Vega et al. (Sanchez-Vega *et al.*, 2018). Afterwards, we consider two features per remaining gene of the filtered list: CNV loss and CNV gain. The copy number data was given on transcript level and we summarize the information on gene level. This means that we neglect all genes that have conflicting transcript level copy number information, i.e. one transcript has a copy number gain while another transcript of the same gene is affected by a copy number loss.

For the drugs with a priori knowledge, we also construct one matrix as described above. In addition to that, we build a matrix with the a priori knowledge. To this end, we additionally define composite features that represent the sensitivity information on specific alterations from the CIViC, OncoKB, CGI and COSMIC databases. In particular, there exist four putative composite features: sensitive (alteration is predictive of positive drug response), not sensitive (alteration has been shown not to be predictive for a positive drug response), resistant (alteration is predictive of negative drug response), not resistant (alteration has been shown not to be predictive for a negative drug response). To obtain the composite features in general, we first annotate the mutation and copy number alterations with a sensitivity-associated state: sensitive, not sensitive, resistant, not resistant, and status unknown. Then we merge all alterations with the same state into one binary feature representing iff at least one of the alterations is present (1) or not (0). However, for the five drugs that we investigated in the main manuscript, only alterations associated with drug sensitivity are present in the GDSC cell line data (see Table 4, Table 5, Table 6, Table 7, and Table 8). Hence, there exists only one composite feature (sensitive) for these drugs. For two of the other drugs in our data set, also additional states are present (Table 10, Table 12).
If an alteration is annotated with a sensitivity- or resistance-associated state except status unknown, then this annotation is prioritized over the oncogenicity-associated annotation. As a consequence, these alterations are present in the respective composite feature of sensitivity/resistance information but not in the composite feature for the oncogenicity information. Note that if the databases state that 'oncogenic loss-of-function mutations' or 'oncogenic gain-of-function mutations' are responsible for the sensitivity or resistance

towards a drug, we take all mutations of the respective gene with a matching oncogenicity label and add them to the composite features. If the databases state that 'truncating mutations' or 'frameshift mutations' are responsible for the sensitivity or resistance towards a drug, we use the additional annotation by the GDSC data set to determine whether a mutation is truncating or induces a frameshift and also add them to the composite feature.

Table 1: This table depicts the number of available cell lines (drug response only) for each drug in GDSC1 and GDSC2.

| Drug name | GDSC1 | GDSC2 |
|---|---|---|
| Rapamycin | 359 | 745 |
| Temsirolimus | 930 | – |
| AZD8055 | 937 | 96 |
| Dactolisib | 934 | 753 |
| Omipalisib | 921 | – |
| Voxtalisib | 926 | – |
| CX-5461 | 913 | – |
| NSC319726 | 880 | – |
| Niraparib | – | 729 |
| Talazoparib | 912 | 750 |
| AS605240 | 916 | – |
| AT7867 | 926 | – |
| GSK690693 | 912 | – |
| GSK1059615 | 912 | – |
| IC-87114 | 920 | – |
| Idelalisib | 921 | – |
| OSI-027 | 907 | 265 |
| PF-4708671 | 932 | 49 |
| PIK-93 | 918 | – |
| Pilaralisib | 926 | – |
| Torin-2 | 878 | – |
| WYE-125132 | 912 | – |
| YM201636 | 919 | – |
| ZSTK474 | 918 | – |
| Apitolisib | 913 | – |
| Alpelisib | – | 805 |
| AZD8186 | – | 753 |
| MK-2206 | 903 | 772 |
| Pictilisib | 936 (883) | 767 |
| Taselisib | – | 806 |
| CZC24832 | – | 729 |
| Buparlisib | – | 753 |
| Afuresertib | – | 752 |
| Ipatasertib | – | 753 |
| GNE-317 | – | 739 |
| AMG-319 | – | 748 |
| LJI308 | – | 746 |
| AKT inhibitor VII | 920 (885) | – |
| AZD6482 | 918 (884) | 48 |
| Uprosertib | – | 745 (736) |

Table 2: For each drug, this table lists the drug-specific IC50 binarization threshold that was calculated using a procedure described by Knijnenburg et al. (Knijnenburg *et al.*, 2016)

| Drug name | Threshold |
|---|---|
| Rapamycin (GDSC2) | -4.25 |
| Temsirolimus (GDSC1) | -3.48 |
| AZD8055 (GDSC1) | -1.19 |
| Dactolisib (GDSC2) | -3.39 |
| Omipalisib (GDSC1) | -3.87 |
| Voxtalisib (GDSC1) | 2.98 |
| CX-5461 (GDSC1) | 3.57 |
| NSC319726 (GDSC1) | -2.83 |
| Niraparib (GDSC2) | 4.03 |
| Talazoparib (GDSC2) | 2.55 |
| Afuresertib (GDSC2) | 0.74 |
| Alpelisib (GDSC2) | 1.59 |
| AMG-319 (GDSC2) | 3.40 |
| Apitolisib (GDSC1) | -1.55 |
| AS605240 (GDSC1) | 1.25 |
| AT13148 (GDSC2) | 1.44 |
| AT7867 (GDSC1) | 1.32 |
| AZD8186 (GDSC2) | 1.72 |
| Buparlisib (GDSC2) | -0.04 |
| CZC24832 (GDSC2) | 3.84 |
| GNE-317 (GDSC2) | -1.03 |
| GSK1059615 (GDSC1) | -1.67 |
| GSK690693 (GDSC1) | 3.46 |
| IC-87114 (GDSC1) | 4.37 |
| Idelalisib (GDSC1) | 2.91 |
| Ipatasertib (GDSC2) | 2.58 |
| LJI308 (GDSC2) | 3.77 |
| MK-2206 (GDSC2) | 1.04 |
| OSI-027 (GDSC1) | -2.26 |
| PF-4708671 (GDSC1) | 2.66 |
| Pictilisib (GDSC2) | -0.09 |
| PIK-93 (GDSC1) | 1.13 |
| Pilaralisib (GDSC1) | 2.00 |
| Taselisib (GDSC2) | -0.09 |
| Torin-2 (GDSC1) | -3.48 |
| WYE-125132 (GDSC1) | -3.15 |
| YM201636 (GDSC1) | 0.58 |
| ZSTK474 (GDSC1) | -1.21 |
| AKT inhibitor VIII (GDSC1 screen 228) | 0.77 |
| AZD6482 (GDSC1 screen 1066) | 0.74 |
| Uprosertib (GDSC2 screen 2106) | 0.58 |

Table 3: This table depicts the dimensions of the input matrices for the different drugs (i.e. excluding the test set cell lines) with and without a priori knowledge. The corresponding database is written in brackets behind the drug name.

| Drug name | Input matrix (a priori) | Input matrix (without a priori) |
|---|---|---|
| Rapamycin (GDSC2) | Cell lines: 588, Features: 1474 | Cell lines: 588, Features: 1480 |
| Temsirolimus (GDSC1) | Cell lines: 723, Features: 1500 | Cell lines: 723, Features: 1502 |
| AZD8055 (GDSC1) | – | Cell lines: 730, Features: 1502 |
| Dactolisib (GDSC2) | Cell lines: 594, Features: 1481 | Cell lines: 594, Features: 1480 |
| Omipalisib (GDSC1) | – | Cell lines: 714, Features: 1504 |
| Voxtalisib (GDSC1) | – | Cell lines: 897, Features: 1501 |
| CX-5461 (GDSC1) | Cell lines: 708, Features: 1500 | Cell lines: 708, Features: 1501 |
| NSC319726 (GDSC1) | – | Cell lines: 684, Features: 1491 |
| Talazoparib (GDSC2) | Cell lines: 591, Features: 1479 | Cell lines: 591, Features: 1480 |
| Niraparib (GDSC2) | – | Cell lines: 579, Features: 1482 |
| Afuresertib (GDSC2) | – | Cell lines: 593, Features: 1475 |
| Alpelisib (GDSC2) | Cell lines: 636 , Features: 1481 | Cell lines: 636 , Features: 1481 |
| AMG-319 (GDSC2) | – | Cell lines: 589, Features: 1473 |
| Apitolisib (GDSC1) | Cell lines: 708, Features: 1497 | Cell lines: 708, Features: 1496 |
| AS605240 (GDSC1) | – | Cell lines: 709, Features: 1504 |
| AT13148 (GDSC2) | – | Cell lines: 589, Features: 1476 |
| AT7867 (GDSC1) | – | Cell lines: 718, Features: 1502 |
| AZD8186 (GDSC2) | Cell lines: 594, Features: 1474 | Cell lines: 594, Features: 1474 |
| Buparlisib (GDSC2) | – | Cell lines: 594, Features: 1474 |
| CZC24832 (GDSC2) | – | Cell lines: 579, Features: 1476 |
| GNE-317 (GDSC2) | – | Cell lines: 583, Features: 1471 |
| GSK1059615 (GDSC1) | – | Cell lines: 710, Features: 1498 |
| GSK690693 (GDSC1) | – | Cell lines: 711, Features: 1504 |
| IC-87114 (GDSC1) | – | Cell lines: 714, Features: 1505 |
| Idelalisib (GDSC1) | – | Cell lines: 715, Features: 1504 |
| Ipatasertib (GDSC2) | – | Cell lines: 594, Features: 1474 |
| LJI308 (GDSC2) | – | Cell lines: 589, Features: 1477 |
| MK-2206 (GDSC2) | Cell lines: 609, Features: 1480 | Cell lines: 609, Features: 1479 |
| OSI-027 (GDSC1) | – | Cell lines: 704, Features: 1505 |
| PF-4708671 (GDSC1) | – | Cell lines: 727, Features: 1500 |
| Pictilisib (GDSC2) | Cell lines: 606, Features: 1477 | Cell lines: 606, Features: 1478 |
| PIK-93 (GDSC1) | – | Cell lines: 711, Features: 1502 |
| Pilaralisib (GDSC1) | – | Cell lines: 718, Features: 1502 |
| Taselisib (GDSC2) | Cell lines: 636, Features: 1481 | Cell lines: 636, Features: 1481 |
| Torin-2 (GDSC1) | – | Cell lines: 681, Features: 1490 |
| WYE-125132 (GDSC1) | – | Cell lines: 709, Features: 1500 |
| YM201636 (GDSC1) | – | Cell lines: 712, Features: 1502 |
| ZSTK474 (GDSC1) | – | Cell lines: 711, Features: 1503 |
| AKT inhibitor VIII (GDSC1 228) | – | Cell lines: 715, Features: 1503 |
| AZD6482 (GDSC1 1066) | – | Cell lines: 716, Features: 1499 |
| Uprosertib (GDSC2 2106) | – | Cell lines: 589, Features: 1477 |

Table 4: This table depicts all biomarkers used as a priori knowledge for Rapamycin.

| Feature name | Association |
|---|---|
| PTEN truncating mutations | sensitive |
| PTEN p.R173H | sensitive |
| PTEN p.R130* | sensitive |
| PTEN p.R233* | sensitive |
| PTEN p.T167A | sensitive |
| PTEN p.Y68H | sensitive |
| PTEN p.D162H | sensitive |
| PTEN p.K128N | sensitive |
| PTEN p.C136R | sensitive |
| PTEN p.Y155C | sensitive |
| PTEN p.R130Q | sensitive |
| PTEN p.F341V | sensitive |
| PTEN p.Q298* | sensitive |
| PTEN p.R173C | sensitive |
| PTEN p.Y16* | sensitive |
| PTEN p.L42R | sensitive |
| PTEN p.W274* | sensitive |
| PTEN p.R159S | sensitive |
| PTEN p.D92H | sensitive |
| PTEN p.G36R | sensitive |
| PTEN p.R335* | sensitive |
| PTEN p.R173P | sensitive |
| PTEN p.C136Y | sensitive |
| PTEN p.K128T | sensitive |
| PTEN p.R130G | sensitive |
| PTEN p.E307* | sensitive |
| PTEN p.R47G | sensitive |
| PTEN p.Q245* | sensitive |
| PTEN CNV loss | sensitive |
| FBXW7 truncating mutations | sensitive |
| NF1 truncating mutations | sensitive |
| NF1 p.R304* | sensitive |
| NF1 p.R1204W | sensitive |
| NF1 p.Y2285* | sensitive |
| NF1 CNV loss | sensitive |
| FBXW7 loss-of-function | sensitive |
| TSC1 frameshift | sensitive |
| PIK3CA p.E542K | sensitive |
| STK11 CNV loss | sensitive |
| FBXW7 CNV loss | sensitive |
| MTOR p.C1483Y | sensitive |

Table 5: This table depicts all biomarkers used as a priori knowledge for Temsirolimus.

| Feature name | Association |
|---|---|
| PTEN R130* | sensitive |
| PTEN CNV loss | sensitive |
| PIK3CA p.H1047R | sensitive |
| ERBB2 gain-of-function | sensitive |
| ERBB2 p.S310F | sensitive |
| ERBB2 p.D277H | sensitive |
| ERBB2 p.R678Q | sensitive |
| ERBB2 p.S335C | sensitive |
| ERBB2 p.L755S | sensitive |
| ERBB2 p.T798I | sensitive |
| ERBB2 p.V842I | sensitive |
| ERBB2 p.S653C | sensitive |
| ERBB2 CNV gain | sensitive |
| PIK3CA p.E542K | sensitive |
| AKT1 p.E17K | sensitive |

Table 6: This table depicts all biomarkers used as a priori knowledge for Dactolisib.

| Feature name | Association |
|---|---|
| PIK3CA p.H1047R | sensitive |

Table 7: This table depicts all biomarkers used as a priori knowledge for Talazoparib.

| Feature name | Association |
|---|---|
| BRCA1 loss-of-function mutations | sensitive |
| BRCA1 truncating mutations | sensitive |
| BRCA2 truncating mutations | sensitive |

Table 8: This table depicts all biomarkers used as a priori knowledge for CX-5461.

| Feature name | Association |
|---|---|
| BRCA1 loss-of-function mutations | sensitive |
| BRCA1 truncating mutations | sensitive |
| BRCA2 truncating mutations | sensitive |

Table 9: This table depicts all biomarkers used as a priori knowledge for Apitolisib.

| Feature name | Association |
|---|---|
| PIK3CA p.E545K | sensitive |
| PIK3CA p.E542K | sensitive |

Table 10: This table depicts all biomarkers used as a priori knowledge for Alpelisib.

| Feature name | Association |
|---|---|
| PIK3CA p.H1047R | sensitive |
| PIK3CA CNV gain | sensitive |
| PTEN CNV loss | resistant |

Table 11: This table depicts all biomarkers used as a priori knowledge for AZD8186.

| Feature name | Association |
|---|---|
| PTEN CNV loss | sensitive |

Table 12: This table depicts all biomarkers used as a priori knowledge for MK-2206.

| Feature name | Association |
|---|---|
| KRAS p.G12D | sensitive |
| PIK3CA p.E545K | sensitive |
| PTEN CNV loss | sensitive |
| AKT1 p.E17K | not sensitive |

Table 13: This table depicts all biomarkers used as a priori knowledge for Pictilisib.

| Feature name | Association |
|---|---|
| BRAF p.V600E | sensitive |
| PIK3CA p.E545K | sensitive |
| ERBB2 CNV gain | sensitive |
| PIK3CA CNV gain | sensitive |

Table 14: This table depicts all biomarkers used as a priori knowledge for Taselisib.

| Feature name | Association |
|---|---|
| PIK3CA p.H1047R | sensitive |
| PIK3CA CNV gain | sensitive |

## References


Iorio, F., Knijnenburg, T., Vis, D. et al. (2016) A landscape of pharmacogenomic interactions in cancer. *Cell*, **166**, 740-754. https://doi.org/10.1016/j.cell.2016.06.017

Tate, J. G., Bamford, S., Jubb, H. C. et al. (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*, **47**, D941-D947. https://doi.org/10.1093/nar/gky1015

Griffith, M., Spies, N., Krysiak, K. et al. (2017) CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet*, **49**, 170–174. https://doi.org/10.1038/ng.3774

Chakravarty, D., Gao, J., Phillips, S. et al. (2017) OncoKB: a precision oncology knowledge base. *JCO Precis Onco*, **1**, 1-16. https://doi.org/10.1200/PO.17.00011

Tamborero, D., Rubio-Perez, C., Deu-Pons, J. et al. (2018) Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med*, **10 (25)**. https://doi.org/10.1186/s13073-018-0531-8

Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J. et al. (2013) IntOGen-mutations identifies cancer drivers across tumor types. *Nat methods*, **1**0, 1081-1082. https://doi.org/10.1038/nmeth.2642

Sanchez-Vega, F., Mina, M., Armenia, J. et al. (2018) Oncogenic signaling pathways in the cancer genome atlas. *Cell*, **173**, 321-337. https://doi.org/10.1016/j.cell.2018.03.035.

Knijnenburg, T., Klau, G., Iorio, F. et al. (2016) Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy. *Sci Rep*, **6**, 1-14. https://doi.org/10.1038/srep36812