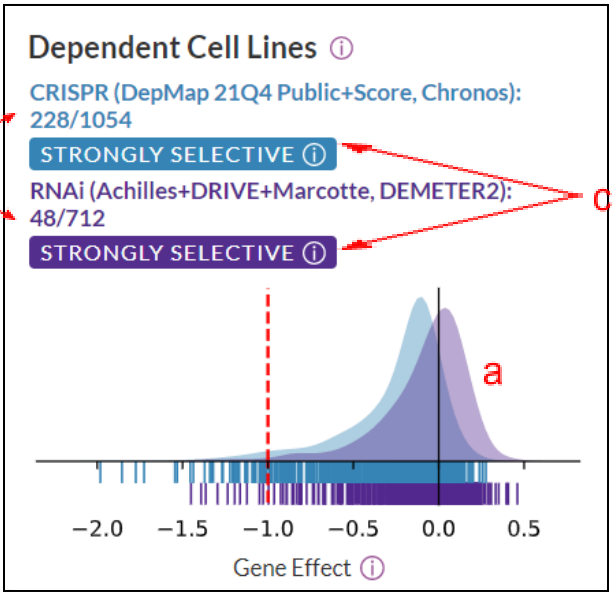
How to Identify Genetic Dependencies Using the DepMap Portal

Genetic Dependencies and Gene Essentiality

A genetic dependency is a gene that is considered necessary for the proliferation and survival of a given cell population. DepMap uses large-scale RNAi and CRISPR screens to identify such dependencies by measuring how essential each gene is for the viability of hundreds of diverse cancer cell line models. In this guide we will show you how to use the DepMap portal to identify and interpret different patterns of gene essentiality, which are shown in the Dependent Cell Lines tile on the Gene Overview summary page for each gene.

The Dependent Cell Lines Tile

The Dependent Cell Lines tile on the gene overview pages provides information on the distribution of gene essentiality across cancer cell lines using both CRISPR and RNAi datasets. There are three components shown in the tile for both the RNAi and CRISPR data (when available), which are: a) the distribution of gene effect scores, b) the fraction of cell lines classified as 'dependent' on the gene, and c) a category assigned to each dependency profile when/as appropriate. Each of these components are described further below.



Dependent Cell Lines

CRISPR (DepMap 21Q4 Public+Score, Chronos): 228/1054

STRONGLY SELECTIVE

RNAi (Achilles+DRIVE+Marcotte, DEMETER2): 48/712

STRONGLY SELECTIVE

Gene Effect Scores

The Dependent Cell Lines tile shows a distribution of the gene effect scores for cell lines from our available CRISPR data (shown in blue), and RNAi data (shown in purple). Each purple and blue dashed line under the distribution curves in the Dependent Cell Lines tile represents a cell line from the corresponding dataset.

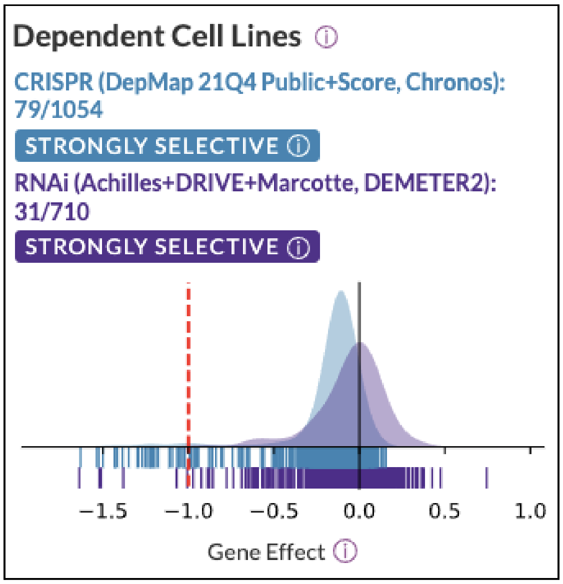
Gene effect scores measure the effect of knocking out or inhibiting the expression of the gene on the viability of a given cell line. These values are calculated using the DEMETER2 (RNAi) or Chronos (CRISPR) algorithms. The data are normalized using annotated sets of common-essential and non-essential genes, such that a score of 0 represents no viability effect, and a score of -1 corresponds to the median effect of known common-essential genes (see here for more information on the normalization method used). Thus, more negative scores represent cell lines that are more dependent on the gene for proliferation.

Categorizing Gene Effect Distributions

There are two classes we use to categorize gene effect distributions that you will see indicated on the Dependent Cell Lines tile for the CRISPR and RNAi data. These two classes are: “strongly selective”, and “common-essential”. We describe these two classes below, as well as what “non-essential” distributions look like, and implications for differences between RNAi and CRISPR distributions.

Strongly Selective Distribution

Distributions like the one shown for PAX8 below are an example of cell lines exhibiting strongly selective dependency on a gene. Here, we see many cell lines for which PAX8 perturbation has little effect on survival (those centered around a gene effect score of 0), as well as a number of cell lines that are strongly dependent on PAX8, with negative scores near or below -1 (representing the median effect for known common-essential genes).


Dependent Cell Lines

CRISPR (DepMap 21Q4 Public+Score, Chronos): 79/1054

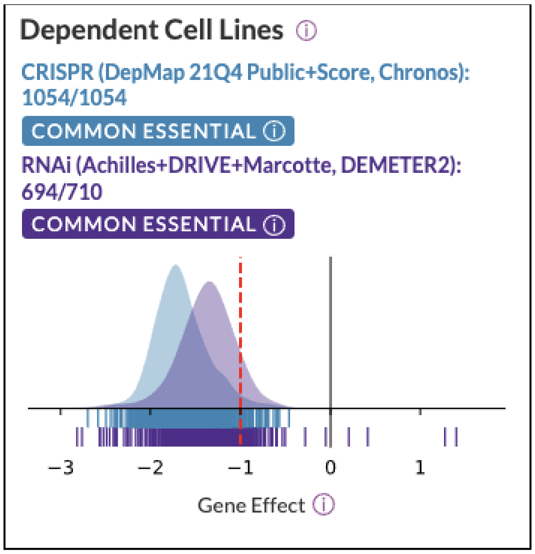
RNAi (Achilles+DRIVE+Marcotte, DEMETER2): 31/710

Defining Strongly Selective Dependencies

A gene that is considered strongly selective is defined as having a dependency distribution that is at least 100 times more likely to have been sampled from a skewed *t*-distribution than a normal distribution (i.e. skewed-LRT value > 100). See here for more information about this metric of selectivity.

Common-Essential Distribution

SF3B1 is an example of a common-essential gene, also referred to as a pan-essential gene. A gene is considered pan-essential if proper functioning of that gene is necessary for the fitness and survival of nearly all cell lines. For pan-essential genes, the gene effect score distribution will be significantly shifted towards and beyond -1.



Dependent Cell Lines ⓘ

CRISPR (DepMap 21Q4 Public+Score, Chronos): 1054/1054

COMMON ESSENTIAL ⓘ

RNAi (Achilles+DRIVE+Marcotte, DEMETER2): 694/710

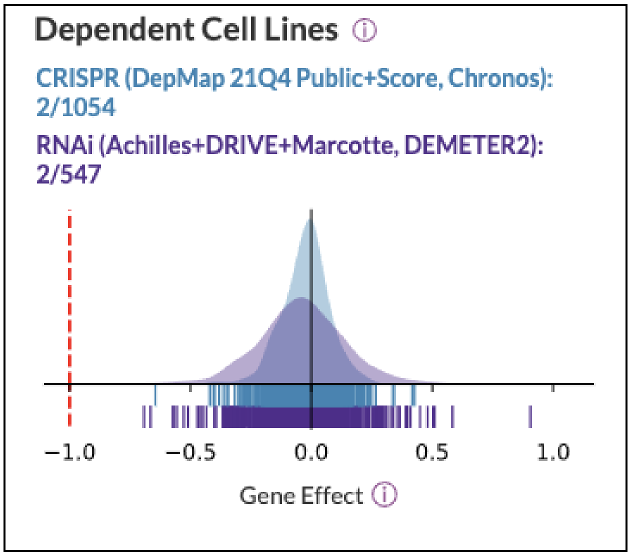
COMMON ESSENTIAL ⓘ

Defining Common-Essential Genes

A common-essential gene is defined as a gene which, in a large, pan-cancer screen, ranks in the top X most depleting genes in at least 90% of cell lines. X is chosen empirically using the minimum of the distribution of gene ranks in the 90th percentile least depleting lines. See here for more information about the methods for classifying common-essential genes.

Non-Essential Distributions

While “non-essential” is not used as an explicit categorization in the Dependent Cell Lines tile, it is important to be able to recognize non-essential gene effect distributions as well. A1BG is an example of a non-essential gene (shown below). The gene effect distribution for a non-essential gene will be centered around a score of 0, suggesting that the gene has little to no effect on the survival of any of the cell lines in DepMap.



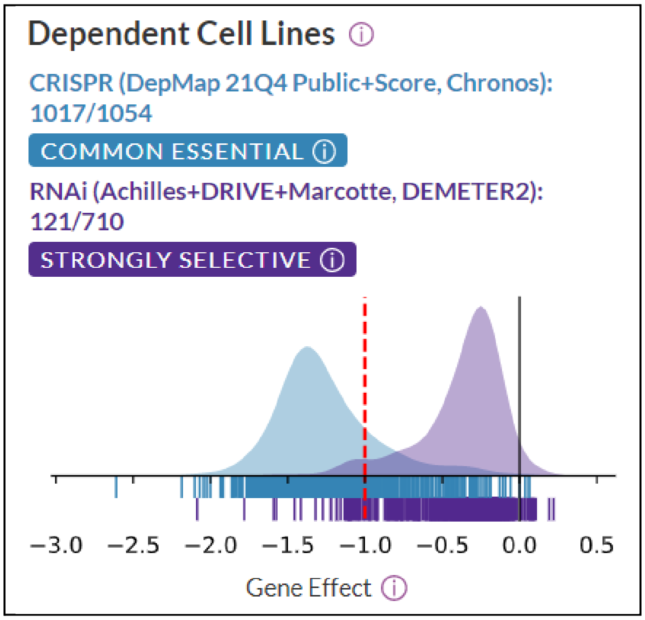
Dependent Cell Lines

CRISPR (DepMap 21Q4 Public+Score, Chronos): 2/1054

RNAi (Achilles+DRIVE+Marcotte, DEMETER2): 2/547

Differences Between RNAi and CRISPR Distributions

When reviewing information in the Dependent Cell Lines tile for a given gene, you'll inevitably come across some genes where RNAi and CRISPR show dissimilar gene effect distributions. We see an example of this when looking at the profile for the gene PRMT5, shown below.



Dependent Cell Lines

CRISPR (DepMap 21Q4 Public+Score, Chronos): 1017/1054

COMMON ESSENTIAL

RNAi (Achilles+DRIVE+Marcotte, DEMETER2): 121/710

STRONGLY SELECTIVE

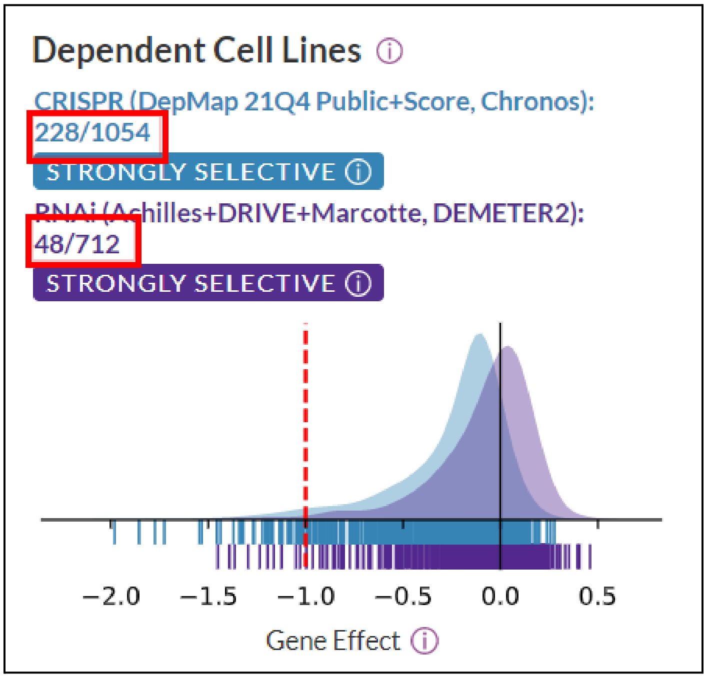
Based on this profile, PRMT5 is considered common-essential according to our CRISPR data, but is strongly selective according to our RNAi data. This pattern is not uncommon, and can highlight biologically relevant differences between the two datasets.

With these CRISPR screens, we're generally looking at the effects of complete knockout of the gene at the DNA-level, versus partial suppression of the gene expression at the mRNA level with RNAi. Hence, differences between these gene perturbation approaches can reflect response to total versus partial suppression of gene activity.

Quantifying the Fraction of Dependent Cell Lines

In addition to estimating the effect size of gene perturbation on each cell line's viability ('gene effect' scores), we also use a method to categorize each cell line as dependent or not on a given gene. The fraction of cell lines called dependent on the gene is indicated as shown for RNAi and CRISPR datasets. These binary calls are based on a method for estimating the 'probability of dependency' for each gene in each cell line, which compares a given gene effect score to those measured for known non-essential and common-essential genes. This approach is similar to [BAGEL](), and is described in detail [here]().

A cell line is considered dependent on a gene if it has a probability of dependency greater than 0.5. The fractions next to each dataset description (outlined in red below) indicate how many cell lines in that dataset are dependent on the gene of interest (i.e. how many cell lines consider the gene to be essential). For example, here we see that 228 of the 1054 cell lines in our CRISPR dataset consider EGFR to be an essential gene.



Dependent Cell Lines ⓘ

CRISPR (DepMap 21Q4 Public+Score, Chronos): 228/1054

STRONGLY SELECTIVE ⓘ

RNAi (Achilles+DRIVE+Marcotte, DEMETER2): 48/712

STRONGLY SELECTIVE ⓘ