1. **Introduction**

This SWEAVE document provides R objects and code for the computational analysis included in the manuscript “pathCHEMO: Uncovering (epi) genomic pathways of chemoresistance in lung adenocarcinoma”. Sections X, and X of this document provide information on data objects (i.e., tables) referenced in the manuscript and sections X,X provide executable R code along with necessary R objects. All R data objects, necessary to execute this document, can be downloaded from <https://doi.org/10.6084/m9.figshare.8041274>

1. **Transcriptomic and epigenomic pathway integration**
   1. **Pathway Signature**

For integrating pathways, you will need GSEA function can be collected from <https://github.com/hejing3283/scripts/blob/master/labScipts/celine_gsea_functions_withPlot_modified.R>

To run GSEA, you will need gene expression and DNA methylation dataset and pathways from MSigDB database, which are provided below.

Pathway analyses of epigenomic Lung Adenocarcinoma Profile correspond to the signature in Table S2B. Epigenomic signatures are individually subjected to pathway enrichment analysis in REACTOME [1] KEGG [2] and BIOCARTA [3] databases

To create *composite expression pathway signature* and *composite methylation pathway* signature, we have used gene expression and DNA methylation datasets from The Cancer Genome Atlas (TCGA-LUAD) project (43) with detailed description of the data available at Genomics Data Commons database (GDC; https://portal.gdc.cancer.gov/)

To run integrative analysis, you will need *composite expression pathway signature* and *composite methylation pathway signature,* which are available as R objects.

{{pathway signature example R code}}

For our validation analysis, we have used gene expression dataset from Tang et al (7) with detailed description of the data available at GSE42127 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42127>

List of pathways is available at C2 database and can be downloaded from MSigDB (<http://software.broadinstitute.org/gsea/msigdb>).

1. Integrative analysis identified (epi) genomic pathways implicated in resistance

Since our goal has been to identify pathways that are significantly affected on both transcriptomic and epigenomic levels, we have employed GSEA to compare composite expression pathway signature and composite methylation pathway signature to identify pathways that are significantly affected on both transcriptomic and epigenomic levels (i.e., belong to the leading edge from the GSEA analysis).

To assure that we can identify pathways which are (i) over-expressed and under-methylated; (ii) under-expressed and over-methylated; (iii) differentially expressed and differentially methylated; etc., each pathway signature was ranked based on the absolute values of their NESs and used for subsequent GSEA comparative analysis.The data object containing signature can be loaded as {{Signature R name}}. First column corresponds to 833 pathway names, second column corresponds to p-value of the pathways, third column corresponds to normalized enrichment score of the pathways (NES), fourth column corresponds to number of genes in the pathway, fifth column corresponds to number of genes present in leading edge and sixth column corresponds to leading edge gene symbols.

So the first step in our analysis pipeline is to load the GSEA package, and the composite expression pathway signature and composite methylation pathway signature.

{{R code}}

1. Datasets for clinical validation (R code and objects)

Gene expression datasets used for clinical validation of Carboplatin-Paclitaxel treatment response include: (i) Tang et al [6]

We have performed Kaplan-Meier Survival Analysis and c-statistics.

The distribution of the correlation coefficients for the negative controls (comparison between randomized versions of the networks), positive controls (comparison between the alternative versions of the same network), and mouse vs. human interactomes can be represented in a graph by:

1. Kaplan-Meier survival analysis (R code and objects)
   1. Tang et al dataset

The Tang cohort data object can be loaded using

Where columns corresponds to samples (i.e., patient) names and rows corresponds to gene names.

First, you need to load the survival data objects (which describes time to prostate cancer specific death)

Libraries necessary for Survival analysis can be loaded as:

Finally, Kaplan-Meier survival analysis depicting all four groups/curves with respect to time to Prostate

cancer-related speci\_c survival (time to death) can be generated as