|  |  |  |  |
| --- | --- | --- | --- |
| Step | Code name | Requirements and Outputs | Functions |
| Data Extraction And Processing | DataProc\_Myeloma.R | **Inputs:**  1. *bortGeo.RData*: pre-downloaded data-sets from the folder, “/Data\_And\_Codes/PreprocessedData/bortezomibData”, which include gene expression data and patients’ response phenotype from GEO data-base.  2. *sensitivity\_data\_for\_drug\_104.csv:* pre-downloaded drug sensitivity data from GDSC website. The source details are inside the code.  3. uses auxiliary subroutines from the folder, “/Data\_And\_Codes/Scripts”, that perform various internal functions, They are elaborated later.  **Outputs:**  1. *Input\_Myeloma.mat*: contains D, C, P, Indresp. Please see the inputs of iRx\_main.m for more details. A copy is stored in “/MetaData/”.  2. DataForFittingMyeloma.Rdata: contains all the data as in Input\_Myeloma.mat + extra details on cleaning exercise, such as transformation details, gene names etc. A copy is stored in “/MetaData/”.  **Others:**  Built on R 3.5.1., Requires R packages with versions as specified within the code to reproduce the output | The function extracts and prepares the required patients gene expression data and response phenotype from GEO website. It uses cell lines’expression data pre-processed from GDSC website. Programs for preprocessing are provided by P. Geeleher given in the folder, “Data\_And\_Codes/PreprocessedData/Codes\_FOR\_PreProc\_P\_Geeleher”.  The program performs mapping probset ids to proper names, subsetting to only CEL files which map uniquely to a cell line, removing low varying genes, summarizing duplicates of gene ids by their means, homogenizing data from different probes of cell lines and patients’ expression, transforming the drug sensitivity values to remove skewness and generates pvalues for significance of association between the drug sensitivity and individual gene expression. |
| Data Extraction And Processing | DataProc\_Breastcancer.R | **Inputs:**  1. *doce\_rma\_syms\_brainArray.RData*: pre-downloaded data-sets from the folder, “/Data\_And\_Codes/PreprocessedData/ docetaxelData/”, which include gene expression data and patients’ response phenotype from GEO data-base, preprocessed using rma function and brainarray mappings, gene symbol mapping.  2. *sensitivity\_data\_for\_drug\_1007.csv:* pre-downloaded drug sensitivity data from GDSC website. The source details are inside the code.  3. uses auxiliary subroutines from the folder, “/Data\_And\_Codes/Scripts”, that perform various internal functions, They are elaborated later.  **Outputs:**  1. *DataForFittinglBreastcancer.Rdata*:  Similar to that of DataProc\_Myeloma.R.  2. *Input\_Breastcancer.mat*: Similar to that of DataProc\_Myeloma.R. | The function uses cell lines’expression and patients’ gene expression data partially pre-processed after being obtained from GDSC and GEO websites respectively. Programs for preprocessing are provided by P. Geeleher given in the folder, “Data\_And\_Codes/PreprocessedData/Codes\_FOR\_PreProc\_P\_Geeleher”.  The program performs mapping probset ids to proper names, Subsetting to only CEL files which map uniquely to a cell line, removing low varying genes, summarizing duplicates of gene ids by their means, homogenizing data from different probes of cell lines and patients’ expression, and transforming the drug sensitivity values to remove skewness. |
| iRx Model Implementation | iRx\_main.m : the main underlying workhorse function | **Inputs:**   1. Design parameters for MCMC: a) nrun : number of MCMC iterations, b)burn: number of samples for burn-in, c)thin: size of thinning, d) seed: random number initialization. 2. Model parameters: a) D: Drug response on cell lines, b) C: gene expression matrix of cell lines, c) P: gene expression matrix of patients. d) IndResp: clinical response to patients (binary, 1 meaning responder while 0 meaning non-responder)   **Outputs:**  iRx,Ni : iRx and NI scores; iRx\_std,Ni\_std : standardized iRx and NI scores; iRx\_P,NI\_P: p-values of t-test significances of iRx and NI scores between classes of responder and non-responder; AUC\_irx, AUC\_NI: AUC’s for both methods; EP\_irx: enrichment probability score; RP,NRP: proportion of true responders and non-responders in the corresponding predicted classes; idx: cluster assignment matrix; L\_mh: standardized factor loading matrix; Lambda\_MCMC: array storing MCMC samples of loading matrix; tau\_MCMC: vector storing MCMC samples of precision parameter of D-C model ; lambda\_MCMC: vector storing MCMC samples of shrinkage parameter; beta\_MCMC: matrix storing MCMC samples of 10 randomly pre-selected beta coefficients; nofout\_MCMC: vector of MCMC samples storing K after burn-in; time: total time taken for execution. | The function is internally called by the wrapper code: “iRx\_main\_real\_implementation”. The function implements adaptive gibbs sampler for C-P model and collapsed Gibbs sampler for D-C by calling two subroutines, “iRx\_spfact.m “ and “iRx\_Bayes\_Ridge.m” respectively. |
| iRx Model Implementation | iRx\_main\_real\_implementation.m | **Inputs:**  Specifies the inputs for iRx main function, once for Myeloma and then for Breast cancer study  **Outputs:** MatOut\_Myeloma.mat (includes all outputs from iRx\_main for Myeloma); MatOut.Breastcancer.mat (includes all outputs from iRx\_main for Breastcancer) | The function calls iRx\_main.m with a given set of MCMC parameters, such as number of iterations, burin-in size etc |
| Implementation of PC\_ind, PC\_joint  +  Visualizing Outputs | Myeloma.R | **Inputs:**  MatOut\_Myeloma.mat, DataForFittingMyeloma.Rdata  **Outputs:**  Fig 4(a): Violin plot of standardized scores (iRx, NI, PC\_ind, and PC\_joint ), Fig 4(c): Density plot of iRx and ½ sigma classifier; Fig 5(a) and 5(b): barplot and chord diagram, and Fig6(a): circular heatmap for genes for myeloma | The function takes inputs as matlab files, which contain cluster details, standardized loading matrix and calibrated values, and produces various plots (see the earlier column) for myeloma study. It also computes the AUC’s of all methods. |
| Implementation of PC\_ind, PC\_joint  +  Visualizing Outputs | Breastcancer.R | **Inputs:**  MatOut\_Breastcancer.mat, DataForFittingBreastcancer.Rdata  **Outputs:**  Fig 4(b): Violin plot of standardized scores (iRx, NI, PC\_ind, and PC\_joint), Fig 4(d): Density plot of iRx and ½ sigma classifier; Fig S5(a) and S5(b) in the supplementary materials: barplot and chord diagram, and Fig6(b): circular heatmap for genes for breastcancer | The function does the same as that of Myeloma.R for Breastcancer study |