## Description of methods used for the benchmarking experiments

For the purposes of this study, only component-based methods that integrated multiple datasets and perform variable selection were considered. Since tuning the number of variables to retain in each model would result in biomarker panels with different numbers of variables, for the purposes of this study all variables were retained in each model. The features were instead ranked based on their absolute value of their loadings (importance) and 60 variables were selected from each omic type, resulting in multi-omic biomarker panels with 180 variables (60 mRNAs, 60 miRNAs and 60 CpGs). Equal numbers of variables allowed for a fair comparison in the gene set enrichment analysis.

|  |  |
| --- | --- |
|  | **Parameter settings** |
| **Supervised** | |
| DIABLO\_null | ncomp = 2 (# of components)  keepX = all variables were retained from each omics dataset    default parameters were used for the other arguments:  scheme="horst",  mode="regression",  scale = TRUE,  init = "svd",  tol = 1e-06,  max.iter = 100 |
| DIABLO\_full | ncomp = 2 (# of components)  keepX = all variables were retained from each omics dataset    default parameters were used for the other arguments:  scheme="horst",  mode="regression",  scale = TRUE,  init = "svd",  tol = 1e-06,  max.iter = 100 |
| Concatenation-sPLSDA | ncomp = 2 (# of components)  keepX = all variables were retained from each omics dataset  default parameters were used for the other arguments:  mode = "regression"  scale = TRUE,  tol = 1e-06,  max.iter = 100 |
| Ensemble\_sPLSDA | ncomp = 2 (# of components)  keepX = all variables were retained from each omics dataset  default parameters were used for the other arguments:  mode = "regression"  scale = TRUE,  tol = 1e-06,  max.iter = 100 |
| **Unsupervised** | |
| sGCCA[1] | ncomp = 2 (# of components)  keepX = all variables were retained from each omics dataset    default parameters were used for the other arguments:  scheme = "horst",  mode="canonical",  scale = TRUE,  init = "svd.single",  tol = .Machine$double.eps,  max.iter=1000, |
| JIVE\*[2] | default parameter settings from the jive() from the r.jive R-package were used:   1. scale = TRUE, center = TRUE 2. method = “perm”   sPCA parameters:  ncomp = 2 (# of components)  keepX = rep(ncol(X),ncomp)(all variables were retained from each omics dataset  default parameters were used for the other arguments:  center = TRUE  scale = TRUE,  max.iter = 500,  tol = 1e-06 |
| MOFA[3] | factors=2 (# of components)  default parameter settings recommended by MOFA were used:   1. likelihoods=( gaussian gaussian gaussian ) 2. Convergence criterion (tolerance=0.01, nostop=0) 3. Training components (startDrop=1 # initial iteration to start shutting down factors, freqDrop=1 # frequency of checking for shutting down factors, dropR2=0.00 # threshold on fraction of variance explained) 4. hyperparameters for the feature-wise spike-and-slab sparsity prior [learnTheta=( 1 1 1 ) # 1 means that sparsity is active whereas 0 means the sparsity is inactivated; each element of the vector corresponds to a view, initTheta=( 1 1 1 ) # initial value of sparsity levels (1 corresponds to a dense model, 0.5 corresponds to factors ); each element of the vector corresponds to a view, startSparsity=250 # initial iteration to activate the spike and slab, we recommend this to be significantly larger than 1]   Intercept was set to TRUE (learnIntercept=1) |

\*since the variable selection functionality has not been added to JIVE R-function, sparse Principal Component Analysis (sPCA) from the mixOmics R-package was applied to the joint variation matrix obtained after applied JIVE to the multi-omics cancer datasets.

**References**

1. Tenenhaus A, Philippe C, Guillemot V, Le Cao K-A, Grill J, Frouin V. Variable selection for generalized canonical correlation analysis. Biostatistics. 2014;15: 569–583. doi:10.1093/biostatistics/kxu001

2. Lock EF, Hoadley KA, Marron JS, Nobel AB. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. Ann Appl Stat. 2013;7: 523–542. doi:10.1214/12-AOAS597

3. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-Omics factor analysis disentangles heterogeneity in blood cancer. bioRxiv. 2017; 217554.