Multiomic data anlysis

Juan R Gonzalez

BRGE - Bioinformatics Research Group in Epidemiology
Barcelona Insitute for Global Health (ISGlobal)
http:// brge. isglobal. org

TASK 1: Consider the SNPs found associated with type 1 diabetes in the paper *Reddy et al. Association between type 1 diabetes and GWAS SNPs in the southeast US Caucasian population. Genes and Immunity, 2001;12:208–212* that are available in this table `https://www.nature.com/articles/gene201070/tables/1`.

1. Have any of those SNPs been associated with any other traits?

2. Is any of those SNPs an eQTL in adiposity tissue (GTEX)?

   HINT: investigate whether the selected SNPs have been annotated in GWAScatalog or have a significant p-value of any gene in GTEX using `haploR` package.

TASK 2: File `nci60.Rdata` contains miRNA, mRNA and protein data of melanoma, leukemia and CNS disease. Data are encapsulated in a list (object `nci60`) where each components stands for a different omic data (NOTE: features are in rows and samples in columns). Data corresponds to cells lines from the NCI-60 panel available at TCGA project. 21 cell lines are providing information about 537 miRNAs, 12,895 gene expression and 7,016 proteins. We are interested in obtaining omic pro
les to characterize those diseases. NOTE: The object `cancer` is a vector containing a factor variable indicating the type of cancer of each sample.

1. Load data into R.

2. Perform a separate PCA of each omic data and give the top-10 features more associated with each tumor.

3. Perform multi coinertia analysis and provide the top-10 features (they can be proteins, miRNA or genes) associated with each tumor.

4. Perform penalized canonical correlation analysis and provide the signi

   cant features associated with each tumor

5. Compare the results