

Exercises 2: Containers; exploring data and metadata

VJ Carey

March 1, 2008

1. ExpressionSet: create a demonstration ExpressionSet instance using simulated data. For example, `x = matrix(rnorm(80),20,4)` can serve as the assay data component for a twenty gene array set. Try to make a meaningful object, choosing featureNames from an annotated platform.
2. racExSet: use GGdata/GGtools to search for eQTL for TBX21 expression on chromosome 17. Use the default approach on chr17GGceuRMA, and then use imputeSNPFixed to generate a racExSet where all missing genotypes are set to homozygous common. use plot_mlp to visualize the association pattern, and top-Snps to find the candidate eQTL. does the imputation make a difference?
3. graph. Obtain a printout of a KEGG pathway from the instructor. Isolate a component with 5-6 nodes and create a graphNEL representing the network. plot it.
4. gene sets. Create a gene set corresponding to the pathway fragment defined in the previous problem. Find the Entrez, hgu133aplus2, and illuminaHumanv2 identifiers for your genes
5. hg18track. Use rawCD4 in encoDnaseI package. Find regions of chr 20 where dnaseI hypersensitivity in CD4 >3, and regions where dnaseI hyp is always < 0.5 Use ph2supp in hm2ceu to see if phase 2 hapmap snps are more common in DNaseI hypersensitive regions on chr20.
6. Biostrings. Use readFASTA to read in the S. pombe chromosome 2 region sequence. Compare the densities of all start positions of pentamers AAAAA and CCCCC in this region.