ComBat Documentation

Description:	Performs batch correction on a dataset containing multiple batches.
Author:	Marc-Danie Nazaire (Broad Institute), gp-help@broadinstitute.org
Version	2.0

Summary

This module runs the ComBat (Combining Batches) R script on a microarray dataset. The ComBat R script uses an Empirical Bayes method to adjust for potential batch effects in the dataset.

A ComBat discussion/help forum exists on GoogleGroups: https://groups.google.com/forum/?fromgroups#!forum/combat-user-forum

References

Johnson WE, Rabinovic A, Li C. Adjusting batch effects in microarray expression data using Empirical Bayes methods. *Biostatistics.* 2007;8(1):118-127.

Parameters

Name	Description
input file	Input file, in RES or GCT format.
sample info file	A file containing information about the samples. The first three columns in the sample info file must be *Array*, *Sample*, and *Batch*. Columns 4 and greater can contain any covariates to include in the analysis. (.txt file)
covariate columns	The subset of covariate columns to use in analysis. This is either set to *all*, *none*, or a list specifying one or more covariate columns from sample info file to use in the analysis (i.e., 4, 5, 7, …). If this is set to *all*, then all the covariate columns are used. If this is set to *none*, then none of the covariate columns are used.
absent calls filter	The filter to apply to genes with absent calls. Filters genes with absent calls in 1 - (absent calls filter) of the samples. This value must be between 0 and 1. Ignored if input file is not a RES file.

create prior plots	Whether to generate prior probability distribution plots. This is ignored if non-parametric method is used.
prior method	Empirical Bayes priors distribution estimation method to use. Either a parametric or non-parametric method can be applied. If unsure of which method to use try the parametric method first since it is faster. If the red and black lines in the plot do not match up well then the non-parametric method should be used instead. NOTE: The non-parametric method may take hours to complete but varies depending on the size of the data.
output file	Name of batch adjusted output file.

Output Files

1. <output file>.<res, gct>

   A file containing the batch effect adjusted dataset in the same file format as the input file.

2. <output file>.plot.<pdf, jpeg>

   A file containing the prior plots in JPEG format if this format is supported, otherwise in PDF format.

Platform Dependencies

Module type:	Preprocess & Utility
CPU type:	any
OS:	any
Language:	R (v. 2.5.0)

GenePattern Module Version Notes

Version	Description
1.0	Released 8/18/2008
2.0	Updated to make it available for any OS.