

Peptide Data Template

June 17, 2021

Upload Data

Expression Data

An **unlabelled** peptide expression data file called “**pep_edata.csv**” was uploaded to pmart. The column that designates unique molecules was marked as “**Mass_Tag_ID**”. The original scale of the data was abundance and was changed to log2. The value to denote missing data was “**NA**” and the expression data **was not already** normalized.

Biomolecule Information

An associated biomolecule information file was also uploaded called pep_emeta.csv and the protein identifier column was designated as Protein.

Sample Information

An associated sample information file was also uploaded called “**pep_fdata.csv**”. Trimmed sample names **were not** used. The column in the sample information file which indicates sample names was designated as “**SampleID**”.

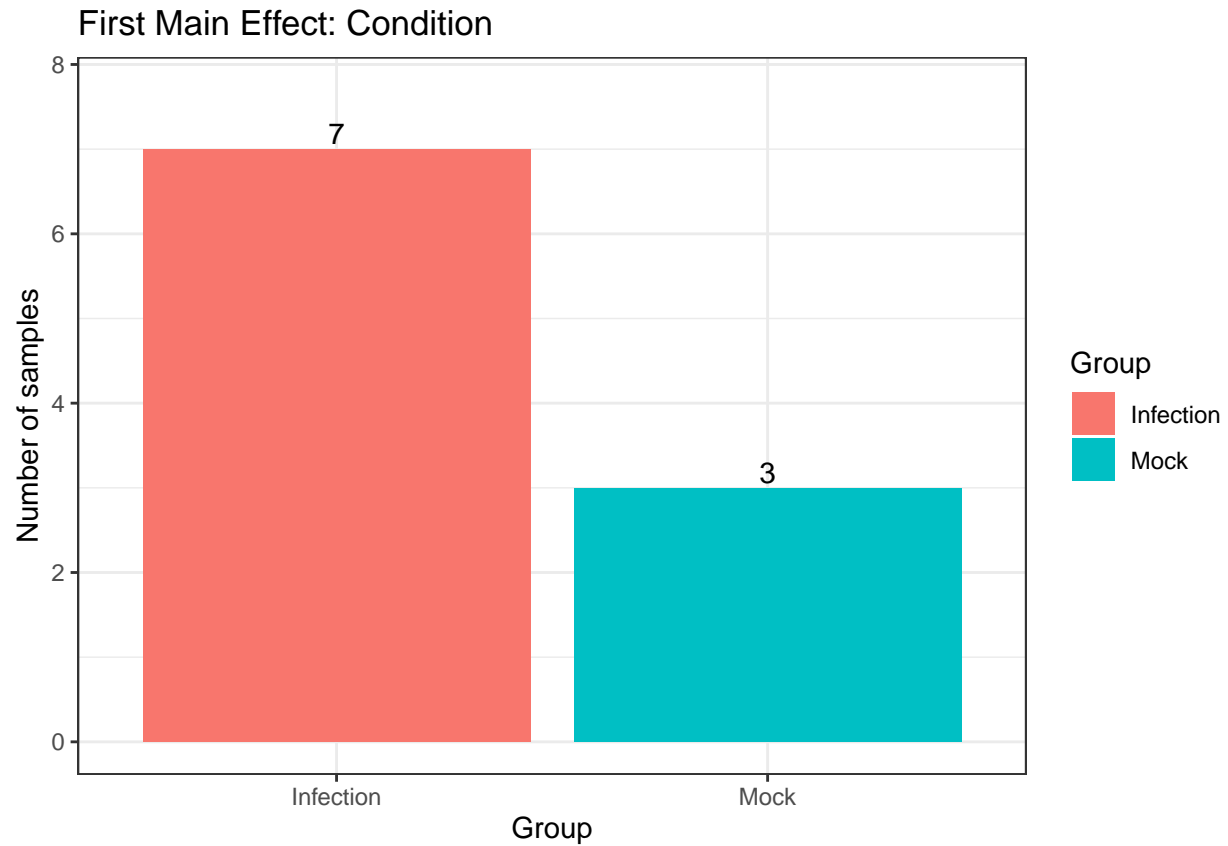
Group Samples

Grouping Information

This table summarizes all the user specified main effects or covariates in pmart. The “Selected Column Name” denotes the name of the column in the sample information file assigned as a main effect or covariate.

Main Effect or Covariate	Selected Column Name
First Main Effect	Condition
Second Main Effect	None selected
First Covariate	None selected
Second Covariate	None selected

Number of Samples per Main Effects



Data Summary

Summary Table

The first column in the table below denotes a property of the peptide data, and the “Data” column states that property’s value.

	Data
Class	pepData
Unique SampleIDs (f_data)	10
Unique Mass_Tag_IDs (e_data)	13022
Unique Proteins (e_meta)	NA
Missing Observations	35317
Proportion Missing	0.271
Samples per group: Infection	7
Samples per group: Mock	3

Missing Value Table

In the table below, the first column denotes the sample and the second is the missing number of observations. The third column represents the second as a percentage of the total number of observations for that sample.

	Missing Observations	Proportion Missing
Infection2	2711	0.208

	Missing Observations	Proportion Missing
Infection3	3447	0.265
Infection4	3978	0.305
Infection6	3535	0.271
Infection7	2896	0.222
Infection8	6014	0.462
Infection9	4953	0.380
Mock1	2682	0.206
Mock2	2551	0.196
Mock3	2550	0.196

Filter

Filters **were** applied. A total of 6 filters were applied. See the table below for a descriptions of the filters and the order.

Summary of Applied Filters

Order	Filter	Type	Parameters	Summary
1	Molecule Filter	Biomolecule	Min Number Molecules: 2	1814 biomolecule(s) were filtered.
2	Proteomics Filter	Biomolecule	Min Number of Peptides: 2 & Degenerate Peptides Removed: Yes	708 biomolecule(s) were filtered and 952 protein(s) were filtered.
3	CV Filter	Biomolecule	Max CV: 150	2 biomolecule(s) were filtered.
4	imd-ANOVA Filter	Biomolecule	Min ANOVA: 2 & Min G-Test: 3	3866 biomolecule(s) were filtered.
5	rMD Filter	Sample	P-Value Threshold: 0.001 & Metrics Used: MAD, Kurtosis, Skewness, Correlation	1 sample(s) were filtered.
6	Custom Filter	Sample or Biomolecule		1 sample(s) were filtered. 0 biomolecule(s) were filtered. 0 protein(s) were filtered.

Molecule Filter

A molecule filter was applied to the data, which removes biomolecule(s) (Mass_Tag_IDs) not having at least the minimum number of samples (Min Number Molecules).

CV Filter

A coefficient of variation (CV) filter was applied to the data which removes biomolecule(s) (Mass_Tag_IDs) with a CV greater than the threshold (Max CV).

imd-ANOVA Filter

An ANOVA filter can be applied to the data which removes biomolecule(s) (Mass_Tag_IDs) not having at least a minimum number of non-missing values per group (Min ANOVA). Additionally, an IMD (independence of missing data) filter can be applied to the data, removing biomolecules not having at least a minimum number of non-missing values (Min G-Test) in at least one of the groups.

Proteomics Filter

A degenerate peptide filter (Degenerate Peptides Removed) can be applied to the data, which identifies biomolecule(s) (Mass_Tag_IDs) not belonging to one protein. Additionally, a protein filter can be applied to the data which identifies proteins not having at least a minimum number of peptides (Min Number of Peptides) mapping to them.

rMD Filter

A robust Mahalanobis distance (rMD) filter was applied to the data, removing sample(s) (SampleIDs) with an associated rMD-associated p-value less than the threshold (P-Value Threshold). Metrics used to calculate the p-value are also included (Metrics Used).

Custom Filter

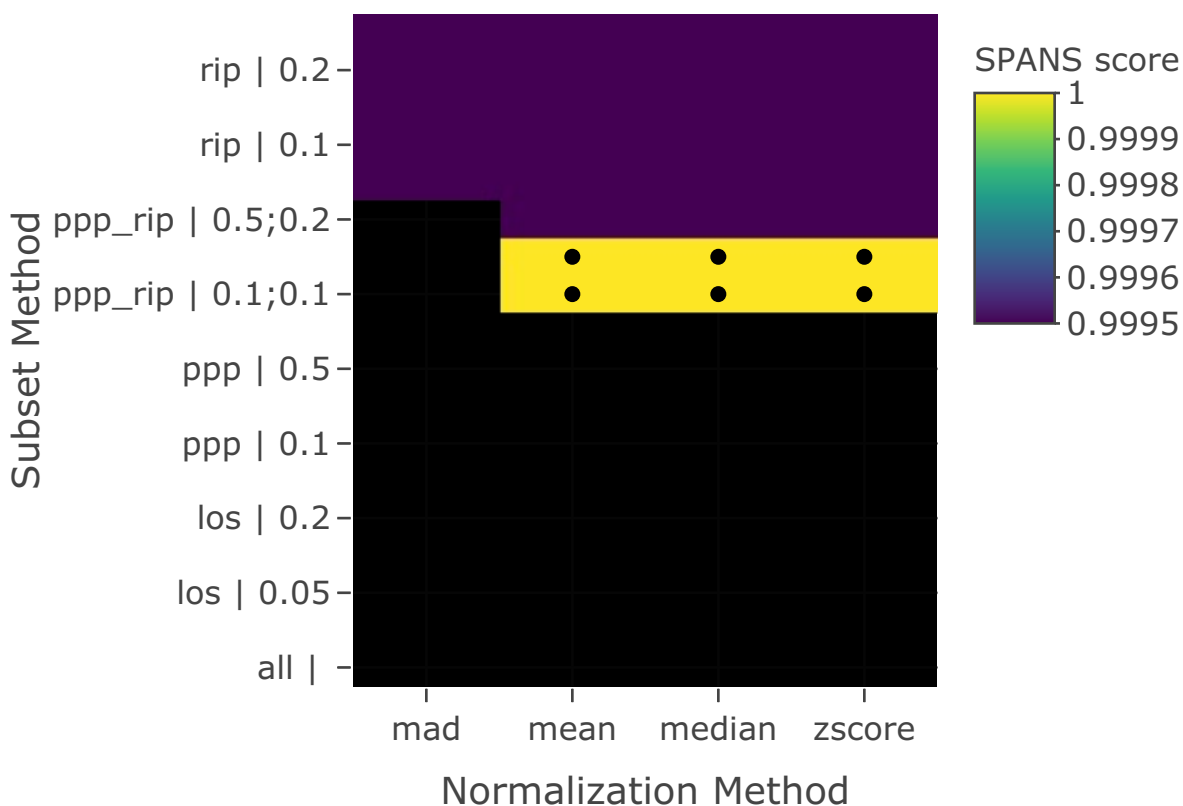
Custom filters can be used to remove samples, biomolecules, or proteins.

Normalization

Peptide data **was** normalized.

SPANS

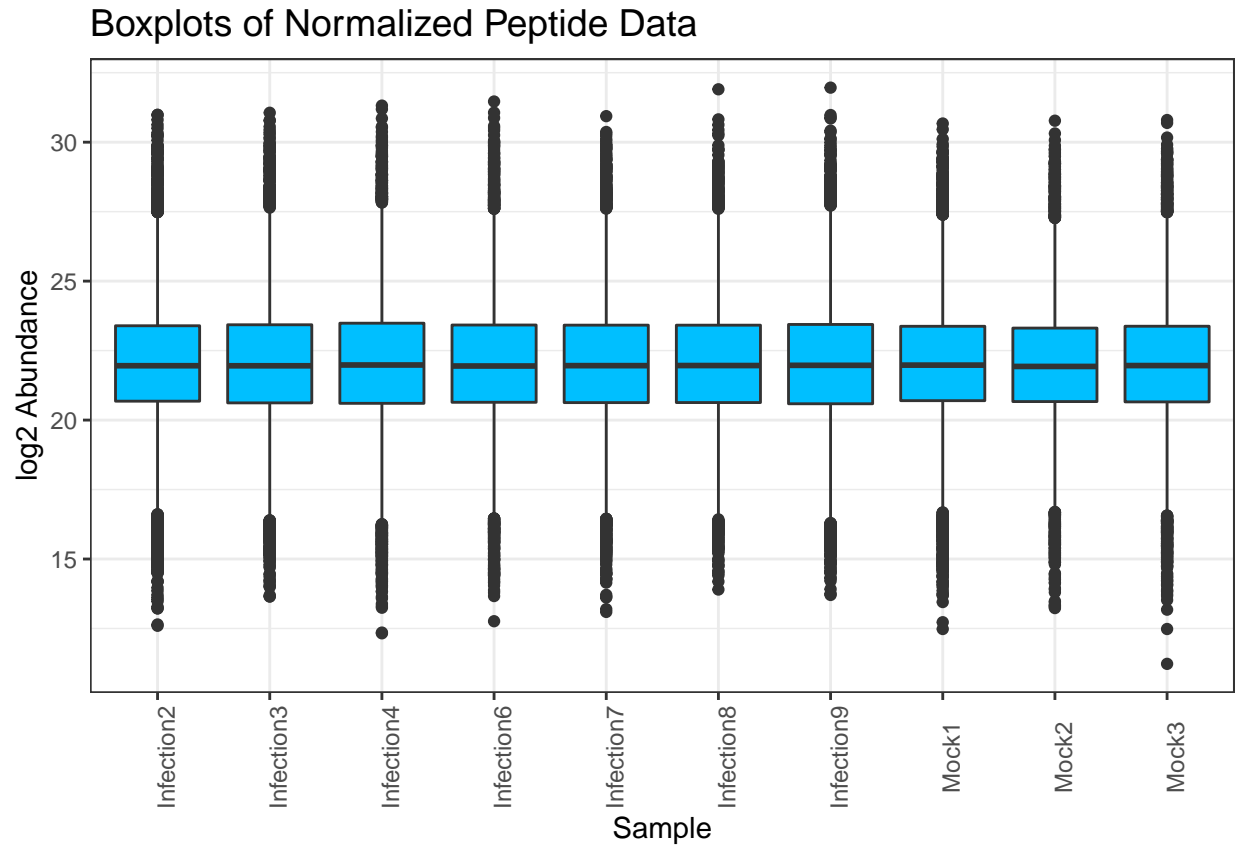
SPANS **was** run to determine optimum normalization parameters.



Manual

Manual normalization **was** used to normalize the data.

Attribute	Value
Subset Function	all
Subset Parameters	
Normalization Function	mean



Statistical Analysis

The statistical analysis step **was** run.

The following groups were compared: Infection_vs_Mock. Reported below are the parameters used in the statistical analysis, followed by the number of significant biomolecules for each comparison.

Attribute	Value
Test Method	anova
Multiple Comparison Adjustment	holm
Significance Threshold	0.05

	Comparison	Up_total	Down_total	Up_anova	Down_anova	Up_gtest	Down_gtest
Infection_vs_Mock	Infection_vs_Mock	1509	2742	1509	2742	0	

