

QRIME Scene Investigation - Overview

Andrew Holding & Matt Eldridge

9 January 2017

Contents

1	SampleSheet	2
2	Peptide intensity data	2
3	Normalization of intensity data	4
4	Protein-level quantification	5
5	Principal Component Analysis	5

1 SampleSheet

SampleSheet and ContrastTable configurations are loaded from tab-delimited format text files in the ‘Sample-Data’ folder. Reports are generated where requested and output to the ‘Reports’ folder. Plain text results are also provided in the ‘Results’ folder.

Group	PR590	PR622	PR650
ER 0min	127N	127N	128N
ER 45min	128C	127C	128C
ER 90min	129N	128N	129N
FOXA1 0min	129C	129C	129C
FOXA1 45min	130N	130N	130N
FOXA1 90min	130C	130C	130C
IgG	126	126	131

Table 1: Isobaric tags used for each sample (group) and run.

Contrast	Report
ER 45min vs ER 0min	Y
ER 90min vs ER 0min	Y
ER 90min vs ER 45min	N
FOXA1 45min vs FOXA1 0min	N
FOXA1 90min vs FOXA1 0min	N
FOXA1 90min vs FOXA1 45min	N

Table 2: Configured Constrasts.

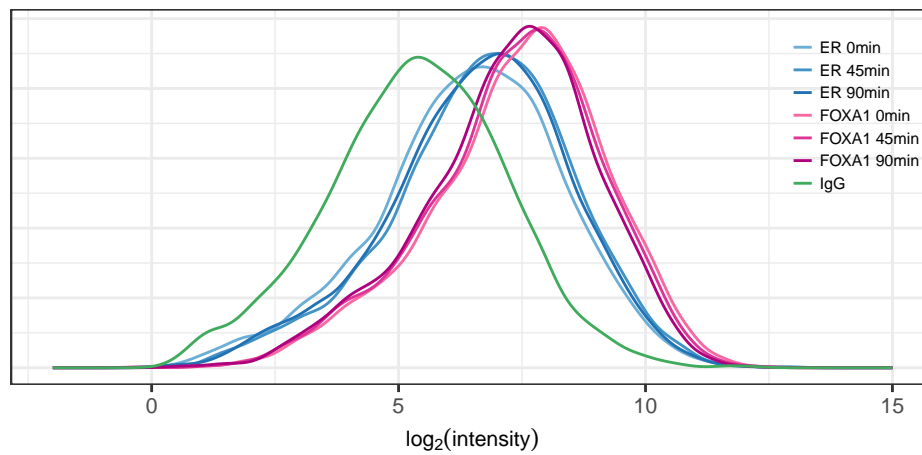
2 Peptide intensity data

	PR590	PR622	PR650	All
Peptides	5822	6952	3183	9321
Peptides with no missing values	5702	6872	3118	9179
Proteins	1125	1227	656	1553
Proteins with no missing values	1112	1218	645	1541

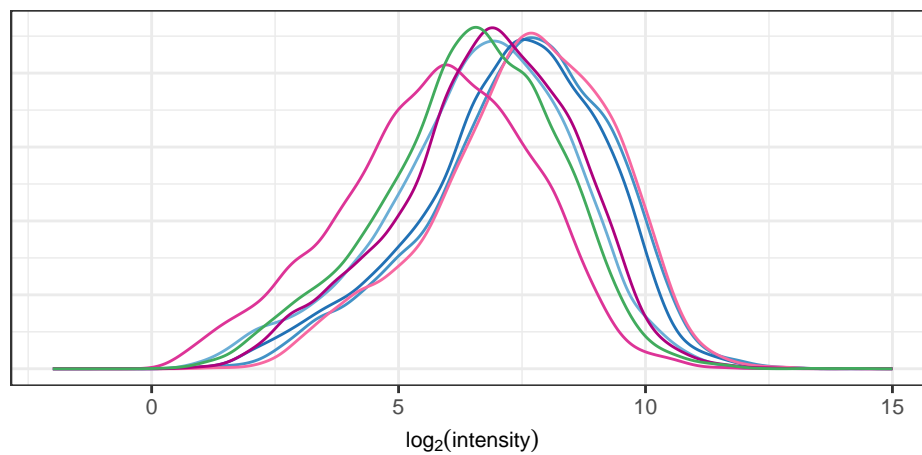
Table 3: Numbers of peptides and proteins observed in each run.

Figure 1 shows the distribution of intensities for each sample within each run.

PR590



PR622



PR650

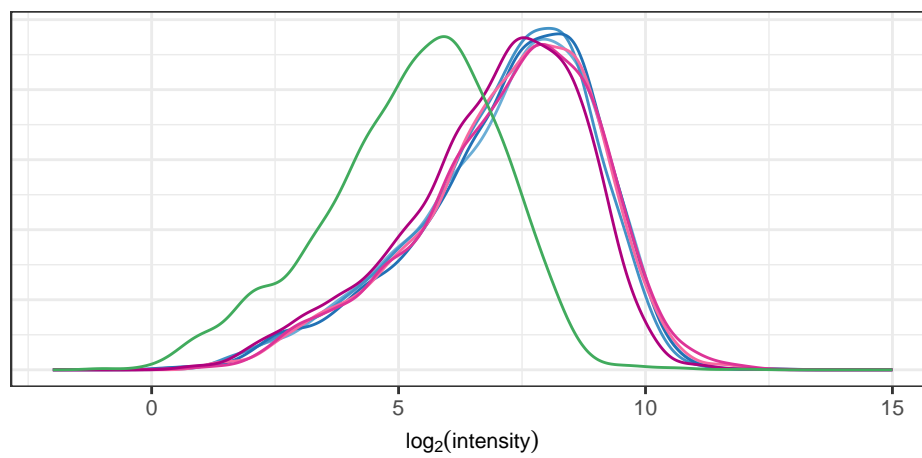


Figure 1: Density plots of intensities from each run.

3 Normalization of intensity data

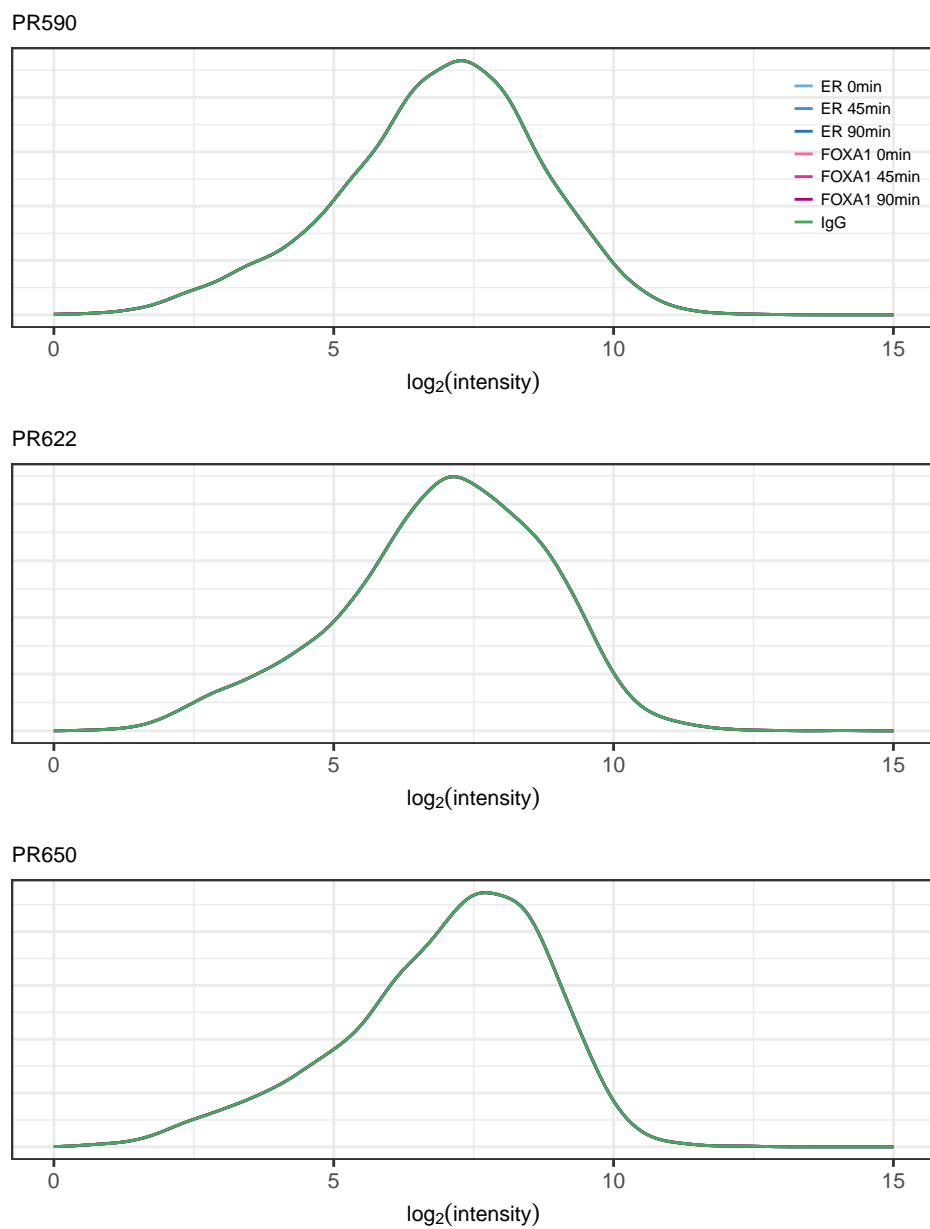


Figure 2: Density plots of normalized intensities from each TMT run where quantile normalization was applied to peptide intensities that include missing values.

4 Protein-level quantification

Runs/replicates	1	2	3	total
Including missing values	625	401	527	1553
Excluding missing values	625	398	518	1541

Table 4: Numbers of proteins identified in differing numbers of runs and total number of proteins identified in all runs.

5 Principal Component Analysis

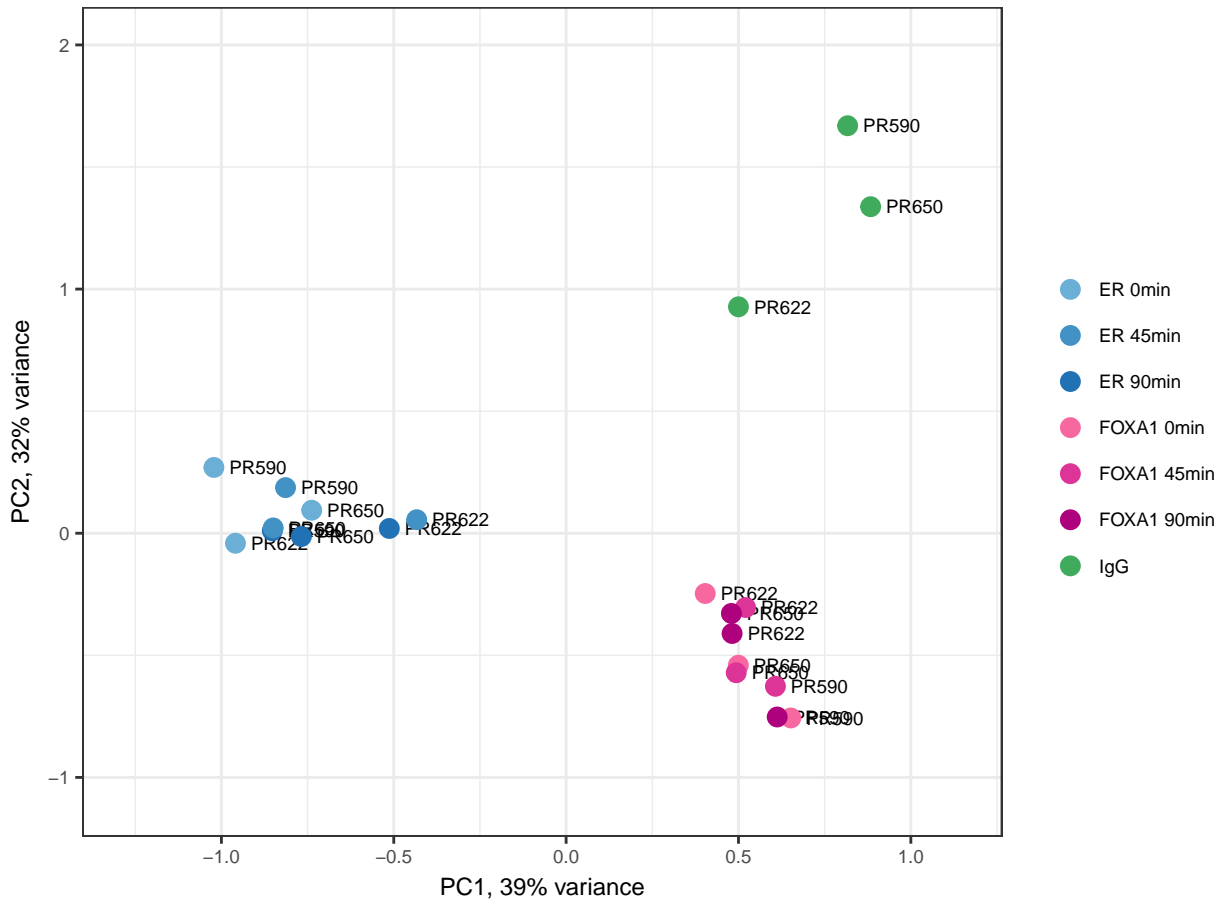


Figure 3: Principal Component Analysis for proteins sampled in all 3 runs. The PCA was based on protein-level data resulting from summation of quantile normalized peptide intensities in which missing values were imputed using KNN-based nearest neighbour averaging. The first two principal components are displayed.