Supplementary

Andrew Holding & Matt Eldridge

9 January 2017

Contents

0.1 Introduction

This report is from an analysis of the RIME-TMT proteomics data for an experiment carried out by Andrew Holding to quantify specific interactors of estrogen receptor (ER) and forkhead box protein A1 (FOXA1) in MCF-7 cells.

The RIME (Rapid Immunoprecipitation of Endogenous Proteins) technique was used to pull down ER and FOXA1 and their interacting proteins, which were subsequently assessed and quantitated using mass spectrometry with tandem mass tags (TMT). TMT runs were carried out for 3 biological replicates, estrogen-starved MCF-7 cells from the same cell line at different times/passages, before and after addition of estrogen with samples taken at 3 different time points – 0, 45 and 90 minutes after estrogen addition. A pull down of Immunoglobulin G (IgG) was also run as a control.

The digests from two of the biological replicates (PR622 and PR650) were analyzed with 3 separate runs of the mass spectrometer to increase coverage; data from the repeated runs were combined prior to the analysis carried out here. The other sample (PR590) was run once only.

Table 1 shows the isobaric tags used for each sample within each of the runs.

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.

0.2 Peptide intensity data

Raw spectra were processed using Proteome Discover 2.1 to produce peptide-level intensity data with a single set of intensity values per distinct peptide. Only peptides with unique high-confidence protein matches were included.

Multiple peptide-spectrum matches (PSMs) for the same peptide were combined using Proteome Discoverer by summing the PSM-level intensities. Peptide sequences with different modifications are treated as distinct peptides and the data provided contain intensities values for each modification identified for a peptide sequence. Several distinct peptide sequences (with modifications) may have been identified for any given protein.

Table 2 shows the numbers of distinct peptides and proteins obtained in each TMT run. Also given are the numbers of peptides and proteins after filtering peptides with missing intensity values in one or more TMT channels.

	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 2: Numbers of peptides and proteins observed in each run.

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

0.2.1 Missing intensity values

Missing intensity values can result even though ions are present at detectable concentrations due to the stochastic nature of the mass spectrometry acquisition method. It is reasonable to expect that these missing values are randomly distributed in the data. Alternatively, missing values may occur when there is a low abundance of ions, below the limit of detection of the instrument. These biologically relevant missing values are not randomly distributed, affecting only those proteins that are expressed at low levels in the samples analysed. The R/Bioconductor package `MSnBase` provides imputation methods for both types of missing value.

In this analysis, missing values have been handled in 3 different ways. The first approach is to exclude all peptide-level mesasurements where there is a missing value for one or more of the samples within a run. In addition, two imputation methods have been performed: one in which the missing values are set to the smallest non-missing value in the data for that run and the other in which k-nearest neighbour (KNN) averaging is applied.

Figure 2 shows the distribution of intensities for each sample within each run following imputation using the smallest non-missing values. A small hump just below zero on the $\log_2$ scale is clearly visible for each run.

0.3 Normalization of intensity data

The intensity distribution for PR622 suggests that the labeled samples have been pooled with differing protein concentrations and that some normalization is required to properly assess differences between groups. Normalization techniques that are commonly applied to microarray expression data and mRNA sequencing data assume that only a few genes, e.g. a few hundred out of tens of thousands, are expressed at very different levels between samples. This assumption may not hold for a RIME experiment where a specific set of interacting proteins is targeted. For example, if the majority of interacting proteins were to bind less strongly at one time point relative to another, quantile normalization or scale normalization approaches

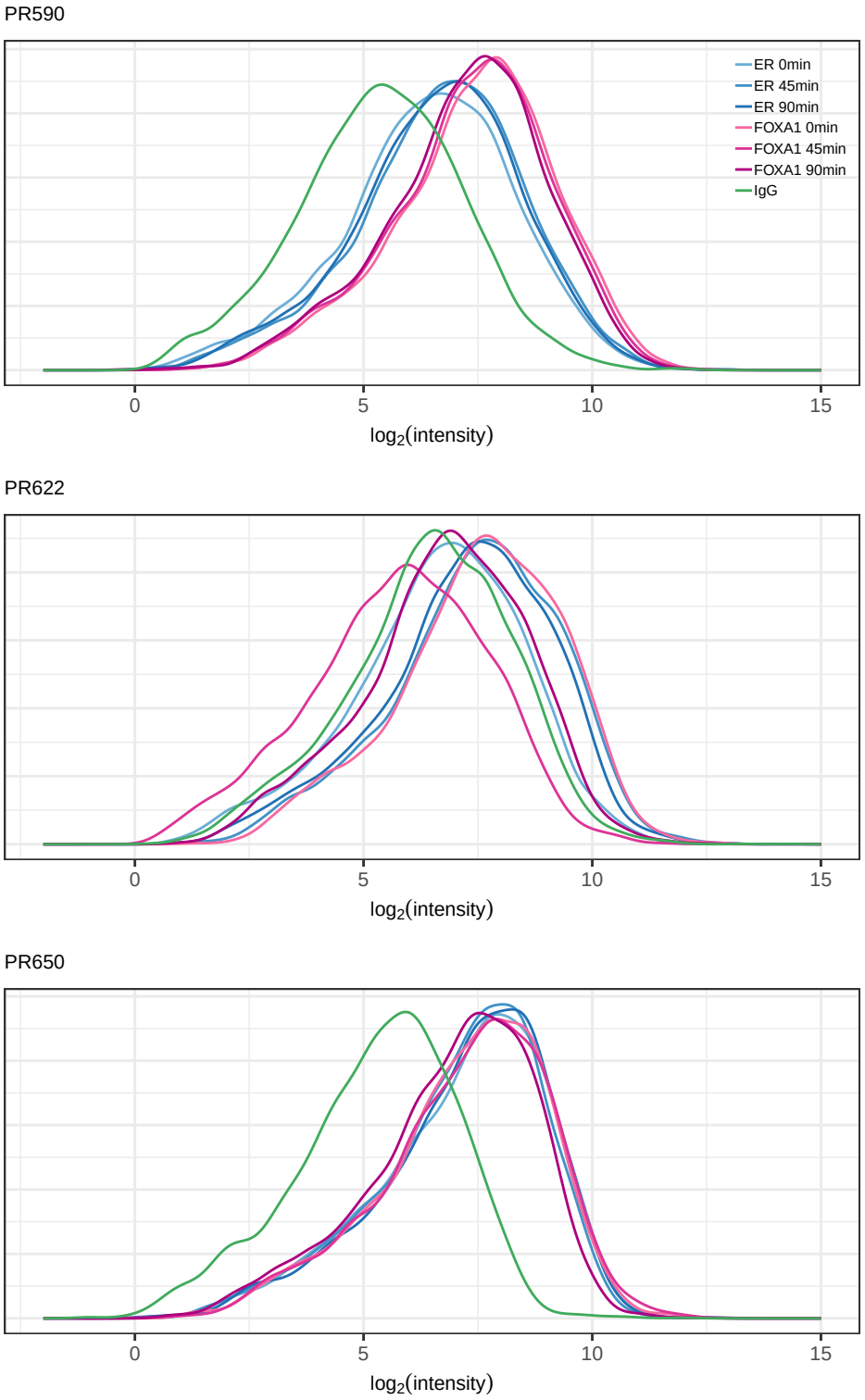

Figure 1: Density plots of intensities from each run.

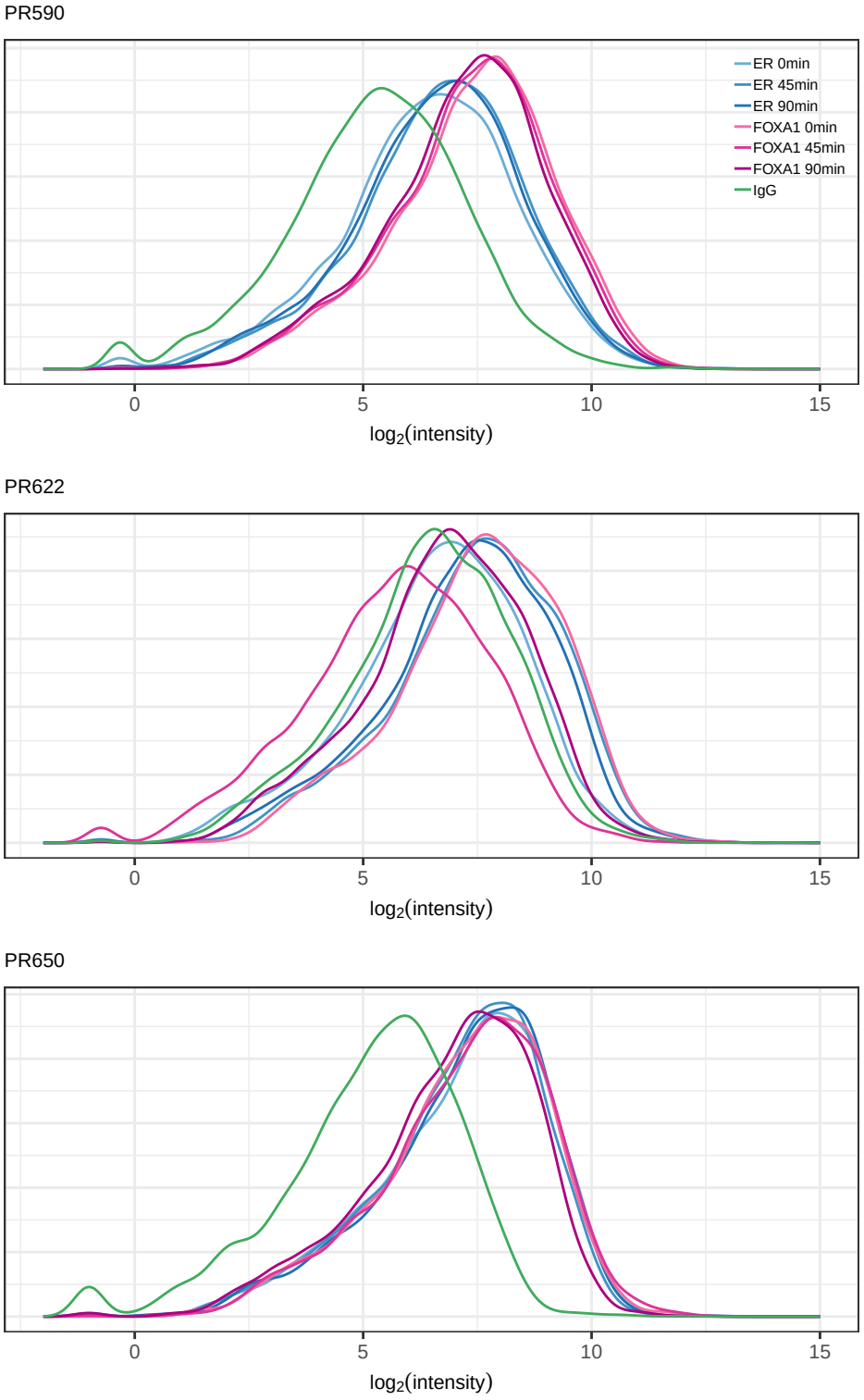

Figure 2: Density plots of intensities from each run following imputation of missing values using the smallest non-missing value in the data for each run.

would have the effect of removing the general trend. In this case, following normalization, those proteins that bind less strongly may appear to be largely unchanged, while proteins with very similar levels of binding at the different time points would have artificially increased binding levels. Therefore some care is required when interpreting the results of a differential binding analysis for RIME TMT experiments.

In this analysis, quantile normalization and scale normalization were applied for each run separately.

The IgG control was used to assess non-specific binding. The binding of a protein was considered to be non-specific if the binding level of that protein in the condition of interest, e.g. ER at 45 minutes, was not significantly above the level observed for the IgG control. Proteins detected with the IgG control are expressed at lower levels in the PR590 and PR650 runs. Normalization of the IgG control along with the other samples would make it more difficult to distinguish between specific and non-specific binding since the intensities of the IgG control would be scaled to a similar level of those of the other samples.

Among a number of normalization approaches tested here, normalizations were carried out that both include and exclude the IgG control. Figure 3 shows the normalized intensity distributions following scale normalization of all peptide intensities including those with missing values. Figure 4 shows the normalized intensity distributions following scale normalization for all samples except the IgG controls.

A further normalization approach was attempted that scales the data from all samples within a run based on a subset of peptides that are considered most likely to be at similar concentrations between the samples. The assumption is that the peptides with the highest intensity values in the IgG control are the result of non-specific binding and that that binding is consistent across all samples. Similar to the scale normalization applied in the first approach, the median intensity within each sample is computed to determine a scaling factor for each sample within a run but the median computed only uses the 10 peptides with the highest IgG measurements. As before, normalization is carried out within each run separately.

Figure 5 shows the resulting normalized intensity distributions. Strikingly, the IgG intensity distribution is shifted to lower values in run PR622, making it more consistent with the other 2 runs.

0.4 Protein-level quantification

Protein-level quantification was carried out for each run separately by summing normalized intensity values for all peptides matching to a particular protein. This is similar in principle to the gene-level quantification that is carried out during microarray analysis or the read-counting approach in RNA-Seq. Essentially, all signal attributable to a particular protein is assigned to that protein. Intensity values vary in magnitude depending on the elution profile for the peptide and when the peptide is sampled. Relative intensities for a peptide between the tags or samples within a run should be consistent but the signal may include some contribution from a contaminating, co-eluting peptide, the extent of which may differ for different PSMs. In summing intensities from multiple measurements for the same peptide, those peptides that have higher levels of intensity values will have a greater weight in determining the overall intensity; higher intensity measurements are likely to more accurately reflect differences between samples so this may be advantageous.

When no imputation of missing values was carried out, peptides with missing values in one or more samples were excluded from the summation. Missing values that arise for technical reasons instead of being at undetectable levels in a sample would otherwise reduce the overall intensity in that sample relative to other samples.

The statistical analysis of differential binding between groups that follows requires at least 3 observations per group. Table 3 gives the number of protein identified in just a single run, two runs or all three runs. A total of proteins were identified across all runs of which 527 were observed in all three runs, allowing for a statistical analysis. $\log_2$ fold changes are still computed for proteins identified only in one or two of the runs, but no measure of the statistical significance is given in the differential binding analysis.

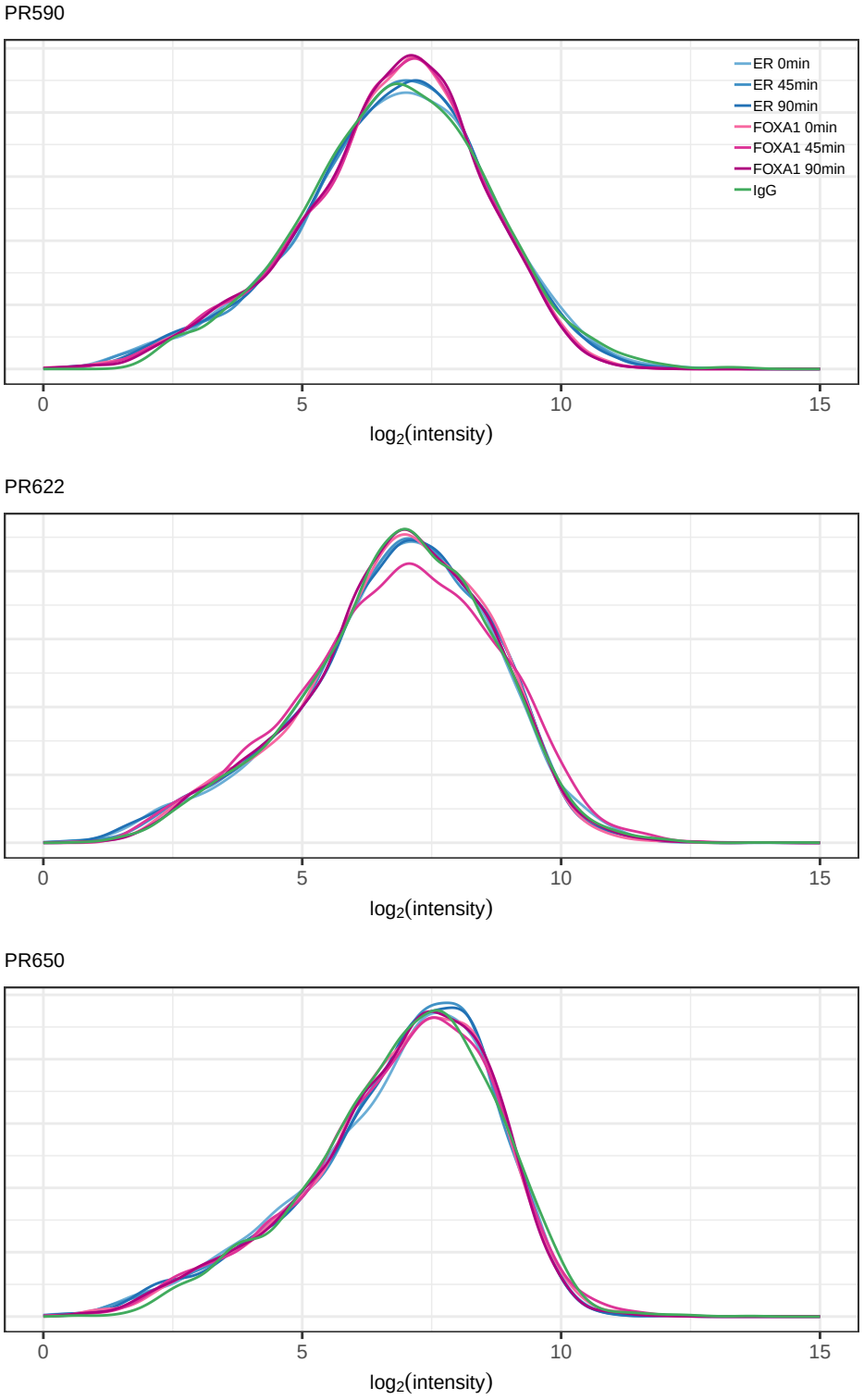

Figure 3: Density plots of normalized intensities from each TMT run where scale normalization was applied to peptide intensities that include misssing values.

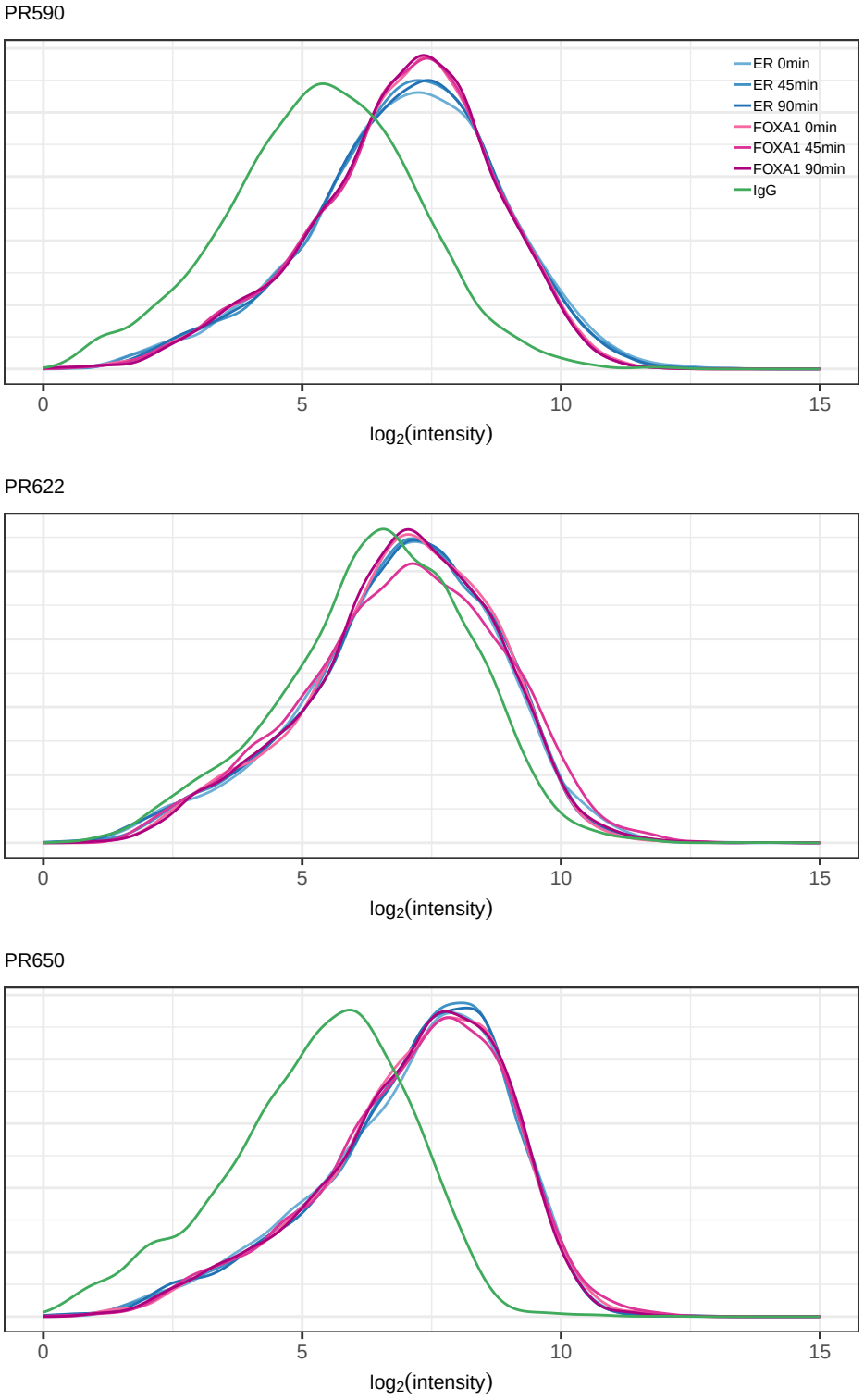

Figure 4: Density plots of normalized intensities from each TMT run where scale normalization was applied to peptide intensities for all samples except the IgG controls.

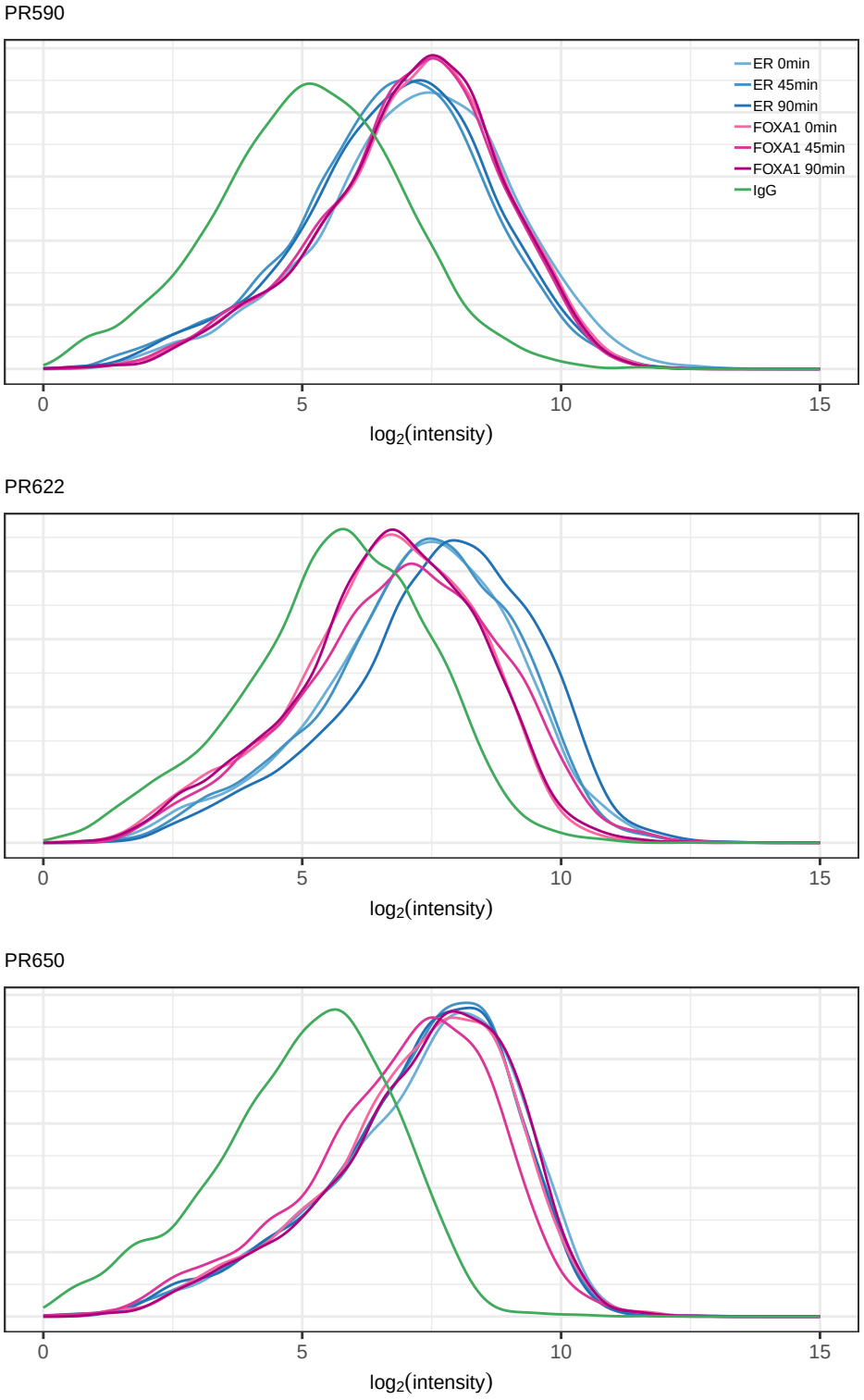

Figure 5: Density plots of normalized intensities from each TMT run where scale normalization is applied based on 10 peptides with the highest IgG intensities in each run.

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

Table 3: Numbers of proteins identified in differing numbers of runs and total number of proteins identified in all runs.

0.5 Principal Component Analysis

Protein-level intensities were scaled to sum to 1.0 for each protein within a run, allowing for comparison across runs in a principal component analysis (PCA).

Figure 6 shows a PCA plot for the first two principal components using all proteins that were sampled in all three runs. The plot shows a clear separation of the ER, FOXA1 and IgG control pull-downs. The 0, 45 and 90 minute timepoints within each of the ER and FOXA1 groups are not completely separated in the first two principal components. Figure 7 shows the PCA plot for just the ER samples.

0.6 Differential Binding

A statistical analysis of differentially-expressed peptides was carried out using **limma**, a R/Bioconductor package commonly used in the analysis of microarray and RNA-seq data, but applicable to data from any quantitative expression technology.

Limma uses linear models to assess differential expression in the context of multifactor experimental designs. It is able to analyze comparisons between many RNA targets simultaneously (or in this case many proteins) and has features that make these analyses stable even for experiments with small numbers of samples; this is achieved by borrowing information across genes (proteins).

In this analysis, limma was used to estimate $\log_2$ fold changes and standard errors by fitting a linear model for each protein where the model includes variables for the group (the ER and FOXA1 pull-downs at 0, 45 and 90 minute time points and the IgG control at 45 minutes) and the run. This is essentially a two-way ANOVA, a generalization of a paired analysis, in which comparisons between pull-downs at different time points are made within each run.

Limma employs an empirical Bayes method to moderate the standard errors of the estimated $\log_2$ fold changes; this results in more stable inference and improved power, especially for experiments with small numbers of replicates.

Multiple testing correction of p-values was applied using the Benjamini-Hochberg method to control the false discovery rate (FDR). The adjusted p-value (also known as a q-value) can be understood as follows. If all proteins with q-value below a threshold of 0.05 are selected as differentially expressed, then the expected proportion of false discoveries in the selected group is controlled to be less than the threshold value, in this case 5%.

The B-statistic (lods or B) is the log odds that the protein is differentially expressed. For example, if B = 1.5, the odds of differential expression is $e^{1.5} = 4.48$, i.e, about four and a half to one. The probability that the protein is differentially expressed is $4.48/(1 + 4.48) = 0.82$. A B-statistic of zero corresponds to a 50-50 chance that the gene is differentially expressed. The B-statistic has been automatically adjusted for multiple testing by assuming that 1% of the proteins are expected to be differentially expressed.

The IgG control was used to assess non-specific binding. The binding of a protein is considered to be non-specific if the $\log_2$ fold change relative to the IgG control is less than 1. In the differential binding analysis, protein intensity values are fitted to a linear model and the fitted values for each condition being compared are also compared to the IgG control. For each comparison of two groups, the maximum $\log_2$ fold change from each of the two groups above the IgG control is used to determine if the binding is specific.

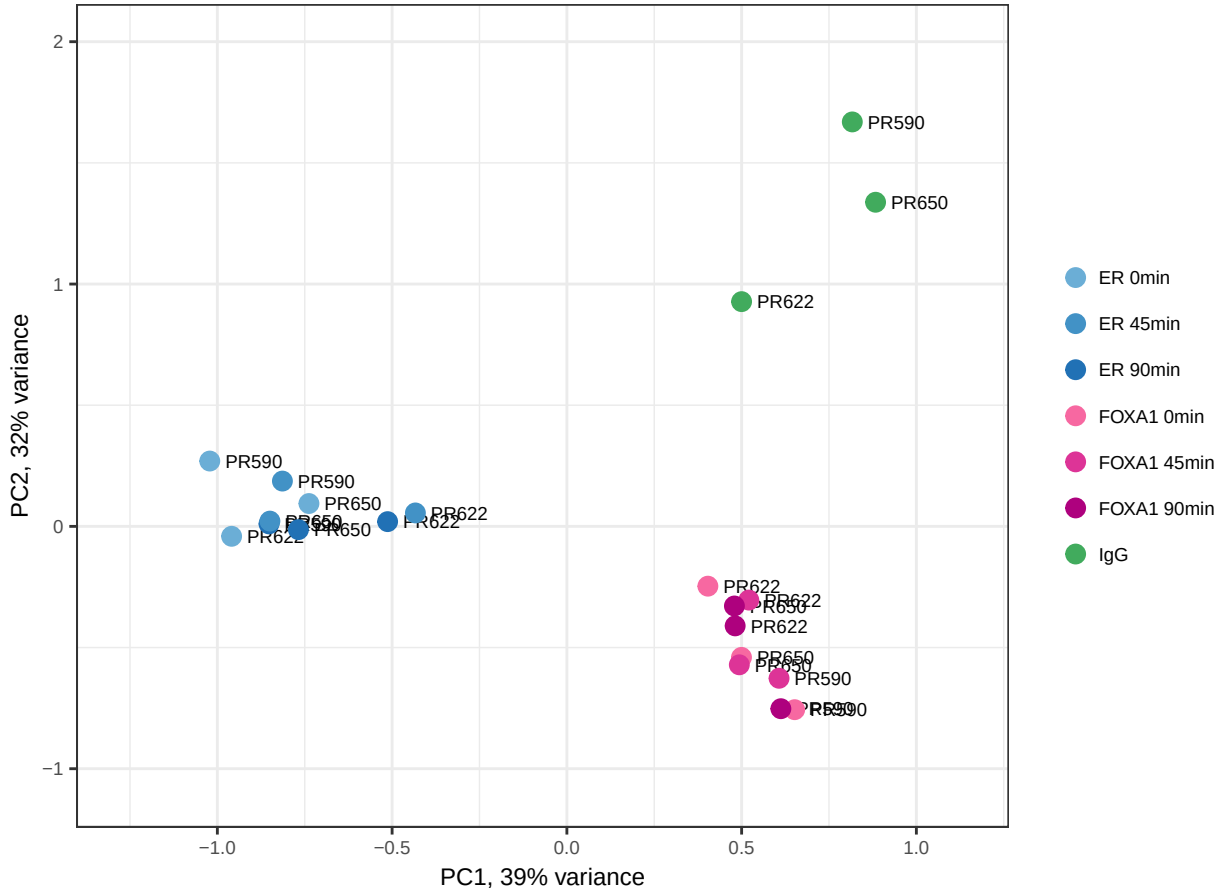

Figure 6: Principal Componenent Analysis for proteins sampled in all 3 runs. The PCA was based on protein-level data resulting from summation of quantile normalized peptide intensities in which missing values were imputed using KNN-based nearest neighbour averaging. The first two principal components are displayed.

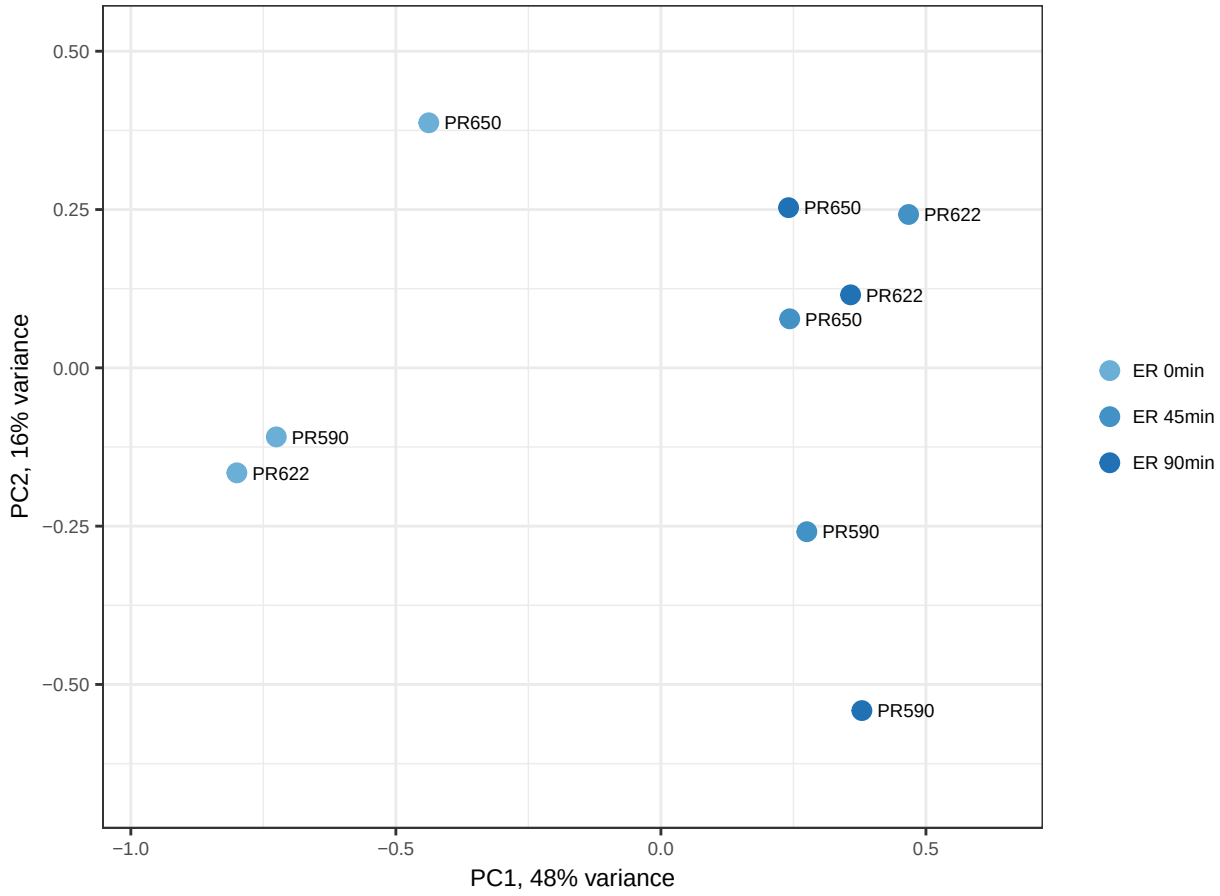

Figure 7: Principal Componenent Analysis for ER-interactng proteins sampled in all 3 runs. The first two principal components are displayed.

0.6.1 ER 45min vs ER 0min, excluding peptides with missing intensities, no normalization

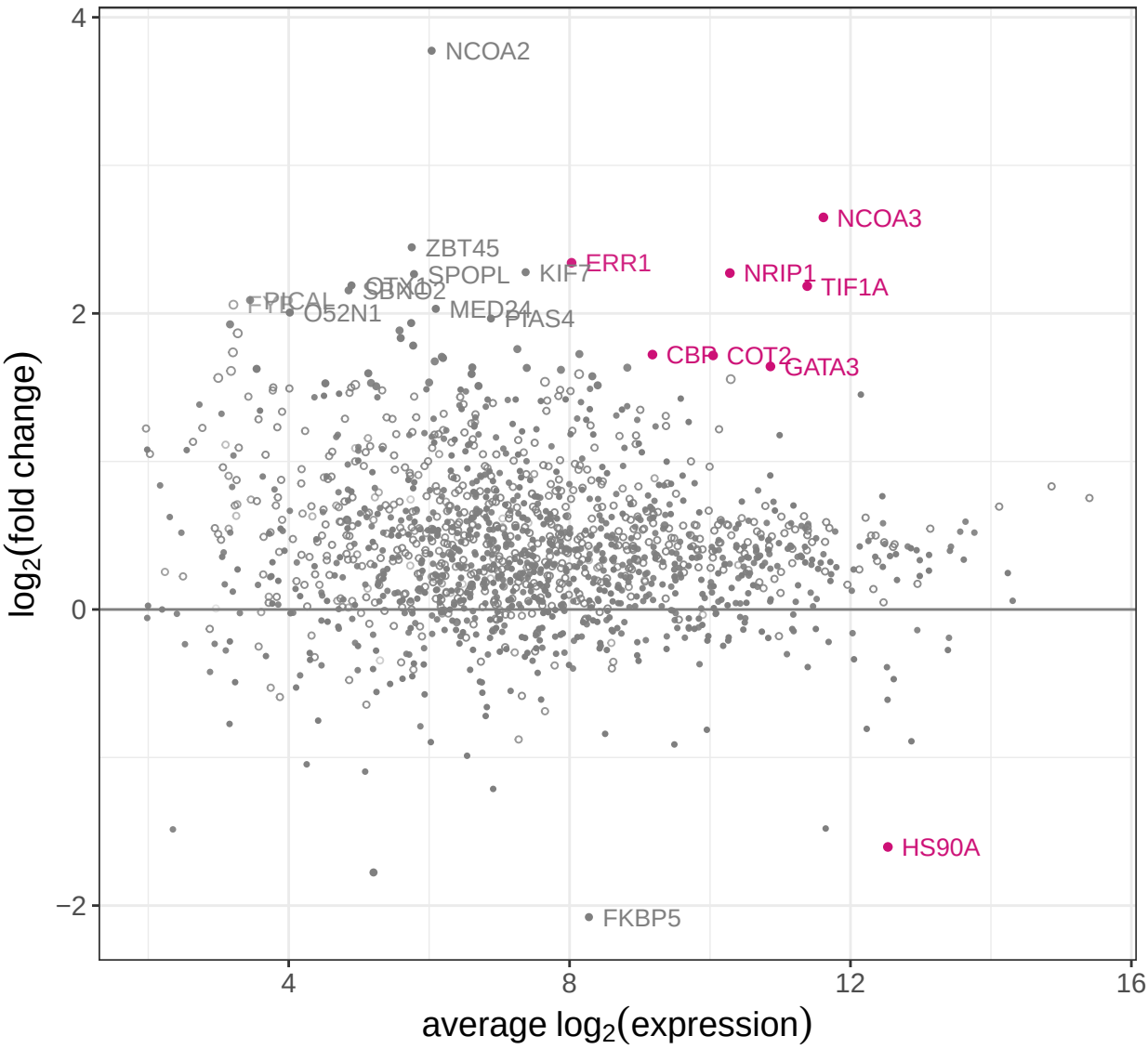

Figure 8: MA plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, no normalization). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

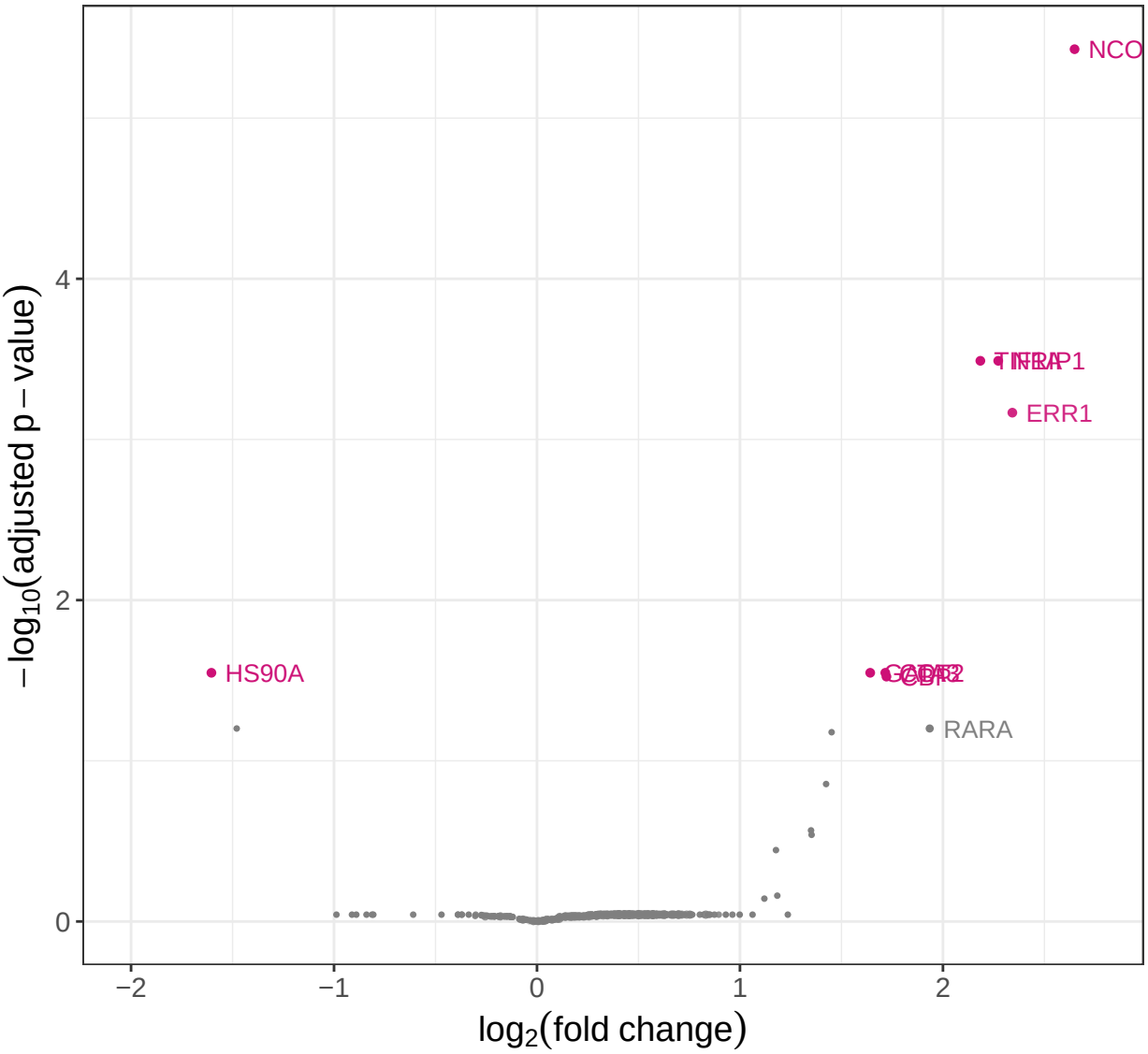

Figure 9: Volcano plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, no normalization). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

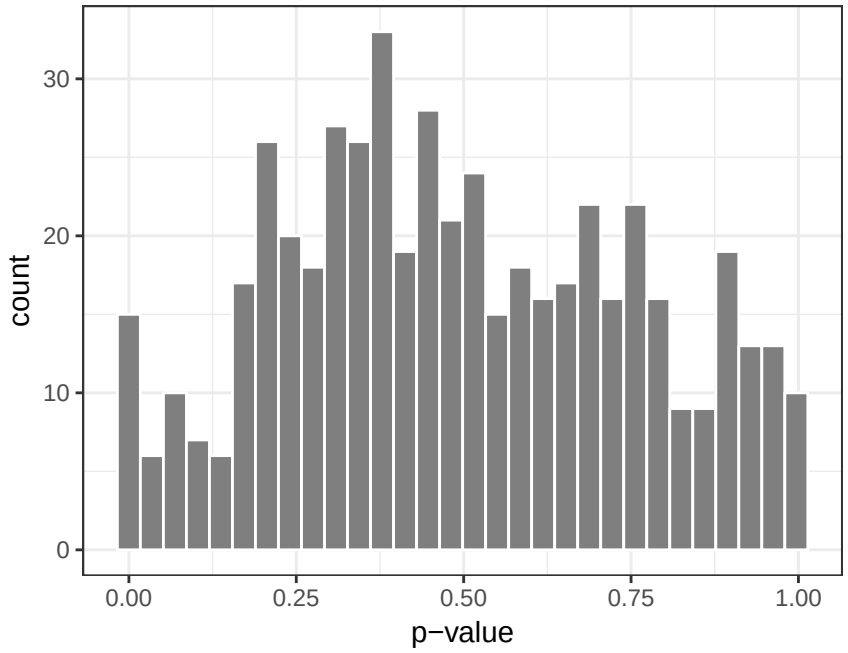

Figure 10: Histogram of p-values for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, no normalization)

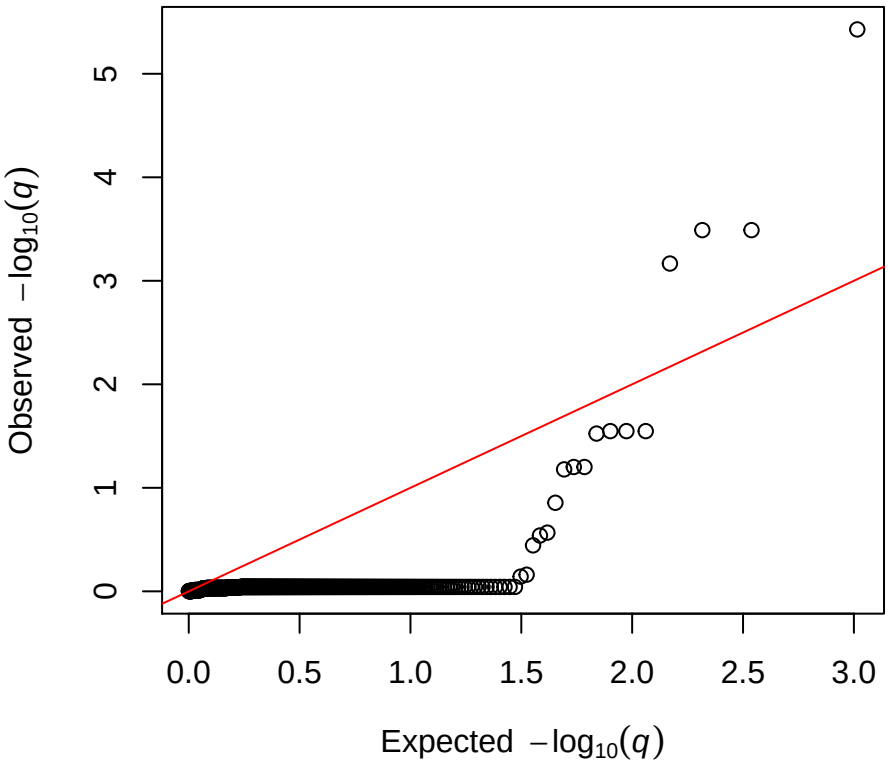

Figure 11: QQ plot of the adjusted p-values for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, no normalization)

0.6.2 ER 45min vs ER 0min, excluding peptides with missing intensities, quantile normalization

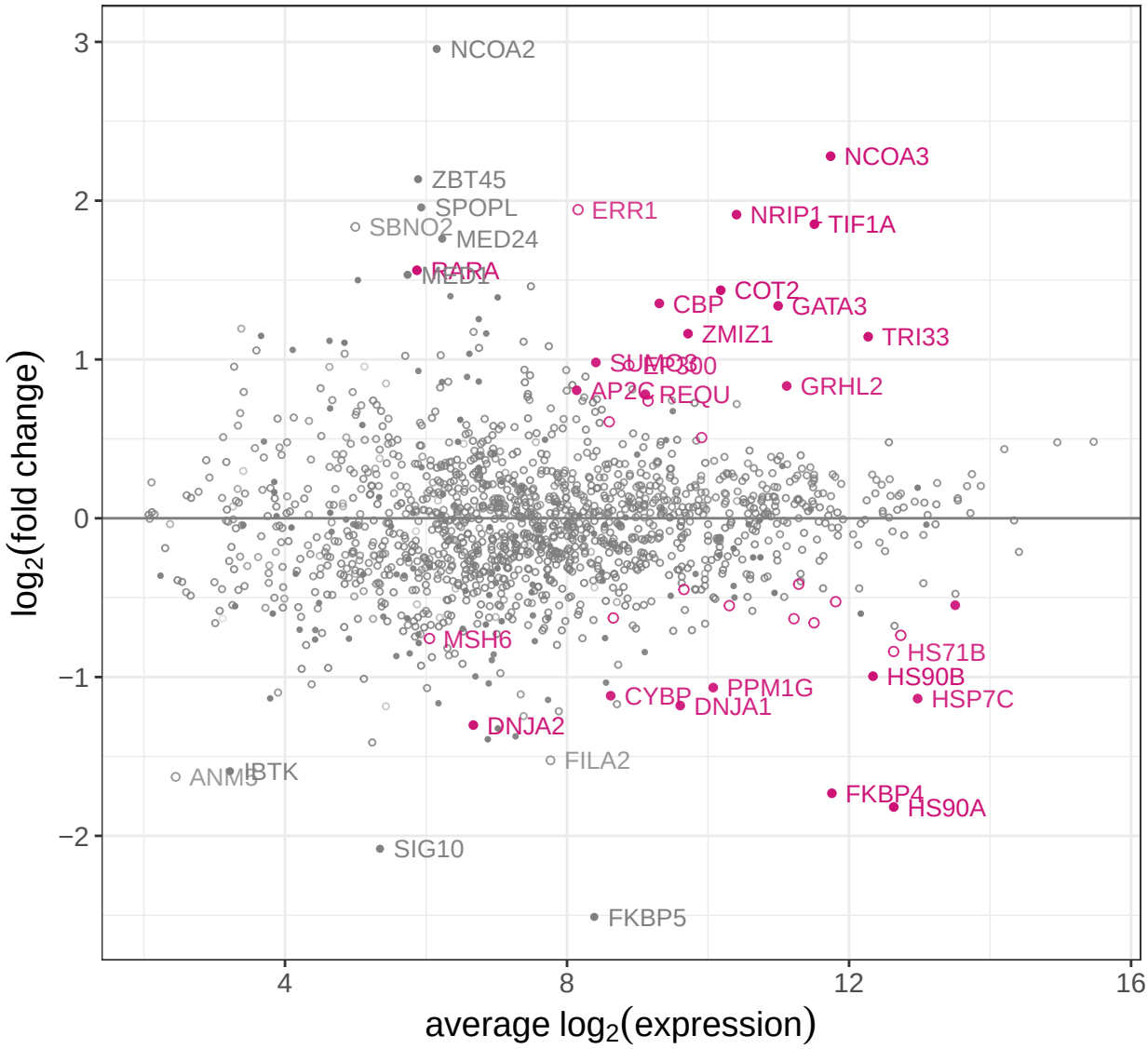

Figure 12: MA plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, quantile normalization). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

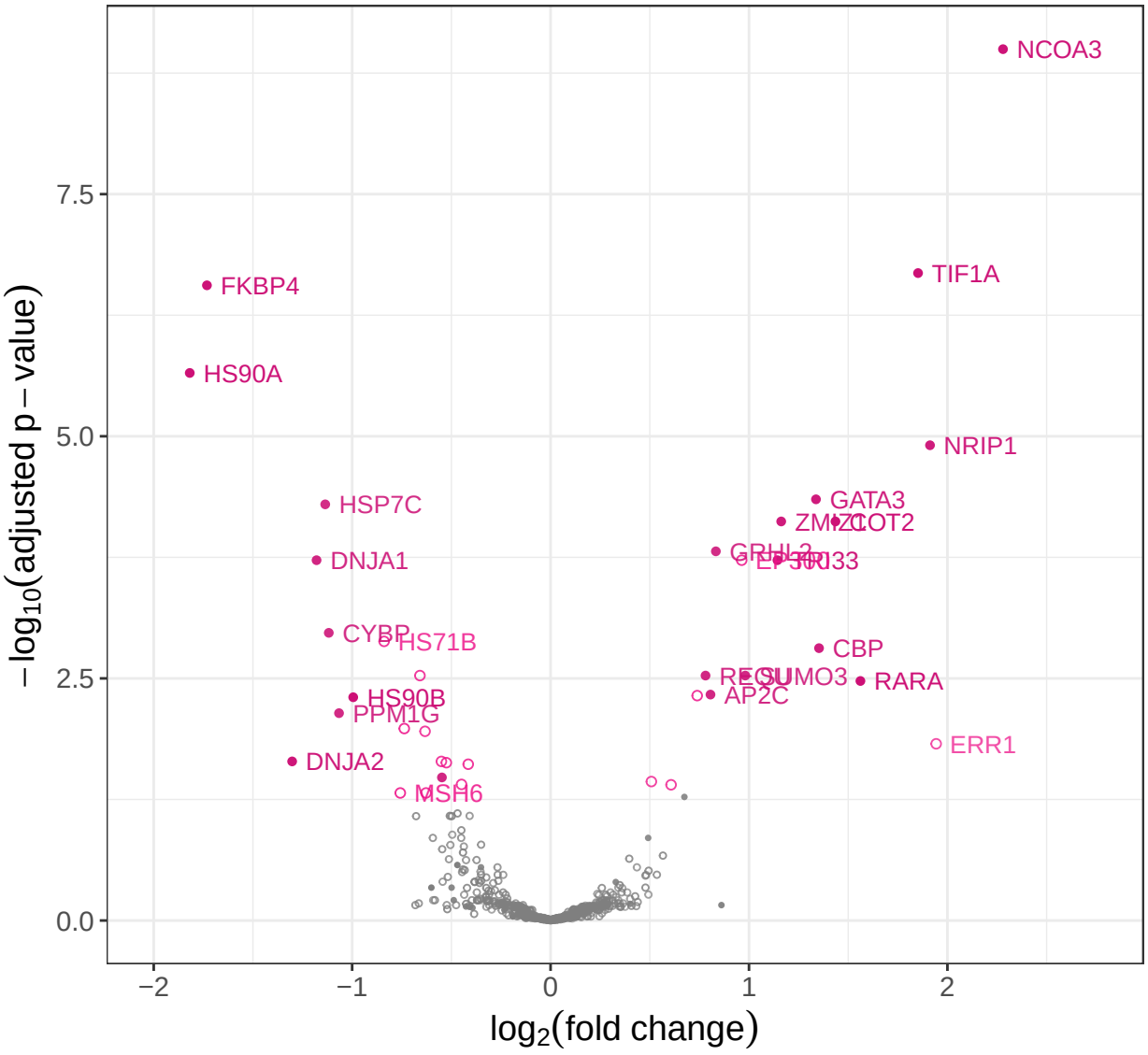

Figure 13: Volcano plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, quantile normalization). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

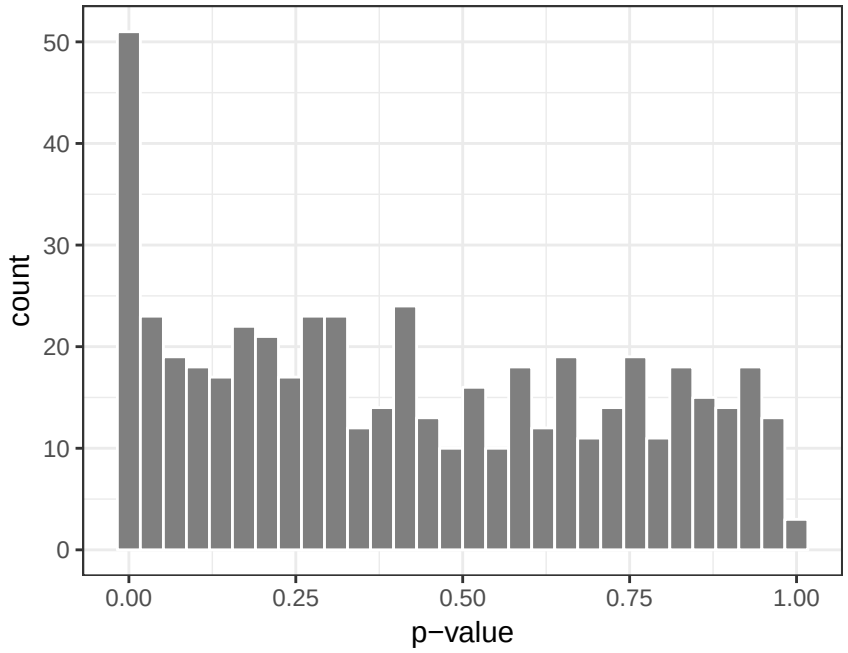

Figure 14: Histogram of p-values for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, quantile normalization)

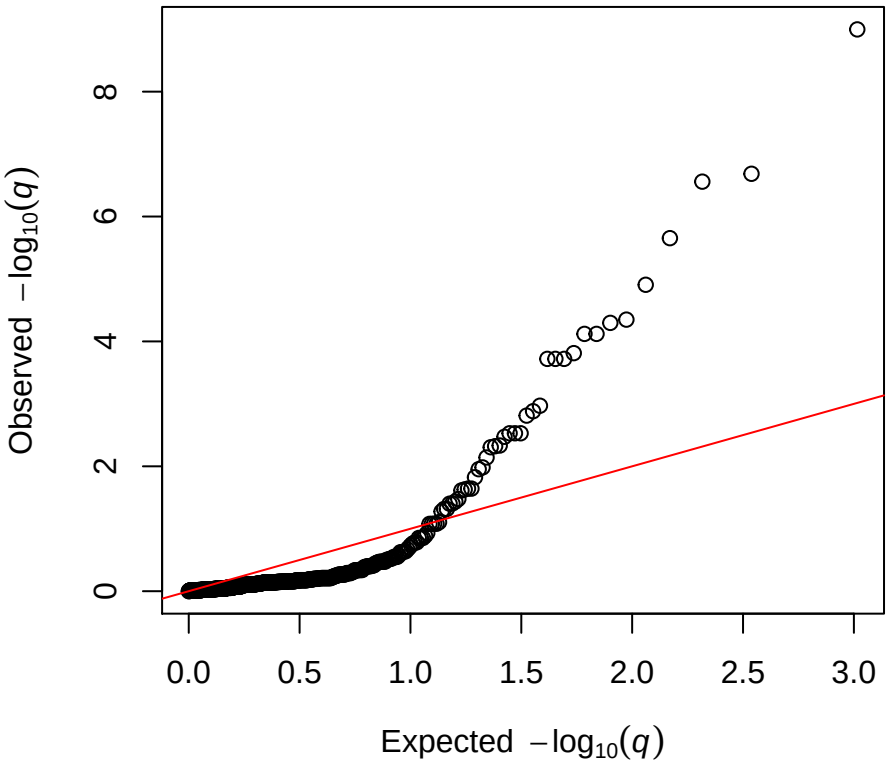

Figure 15: QQ plot of the adjusted p-values for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, quantile normalization)

0.6.3 ER 45min vs ER 0min, excluding peptides with missing intensities, quantile normalization excluding IgG control

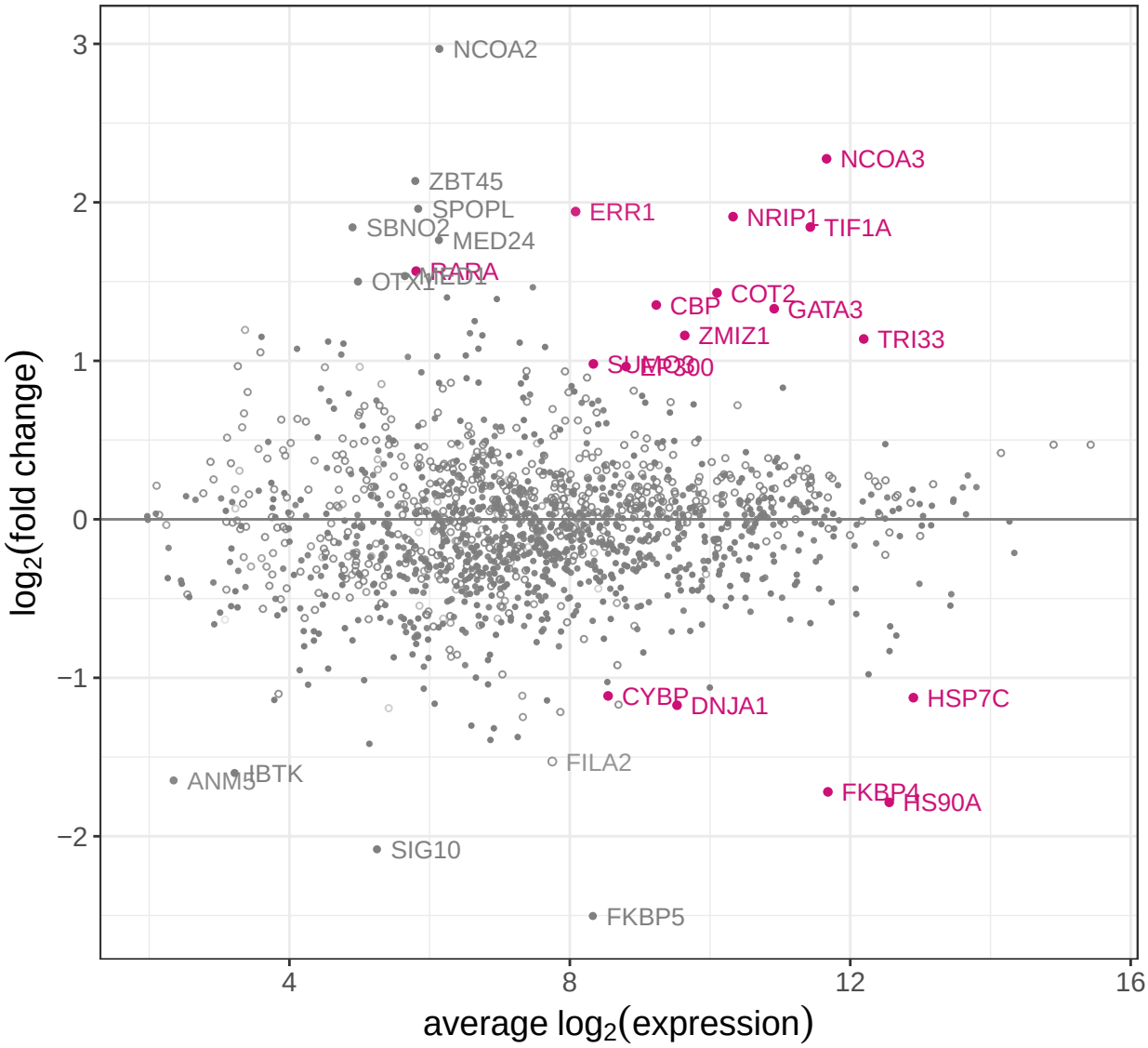

Figure 16: MA plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, quantile normalization excluding IgG control). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

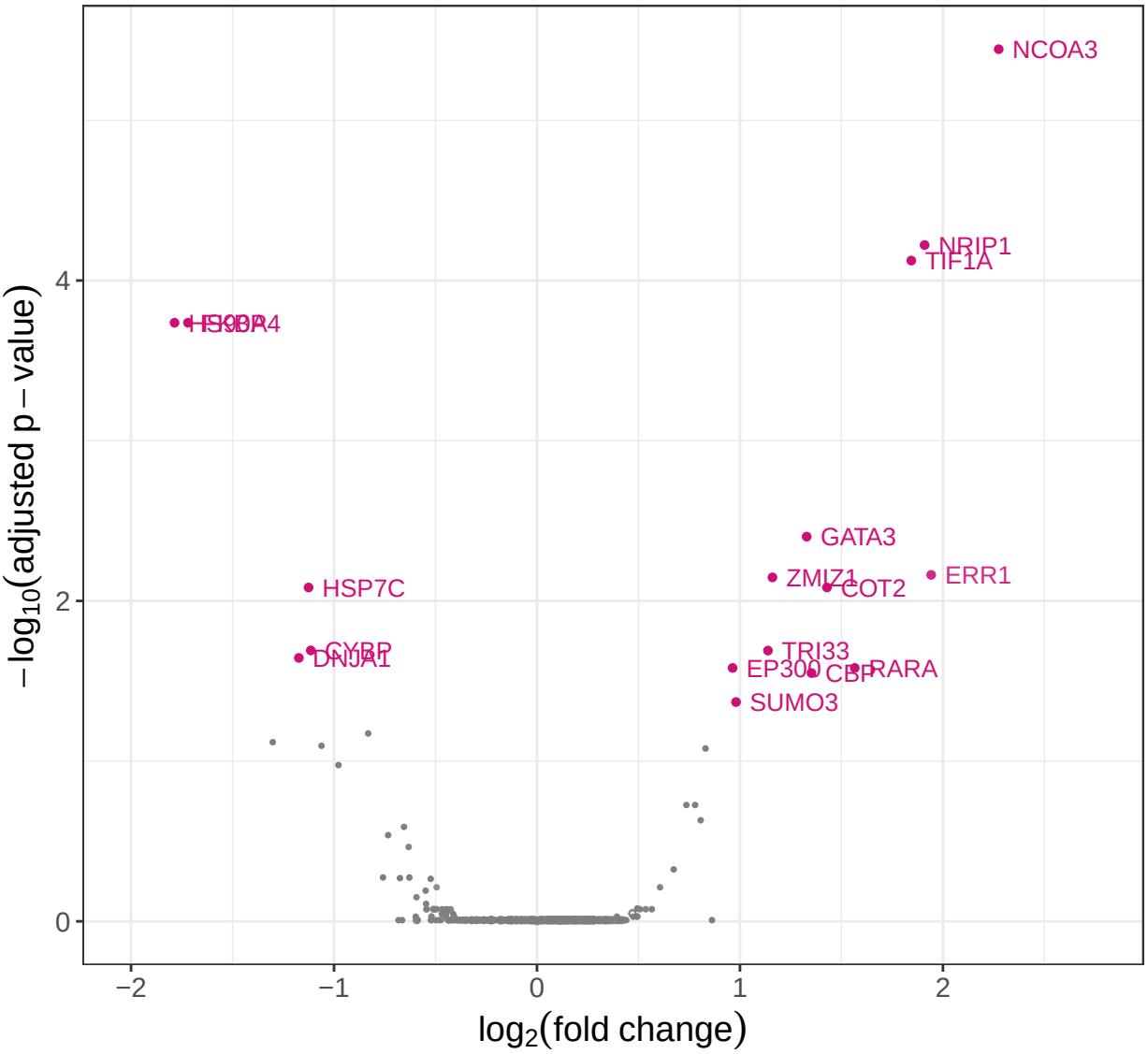

Figure 17: Volcano plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, quantile normalization excluding IgG control). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

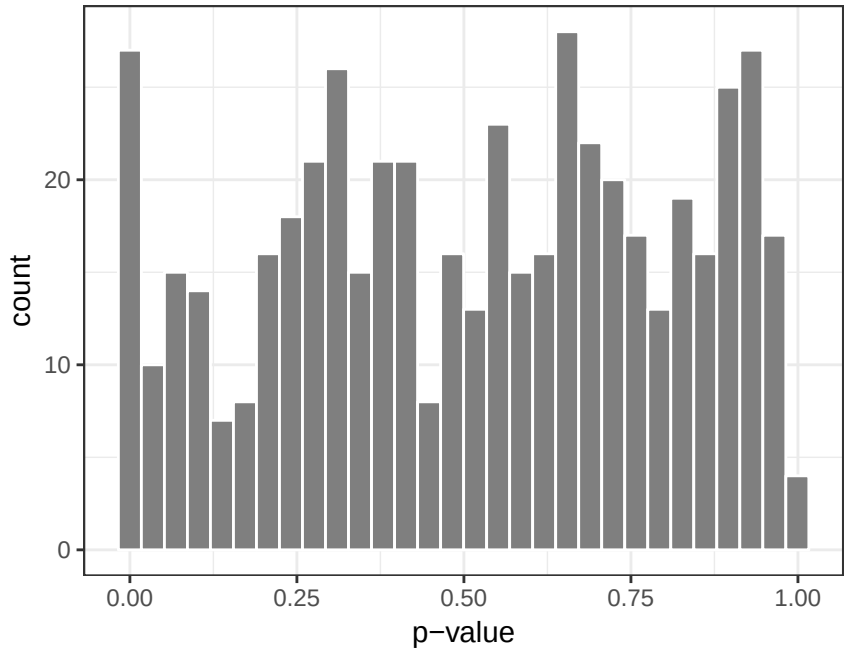

Figure 18: Histogram of p-values for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, quantile normalization excluding IgG control)

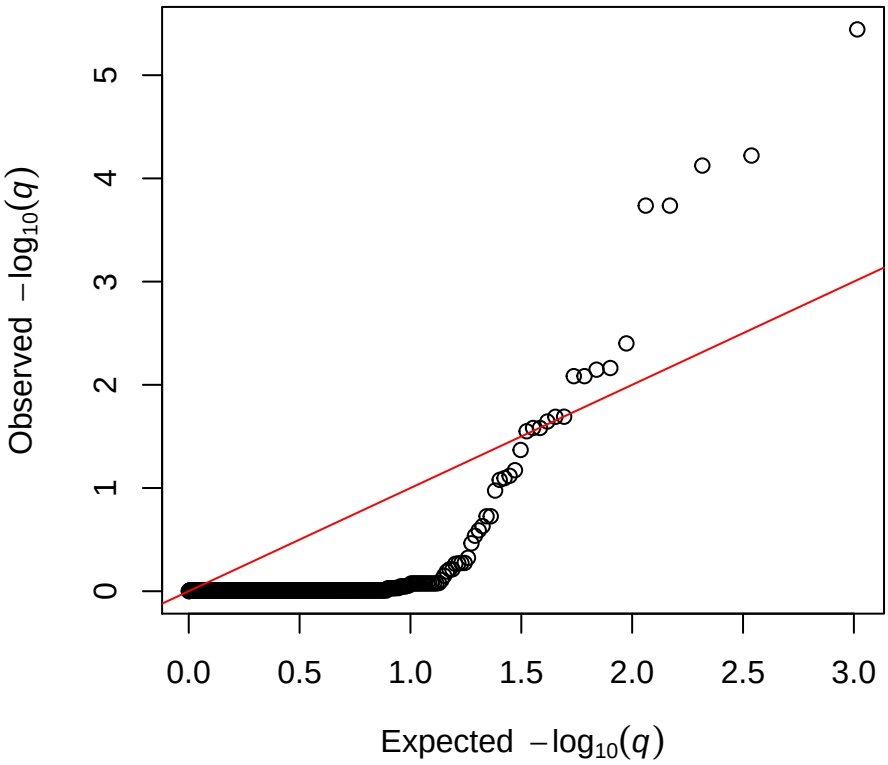

Figure 19: QQ plot of the adjusted p-values for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, quantile normalization excluding IgG control)

0.6.4 ER 45min vs ER 0min, excluding peptides with missing intensities, scale normalization

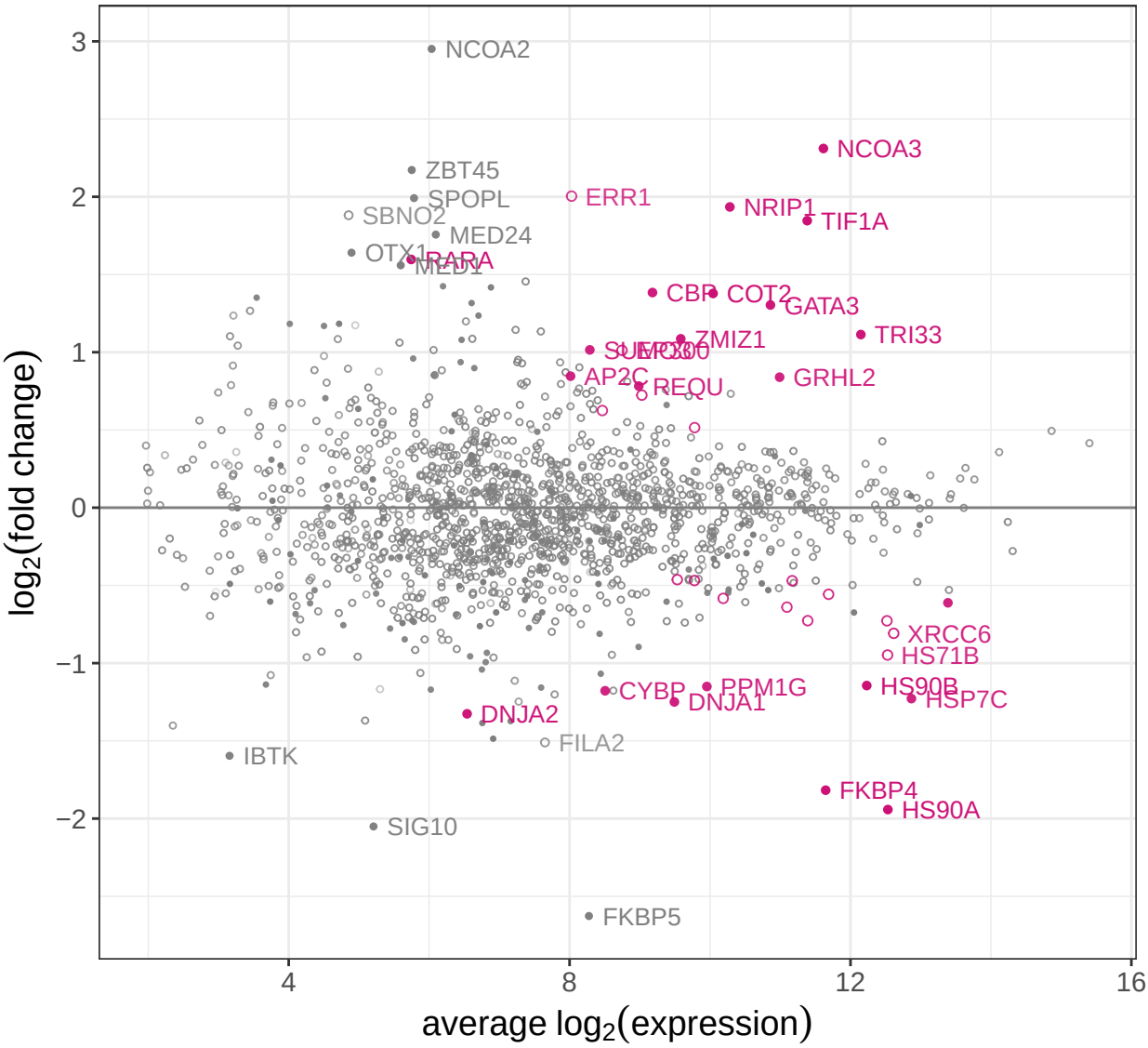

Figure 20: MA plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, scale normalization). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

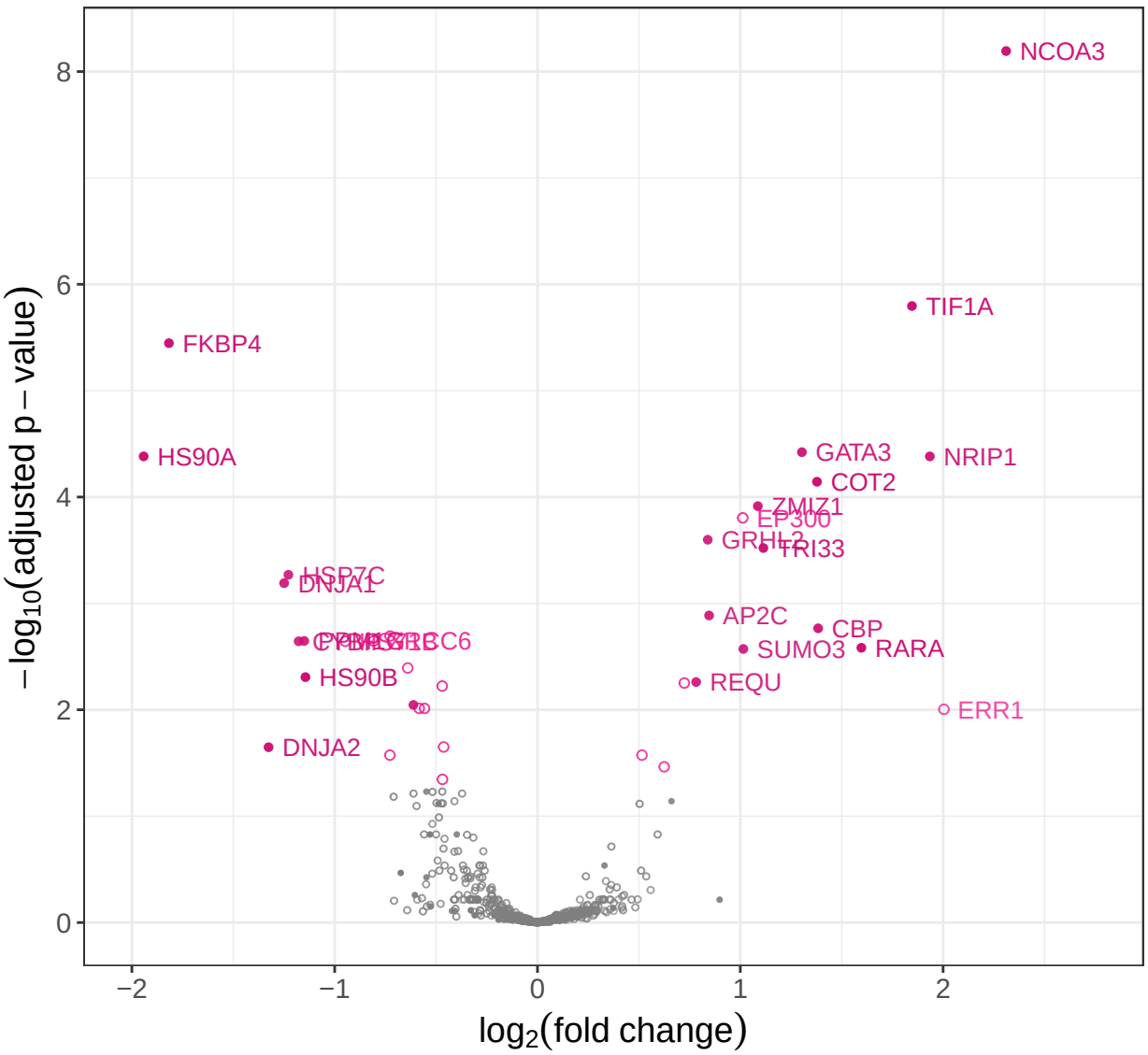

Figure 21: Volcano plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, scale normalization). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

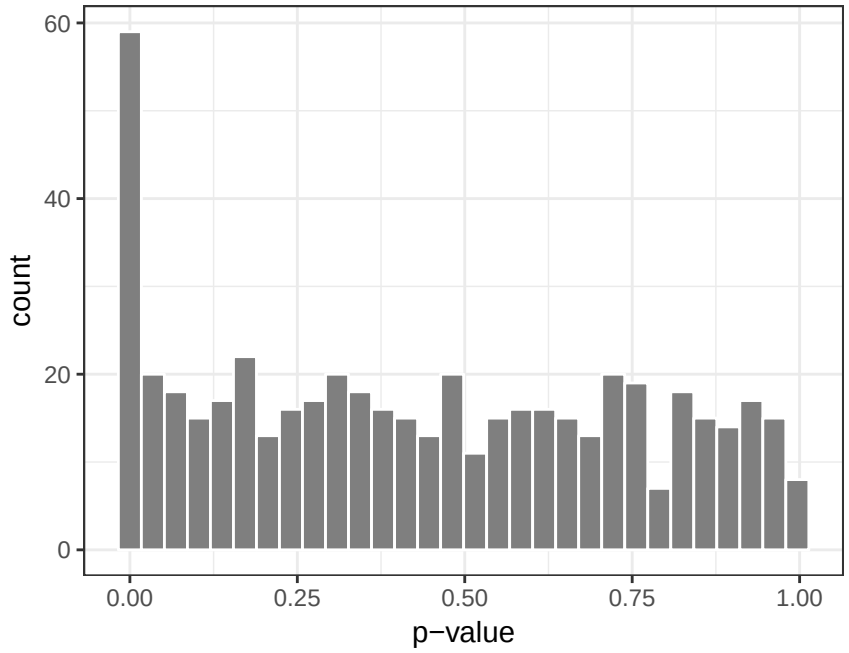

Figure 22: Histogram of p-values for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, scale normalization)

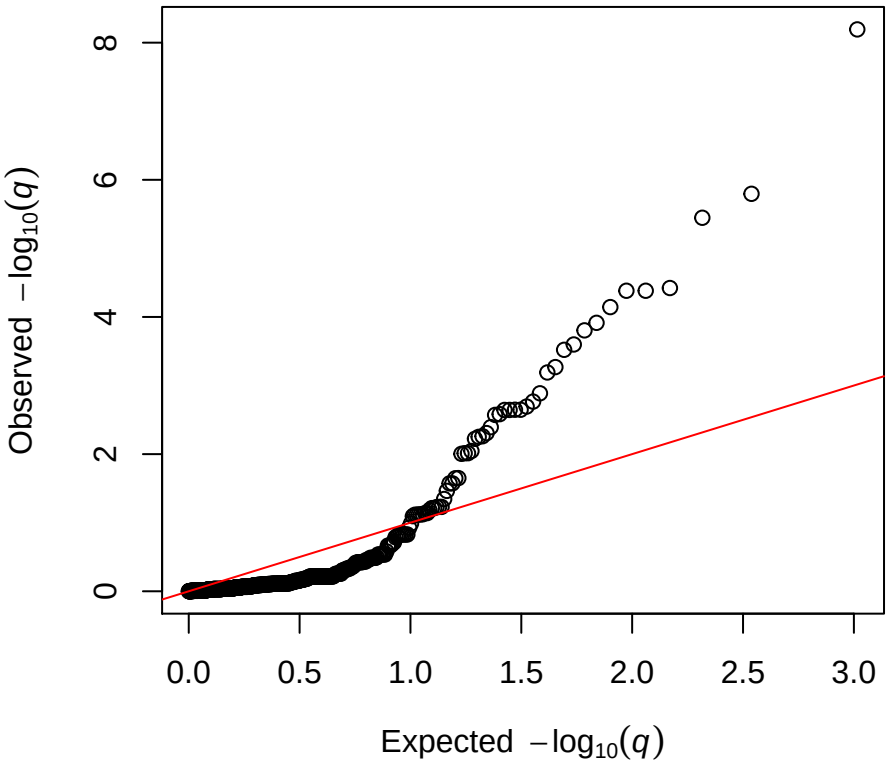

Figure 23: QQ plot of the adjusted p-values for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, scale normalization)

0.6.5 ER 45min vs ER 0min, excluding peptides with missing intensities, scale normalization using top 10 IgG peptides

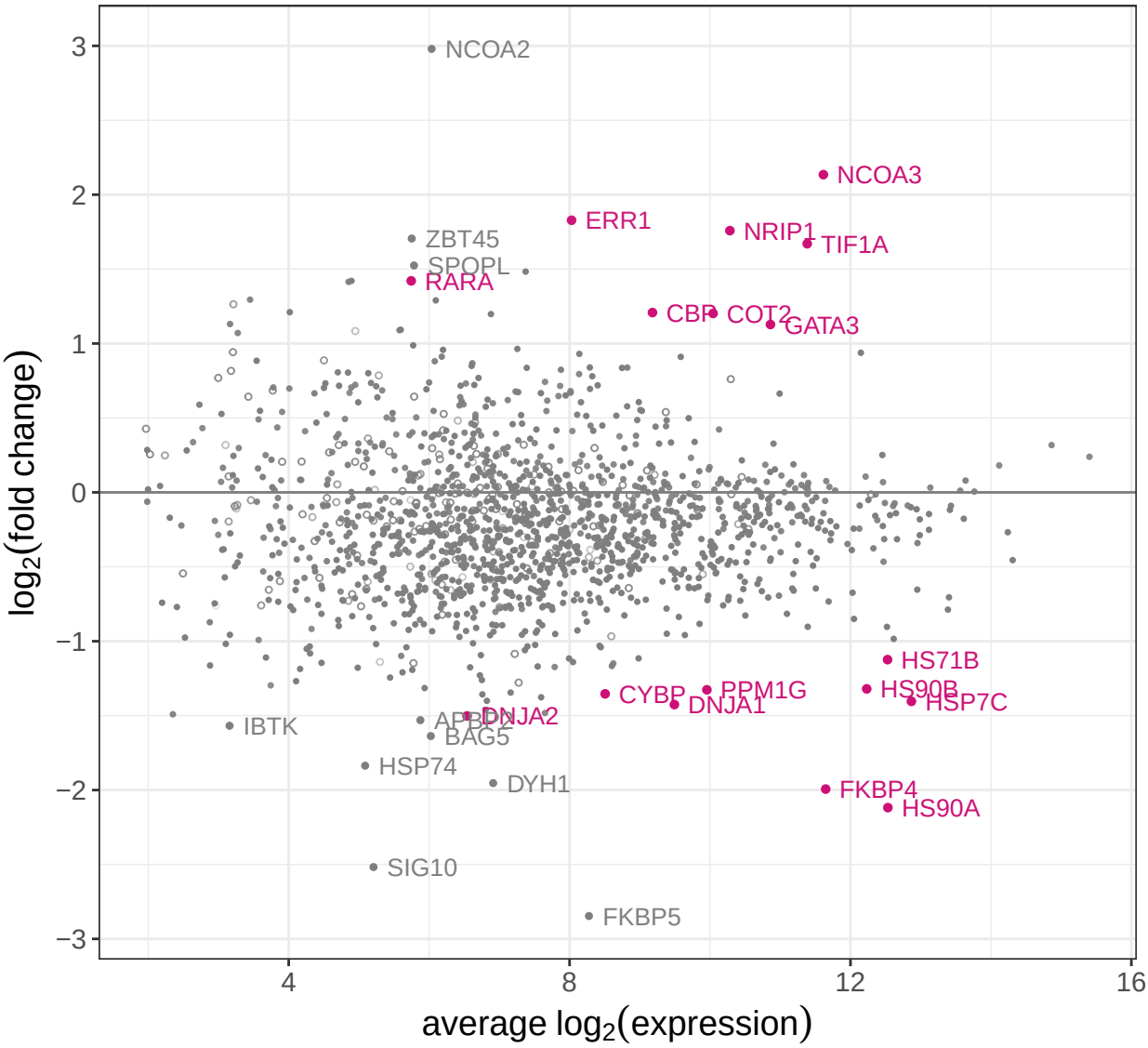

Figure 24: MA plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, scale normalization using top 10 IgG peptides). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

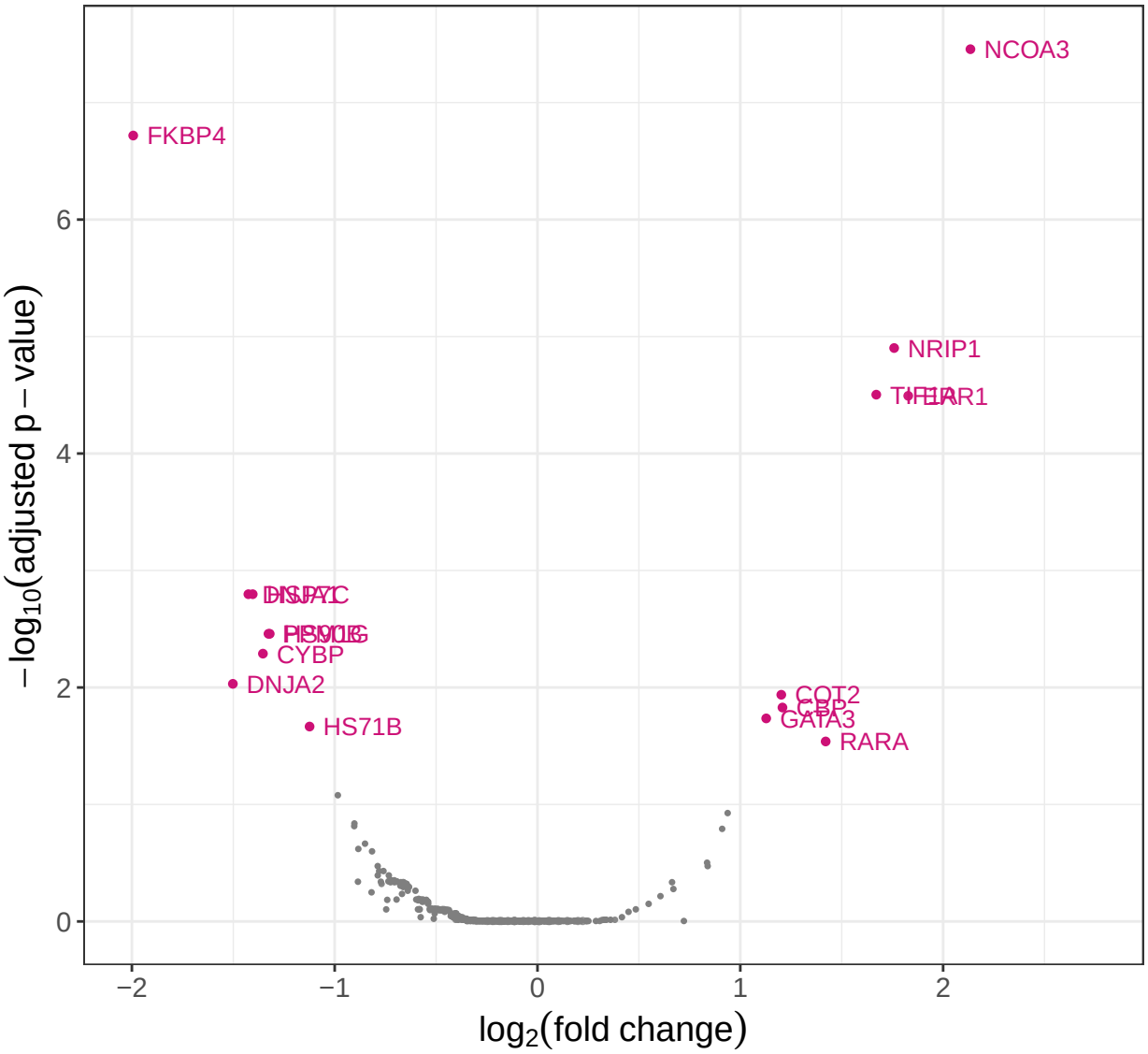

Figure 25: Volcano plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, scale normalization using top 10 IgG peptides). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

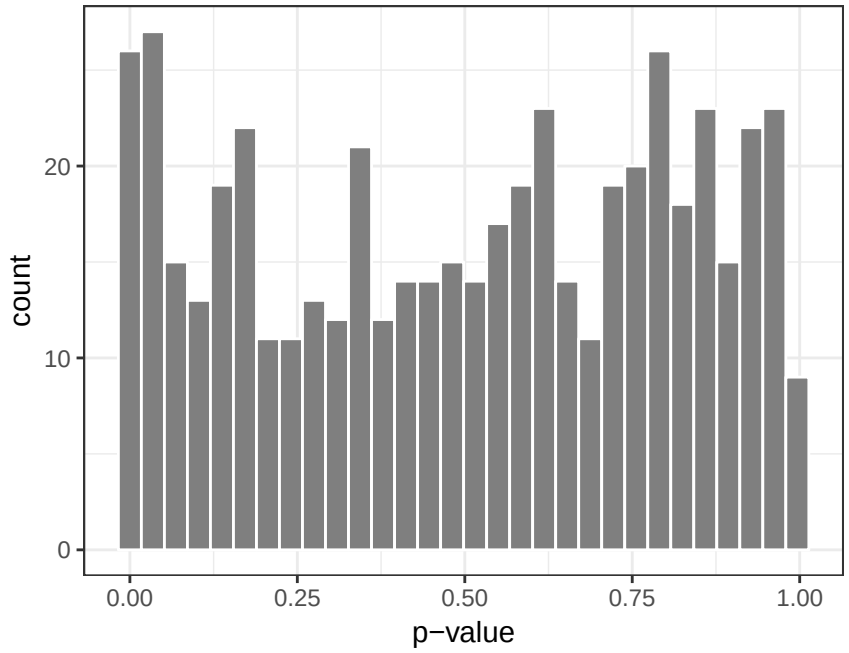

Figure 26: Histogram of p-values for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, scale normalization using top 10 IgG peptides)

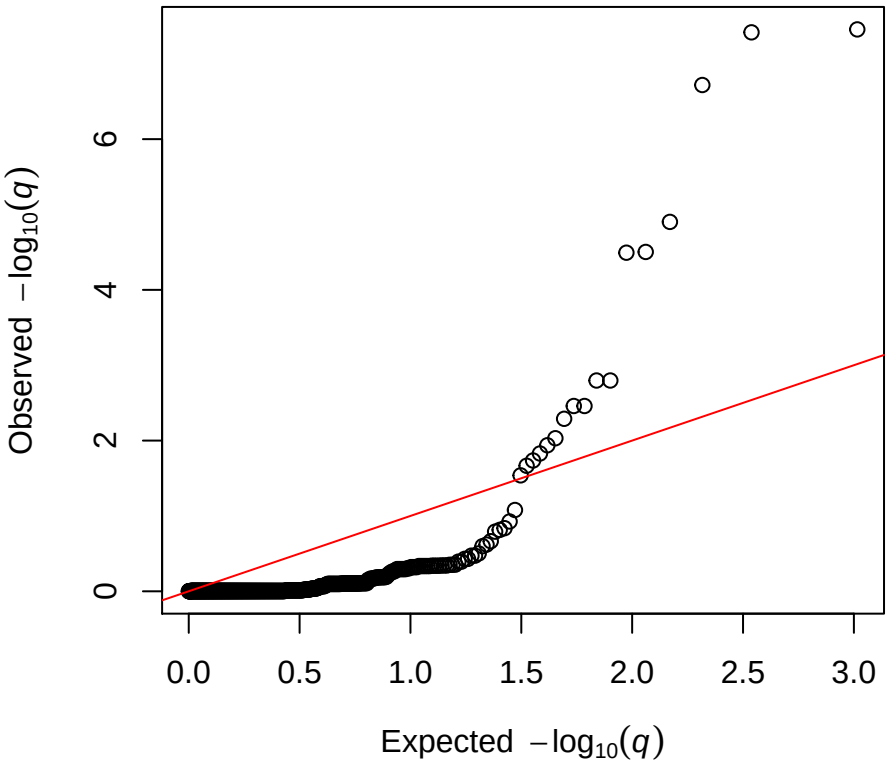

Figure 27: QQ plot of the adjusted p-values for the ER 45min vs ER 0min comparison (excluding peptides with missing intensities, scale normalization using top 10 IgG peptides)

0.6.6 ER 45min vs ER 0min, KNN nearest neighbour averaging imputation of missing values, quantile normalization

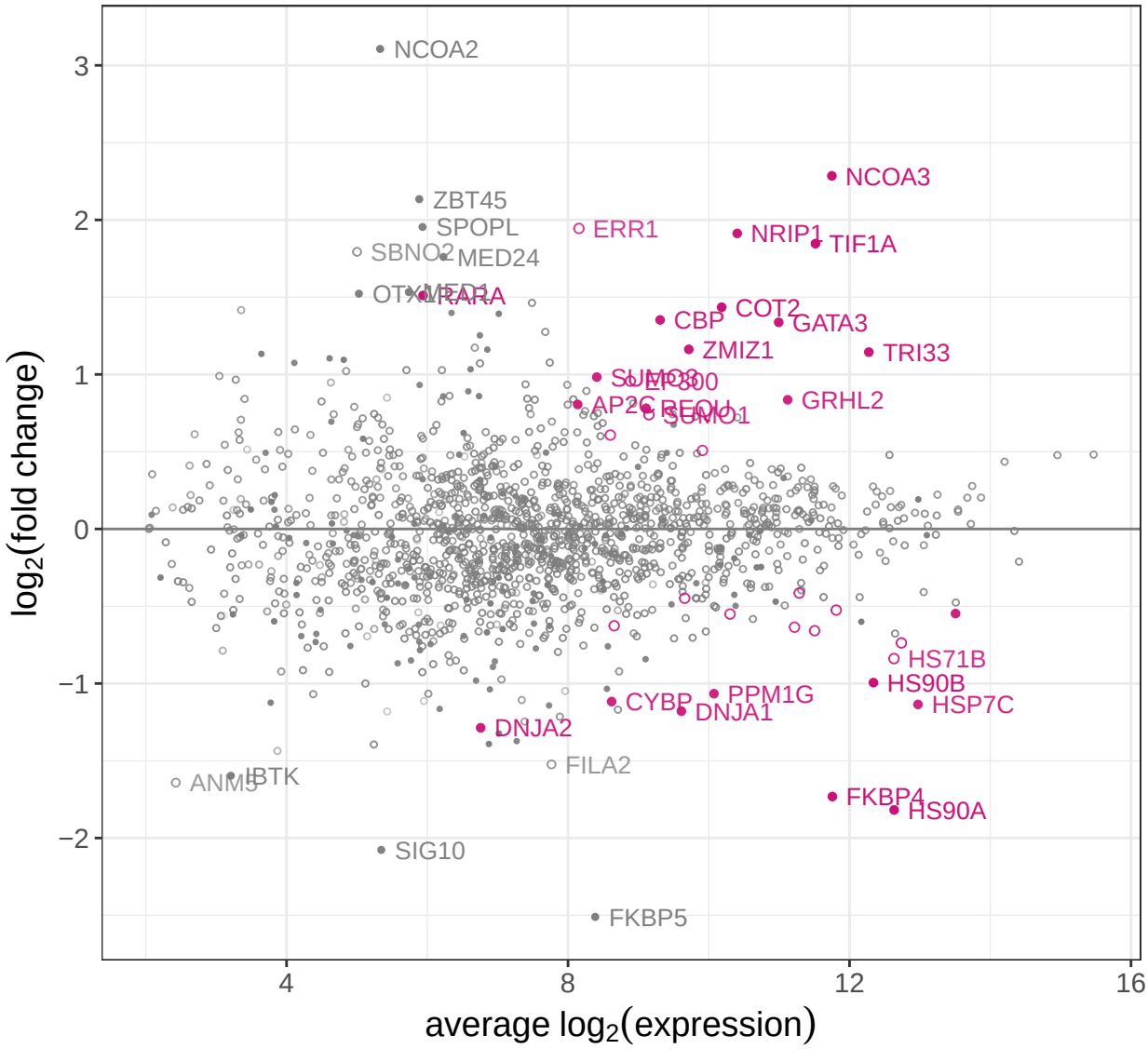

Figure 28: MA plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (KNN nearest neighbour averaging imputation of missing values, quantile normalization). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

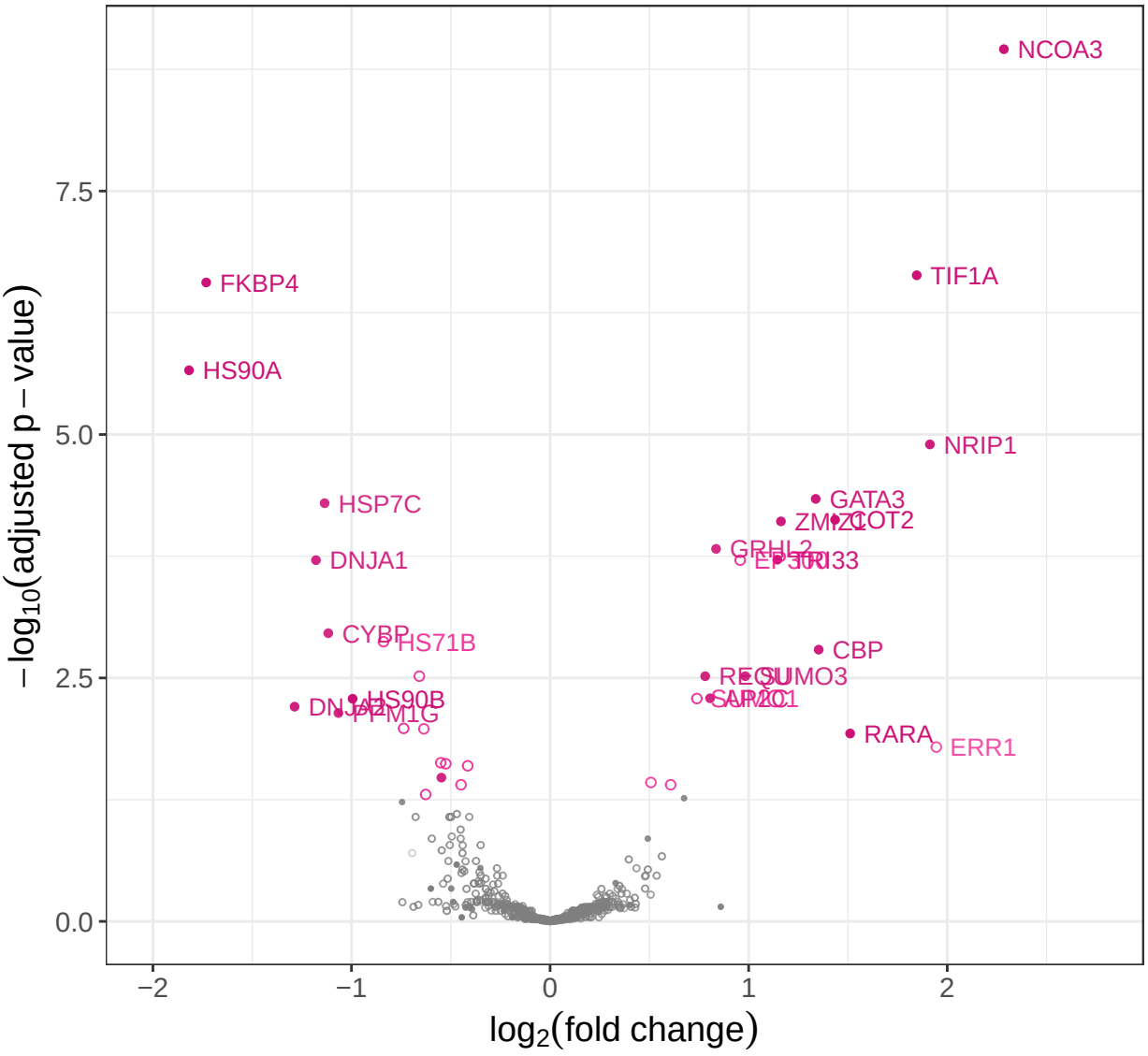

Figure 29: Volcano plot of the average intensity against $\log_2$ fold change for the ER 45min vs ER 0min comparison (KNN nearest neighbour averaging imputation of missing values, quantile normalization). Top ranking differentially-expressed proteins with false discovery rate below 0.05 are highlighted in pink. Open circles indicate that the protein is non-specific from the IgG control comparison.

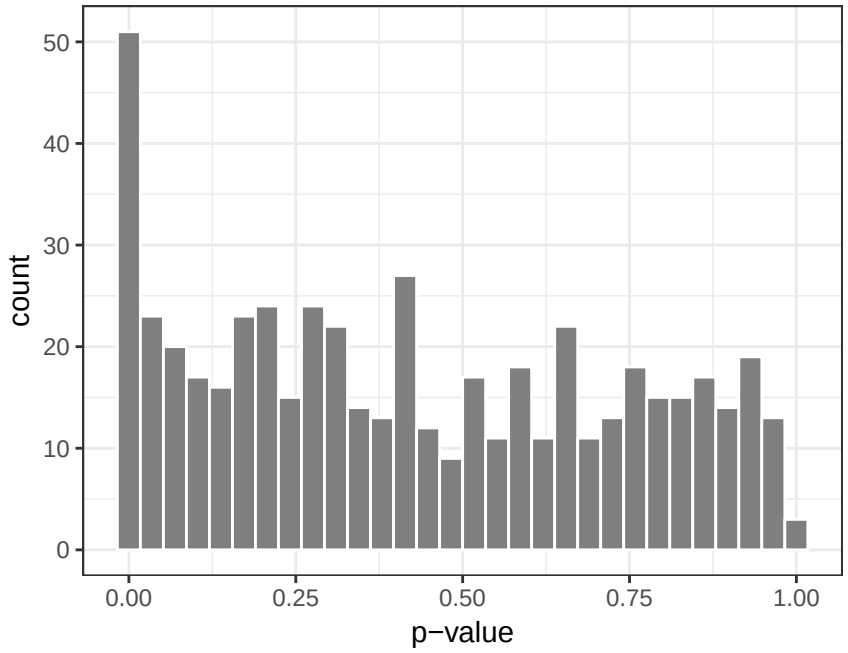

Figure 30: Histogram of p-values for the ER 45min vs ER 0min comparison (KNN nearest neighbour averaging imputation of missing values, quantile normalization)

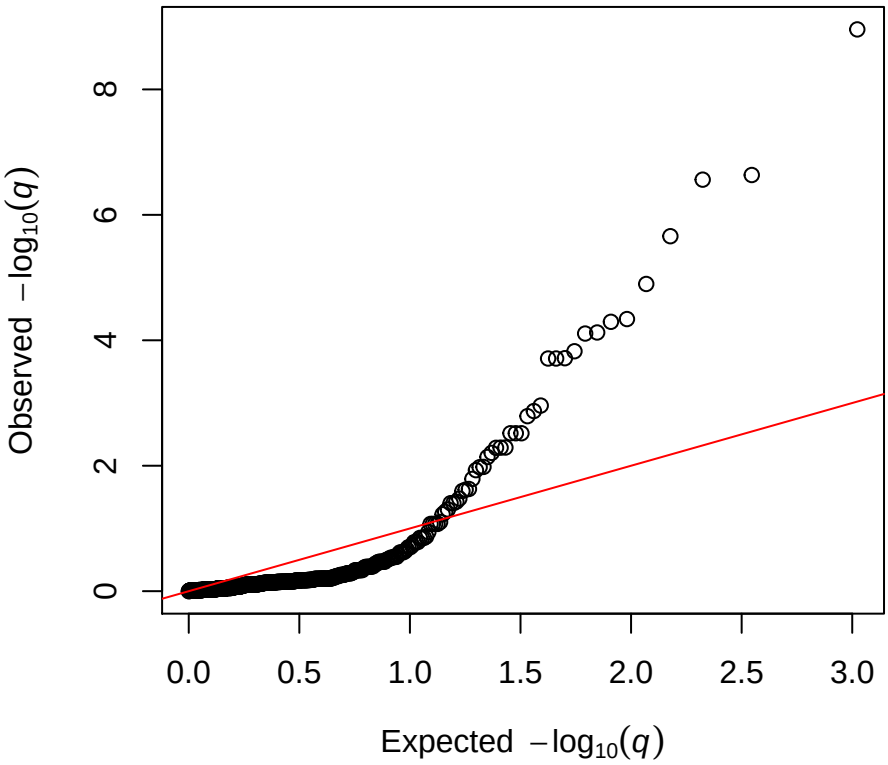

Figure 31: QQ plot of the adjusted p-values for the ER 45min vs ER 0min comparison (KNN nearest neighbour averaging imputation of missing values, quantile normalization)

0.7 Comparison of differential binding analysis approaches

In this section, comparisons of the differential binding results obtained using differing methods for handling missing values and varying normalization techniques are presented. The B-statistic (log odds) is used to rank the proteins from most significantly differentially expressed to least significant. Scatter plots of both the B-statistic generated for the two methods being compared and the rank are presented. The protein with the highest B-statistic, i.e. the most significantly differentially expressed protein, is given a rank of 1.

Differing imputation methods have relatively little effect on the results for most proteins. This is largely because only a small proportion of the proteins have peptide-level measurements containing missing values. There are a few proteins that are only represented in the differential binding results when imputation is carried out, for which all peptide-level measurements contain missing values. These proteins are not represented in the comparions with the analysis in which measurements with missing values are removed.

Different normalization approaches have a more substantial impact on the differential binding analysis.

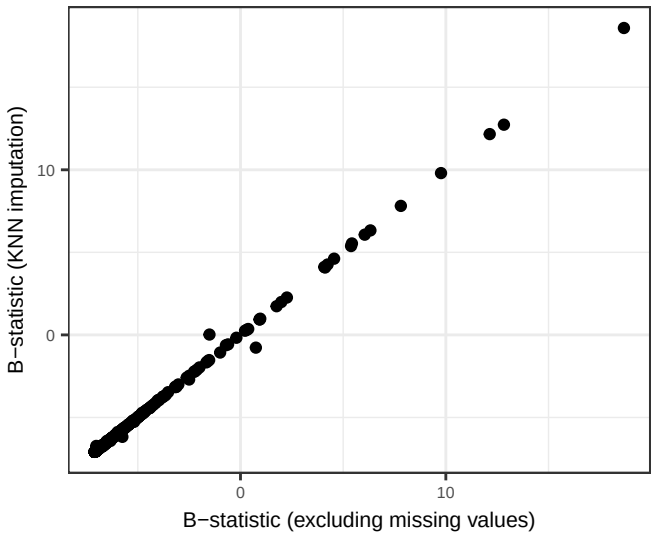

Figure 32: Plot of B-statistic (log odds) for differential expression of proteins for the ER 45min vs ER 0min contrast, comparing results following exclusion of missing values and KNN imputation.

0.7.1 Removal of missing values vs KNN imputation (quantile normalization in both)

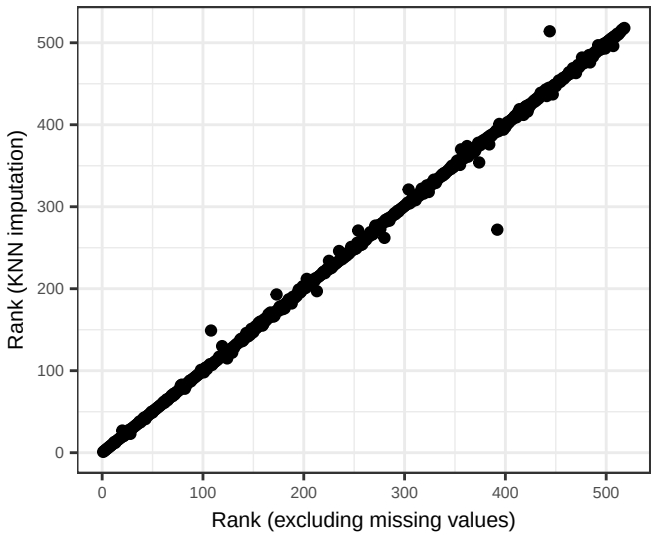

Figure 33: Plot of the rank of differentially-expressed proteins for the ER 45min vs ER 0min contrast, comparing results following exclusion of missing values and KNN imputation.

0.7.2 Removal of missing values vs lowest value imputation (quantile normalization in both)

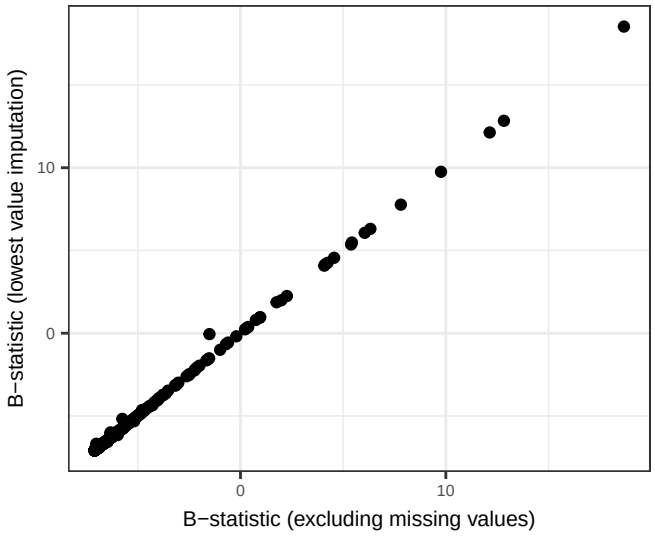

Figure 34: Plot of B-statistic (log odds) for differential expression of proteins for the ER 45min vs ER 0min contrast, comparing results following exclusion of missing values and lowest value imputation.

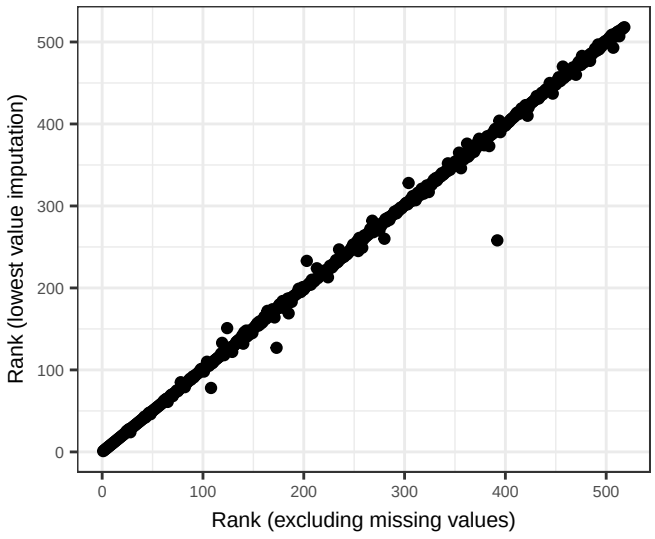

Figure 35: Plot of the rank of differentially-expressed proteins for the ER 45min vs ER 0min contrast, comparing results following exclusion of missing values and lowest value imputation.

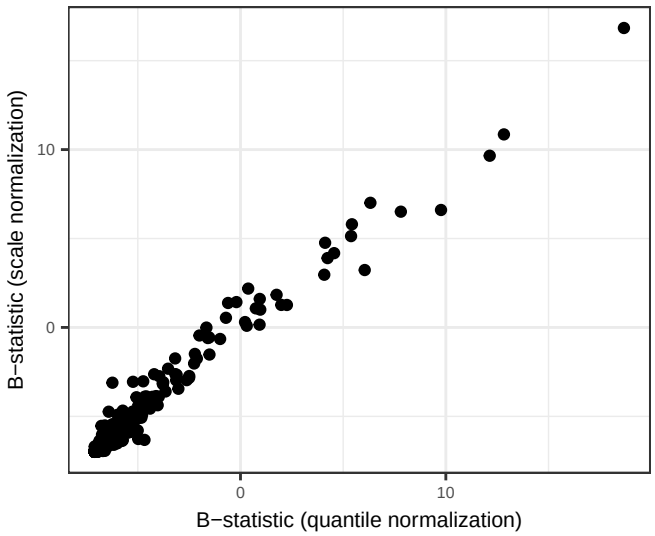

Figure 36: Plot of B-statistic (log odds) for differential expression of proteins for the ER 45min vs ER 0min contrast, comparing results following quantile and scale normalization.

0.7.3 Quantile vs Scale normalization (missing values excluded in both)

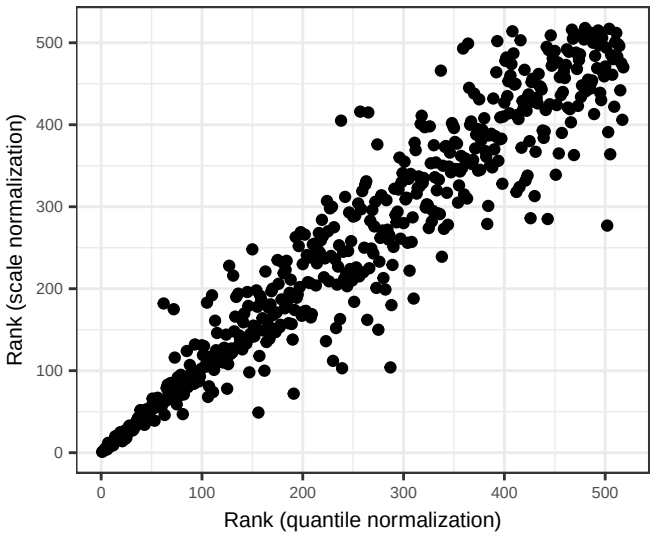

Figure 37: Plot of the rank of differentially-expressed proteins for the ER 45min vs ER 0min contrast, comparing results following quantile and scale normalization.

0.7.4 Scale normalization based on all peptides vs peptides with highest IgG intensities (missing values excluded in both)

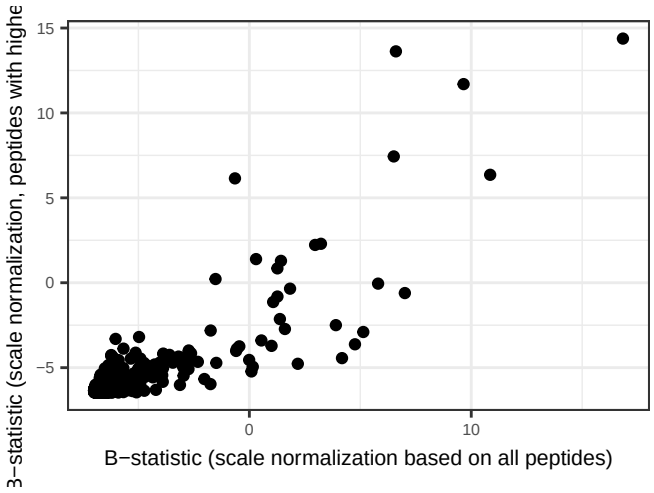

Figure 38: Plot of B-statistic (log odds) for differential expression of proteins for the ER 45min vs ER 0min contrast, comparing results following scale normalization based on all peptides and those with the highest IgG intensities.

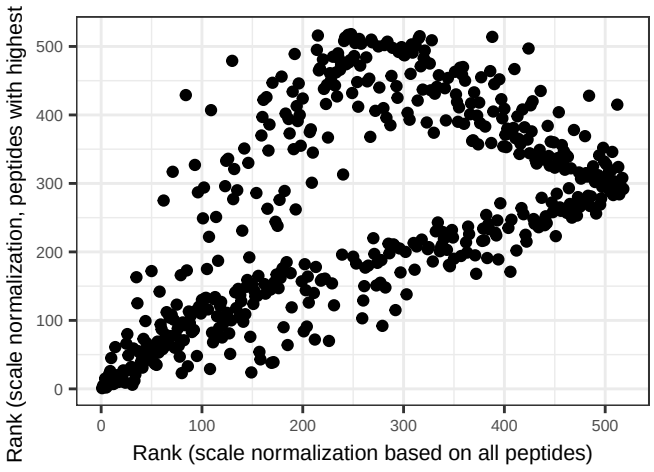

Figure 39: Plot of the rank of differentially-expressed proteins for the ER 45min vs ER 0min contrast, comparing results following scale normalization based on all peptides and those with the highest IgG intensities.

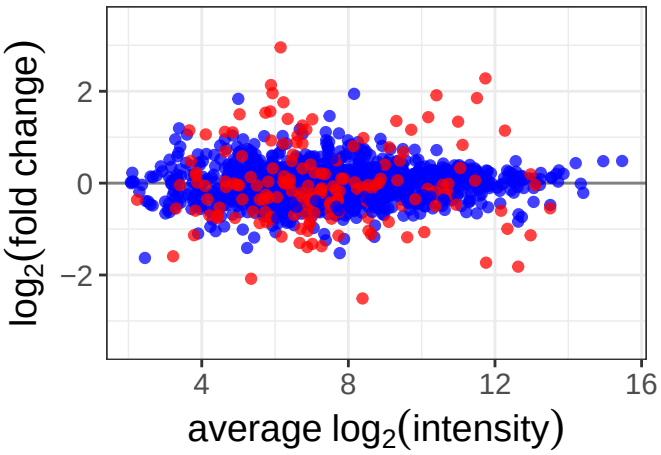

Figure 40: MA plot showing the distribution of non-specific binding. Non-specific binding was identified as proteins that showed less than a log2 enrichment over the IgG channel at either 0 or 45 minutes. The majority of non-specific binding (blue) is found to show less change between time points whereas the specific interactions (red) shows a broader distribution.

0.8 Comparisoin of variance between specfic and non-specific binding

```
##
##  F test to compare two variances
##
## data:  tmp$logFC[tmp$group == 1] and tmp$logFC[tmp$group == 0]
## F = 0.18045, num df = 1373, denom df = 166, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is less than 1
## 95 percent confidence interval:
##  0.0000000 0.2167525
## sample estimates:
## ratio of variances
##          0.1804544
```

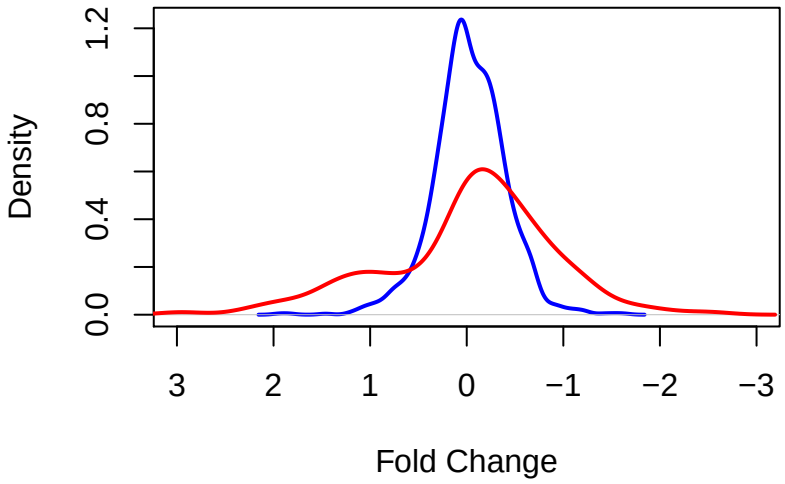

Figure 41: Density plot right showing the distribution of non-specific binding. Non-specific binding was identified as proteins that showed less than a log2 enrichment over the IgG channel at either 0 or 45 minutes. The majority of non-specific binding (blue) is found to show less change between time points whereas the specific interactions (red) shows a broader distribution.

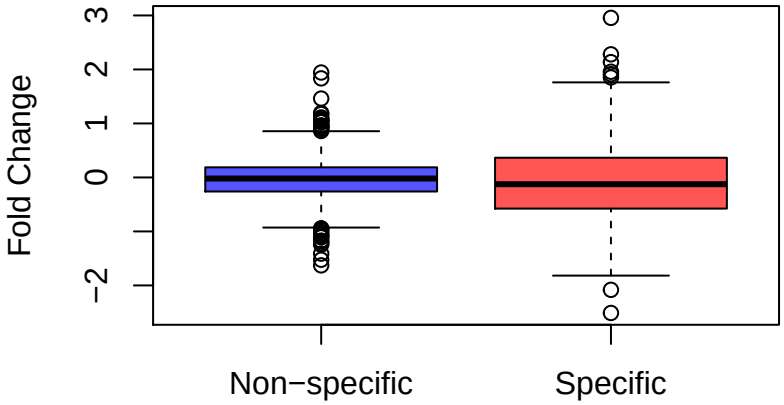

Figure 42: Box plot showing the variance of non-specific vs non-specific binding. Non-specific binding was identified as proteins that showed less than a log2 enrichment over the IgG channel at either 0 or 45 minutes. The the variance of the log fold changes for the non-specific binding interactions was significantly less than that of the specific interactions (F-test, one-sided, p-value $< 2.2e^{*}$-16).

0.9 Differential binding results tables

The following tables contain the top ranking differentially expressed proteins for each comparison. Included are all proteins that reach a statistical signficance of 0.05 in terms of the adjusted p-value and those with an absolute $\log_2$ fold change of 1 or above.

The IgG column gives the larger of the $\log_2$ fold changes for the two groups against the IgG control and an asterisk indicates specific binding where this $\log_2$ fold change is above a threshold of 1. N is the number of replicates in which the protein was observed.

In all cases, peptide intensities were quantile normalized and measurements with missing values were removed prior to summarization at the protein level.

0.9.1 ER 45min vs ER 0min

Protein	Gene	N	log2FC	Avg Expr	p-value	B	IgG	
Q15596	NCOA2	1	2.96	6.15			2.12	*
Q13451	FKBP5	2	-2.51	8.39			2.55	*
Q9Y6Q9	NCOA3	3	2.28	11.74	1e-09	18.67	1.90	*
Q96K62	ZBT45	1	2.14	5.89			2.58	*
Q96LC7	SIG10	1	-2.08	5.35			1.79	*
Q6IQ16	SPOPL	1	1.96	5.93			1.59	*
P11474	ERR1	3	1.94	8.16	0.015	-0.99	-0.14	
P48552	NRIP1	3	1.91	10.40	1.2e-05	7.81	1.56	*
O15164	TIF1A	3	1.85	11.51	2.1e-07	12.83	2.15	*
Q9Y2G9	SBNO2	1	1.83	5.00			0.09	
P07900	HS90A	3	-1.82	12.63	2.2e-06	9.77	2.23	*
O75448	MED24	1	1.76	6.23			1.66	*
Q02790	FKBP4	3	-1.73	11.76	2.8e-07	12.14	1.89	*
O14744	ANM5	1	-1.63	2.45			-0.44	
Q9P2D0	IBTK	1	-1.59	3.22			1.84	*
P10276	RARA	3	1.56	5.87	0.0034	0.75	2.75	*
Q15648	MED1	1	1.53	5.74			1.97	*
Q5D862	FILA2	1	-1.52	7.77			-0.29	
P32242	OTX1	2	1.50	5.03			1.88	*
Q2M1P5	KIF7	1	1.46	7.49			0.71	
P24468	COT2	3	1.44	10.18	7.6e-05	5.43	2.07	*
P34932	HSP74	1	-1.41	5.24			0.54	
O14770	MEIS2	1	1.40	6.35			1.08	*
P61960	UFM1	1	-1.39	6.88			1.92	*
Q8N2W9	PIAS4	2	1.39	7.02			1.68	*
Q86UV5	UBP48	1	-1.37	7.27			1.36	*
Q92793	CBP	3	1.35	9.31	0.0015	1.76	1.48	*
P23771	GATA3	3	1.34	10.99	4.5e-05	6.33	1.61	*
Q9P2D7	DYH1	1	-1.32	7.02			2.63	*
O60884	DNJA2	3	-1.30	6.67	0.023	-1.51	1.76	*
Q9UBW7	ZMYM2	1	1.25	6.75			1.46	*
Q01546	K22O	2	-1.25	7.38			-1.21	
O60809	PRA10	1	-1.22	7.88			-0.24	
O15117	FYB	1	1.19	3.38			-1.88	
P53990	IST1	1	-1.18	5.43			-4.76	
P31689	DNJA1	3	-1.18	9.61	0.00019	4.08	1.20	*
O00712	NFIB	1	1.17	6.67			-0.34	
Q8N283	ANR35	1	-1.17	8.71			-0.15	
Q9UL15	BAG5	1	-1.16	6.18			1.11	*
Q6PHW0	IYD1	1	1.16	6.85			1.52	*
Q9ULJ6	ZMIZ1	3	1.16	9.72	7.6e-05	5.38	1.61	*
P15408	FOSL2	1	1.15	3.66			1.69	*
Q9UNE7	CHIP	2	-1.14	7.73			1.05	*
Q9UPN9	TRI33	3	1.14	12.27	0.00019	4.24	1.73	*
P11142	HSP7C	3	-1.14	12.97	5e-05	6.05	1.00	*
Q9NWS6	F118A	1	-1.13	3.79			1.60	*
Q9HB71	CYBP	3	-1.12	8.62	0.0011	2.26	1.02	*
Q8TCU4	ALMS1	1	1.12	4.63			1.44	*
O14686	KMT2D	1	1.11	7.38			0.39	

Continued on next page

Protein	Gene	N	log2FC	Avg Expr	p-value	B	IgG	
Q7Z794	K2C1B	1	-1.11	7.34			-0.84	
Q6KC79	NIPBL	1	1.11	4.84			2.63	*
Q68E01	INT3	2	-1.10	3.90			-0.67	
Q8NFD5	ARI1B	1	1.08	7.74			0.56	
Q9Y6X2	PIAS3	2	1.07	6.75			0.71	
Q92624	APBP2	1	-1.07	6.01			0.05	
O15355	PPM1G	3	-1.07	10.07	0.0072	-0.20	1.09	*
Q8NH53	O52N1	1	1.06	4.11			1.28	*
Q13492	PICAL	1	1.06	3.60			0.12	
Q4L235	ACSF4	1	-1.05	4.38			-0.13	
O00170	AIP	2	-1.04	6.89			1.13	*
P14625	ENPL	2	1.04	4.85			-0.35	
O60244	MED14	1	1.04	6.62			1.56	*
Q15185	TEBP	1	-1.03	8.55			1.29	*
Q6P2C8	MED27	1	1.03	6.21			-0.01	
Q5VTD9	GFI1B	1	1.02	5.70			0.42	
O14929	HAT1	2	-1.01	5.12			0.70	
P08238	HS90B	3	-0.99	12.34	0.005	0.22	2.04	*
P55854	SUMO3	3	0.98	8.41	0.003	0.96	1.13	*
Q09472	EP300	3	0.96	8.88	0.00019	4.12	0.80	
P0DMV9	HS71B	3	-0.84	12.63	0.0013	1.99	0.65	
Q6ISB3	GRHL2	3	0.83	11.12	0.00015	4.56	1.04	*
Q92754	AP2C	3	0.81	8.14	0.0046	0.38	1.15	*
Q92785	REQU	3	0.78	9.11	0.003	0.93	1.16	*
P52701	MSH6	3	-0.76	6.05	0.048	-2.50	0.97	
P63165	SUMO1	3	0.74	9.15	0.0048	0.31	0.72	
P12956	XRCC6	3	-0.74	12.73	0.01	-0.60	0.90	
P31948	STIP1	3	-0.66	11.50	0.003	0.94	0.59	
P78527	PRKDC	3	-0.63	11.22	0.011	-0.71	0.36	
Q9HAV4	XPO5	3	-0.63	8.66	0.048	-2.50	0.72	
O14497	ARI1A	3	0.61	8.60	0.04	-2.26	0.29	
Q9Y383	LC7L2	3	-0.55	10.30	0.023	-1.53	0.51	
Q13263	TIF1B	3	-0.55	13.51	0.033	-2.00	1.17	*
P09874	PARP1	3	-0.53	11.81	0.023	-1.59	0.63	
Q92925	SMRD2	3	0.51	9.91	0.037	-2.12	0.23	
P25685	DNJB1	3	-0.45	9.66	0.039	-2.22	0.60	
Q99873	ANM1	3	-0.42	11.29	0.024	-1.66	0.65	

Table 4: Top ranking differentially expressed proteins from the ER 45min vs ER 0min comparison, sorted by log2 fold change.

0.9.2 ER 90min vs ER 0min

Protein	Gene	N	log2FC	Avg Expr	p-value	B	IgG	
Q15596	NCOA2	1	2.82	6.15			1.98	*
Q13451	FKBP5	2	-2.39	8.39			2.55	*
Q96LC7	SIG10	1	-2.25	5.35			1.79	*
Q9Y6Q9	NCOA3	3	2.22	11.74	1.5e-09	18.26	1.84	*
Q96K62	ZBT45	1	2.19	5.89			2.63	*
Q6IQ16	SPOPL	1	2.16	5.93			1.79	*
Q9Y2G9	SBNO2	1	2.15	5.00			0.41	
P48552	NRIP1	3	2.03	10.40	5e-06	8.73	1.68	*
Q6UX73	CP089	1	2.00	8.00			1.27	*
O75448	MED24	1	1.88	6.23			1.77	*
P20393	NR1D1	1	1.86	8.26			1.17	*
Q68CL5	TPGS2	2	-1.85	3.12			-6.76	
O15164	TIF1A	3	1.84	11.51	2.4e-07	12.70	2.14	*
P07900	HS90A	3	-1.82	12.63	2.3e-06	9.75	2.23	*
P11474	ERR1	3	1.81	8.16	0.024	-1.59	-0.28	
O14770	MEIS2	1	1.76	6.35			1.44	*
Q2M1P5	KIF7	1	1.74	7.49			0.99	
Q02790	FKBP4	3	-1.72	11.76	3.2e-07	11.99	1.89	*
Q15648	MED1	1	1.66	5.74			2.09	*
P10276	RARA	3	1.58	5.87	0.0033	0.87	2.77	*
O60229	KALRN	1	1.57	6.32			0.86	
P15408	FOSL2	1	1.57	3.66			2.11	*
Q9UBW7	ZMYM2	1	1.55	6.75			1.76	*
P81605	DCD	1	1.47	4.63			-1.79	
P24468	COT2	3	1.45	10.18	7.4e-05	5.53	2.08	*
Q86UV5	UBP48	1	-1.44	7.27			1.36	*
P32242	OTX1	2	1.43	5.03			1.81	*
Q92624	APBP2	1	-1.42	6.01			0.05	
Q6KC79	NIPBL	1	1.42	4.84			2.94	*
Q92793	CBP	3	1.41	9.31	0.0011	2.27	1.54	*
P23771	GATA3	3	1.38	10.99	2.8e-05	6.79	1.65	*
Q9P2D7	DYH1	1	-1.37	7.02			2.63	*
O15117	FYB	1	1.36	3.38			-1.71	
P34932	HSP74	1	-1.29	5.24			0.54	
O60884	DNJA2	3	-1.28	6.67	0.024	-1.61	1.76	*
O00712	NFIB	1	1.28	6.67			-0.24	
O76094	SRP72	1	1.25	3.41			0.35	
Q9NWS6	F118A	1	-1.25	3.79			1.60	*
O60934	NBN	1	-1.24	4.24			0.84	
Q13492	PICAL	1	1.24	3.60			0.30	
Q8WX92	NELFB	2	-1.23	6.70			1.26	*
Q8N2W9	PIAS4	2	1.21	7.02			1.50	*
Q8N283	ANR35	1	-1.20	8.71			-0.15	
Q6KB66	K2C80	2	1.19	11.02			0.73	
Q9UNE7	CHIP	2	-1.17	7.73			1.05	*
O60809	PRA10	1	-1.17	7.88			-0.24	
P31689	DNJA1	3	-1.16	9.61	0.00026	3.87	1.20	*
Q9Y6X2	PIAS3	2	1.15	6.75			0.79	
Q4L235	ACSF4	1	-1.14	4.38			-0.13	

Continued on next page

Protein	Gene	N	log2FC	Avg Expr	p-value	B	IgG	
Q15126	PMVK	1	-1.14	5.45			0.49	
Q9Y4C1	KDM3A	1	-1.11	5.93			0.91	
Q13158	FADD	1	-1.11	2.11			-0.90	
Q9ULL5	PRR12	1	1.11	3.33			-0.10	
P11142	HSP7C	3	-1.11	12.97	7.3e-05	5.69	1.00	*
Q9HB71	CYBP	3	-1.10	8.62	0.0013	2.04	1.02	*
Q6P2C8	MED27	1	1.10	6.21			0.06	
Q9UL15	BAG5	1	-1.09	6.18			1.11	*
Q6PHW0	IYD1	1	1.09	6.85			1.45	*
P07205	PGK2	1	-1.08	6.73			1.03	*
Q96QK1	VPS35	1	1.08	4.72			0.59	
O15355	PPM1G	3	-1.07	10.07	0.0065	-0.12	1.09	*
Q9UPN9	TRI33	3	1.05	12.27	0.00045	3.22	1.64	*
P61960	UFM1	1	-1.05	6.88			1.92	*
Q9ULJ6	ZMIZ1	3	1.04	9.72	0.00025	3.98	1.49	*
O43866	CD5L	2	-1.04	6.17			-2.51	
O14686	KMT2D	1	1.04	7.38			0.32	
Q5D862	FILA2	1	-1.03	7.77			-0.29	
Q09472	EP300	3	1.03	8.88	0.00011	5.00	0.86	
Q8NFD5	ARI1B	1	1.02	7.74			0.50	
P08238	HS90B	3	-1.00	12.34	0.0047	0.32	2.04	*
P55854	SUMO3	3	0.96	8.41	0.0034	0.71	1.11	*
P52701	MSH6	3	-0.93	6.05	0.012	-0.78	0.97	
Q6ISB3	GRHL2	3	0.85	11.12	0.00012	4.82	1.05	*
Q92754	AP2C	3	0.84	8.14	0.0033	0.88	1.19	*
P63165	SUMO1	3	0.79	9.15	0.0033	0.98	0.77	
Q92785	REQU	3	0.77	9.11	0.0034	0.81	1.15	*
P0DMV9	HS71B	3	-0.75	12.63	0.0034	0.76	0.65	
P61956	SUMO2	3	0.73	9.50	0.033	-2.05	1.10	*
P12956	XRCC6	3	-0.72	12.73	0.012	-0.75	0.90	
O14497	ARI1A	3	0.72	8.60	0.012	-0.81	0.41	
P78527	PRKDC	3	-0.68	11.22	0.0061	-0.03	0.36	
Q9HAV4	XPO5	3	-0.65	8.66	0.037	-2.19	0.72	
P31948	STIP1	3	-0.62	11.50	0.0047	0.29	0.59	
Q04724	TLE1	3	0.61	9.40	0.041	-2.32	1.12	*
Q92925	SMRD2	3	0.58	9.91	0.014	-1.00	0.31	
P55060	XPO2	3	-0.56	10.10	0.044	-2.43	0.86	
Q9Y383	LC7L2	3	-0.51	10.30	0.037	-2.18	0.51	
P09874	PARP1	3	-0.50	11.81	0.033	-2.01	0.63	
Q99873	ANM1	3	-0.40	11.29	0.033	-1.99	0.65	

Table 5: Top ranking differentially expressed proteins from the ER 90min vs ER 0min comparison, sorted by log2 fold change.

0.9.3 ER 90min vs ER 45min

Protein	Gene	N	log2FC	Avg Expr	p-value	B	IgG	
Q13158	FADD	1	-1.34	2.11			-0.67	
Q68CL5	TPGS2	2	-1.21	3.12			-7.40	
Q6UX73	CP089	1	1.20	8.00			1.27	*
P33176	KINH	1	1.09	4.64			0.78	
O14744	ANM5	1	1.06	2.45			-1.01	
Q8WXG9	GPR98	1	1.02	6.85			2.30	*

Table 6: Top ranking differentially expressed proteins from the ER 90min vs ER 45min comparison, sorted by log2 fold change.

0.9.4 FOXA1 45min vs FOXA1 0min

Protein	Gene	N	log2FC	Avg Expr	p-value	B	IgG	
Q8NH53	O52N1	1	4.29	4.11			4.90	*
Q96QK1	VPS35	1	3.72	4.72			4.08	*
A6NNA2	SRRM3	1	3.69	7.45			4.27	*
P57773	CXA9	1	3.12	8.22			2.52	*
A5PLK6	RGSL	1	-3.03	6.46			0.16	
P42356	PI4KA	1	-2.64	4.49			-1.42	
Q9NVG8	TBC13	1	-2.39	6.80			0.41	
O94844	RHBT1	1	-2.35	5.71			0.20	
P57678	GEMI4	1	-2.32	7.40			-0.16	
Q86UP2	KTN1	1	2.23	6.69			2.02	*
P08729	K2C7	1	-2.19	3.89			0.04	
Q9NW13	RBM28	1	2.06	6.68			2.13	*
O94992	HEXI1	2	-1.70	4.22			0.53	
Q86WI1	PKHL1	1	-1.69	4.09			-0.70	
Q9H2Y7	ZN106	1	1.66	7.50			-1.73	
Q01813	PFKAP	1	-1.62	3.27			-0.69	
Q8IV04	TB10C	1	-1.61	6.35			0.15	
Q9H3P2	NELFA	1	-1.54	3.92			0.55	
Q9NWS6	F118A	1	-1.49	3.79			0.73	
Q9Y3E5	PTH2	1	1.49	3.11			0.89	
Q8TDN6	BRX1	1	1.46	7.29			1.55	*
Q8NF37	PCAT1	1	-1.41	7.34			-0.42	
Q9H0A0	NAT10	1	1.38	2.30			1.61	*
Q96KQ4	ASPP1	1	-1.38	3.41			-0.79	
P33176	KINH	1	-1.36	4.64			-0.26	
P20393	NR1D1	1	-1.32	8.26			0.77	
O60610	DIAP1	1	-1.32	4.23			0.08	
Q9BQG0	MBB1A	1	1.30	8.49			1.14	*
Q9HA92	RSAD1	1	-1.30	7.23			0.48	
Q9UHB6	LIMA1	1	1.30	3.12			-0.08	
Q9BXD5	NPL	1	-1.29	4.27			0.16	
Q5VTD9	GFI1B	1	-1.29	5.70			0.24	
Q13492	PICAL	1	1.26	3.60			1.43	*
Q15413	RYR3	1	-1.26	6.19			-5.07	
Q96HY6	DDRGK	1	-1.23	3.95			-0.17	
O00487	PSDE	1	1.23	2.15			0.67	
O15269	SPTC1	1	-1.18	3.96			0.19	
Q9HDC9	APMAP	1	-1.14	7.62			-0.44	
O75179	ANR17	1	1.14	3.36			0.61	
Q8ND56	LS14A	1	1.14	3.28			0.07	
O60229	KALRN	1	-1.13	6.32			0.23	
O95433	AHSA1	1	1.13	3.19			1.67	*
Q5T0W9	FA83B	1	1.13	6.88			0.78	
P00450	CERU	1	1.12	6.76			0.04	
Q96CT7	CC124	1	-1.12	6.21			-0.32	
Q14677	EPN4	1	-1.09	6.32			-1.86	
Q08J23	NSUN2	2	-1.09	4.80			0.50	
Q14966	ZN638	1	1.08	7.79			1.09	*
Q9H9B1	EHMT1	1	1.07	3.76			1.84	*

Continued on next page

Protein	Gene	N	log2FC	Avg Expr	p-value	B	IgG
A5YKK6	CNOT1	1	-1.06	5.89			-0.60
Q9HCH5	SYTL2	1	-1.04	6.63			0.77
P61966	AP1S1	1	1.01	3.79			0.46
Q06787	FMR1	1	-1.01	7.26			-0.15
O15117	FYB	1	1.01	3.38			0.64
P29083	T2EA	1	1.01	5.86			-3.44
Q8IVT2	MISP	1	1.01	5.52			0.40

Table 7: Top ranking differentially expressed proteins from the FOXA1 45min vs FOXA1 0min comparison, sorted by log2 fold change.

0.9.5 FOXA1 90min vs FOXA1 0min

Protein	Gene	N	log2FC	Avg Expr	p-value	B	IgG
A5PLK6	RGSL	1	-1.53	6.46			0.16
Q5T0F9	C2D1B	1	-1.28	3.34			0.00
P04433	KV309	1	-1.24	8.54			0.36
P55036	PSMD4	1	-1.22	4.17			0.15
O95433	AHSA1	1	-1.13	3.19			0.54
O00487	PSDE	1	-1.08	2.15			-0.56

Table 8: Top ranking differentially expressed proteins from the FOXA1 90min vs FOXA1 0min comparison, sorted by log2 fold change.

0.9.6 FOXA1 90min vs FOXA1 45min

Protein	Gene	N	log2FC	Avg Expr	p-value	B	IgG	
Q8NH53	O52N1	1	-4.27	4.11			4.90	*
Q96QK1	VPS35	1	-4.20	4.72			4.08	*
P42356	PI4KA	1	3.49	4.49			-0.57	
P57773	CXA9	1	-2.97	8.22			2.52	*
A6NNA2	SRRM3	1	-2.88	7.45			4.27	*
Q86UP2	KTN1	1	-2.69	6.69			2.02	*
Q9NVG8	TBC13	1	2.67	6.80			0.69	
O94844	RHBT1	1	2.51	5.71			0.36	
O00487	PSDE	1	-2.31	2.15			0.67	
O95433	AHSA1	1	-2.26	3.19			1.67	*
Q15413	RYR3	1	2.13	6.19			-4.20	
P08729	K2C7	1	2.04	3.89			-0.11	
Q9H0A0	NAT10	1	-1.98	2.30			1.61	*
O75179	ANR17	1	-1.91	3.36			0.61	
Q9HDC9	APMAP	1	1.89	7.62			0.31	
P57678	GEMI4	1	1.88	7.40			-0.60	
Q8ND56	LS14A	1	-1.81	3.28			0.07	
Q9H3P2	NELFA	1	1.74	3.92			0.76	
O94992	HEXI1	2	1.69	4.22			0.51	
Q8IV04	TB10C	1	1.67	6.35			0.20	
P33176	KINH	1	1.65	4.64			0.03	
P61966	AP1S1	1	-1.62	3.79			0.46	
Q9NWS6	F118A	1	1.51	3.79			0.75	
A5PLK6	RGSL	1	1.50	6.46			-1.37	
Q13492	PICAL	1	-1.46	3.60			1.43	*
Q8TE85	GRHL3	1	1.43	4.85			-0.03	
Q9HCH5	SYTL2	1	1.43	6.63			1.16	*
Q9H2Y7	ZN106	1	-1.40	7.50			-1.73	
Q9Y6E0	STK24	1	-1.39	3.66			0.52	
Q14677	EPN4	1	1.37	6.32			-1.58	
P23677	IP3KA	1	1.35	7.05			-0.42	
Q9NW13	RBM28	1	-1.33	6.68			2.13	*
Q9Y3E5	PTH2	1	-1.32	3.11			0.89	
Q5T0F9	C2D1B	1	-1.32	3.34			0.04	
Q9HA92	RSAD1	1	1.31	7.23			0.49	
P04004	VTNC	1	1.29	6.39			-0.30	
O60229	KALRN	1	1.29	6.32			0.38	
Q8NF37	PCAT1	1	1.28	7.34			-0.55	
Q86WI1	PKHL1	1	1.26	4.09			-1.13	
Q5T1M5	FKB15	1	-1.25	6.17			-0.05	
Q8IUE6	H2A2B	2	1.23	7.40			0.00	
Q9H0S4	DDX47	1	1.18	7.09			-0.14	
Q96KQ4	ASPP1	1	1.16	3.41			-1.01	
Q8IVT2	MISP	1	-1.15	5.52			0.40	
Q9H9B1	EHMT1	1	-1.13	3.76			1.84	*
Q13158	FADD	1	-1.10	2.11			0.37	
P07384	CAN1	1	-1.09	5.85			0.49	
P32242	OTX1	2	-1.09	5.03			2.36	*
P02656	APOC3	3	1.08	6.97	1	-4.11	-0.76	

Continued on next page

Protein	Gene	N	log2FC	Avg Expr	p-value	B	IgG	
P35580	MYH10	1	1.07	7.89			-0.23	
O00442	RTCA	1	-1.06	2.48			1.26	*
P39019	RS19	2	-1.06	8.37			-0.10	
P14770	GPIX	2	1.06	6.28			-0.01	
Q9UHB6	LIMA1	1	-1.03	3.12			-0.08	
O14745	NHRF1	1	-1.02	5.90			0.31	

Table 9: Top ranking differentially expressed proteins from the FOXA1 90min vs FOXA1 45min comparison, sorted by log2 fold change.