

# Antibody Profiling of Prostate Cancer Patients Between Disease Stages and Following Treatment

*Tun Lee Ng and Michael A. Newton*

*July 07, 2020*

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Section I: Antibody Responses between Disease Stages</b>	<b>1</b>
2.1	Preamble . . . . .	1
2.2	Normalization of Fluorescence Data . . . . .	2
2.3	Reproducibility of Replicates . . . . .	3
2.4	Tests on Binary Calls . . . . .	4
2.5	Tests on Fluorescence Levels . . . . .	6
2.6	Pairwise Comparisons . . . . .	9
2.7	Visualization . . . . .	11
2.8	Gene-Set-Analyses . . . . .	13
<b>3</b>	<b>Section II: Antibody Responses over Time after Treatments</b>	<b>27</b>
3.1	Preamble . . . . .	27
3.2	Normalization of Fluorescence Data . . . . .	27
3.3	Tests on Time Effect . . . . .	28
3.4	Visualization . . . . .	30
3.5	Gene-Set-Analysis . . . . .	33
<b>4</b>	<b>Conclusion</b>	<b>35</b>

## 1 Introduction

This supplemental analysis consists of two major sections:

- Section I focuses on characterizing antibody responses to a wide variety of proteins in prostate cancer patients at different stages of the disease.
- Section II focuses on analyzing whether treatments induces different changes in antibody repertoires in individuals over time.

## 2 Section I: Antibody Responses between Disease Stages

### 2.1 Preamble

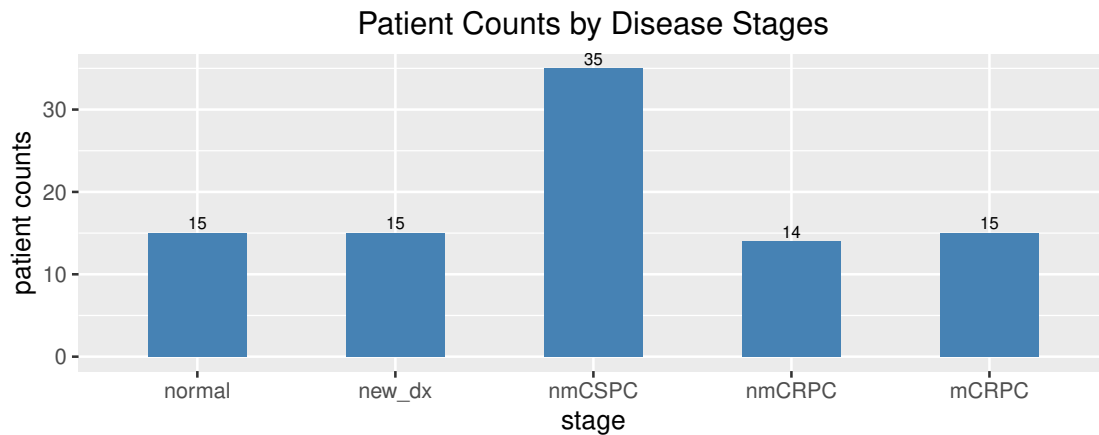
In this section, we consider a study that involved healthy subjects and patients with different stages of prostate cancer

- `new_dx`: newly diagnosed,
- `nmCSPC`: non-metastatic castration-sensitive
- `mCSPC`: metastatic castration-sensitive,
- `nmCRPC`: non-metastatic castration-resistant,
- `mCRPC`: metastatic castration-resistant

Each patient's serum was assayed in a number of replicates, **rep**, which were 1, 2, or 3, for peptide-specific IgG responses using a microarray: 16-mer peptides spanning the amino acid sequences of these 1611 gene products, and overlapping by 12 amino acids, were used to generate the microarray comprising 177,604 peptides. We also considered peptides with fluorescence intensity of at least  $2^{12}$ , and a sliding-window p-value of less than 0.05 (indicating high signal in adjacent peptides), in at least 2 of the 3 technical replicates to be called positive.

We remove patients with **rep** = 1. The criterion for a positive call on a peptide for a patient was that they had to meet the signal (fluorescence level) threshold in at least two of the technical replicates. Since this is not possible for patients without technical replicates, we exclude them for consistency.

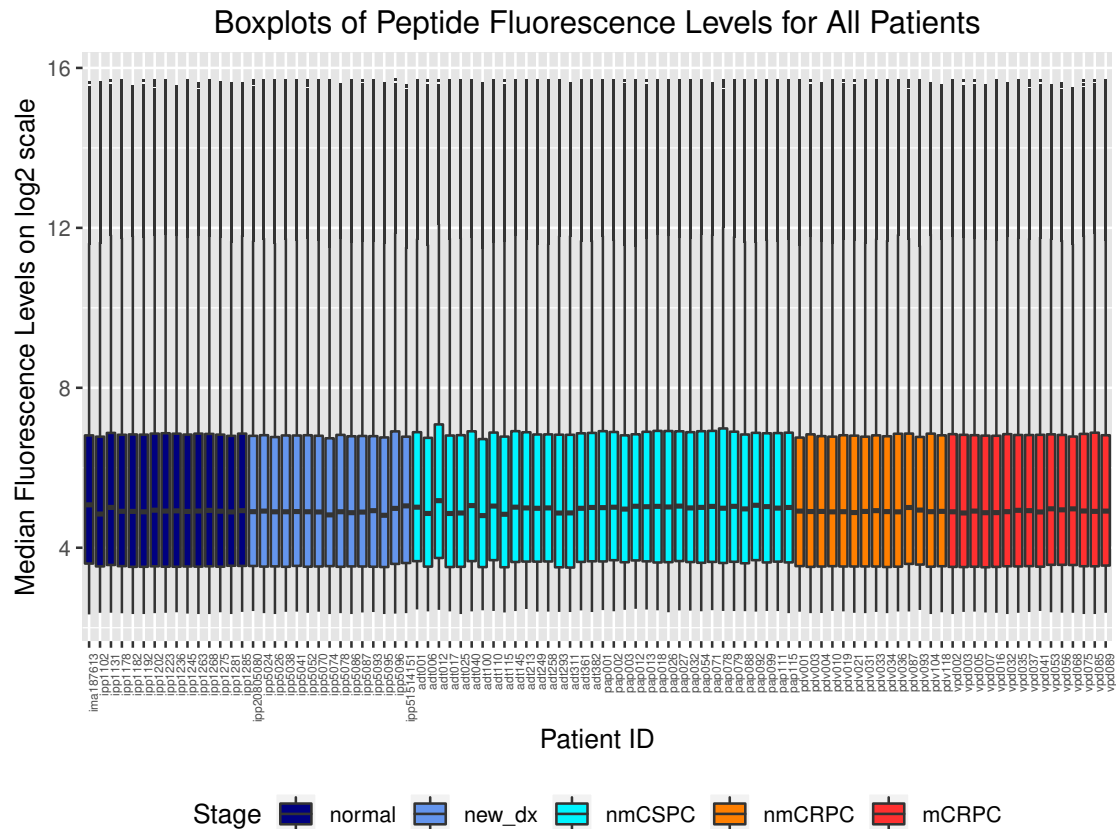
Note that there were 6 patients who were measured at two different stages of prostate cancer. We removed their earlier-stage records, and finally arrive at 94 distinct patients.



We will utilize both binary calls data and fluorescence levels data to **investigate if patients at different stages of prostate cancer exhibit different antibody responses to certain peptide chains or proteins**. We take  $\log_2$  transformation on the fluorescence levels prior to subsequent steps in our analysis.

## 2.2 Normalization of Fluorescence Data

In order to verify normalization of the fluorescence level, we also plot the boxplots of median (across replicates)  $\log_2$  fluorescence level of all peptides for each patient.



It appears that the fluorescence levels of the peptides are normalized accordingly.

### 2.3 Reproducibility of Replicates

We have assessed the issue of replicate reproducibility by looking at (Pearson) correlation coefficients between patients' fluorescence levels. Another approach is to measure how much variation the technical replicates are contributing to the overall variation in the data. Everytime when the fluorescence levels were measured (with replicates) for patient's stage effects, there are two sources of random variation at play, namely

- patient/subject **random effect**: this reflects the biological variation of a patient (as opposed to the **fixed effect** term, which would be the cancer stage effect in this experiment)
- (residual) random error: measuring replicates of a patient is itself a source of technical variation.

Specifically,

$$y_{ijk} = \mu + \beta_i + b_j + \epsilon_{ijk},$$

where

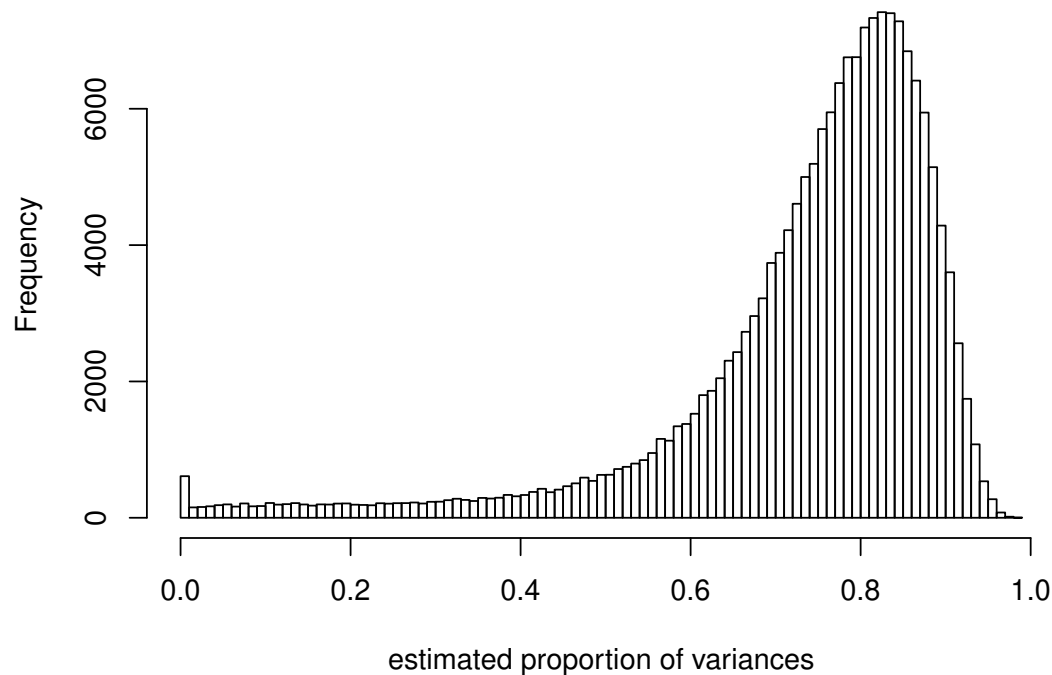
- $y_{ijk}$  denotes the  $\log_2$  fluorescence level of a replicate,
- $\mu$  denotes the grand mean/intercept,
- $\beta_i$  denotes the fixed effect term, ie. cancer stage, with  $i$  indexing the patients' cancer stage,
- $b_j$  denotes the random effect term, ie. individual patient, with  $j$  indexing the patients,
- $\epsilon_{ijk}$  denotes the (residual) random error of the model, with  $k$  indexing the replicates.

This is the linear mixed-effects model, which we deploy using the R package `lme4` [Bates et al., 2015]. The model estimates the two sources of variation:  $\hat{\sigma}_b^2$  (biological variation) and  $\hat{\sigma}_\epsilon^2$  (technical variation). Ideally, biological variation should dominate technical variation since the replicates' variance  $\hat{\sigma}_\epsilon^2$  should be minimal. Hence, we are interested in the estimated proportion of random-effect variance to total variance

$$\frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_\epsilon^2},$$

and we would like to see if this ratio is close to one. For each of the 177,604 peptides, we deploy this mixed-effect model, and plot the histogram of the estimated proportions of variances.

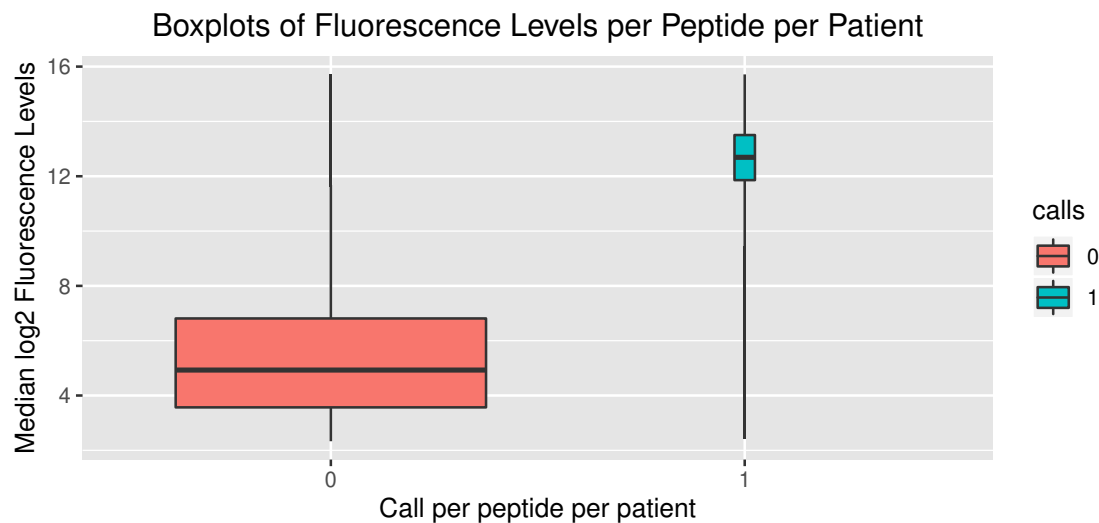
### Histogram of peptide-level proportion of random-effect variance to total variance



As expected, the histogram amasses at values near one, indicating that most of the variation in the ( $\log_2$ ) fluorescence data is attributable to the biological variation of the patients and not the technical replicates themselves, which also suggests reproducibility of the replicates.

#### 2.4 Tests on Binary Calls

The binary calls on a peptide of a patient are conservative – out of 177,604 peptides, only 37919 of them have at least one call among all patients. To verify that positive calls are associated with stronger signals (remember that `call` = 1 if fluorescence levels meet a certain signal threshold in at least two of the replicates), we plot the boxplot of  $\log_2$  fluorescence levels for all peptides across all patients, comparing between those that are associated with positive calls and those with zero-calls. Boxplots are plotted with their width reflecting the sample size in each group (positive or zero call).



For each of these 37919 peptides, we run a logistic regression based on these binary calls of the patients in order **to determine if calls are significantly different among patients of different cancer stages.**

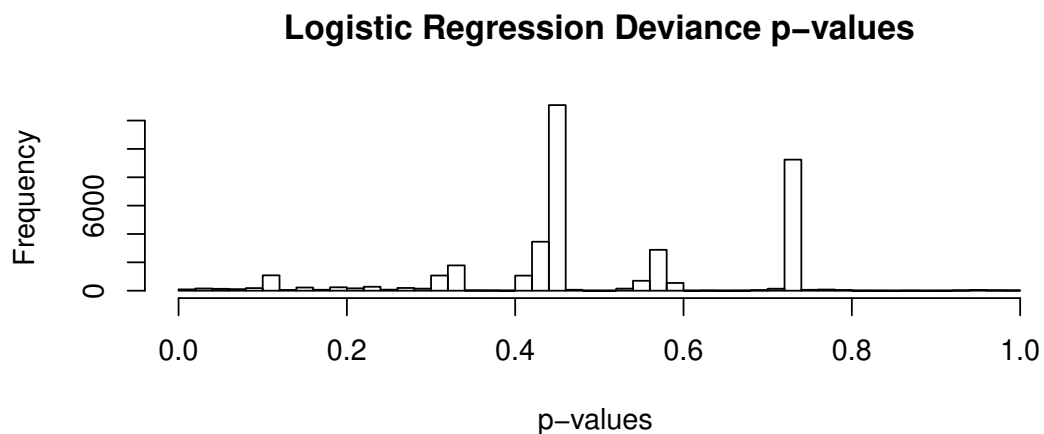
Specifically, for each of these 37919 peptides, we fit the following model:

$$\text{logit}(y_{ij}^{\text{calls}}) = \mu + \beta_i + \epsilon_{ij},$$

where

- $y_{ij}^{\text{calls}}$  denotes the binary call of the peptide of a patient: 1 if the fluorescence levels meet the signal threshold in at least two replicates of the patient, and 0 otherwise,
- $\mu$  denotes the grand mean/intercept,
- $\beta_i$  denotes the cancer stage,
- $\epsilon_{ij}$  denotes the random error of the model, with  $j$  indexing the patients,

and compute the deviance p-values:  $(\text{null\_deviance} - \text{residual\_deviance}) \sim \chi^2$  with 4 degrees of freedom. We plot the histogram of the 37919 p-values.

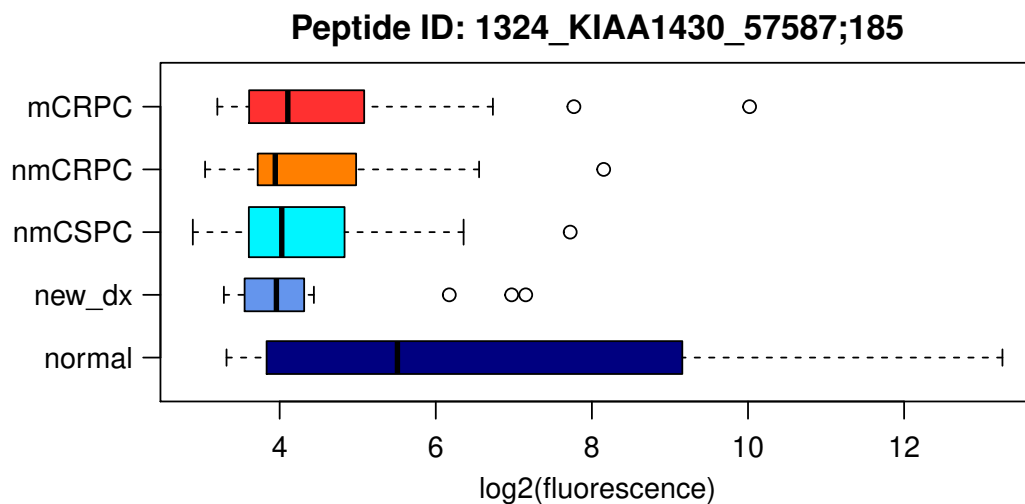


It appears that there are hardly any signals of different calls pattern among patients of different cancer stages, which corroborates with the results in the main manuscript. As expected, after correcting for false discovery rate, no peptides appear to be significant.

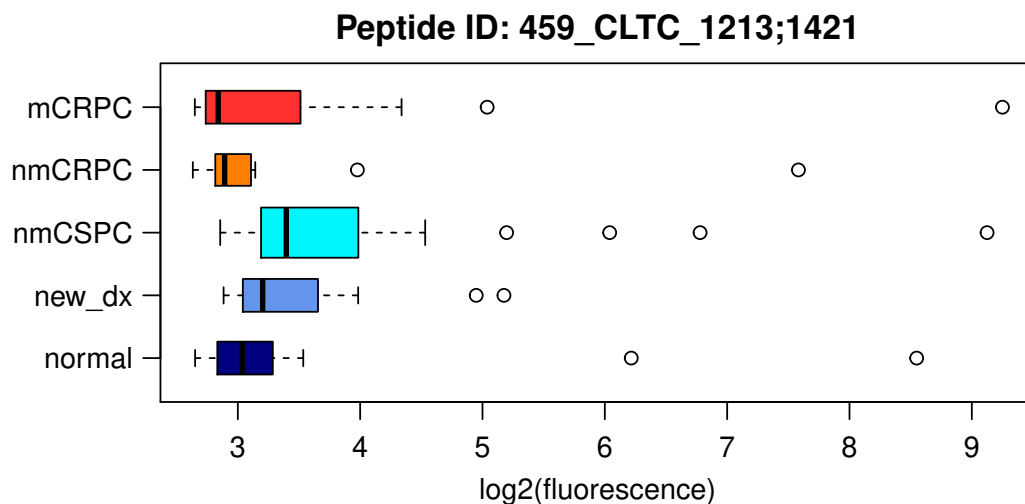
## 2.5 Tests on Fluorescence Levels

We hypothesized that while the overall number of peptides recognized may not change with disease stages, the composition of peptides recognized may be different. In this section, we will instead utilize the fluorescence data to investigate our hypothesis.

We are aware that the  $\log_2$  fluorescence data among the prostate cancer patients of different disease stages may violate the assumptions in the normal-theory one-way analysis of variance (ANOVA). For one, the variation of fluorescence levels among patients of different stages may not be similar, as illustrated by the boxplots of the peptide *1324\_KIAA1430\_57587;185* as an example.



Presence of outliers may also distort inference by the ANOVA. An example would be the peptide *459\_CLTC\_1213;1421*.



To avoid making any distributional assumptions on the fluorescence levels, we adopt the nonparametric Kruskal-Wallis test [McDonald, 2014] on each of the 177,604 peptides to test:

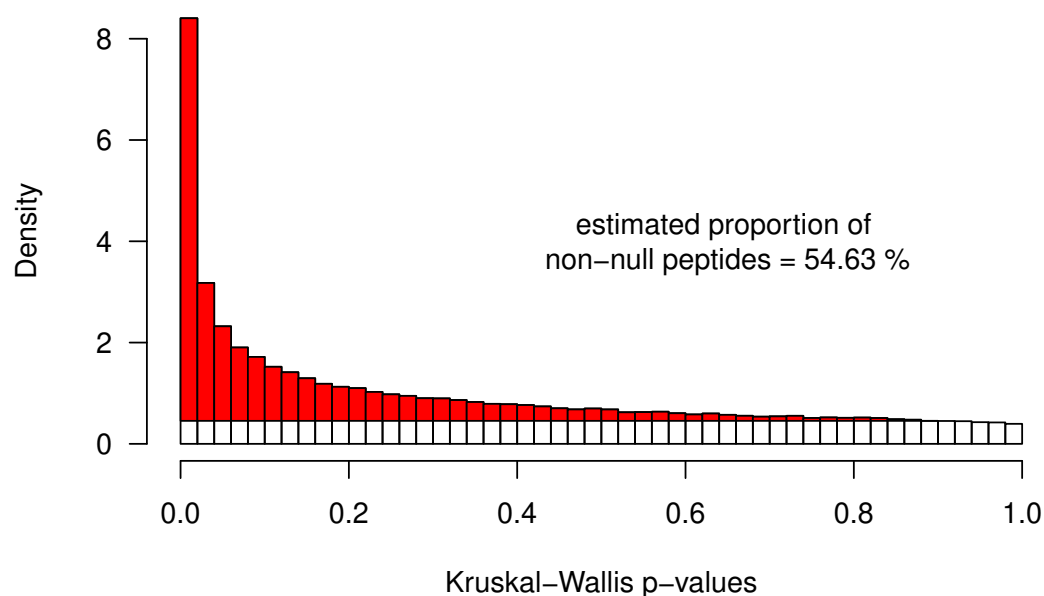
$H_0$ : The antibody response levels (in terms of  $\log_2$  fluorescence levels) for each disease stage are stochastically equal, ie. Once the fluorescence levels from all groups are ranked, the probability of an observation from one group being higher than an observation from another group is 0.5.

$H_1$ : The antibody response levels for at least one disease stage are stochastically dominant than those of other groups in the study.

Note that this is not a test of medians of the fluorescence levels since we are not making any distributional assumptions (shape and spread) on the fluorescence levels [McDonald, 2014].

After getting p-values for all the peptides, we plot the p-value histogram.

### p-values distribution for 177604 peptides



If cancer-stage effect is not present in our peptide array data, then the p-values from the Kruskal-Wallis tests would have a uniform distribution between 0 and 1, and we expect to see a rather flat-shaped histogram of p-values.

However, the p-values histogram exhibits large counts of significant p-values (p-values close to zero), and the shape of histogram flattens off exponentially with larger p-values. Such a large count of significant p-values may not be explained by false discovery alone, and that perhaps cancer-stage effect is indeed present in some of the peptides in our profile. The red-shaded regions of the histogram represents the estimated proportion of non-null peptides in the data based on Storey's q-values [Storey and Tibshirani, 2003] calculation obtained via the R package `fdrtool` [Strimmer, 2008].

We apply the Benjamini-Hochberg (BH) method [Benjamini and Hochberg, 1995] on the Kruskal-Wallis p-values to control for false discovery rate (FDR). The peptide counts at various BH FDR thresholds are tabulated below.

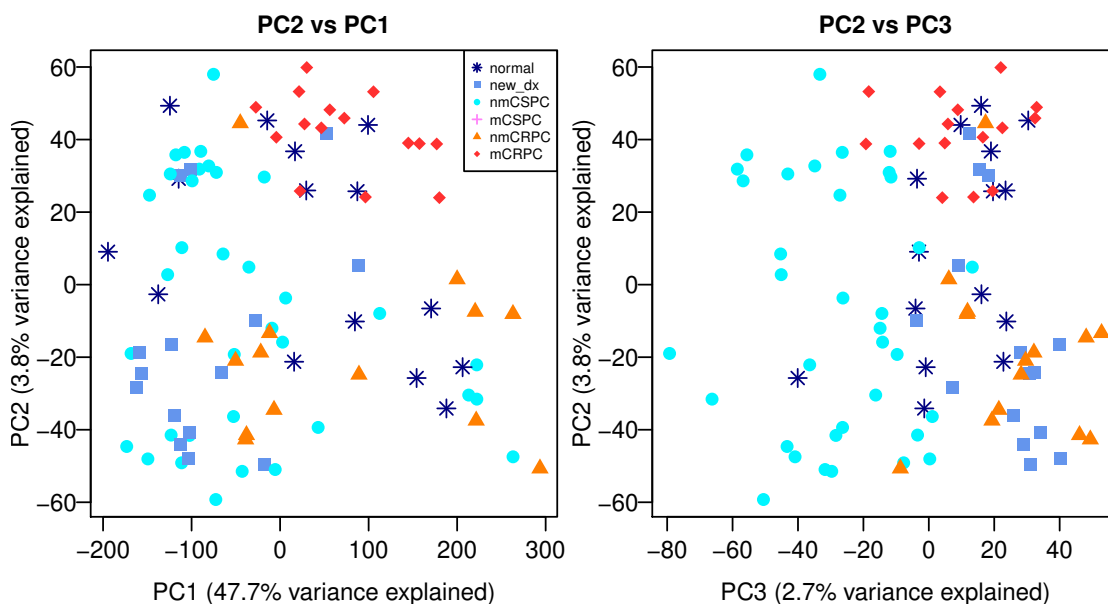
We could obtain a graphical representation to illustrate how the  $\log_2$  fluorescence levels differ across different

FDR threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Peptide counts	522	3499	7301	10826	13729	16515	18940	21401	23640	25828

cancer stages for these peptides via the Principal Component Analysis (PCA). For each peptide, we remove the grand mean (row mean) of the  $\log_2$  fluorescence levels for all patients before performing PCA on the residuals. If there is no cancer-stage effect, we expect these residual  $\log_2$  fluorescence to be random noises. Any observed (clustering) patterns among these residual data points reveal the effects of various stages of prostate cancer.

For purpose of uniformity, we also use the same color scheme to distinguish the different stages of cancer patients (notice how the spectrum of colors changes with severity of the cancer stages):

- navy for healthy subjects
- cornflower\_blue for `new_dx` newly diagnosed patients
- turquoise for `nmCSPC` patients
- light\_pink for `mCSPC` patients – these patients have no technical replicates and are excluded from this analysis
- dark\_orange for `nmCRPC` patients
- dark\_red for `mCRPC` patients



From the “*PC2 vs PC1*” plot, we observe that all `mCRPC` points are clustered at the topright of the panel, whereas newly-diagnosed and `nmCRPC` observations hover at the bottom of panel. The percentage of variance explained for each principal component (PC) is shown on the axis. Note that the first principal component manages to capture most of the variation in the data.



## 2.6 Pairwise Comparisons

Based on the Kruskal-Wallis tests, we identified 13729 peptides for which at least one group of patients stochastically dominates patients from the other disease stages at 5% BH FDR. Among these “interesting” peptides, we are interested in making some further pairwise comparisons between the groups of patients. In particular, we would like to analyze if antibody responses are different between cancer patients and healthy subjects. Besides that, the PCA plot has revealed that the mCRPC (worst-case scenario) patients are clustered away from the other patients and it may be interesting to compare how the antibody profiles of the mCRPC (worst-case-scenario) patients could be different from the other subjects. In addition, we would like to make pairwise comparison between consecutive groups of patients in terms of disease severity, namely:

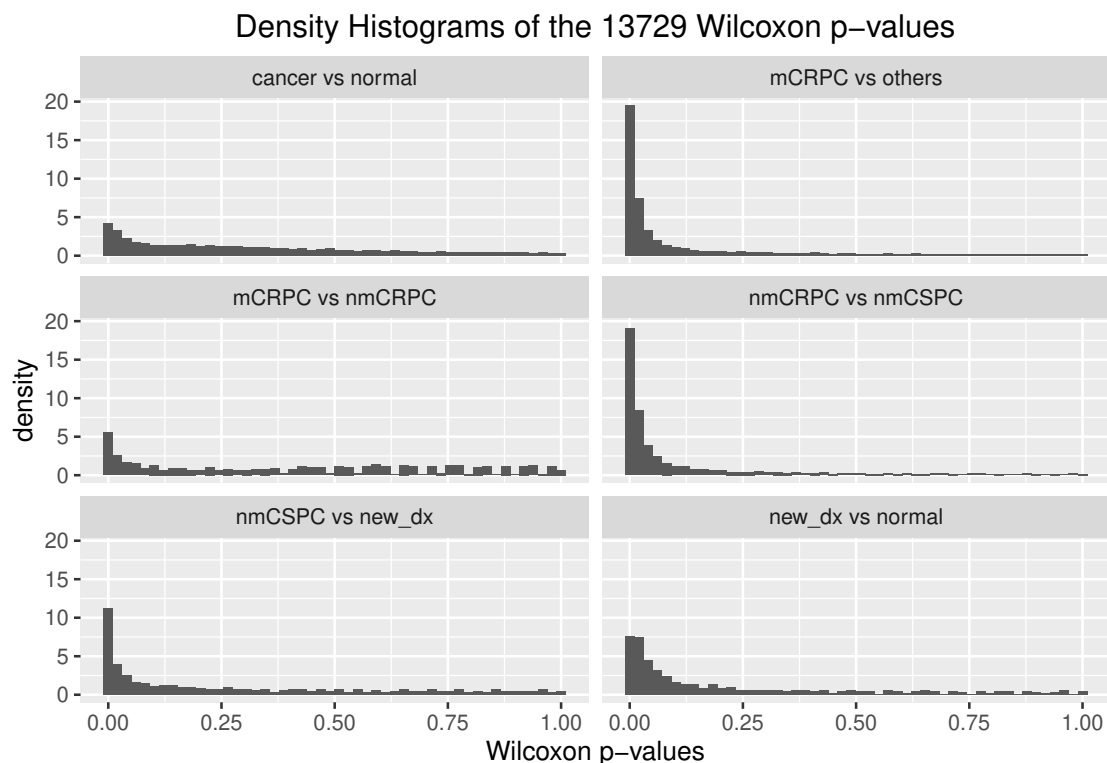
- between mCRPC and nmCRPC patients
- between nmCRPC and nmCSPC patients
- between nmCSPC and newly-diagnosed patients
- between newly-diagnosed and healthy subjects

For each of the 13729 peptides, we will perform a two-sided Wilcoxon-Rank-Sum (henceforth known as Wilcoxon) test [Winner, 2004] for each of the 6 contrasts as mentioned above, to test the following hypothesis:

$H_0$ : The antibody responses of both groups of patients are stochastically equal.

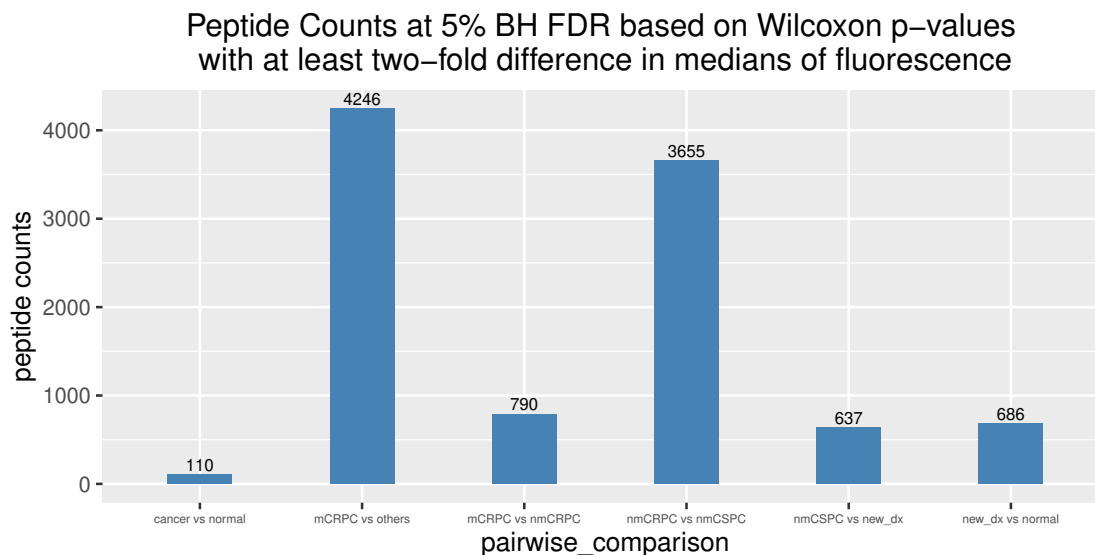
$H_1$ : The antibody responses of both groups of patients are NOT stochastically equal.

Exact p-values are computed for each Wilcoxon test whenever possible – if there are ties in the fluorescence levels, then normal approximation is used to obtain the p-values. After getting the 13729 p-values for each of the 6 contrasts, we plot their p-value density histograms at the same scale.



The BH procedure is also performed **on the 13729 Wilcoxon p-values separately for each contrast** to control FDR within each contrast (at 5%). On top of that, we also require at least a two-fold difference between the medians of the two groups' fluorescence levels, ie. the absolute difference of the medians of the

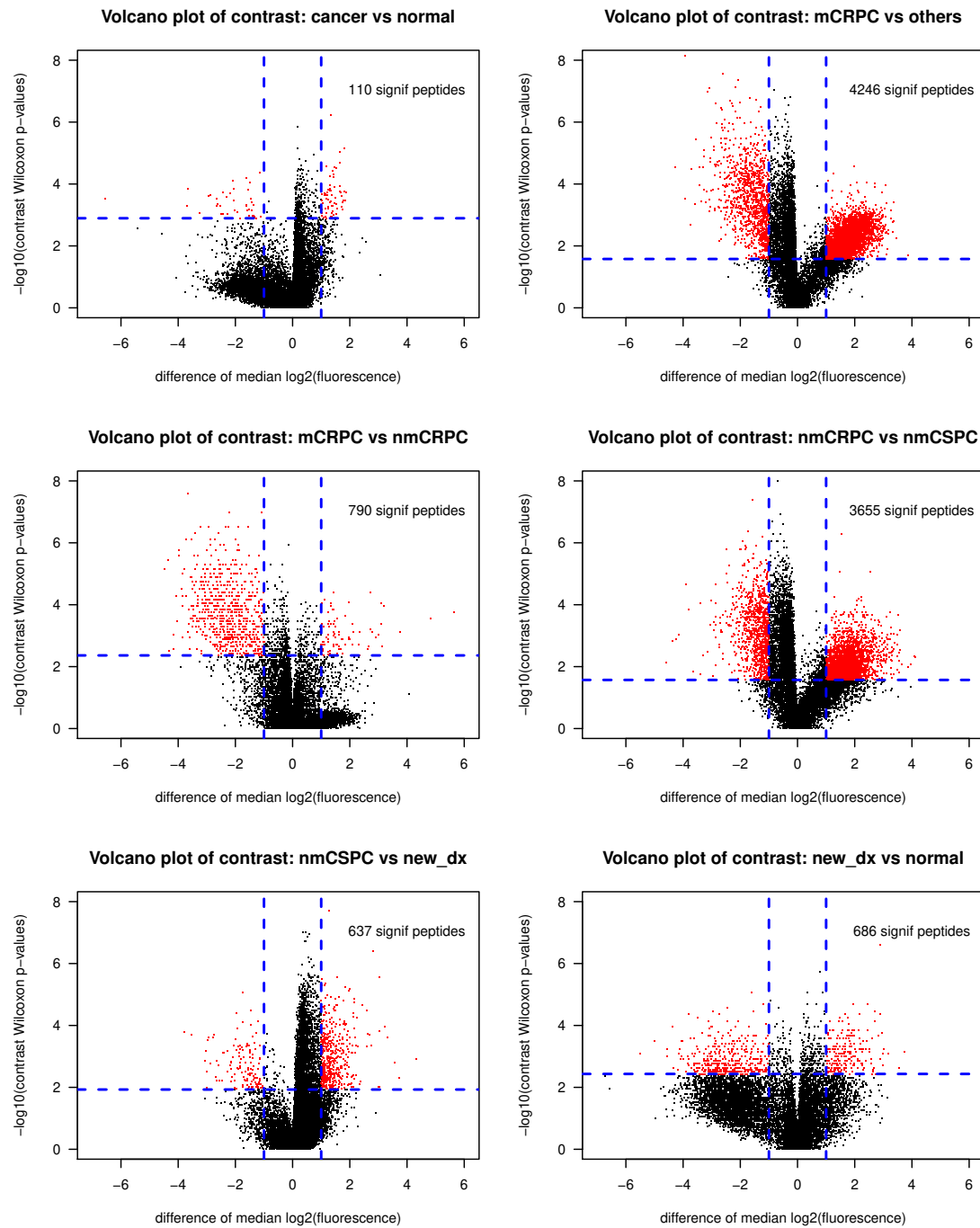
$\log_2$  fluorescence  $\geq 1$ . We graph the number of peptides that fulfill these two secondary cut-offs.



The volcano plots for these contrasts are also obtained. Each volcano plot has 13729 points, which are the peptides identified by the omnibus Kruskal-Wallis tests at 5% BH FDR. The Wilcoxon p-values of each contrast are plotted at  $-\log_{10}$  scale. The peptides which meet the secondary cutoffs for each contrast are colored red. The vertical blue dashed lines refer to the two-fold difference in medians. The horizontal blue dashed line refers to the minimum  $-\log_{10}$  (p-value) at which the peptides meet the secondary 5% BH FDR cutoff based on the Wilcoxon p-values.

From the bar chart of peptide counts and volcano plots of the pairwise comparisons of consecutive groups, it appears that as disease stage worsens, some peptides exhibit significantly higher median fluorescence levels (especially from nmCSPC to nmCRPC) whereas many peptides also display lower median fluorescence levels (especially from nmCRPC to mCRPC). Overall, if we compare the worst-case-scenario mCRPC against the other stages, many peptides exhibit significantly different (could be higher or lower) median fluorescence. Such changes across disease stages may explain why fewer significant peptides show up when we compare all cancer patients against healthy subjects.

The lists of significant/interesting peptides are also exported to the spreadsheet “09\_Significant\_Peptides.xlsx”. Specifically, the “Kruskal-Wallis” sheet contains the 13729 peptides at 5% BH FDR based on the Kruskal-Wallis tests on all 177,604 peptides. The other contrast sheets (for example, the “cancer vs normal” sheet) contain the lists of peptides at 5% BH FDR (based on the Wilcoxon tests on the 13729 peptides) which also exhibit at least a two-fold difference in medians of the two groups.



## 2.7 Visualization

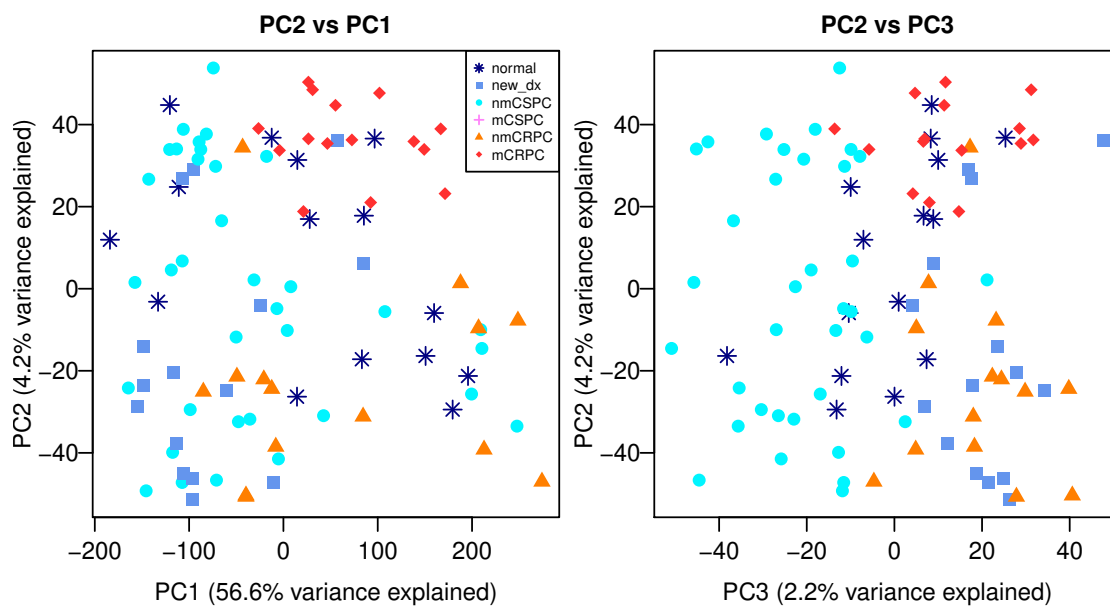
We are interested in peptides that meet the secondary cutoffs (at least a two-fold difference in medians and 5% BH FDR based on the Wilcoxon p-values) in at least one of the 6 contrasts. Out of the 13729 peptides at

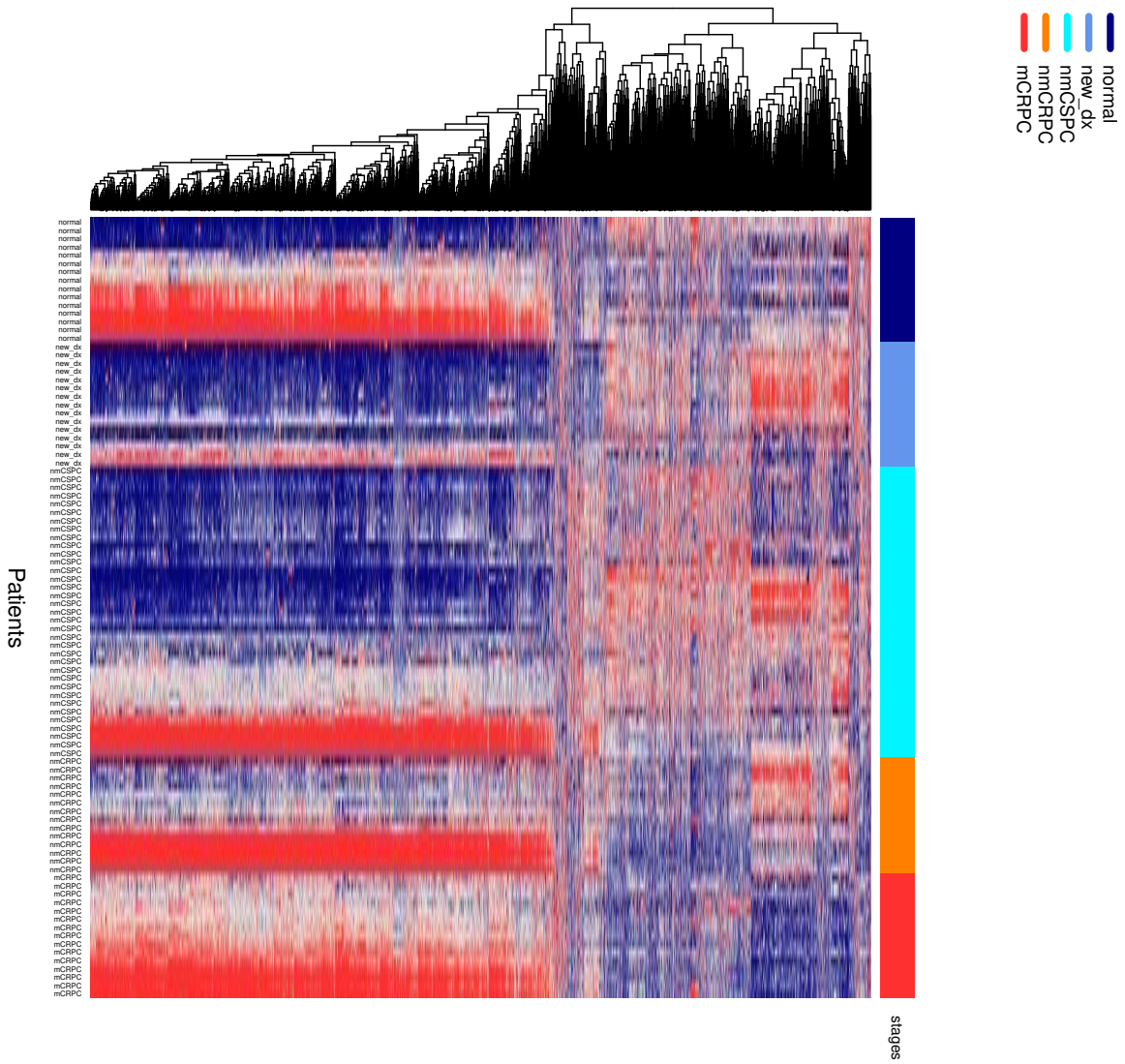
5% BH FDR based on the Kruskal-Wallis p-values, only 6708 of them also meet the secondary cutoff. We shall use these 6708 peptides to illustrate the effects of cancer stages via heatmap.

We remove the grand mean of each row of  $\log_2$  fluorescence. The fluorescence residuals are then winsorized at -2 and 2, which correspond to roughly bottom 12% and top 15% of the residuals. We then use these winsorized fluorescence residuals to plot the heatmap without any row-wise scaling. The color scheme of the heatmap is specified as navy for -2 which gradually transitions to firebrick for 2.

From the heatmap, we observe a clear pattern. The bottom part of the heatmap consists of peptides that show higher fluorescence levels consistently among all mCRPC patients compared to other groups of patients. Interestingly, there seem to be equal number of healthy subjects who display either higher or lower antibody responses in these peptides. Meanwhile, the upper part of the heatmap consists of peptides that show lower antibody responses consistently among mCRPC as well as nmCRPC patients.

We also could also reproduce the PCA plot for these unwinsorized fluorescence residuals of these 6708 peptides. Interestingly, they largely preserve the clustering pattern that we observe in the previous PCA plot when we use all the 13729 peptides at 5% BH FDR based on the Kruskal-Wallis p-values.





## 2.8 Gene-Set-Analyses

Here, we shall perform the gene-set analysis based on the interesting/significant peptides identified by the Wilcoxon-BH-FDR and absolute-difference-of-medians cutoffs for each contrast or pairwise comparison. Recall

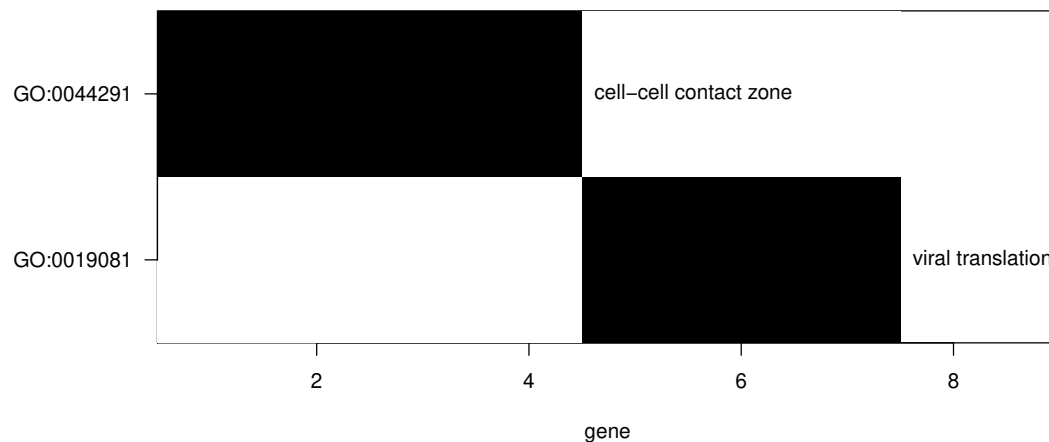
that the 177604 peptides correspond to 1611 proteins, and 1463 of these proteins have matching genes in “*uniprot\_gene\_entrez.csv*”. In this analysis, we deem a protein to be significant if it has at least one significant associated peptide.

Specifically, we investigate if there are any pre-specified gene-sets that are enriched for the genes associated with the list of significant peptides for each contrast or pairwise comparison. These pre-determined gene-sets are defined based on their functional categories or biological properties, such as the Gene Ontology (GO) annotations. Enriched gene-sets could reflect the biological signals in the peptide microarray data. The gene-set-analysis is performed with the R package `allez` [Newton et al., 2018]. We shall consider gene-sets containing at least 2 interesting/significant genes (`n.cell = 2`) with Bonferroni-corrected enrichment p-values not exceeding 5% (`nominal.alpha = 0.05`). We also limit our analysis to those GO gene-sets which contain at least 5 genes (`n.low = 5`) and at most 300 genes (`n.upp = 300`).

We present our gene-set-analysis results for each contrast in the following subsections.

### 2.8.1 Cancer vs Normal

Recall that we identified 110 interesting/significant peptides for this contrast. Based on these peptides, the gene-set-analysis yields the following waterfall plot.



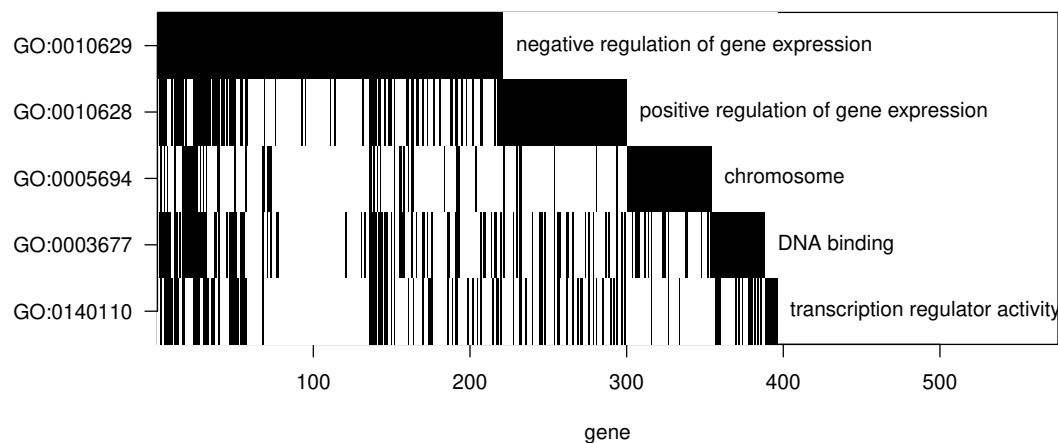
The waterfall plot was constructed by finding the significant (Bonferroni-corrected p-value < 0.05) GO term having the largest overlap with genes associated with proteins that have at least one significant peptide with at least one call among all patients (cell-cell contact zone GO:0044291) and placing it in the top row of the figure. We next removed these genes from the list and found the significant GO term having the highest overlap with the remainder of the list (viral translation GO: 0019081). This process is repeated, and genes identified by this sequential process are counted along the x-axis, and the overlap between the GO terms can be visually assessed. Shading under the ‘waterfall’ component of the graph indicates genes that were annotated to previously named categories.

We also tabulate the enriched/overrepresented GO terms. The last column of the table shows the genes associated with proteins that have at least one significant peptide in the contrast or pairwise comparison.

Term	Ontology	set.mean	set.size	z.score	in.genes
viral translation	BP	0.5000	3/6	4.2446	EIF3A; EIF3D; EIF3L
T-tubule	CC	0.5000	3/6	4.2446	ANK3; ATP1A1; AHNAK
cell-cell contact zone	CC	0.4444	4/9	4.5363	ANK3; ATP1A1; AFDN; AHNAK

### 2.8.2 mCRPC vs others

Recall that we identified 4246 interesting/significant peptides for this contrast. Based on these peptides, the gene-set-analysis yields the following waterfall plot.



Interpretation for the waterfall plot remains the same as above. We also tabulate the enriched/overrepresented GO terms. The last column of the table shows the genes associated with proteins that have at least one significant peptide in the contrast or pairwise comparison.

Term	Ontology	set.mean	set.size	z.score	in.genes
chromatin	CC	0.9333	70/75	4.6899	ACTB; AR; CEBPB; DHX9; EZH2; MSH6; H1F0; HIST1H1C; HIST1H2AD; H3F3A; H3F3B; HDAC1; HMGB2; HMG1; HMG2; HNRNPC; HNRNPK; HSF1; EIF3E; JUN; JUNB; JUND; MCM7; PRM2; RAD21; RAN; RBBP4; RBBP7; UPF1; SMARCA1; SMARCA4; SMARCC2; TCF3; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST2H2AC; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; HIST1H2AG; MTA1; MAGED1; H2AFY; NCOR2; IST1; MORF4L1; CBX3; POGZ; PDS5A; TARDBP; SUZ12; NOP53; BICRA; HP1BP3; PHF10; H2AFJ; FAM111A; NUCKS1; HIST1H2AH; HIST1H2BK; HIST3H2A; H2AFV; HIST2H2AB; H3F3C; HIST2H2AA4
chromosome	CC	0.8655	103/119	4.3386	ACTB; PARP1; AR; BCL6; CEBPB; CENPE; DHX9; DYNC1LI2; FBL; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HIST1H2AD; H3F3A; H3F3B; HDAC1; HMGB1; HMGB2; HMG1; HMG2; HNRNPC; HNRNPK; HSF1; EIF3E; JUN; JUNB; JUND; MCM7; MCM7; SEPTIN2; NKX3-1; PAFAH1B1; PHF2; PPP1CC; PRM2; PURB; RAD21; RAN; RBBP4; RBBP7; UPF1; RPA1; CLIP1; SMARCA1; SMARCA4; SMARCC2; SP100; SSB; SSRP1; TCF3; VCP; KAT6A; USP11; HIST3H3; HIST1H2AK; HIST1H2AM; HIST2H2AC; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; HIST1H2AG; MTA1; MAGED1; H2AFY; NCOR2; IST1; ARPC3; ARPC2; PCGF3; P3H4; MORF4L1; CBX3; POGZ; PDS5A; TARDBP; SUZ12; SPIDR; ORC6; REPIN1; NOP53; BICRA; HP1BP3; PHF10; H2AFJ; NSFL1C; THOC2; FAM111A; NUCKS1; MEAF6; HIST1H2AH; HIST1H2BK; HIST3H2A; H2AFV; TOP1MT; ANAPC16; HIST2H2AB; H3F3C; HIST2H2AA4

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
transcription regulator activity	MF	0.8571	114/133	4.3918	ACTN1; ACTN2; PARP1; AR; ATF4; BCL6; ZFP36L1; ZFP36L2; C1QBP; CEBPB; CEBPD; CTBP2; DDX1; DHX9; EPAS1; EZH2; GOLGB1; GTF2I; GTF3A; HDAC1; HMGB1; HMGB2; FOXA1; HNRNPK; HES1; HSF1; HSPA1A; DNAJB1; ID1; RBPJ; JUN; JUNB; JUND; MAFG; KMT2A; NFIA; NFE2L1; NFIB; NFIC; NFIL3; NKX3-1; NONO; CNOT2; NPAS2; NPM1; YBX1; PA2G4; PHF2; PURB; SMARCA1; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; SSRP1; SSX1; TAF7; TCF3; TDG; NR2F2; TSG101; SF1; ZNF24; VEZF1; KAT6A; EDF1; TSC22D1; MTA1; ZRANB2; MAGED1; IER2; NCOR2; MAML1; THRAP3; SAP18; RBM14; N4BP2L2; TADA3; HOXB13; SUB1; FOXJ3; TCF25; POGZ; WWC1; RYBP; TARDBP; EHF; NUPR1; SND1; HIPK2; BICRA; GMNN; LEF1; TDP2; ARID4B; YEATS2; ZNF395; SLC2A4RG; ENY2; BBX; MRTFB; ZNF462; ZNF350; NUCKS1; NIBAN2; IRX3; TBL1XR1; RAX2; ZNF664; CREB3L4; ZNF525; ZFP62
positive regulation of RNA metabolic process	BP	0.8519	138/162	4.7482	ACTN1; ACTN2; PARP1; AGT; APP; AR; ARF4; ATF4; BMPR1B; ZFP36L1; ZFP36L2; CEBPB; CEBPD; CTBP2; DDX3X; DDX5; DHX9; DVL1; EPAS1; FLT3LG; XRCC6; HDAC1; HMGB1; HMGB2; HMGNI; FOXA1; HNRNPD; HNRNPK; HES1; HSF1; HSPA1A; HSPA8; RBPJ; ILF3; INSIG1; JUN; JUNB; JUND; EPCAM; MAFG; MARS; MDK; MAP3K5; KMT2A; MYO6; NCL; NFIA; NFE2L1; NFIB; NFIC; NFIL3; NKX3-1; NOS1; NPAS2; NPM1; YBX1; PFKM; PHF2; PPP3CA; PPP3R1; MAP2K3; RAN; UPP1; RPS27A; SRSF5; TRA2B; SMARCA1; SMARCA4; SMARCC2; SNRNP70; SOX4; SP3; SP100; SREBF1; TAF7; TCEA1; TCF3; NR2F2; TSG101; UBA52; ZNF24; VEZF1; KAT6A; TAF15; OGT; KHSRP; EDF1; MTA1; MAGED1; PRDX6; IER2; MORF4L2; MICAL2; PUM1; BCLAF1; MAML1; THRAP3; NAMPT; HNRNPR; RBM14; TADA3; CAMKK2; RAI1; GCN1; FOXJ3; PHF8; WWC1; RYBP; TARDBP; NUP62; AUTS2; EHF; GNL3; NUPR1; PABPC1; HIPK2; BICRA; LEF1; WAC; YTHDF2; RTRAF; ARID4B; BANP; CHD7; ENY2; ZMIZ1; MRTFB; MAVS; CHD8; NUCKS1; NIBAN2; PAGR1; TBL1XR1; LBH; ING5; RAX2; CREB3L4; IRF2BP2
positive regulation of transcription, DNA-templated	BP	0.8478	117/138	4.2336	PARP1; AGT; APP; AR; ARF4; ATF4; BMPR1B; CEBPB; CEBPD; CTBP2; DDX3X; DHX9; DVL1; EPAS1; FLT3LG; XRCC6; HDAC1; HMGB1; HMGB2; HMGNI; FOXA1; HNRNPD; HNRNPK; HES1; HSF1; RBPJ; ILF3; INSIG1; JUN; JUNB; JUND; EPCAM; MAFG; MARS; MDK; MAP3K5; KMT2A; MYO6; NCL; NFIA; NFE2L1; NFIB; NFIC; NFIL3; NKX3-1; NOS1; NPAS2; NPM1; YBX1; PFKM; PHF2; PPP3CA; PPP3R1; MAP2K3; RAN; RPS27A; SMARCA1; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TAF7; TCF3; NR2F2; UBA52; ZNF24; VEZF1; KAT6A; TAF15; OGT; EDF1; MAGED1; IER2; MORF4L2; MICAL2; BCLAF1; MAML1; THRAP3; NAMPT; RBM14; TADA3; CAMKK2; RAI1; GCN1; FOXJ3; PHF8; WWC1; RYBP; TARDBP; NUP62; AUTS2; EHF; GNL3; HIPK2; BICRA; LEF1; WAC; RTRAF; ARID4B; BANP; CHD7; ENY2; ZMIZ1; MRTFB; MAVS; CHD8; NUCKS1; NIBAN2; PAGR1; TBL1XR1; LBH; ING5; RAX2; CREB3L4; IRF2BP2
positive regulation of RNA biosynthetic process	BP	0.8472	122/144	4.3183	ACTN1; ACTN2; PARP1; AGT; APP; AR; ARF4; ATF4; BMPR1B; CEBPB; CEBPD; CTBP2; DDX3X; DHX9; DVL1; EPAS1; FLT3LG; XRCC6; HDAC1; HMGB1; HMGB2; HMGNI; FOXA1; HNRNPD; HNRNPK; HES1; HSF1; RBPJ; ILF3; INSIG1; JUN; JUNB; JUND; EPCAM; MAFG; MARS; MDK; MAP3K5; KMT2A; MYO6; NCL; NFIA; NFE2L1; NFIB; NFIC; NFIL3; NKX3-1; NOS1; NPAS2; NPM1; YBX1; PFKM; PHF2; PPP3CA; PPP3R1; MAP2K3; RAN; RPS27A; SMARCA1; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TAF7; TCF3; NR2F2; TSG101; UBA52; ZNF24; VEZF1; KAT6A; TAF15; OGT; EDF1; MTA1; MAGED1; IER2; MORF4L2; MICAL2; BCLAF1; MAML1; THRAP3; NAMPT; RBM14; TADA3; CAMKK2; RAI1; GCN1; FOXJ3; PHF8; WWC1; RYBP; TARDBP; NUP62; AUTS2; EHF; GNL3; NUPR1; HIPK2; BICRA; LEF1; WAC; RTRAF; ARID4B; BANP; CHD7; ENY2; ZMIZ1; MRTFB; MAVS; CHD8; NUCKS1; NIBAN2; PAGR1; TBL1XR1; LBH; ING5; RAX2; CREB3L4; IRF2BP2



(continued)

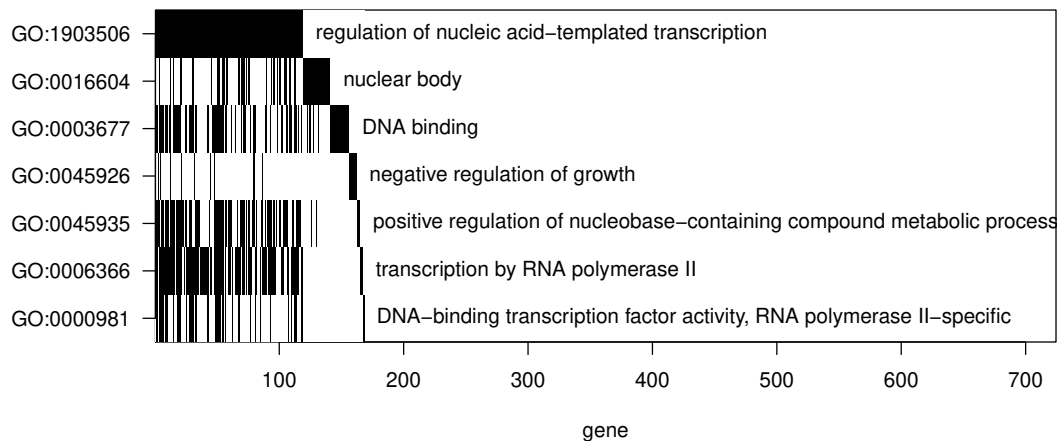
Term	Ontology	set.mean	set.size	z.score	in.genes
positive regulation of nucleic acid-templated transcription	BP	0.8472	122/144	4.3183	ACTN1; ACTN2; PARP1; AGT; APP; AR; ARF4; ATF4; BMPR1B; CEBPB; CEBPD; CTBP2; DDX3X; DHX9; DVL1; EPAS1; FLT3LG; XRCC6; HDAC1; HMGB1; HMGB2; HMGN1; FOXA1; HNRNPD; HNRNPK; HES1; HSF1; RBPJ; ILF3; INSIG1; JUN; JUNB; JUND; EPCAM; MAFG; MARS; MDK; MAP3K5; KMT2A; MYO6; NCL; NFIA; NFE2L1; NFIB; NFIC; NFIL3; NKX3-1; NOS1; NPAS2; NPM1; YBX1; PFKM; PHF2; PPP3CA; PPP3R1; MAP2K3; RAN; RPS27A; SMARCA1; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TAF7; TCF3; NR2F2; TSG101; UBA52; ZNF24; VEZF1; KAT6A; TAF15; OGT; EDF1; MTA1; MAGED1; IER2; MORF4L2; MICAL2; BCLAF1; MAML1; THRAP3; NAMPT; RBM14; TADA3; CAMKK2; RAI1; GCN1; FOXJ3; PHF8; WWC1; RYBP; TARDBP; NUP62; AUTS2; EHF; GNL3; NUPR1; HIPK2; BICRA; LEF1; WAC; RTRAF; ARID4B; BANP; CHD7; ENY2; ZMIZ1; MRTFB; MAVS; CHD8; NUCKS1; NIBAN2; PAGR1; TBL1XR1; LBH; ING5; RAX2; CREB3L4; IRF2BP2
DNA binding	MF	0.8378	155/185	4.6798	ACTB; ADAR; PARP1; APLP2; APP; AR; ATF4; BCL6; ZFP36L1; ZFP36L2; CEBPB; CEBPD; CUX1; DDX1; DDX3X; DHX9; EPAS1; EZH2; XRCC6; GOLGB1; MSH6; GTF2I; GTF3A; H1F0; HIST1H1C; H3F3A; H3F3B; HDAC1; HMGB1; HMGB2; HMGN1; HMGN2; FOXA1; HNRNPC; HNRNPD; HNRNPK; HES1; HSF1; HSPD1; RBPJ; ILF3; JUN; JUNB; JUND; MCM3; MCM7; KMT2A; NACA; NCL; NFIA; NFE2L1; NFIB; NFIC; NFIL3; NKX3-1; NONO; NPAS2; NPM1; YBX1; PA2G4; PNN; PRM2; PURB; RAD23B; RBBP4; UPP1; RPA1; RPL6; RPL7; RPS15; RPS27; SET; SMARCA1; SMARCA4; SMARCC2; SON; SOX4; SP3; SP100; SREBF1; SSRP1; TAF7; TCEA1; TCF3; TDG; NR2F2; TSG101; ZNF24; VEZF1; ZFAND5; KAT6A; TAF15; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; KHSRP; DDX3Y; EDF1; EED; TAF1C; MTA1; H2AFY; IER2; BCLAF1; THRAP3; DNAJB6; AKAP9; RBM5; SRRM1; ZMPSTE24; HOXB13; RAI1; SUB1; FOXJ3; TCF25; SMG1; RYBP; TARDBP; SUZ12; LSM14A; EHF; NUPR1; REPIN1; HP1BP3; SIDT2; LEF1; CXXC5; TDP2; SRRT; XRN1; ZFAND6; BANP; IFT57; STRBP; CHD7; ZNF395; SLC2A4RG; BBX; SCYL1; CHD8; ZNF350; NUCKS1; IRX3; TBL1XR1; RAX2; HIST3H2A; GTF3C6; TOP1MT; ZNF664; CREB3L4; ZMAT2; ZNF525; H3F3C
positive regulation of gene expression	BP	0.8223	162/197	4.3466	ACTB; ADAR; PARP1; AGT; ANK3; APP; AR; ARF4; ATF4; BMPR1B; C1QBP; CEBPB; CEBPD; CTBP2; DDX3X; DDX5; DHX9; DVL1; EPAS1; FLT3LG; XRCC6; HDAC1; HMGB1; HMGB2; HMGN1; FOXA1; HNRNPC; HNRNPD; HNRNPK; HES1; HSF1; HSPA1A; HSPA8; ID1; RBPJ; ILF3; INSIG1; EIF3E; JUN; JUNB; JUND; KRAS; LDLR; LIMS1; EPCAM; MAFG; MARS; MDK; MAP3K5; KMT2A; AFDN; MYH9; MYO1C; MYO6; NCL; NFIA; NFE2L1; NFIB; NFIC; NFIL3; NKX3-1; NOS1; NPAS2; NPM1; YBX1; PFKM; PHF2; PPP3CA; PPP3R1; MAP2K3; RAN; RPL5; RPL26; RPL30; RPS4X; RPS7; RPS27A; SRSF5; TRA2B; SMARCA1; SMARCA4; SMARCC2; SNRNP70; SOX4; SP3; SP100; SREBF1; TAF7; TCF3; NR2F2; UBA52; EZR; ZNF24; VEZF1; KAT6A; TAF15; OGT; EIF3C; EIF3D; EDF1; TAF1C; HGS; RPL23; MAGED1; H2AFY; PRDX6; IER2; NCOR2; MORF4L2; MICAL2; BCLAF1; EIF4A3; MAML1; THRAP3; NAMPT; ZMPSTE24; RBM14; TADA3; SYNCRIP; CAMKK2; RAI1; GCN1; PKP3; FOXJ3; PHF8; WWC1; RYBP; TARDBP; SF3B1; NUP62; AUTS2; EHF; GNL3; PABPC1; HIPK2; BICRA; RPS27L; LEF1; WAC; YTHDF2; RTRAF; ARID4B; YTHDF1; BANP; DNAJA4; CHD7; ENY2; ZMIZ1; MRTFB; MAVS; CHD8; NUCKS1; NIBAN2; SECISBP2; PAGR1; TBL1XR1; LBH; ING5; RAX2; NIBAN1; CREB3L4; IRF2BP2

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
negative regulation of gene expression	BP	0.7950	221/278	4.2448	A2M; ADAR; PARP1; APP; AR; ATF4; BCL6; ZFP36L1; ZFP36L2; C1QBP; CAST; CEBPB; CEBPD; CTBP2; DDX3X; DDX5; DHX9; EZH2; XRCC6; H1F0; HIST1H1C; H3F3A; H3F3B; HDAC1; HMGB1; HMGB2; FOXA1; HNRNPC; HNRNPD; HNRNPK; HES1; HSF1; HSPA1A; HSPAS; DNAJB1; ID1; RBPJ; ILF3; EIF3E; JUN; RPSA; LDLR; LIMS1; CAPRIN1; NCL; RPL10A; NFIB; NFIC; NFIL3; NKX3-1; NONO; CNOT2; NOTCH2; NPM1; YBX1; PA2G4; PHF2; PPP3CA; PRNP; PSMA1; PSMA6; PSMB6; PSMC1; PSMD1; PSMD3; PSMD4; PSME1; PURB; RAN; RANBP2; RBBP4; RBBP7; UPP1; RNH1; RPL3; RPL5; RPL6; RPL7; RPL7A; RPL8; RPL9; RPL12; RPL13; RPL15; RPL17; RPL18; RPL18A; RPL19; RPL21; RPL22; RPL23A; RPL24; RPL26; RPL27; RPL30; RPL27A; RPL28; RPL29; RPL31; RPL32; RPL34; RPL37; RPL37A; RPL38; RPL39; RPL41; RPL36A; RPLP0; RPS2; RPS3A; RPS4X; RPS4Y1; RPS6; RPS7; RPS8; RPS10; RPS11; RPS12; RPS13; RPS14; RPS15; RPS15A; RPS17; RPS18; RPS19; RPS20; RPS23; RPS24; RPS25; RPS26; RPS27; RPS27A; SET; SRSF4; SRSF7; SMARCA4; SMARCC2; SP3; SP100; SREBF1; SSB; TAF7; TCF3; TDG; TMBIM6; NR2F2; TSG101; UBA52; EZR; ZNF24; CSDE1; KAT6A; FXR1; USP9X; HIST1H4C; HIST1H4L; KHSRP; EED; RPL14; RPL23; MAGED1; TMEM59; H2AFY; NCOR2; PUM1; BCLAF1; EIF4A3; POM121; RBM8A; THRAP3; DNAJB6; HNRNPR; ZMPSTE24; SAP18; N4BP2L2; HOXB13; SYNCRIP; HEXIM1; CELF1; SRSF10; PKP3; RPL35; PHB2; CBX3; CASC3; TCF25; SMG1; PHF8; WWC1; NEDD4L; RYBP; TARDBP; SUZ12; RPL13A; NUP62; RPL36; SERBP1; PABPC1; SND1; EIF2AK1; PDCC4; HIPK2; SNX12; NOP53; GMNN; ZNF706; LEF1; YTHDF2; CXXC5; SRRT; PTRH2; RASD1; UIMC1; XRN1; YEATS2; CHD8; ZNF350; NIBAN2; SECISBP2; TBL1XR1; LBH

### 2.8.3 mCRPC vs nmCRPC

Recall that we identified 790 interesting/significant peptides for this contrast. Based on these peptides, the gene-set-analysis yields the following waterfall plot.



Interpretation for the waterfall plot remains the same as above. We also tabulate the enriched/overrepresented GO terms. The last column of the table shows the genes associated with proteins that have at least one significant peptide in the contrast or pairwise comparison.

Term	Ontology	set.mean	set.size	z.score	in.genes
negative regulation of growth	BP	0.7143	15/21	4.3829	AGT; BCL6; DDX3X; DNAJB2; HSPA1A; MT1E; MT1G; MT1H; MT1X; MT2A; NOTCH2; RBBP7; SMARCA4; RAI1; WWC1

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
DNA-binding transcription factor activity, RNA polymerase II-specific	MF	0.5200	39/75	4.6219	PARP1; AR; BCL6; ZFP36L1; ZFP36L2; CEBPB; KLF6; EPAS1; FOXA1; HES1; HSF1; JUN; KMT2A; NFIA; NFIC; NONO; NPAS2; YBX1; PURB; SMARCC2; SOX4; SP3; SP100; SREBF1; TCF3; VEZF1; TSC22D1; NCOR2; HOXB13; SUB1; POGZ; EHF; YEATS2; BBX; ZNF462; ZNF350; NUCKS1; ZNF525; ZFP62
nuclear body	CC	0.4854	50/103	4.6676	ADD1; AR; DDX3X; EPAS1; MKNK2; HSF1; HSPA1A; NBR1; NONO; PNN; PKN2; BRD2; RPA1; SRSF5; SNRPC; SON; SP3; SP100; TCF3; SF1; KAT6A; AKAP17A; TRIP12; NCOR2; PUM1; BCLAF1; MAML1; THRAP3; HIPK3; NAMPT; SRRM1; RBM14; ATXN2L; CASC3; FBNP4; SUZ12; MORC3; SRRM2; PNISR; VIRMA; GNL3; HIPK2; HP1BP3; WAC; UIMC1; BANP; SLC2A4RG; THOC2; NUFIP2; ZNF350
positive regulation of RNA metabolic process	BP	0.4568	74/162	5.1325	ACTN1; PARP1; AGT; AR; ZFP36L1; ZFP36L2; CEBPB; KLF6; CTBP2; DDX3X; DVL1; EPAS1; XRCC6; HMGNI; FOXA1; HES1; HSF1; HSPA1A; HSPA8; ILF3; JUN; KMT2A; NFIA; NFIC; NPAS2; YBX1; PFKM; MAP2K3; RAN; SRSF5; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCEA1; TCF3; VEZF1; KAT6A; OGT; KHSRP; MAGED1; MICAL2; PUM1; BCLAF1; MAML1; THRAP3; NAMPT; RBM14; CAMKK2; RAI1; GCN1; PHF8; WWC1; RYBP; NUP62; AUTS2; EHF; GNL3; PABPC1; HIPK2; BICRA; WAC; YTHDF2; RTRAF; BANP; CHD7; MRTFB; MAVS; NUCKS1; PAGR1; TBL1XR1; LBH
positive regulation of transcription, DNA-templated	BP	0.4565	63/138	4.6850	PARP1; AGT; AR; CEBPB; KLF6; CTBP2; DDX3X; DVL1; EPAS1; XRCC6; HMGNI; FOXA1; HES1; HSF1; ILF3; JUN; KMT2A; NFIA; NFIC; NPAS2; YBX1; PFKM; MAP2K3; RAN; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCF3; VEZF1; KAT6A; OGT; MAGED1; MICAL2; BCLAF1; MAML1; THRAP3; NAMPT; RBM14; CAMKK2; RAI1; GCN1; PHF8; WWC1; RYBP; NUP62; AUTS2; EHF; GNL3; HIPK2; BICRA; WAC; RTRAF; BANP; CHD7; MRTFB; MAVS; NUCKS1; PAGR1; TBL1XR1; LBH
transcription regulator activity	MF	0.4511	60/133	4.4458	ACTN1; PARP1; AR; BCL6; ZFP36L1; ZFP36L2; C1QBP; CEBPB; KLF6; CTBP2; EPAS1; GOLGB1; FOXA1; HES1; HSF1; HSPA1A; DNAJB1; JUN; KMT2A; NFIA; NFIC; NONO; NPAS2; YBX1; PURB; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCF3; SF1; VEZF1; KAT6A; TSC22D1; MAGED1; NCOR2; MAML1; THRAP3; RBM14; HOXB13; SUB1; POGZ; WWC1; RYBP; EHF; HIPK2; BICRA; GMNN; YEATS2; SLC2A4RG; BBX; MRTFB; ZNF462; ZNF350; NUCKS1; TBL1XR1; ZNF525; ZFP62
positive regulation of RNA biosynthetic process	BP	0.4444	64/144	4.4587	ACTN1; PARP1; AGT; AR; CEBPB; KLF6; CTBP2; DDX3X; DVL1; EPAS1; XRCC6; HMGNI; FOXA1; HES1; HSF1; ILF3; JUN; KMT2A; NFIA; NFIC; NPAS2; YBX1; PFKM; MAP2K3; RAN; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCF3; VEZF1; KAT6A; OGT; MAGED1; MICAL2; BCLAF1; MAML1; THRAP3; NAMPT; RBM14; CAMKK2; RAI1; GCN1; PHF8; WWC1; RYBP; NUP62; AUTS2; EHF; GNL3; HIPK2; BICRA; WAC; RTRAF; BANP; CHD7; MRTFB; MAVS; NUCKS1; PAGR1; TBL1XR1; LBH
positive regulation of nucleic acid-templated transcription	BP	0.4444	64/144	4.4587	ACTN1; PARP1; AGT; AR; CEBPB; KLF6; CTBP2; DDX3X; DVL1; EPAS1; XRCC6; HMGNI; FOXA1; HES1; HSF1; ILF3; JUN; KMT2A; NFIA; NFIC; NPAS2; YBX1; PFKM; MAP2K3; RAN; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCF3; VEZF1; KAT6A; OGT; MAGED1; MICAL2; BCLAF1; MAML1; THRAP3; NAMPT; RBM14; CAMKK2; RAI1; GCN1; PHF8; WWC1; RYBP; NUP62; AUTS2; EHF; GNL3; HIPK2; BICRA; WAC; RTRAF; BANP; CHD7; MRTFB; MAVS; NUCKS1; PAGR1; TBL1XR1; LBH
regulation of transcription by RNA polymerase II	BP	0.4352	84/193	4.9595	PARP1; AR; BCL6; ZFP36L1; ZFP36L2; C1QBP; CEBPB; KLF6; CTBP2; CUX1; DDX3X; EPAS1; XRCC6; HMGNI; FOXA1; HES1; HSF1; HSPA1A; DNAJB1; JUN; MAGEA1; KMT2A; NFIA; NFIC; NONO; NPAS2; YBX1; PFKM; PSMA6; PSMB6; PSMD1; PSMD3; PSMD4; PSME1; PURB; RBBP7; BRD2; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCF3; VEZF1; USP9X; OGT; TSC22D1; RPL23; MAGED1; NCOR2; MICAL2; MAML1; THRAP3; NAMPT; RBM14; HOXB13; HEXIM1; SUB1; GCN1; POGZ; WWC1; NEDD4L; RYBP; SUZ12; AUTS2; EHF; GNL3; PDCC4; HIPK2; NOP53; CXXC5; RTRAF; CHD7; YEATS2; BBX; MRTFB; MAVS; ZNF350; NUCKS1; PAGR1; TBL1XR1; ZNF525; ZFP62

(continued)

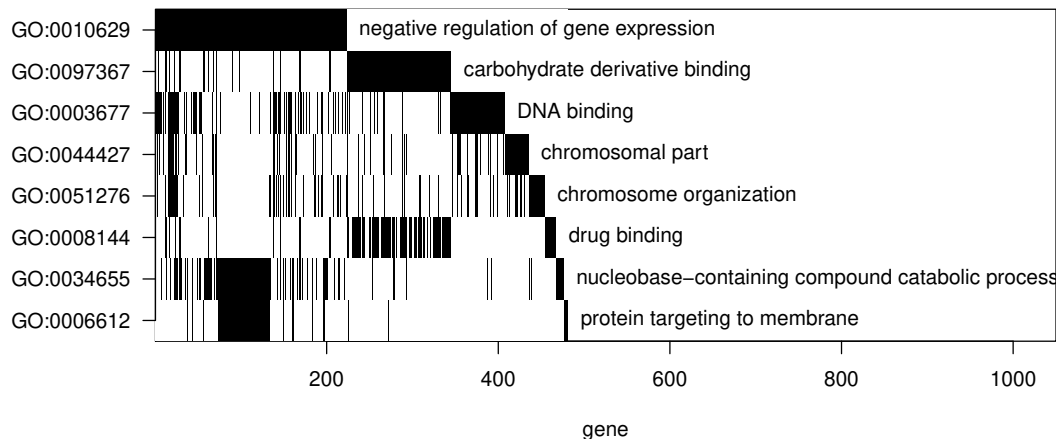
Term	Ontology	set.mean	set.size	z.score	in.genes
transcription by RNA polymerase II	BP	0.4272	91/213	4.9729	PARP1; AR; BCL6; ZFP36L1; ZFP36L2; BTF3; C1QBP; CEBPB; KLF6; CTBP2; CUX1; DDX3X; DVLI; EPAS1; XRCC6; HMG1; FOXA1; HES1; HSF1; HSPA1A; DNAJB1; JUN; MAGEA1; KMT2A; NFIA; NFIC; NONO; NOTCH2; NPAS2; YBX1; PFKM; PSMA6; PSMB6; PSMD1; PSMD3; PSMD4; PSME1; PURB; RBBP7; BRD2; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCEA1; TCF3; VEZF1; USP9X; OGT; TSC22D1; TAF1C; RPL23; MAGED1; NCOR2; MICAL2; MAML1; THRAP3; NAMPT; RBM14; HOXB13; HEXIM1; SUB1; GCN1; POGZ; WWC1; NEDD4L; RYBP; SUZ12; AUTS2; EHF; GNL3; PDCC4; HIPK2; NOP53; CXXC5; PCF11; SRRT; RTRAF; CHD7; YEATS2; BBX; MRTFB; MAVS; ZNF350; NUCKS1; PAGR1; TBL1XR1; ZNF525; ZFP62
positive regulation of nucleobase-containing compound metabolic process	BP	0.4254	77/181	4.4664	ACTN1; PARP1; AGT; AR; ZFP36L1; ZFP36L2; CCT6A; CEBPB; KLF6; CTBP2; DDX3X; DVLI; EPAS1; XRCC6; HMG1; FOXA1; HES1; HSF1; HSPA1A; HSPA8; ILF3; JUN; KMT2A; NFIA; NFIC; NPAS2; YBX1; PFKM; MAP2K3; RAN; SRSF5; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCEA1; TCF3; VEZF1; KAT6A; USP9X; OGT; KHSRP; MAGED1; MICAL2; PUM1; BCLAF1; MAML1; THRAP3; NAMPT; RBM14; CAMKK2; RAI1; GCN1; PHF8; WWC1; RYBP; NUP62; AUTS2; EHF; GNL3; PABPC1; HIPK2; BICRA; WAC; YTHDF2; RTRAF; UIMC1; BANP; CHD7; MRTFB; MAVS; NUCKS1; PAGR1; TBL1XR1; LBH ADAR; PARP1; AR; BCL6; ZFP36L1; ZFP36L2; CEBPB; KLF6; CUX1; DDX3X; EPAS1; XRCC6; GOLGB1; HMG1; FOXA1; HES1; HSF1; ILF3; JUN; MCM3; KMT2A; NACA; NFIA; NFIC; NONO; NPAS2; YBX1; NUCB2; PNN; PURB; RAD23B; RPA1; SET; SMARCA4; SMARCC2; SON; SOX4; SP3; SP100; SREBF1; TCEA1; TCF3; VEZF1; ZFAND5; KAT6A; KHSRP; DDX3Y; TAF1C; BCLAF1; THRAP3; DNAJB6; AKAP9; RBM5; SRRM1; HOXB13; RAI1; SUB1; SMG1; RYBP; SUZ12; LSM14A; EHF; REPIN1; HP1BP3; CXXC5; SRRT; ZFAND6; BANP; STRBP; CHD7; SLC2A4RG; BBX; SCYL1; ZNF350; NUCKS1; TBL1XR1; ZNF525
DNA binding	MF	0.4162	77/185	4.2259	PARP1; AR; BCL6; ZFP36L1; ZFP36L2; C1QBP; CEBPB; KLF6; CTBP2; CUX1; DDX3X; DVLI; EPAS1; XRCC6; GOLGB1; HMG1; FOXA1; HES1; HSF1; HSPA1A; HSPA8; DNAJB1; ILF3; JUN; MAGEA1; KMT2A; NFIA; NFIC; NONO; NOTCH2; NPAS2; YBX1; PFKM; MAP2K3; PSMA6; PSMB6; PSMD1; PSMD3; PSMD4; PSME1; PURB; RAN; RBBP7; BRD2; SET; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCEA1; TCF3; VEZF1; KAT6A; AKAP17A; USP9X; OGT; KHSRP; TSC22D1; TAX1BP1; TAF1C; RPL23; MAGED1; NCOR2; MICAL2; BCLAF1; MAML1; THRAP3; DNAJB6; NAMPT; PCGF3; RBM14; HOXB13; HEXIM1; CAMKK2; RAI1; SUB1; GCN1; PHB2; TAB2; POGZ; PHF8; WWC1; NEDD4L; RYBP; SUZ12; NUP62; AUTS2; EHF; GNL3; PDCC4; HIPK2; NOP53; BICRA; HP1BP3; GMNN; WAC; CXXC5; SRRT; RTRAF; UIMC1; BANP; CHD7; YEATS2; SLC2A4RG; BBX; MRTFB; MAVS; ZNF350; NUCKS1; PAGR1; TBL1XR1; LBH; ZNF525; ZFP62
regulation of transcription, DNA-templated	BP	0.4021	117/291	4.9447	ACTN1; PARP1; AGT; AR; BCL6; ZFP36L1; ZFP36L2; C1QBP; CEBPB; KLF6; CTBP2; CUX1; DDX3X; DVLI; EPAS1; XRCC6; GOLGB1; HMG1; FOXA1; HES1; HSF1; HSPA1A; HSPA8; DNAJB1; ILF3; JUN; MAGEA1; KMT2A; NFIA; NFIC; NONO; NOTCH2; NPAS2; YBX1; PFKM; MAP2K3; PSMA6; PSMB6; PSMD1; PSMD3; PSMD4; PSME1; PURB; RAN; RBBP7; BRD2; SET; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCEA1; TCF3; SF1; VEZF1; KAT6A; AKAP17A; USP9X; OGT; KHSRP; TSC22D1; TAX1BP1; TAF1C; RPL23; MAGED1; NCOR2; MICAL2; BCLAF1; MAML1; THRAP3; DNAJB6; NAMPT; PCGF3; RBM14; HOXB13; GCN1; PHB2; TAB2; POGZ; PHF8; WWC1; NEDD4L; RYBP; SUZ12; NUP62; AUTS2; EHF; GNL3; PDCC4; HIPK2; NOP53; BICRA; HP1BP3; GMNN; WAC; CXXC5; SRRT; RTRAF; UIMC1; BANP; CHD7; YEATS2; SLC2A4RG; BBX; MRTFB; MAVS; ZNF350; NUCKS1; PAGR1; TBL1XR1; LBH; ZNF525; ZFP62
regulation of nucleic acid-templated transcription	BP	0.3980	119/299	4.8548	ACTN1; PARP1; AGT; AR; BCL6; ZFP36L1; ZFP36L2; C1QBP; CEBPB; KLF6; CTBP2; CUX1; DDX3X; DVLI; EPAS1; XRCC6; GOLGB1; HMG1; FOXA1; HES1; HSF1; HSPA1A; HSPA8; DNAJB1; ILF3; JUN; MAGEA1; KMT2A; NFIA; NFIC; NONO; NOTCH2; NPAS2; YBX1; PFKM; MAP2K3; PSMA6; PSMB6; PSMD1; PSMD3; PSMD4; PSME1; PURB; RAN; RBBP7; BRD2; SET; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCEA1; TCF3; SF1; VEZF1; KAT6A; AKAP17A; USP9X; OGT; KHSRP; TSC22D1; TAX1BP1; TAF1C; RPL23; MAGED1; NCOR2; MICAL2; BCLAF1; MAML1; THRAP3; DNAJB6; NAMPT; PCGF3; RBM14; HOXB13; HEXIM1; CAMKK2; RAI1; SUB1; GCN1; PHB2; TAB2; POGZ; PHF8; WWC1; NEDD4L; RYBP; SUZ12; NUP62; AUTS2; EHF; GNL3; PDCC4; HIPK2; NOP53; BICRA; HP1BP3; GMNN; WAC; CXXC5; SRRT; RTRAF; UIMC1; BANP; CHD7; YEATS2; SLC2A4RG; BBX; MRTFB; MAVS; ZNF350; NUCKS1; PAGR1; TBL1XR1; LBH; ZNF525; ZFP62

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
regulation of RNA biosynthetic process	BP	0.3980	119/299	4.8548	ACTN1; PARP1; AGT; AR; BCL6; ZFP36L1; ZFP36L2; C1QBP; CEBPB; KLF6; CTBP2; CUX1; DDX3X; DVL1; EPAS1; XRCC6; GOLGB1; HMGN1; FOXA1; HES1; HSF1; HSPA1A; HSPA8; DNAJB1; ILF3; JUN; MAGEA1; KMT2A; NFIA; NFIC; NONO; NOTCH2; NPAS2; YBX1; PFKM; MAP2K3; PSMA6; PSMB6; PSMD1; PSMD3; PSMD4; PSME1; PURB; RAN; RBBP7; BRD2; SET; SMARCA4; SMARCC2; SOX4; SP3; SP100; SREBF1; TCEA1; TCF3; SF1; VEZF1; KAT6A; AKAP17A; USP9X; OGT; KHSRP; TSC22D1; TAX1BP1; TAF1C; RPL23; MAGED1; NCOR2; MICAL2; BCLAF1; MAML1; THRAP3; DNAJB6; NAMPT; PCGF3; RBM14; HOXB13; HEXIM1; CAMKK2; RAI1; SUB1; GCN1; PHB2; TAB2; POGZ; PHF8; WWC1; NEDD4L; RYBP; SUZ12; NUP62; AUTS2; EHF; GNL3; PDCC4; HIPK2; NOP53; BICRA; HP1BP3; GMNN; WAC; CXXC5; SRRT; RTRAF; UIMC1; BANP; CHD7; YEATS2; SLC2A4RG; BBX; MRTFB; MAVS; ZNF350; NUCKS1; PAGR1; TBL1XR1; LBH; ZNF525; ZFP62

### 2.8.4 nmCRPC vs nmCSPC

Recall that we identified 3655 interesting/significant peptides for this contrast. Based on these peptides, the gene-set-analysis yields the following waterfall plot.



Interpretation for the waterfall plot remains the same as above. We also tabulate the enriched/overrepresented GO terms. The last column of the table shows the genes associated with proteins that have at least one significant peptide in the contrast or pairwise comparison.

Term	Ontology	set.mean	set.size	z.score	in.genes
chromatin	CC	0.9333	70/75	4.8602	ACTB; AR; CEBPB; DHX9; EZH2; MSH6; H1F0; HIST1H1C; HIST1H2AD; H3F3A; H3F3B; HDAC1; HMGB2; HMGN1; HMGN2; HNRNPC; HNRNPK; HSF1; EIF3E; JUN; JUNB; JUND; MCM7; MYC; PRM2; RAD21; RAN; RBBP4; RBBP7; UPF1; SMARCA1; SMARCA4; SMARCC2; TCF3; TCP1; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST2H2AC; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; HIST1H2AG; MTA1; MAGED1; H2AFY; NCOR2; MORF4L1; PARK7; CBX3; POGZ; PDS5A; SUZ12; NOP53; BICRA; HP1BP3; PHF10; H2AFJ; FAM111A; NUCKS1; HIST1H2AH; HIST1H2BK; HIST3H2A; H2AFV; H3F3C; HIST2H2AA4

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	BP	0.9059	77/85	4.6352	EIF3E; RPSA; RPL10A; UPF1; RPL3; RPL5; RPL6; RPL7; RPL7A; RPL8; RPL9; RPL10; RPL12; RPL13; RPL15; RPL17; RPL18; RPL18A; RPL19; RPL21; RPL22; RPL23A; RPL24; RPL26; RPL27; RPL30; RPL27A; RPL28; RPL29; RPL31; RPL32; RPL34; RPL35A; RPL37; RPL37A; RPL38; RPL39; RPL41; RPL36A; RPLP0; RPS2; RPS3; RPS3A; RPS4Y1; RPS6; RPS7; RPS8; RPS10; RPS11; RPS12; RPS13; RPS14; RPS15; RPS15A; RPS16; RPS17; RPS18; RPS19; RPS20; RPS23; RPS24; RPS25; RPS26; RPS27A; UBA52; RPL14; RPL23; EIF4A3; RBM8A; RPL35; CASC3; SMG1; RPL13A; RPL36; PABPC1; MAGOHB; SECISBP2
SRP-dependent cotranslational protein targeting to membrane	BP	0.8987	71/79	4.3187	RPSA; RPL10A; RPL3; RPL5; RPL6; RPL7; RPL7A; RPL8; RPL9; RPL10; RPL12; RPL13; RPL15; RPL17; RPL18; RPL18A; RPL19; RPL21; RPL22; RPL23A; RPL24; RPL26; RPL27; RPL30; RPL27A; RPL28; RPL29; RPL31; RPL32; RPL34; RPL35A; RPL37; RPL37A; RPL38; RPL39; RPL41; RPL36A; RPLP0; RPS2; RPS3; RPS3A; RPS4Y1; RPS6; RPS7; RPS8; RPS10; RPS11; RPS12; RPS13; RPS14; RPS15; RPS15A; RPS16; RPS17; RPS18; RPS19; RPS20; RPS23; RPS24; RPS25; RPS26; RPS27A; SRP14; SRPRA; UBA52; RPL14; RPL23; RPL35; TRAM1; RPL13A; RPL36
protein targeting to membrane	BP	0.8953	77/86	4.4486	AKT2; ANK3; RPSA; MYO1C; RPL10A; PRNP; RPL3; RPL5; RPL6; RPL7; RPL7A; RPL8; RPL9; RPL10; RPL12; RPL13; RPL15; RPL17; RPL18; RPL18A; RPL19; RPL21; RPL22; RPL23A; RPL24; RPL26; RPL27; RPL30; RPL27A; RPL28; RPL29; RPL31; RPL32; RPL34; RPL35A; RPL37; RPL37A; RPL38; RPL39; RPL41; RPL36A; RPLP0; RPS2; RPS3; RPS3A; RPS4Y1; RPS6; RPS7; RPS8; RPS10; RPS11; RPS12; RPS13; RPS14; RPS15; RPS15A; RPS16; RPS17; RPS18; RPS19; RPS20; RPS23; RPS24; RPS25; RPS26; RPS27A; SRP14; SRPRA; UBA52; RPL14; RPL23; RPL35; CHP1; TRAM1; RPL13A; RPL36; RAB3IP
nuclear-transcribed mRNA catabolic process	BP	0.8750	84/96	4.2762	ZFP36L2; DDX5; EIF3E; RPSA; RPL10A; CNOT2; UPF1; RPL3; RPL5; RPL6; RPL7; RPL7A; RPL8; RPL9; RPL10; RPL12; RPL13; RPL15; RPL17; RPL18; RPL18A; RPL19; RPL21; RPL22; RPL23A; RPL24; RPL26; RPL27; RPL30; RPL27A; RPL28; RPL29; RPL31; RPL32; RPL34; RPL35A; RPL37; RPL37A; RPL38; RPL39; RPL41; RPL36A; RPLP0; RPS2; RPS3; RPS3A; RPS4Y1; RPS6; RPS7; RPS8; RPS10; RPS11; RPS12; RPS13; RPS14; RPS15; RPS15A; RPS16; RPS17; RPS18; RPS19; RPS20; RPS23; RPS24; RPS25; RPS26; RPS27A; SSB; UBA52; CSDE1; RPL14; RPL23; EIF4A3; RBM8A; THRAP3; RPL35; CASC3; SMG1; RPL13A; RPL36; PABPC1; XRN1; MAGOHB; SECISBP2
chromosomal part	CC	0.8727	96/110	4.5488	ACTB; PARP1; AR; BCL6; CEBPB; CENPE; DDB1; DHX9; DYNC1L1; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HIST1H2AD; H3F3A; H3F3B; HDAC1; HMGB2; HMG1; HMG2; HNRNPC; HNRNPK; HSF1; EIF3E; JUN; JUNB; JUND; MCM3; MCM7; MYC; NKX3-1; PAFAH1B1; PHF2; PPP1CC; PPP2CB; PRM2; PURB; RAD21; RAN; RBBP4; RBBP7; UPF1; CLIP1; SEC13; SMARCA1; SMARCA4; SMARCC2; SP100; SSB; TCF3; TCP1; VCP; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST2H2AC; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; HIST1H2AG; MTA1; MAGED1; H2AFY; NCOR2; ARPC3; ARPC2; P3H4; MORF4L1; PARK7; CBX3; POGZ; PDS5A; SUZ12; ORC6; REPIN1; NOP53; BICRA; HP1BP3; GARI; PHF10; H2AFJ; THOC2; FAM111A; NUCKS1; MEAF6; HIST1H2AH; HIST1H2BK; HIST3H2A; H2AFV; H3F3C; HIST2H2AA4
chromosome organization	BP	0.8525	104/122	4.3118	ACTB; PARP1; BCL6; CENPE; DDB1; DDX1; DDX3X; DHX9; EZH2; XRCC6; H1F0; HIST1H1C; H3F3A; H3F3B; HDAC1; HMGB1; HMGB2; HMG1; HNRNPC; HNRNPK; HSP90AA1; IGF2; KPNB1; MCM7; KMT2A; MYC; NAP1L1; NOS1; NPM1; PHF2; PRM2; RAD21; RAD23B; RAN; RBBP4; RBBP7; UPF1; BRD2; RPS27A; SET; SMARCA1; SMARCA4; SMARCC2; SP100; SREBF1; TAF7; TCP1; TDG; UBA52; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; OGT; COPS3; EED; TRIP12; H2AFY; ZMPSTE24; PCGF3; RBM14; TADA3; CCT7; CCT4; CCT2; P3H4; MORF4L1; PHB2; CBX3; SMG1; POGZ; PHF8; PDS5A; TSPYL4; SUN1; RYBP; SUZ12; NUP62; BRD1; AUTS2; GNL3; HP1BP3; LEF1; UIMC1; ARID4B; GARI; XRN1; BANP; PHF10; NOP10; CHD7; YEATS2; ENY2; CHD8; ZNF462; NUCKS1; MEAF6; HDAC10; ING5; TSPYL5; HIST3H2A

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
drug binding	MF	0.8521	121/142	4.6789	ABAT; ACTB; ACTG1; AKT2; ASNS; ATP1A1; ATP1B1; ATP6V1A; ATP6AP1; ATP5PO; BMPR1B; DDR1; CBS; CENPE; CHKA; CSNK1D; CSNK1E; CYP1B1; DDX1; DDX3X; DDX5; DHX9; CYB5R3; DYNC1L1; FKBP2; FKBP5; XRCC6; MKNK2; MSH6; HBB; HK2; HMGB2; DNAJA1; HSPA1A; HSPA8; HSP90AA1; HSPD1; IARS; IGF1R; ILF2; ITPK1; KIF5C; MARS; MAT2A; MCM3; MCM7; MAP3K5; MT2A; MMUT; MYH9; MYH11; MYO1C; MYO6; NKTR; NME3; NOS1; PDPK1; PFKM; PGK1; PPP3CA; PPP3R1; PKN2; MAP2K3; PSMC1; RARS; UPF1; SGK1; SMARCA1; SMARCA4; TARS; TCP1; TDG; HSP90B1; VCP; CXCR4; PIP4K2B; ULK1; STK24; DGKD; DDX3Y; EIF4A3; THRAP3; FARSB; ABCC5; ATP9A; HIPK3; NAMPT; UBE2E3; CCT7; CCT4; CCT2; CAMKK2; HSPH1; FASTK; SNRNP200; SMG1; KIF13B; ATP2C1; EIF2AK1; HIPK2; DDX47; IP6K2; RTCB; RIPK4; DNAJA4; UBE2Q1; CHD7; MCCC1; ATP8B2; SCYL1; CHD8; WNK1; UBE2Z; DDX50; ATP13A3; MYO19; ALPK1; ACSS1; NEK9; ANKK1; NRBP2; ACAT1; AHCY; ZFP36L2; ENTPD6; DDX5; DHX9; GPI; H1F0; HINT1; HK2; HMGB1; HMGB2; HNRNPC; HNRNP; HPRT1; HSF1; HSPA1A; HSPA8; HSPB1; EIF3E; KPNB1; RPSA; LDHA; RPL10A; CNOT2; NPM1; YBX1; PDE4C; PFKM; PGK1; PSMA1; PSMA6; PSMB6; PSMB7; PSMC1; PSMD1; PSMD3; PSMD4; PSME1; RANBP2; UPF1; RNH1; RPL3; RPL5; RPL6; RPL7; RPL7A; RPL8; RPL9; RPL10; RPL12; RPL13; RPL15; RPL17; RPL18; RPL18A; RPL19; RPL21; RPL22; RPL23A; RPL24; RPL26; RPL27; RPL30; RPL27A; RPL28; RPL29; RPL31; RPL32; RPL34; RPL35A; RPL37; RPL37A; RPL38; RPL39; RPL41; RPL36A; RPLP0; RPS2; RPS3; RPS3A; RPS4Y1; RPS6; RPS7; RPS8; RPS10; RPS11; RPS12; RPS13; RPS14; RPS15; RPS15A; RPS16; RPS17; RPS18; RPS19; RPS20; RPS23; RPS24; RPS25; RPS26; RPS27A; SEC13; SET; SSB; TDG; UBA52; VCP; CSDE1; OGT; KHSRP; RPL14; RPL23; PUM1; EIF4A3; RBM8A; THRAP3; HNRNPR; SYNCRIP; PKP3; RPL35; CASC3; SMG1; RPL13A; NUP62; RPL36; SERBP1; NUPR1; PABPC1; SND1; SIDT2; YTHDF2; XRN1; DDIT4; MAGOHB; SECISBP2
nucleobase-containing compound catabolic process	BP	0.8293	136/164	4.4069	ACAT1; AHCY; ZFP36L2; ENTPD6; DDX5; DHX9; GPI; H1F0; HINT1; HK2; HMGB1; HMGB2; HNRNPC; HNRNP; HPRT1; HSF1; HSPA1A; HSPA8; HSPB1; EIF3E; KPNB1; RPSA; LDHA; RPL10A; CNOT2; NPM1; YBX1; PDE4C; PFKM; PGK1; PSMA1; PSMA6; PSMB6; PSMB7; PSMC1; PSMD1; PSMD3; PSMD4; PSME1; RANBP2; UPF1; RNH1; RPL3; RPL5; RPL6; RPL7; RPL7A; RPL8; RPL9; RPL10; RPL12; RPL13; RPL15; RPL17; RPL18; RPL18A; RPL19; RPL21; RPL22; RPL23A; RPL24; RPL26; RPL27; RPL30; RPL27A; RPL28; RPL29; RPL31; RPL32; RPL34; RPL35A; RPL37; RPL37A; RPL38; RPL39; RPL41; RPL36A; RPLP0; RPS2; RPS3; RPS3A; RPS4Y1; RPS6; RPS7; RPS8; RPS10; RPS11; RPS12; RPS13; RPS14; RPS15; RPS15A; RPS16; RPS17; RPS18; RPS19; RPS20; RPS23; RPS24; RPS25; RPS26; RPS27A; SEC13; SET; SSB; TDG; UBA52; VCP; CSDE1; OGT; KHSRP; RPL14; RPL23; PUM1; EIF4A3; RBM8A; THRAP3; HNRNPR; SYNCRIP; PKP3; RPL35; CASC3; SMG1; RPL13A; NUP62; RPL36; SERBP1; NUPR1; PABPC1; SND1; SIDT2; YTHDF2; XRN1; DDIT4; MAGOHB; SECISBP2
cellular nitrogen compound catabolic process	BP	0.8242	136/165	4.2752	ACAT1; AHCY; ZFP36L2; ENTPD6; DDX5; DHX9; GPI; H1F0; HINT1; HK2; HMGB1; HMGB2; HNRNPC; HNRNP; HPRT1; HSF1; HSPA1A; HSPA8; HSPB1; EIF3E; KPNB1; RPSA; LDHA; RPL10A; CNOT2; NPM1; YBX1; PDE4C; PFKM; PGK1; PSMA1; PSMA6; PSMB6; PSMB7; PSMC1; PSMD1; PSMD3; PSMD4; PSME1; RANBP2; UPF1; RNH1; RPL3; RPL5; RPL6; RPL7; RPL7A; RPL8; RPL9; RPL10; RPL12; RPL13; RPL15; RPL17; RPL18; RPL18A; RPL19; RPL21; RPL22; RPL23A; RPL24; RPL26; RPL27; RPL30; RPL27A; RPL28; RPL29; RPL31; RPL32; RPL34; RPL35A; RPL37; RPL37A; RPL38; RPL39; RPL41; RPL36A; RPLP0; RPS2; RPS3; RPS3A; RPS4Y1; RPS6; RPS7; RPS8; RPS10; RPS11; RPS12; RPS13; RPS14; RPS15; RPS15A; RPS16; RPS17; RPS18; RPS19; RPS20; RPS23; RPS24; RPS25; RPS26; RPS27A; SEC13; SET; SSB; TDG; UBA52; VCP; CSDE1; OGT; KHSRP; RPL14; RPL23; PUM1; EIF4A3; RBM8A; THRAP3; HNRNPR; SYNCRIP; PKP3; RPL35; CASC3; SMG1; RPL13A; NUP62; RPL36; SERBP1; NUPR1; PABPC1; SND1; SIDT2; YTHDF2; XRN1; DDIT4; MAGOHB; SECISBP2
heterocycle catabolic process	BP	0.8242	136/165	4.2752	ACAT1; AHCY; ZFP36L2; ENTPD6; DDX5; DHX9; GPI; H1F0; HINT1; HK2; HMGB1; HMGB2; HNRNPC; HNRNP; HPRT1; HSF1; HSPA1A; HSPA8; HSPB1; EIF3E; KPNB1; RPSA; LDHA; RPL10A; CNOT2; NPM1; YBX1; PDE4C; PFKM; PGK1; PSMA1; PSMA6; PSMB6; PSMB7; PSMC1; PSMD1; PSMD3; PSMD4; PSME1; RANBP2; UPF1; RNH1; RPL3; RPL5; RPL6; RPL7; RPL7A; RPL8; RPL9; RPL10; RPL12; RPL13; RPL15; RPL17; RPL18; RPL18A; RPL19; RPL21; RPL22; RPL23A; RPL24; RPL26; RPL27; RPL30; RPL27A; RPL28; RPL29; RPL31; RPL32; RPL34; RPL35A; RPL37; RPL37A; RPL38; RPL39; RPL41; RPL36A; RPLP0; RPS2; RPS3; RPS3A; RPS4Y1; RPS6; RPS7; RPS8; RPS10; RPS11; RPS12; RPS13; RPS14; RPS15; RPS15A; RPS16; RPS17; RPS18; RPS19; RPS20; RPS23; RPS24; RPS25; RPS26; RPS27A; SEC13; SET; SSB; TDG; UBA52; VCP; CSDE1; OGT; KHSRP; RPL14; RPL23; PUM1; EIF4A3; RBM8A; THRAP3; HNRNPR; SYNCRIP; PKP3; RPL35; CASC3; SMG1; RPL13A; NUP62; RPL36; SERBP1; NUPR1; PABPC1; SND1; SIDT2; YTHDF2; XRN1; DDIT4; MAGOHB; SECISBP2

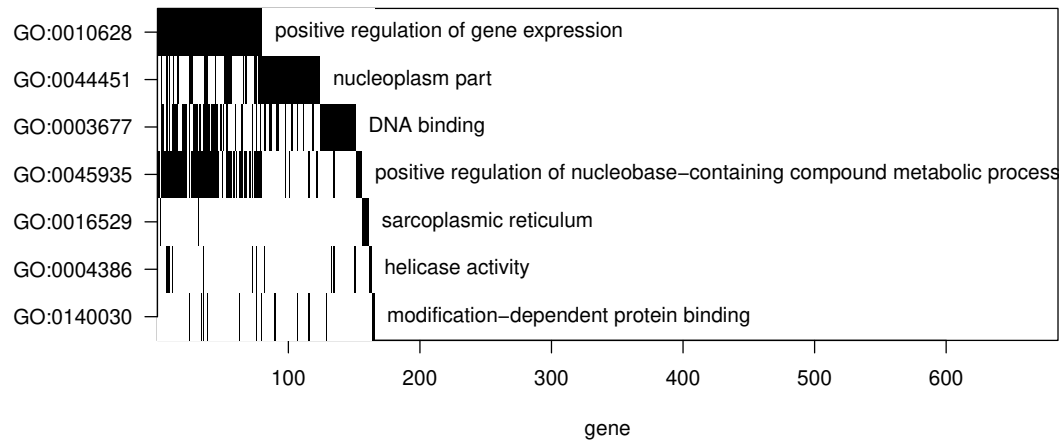
(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
DNA binding	MF	0.8216	152/185	4.4818	ACTB; ADAR; PARP1; APLP2; APP; AR; ATF4; BCL6; ZFP36L2; CEBPB; CUX1; DDB1; DDX1; DDX3X; DHX9; EEF1D; EPAS1; EZH2; XRCC6; GOLGB1; MSH6; GTF2I; GTF3A; H1F0; HIST1H1C; H3F3A; H3F3B; HDAC1; HMGB1; HMGB2; HMGN1; HMGN2; HNRNPC; HNRNPD; HNRNPK; HSF1; HSPD1; RBPJ; ILF2; ILF3; JUN; JUNB; JUND; MCM3; MCM7; KMT2A; MYC; NACA; NCL; NFIA; NFE2L1; NFIB; NFIC; NFIL3; NKX3-1; NONO; NPAS2; NPM1; YBX1; NUCB2; PA2G4; PCBP1; PNN; PRM2; PURB; RAD23B; RBBP4; UPP1; RPL6; RPL7; RPS3; RPS15; SET; SMARCA1; SMARCA4; SMARCC2; SON; SOX4; SP3; SP100; SREBF1; SSRP1; TAF7; TCEA1; TCF3; TDG; NR2F2; ZNF24; ZKSCAN1; VEZF1; ZFAND5; KAT6A; TAF15; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; KHSRP; DDX3Y; EDF1; EED; TAF1C; MTA1; H2AFY; BCLAF1; THRAP3; DNAJB6; AKAP9; RBM5; SRRM1; ZMPSTE24; HOXB13; KHDRBS1; RAI1; ZNF275; FOXJ3; TCF25; SMG1; RYBP; SUZ12; LSM14A; EHF; NUPR1; FOXP1; REPIN1; IRX4; HP1BP3; SIDT2; LEF1; CXXC5; TDP2; SRRT; XRN1; ZFAND6; BANP; STRBP; CHD7; ZNF395; BBX; SCYL1; CHD8; ZNF350; NUCKS1; IRX3; TBL1XR1; HIST3H2A; ZNF664; CREB3L4; ZMAT2; ZNF525; H3F3C
carbohydrate derivative binding	MF	0.8192	145/177	4.2962	ACTB; ACTG1; AKT2; APLP2; APP; ARF1; ARF4; ASNS; ATP1A1; ATP1B1; ATP6V1A; ATP6AP1; BMPR1B; DDR1; CENPE; CHKA; CSNK1D; CSNK1E; CTSB; DDX1; DDX3X; DDX5; DHX9; CYB5R3; DYNC1LI2; DPYSL3; EEF1A1; EIF5; XRCC6; GNAQ; GNAS; MKNK2; MSH6; GUCY1A1; HK2; HMGB1; DNAJA1; HSPA1A; HSPA8; HSP90AA1; HSPD1; IARS; IGF1R; ILF2; ITPK1; KIF5C; KRAS; LRPAP1; MARS; MAT2A; MCM3; MCM7; MDK; MAP3K5; MYH9; MYH11; MYO1C; MYO6; NME3; NOS1; PAFAH1B1; PDPK1; PFKM; PGK1; PRKAR2A; PKN2; MAP2K3; PRNP; PSMA1; PSMC1; PTPRF; RAB5A; RAN; RAP1B; RARS; UPP1; RPL22; RPL29; SGK1; SMARCA1; SMARCA4; SRPRA; TARS; TCP1; TDG; HSP90B1; VCP; DAP3; MANF; PIP4K2B; ULK1; STK24; DGKD; DDX3Y; ADGRG1; EIF4A3; MFN2; THRAP3; FARSB; ABCC5; ATP9A; HIPK3; ECI2; UBE2E3; CCT7; CCT4; CCT2; CAMKK2; RRAGA; HSPH1; FASTK; RAB35; ADAMTS5; SNRNP200; SMG1; KIF13B; GTPBP4; GNL3; ATP2C1; EIF2AK1; HIPK2; SAR1B; HSD17B12; DDX47; IP6K2; RTCB; RIPK4; DNAJA4; UBE2Q1; CHD7; MCCC1; ATP8B2; SCYL1; CHD8; RRAGC; WNK1; UBE2Z; DDX50; ATP13A3; MYO19; ALPK1; ACSS1; NEK9; ANKK1; NRBP2
negative regulation of gene expression	BP	0.8022	223/278	4.9399	A2M; ADAR; PARP1; APP; AR; ATF4; BCL6; ZFP36L2; CAST; CEBPB; CTBP2; DDX3X; DDX5; DHX9; EIF4EBP2; EZH2; XRCC6; H1F0; HIST1H1C; H3F3A; H3F3B; HDAC1; HMGB1; HMGB2; HNRNPC; HNRNPD; HNRNPK; HSF1; HSPA1A; HSPA8; HSPB1; DNAJB1; IGF2; RBPJ; ILF3; EIF3E; JUN; RPSA; LDLR; LIMS1; CAPRIN1; MYC; NCL; RPL10A; NFIB; NFIC; NFIL3; NKX3-1; NONO; CNOT2; NOTCH2; NPM1; YBX1; PA2G4; PHF2; PPP3CA; PRNP; PSMA1; PSMA6; PSMB6; PSMB7; PSMC1; PSMD1; PSMD3; PSMD4; PSME1; PURB; RAN; RANBP2; RBBP4; RBBP7; UPP1; RNH1; RPL3; RPL5; RPL6; RPL7; RPL7A; RPL8; RPL9; RPL10; RPL12; RPL13; RPL15; RPL17; RPL18; RPL18A; RPL19; RPL21; RPL22; RPL23A; RPL24; RPL26; RPL27; RPL30; RPL27A; RPL28; RPL29; RPL31; RPL32; RPL34; RPL35A; RPL37; RPL37A; RPL38; RPL39; RPL41; RPL36A; RPLP0; RPS2; RPS3; RPS3A; RPS4Y1; RPS6; RPS7; RPS8; RPS10; RPS11; RPS12; RPS13; RPS14; RPS15; RPS15A; RPS16; RPS17; RPS18; RPS19; RPS20; RPS23; RPS24; RPS25; RPS26; RPS27A; SEC13; SET; SRSF4; SRSF7; SMARCA4; SMARCC2; SP3; SP100; SREBF1; SSB; TAF7; TCF3; TDG; TMBIM6; NR2F2; TXN; UBA52; ZNF24; CSDE1; KAT6A; FXR1; USP9X; HIST1H4C; HIST1H4L; KHSRP; EED; RPL14; RPL23; MAGED1; TMEM59; H2AFY; NCOR2; PUM1; BCLAF1; EIF4A3; RBM8A; THRAP3; DNAJB6; HNRNPR; ZMPSTE24; SFP18; CNPY2; N4BP2L2; HOXB13; SYNCRIP; HEXIM1; KHDRBS1; CELF1; SRSF10; PKP3; RPL35; PARK7; PHB2; CBX3; CASC3; TCF25; SMG1; PHF8; WWC1; NEDD4L; RYBP; SUZ12; RPL13A; NUP62; RPL36; SERBP1; PABPC1; SND1; FOXP1; EIF2AK1; PDCC4; HIPK2; NOP53; ZNF706; LEF1; YTHDF2; CXXC5; SRRT; PTRH2; UIMC1; XRN1; MAGOHB; YEATS2; VPS35; CHD8; ZNF350; NIBAN2; SECISBP2; TBL1XR1; HDAC10



### 2.8.5 nmCSPC vs new\_dx

Recall that we identified 637 interesting/significant peptides for this contrast. Based on these peptides, the gene-set-analysis yields the following waterfall plot.



Interpretation for the waterfall plot remains the same as above. We also tabulate the enriched/overrepresented GO terms. The last column of the table shows the genes associated with proteins that have at least one significant peptide in the contrast or pairwise comparison.

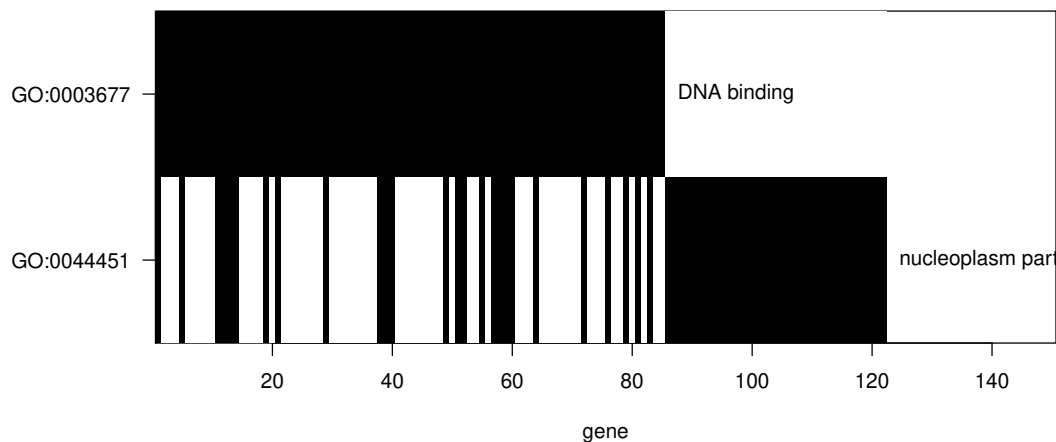
Term	Ontology	set.mean	set.size	z.score	in.genes
sarcoplasmic reticulum	CC	1.0000	7/7	4.3065	ANK3; HK2; NOS1; SRI; MANF; CHERP; RASD1
promoter-specific chromatin binding	MF	1.0000	9/9	4.8866	DDX5; DHX9; EZH2; HDAC1; HSF1; NFE2L1; H2AFY; SUZ12; CHD7
helicase activity	MF	0.7222	13/18	4.2777	DDX1; DDX3X; DDX5; DHX9; EIF4A1; XRCC6; MCM7; UPF1; SMARCA4; DDX3Y; CHD7; CHD8; DDX50
modification-dependent protein binding	MF	0.7222	13/18	4.2777	MSH6; DNAJB2; KMT2A; PHF2; BRD2; SMARCA4; TAF7; TAB2; PHF8; SUZ12; UIMC1; CHD8; ING5
chromatin binding	MF	0.5283	28/53	4.2125	AR; BCL6; CTBP2; DDX1; DDX5; DHX9; EZH2; MSH6; HDAC1; HNRNPD; HSF1; JUN; NFIA; NFE2L1; NONO; YBX1; UPF1; BRD2; SMARCA4; SP3; MTA1; H2AFY; PHF8; SUZ12; AUTS2; HP1BP3; CHD7; CHD8
nucleoplasm part	CC	0.5109	70/137	6.5131	ADD1; AR; DDX1; DDX3X; DHX9; EZH2; HDAC1; HSF1; HSPA1A; NBR1; KMT2A; MYO1C; MYO6; HNRNPM; NONO; PNN; PKN2; RBBP7; BRD2; RPA1; SON; SP3; SP100; TAF7; TDG; TPP2; U2AF1; SF1; KAT6A; AKAP17A; MTA1; TRIP12; RBM39; NCOR2; PUM1; BCLAF1; THRAP3; HIPK3; ALYREF; SRRM1; SAP18; RBM14; TADA3; UBOX5; ELL2; FBNP4; SF3B1; SUZ12; MORC3; SRRM2; BRD1; PNISR; VIRMA; GNL3; HP1BP3; ARL6IP4; PCF11; NOP58; UIMC1; NXF2; ZMIZ1; THOC2; CHD8; ZNF350; MEAF6; PAGR1; TBL1XR1; LAS1L; HDAC10; ING5
nuclear body	CC	0.5049	52/103	5.4300	ADD1; AR; DDX1; DDX3X; DHX9; HSF1; HSPA1A; NBR1; MYO1C; HNRNPM; NONO; PNN; PKN2; BRD2; RPA1; SON; SP3; SP100; TDG; TPP2; U2AF1; SF1; KAT6A; AKAP17A; TRIP12; RBM39; NCOR2; PUM1; BCLAF1; THRAP3; HIPK3; ALYREF; SRRM1; SAP18; RBM14; UBOX5; FBNP4; SF3B1; SUZ12; MORC3; SRRM2; BRD1; PNISR; VIRMA; GNL3; HP1BP3; ARL6IP4; NOP58; UIMC1; ZMIZ1; THOC2; ZNF350
DNA binding	MF	0.4216	78/185	4.7975	ADAR; AR; ATF4; BCL6; CUX1; DDX1; DDX3X; DHX9; EEZF1; EZH2; XRCC6; GOLGB1; MSH6; GTF2I; GTF3A; HDAC1; HNRNPD; HES1; HSF1; ILF2; ILF3; JUN; JUNB; LBR; MCM7; KMT2A; NACA; NCL; NFIA; NFE2L1; NFIC; NONO; YBX1; PNN; PURB; UPF1; RPA1; RPS27; SMARCA4; SON; SP3; SP100; TAF7; TDG; NR2F2; ZFP36; ZNF24; ZKSCAN1; VEZF1; KAT6A; TAF15; DDX3Y; TAF1C; MTA1; H2AFY; BCLAF1; THRAP3; AKAP9; SRRM1; KHDRBS1; SMG1; RYBP; SUZ12; REPIN1; IRX4; HP1BP3; SRRT; IFT57; CHD7; BBX; SCYL1; CHD8; ZNF350; ZSCAN18; TBL1XR1; ZNF587; ZNF664; ZNF525

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
positive regulation of RNA metabolic process	BP	0.4136	67/162	4.2046	AGT; AR; ATF4; CDKN1C; CTBP2; DDX3X; DDX5; DHX9; FLT3LG; XRCC6; HDAC1; HNRNPD; HES1; HSF1; HSPA1A; HSPA8; IGF2; ILF2; ILF3; JUN; JUNB; KMT2A; MYO6; NCL; NFIA; NFE2L1; NFIC; NOS1; YBX1; PHF2; UPP1; TRA2B; SMARCA4; SP3; SP100; TAF7; NR2F2; UBA52; ZFP36; ZNF24; VEZF1; KAT6A; TAF15; FZD4; MTA1; MAGED1; PUM1; BCLAF1; THRAP3; ALYREF; RBM14; TADA3; CAMKK2; GCN1; PHF8; WWC1; RYBP; AUTS2; GNL3; YTHDF2; CHD7; ZMIZ1; CHD8; NIBAN2; PAGR1; TBL1XR1; ING5
positive regulation of nucleobase-containing compound metabolic process	BP	0.4088	74/181	4.3255	AGT; AR; ATF4; CDKN1C; CTBP2; DDX3X; DDX5; DHX9; FLT3LG; XRCC6; HDAC1; HNRNPD; HES1; HSF1; HSPA1A; HSPA8; HSP90AA1; IGF2; ILF2; ILF3; JUN; JUNB; KMT2A; COX2; MYO6; NCL; NFIA; NFE2L1; NFIC; NOS1; YBX1; PHF2; UPP1; TRA2B; SMARCA4; SP3; SP100; TAF7; TCP1; NR2F2; UBA52; ZFP36; ZNF24; VEZF1; KAT6A; TAF15; USP9X; FZD4; MTA1; MAGED1; PUM1; BCLAF1; THRAP3; ALYREF; RBM14; TADA3; CCT2; CAMKK2; GCN1; PHF8; WWC1; RYBP; AUTS2; GNL3; YTHDF2; UIMC1; CHD7; ZMIZ1; CHD8; NIBAN2; PAGR1; TBL1XR1; HDAC10; ING5
positive regulation of gene expression	BP	0.4061	80/197	4.4491	ADAR; AGT; ANK3; AR; ATF4; CDKN1C; CTBP2; DDX3X; DDX5; DHX9; FLT3LG; XRCC6; HDAC1; HNRNPD; HES1; HSF1; HSPA1A; HSPA8; IGF2; ILF2; ILF3; JUN; JUNB; LDLR; KMT2A; MYO1C; MYO6; NCL; NFIA; NFE2L1; NFIC; NOS1; YBX1; PHF2; TRA2B; SMARCA4; SP3; SP100; TAF7; NR2F2; UBA52; ZFP36; ZNF24; VEZF1; KAT6A; TAF15; FZD4; EIF3D; TAF1C; MAGED1; H2AFY; NCOR2; BCLAF1; THRAP3; ALYREF; RBM14; TADA3; SYNCRIP; CAMKK2; KHDRBS1; GCN1; PKP3; PHF8; WWC1; RYBP; SF3B1; AUTS2; GNL3; COA3; RPS27L; YTHDF2; DNAJA4; CHD7; VPS35; ZMIZ1; CHD8; NIBAN2; PAGR1; TBL1XR1; ING5

### 2.8.6 new\_dx vs normal

Recall that we identified 686 interesting/significant peptides for this contrast. Based on these peptides, the gene-set-analysis yields the following waterfall plot.



Interpretation for the waterfall plot remains the same as above. We also tabulate the enriched/overrepresented GO terms. The last column of the table shows the genes associated with proteins that have at least one significant peptide in the contrast or pairwise comparison.

Term	Ontology	set.mean	set.size	z.score	in.genes
nucleoplasm part	CC	0.4599	63/137	4.2887	ACTB; ADD1; AR; DDX1; DHX9; EPAS1; EZH2; H1F0; HDAC1; HSPA1A; NBR1; KMT2A; AFDN; MYO6; PCBP1; PNN; POLR2E; BRD2; SRSF1; SNRNP70; SNRPB2; SON; SP3; SP100; TPP2; U2AF1; KAT6A; EED; MTA1; TRIP12; RBM39; NCOR2; PUM1; BCLAF1; MAML1; THRAP3; HIPK3; SRRM1; RBM14; TADA3; ATXN2L; CASC3; FBNP4; SF3B1; SRRM2; BRD1; PNISR; VIRMA; TDP2; PCF11; NOP58; UIMC1; YEATS2; SLC2A4RG; ENY2; ZMIZ1; THOC2; NUFIP2; CHD8; MEAF6; TBL1XR1; HDAC10; GTF3C6
DNA binding	MF	0.4595	85/185	5.0673	ACTB; ADAR; PARP1; APLP2; AR; ATF4; BCL6; ZFP36L2; CEBPD; CUX1; DDX1; DHX9; EPAS1; EZH2; XRCC6; GOLGB1; MSH6; GTF2I; H1F0; HIST1H1C; HDAC1; HNRNPD; HSPD1; RBPJ; ILF3; JUNB; JUND; MCM3; KMT2A; NCL; NFIA; NFE2L1; NFIB; NFIC; NME1; NME2; NPAS2; PCBP1; PNN; POLR2E; UPF1; RPL6; RPL7; RPS3; RPS15; SMARCA1; SMARCA4; SMARCC2; SON; SOX4; SP3; SP100; SSRP1; ZFP36; KAT6A; TAF15; DDX3Y; EED; MTA1; BCLAF1; THRAP3; DNAJB6; AKAP9; RBM5; SRRM1; ZMPSTE24; HOXB13; RAI1; SMG1; NUPR1; REPIN1; SIDT2; TDP2; SRRT; CHD7; ZNF395; SLC2A4RG; BBX; SCYL1; CHD8; ZSCAN18; TBL1XR1; ZNF587; GTF3C6; CREB3L4

### 3 Section II: Antibody Responses over Time after Treatments

#### 3.1 Preamble

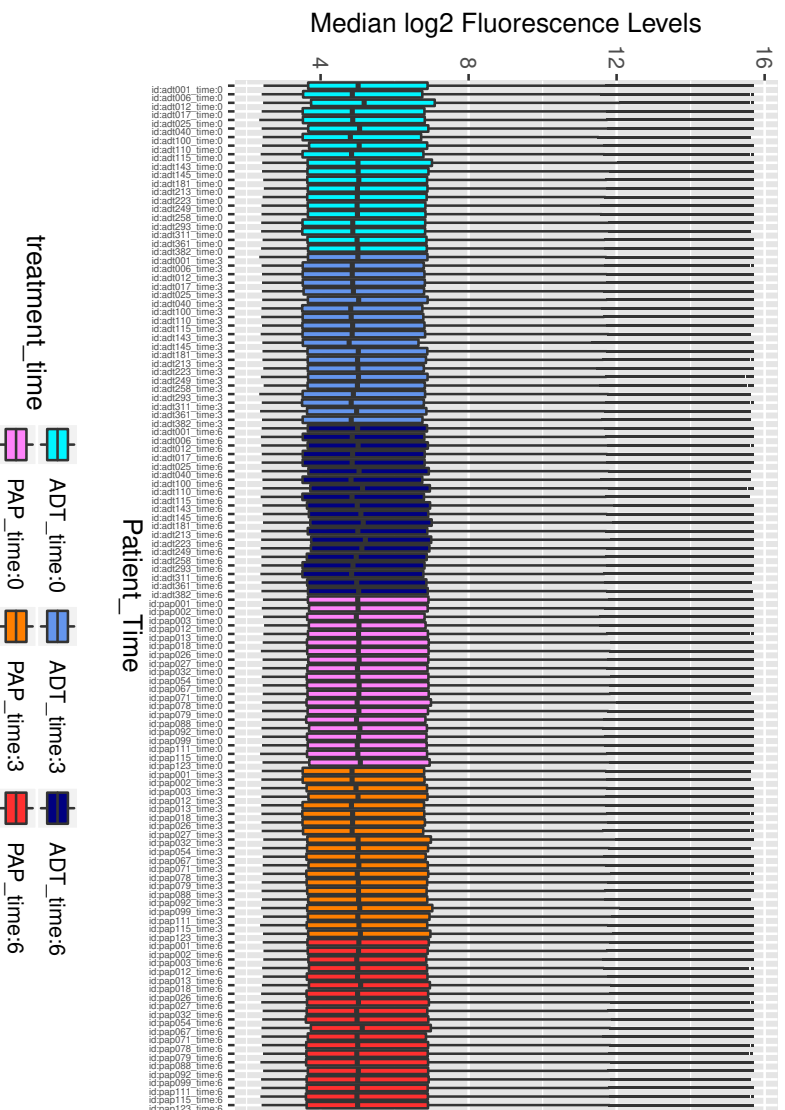
Now, we want to investigate how treatments induce changes in antibody repertoires in individuals over time. To address this question, we used serum samples available from the 40 patients with nmCSPC who were treated with one of two therapies. 20 patients received standard androgen deprivation therapy (ADT; GnRh analogue given every 3 months), and 20 patients received a DNA vaccine encoding prostatic-acid phosphatase (PAP; pTVG-HP given every 14 days for 6 administrations). Samples were collected (3 replicates) from each of these patients at baseline, 3 months, and 6 months following initiation of treatment.

Again, we take  $\log_2$  transformation on the fluorescence levels prior to subsequent steps in our analysis.

#### 3.2 Normalization of Fluorescence Data

In order to verify normalization of the fluorescence level, we also plot the boxplots of median (across replicates)  $\log_2$  fluorescence level of all peptides for each patient at each time point.

Boxplots of Peptide Fluorescence Levels for Patients at 3 time points



It appears that the fluorescence levels of the peptides are normalized.

### 3.3 Tests on Time Effect

To analyze how the two treatments (PAP or ADT) induces changes in antibody responses in individuals over time, we analyze the time effect of treatments in the patients for this study separately for the PAP group and for the ADT group. Specifically, for each peptide and each of the two treatment groups (consisting of 20 patients), we fit the following linear mixed model

$$y_{i\tau} = \beta_0 + \beta_1 \tau + b_{0i} + b_{1i} \tau + \epsilon_i,$$

where

- $y_{i\tau}$  be the median fluorescence level on  $\log_2$  scale for the  $i^{th}$  patient at time  $\tau$
- $i = 1, \dots, 20$  and  $\tau = 0, 3$  or 6 months
- $\beta_0$  = the baseline antibody response level for all patients in the treatment group
- $b_{0i}$  is the random intercept of the  $i^{th}$  patient
- $b_{1i}$  is the random slope of the  $i^{th}$  patient
- $\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N_3 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 & 0 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_2^2 \end{bmatrix} \right)$

For each peptide and for each treatment group, we test the following:

- $H_0$ :  $\beta_1 = 0$ , ie. Treatment does not induce changes in antibody response over time.
- $H_1$ :  $\beta_1 \neq 0$ , ie. Treatment induces changes in antibody response over time.

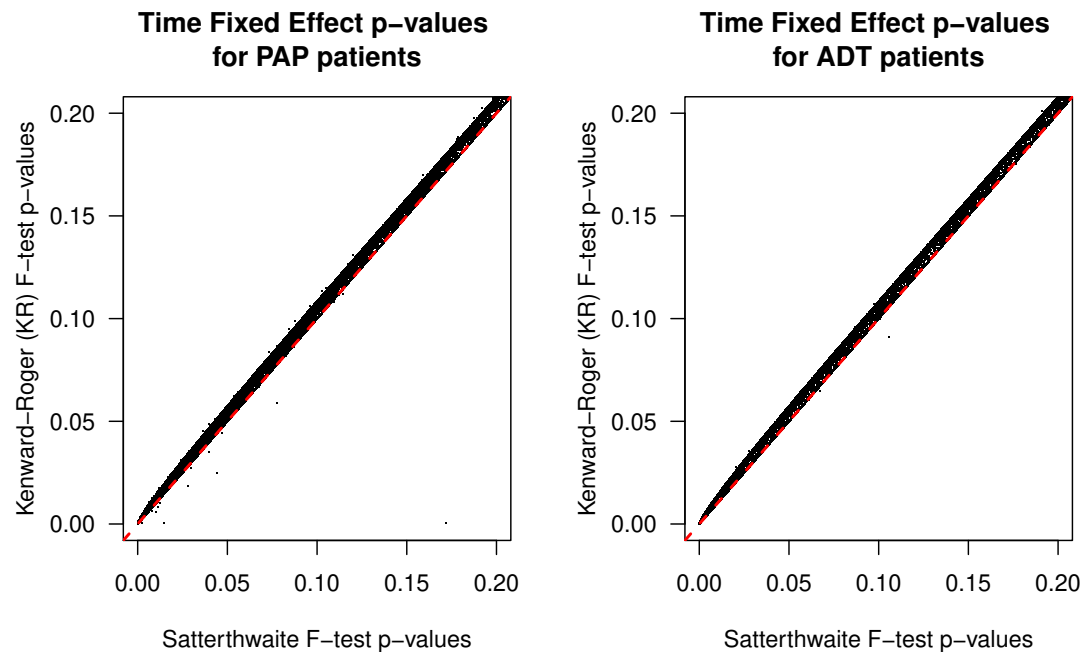
**Rationale of the model:** Since each patient has multiple measurements, the random effects of the mixed model allow us to capture the within-subject interdependencies. Every patient's antibody response is unique and possibly changes across time due to individual circumstances, so we want our model to include random intercept (representing patient-specific randomness) and random slope (of time). Since measurements were taken across only 3 time points, we refrain from considering more complicated terms involving time effect (eg. higher-order polynomial function of time).

**Model-fitting and Test Statistics:** Hypothesis testing in linear mixed-models is still an active area of research. Due to the large number of peptides, any non-parametric tests like permutation tests (shuffling treatment identifiers among patients by respecting time blocks) are prohibitively expensive in terms of computation. There are three usual parametric approximate tests for fixed effects in linear mixed models [Luke, 2017]:

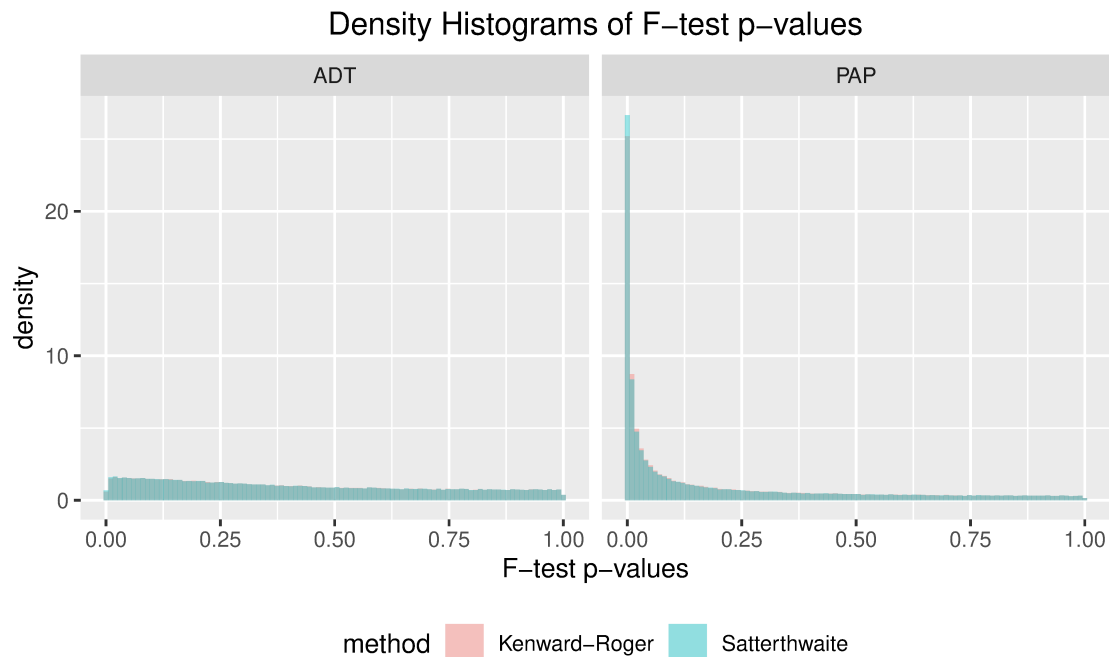
- Kenward-Roger (KR) approximate F-test, with model estimates fitted using the Restricted Maximum Likelihood (REML) approach,
- Satterthwaite approximate F-test, with model estimates also fitted with REML, and
- likelihood ratio test (LRT), with model estimates fitted using the usual Maximum Likelihood (ML) approach.

Roughly, unlike the ML approach, the REML method gives unbiased estimate of  $\hat{\Sigma}$ . This is imperative, since  $\hat{\Sigma}$  feeds into the F-test calculations. Both Kenward\_Roger and Satterthwaite approximations aim to adjust the degrees-of-freedom in the F-test to account for the additional estimation of covariance terms in the random effects of mixed models, as compared to a vanilla F-test in basic linear models [Luke, 2017]. Likelihood ratio test is only meaningful when parameter estimates are fitted with ML, otherwise the likelihood ratio test statistic may even end up as a negative value.

The consensus is that likelihood ratio test (LRT) could be slightly more liberal than the other two methods [Luke, 2017]. KR and Satterthwaite approximations usually give comparable results, and the Satterthwaite method is also the default linear-mixed-model setting in SAS and in the R package *lmerTest* [Kuznetsova et al., 2017]. We deploy only the KR and Satterthwaite approximate F-tests and compare their p-values. We zoom-in the plots to consider p-values  $\leq 0.2$ .



It appears that the KR-approximation is slightly more conservative than the Satterthwaite approximation in most cases. We also plot the density histograms of both sets of F-test p-values for both treatment groups at the same scale.



Again, the KR F-test p-values are slightly more conservative than the Satterthwaite approximation for the PAP patients. Where the ADT group is concerned, the p-value histograms are relatively flat for both approximation methods. After applying the BH method, no peptides from the ADT group are found to be significant even at 20% FDR for either of the two approximation methods. For the PAP group, we tabulate the peptide counts at various BH FDR thresholds.

BH_FDR_thresholds	Peptide_counts_KR	Peptide_counts_Satterthwaite
0.01	35039	39071
0.02	45356	48858
0.03	52742	55816
0.04	58713	61466
0.05	63747	66252

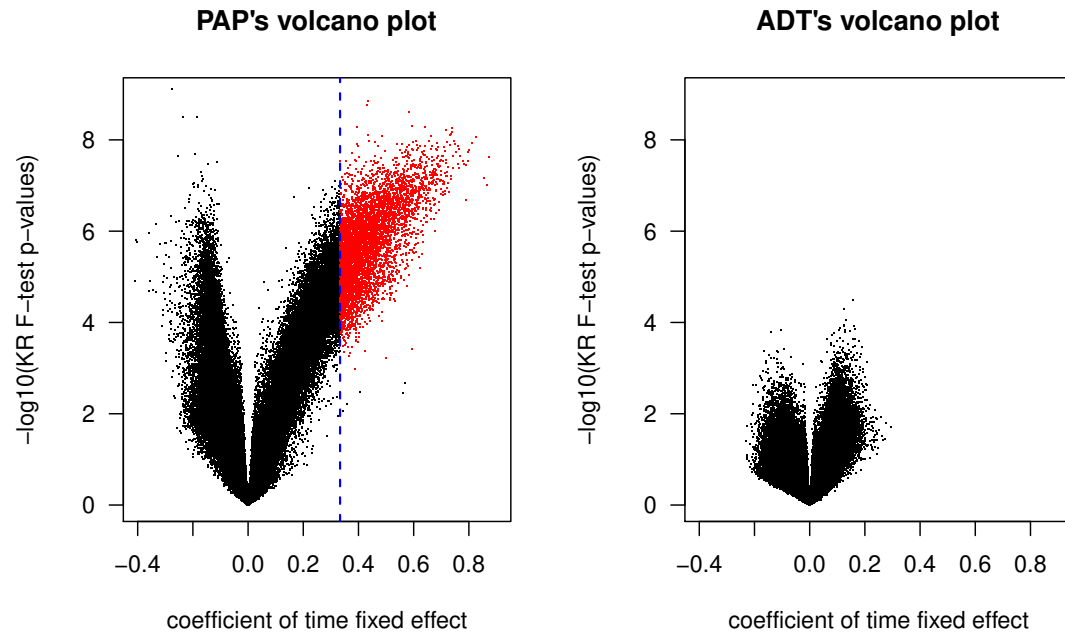
For instance, out of the 63747 peptides at 5% FDR based on KR p-values, 35034 of them are among the 63745 peptides at 5% FDR based on the Satterthwaite p-values. Where the PAP group is concerned, we will only consider peptides that meet **5% BH FDR cut-off for both KR and Satterthwaite methods**. In addition, we are only interested in peptides that demonstrate **at least two-fold increase in fluorescence after every 3-months, ie.  $\beta_1 \geq 0.3333$** . There are 5680 peptides which meet these two requirements. The list of these peptides is also exported to the sheet “*PAP\_Longitudinal*” in the Excel file “*09\_Significant\_Peptides.xlsx*”.

### 3.4 Visualization

We first obtain the volcano plots of  $-\log_{10}$  (KR) F-test p-values versus  $\hat{\beta}_1$  for the PAP and ADT groups at the same scale. The 5680 significant peptides that meet the 5% BH FDR and estimated time effect cut-offs

are colored red. The vertical blue dashed line represents the 0.3333 threshold of estimated coefficient of time fixed effect.

The volcano plots corroborate with the patterns we observe from the p-value density histograms. More patients in the PAP groups exhibit more significantly higher changes in antibody responses over time.



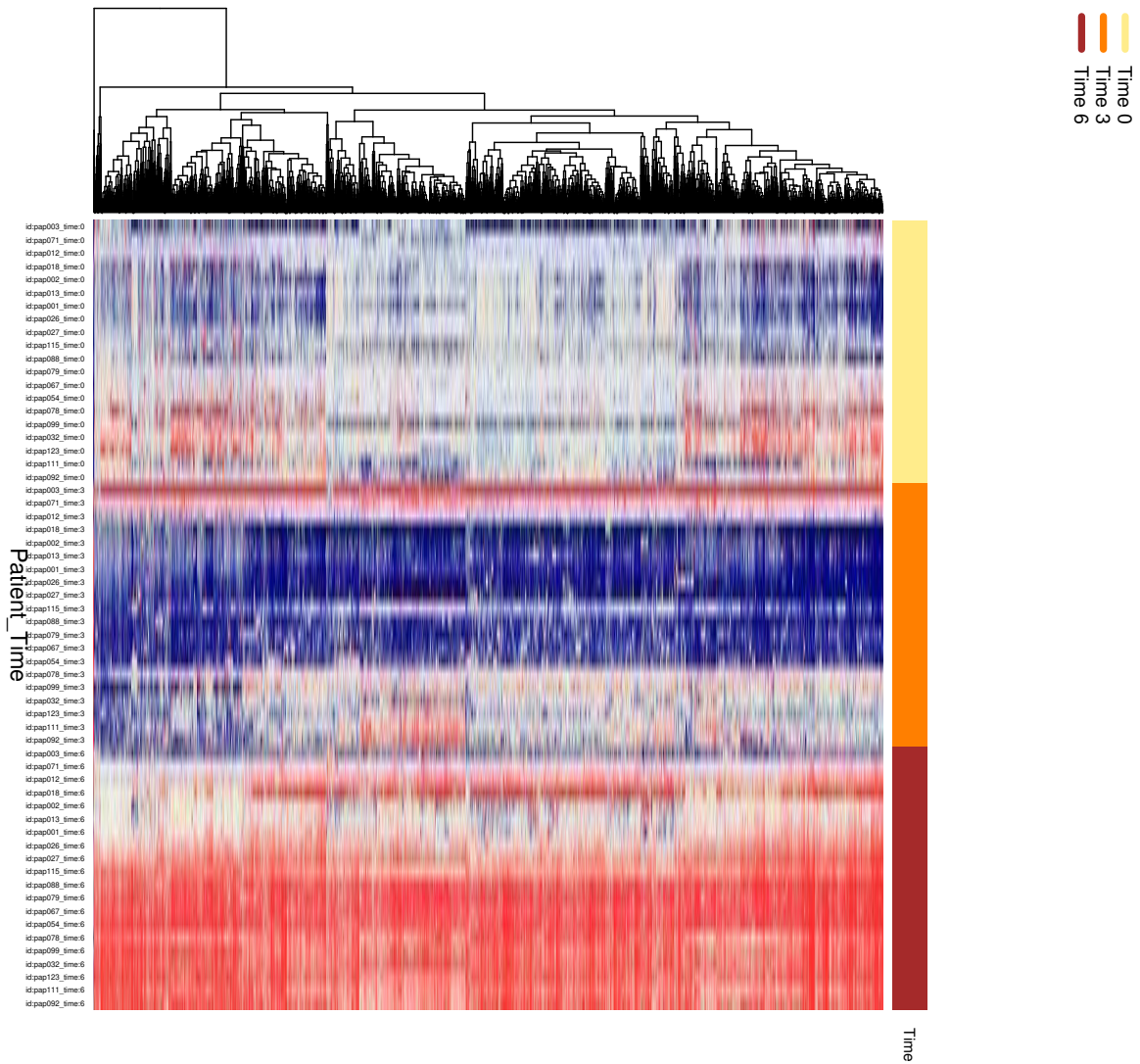
Next, we will illustrate the time fixed effect for the PAP patients among these 5680 significant peptides via a heatmap. First, we obtain the estimated residuals from null model, ie. for each of the 5680 peptides among the PAP patients, we fit the model

$$y_{i\tau} = \beta_0 + b_{0i} + b_{1i}\tau + \epsilon_i,$$

(which corresponds to setting  $\beta_1 = 0$ ) and obtain the residuals for the PAP patients. All the other terms in the model are left unchanged so they retain the same explanation from above. Any pattern among these residuals will demonstrate the time fixed effect not covered in the null model.

The fluorescence residuals are then winsorized at -1.7 and 1.7, which correspond to roughly bottom 5% and top 5% of the residuals. We then use these winsorized fluorescence residuals to plot the heatmap without any row-wise scaling. The color scheme of the heatmap is specified as navy for -1.7 which gradually transitions to firebrick for 1.7. Note that the order of the patients are the same across the 3 time points to show how these individuals' antibody response changes over time. Overall, the heatmap clearly illustrates that the individuals' antibody response levels increase over time.

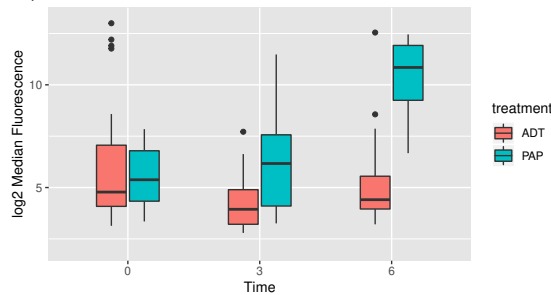
The heatmap also illustrates the fact that each patient's antibody response across the 3 time points is still different. For example, fluorescence levels of (most of) the 5680 peptides for patients with ID *pap078*, *pap099*, *pap032* and *pap111* are pretty 'flat' across the first 3 months before rising profusely in the next 3 months. Patients with ID *pap018*, *pap002*, *pap013*, *pap001*, *pap026*, *pap027*, *pap115*, *pap088*, *pap079* and *pap067* exhibit a drop in antibody response level at time 3 months followed by a substantive increase at time 6 months. In a way, these peptides made the aforementioned dual-cutoff (at most 5% BH FDR and time fixed-effect coefficient  $\geq 0.3333$ ) because antibody response levels increase tremendously by time 6 months for all 20 patients.



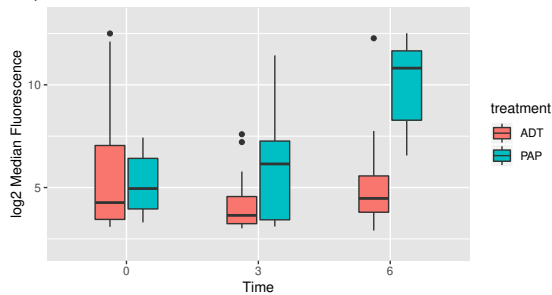
We also provide boxplots of fluorescence of a few of these 5680 peptides.



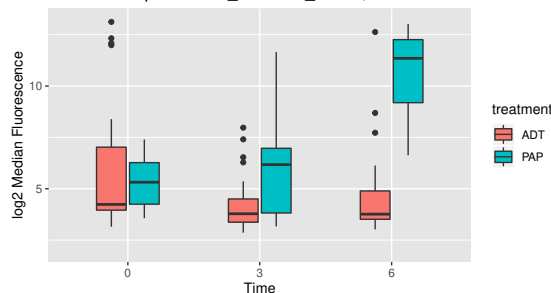
Boxplots of Fluorescence Levels for Peptide: CAT151.3\_T027960\_G006287\_2\_15387\_15731\_345;101



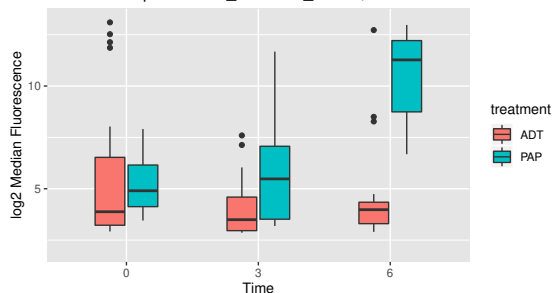
Boxplots of Fluorescence Levels for Peptide: FBXL19\_AS1\_T130509\_G030577\_1\_4499\_4918\_420;5



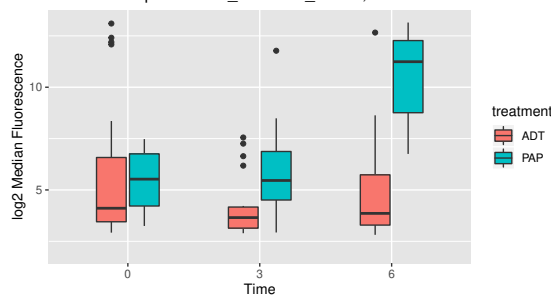
Boxplots of Fluorescence Levels for Peptide: 1052\_ARL6IP4\_51329;285



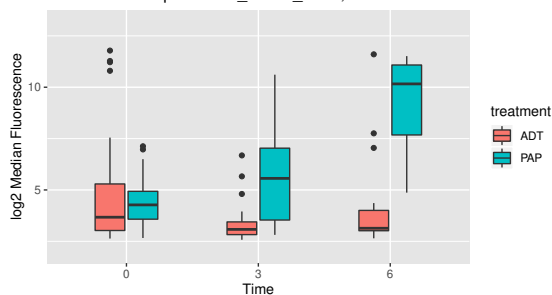
Boxplots of Fluorescence Levels for Peptide: 1052\_ARL6IP4\_51329;289



Boxplots of Fluorescence Levels for Peptide: 1052\_ARL6IP4\_51329;293



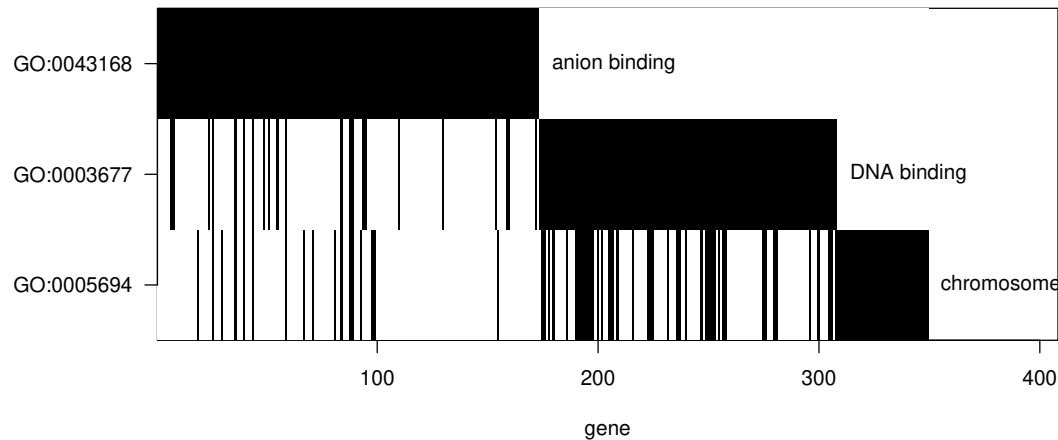
Boxplots of Fluorescence Levels for Peptide: 946\_NKTR\_4820;425



### 3.5 Gene-Set-Analysis

We also perform gene-set analysis based on the 5680 interesting/significant peptides identified for the PAP group. Again, the explanations for gene-set-analysis remain the same. Again, we use the same parameters for the gene-set-analysis: We shall consider gene-sets containing at least 2 interesting/significant genes (`n.cell = 2`) with Bonferroni-corrected enrichment p-values not exceeding 5% (`nominal.alpha = 0.05`). We also limit our analysis to those GO gene-sets which contain at least 5 genes (`n.low = 5`) and at most 300 genes (`n.upp = 300`).

The gene-set-analysis yields the following waterfall plot. Interpretation for the waterfall plot is similar as before.



We also tabulate the enriched/overrepresented GO terms. The last column of the table shows the genes associated with proteins that have at least one significant peptide in the contrast or pairwise comparison.

Term	Ontology	set.mean	set.size	z.score	in.genes
chromatin	CC	0.9200	69/75	4.3797	AR; CEBPB; DHX9; EZH2; MSH6; H1F0; HIST1H1C; HIST1H2AD; H3F3A; H3F3B; HDAC1; HMGB2; HMGN1; HMGN2; HNRNPC; HNRNPK; HSF1; EIF3E; JUN; JUNB; JUND; MYC; PRM2; RAD21; RAN; UPF1; SMARCA1; SMARCA4; SMARCC2; TCF3; TCP1; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST2H2AC; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; HIST1H2AG; MTA1; MAGED1; H2AFY; NCOR2; IST1; MORF4L1; CBX1; CBX3; POGZ; PDS5A; TARDBP; SUZ12; NOP53; BICRA; HP1BP3; PHF10; H2AFJ; FAM111A; NUCKS1; HIST1H2AH; HIST1H2BK; HIST3H2A; H2AFV; HIST2H2AB; H3F3C; HIST2H2AA4
chromosome	CC	0.8739	104/119	4.4726	PARP1; AR; BCL6; CEBPB; CENPE; DHX9; DYNC1L12; FBL; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HIST1H2AD; H3F3A; H3F3B; HDAC1; HMGB1; HMGB2; HMGN1; HMGN2; HNRNPC; HNRNPK; HSF1; EIF3E; JUN; JUNB; JUND; MCM3; MYC; SEPTIN2; NKX3-1; PAFAH1B1; PHF2; PPP1CC; PRM2; PURB; RAD21; RAN; RBBP6; UPF1; RPA1; CLIP1; SMARCA1; SMARCA4; SMARCC2; SP100; SSB; SSRP1; TCF3; TCP1; UBE2I; VCP; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST2H2AC; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; HIST1H2AG; MTA1; MAGED1; H2AFY; NCOR2; IST1; ARPC3; PCGF3; P3H4; PTGES3; MORF4L1; CBX1; CBX3; POGZ; PDS5A; TARDBP; SUZ12; SPIDR; ORC6; REPIN1; NOP53; BICRA; HP1BP3; GARI; PHF10; H2AFJ; NSFL1C; THOC2; FAM111A; NUCKS1; MEAF6; HIST1H2AH; HIST1H2BK; HIST3H2A; H2AFV; TOP1MT; CENPX; HIST2H2AB; H3F3C; HIST2H2AA4
chromosomal part	CC	0.8727	96/110	4.2564	PARP1; AR; BCL6; CEBPB; CENPE; DHX9; DYNC1L12; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HIST1H2AD; H3F3A; H3F3B; HDAC1; HMGB2; HMGN1; HMGN2; HNRNPC; HNRNPK; HSF1; EIF3E; JUN; JUNB; JUND; MCM3; MYC; SEPTIN2; NKX3-1; PAFAH1B1; PHF2; PPP1CC; PRM2; PURB; RAD21; RAN; UPF1; RPA1; CLIP1; SMARCA1; SMARCA4; SMARCC2; SP100; SSB; TCF3; TCP1; UBE2I; VCP; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST2H2AC; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; HIST1H2AG; MTA1; MAGED1; H2AFY; NCOR2; IST1; ARPC3; P3H4; PTGES3; MORF4L1; CBX1; CBX3; POGZ; PDS5A; TARDBP; SUZ12; ORC6; REPIN1; NOP53; BICRA; HP1BP3; GARI; PHF10; H2AFJ; THOC2; FAM111A; NUCKS1; MEAF6; HIST1H2AH; HIST1H2BK; HIST3H2A; H2AFV; CENPX; HIST2H2AB; H3F3C; HIST2H2AA4

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
DNA binding	MF	0.8486	157/185	4.9247	ADAR; PARP1; APLP2; APP; AR; ATF4; BCL6; ZFP36L2; CEBPB; CEBPD; CUX1; DDX1; DHX9; EEF1D; EPAS1; ERH; EZH2; XRCC6; GOLGB1; MSH6; GTF2I; GTF3A; H1FO; HIST1H1C; H3F3A; H3F3B; HDAC1; HMGB1; HMGB2; HMGN1; HMGN2; HNRNPC; HNRNPD; HNRNPK; HES1; HSF1; HSPD1; RBPJ; ILF2; ILF3; JUN; JUNB; JUND; LBR; MCM3; KMT2A; MYC; NACA; NCL; NFIA; NFE2L1; NFIB; NFIC; NFIL3; NKX3-1; NPAS2; NPM1; YBX1; PA2G4; PNN; POLR2L; PRM2; PURB; UPP1; RPA1; RPL6; RPL7; RPS3; RPS15; RPS27; SET; SMARCA1; SMARCA4; SMARCC2; SON; SOX4; SP3; SP100; SSRP1; TAF7; TCEA1; TCF3; TDG; NR2F2; TSG101; ZFP36; ZNF24; ZKSCAN1; VEZF1; ZFAND5; KAT6A; TAF15; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; DDX3Y; EDF1; EED; TAF1C; MTA1; H2AFY; IER2; BCLAF1; THRAP3; DNAJB6; AKAP9; RBM5; SRRM1; ZMPSTE24; BASP1; HOXB13; KHDRBS1; RAI1; ZNF275; SUB1; FOXJ3; TCF25; SMG1; RYBP; TARDBP; SUZ12; EHF; NUPR1; FOXP1; REPIN1; HP1BP3; SIDT2; LEF1; CXXC5; TDP2; SRR1; XRN1; ZFAND6; BANP; IFT57; STRBP; CHD7; ZNF395; SLC2A4RG; BBX; SCYL1; CHD8; ZNF350; NUCKS1; IRX3; TBL1XR1; RAX2; HIST3H2A; GTF3C6; TOP1MT; ZNF664; CREB3L4; ZMAT2; CENPX; H3F3C ABAT; ACADVL; ACTN2; AK2; ALDH1A3; ANXA1; APLP2; APP; ARF1; ARF4; ASNS; ATP1A1; ATP1B1; ATP6V1A; BMPR1B; DDR1; CBS; CCT6A; CENPE; CHKA; CKB; CSNK1D; CSNK1E; DDX1; DDX5; DHX9; DHCR24; CYB5R3; DLD; DYNC1L1; DPYSL3; EEF1A1; EIF5; FABP5; FOLH1; XRCC6; GNAQ; GNAS; MKNK2; MSH6; GUCY1A1; HADH; HK2; HMGB1; DNAJA1; HSPA1A; HSPA8; HSP90AA1; HSPD1; IARS; ILF2; ITPK1; KIF5C; KRAS; LBR; LRPAP1; MARS; MAT2A; MCM3; MDK; MAP3K5; MGST1; MYH9; MYH11; MYO1C; MYO6; SEPTIN2; NME3; NOS1; PEBP1; PAFAH1B1; PDPK1; PFKM; PITPNA; PRKAR2A; PKN2; MAP2K3; PSMA6; PSMC1; PTPRF; RAN; RAP1B; RARS; UPP1; RPL22; RPL29; SGK1; SMARCA1; SMARCA4; SQLE; SRPRA; TARS; TCP1; TDG; NR2F2; HSP90B1; CCT3; UBE2I; VCP; PIP4K2B; ULK1; STK24; OGT; DGKD; DDX3Y; SNX3; RAB3D; MICAL2; EIF4A3; MFN2; THRAP3; FARSB; ABCC5; ATP9A; HIPK3; ECI2; UBE2E3; CCT7; CCT4; CAMKK2; PMVK; Rraga; SEPTIN9; HSPH1; RAB10; FASTK; RAB35; UBE2C; ADAMTS5; SNRNP200; SMG1; KIF13B; GTPBP4; WIPI2; GNL3; ATP2C1; EIF2AK1; ARFIP1; ACAD9; HIPK2; SNX12; CRYL1; SAR1B; HSD17B12; DDX47; IP6K2; RTCB; CHMP3; RASD1; RIPK4; LAPTM4B; DNAJA4; DHTKD1; UBE2Q1; CHD7; NSFL1C; WDR45B; MCCC1; ATP8B2; SCYL1; CHD8; GOLPH3; RRAGC; WNK1; DDX50; ATP13A3; MYO19; ALPK1; ACSS1; NEK9; BBS5; ANKK1; NRPB2
anion binding	MF	0.8357	173/207	4.8214	

## 4 Conclusion

To investigate whether different prostate cancer stages lead to different antibody responses:

- We utilized both calls data and median fluorescence (across replicates) data.
- Calls data is very conservative – most calls are zero and nothing interesting pops up.
- For median fluorescence data, we deployed the Kruskal-Wallis tests and performed the Benjamini-Hochberg (BH) procedure to control for false discovery rate (FDR). With this approach, we identified 13729 peptides at 5% FDR.
- For these 13729 peptides, we zoomed-in on the following 6 contrasts or pairwise-comparisons. For each contrast, we deployed the Wilcoxon-Rank-Sum tests and performed the BH procedure on these 13729 Wilcoxon p-values. We are interested in peptides that meet the 5% BH-FDR cutoff based on these Wilcoxon p-values as well as having at least a two-fold difference between the medians of the two groups. The counts of peptides that fulfill the two conditions are tabulated below.

pairwise_comparison	peptide_counts
cancer vs normal	110
mCRPC vs others	4246
mCRPC vs nmCRPC	790
nmCRPC vs nmCSPC	3655
nmCSPC vs new_dx	637
new_dx vs normal	686

- We further performed gene-set-analyses based on the peptides identified as interesting/significant for each contrast.
- Visualization techniques via the heatmap and PCA (principal component analysis) reveal the effects of cancer stages on individuals' antibody responses.

To analyze how treatments (PAP vaccine or ADT) influence change in antibody responses over time,

- We deployed linear mixed effects model for each peptide, separately for the group of PAP-vaccinated patients and for the group of patients administered with ADT.
- We also applied the BH procedure on the F-test p-values of the time fixed effects for both groups of patients.
- No significant peptides are identified for the ADT group, even at 20% FDR.
- For the PAP group, we identified 5680 peptides at 5% BH FDR which also exhibit at least two-fold increase in median fluorescence levels every 3 months. Gene-set analysis is also performed based on the proteins associated with this list of peptides.
- The heatmap clearly illustrates that PAP-vaccinated patients had significantly higher antibody responses (measured by those peptides) over the course of 6 months.

The lists of significant peptides identified in both studies are exported to the Excel file “09\_Significant\_Peptides.xlsx”. Boxplots of median fluorescence levels of some example peptides for different cancer stages are generated in .png images.

## References

- Douglas Bates, Martin Machler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.
- Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.
- Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26, 2017.
- Steven G. Luke. Evaluating significance in linear mixed-effects models in r. *Behavior Research Methods*, 49:1494–1502, 2017.
- J.H. McDonald. *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, Maryland, third edition, 2014.
- Michael Newton, Subhrangshu Nandi, Ning Leng, and Aimee Teo Broman. *allez: Random-set calibration of gene-set statistics*, 2018. R package version 2.0.7.
- John D. Storey and Robert Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16):9440–9445, 2003.
- Korbinian Strimmer. A unified approach to false discovery rate estimation. *BMC Bioinformatics*, 9(303), 2008.
- Larry Winner. *Introduction to Biostatistics*. Department of Statistics, University of Florida, 2004. URL [http://users.stat.ufl.edu/~winner/sta6934/st4170\\_int.pdf](http://users.stat.ufl.edu/~winner/sta6934/st4170_int.pdf).