Antibody Responses to Different Proteins in Prostate Cancer Patients

Tun Lee Ng and Michael A. Newton

April 06, 2020

Contents

1 Introduction

This project aims to characterize antibody responses to a wide variety of proteins in prostate cancer patients at different stages of the disease. 16-mer peptides spanning the amino acid sequences of these 1611 gene products, and overlapping by 12 amino acids, were used to generate a microarray comprising 177,604 peptides. In this study, there were 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

stage	n
normal	17
new_dx	19
nmCSPC	52
mCSPC	16
nmCRPC	15
mCRPC	35
binding_buffer	1

Note that these are not distinct patient counts, because there were 11 patients who were measured at two different stages. Number of replicates for each patient, `rep` could 1, 2, or 3. Hemanth dropped patients with `rep` = 1. He also kept the latest distinct patient records only, so the distinct patient counts are:

Hemanth has already checked that replicates for each patient largely "agree with one another". We compute the median fluorescence level for each patient and take $\log_2$ transformation of the median fluorescence intensity
.

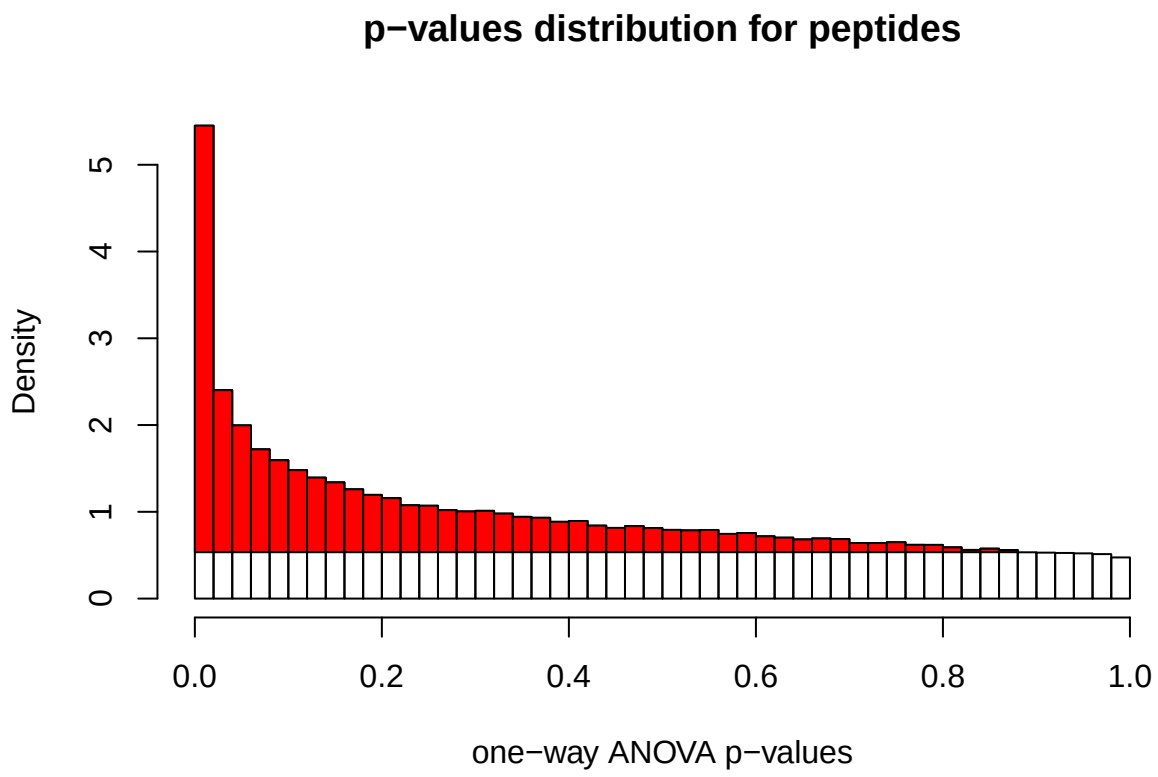
stage	n
normal	15
new_dx	15
nmCSPC	33
nmCRPC	13
mCRPC	15

2 One-Way ANOVA

We would like to investigate if patients at different stages of prostate cancer exhibit different antibody responses to certain peptide chains or proteins. Let μ_i be the average fluorescence level (on $\log_2$ scale) of patients, with subscript i indexing the different stages of prostate cancer as explained in Introduction section. We want to test

H_0: All μ_i's are the same, ie. Antibody responses are the same for patients at different stages of prostate cancer.
H_1: NOt all μ_i's are the same, ie. Antibody responses are not the same for patients at different stages of prostate cancer.

For each peptide, we perform one-way ANOVA (analysis of variance). After getting p-values for all 177k peptides, we plot the p-value histogram.



The estimated proportion of non-null probes in the dataset based on ANOVA is

```
## [1] 0.4657292
```

The estimation is based on Storey's q-values computed with the R package `fdrtool`. The q-value is similar

to the well known p-value, except it is a measure of significance in terms of the false discovery rate rather than the false positive rate. The peptide counts at various FDR thresholds are tabulated below.

FDR threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Peptide counts	176	2688	5202	7428	9493	11543	13511	15572	17839	19791

As a comparison, we also apply the Benjamini-Hochberg (BH) method on the ANOVA p-values to control for false discovery rate. The peptide counts at various FDR thresholds are tabulated below.

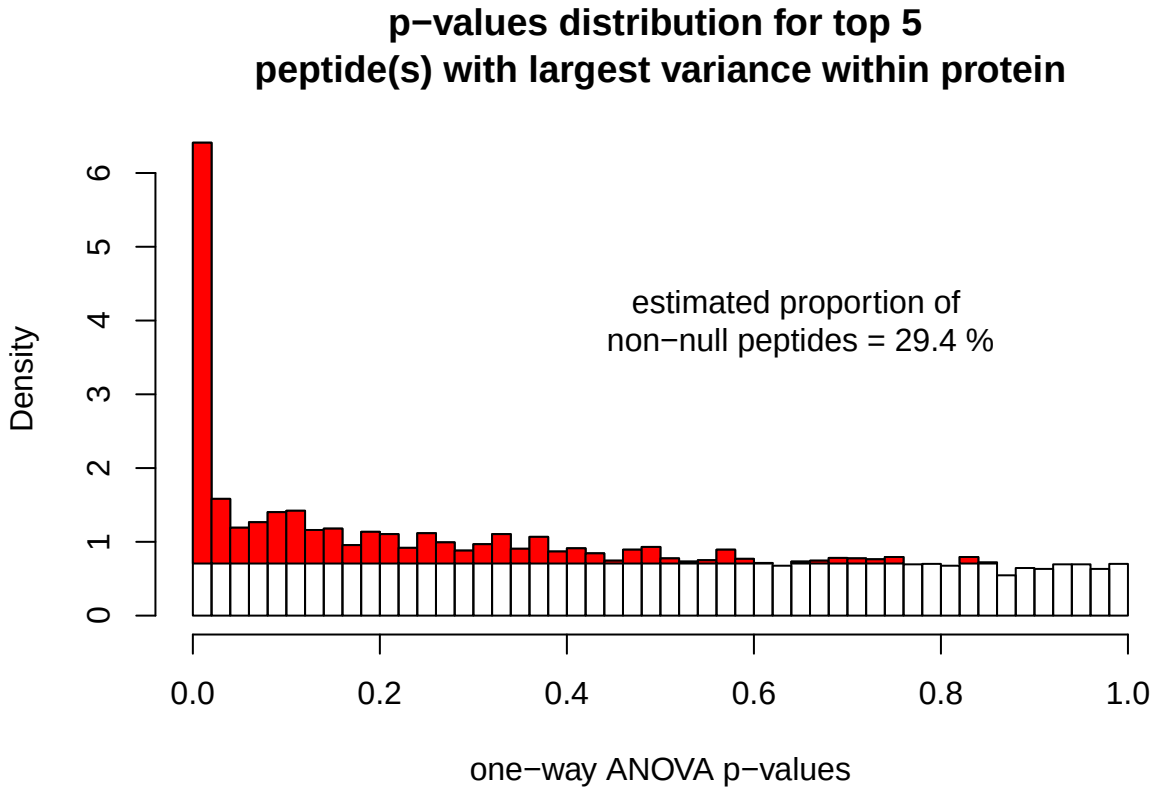
FDR threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Peptide counts	38	181	1420	3082	4549	5656	6779	7994	9123	10183

3 Marginal Variance Filtering

Next, for every protein, we would like to filter the top few peptides with the largest marginal variance of $\log_2$ fluorescence among all patients.

3.1 Top 5 peptides with largest marginal variance

First, we filter top 5 peptides with largest marginal variance in each protein. Then we plot histogram of p-values.



Next, we tabulate peptide counts at different BH-adjusted p-values thresholds.

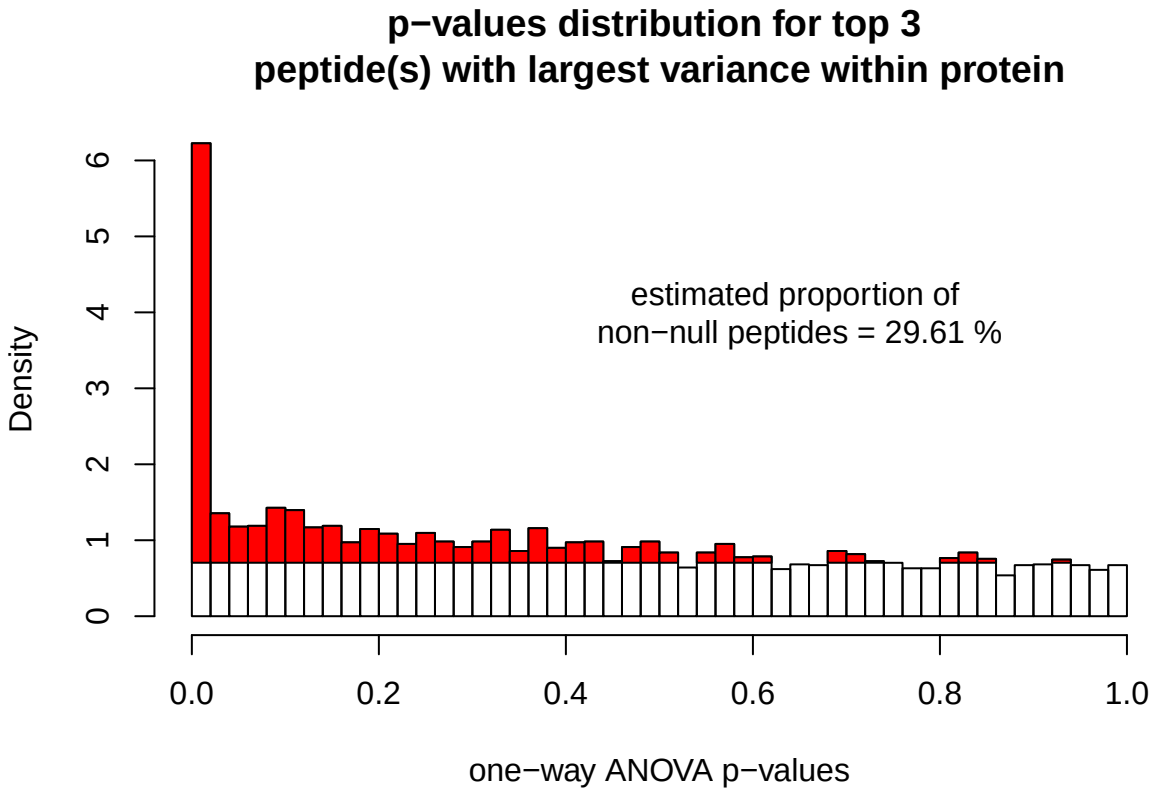
Then, we tabulate peptide counts at different Storey's q-values thresholds.

FDR threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Peptide counts	5	425	538	612	684	736	782	814	843	876

FDR threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Peptide counts	212	527	641	725	785	829	873	918	947	988

3.2 Top 3 peptides with largest marginal variance

Now, we filter top 3 peptides with largest marginal variance in each protein. Then we plot histogram of p-values.



Next, we tabulate peptide counts at different BH-adjusted p-values thresholds.

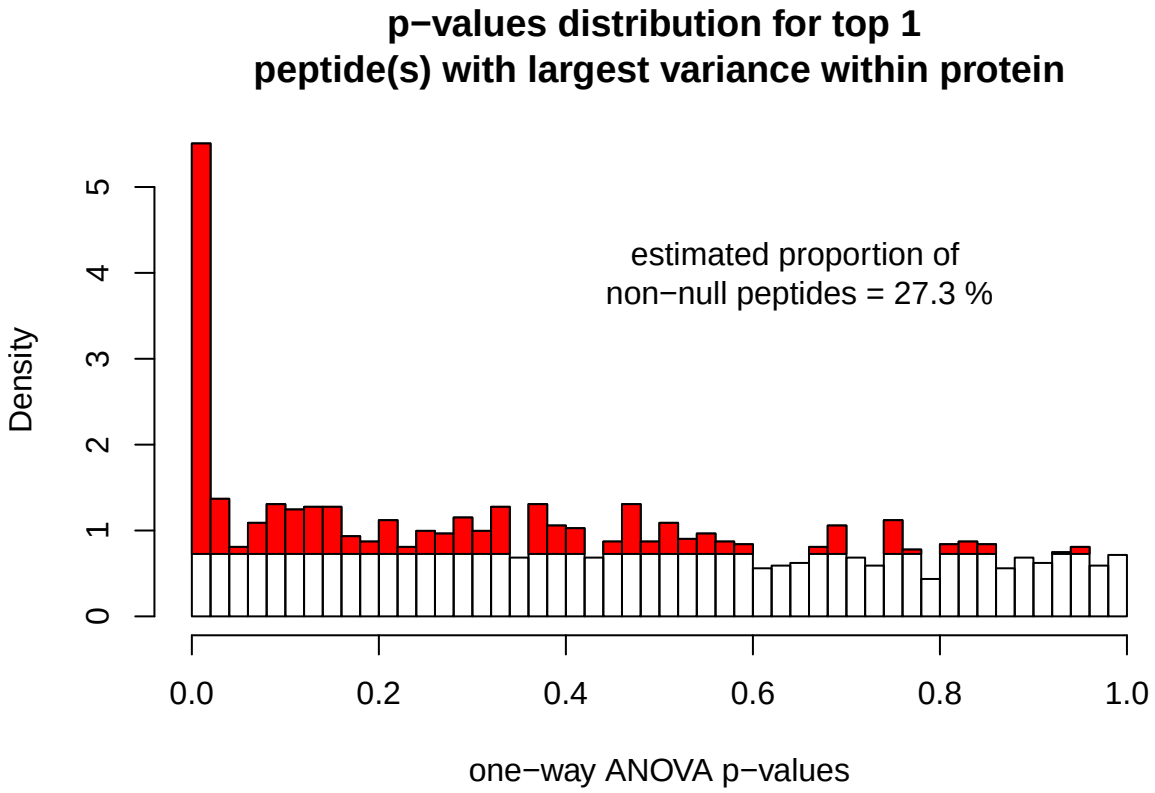
FDR threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Peptide counts	52	273	328	369	412	438	458	480	498	517

Then, we tabulate peptide counts at different Storey's q-values thresholds.

FDR threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Peptide counts	143	323	388	434	464	491	516	541	559	580

3.3 Top peptide with largest marginal variance (As representative of protein)

Now, we filter top peptide with largest marginal variance in each protein. This can be taken as protein-level analysis as we take the peptide with largest marginal variance a representative of its corresponding protein. Then we plot histogram of p-values.



Next, we tabulate peptide counts at different BH-adjusted p-values thresholds.

FDR threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Peptide counts	27	69	93	101	107	112	122	122	130	137

Then, we tabulate peptide counts at different Storey's q-values thresholds.

FDR threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Peptide counts	48	88	104	112	122	126	137	147	157	162