Antibody Responses to Different Proteins in Prostate Cancer Patients

Tun Lee Ng and Michael A. Newton

May 21, 2020

Contents

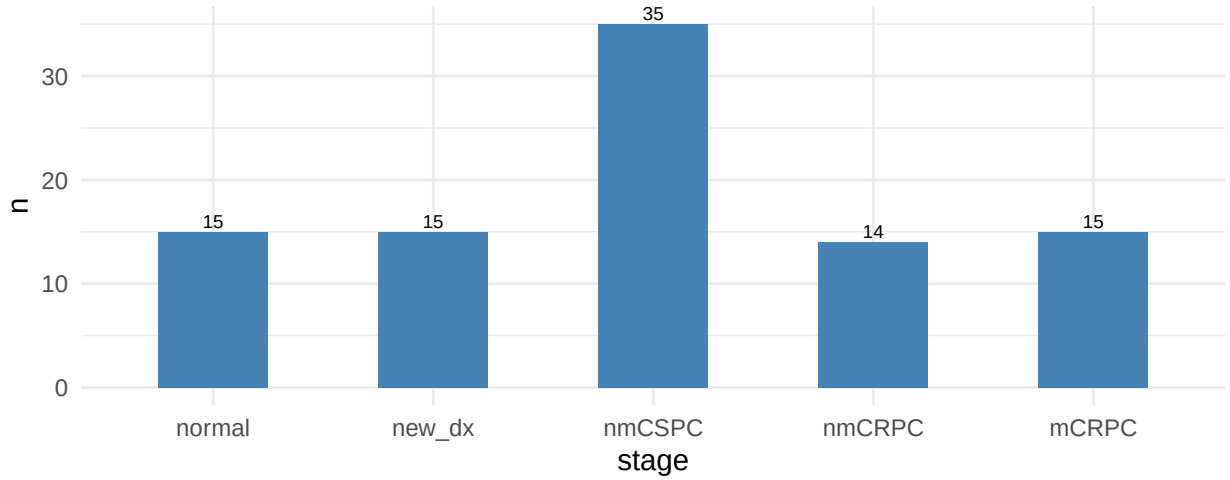
1 Introduction

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

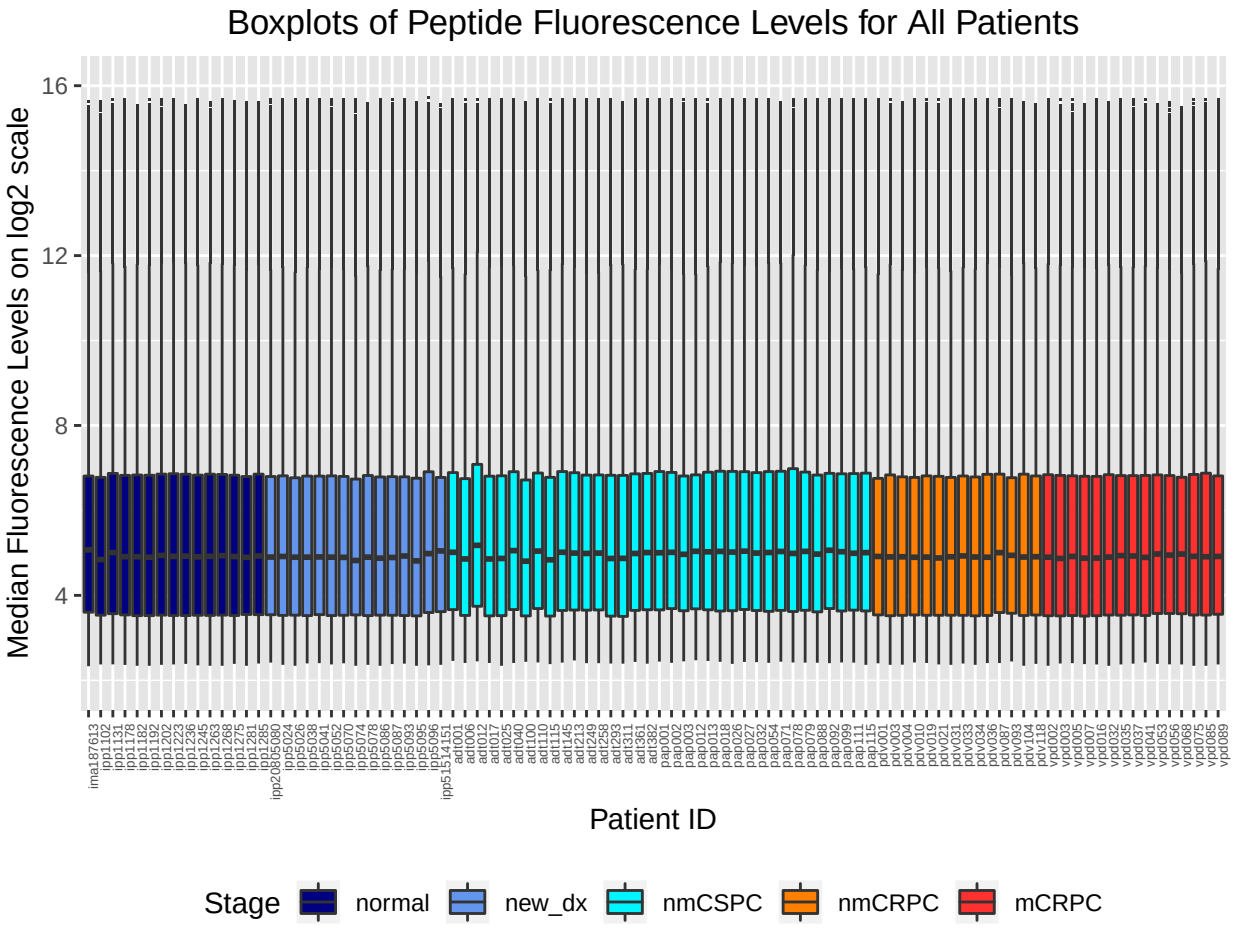
- `new_dx`: newly diagnosed,
- `nmCSPC`: non-metastatic castration-sensitive
- `mCSPC`: metastatic castration-sensitive,
- `nmCRPC`: non-metastatic castration-resistant,
- `mCRPC`: metastatic castration-resistant

Each patient's serum was assayed in a number of replicates, `rep`, which were 1, 2, or 3, for peptide-specific IgG responses using the microarray. 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 these 94 patient-records do not represent distinct patients, as certain patients were measured at two different stages of prostate cancer. There are 6 patients who had replicates for both stages. We removed their earlier-stage records, and finally arrive at 94 distinct patients.



We will utilize both fluorescence levels data and patients' binary calls 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. 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.

2 Reproducibility of Replicates

We have assessed the issue of replicate reproducibility by looking at (Pearson and Spearman's Rank) correlation coefficients between patients' fluorescence levels (both with and without $\log_2$ transformation). Another approach is to measure how much variation the technical replicates are contributing to the overall variation in the data. If the replicates are reproducible, ie. they "largely agree with one another", then the technical variation in the dataset should be minimal.

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,

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

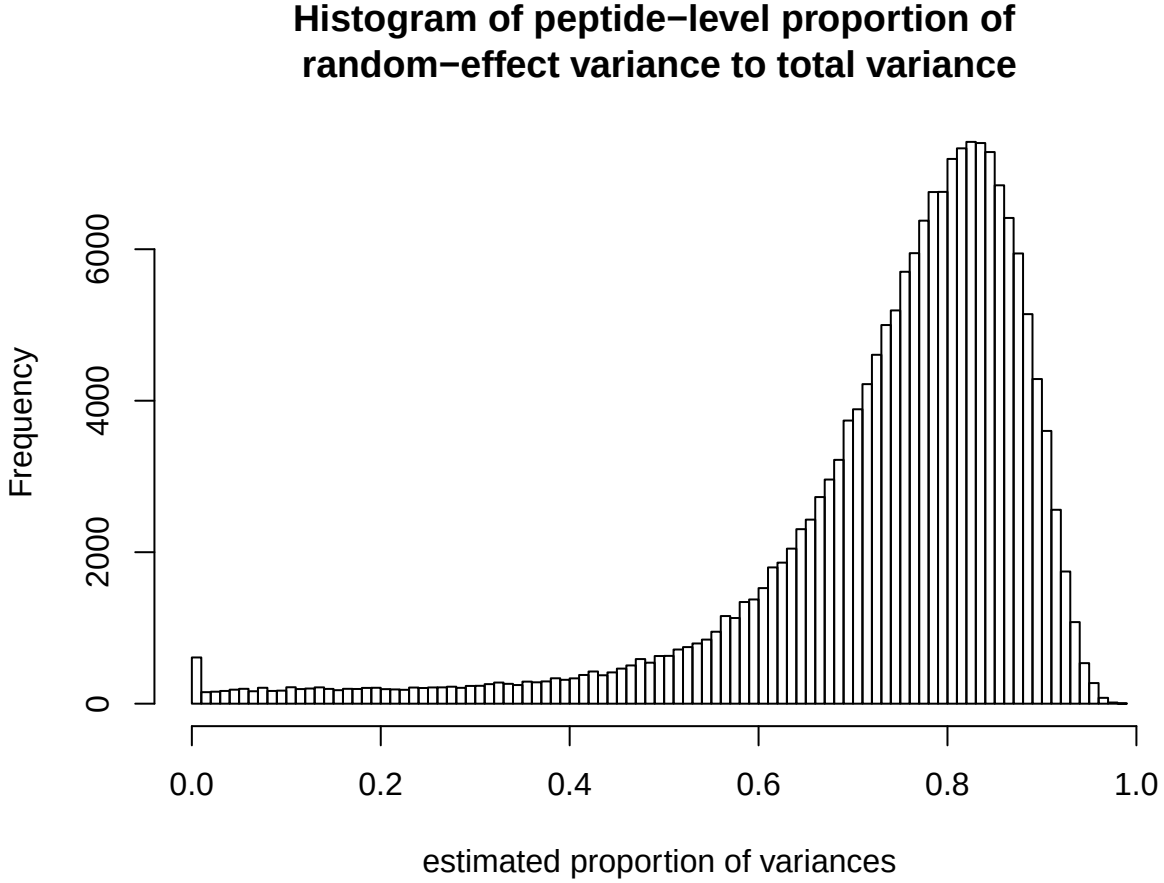
where

- y_{ijk} denotes the $\log_2$ fluorescence level of a replicate,
- μ denotes the grand mean/intercept,
- β_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,
- ϵ_{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`. 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.



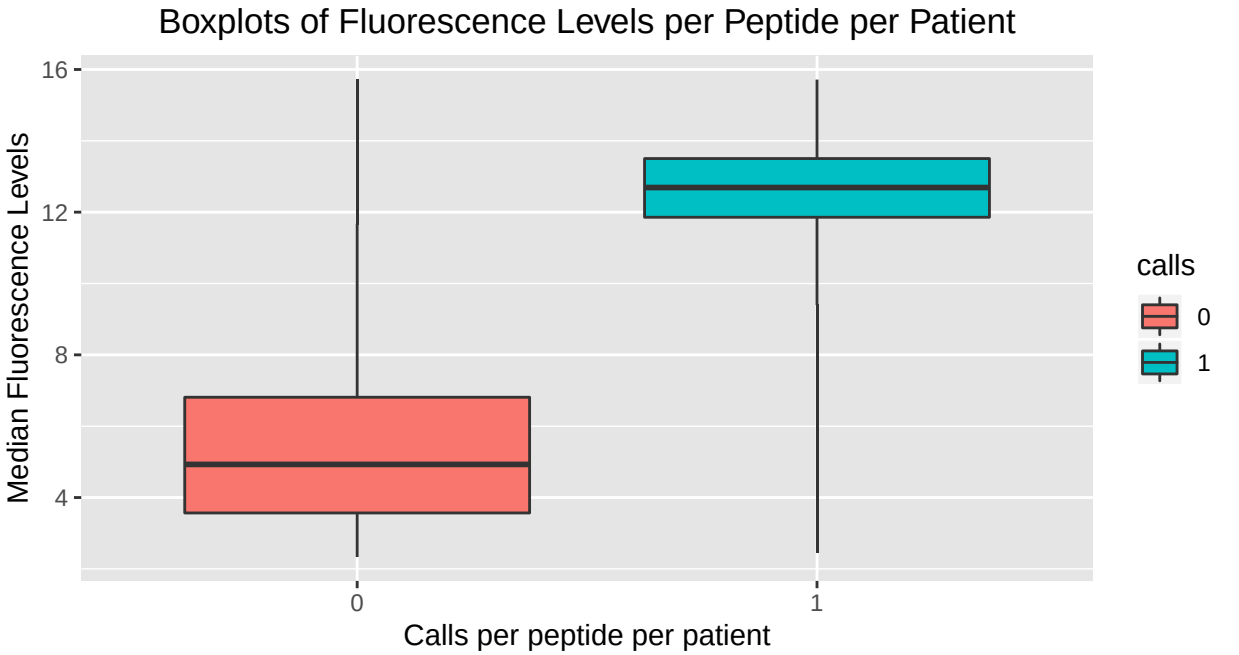
As expected, the histogram amasses at values near one, indicating that most of the variation in the ($\log_2$) fluorescence data is attributable to the biological variation of the patients and not the technical replicates themselves, which also suggests reproducibility of the replicates. The little spike at zero estimated proportions is due to the 477 singular cases where the fitted random-effect variance $\hat{\sigma}_b^2$ is close/equal to zero.

3 Tests on Binary Calls

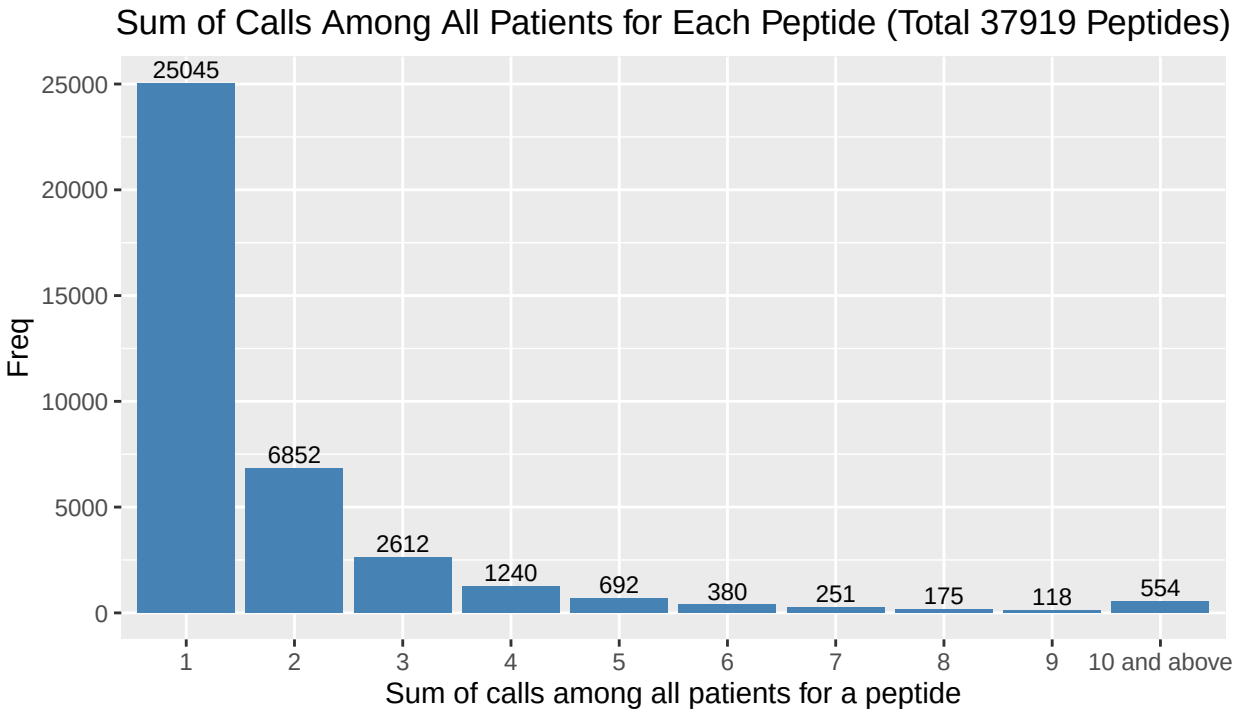
We acknowledge the consensus that the binary calls on a peptide of a patient could be conservative – out of 177,604 peptides, only 37919 of them have at least one call among all patients. If we collapse all patients' peptides to quantify the number of zero- and one- calls, we obtain the following counts:

calls	patient_peptide_counts	percent
0	16622577	99.6%
1	72199	0.4%

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.



Even among the 37919 peptides (that have at least one call), most of them have very few positive calls among all patients. We plot the distribution of the sum of calls among all patients for each of these peptides.



For example, there are 25045 peptides with only one positive call among all 94 patients. 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**.

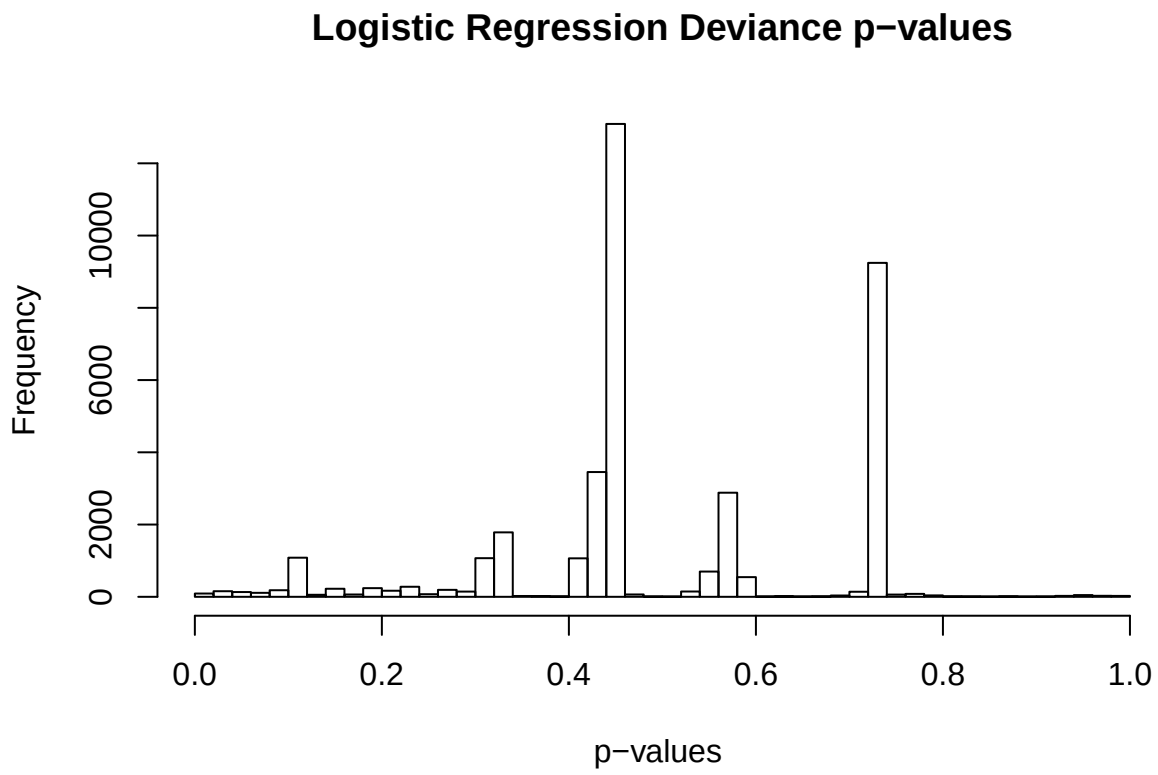
Specifically, for each of these 37919 peptides, we fit the following model:

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

where

- y_{ij}^{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,
- μ denotes the grand mean/intercept,
- β_i denotes the cancer stage,
- ϵ_{ij} denotes the random error of the model, with j indexing the patients,

and compute the deviance p-values: (null_deviance - residual_deviance) $\sim \chi^2$ with 4 degrees of freedom. We plot the histogram of the 37919 p-values.

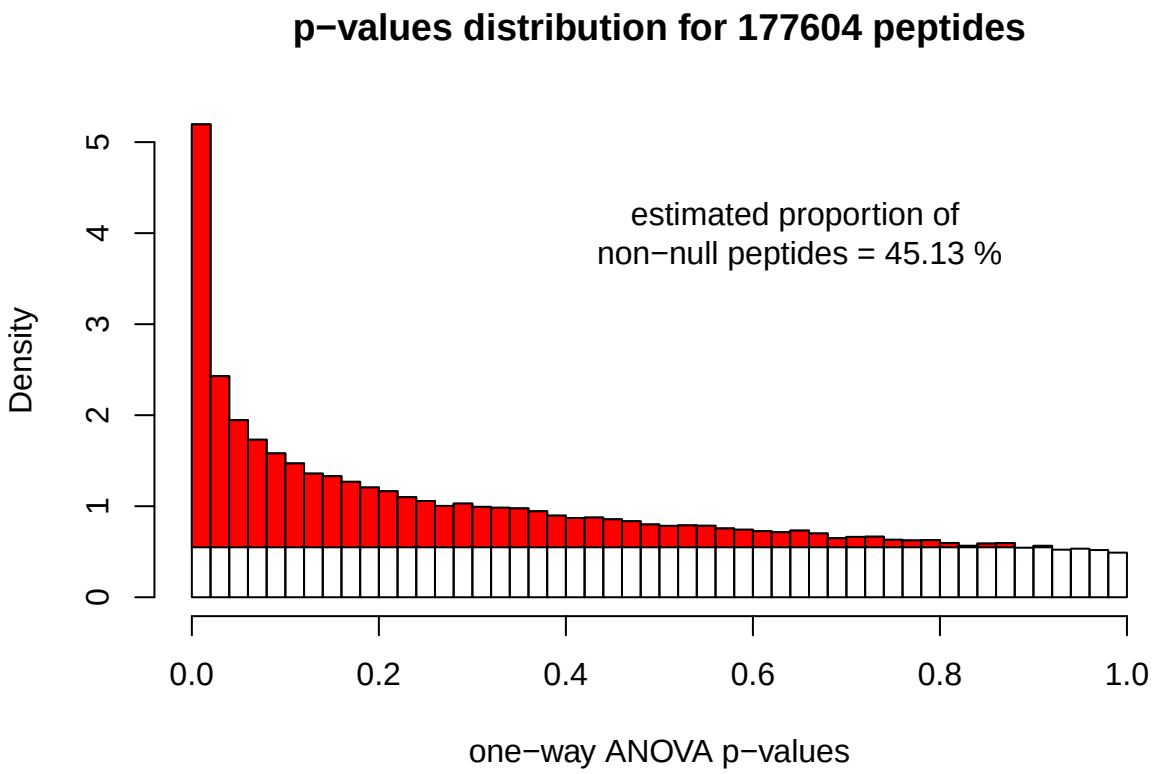


It appears that there are hardly any signals of different calls pattern among patients of different cancer stages, which corroborates with the results in Figures 2, 3 and 4 in the main manuscript. After correcting for false discovery rate, no peptides appear to be significant.

4 Tests on Fluorescence Levels

We lose a lot of information due to the conservative signal threshold of the calls data. Instead, in this section, we will investigate antibody response level among patients of different cancer stages using the fluorescence data. In the next section, we will utilize both fluorescence and calls data to investigate our hypothesis and compare our results. First, we perform one-way ANOVA (analysis of variance) to test the following:
H_0: Antibody responses are the same for patients at different stages of prostate cancer.
H_1: Antibody responses are not the same for patients at different stages of prostate cancer.

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



If cancer-stage effect is not present in our peptide array data, then the p-values from the ANOVA 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 dataset based on Storey's q-values calculation obtained via the R package `fdrtool`. The q-value is similar to the well known p-value, except that it is a measure of significance in terms of the false discovery rate rather than the false positive rate.

We apply the Benjamini-Hochberg (BH) method on the ANOVA p-values to control for false discovery rate (FDR). The peptide counts at various BH 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	82	204	738	1807	3128	4396	5562	6861	7858	8885

Here are also the peptide counts at different 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	169	1380	3732	5938	7977	9897	11708	13641	15513	17395

It appears that the BH method gives a more conservative significant peptide counts.

4.1 Visualizing Results

It appears that we have identify 3128 significant peptides at 5% BH FDR. We would like to obtain some graphical representations to illustrate how the $\log_2$ fluorescence levels differ across different cancer stages for these peptides.

For each peptide, we remove the grand mean (row mean) of the $\log_2$ fluorescence levels for all patients, before applying the following visualization techniques:

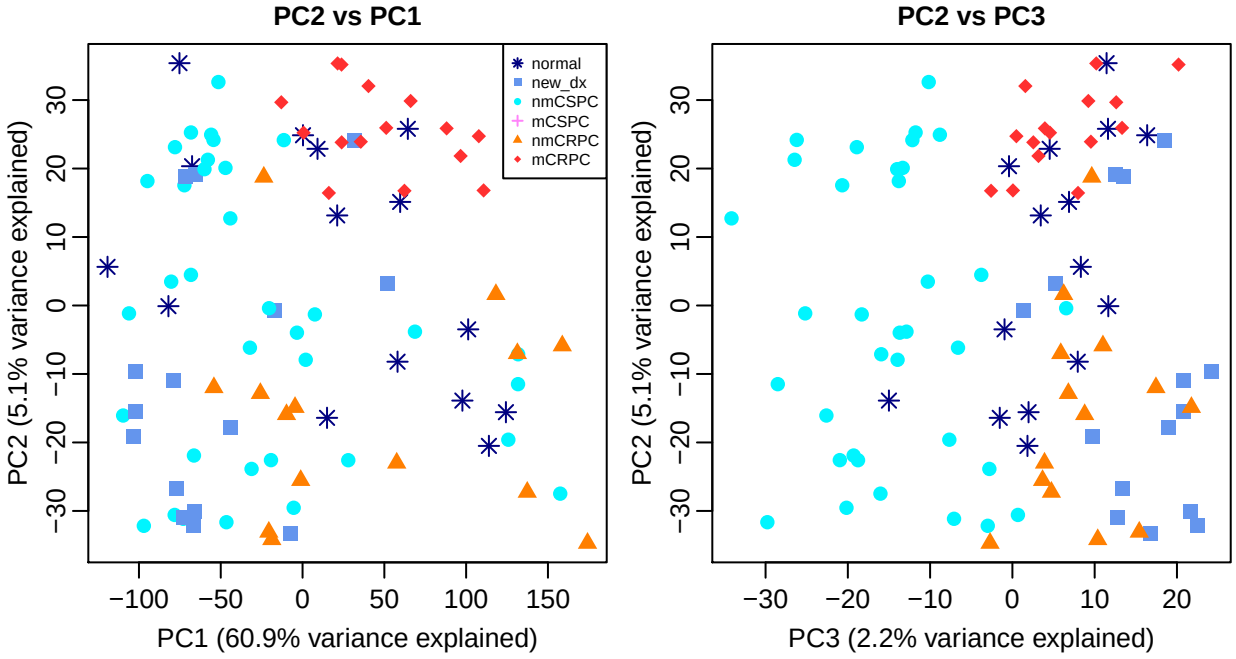
- Principal Component Analysis (PCA)
- t-distributed Stochastic Neighbor Embedding (t-SNE)
- Heatmap

Based on our one-way-ANOVA model assumption, 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):

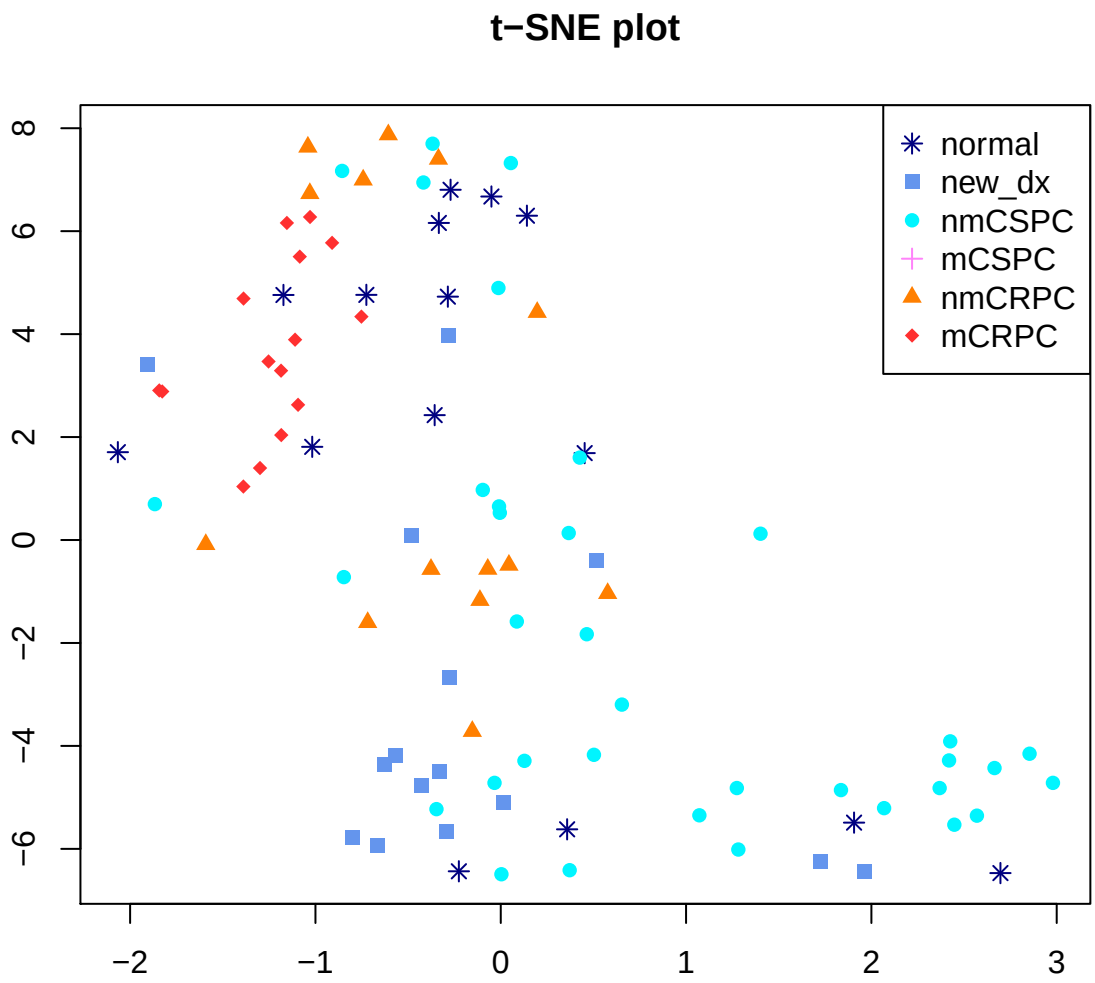
- navy for healthy subjects
- cornflower_blue for `new_dx` newly diagnosed patients
- turquoise for `nmCSPC` patients
- light_pink for `mCSPC` patients – these patients have no technical replicates and are excluded from this analysis
- dark_orange for `nmCSPC` patients
- dark_red for `mCSPC` patients

First, the PCA plot is given by



From the "*PC2 vs PC1*" plot, we observe that all mCRPC points are clustered at the topright of the panel, whereas 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.

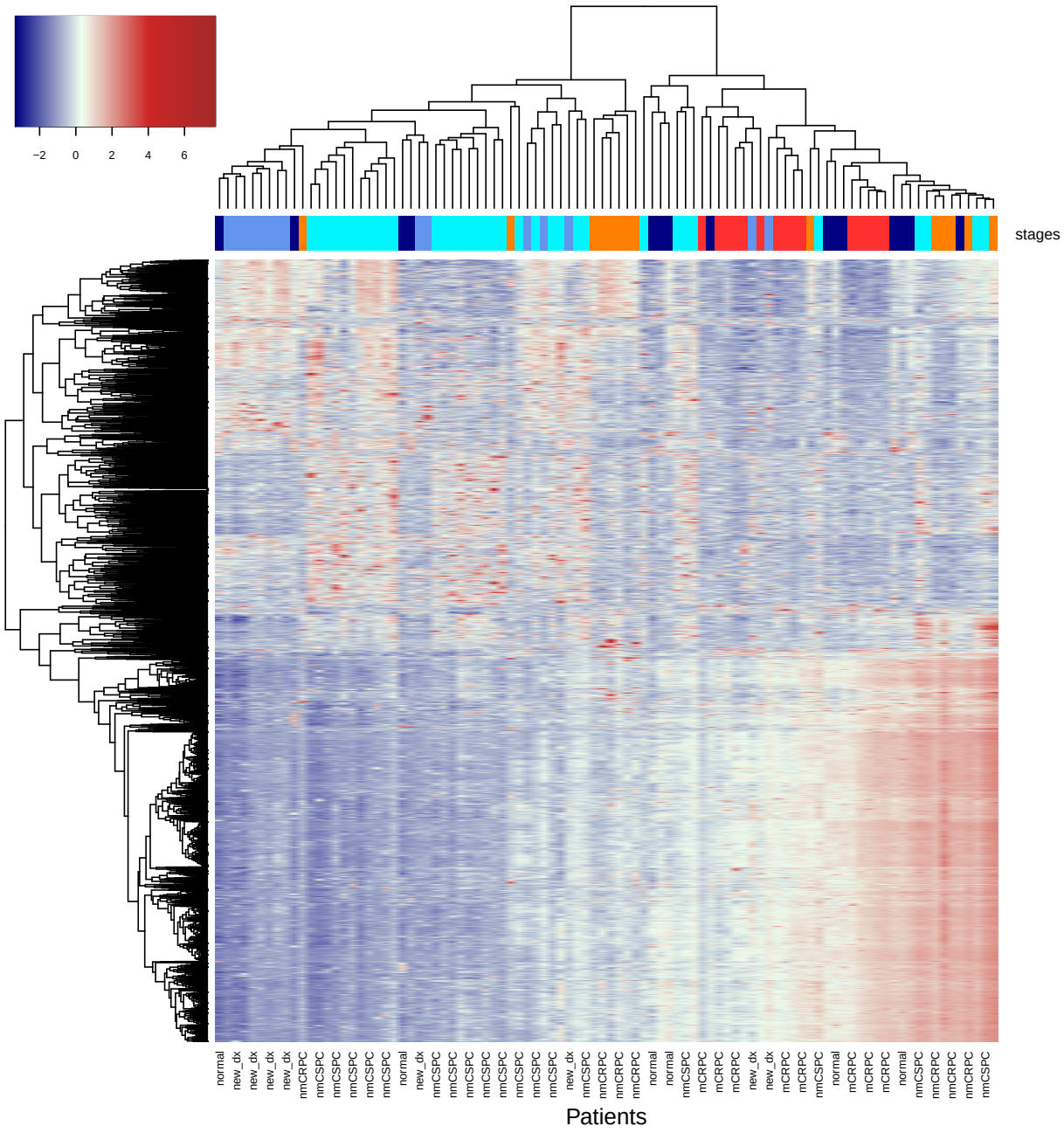
Next, we obtain the t-SNE plot. Just like PCA, t-SNE is a dimensionality reduction technique which was first introduced by van der Maaten and Hinton in 2008.



From the t-SNE plot, we observe that:

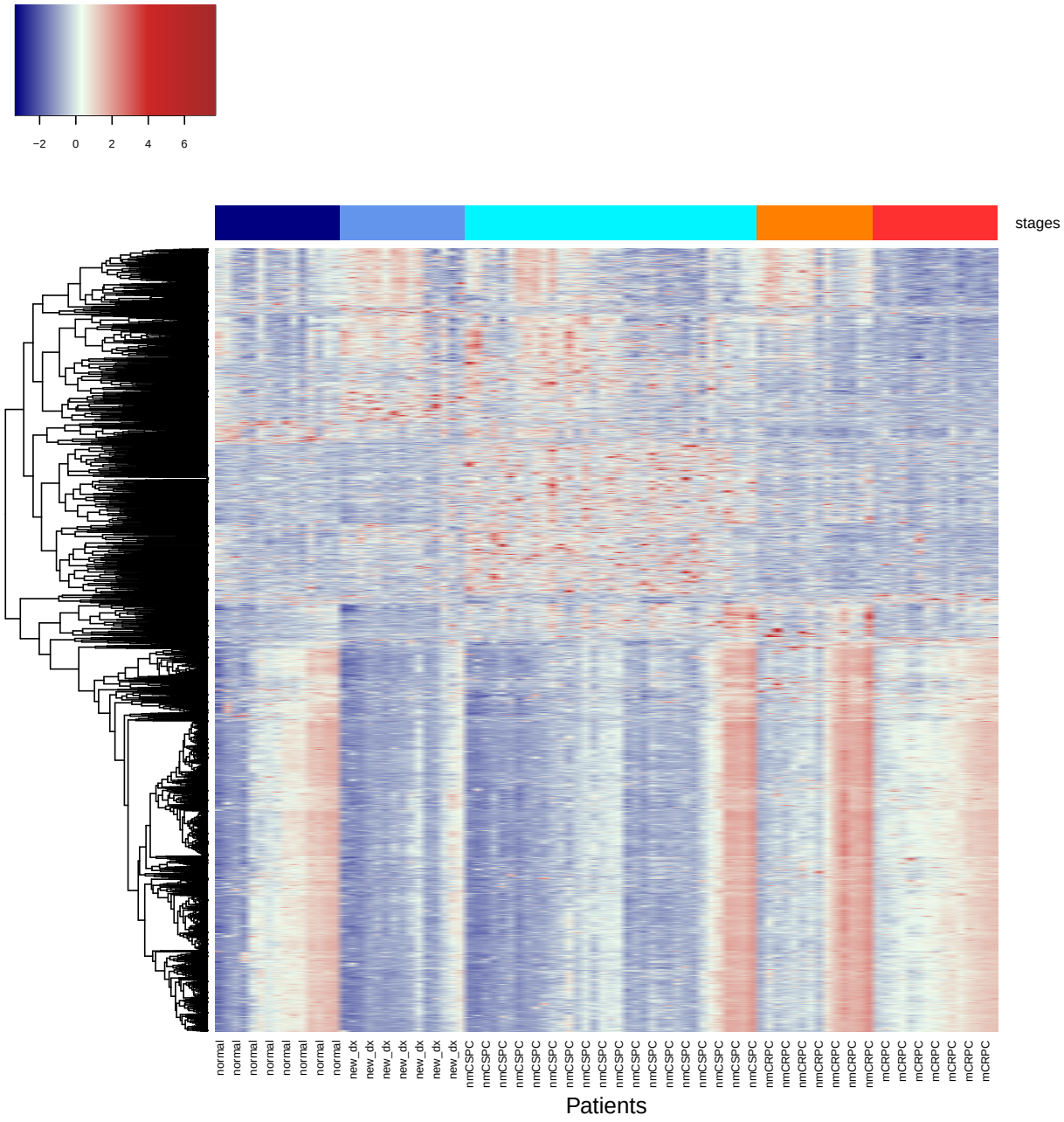
- The mCRPC points are clustered together near top-left of the plot.
- Most of the nmCRPC patients are not too far off from the mCRPC subjects. There seems to be 2 clusters of nmCRPC patients.
- Most of the new_dx and nmCSPC patients are clustered at the bottom left and middle of the plot.
- Normal subjects are somewhat "all over the place".

Finally, we plot the heatmap of the $\log_2$ fluorescence (after removing row means) of these peptides at 5% BH-FDR.



Again, we observe similar patterns that nmCSPC (colored turquoise) and new_dx (colored cornflowerblue) subjects are mostly clustered to the left part of the heatmap, whereas the mCRPC (colored red) patients and most of the nmCRPC patients (colored darkorange) are clustered to the right part of the heatmap. Again, normal subjects (colored navy) are "all over the place".

We also plot the heatmap where we order the columns according to patients' stages instead of the default cluster dendrogram order.



4.2 Gene-Set Analysis

Recall that the 177604 peptides correspond to 1611 proteins, and 1462 of these proteins have matching genes in *uniprot_gene_entrez.csv*. The 3128 significant peptides at 5% BH-FDR are associated with 993 proteins with matching genes. In this analysis, we deem a protein to be significant if it has at least one significant associated peptide. We shall utilize this information to perform gene-set enrichment analysis.

Specifically, we investigate if there are any pre-specified gene-sets that are enriched for the genes associated with the list of significant peptides. 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`. We shall consider gene-sets containing at least 2 genes with Bonferroni-corrected enrichment p-values not exceeding 5% .

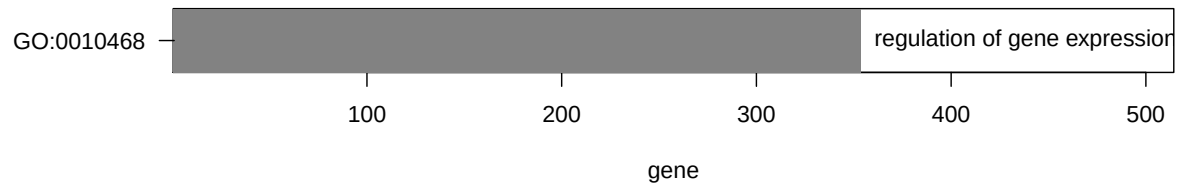
```
# gene-set analysis via allez!
allez.go <- allez(seq_id_ok, lib = "org.Hs.eg", idtype = "ENTREZID", sets = "GO")

# nominal alpha for enrichment score
nom.alpha <- 0.05

# minimum number of genes in gene-set to be considered
min.num.gene <- 2
```

We display the waterfall plot of the gene-set-analysis results. It turns out that in this case, there is only one significantly enriched gene set, since most of the proteins considered in this case are associated with at least one significant peptide.

```
# Waterfall plot
allezPlot(allez.go, n.cell = min.num.gene, nominal.alpha = nom.alpha)
```



4.3 Comparing Healthy Subjects vs Cancer Patients

The one-way ANOVA is helpful in revealing peptides that exhibit significant difference in the group means of $\log_2$ fluorescence levels among patients at different stages of cancer. We may be interested to find out, for example, **if antibody response is significantly higher in cancer patients than in healthy subjects**. Using the linear-contrast method, we are able to further test this hypothesis after computing the ANOVA.

Specifically, it is a one-sided t-test comparing the averages of the $\log_2$ fluorescence levels among cancer patients and among healthy subjects, except that now, the degrees of freedom and the standard error of the t-test statistic are based on the one-way ANOVA computed previously. This allows us to pull samples from different cancer stages and increase power – with the caveat/assumption that the patients of different cancer stages share the same level of variation in antibody response. (Note that we could avoid doing the multiplicity control like Tukey's Honest Significant Difference if we are only interested in a particular contrast/comparison and not every possible combination of pairwise comparisons)

There are 356 peptides that are associated with higher antibody response in cancer patients than in normal subjects. We tabulate these peptides with their differences (of average $\log_2$ fluorescence levels between cancer

and healthy subjects) and p-values of their t-statistics.

peptide_id	difference	tstat_pvalues
539_MAML1_9794;657	1.8619	0.0071
1386_RALGAPA2_57186;997	1.4539	0.0076
788_IVNS1ABP_10625;305	0.1645	0.0084
255_RANBP2_5903;1441	1.3179	0.0085
646_FASTK_10922;153	0.2059	0.0087
385_EIF3D_8664;45	1.4657	0.0089
190_ND5_4540;249	0.2192	0.0094
912_EIF2B1_1967;125	0.1535	0.0095
1242_THOC2_57187;1189	1.7438	0.0099
CAT297_T196839_G045377_1_2510_3112_603;1	0.8343	0.0101
114_SRRM2_23524;1569	1.3350	0.0102
1081_DDX3Y_8653;301	1.6009	0.0106
1429_FRMD8_83786;213	1.0546	0.0106
191_CMTM6_54918;25	0.3225	0.0107
584_TUBA1C_84790;273	2.0708	0.0108
882_ARL8B_55207;1	0.3196	0.0115
453_NCL_4691;597	1.3914	0.0116
1026_CLIP1_6249;337	1.1904	0.0122
1204_TAF15_8148;581	1.3479	0.0126
654_RPS3_6188;109	0.2659	0.0129
385_EIF3D_8664;49	1.5236	0.0130
200_BRI3_25798;93	0.1482	0.0131
500_F11R_50848;9	0.1297	0.0132
662_AUTS2_26053;573	1.6983	0.0133
135_TRMT112_51504;5	0.2724	0.0135
489_MTMR4_9110;869	0.8476	0.0135
1098_ARGLU1_55082;197	1.7184	0.0140
471_NDUFB8_4714;81	1.3128	0.0142
556_POGZ_23126;869	1.1663	0.0143
1045_STRBP_55342;213	0.1353	0.0144
859_KIAA1522_57648;229	1.4817	0.0148
PRO30;853	0.1989	0.0150
979_RPS2_6187;153	0.3368	0.0151
CAT583_T271074_G062826_2_704_1135_432;37	0.1400	0.0153
PRO25;25	1.0527	0.0157
1046_SRRT_51593;9	1.4562	0.0159
909_RNF149_284996;281	0.1877	0.0159
1269_SLC37A3_84255;117	0.1341	0.0160
1201_UIMC1_51720;225	1.5096	0.0161
1330_PDIA4_9601;5	0.1795	0.0161
797_MARCH6_10299;465	0.1788	0.0162
1422_CADM1_23705;385	0.1290	0.0165
640_VPS35_55737;193	0.7225	0.0167
19_NPIP_9284;213	0.7591	0.0168
1063_ITGAV_3685;453	0.2723	0.0169
1173_ATP1B1_481;241	0.4046	0.0173
591_ECHDC2_55268;5	0.4165	0.0178
1308_DPM1_8813;85	1.6117	0.0179
120_NPIPL3_23117;649	1.1170	0.0183
120_NPIPL3_23117;733	1.1170	0.0183
30_LOC100132247_100132247;649	1.1170	0.0183
30_LOC100132247_100132247;733	1.1170	0.0183
306_LBR_3930;313	0.1955	0.0184
682_ARFGAP3_26286;49	0.1919	0.0185
1268_XYLT2_64132;301	0.7833	0.0187
578_WSB1_26118;81	0.9738	0.0187
101_HSPH1_10808;421	0.1672	0.0190
130_HLA-C_3107;213	0.7960	0.0190
336_STEAP1_26872;257	0.0973	0.0190
1334_HGSNAT_138050;193	0.1127	0.0192
1419_TBL1XR1_79718;285	1.0590	0.0192
289_NDUFA8_4702;89	0.8656	0.0194
402_C12orf62_84987;37	1.0040	0.0194
571_CMTM4_146223;133	0.2155	0.0196
639_HECTD1_25831;429	0.5327	0.0196
523_TRMT5_57570;49	1.1709	0.0197
1010_LDLR_3949;797	0.1348	0.0200
1343_DAAM1_23002;381	0.5539	0.0203
1386_RALGAPA2_57186;85	0.2588	0.0203
979_RPS2_6187;21	0.6197	0.0204
556_POGZ_23126;1101	0.9573	0.0206
655_CCT2_10576;517	1.9199	0.0207
65_ZKSCAN1_7586;473	1.4761	0.0207
559_BSDC1_55108;169	1.3160	0.0208
653_NUFIP2_57532;9	1.0511	0.0208
656_ANK3_288;797	0.1351	0.0209

(continued)

peptide_id	difference	tstat_pvalues
872_RASD1_51655;261	1.2922	0.0209
885_TBCD_6904;969	0.1575	0.0209
198_ATP9A_10079;325	0.1770	0.0213
93_PABPC3_5042;421	0.7367	0.0213
386_LRRC8A_56262;265	0.1778	0.0216
854_SORL1_6653;1365	0.9160	0.0217
PCAT1_T351126_G082910_2_1371_1655_285;25	0.3069	0.0219
538_THRAP3_9967;73	0.3852	0.0219
1321_MLLT4_4301;765	0.3395	0.0221
474_RPL18A_6142;17	0.2767	0.0222
ADT12;129	0.6684	0.0223
2_AR_367;325	1.0911	0.0223
61_KLC3_147700;145	1.1848	0.0224
731_F5_2153;117	1.0325	0.0224
1178_C12orf51_283450;649	0.1925	0.0225
163_ATP13A3_79572;121	1.0535	0.0225
966_TARS_6897;433	1.0728	0.0225
683_MRPL3_11222;25	0.1831	0.0226
830_ZNF664_144348;97	0.8137	0.0228
266_UTRN_7402;401	0.9203	0.0230
1196_ZFP62_643836;265	1.3867	0.0231
1196_ZFP62_643836;293	1.3867	0.0231
314_TUG1_55000;25	0.2978	0.0232
677_ZNF598_90850;753	0.6518	0.0232
1046_SRRT_51593;605	0.2071	0.0233
978_CHERP_10523;413	1.1558	0.0235
448_GOLGA4_2803;229	0.9787	0.0236
646_FASTK_10922;473	0.1410	0.0239
1032_C1orf9_51430;161	0.1158	0.0241
1158_SEMA4C_54910;789	1.0786	0.0245
893_UPF1_5976;693	0.2995	0.0245
1386_RALGAPA2_57186;81	0.3405	0.0246
1062_BCLAF1_9774;73	0.8133	0.0247
843_DDB1_1642;1065	0.7300	0.0247
1146_SSBP1_6742;1	0.1938	0.0248
608_CCDC6_8030;301	1.1939	0.0248
ADT3;513	1.0341	0.0249
1174_GAS5_60674;329	0.2048	0.0249
610_USP9X_8239;1857	0.1192	0.0250
1269_SLC37A3_84255;209	0.1080	0.0254
311_EFNA1_1942;173	0.1627	0.0254
677_ZNF598_90850;285	0.6974	0.0256
872_RASD1_51655;89	0.1083	0.0256
1037_SUZ12_23512;89	0.3501	0.0257
1387_RPA1_6117;205	0.6860	0.0257
630_AHCYL1_10768;349	1.2614	0.0258
923_MRPS6_64968;57	0.1107	0.0258
497_STARD3NL_83930;101	0.1306	0.0259
1133_SMARCA1_6594;109	1.0842	0.0260
1318_GANAB_23193;773	0.9192	0.0260
529_BTBD6_90135;341	0.5582	0.0260
1258_SRSF10_10772;105	0.5432	0.0263
692_ANAPC5_51433;269	0.1629	0.0266
1055_KIAA0146_23514;89	1.0008	0.0268
775_KIAA2013_90231;137	0.1772	0.0268
PRCAT104.4_T230577_G053084_2_3913_4350_438;9	0.4732	0.0269
1072_CTSB_1508;197	1.0329	0.0269
632_NOS1_4842;81	0.2329	0.0272
781_ATP1A1_476;301	0.1147	0.0274
MIR205HG.2_T029617_G006597_1_387_779_393;21	0.1765	0.0275
PRCAT3.1_T183162_G042038_1_6945_7547_603;133	0.3818	0.0275
1378_CHD7_55636;2017	0.7821	0.0278
PRCAT236_T273012_G063396_2_1244_1561_318;53	0.1859	0.0280
144_SPON2_10417;233	0.1545	0.0280
CTA19;117	0.7004	0.0281
285_NPRL3_8131;397	0.1337	0.0281
21_GTF2I_2969;777	0.6205	0.0282
354_SLC12A7_10723;981	0.8015	0.0283
407_EDEM3_80267;645	0.3003	0.0284
974_EIF3L_51386;349	0.7739	0.0284
1076_TANC1_85461;33	1.4025	0.0285
1256_C19orf28_126321;325	0.5919	0.0286
1007_SEL1L3_23231;1037	0.1395	0.0287
1007_SEL1L3_23231;169	0.2695	0.0289
951_ENTPD6_955;61	0.2075	0.0289
1196_ZFP62_643836;381	1.3175	0.0290
477_SRSF5_6430;89	0.9165	0.0293
1431_ZNF462_58499;1653	0.1964	0.0297

(continued)

peptide_id	difference	tstat_pvalues
17_CYTB_4519;89	0.1066	0.0297
692_ANAPC5_51433;709	0.1669	0.0299
188_TOMM7_54543;33	1.1299	0.0301
389_RPL6_6128;41	0.2497	0.0304
470_SCCPDH_51097;293	0.1407	0.0304
1032_C1orf9_51430;781	0.1501	0.0305
912_EIF2B1_1967;25	1.0086	0.0305
1064_TMEM184B_25829;305	0.1544	0.0306
37_RPL34_6164;73	0.5731	0.0306
77_SP100_6672;281	0.1798	0.0307
PRO3;525	0.2980	0.0308
CTA6;93	0.8390	0.0309
1062_BCLAF1_9774;61	0.5798	0.0312
1062_BCLAF1_9774;85	0.4328	0.0313
946_NKTR_4820;485	0.6716	0.0313
1433_SASH1_23328;993	0.8614	0.0315
CAT1686_T107057_G025268_2_4537_5037_501;137	0.4395	0.0316
818_ILF3_3609;237	0.9665	0.0316
1340_VSIG10L_147645;789	0.1461	0.0317
178_ZC3H11A_9877;697	1.2809	0.0317
449_GALNT11_63917;17	0.1474	0.0318
PRCAT28.1_T183157_G042036_1_3585_4718_1134;5	0.4663	0.0319
PRCAT28.4_T183159_G042036_1_1382_2515_1134;5	0.4663	0.0319
833_ZMPSTE24_10269;381	0.0905	0.0319
1306_SLC38A1_81539;361	0.1330	0.0320
1335_KTN1_3895;1101	1.3326	0.0320
379_ELP2_55250;201	0.8238	0.0320
PRCAT193_T143309_G033348_1_7398_8183_786;237	0.5505	0.0324
1041_MLL_4297;2129	0.2140	0.0326
267_ACTB_60;113	1.2223	0.0327
431_ACTG1_71;113	1.2223	0.0327
1392_C20orf108_116151;149	0.2062	0.0328
309_TMC8_147138;537	0.0883	0.0328
1248_DHTKD1_55526;21	0.3491	0.0329
749_STEAP2_261729;305	0.1247	0.0331
29_HSP90B1_7184;685	0.2225	0.0333
1072_CTSB_1508;53	0.6642	0.0335
411_PPDPF_79144;5	0.4175	0.0337
854_SORL1_6653;481	0.1991	0.0338
847_TMC5_79838;741	0.2156	0.0340
163_ATP13A3_79572;245	0.2478	0.0341
543_RHOU_58480;13	0.6965	0.0342
ADT14;669	1.1259	0.0343
231_B4GALT1_2683;21	0.1181	0.0343
611_KIAA1429_25962;121	0.4611	0.0344
959_NCOR2_9612;1605	0.8209	0.0344
1080_IRX4_50805;501	1.3802	0.0345
191_CMTM6_54918;137	0.1211	0.0345
1082_MCM7_4176;513	0.9584	0.0346
1178_C12orf51_283450;1365	0.6867	0.0347
1137_KAT6A_7994;733	0.1366	0.0348
314_TUG1_55000;41	1.0880	0.0348
749_STEAP2_261729;405	0.1576	0.0349
5_GGT6_124975;185	0.2371	0.0353
491_KDELR2_11014;145	0.1176	0.0355
918_MT1G_4495;41	1.0365	0.0355
935_HIPK3_10114;1049	0.8737	0.0355
1362_ADD1_118;381	0.3540	0.0357
591_ECHDC2_55268;1	0.1945	0.0359
869_COX7B_1349;41	0.1017	0.0363
165_LSM14A_26065;397	0.6848	0.0364
CAT634.1_T277460_G064362_2_111_689_579;69	0.6698	0.0367
1137_KAT6A_7994;29	0.9300	0.0369
1076_TANC1_85461;705	0.1408	0.0371
33_ANKK1_255239;677	0.6941	0.0372
676_EIF3A_8661;481	0.8135	0.0372
77_SP100_6672;305	0.9443	0.0374
PRO30;837	1.0270	0.0375
1108_CUX1_1523;1285	0.1900	0.0375
490_ABCC5_10057;329	0.2442	0.0375
1178_C12orf51_283450;257	0.6901	0.0376
833_ZMPSTE24_10269;209	0.1445	0.0376
478_VDAC1_7416;37	1.2779	0.0377
562_COX17_10063;53	0.1181	0.0377
1361_IARS_3376;725	0.4215	0.0380
1196_ZFP62_643836;217	1.1239	0.0383
1375_ARID4B_51742;601	0.0943	0.0383

(continued)

peptide_id	difference	tstat_pvalues
692_ANAPC5_51433;129	0.1322	0.0383
1099_WDR90_197335;585	0.3226	0.0384
881_PYCR1_5831;197	0.7078	0.0384
1144_ATP5I_521;5	0.1766	0.0385
1156_PREPL_9581;509	0.7617	0.0385
306_LBR_3930;329	0.0773	0.0388
611_KIAA1429_25962;1717	0.7533	0.0388
96_ND1_4535;1837	0.1898	0.0389
1041_MLL_4297;3925	0.1623	0.0391
1037_SUZ12_23512;109	0.2470	0.0393
PRO14;2497	1.0935	0.0394
743_AK4_205;41	1.4307	0.0394
364_TCF25_22980;605	0.0923	0.0395
1419_TBL1XR1_79718;421	1.0343	0.0397
884_SAP18_10284;141	0.5235	0.0398
1010_LDLR_3949;737	1.0057	0.0400
843_DDB1_1642;1069	0.6412	0.0401
975_AGT_183;45	0.8280	0.0401
430_NPC2_10577;1	0.1377	0.0402
CAT1454_T071522_G016513_1_1122_1892_771;137	0.0823	0.0403
519_NBPF11_200030;497	0.4605	0.0403
1138_SNRNP200_23020;1081	0.2192	0.0404
86_MCL1_4170;177	0.5519	0.0404
542_TM2D3_80213;1	0.5211	0.0405
183_CD24_100133941;65	0.1017	0.0406
456_PDCD6_10016;153	0.1276	0.0406
56_ALPK1_80216;1233	0.9445	0.0408
725_GAA_2548;265	0.1014	0.0409
CAT1686_T107057_G025268_1_5122_5934_813;113	0.8336	0.0410
1207_PAFAH1B1_5048;341	0.1112	0.0411
141_CAMKK2_10645;21	1.1280	0.0411
29_HSP90B1_7184;505	0.8768	0.0411
1199_PRKAR2A_5576;297	0.1654	0.0412
459_CLTC_1213;785	0.0893	0.0413
1242_THOC2_57187;385	0.1807	0.0414
653_NUFIP2_57532;381	1.0389	0.0416
BOLA3_AS1_T191973_G044207_2_402_665_264;1	0.7923	0.0417
CTA1;69	0.7611	0.0418
1108_CUX1_1523;1289	0.1519	0.0419
1280_FZD4_8322;265	0.1153	0.0420
739_RGS10_6001;81	0.6912	0.0421
682_ARFGAP3_26286;357	0.7731	0.0423
PRO18;209	0.5309	0.0424
1039_ABAT_18;425	0.2485	0.0424
842_NEURL1B_54492;517	0.3419	0.0425
797_MARCH6_10299;373	0.1323	0.0427
1236_KLHDC10_23008;73	0.2909	0.0428
1285_MICAL2_9645;969	1.0813	0.0428
639_HECTD1_25831;985	0.2287	0.0428
1160_CYFIP1_23191;209	0.4152	0.0429
255_RANBP2_5903;301	0.6503	0.0429
1261_PIGT_51604;381	0.0794	0.0431
639_HECTD1_25831;1841	0.1438	0.0431
1108_CUX1_1523;1405	1.0791	0.0432
549_REPIN1_29803;69	0.1650	0.0433
1269_SLC37A3_84255;93	0.1795	0.0435
660_GRB10_2887;389	0.1690	0.0435
701_ECI1_1632;13	0.3184	0.0435
PRCAT116_T182782_G041985_2_2888_3328_441;85	0.4604	0.0436
1178_C12orf51_283450;2965	0.1336	0.0439
552_ADAR_103;909	0.1193	0.0439
PRCAT282_T344692_G081083_1_3008_3415_408;117	0.2234	0.0440
366_TMEFF2_23671;305	0.1232	0.0440
CTA15;193	0.0882	0.0441
1004_TRIP12_9320;1145	0.6571	0.0441
337_ATP1B3_483;89	0.7357	0.0441
820_MON1B_22879;353	0.4067	0.0441
1196_ZFP62_643836;741	1.0218	0.0443
150_YWHAG_7532;1	1.1831	0.0443
516_TACSTD2_4070;121	0.2607	0.0443
1406_CCDC85C_317762;109	1.2628	0.0445
1377_TRIM2_23321;601	0.3587	0.0447
CTA11;49	0.3507	0.0448
5_GGT6_124975;181	0.1992	0.0449
351_CCT4_10575;469	0.6923	0.0452
1221_COMMD3_23412;129	0.6045	0.0453
422_EIF3C_8663;569	0.1625	0.0454
CAT2217.1_T227650_G052508_2_19310_20443_1134;33	0.1023	0.0457

(continued)

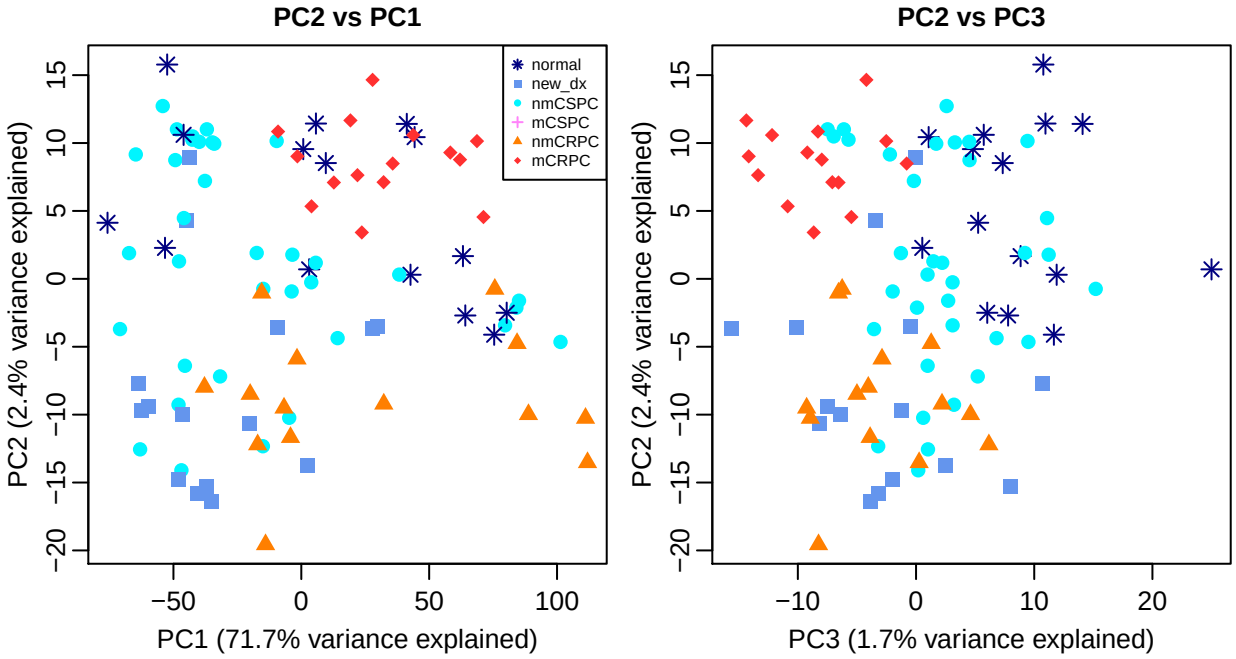
peptide_id	difference	tstat_pvalues
CAT2217.3_T227648_G052508_2_18684_19817_1134;33	0.1023	0.0457
CAT2217.4_T227647_G052508_2_13944_15077_1134;33	0.1023	0.0457
633_USP54_159195;1001	1.0586	0.0457
120_NPIPL3_23117;85	0.3616	0.0459
30_LOC100132247_100132247;85	0.3616	0.0459
152_FAM162A_26355;73	0.8205	0.0460
449_GALNT11_63917;405	1.0908	0.0460
578_WSB1_26118;409	0.1331	0.0462
1364_DHCR7_1717;337	0.1251	0.0464
1405_ATXN2_6311;1221	0.8494	0.0468
514_NR2F2_7026;377	0.0839	0.0468
7_HDAC10_83933;589	0.2989	0.0468
818_ILF3_3609;225	0.1098	0.0469
865_PSME1_5720;157	0.7166	0.0469
1402_A2M_2;357	0.1071	0.0470
CALML3_AS1.5_T036343_G008259_2_385_1059_675;1	0.3479	0.0472
1151_SAV1_60485;33	0.2578	0.0472
CAT625.2_T275901_G064124_1_899_2041_1143;221	0.0805	0.0473
1431_ZNF462_58499;1389	0.9080	0.0473
251_COX2_4513;33	0.5396	0.0473
549_REPIN1_29803;389	0.7453	0.0473
1108_CUX1_1523;837	1.1472	0.0474
985_CALD1_800;509	1.0753	0.0474
1189_HDAC1_3065;21	0.4489	0.0476
689_ROMO1_140823;33	0.1456	0.0476
ADT14;381	0.1201	0.0477
1129_EZH2_2146;273	0.1525	0.0478
1380_PSMD3_5709;97	0.1774	0.0480
16_ORAI1_84876;121	0.0922	0.0481
300_GNL3_26354;241	0.6180	0.0481
1252_SMG1_23049;1285	0.1774	0.0483
284_DSP_1832;1533	0.6084	0.0486
985_CALD1_800;549	1.1548	0.0486
813_RPS29_6235;5	0.5594	0.0487
1031_SOCS7_30837;293	1.1408	0.0488
PRO3;37	0.5806	0.0489
894_PPP1R16A_84988;321	0.3836	0.0489
CTA11;133	1.0684	0.0492
587_BTF3_689;161	0.7396	0.0492
1127_IGF1R_3480;209	0.5713	0.0494
46_ERGIC1_57222;129	0.8547	0.0494
PRCAT67_T183175_G042042_1_16548_18869_2322;317	0.6842	0.0496
1424_ITPRIPL2_162073;253	0.2101	0.0496
PCA29;305	0.5536	0.0498
1034_OCIAD2_132299;97	0.1671	0.0498
361_EIF1_10209;25	0.7168	0.0499
369_KIAA1244_57221;1709	0.1818	0.0499

5 Tests on Fluorescence Levels with Binary-calls-filtering

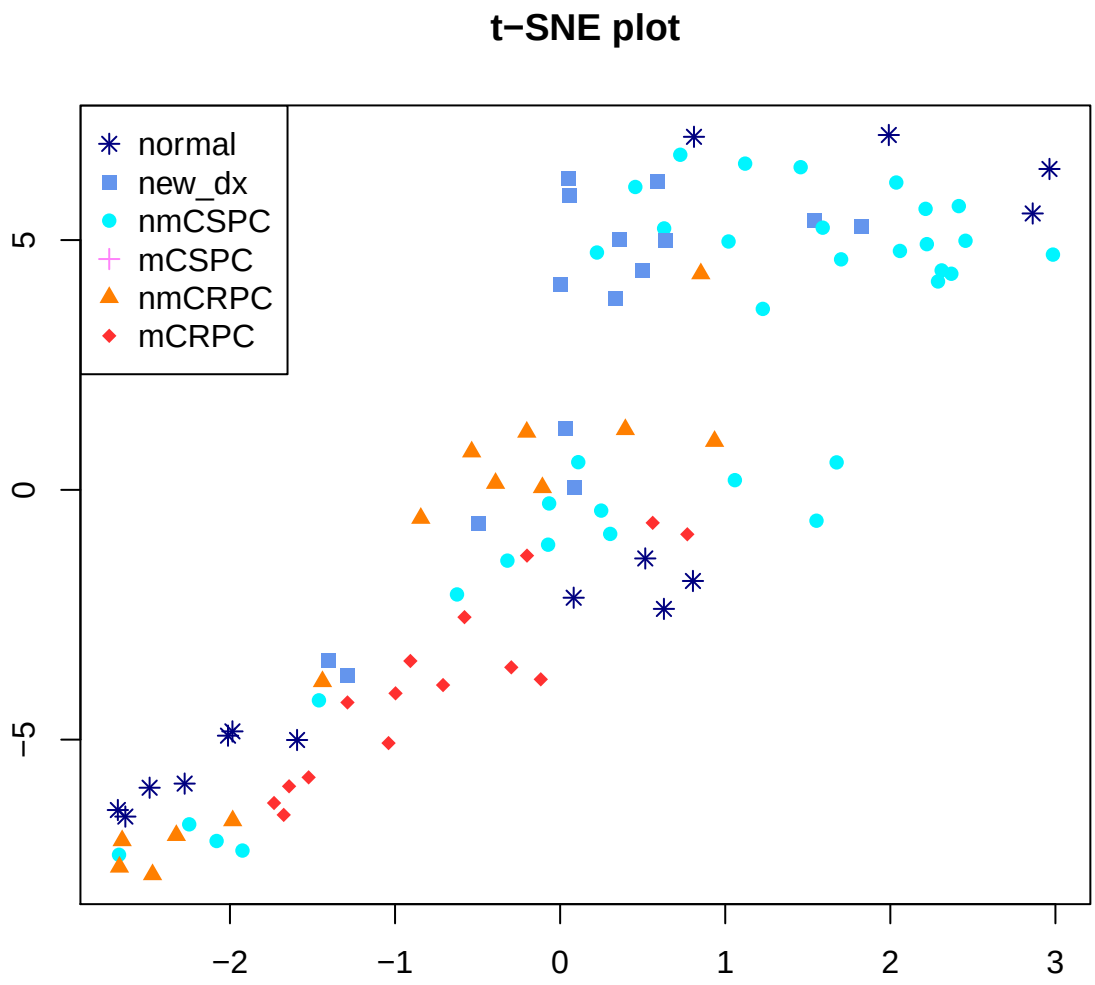
Previously we identified 3128 significant peptides at 5% BH FDR. Out of these peptides, 698 of them also have at least one call among all patients. We will use these calls as filtering criterion on the significant peptides – in this section, we will examine whether zooming in on these 698 significant peptides help to improve clustering patterns that we would observe in PCA, t-SNE and heatmap.

5.1 Visualization based on Filtered Significant Peptides

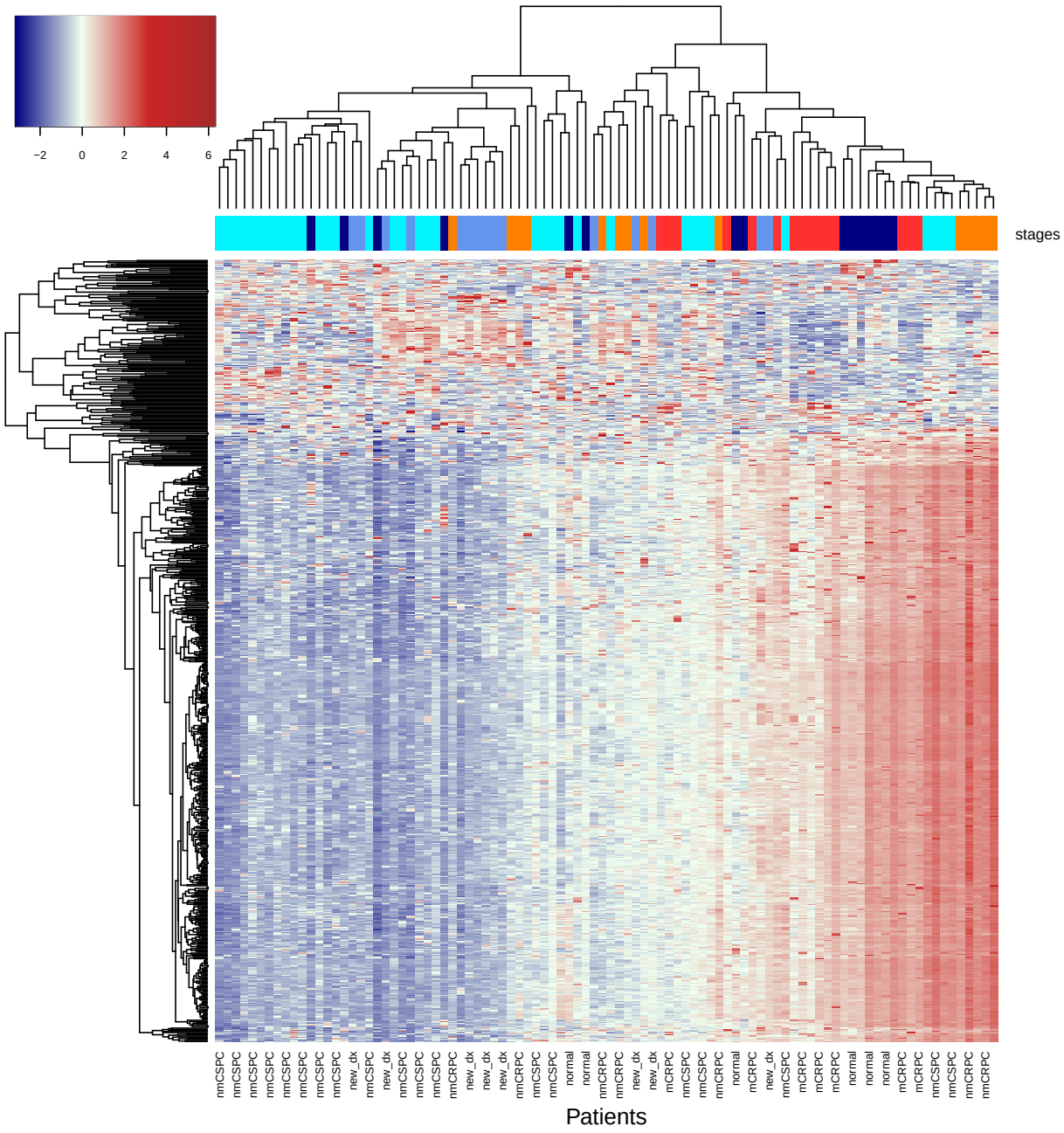
Based on the 698 significant peptides, the PCA plots become



The clustering patterns here is similar to the pattern we observe in the PCA plot for all 3128 peptides. The t-SNE plot now becomes

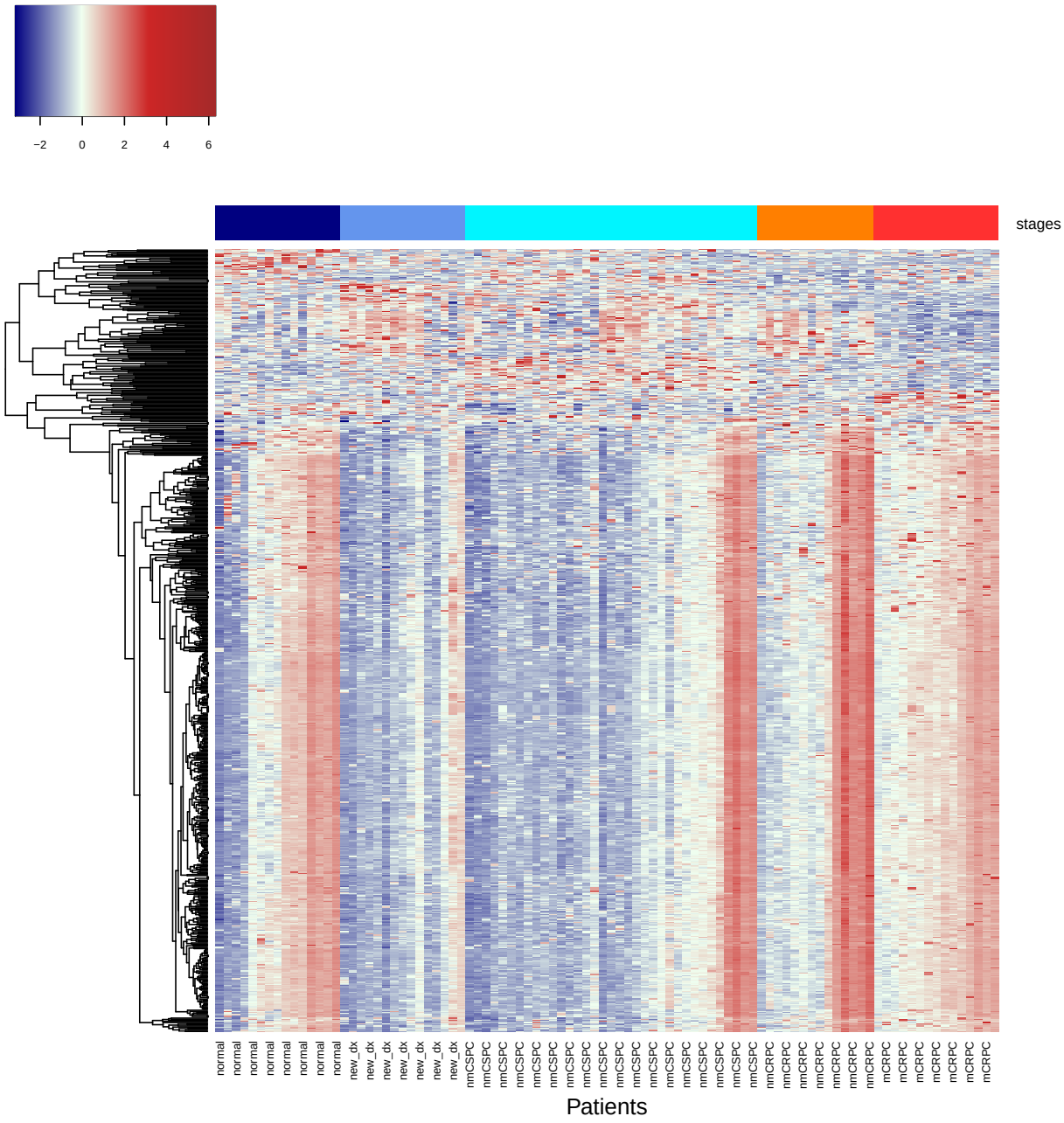


Again, clustering pattern is somewhat similar. Finally, the heatmap becomes



The clustering pattern that reveals cancer stage effect is more obvious in the previous section (when all 3128 peptides are considered) than what we observe here.

We also plot the heatmap where we order the columns according to patients' stages instead of the default cluster dendrogram order.



5.2 Gene Set Analysis

We repeat our gene set enrichment analysis, but only with the proteins that are associated with peptides with at least one call among all patients. Hence, in this case, we only consider 1437 proteins with matching

genes, and 395 of them are associated with significant peptides with at least one call among all patients. We use this information to perform our gene set analysis.

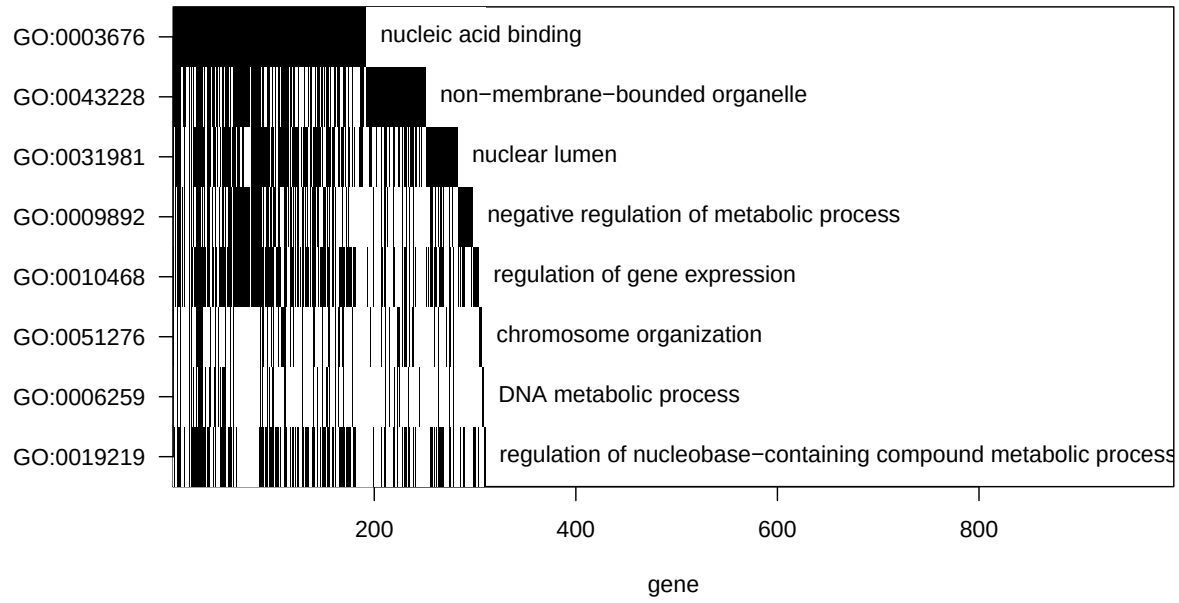
```
# gene-set analysis via allez!
allez.go <- allez(seq_id_ok, lib = "org.Hs.eg", idtype = "ENTREZID", sets = "GO")

# nominal alpha for enrichment score
nom.alpha <- 0.05

# minimum number of genes in gene-set to be considered
min.num.gene <- 2
```

We display the waterfall plot of the gene-set-analysis results.

```
# Waterfall plot
allezPlot(allez.go, n.cell = min.num.gene, nominal.alpha = nom.alpha)
```



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 (nucleic acid binding GO:0003676) 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 (non-membrane-bounded organelle GO: 0043228). 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 with at least one call among all patients.

Term	Ontology	set.mean	set.size	z.score	in.genes
nucleosome	CC	0.7692	20/26	5.6814	H1F0; HIST1H1C; HIST1H2AD; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; HIST1H2AG; H2AFY; HP1BP3; H2AFJ; HIST1H2BK; HIST3H2A; H2AFV; HIST2H2AA4

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
DNA packaging complex	CC	0.7692	20/26	5.6814	H1F0; HIST1H1C; HIST1H2AD; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; HIST1H2AG; H2AFY; HP1BP3; H2AFJ; HIST1H2BK; HIST3H2A; H2AFV; HIST2H2AA4
histone binding	MF	0.7000	14/20	4.2746	MSH6; KMT2A; NPM1; PHF2; BRD2; TAF7; PHF8; BRD1; LEF1; UIMC1; YEATS2; CHD8; TBL1XR1; ING5
nucleosome assembly	BP	0.6957	16/23	4.5420	H1F0; HIST1H1C; HMGB2; NAP1L1; NPM1; BRD2; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; H2AFY; HP1BP3; TSPYL5
chromatin assembly	BP	0.6957	16/23	4.5420	H1F0; HIST1H1C; HMGB2; NAP1L1; NPM1; BRD2; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; H2AFY; HP1BP3; TSPYL5
chromatin assembly or disassembly	BP	0.6923	18/26	4.7959	H1F0; HIST1H1C; HMGB2; NAP1L1; NPM1; BRD2; SMARCC2; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; H2AFY; HP1BP3; TSPYL5; HIST3H2A
protein-DNA complex	CC	0.6842	26/38	5.7104	PARP1; XRCC6; H1F0; HIST1H1C; HIST1H2AD; MCM3; NPM1; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; HIST1H2AG; H2AFY; NUPR1; HP1BP3; LEF1; H2AFJ; HIST1H2BK; HIST3H2A; H2AFV; HIST2H2AA4
DNA packaging	BP	0.6667	16/24	4.3210	H1F0; HIST1H1C; HMGB2; NAP1L1; NPM1; BRD2; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; H2AFY; HP1BP3; TSPYL5
nucleosome organization	BP	0.6667	18/27	4.5882	H1F0; HIST1H1C; HMGB2; NAP1L1; NPM1; BRD2; SMARCC2; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; H2AFY; HP1BP3; TSPYL5; HIST3H2A
chromatin organization	BP	0.6076	48/79	6.7931	ACTB; BCL6; EZH2; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; KMT2A; NAP1L1; NPM1; PHF2; BRD2; SMARCA1; SMARCC2; TAF7; TDG; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; TRIP12; H2AFY; ZMPSTE24; MORF4L1; CBX3; PHF8; RYBP; BRD1; HP1BP3; LEF1; UIMC1; ARID4B; PHF10; CHD7; YEATS2; CHD8; ZNF462; NUCKS1; ING5; TSPYL5; HIST3H2A
protein-DNA complex assembly	BP	0.5789	22/38	4.2391	PARP1; CENPE; H1F0; HIST1H1C; HMGB1; HMGB2; NAP1L1; NPM1; BRD2; RPS27A; UBA52; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; H2AFY; POGZ; HP1BP3; TSPYL5
DNA conformation change	BP	0.5750	23/40	4.2958	PARP1; XRCC6; H1F0; HIST1H1C; HMGB1; HMGB2; NAP1L1; NPM1; BRD2; RPS27A; SMARCA1; UBA52; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; H2AFY; HP1BP3; CHD8; TSPYL5
nuclear chromatin	CC	0.5714	24/42	4.3526	ACTB; AR; EZH2; MSH6; H1F0; HIST1H1C; HDAC1; HMGB2; HNRNPC; JUN; SMARCA1; SMARCC2; HIST3H3; HIST1H4C; HIST1H4L; EED; H2AFY; NCOR2; MORF4L1; CBX3; POGZ; PHF10; NUCKS1; HIST3H2A
nuclear chromosome part	CC	0.5667	34/60	5.1530	ACTB; PARP1; AR; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HDAC1; HMGB2; HNRNPC; JUN; MCM3; PPP1CC; UPF1; SMARCA1; SMARCC2; SP100; SSB; HIST3H3; HIST1H4C; HIST1H4L; EED; H2AFY; NCOR2; P3H4; MORF4L1; CBX3; POGZ; REPIN1; PHF10; THOC2; NUCKS1; HIST3H2A
nuclear chromosome	CC	0.5645	35/62	5.2034	ACTB; PARP1; AR; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HDAC1; HMGB2; HNRNPC; JUN; MCM3; PPP1CC; UPF1; SMARCA1; SMARCC2; SP100; SSB; HIST3H3; HIST1H4C; HIST1H4L; EED; H2AFY; NCOR2; P3H4; MORF4L1; CBX3; POGZ; SPIDR; REPIN1; PHF10; THOC2; NUCKS1; HIST3H2A
chromatin	CC	0.5600	42/75	5.6612	ACTB; AR; EZH2; MSH6; H1F0; HIST1H1C; HIST1H2AD; HDAC1; HMGB2; HMGN1; HNRNPC; JUN; RAN; UPF1; SMARCA1; SMARCC2; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; HIST1H2AG; H2AFY; NCOR2; MORF4L1; CBX3; POGZ; NOP53; HP1BP3; PHF10; H2AFJ; FAM111A; NUCKS1; HIST1H2BK; HIST3H2A; H2AFV; HIST2H2AA4
protein-DNA complex subunit organization	BP	0.5581	24/43	4.2078	PARP1; CENPE; H1F0; HIST1H1C; HMGB1; HMGB2; NAP1L1; NPM1; BRD2; RPS27A; SMARCC2; UBA52; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; H2AFY; POGZ; HP1BP3; TSPYL5; HIST3H2A
chromatin binding	MF	0.5472	29/53	4.5068	ACTB; AR; BCL6; DDX5; EZH2; MSH6; H1F0; HIST1H1C; HDAC1; HMGN1; HNRNPC; RBPJ; JUN; NFIA; NFE2L1; RAN; UPF1; BRD2; SMARCA1; SMARCC2; SSRP1; H2AFY; PHF8; NUP62; NUPR1; HP1BP3; CHD7; CHD8; NUCKS1
chromosome organization	BP	0.5410	66/122	6.8609	ACTB; PARP1; BCL6; CENPE; EZH2; XRCC6; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; KPNB1; KMT2A; NAP1L1; NPM1; PHF2; RAN; UPF1; BRD2; RPS27A; SMARCA1; SMARCC2; SP100; TAF7; TDG; UBA52; KAT6A; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; TRIP12; H2AFY; ZMPSTE24; CCT2; P3H4; MORF4L1; CBX3; SMG1; POGZ; PHF8; SUN1; RYBP; NUP62; BRD1; GNL3; HP1BP3; LEF1; UIMC1; ARID4B; XRN1; PHF10; NOP10; CHD7; YEATS2; CHD8; ZNF462; NUCKS1; ING5; TSPYL5; HIST3H2A

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
chromosomal part	CC	0.5182	57/110	5.9269	ACTB; PARP1; AR; BCL6; CENPE; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HIST1H2AD; HDAC1; HMGB2; HMGN1; HNRNPC; JUN; MCM3; SEPTIN2; PHF2; PPP1CC; PPP2CB; RAN; UPF1; SEC13; SMARCA1; SMARCC2; SP100; SSB; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; HIST1H2AG; H2AFY; NCOR2; P3H4; MORF4L1; CBX3; POGZ; REPIN1; NOP53; HP1BP3; PHF10; H2AFJ; THOC2; FAM111A; NUCKS1; HIST1H2BK; HIST3H2A; H2AFV; HIST2H2AA4
chromosome	CC	0.5126	61/119	6.0441	ACTB; PARP1; AR; BCL6; CENPE; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HIST1H2AD; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; JUN; MCM3; SEPTIN2; PHF2; PPP1CC; PPP2CB; RAN; UPF1; SEC13; SMARCA1; SMARCC2; SP100; SSB; SSRP1; KAT6A; HIST3H3; HIST1H2AK; HIST1H2AM; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; EED; HIST1H2AG; H2AFY; NCOR2; P3H4; MORF4L1; CBX3; POGZ; SPIDR; REPIN1; NOP53; HP1BP3; PHF10; H2AFJ; THOC2; FAM111A; NUCKS1; HIST1H2BK; HIST3H2A; H2AFV; TOP1MT; HIST2H2AA4
double-stranded DNA binding	MF	0.5000	34/68	4.2418	ACTB; PARP1; AR; ATF4; EZH2; XRCC6; MSH6; H1F0; HDAC1; HMGB1; HMGB2; HNRNPC; HSPD1; RBPJ; JUN; KMT2A; NFIA; NFE2L1; NFIB; NFIC; NPAS2; SMARCC2; SOX4; TDG; EED; H2AFY; THRAP3; ZMPSTE24; LSM14A; LEF1; CHD7; ZNF395; NUCKS1; CREB3L4
DNA metabolic process	BP	0.4867	55/113	5.2342	PARP1; BCL6; EZH2; XRCC6; GNAS; MSH6; H1F0; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HSPD1; JUN; KPNB1; MCM3; KMT2A; NAP1L1; NFIA; NFIB; NFIC; NPAS2; NPM1; RAN; UPF1; RPS27A; SMARCA1; SP100; SSRP1; TCEA1; TDG; UBA52; HIST3H3; HIST1H4C; HIST1H4L; TRIP12; MORF4L2; ZMPSTE24; CCT2; MORF4L1; SMG1; SPIDR; GTPBP4; GNL3; REPIN1; NOP53; LEF1; UIMC1; XRN1; NOP10; FAM111A; NUCKS1; ING5; TBRG1; HIST3H2A
nucleolus	CC	0.4836	59/122	5.3772	ADAR; PARP1; CAPG; CTSB; DDX5; XRCC6; HMGB2; RBPJ; ILF3; MARS; HNRNPM; NFIB; NFIC; NPM1; PHF2; PPP1CC; RAN; RPL3; RPL5; RPL7; RPL12; RPL13; RPL23A; RPL26; RPS3A; RPS6; RPS7; RPS25; RPS27A; SSRP1; TCEA1; KAT6A; FXR1; USO1; EIF3A; H2AFY; MORF4L2; NOP56; P3H4; RPL35; PHF8; RPL13A; GTPBP4; RPL36; DCAF13; SPATS2L; GNL3; NOP53; ARL6IP4; MAGEC2; IP6K2; NOP58; NOP10; CHD7; TCIM; FAM111A; NUCKS1; DDX50; TSEN34
DNA-binding transcription factor activity	MF	0.4824	41/85	4.3977	PARP1; AR; ATF4; BCL6; EZH2; GOLGB1; GTF2I; GTF3A; HDAC1; RBPJ; JUN; KMT2A; NFIA; NFE2L1; NFIB; NFIC; NPAS2; SMARCA1; SMARCC2; SOX4; SP100; SSRP1; TAF7; ZKSCAN1; TSC22D1; ZRANB2; NCOR2; TCF25; POGZ; LEF1; ARID4B; YEATS2; ZNF395; BBX; ZNF462; NUCKS1; IRX3; ZNF587; ZNF664; CREB3L4; ZFP62
DNA binding	MF	0.4703	87/185	6.3564	ACTB; ADAR; PARP1; AR; ATF4; BCL6; EZH2; XRCC6; GOLGB1; MSH6; GTF2I; GTF3A; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HSPD1; RBPJ; ILF3; JUN; MCM3; KMT2A; NACA; NFIA; NFE2L1; NFIB; NFIC; NPAS2; NPM1; PNN; UPF1; RPL6; RPL7; SMARCA1; SMARCC2; SON; SOX4; SP100; SSRP1; TAF7; TCEA1; TDG; ZKSCAN1; KAT6A; TAF15; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; KHSRP; DDX3Y; EED; H2AFY; BCLAF1; THRAP3; DNAJB6; SRRM1; ZMPSTE24; RAI1; TCF25; SMG1; RYBP; LSM14A; NUPR1; REPIN1; HP1BP3; LEF1; CXXC5; SRRT; XRN1; ZFAND6; CHD7; ZNF395; BBX; CHD8; NUCKS1; IRX3; TBL1XR1; ZNF587; HIST3H2A; TOP1MT; ZNF664; CREB3L4
negative regulation of nucleobase-containing compound metabolic process	BP	0.4400	66/150	4.7648	PARP1; AR; BCL6; CDKN1C; DDX5; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HNRNPC; HSPA1A; RBPJ; ILF3; JUN; LIMS1; MAGEA1; KMT2A; NFIB; NFIC; PHF2; RPS26; RPS27A; SRSF4; SMARCC2; SP100; SSX1; TAF7; TDG; UBA52; SF1; KAT6A; HIST1H4C; HIST1H4L; EED; TRIP12; H2AFY; NCOR2; BCLAF1; THRAP3; DNAJB6; ZMPSTE24; SYNCRIP; HEXIM1; SRSF10; PKP3; CBX3; TCF25; PHF8; RYBP; GTPBP4; NUPR1; PABPC1; PDCD4; NOP53; LEF1; CXXC5; UIMC1; XRN1; YEATS2; CHD8; SECISBP2; TBL1XR1
negative regulation of RNA metabolic process	BP	0.4317	60/139	4.3347	PARP1; AR; BCL6; CDKN1C; DDX5; EZH2; XRCC6; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HNRNPC; HSPA1A; RBPJ; ILF3; JUN; LIMS1; MAGEA1; NFIB; NFIC; PHF2; RPS26; RPS27A; SRSF4; SMARCC2; SP100; SSX1; TAF7; TDG; UBA52; SF1; KAT6A; HIST1H4C; HIST1H4L; EED; H2AFY; NCOR2; BCLAF1; THRAP3; DNAJB6; ZMPSTE24; SYNCRIP; HEXIM1; SRSF10; PKP3; CBX3; TCF25; PHF8; RYBP; PABPC1; PDCD4; NOP53; LEF1; CXXC5; UIMC1; YEATS2; CHD8; SECISBP2; TBL1XR1

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
negative regulation of gene expression	BP	0.4051	111/274	5.3463	A2M; ADAR; PARP1; AR; ATF4; BCL6; CAST; CDKN1C; DDX5; EZH2; XRCC6; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HNRNPC; HSPA1A; RBPJ; ILF3; JUN; LDLR; LIMS1; CAPRIN1; MAGEA1; RPL10A; NFIB; NFIC; NOTCH2; NPM1; PHF2; PSMA6; PSMB7; PSMD1; PSMD4; RAN; RANBP2; UPF1; RNH1; RPL3; RPL5; RPL6; RPL7; RPL12; RPL13; RPL17; RPL22; RPL23A; RPL26; RPL27; RPL29; RPL32; RPL37; RPL39; RPL36A; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SEC13; SRSF4; SMARCC2; SP100; SSB; TAF7; TDG; UBA52; KAT6A; FXR1; HIST1H4C; HIST1H4L; KHSRP; EED; RPL14; H2AFY; NCOR2; BCLAF1; THRAP3; DNAJB6; ZMPSTE24; SYNCRIP; HEXIM1; SRSF10; PKP3; RPL35; CBX3; TCF25; SMG1; PHF8; RYBP; RPL13A; NUP62; RPL36; PABPC1; PDCD4; NOP53; LEF1; CXXC5; SRRT; UIMC1; XRN1; YEATS2; CHD8; SECISBP2; TBL1XR1
nucleic acid binding	MF	0.4000	192/480	7.5246	ACTB; ACTN1; ADAR; ADD1; PARP1; AR; ATF4; BCL6; BTF3; CAST; CTNNA1; DDX5; DSP; EEF1B2; EIF1AX; EIF4A1; EZH2; FAU; XRCC6; GOLGB1; MSH6; GTF2I; GTF3A; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HNRNPH1; HSPA1A; HSPD1; IARS; RBPJ; ILF3; JUN; KPNB1; KTN1; CAPRIN1; MARS; MCM3; KMT2A; MYH9; NACA; HNRNPM; NAP1L1; RPL10A; NFIA; NFE2L1; NFIB; NFIC; NPAS2; NPM1; PNN; PPP1CC; PSMA6; PSMD4; RAN; RANBP2; RARS; UPF1; RPL3; RPL5; RPL6; RPL7; RPL12; RPL13; RPL17; RPL22; RPL23A; RPL26; RPL27; RPL29; RPL32; RPL37; RPL39; RPL36A; RPN1; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SRSF4; SMARCA1; SMARCC2; SNRPD2; SON; SOX4; SP100; SSB; SSRP1; SSX1; TAF7; TCEA1; TDG; YWHAG; SF1; ZKSCAN1; KAT6A; FXR1; TAF15; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; KHSRP; USO1; DDX3Y; EIF3A; EIF3J; EED; RPL14; ZRANB2; H2AFY; RBM39; BZW1; BCLAF1; ZC3H11A; THRAP3; DNAJB6; FARSB; EIF1; SRRM1; ZMPSTE24; SYNCRIP; NOP56; HEXIM1; RAI1; SRSF10; GCN1; IMMT; RPL35; ATXN2L; TCF25; SMG1; RYBP; RPL13A; SRRM2; GTPBP4; RPL36; DCAF13; PNISR; VIRMA; SPATS2L; LSM14A; GNL3; NUPR1; PABPC1; PDCD4; REPIN1; NOP53; HP1BP3; LEF1; ARL6IP4; EIF3L; CXXC5; SRRT; NOP58; XRN1; ZFAND6; ANKHD1; MRPL20; NOP10; CHD7; ZNF395; BBX; NUFIP2; CHD8; NUCKS1; DDX50; AHNAK; TSEN34; SECISBP2; IRX3; TBL1XR1; ESRP2; ZNF587; ZNF598; HIST3H2A; TOP1MT; ZNF664; CREB3L4; ZFP62; NBPF10
transcription by RNA polymerase II	BP	0.3962	84/212	4.2634	PARP1; AR; ATF4; BCL6; BTF3; CDKN1C; DDX5; DVL1; EZH2; XRCC6; GTF2I; GTF3A; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HSPA1A; HSPD1; RBPJ; JUN; MAGEA1; KMT2A; NFIA; NFE2L1; NFIB; NFIC; NOTCH2; NPAS2; NPM1; PSMA6; PSMB7; PSMD1; PSMD4; BRD2; RPS27A; SMARCA1; SMARCC2; SOX4; SP100; SSRP1; TAF7; TCEA1; ELOC; TDG; UBA52; ZKSCAN1; TAF15; TSC22D1; H2AFY; NCOR2; MORF4L2; MAML1; THRAP3; CTDSP2; NAMPT; HEXIM1; GCN1; ELL2; TCF25; POGZ; RYBP; GNL3; PDCD4; NOP53; LEF1; CXXC5; PCF11; SRRT; ARID4B; CHD7; YEATS2; ZNF395; BBX; MRTFB; CHD8; NUCKS1; IRX3; TBL1XR1; ZNF587; ZNF664; CREB3L4; ZFP62
RNA binding	MF	0.3848	137/356	5.3387	ACTN1; ADAR; ADD1; PARP1; BTF3; CAST; CTNNA1; DDX5; DSP; EEF1B2; EIF1AX; EIF4A1; EZH2; FAU; XRCC6; GOLGB1; GTF3A; H1F0; HIST1H1C; HMGB1; HMGB2; HNRNPC; HNRNPH1; HSPA1A; HSPD1; IARS; ILF3; JUN; KPNB1; KTN1; CAPRIN1; MARS; MYH9; HNRNPM; NAP1L1; RPL10A; NPM1; PNN; PPP1CC; PSMA6; PSMD4; RAN; RANBP2; RARS; UPF1; RPL3; RPL5; RPL6; RPL7; RPL12; RPL13; RPL17; RPL22; RPL23A; RPL26; RPL27; RPL29; RPL32; RPL37; RPL39; RPL36A; RPN1; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SRSF4; SNRPD2; SON; SSB; SSRP1; YWHAG; SF1; FXR1; TAF15; HIST1H4C; HIST1H4L; KHSRP; USO1; DDX3Y; EIF3A; EIF3J; RPL14; ZRANB2; RBM39; BZW1; BCLAF1; ZC3H11A; THRAP3; FARSB; EIF1; SRRM1; SYNCRIP; NOP56; HEXIM1; SRSF10; GCN1; IMMT; RPL35; ATXN2L; SMG1; RPL13A; SRRM2; GTPBP4; RPL36; DCAF13; PNISR; VIRMA; SPATS2L; LSM14A; GNL3; PABPC1; PDCD4; REPIN1; NOP53; ARL6IP4; EIF3L; SRRT; NOP58; XRN1; ANKHD1; MRPL20; NOP10; NUFIP2; NUCKS1; DDX50; AHNAK; SECISBP2; ESRP2; ZNF598; NBPF10

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
regulation of gene expression	BP	0.3839	177/461	6.3564	A2M; ACTB; ADAR; PARP1; ANK3; AR; ATF4; BCL6; CAST; CDKN1C; DDX5; DVL1; EZH2; XRCC6; GOLGB1; MKNK2; GTF2I; GTF3A; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HNRNPH1; HSPA1A; HSPD1; RBPJ; ILF3; JUN; LDLR; LIMS1; CAPRIN1; MAGEA1; MARS; KMT2A; AFDN; MYH9; RPL10A; NFIA; NFE2L1; NFIB; NFIC; NOTCH2; NPAS2; NPM1; PHF2; PPP1CC; PPP2CB; PSMA6; PSMB7; PSMD1; PSMD4; RAN; RANBP2; UPF1; BRD2; RNH1; RPL3; RPL5; RPL6; RPL7; RPL12; RPL13; RPL17; RPL22; RPL23A; RPL26; RPL27; RPL29; RPL32; RPL37; RPL39; RPL36A; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SEC13; SRSF4; SMARCA1; SMARCC2; SON; SOX4; SP100; SSB; SSRP1; SSX1; TAF7; TCEA1; ELOC; TDG; UBA52; ZKSCAN1; KAT6A; FXR1; TAF15; HIST1H4C; HIST1H4L; KHSRP; EED; TSC22D1; TAX1BP1; RPL14; ZRANB2; H2AFY; PRDX6; NCOR2; MORF4L2; BCLAF1; MAML1; THRAP3; DNAJB6; CTDSP2; NAMPT; EIF1; ZMPSTE24; SYNCRIP; HEXIM1; CAMKK2; RAI1; SRSF10; GCN1; PKP3; RPL35; CBX3; TCF25; SMG1; POGZ; PHF8; RYBP; RPL13A; NUP62; RPL36; VIRMA; LSM14A; GNL3; NUPR1; PABPC1; PDCD4; IGKV3-20; NOP53; HP1BP3; LEF1; CXXC5; SRRT; UIMC1; ARID4B; XRN1; DNAJA4; CHD7; YEATS2; ZNF395; TCIM; BBX; MRTFB; CHD8; NUCKS1; AHNAK; SECISBP2; IRX3; TBL1XR1; ESRP2; ING5; ZNF587; ZNF598; NIBAN1; ZNF664; CREB3L4; ZFP62
regulation of RNA metabolic process	BP	0.3825	127/332	4.9895	ACTN1; PARP1; AR; ATF4; BCL6; CDKN1C; DDX5; DVL1; EZH2; XRCC6; GOLGB1; GTF2I; GTF3A; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HNRNPH1; HSPA1A; HSPD1; RBPJ; ILF3; JUN; LIMS1; MAGEA1; MARS; KMT2A; NFIA; NFE2L1; NFIB; NFIC; NOTCH2; NPAS2; NPM1; PHF2; PSMA6; PSMB7; PSMD1; PSMD4; RAN; UPF1; BRD2; RPL6; RPS26; RPS27A; SRSF4; SMARCA1; SMARCC2; SON; SOX4; SP100; SSRP1; SSX1; TAF7; TCEA1; ELOC; TDG; UBA52; SF1; ZKSCAN1; KAT6A; TAF15; HIST1H4C; HIST1H4L; KHSRP; EED; TSC22D1; TAX1BP1; ZRANB2; H2AFY; PRDX6; NCOR2; MORF4L2; BCLAF1; MAML1; THRAP3; DNAJB6; CTDSP2; NAMPT; ZMPSTE24; SYNCRIP; HEXIM1; CAMKK2; RAI1; SRSF10; GCN1; PKP3; CBX3; TCF25; POGZ; PHF8; RYBP; NUP62; VIRMA; GNL3; NUPR1; PABPC1; PDCD4; NOP53; HP1BP3; LEF1; CXXC5; SRRT; UIMC1; ARID4B; XRN1; CHD7; YEATS2; ZNF395; TCIM; BBX; MRTFB; CHD8; NUCKS1; AHNAK; SECISBP2; IRX3; TBL1XR1; ESRP2; ING5; ZNF587; ZNF664; CREB3L4; ZFP62
nuclear lumen	CC	0.3824	187/489	6.5530	ACTB; ADAR; ADD1; PARP1; AR; ATF4; BCL6; CAPG; CENPE; CTSB; DDX5; EZH2; XRCC6; MKNK2; MSH6; GTF2I; GTF3A; H1F0; HIST1H1C; HADH; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HNRNPH1; HSPA1A; RBPJ; ILF3; JUN; KPNB1; NBR1; MARS; MCM3; KMT2A; AFDN; HNRNPM; NFIA; NFIB; NFIC; NKTR; NOTCH2; NPAS2; NPM1; PHF2; PNN; PPP1CC; PSMA6; PSMB7; PSMD1; PSMD4; PTPN11; RAN; RARS; UPF1; BRD2; RNH1; RPL3; RPL5; RPL7; RPL12; RPL13; RPL23A; RPL26; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SEC13; SRSF4; SMARCA1; SMARCC2; SNRPD2; SON; SOX4; SP100; SSB; SSRP1; TAF7; TCEA1; ELOC; TDG; UBA52; UQCRC2; UTRN; SF1; KAT6A; FXR1; TAF15; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; KHSRP; USO1; EIF3A; EED; TRIP12; ZRANB2; H2AFY; RBM39; NCOR2; MORF4L2; BCLAF1; MAML1; ZC3H11A; THRAP3; DNAJB6; CTDSP2; NAMPT; SRRM1; ZNHIT1; SYNCRIP; NOP56; P3H4; HEXIM1; CAMKK2; RAI1; SRSF10; MORF4L1; PKP3; RPL35; ATXN2L; CBX3; ELL2; KLHDC10; SMG1; NPIPB3; POGZ; PHF8; FNBP4; RYBP; SPIDR; RPL13A; SRRM2; GTPBP4; BRD1; RPL36; DCAF13; PNISR; VIRMA; SPATS2L; GNL3; ANKRD11; REPIN1; NOP53; HP1BP3; LEF1; ARL6IP4; MAGEC2; IP6K2; CXXC5; PCF11; SRRT; NOP58; UIMC1; ARID4B; ARGLU1; PHF10; NOP10; CHD7; YEATS2; TCIM; BBX; THOC2; NUFIP2; CHD8; FAM111A; NUCKS1; DDX50; TSEN34; TBL1XR1; ESRP2; RUFY1; DUSP16; ING5; HIST1H2BK; HIST3H2A; CREB3L4

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
regulation of nucleobase-containing compound metabolic process	BP	0.3810	136/357	5.1587	ACTN1; PARP1; AR; ATF4; BCL6; CDKN1C; DDX5; DVL1; EZH2; XRCC6; GOLGB1; MSH6; GTF2I; GTF3A; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HNRNPH1; HSPA1A; HSPD1; RBPJ; ILF3; JUN; LIMS1; MAGEA1; MARS; KMT2A; COX2; NFIA; NFE2L1; NFIB; NFIC; NOTCH2; NPAS2; NPM1; PHF2; PSMA6; PSMB7; PSMD1; PSMD4; RAN; RANBP2; UPF1; BRD2; RPL6; RPS26; RPS27A; SEC13; SRSF4; SMARCA1; SMARCC2; SON; SOX4; SP100; SSRP1; SSX1; TAF7; TCEA1; ELOC; TDG; UBA52; SF1; ZKSCAN1; KAT6A; TAF15; HIST1H4C; HIST1H4L; KHSRP; EED; TSC22D1; TAX1BP1; TRIP12; ZRANB2; H2AFY; PRDX6; NCOR2; MORF4L2; BCLAF1; MAML1; THRAP3; DNAJB6; CTDSP2; NAMPT; ZMPSTE24; SYNCRIP; CCT2; HEXIM1; CAMKK2; RAI1; SRSF10; GCN1; PKP3; CBX3; TCF25; SMG1; POGZ; PHF8; RYBP; SPIDR; GTPBP4; NUP62; VIRMA; GNL3; NUPR1; PABPC1; PDCD4; NOP53; HP1BP3; LEF1; CXXC5; SRRT; UIMC1; ARID4B; XRN1; CHD7; YEATS2; ZNF395; TCIM; BBX; MRTFB; CHD8; NUCKS1; AHNAK; SECISBP2; IRX3; TBL1XR1; ESRP2; ING5; ZNF587; ZNF664; CREB3L4; ZFP62
regulation of transcription, DNA-templated	BP	0.3793	110/290	4.4358	PARP1; AR; ATF4; BCL6; CDKN1C; DDX5; DVL1; EZH2; XRCC6; GOLGB1; GTF2I; GTF3A; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HSPA1A; HSPD1; RBPJ; ILF3; JUN; LIMS1; MAGEA1; MARS; KMT2A; NFIA; NFE2L1; NFIB; NFIC; NOTCH2; NPAS2; NPM1; PHF2; PSMA6; PSMB7; PSMD1; PSMD4; RAN; BRD2; RPL6; RPS27A; SMARCA1; SMARCC2; SOX4; SP100; SSRP1; SSX1; TAF7; TCEA1; ELOC; TDG; UBA52; ZKSCAN1; KAT6A; TAF15; HIST1H4C; HIST1H4L; KHSRP; EED; TSC22D1; TAX1BP1; ZRANB2; H2AFY; NCOR2; MORF4L2; BCLAF1; MAML1; THRAP3; DNAJB6; CTDSP2; NAMPT; ZMPSTE24; HEXIM1; CAMKK2; RAI1; SRSF10; GCN1; CBX3; TCF25; POGZ; PHF8; RYBP; NUP62; GNL3; NUPR1; PDCD4; NOP53; HP1BP3; LEF1; CXXC5; SRRT; UIMC1; ARID4B; CHD7; YEATS2; ZNF395; TCIM; BBX; MRTFB; CHD8; NUCKS1; IRX3; TBL1XR1; ING5; ZNF587; ZNF664; CREB3L4; ZFP62
RNA metabolic process	BP	0.3774	180/477	6.1251	ACTN1; ADAR; PARP1; AR; ATF4; BCL6; BTF3; CDKN1C; DDX5; DVL1; EZH2; XRCC6; GOLGB1; GTF2I; GTF3A; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HNRNPH1; HSPA1A; HSPD1; IARS; RBPJ; ILF3; JUN; LIMS1; MAGEA1; MARS; KMT2A; HNRNPM; RPL10A; NFIA; NFE2L1; NFIB; NFIC; NOTCH2; NPAS2; NPM1; PHF2; PNN; PSMA6; PSMB7; PSMD1; PSMD4; RAN; RARS; UPF1; BRD2; RNH1; RPL3; RPL5; RPL6; RPL7; RPL12; RPL13; RPL17; RPL22; RPL23A; RPL26; RPL27; RPL29; RPL32; RPL37; RPL39; RPL36A; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SRSF4; SMARCA1; SMARCC2; SNRPD2; SON; SOX4; SP100; SSB; SSRP1; SSX1; TAF7; TCEA1; ELOC; TDG; UBA52; SF1; ZKSCAN1; KAT6A; TAF15; HIST1H4C; HIST1H4L; KHSRP; EED; TSC22D1; TAX1BP1; RPL14; ZRANB2; H2AFY; RBM39; PRDX6; NCOR2; MORF4L2; BCLAF1; MAML1; ZC3H11A; THRAP3; DNAJB6; FARSB; CTDSP2; NAMPT; SRRM1; ZMPSTE24; SYNCRIP; NOP56; HEXIM1; CAMKK2; RAI1; SRSF10; GCN1; PKP3; RPL35; CBX3; ELL2; TCF25; SMG1; POGZ; PHF8; RYBP; RPL13A; SRRM2; GTPBP4; NUP62; RPL36; DCAF13; VIRMA; GNL3; NUPR1; PABPC1; PDCD4; NOP53; HP1BP3; LEF1; ARL6IP4; CXXC5; PCF11; SRRT; NOP58; UIMC1; ARID4B; XRN1; NOP10; CHD7; YEATS2; ZNF395; TCIM; BBX; THOC2; MRTFB; CHD8; NUCKS1; AHNAK; TSEN34; SECISBP2; IRX3; TBL1XR1; ESRP2; ING5; ZNF587; ZNF664; CREB3L4; ZFP62
regulation of nucleic acid-templated transcription	BP	0.3758	112/298	4.3618	ACTN1; PARP1; AR; ATF4; BCL6; CDKN1C; DDX5; DVL1; EZH2; XRCC6; GOLGB1; GTF2I; GTF3A; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HSPA1A; HSPD1; RBPJ; ILF3; JUN; LIMS1; MAGEA1; MARS; KMT2A; NFIA; NFE2L1; NFIB; NFIC; NOTCH2; NPAS2; NPM1; PHF2; PSMA6; PSMB7; PSMD1; PSMD4; RAN; BRD2; RPL6; RPS27A; SMARCA1; SMARCC2; SOX4; SP100; SSRP1; SSX1; TAF7; TCEA1; ELOC; TDG; UBA52; SF1; ZKSCAN1; KAT6A; TAF15; HIST1H4C; HIST1H4L; KHSRP; EED; TSC22D1; TAX1BP1; ZRANB2; H2AFY; NCOR2; MORF4L2; BCLAF1; MAML1; THRAP3; DNAJB6; CTDSP2; NAMPT; ZMPSTE24; HEXIM1; CAMKK2; RAI1; SRSF10; GCN1; CBX3; TCF25; POGZ; PHF8; RYBP; NUP62; GNL3; NUPR1; PDCD4; NOP53; HP1BP3; LEF1; CXXC5; SRRT; UIMC1; ARID4B; CHD7; YEATS2; ZNF395; TCIM; BBX; MRTFB; CHD8; NUCKS1; IRX3; TBL1XR1; ING5; ZNF587; ZNF664; CREB3L4; ZFP62

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
regulation of RNA biosynthetic process	BP	0.3758	112/298	4.3618	ACTN1; PARP1; AR; ATF4; BCL6; CDKN1C; DDX5; DVL1; EZH2; XRCC6; GOLGB1; GTF2I; GTF3A; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HMGN1; HSPA1A; HSPD1; RBPJ; ILF3; JUN; LIMS1; MAGEA1; MARS; KMT2A; NFIA; NFE2L1; NFIB; NFIC; NOTCH2; NPAS2; NPM1; PHF2; PSMA6; PSMB7; PSMD1; PSMD4; RAN; BRD2; RPL6; RPS27A; SMARCA1; SMARCC2; SOX4; SP100; SSRP1; SSX1; TAF7; TCEA1; ELOC; TDG; UBA52; SF1; ZKSCAN1; KAT6A; TAF15; HIST1H4C; HIST1H4L; KHSRP; EED; TSC22D1; TAX1BP1; ZRANB2; H2AFY; NCOR2; MORF4L2; BCLAF1; MAML1; THRAP3; DNAJB6; CTDSP2; NAMPT; ZMPSTE24; HEXIM1; CAMKK2; RAI1; SRSF10; GCN1; CBX3; TCF25; POGZ; PHF8; RYBP; NUP62; GNL3; NUPR1; PDCD4; NOP53; HP1BP3; LEF1; CXXC5; SRRT; UIMC1; ARID4B; CHD7; YEATS2; ZNF395; TCIM; BBX; MRTFB; CHD8; NUCKS1; IRX3; TBL1XR1; ING5; ZNF587; ZNF664; CREB3L4; ZFP62
non-membrane-bounded organelle	CC	0.3721	176/473	5.7710	ACTB; ACTG1; ACTN1; ADAR; ADD1; PARP1; ANK3; AR; ATF4; BCL6; CALD1; CAPG; CENPE; CTNNA1; CTSB; DDX5; CYB5R3; DSP; DVL1; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HIST1H2AD; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HSPA1A; RBPJ; ILF3; JUN; KIF5C; KPNB1; CAPRIN1; NBR1; MARCKS; MARS; MCM3; MYH9; MYH11; MYL6; HNRNPM; SEPTIN2; RPL10A; NFIB; NFIC; NPM1; PHF2; PNN; PPP1CC; PPP2CB; PSMA6; RAN; RAP1B; UPF1; RPL3; RPL5; RPL6; RPL7; RPL12; RPL13; RPL17; RPL22; RPL23A; RPL26; RPL27; RPL29; RPL32; RPL37; RPL39; RPL36A; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SEC13; SMARCA1; SMARCC2; SP100; SSB; SSRP1; TCEA1; UBA52; UTRN; VDAC1; SF1; KAT6A; CCDC6; FXR1; HIST3H3; HIST1H2AK; HIST1H2AM; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; KHSRP; USO1; EIF3A; EED; HIST1H2AG; RPL14; H2AFY; RBM39; NCOR2; MORF4L2; MFN2; DNAJB6; NDRG1; NOP56; CCT2; P3H4; MORF4L1; GCN1; DSTN; AKAP13; RPL35; ATXN2L; CBX3; DAAM1; POGZ; PHF8; EXOC7; SPIDR; RPL13A; GTPBP4; NUP62; RPL36; DCAF13; SPATS2L; LSM14A; GNL3; PABPC1; REPIN1; NOP53; HP1BP3; ARL6IP4; MAGEC2; IP6K2; NOP58; XRN1; MRPL20; PHF10; NOP10; CHD7; YEATS2; H2AFJ; TCIM; THOC2; NUFIP2; SHROOM3; RHOU; FAM111A; NUCKS1; DDX50; AHNAK; TSEN34; TBL1XR1; TUBA1C; HIST1H2BK; CEP95; HIST3H2A; H2AFV; TOP1MT; KLC3; HIST2H2AA4
intracellular non-membrane-bounded organelle	CC	0.3721	176/473	5.7710	ACTB; ACTG1; ACTN1; ADAR; ADD1; PARP1; ANK3; AR; ATF4; BCL6; CALD1; CAPG; CENPE; CTNNA1; CTSB; DDX5; CYB5R3; DSP; DVL1; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HIST1H2AD; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HSPA1A; RBPJ; ILF3; JUN; KIF5C; KPNB1; CAPRIN1; NBR1; MARCKS; MARS; MCM3; MYH9; MYH11; MYL6; HNRNPM; SEPTIN2; RPL10A; NFIB; NFIC; NPM1; PHF2; PNN; PPP1CC; PPP2CB; PSMA6; RAN; RAP1B; UPF1; RPL3; RPL5; RPL6; RPL7; RPL12; RPL13; RPL17; RPL22; RPL23A; RPL26; RPL27; RPL29; RPL32; RPL37; RPL39; RPL36A; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SEC13; SMARCA1; SMARCC2; SP100; SSB; SSRP1; TCEA1; UBA52; UTRN; VDAC1; SF1; KAT6A; CCDC6; FXR1; HIST3H3; HIST1H2AK; HIST1H2AM; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; KHSRP; USO1; EIF3A; EED; HIST1H2AG; RPL14; H2AFY; RBM39; NCOR2; MORF4L2; MFN2; DNAJB6; NDRG1; NOP56; CCT2; P3H4; MORF4L1; GCN1; DSTN; AKAP13; RPL35; ATXN2L; CBX3; DAAM1; POGZ; PHF8; EXOC7; SPIDR; RPL13A; GTPBP4; NUP62; RPL36; DCAF13; SPATS2L; LSM14A; GNL3; PABPC1; REPIN1; NOP53; HP1BP3; ARL6IP4; MAGEC2; IP6K2; NOP58; XRN1; MRPL20; PHF10; NOP10; CHD7; YEATS2; H2AFJ; TCIM; THOC2; NUFIP2; SHROOM3; RHOU; FAM111A; NUCKS1; DDX50; AHNAK; TSEN34; TBL1XR1; TUBA1C; HIST1H2BK; CEP95; HIST3H2A; H2AFV; TOP1MT; KLC3; HIST2H2AA4

(continued)

Term	Ontology	set.mean	set.size	z.score	in.genes
nucleoplasm	CC	0.3698	159/430	5.2484	ACTB; ADAR; ADD1; PARP1; AR; ATF4; BCL6; CAPG; CENPE; DDX5; EZH2; XRCC6; MKNK2; MSH6; GTF2I; GTF3A; H1F0; HADH; HDAC1; HMGB1; HMGB2; HMGN1; HNRNPC; HNRNPH1; HSPA1A; RBPJ; ILF3; JUN; KPNB1; NBR1; MCM3; KMT2A; AFDN; HNRNPM; NFIA; NKTR; NOTCH2; NPAS2; NPM1; PHF2; PNN; PPP1CC; PSMA6; PSMB7; PSMD1; PSMD4; PTPN11; RAN; RARS; UPF1; BRD2; RNH1; RPL5; RPL26; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SEC13; SRSF4; SMARCA1; SMARCC2; SNRPD2; SON; SOX4; SP100; SSRP1; TAF7; TCEA1; ELOC; TDG; UBA52; UQCRC2; UTRN; SF1; KAT6A; TAF15; HIST3H3; HIST1H2BL; HIST1H2BF; HIST1H2BH; HIST1H4C; HIST1H4L; KHSRP; EED; TRIP12; ZRANB2; RBM39; NCOR2; MORF4L2; BCLAF1; MAML1; ZC3H11A; THRAP3; DNAJB6; CTDSP2; NAMPT; SRRM1; ZNHIT1; SYNCRIP; NOP56; HEXIM1; CAMKK2; RAI1; SRSF10; MORF4L1; PKP3; ATXN2L; CBX3; ELL2; KLHDC10; SMG1; NPIPB3; POGZ; PHF8; FNBP4; RYBP; SPIDR; SRRM2; BRD1; DCAF13; PNISR; VIRMA; GNL3; ANKRD11; NOP53; HP1BP3; LEF1; ARL6IP4; IP6K2; CXXC5; PCF11; SRRT; NOP58; UIMC1; ARID4B; ARGLU1; NOP10; CHD7; YEATS2; TCIM; BBX; THOC2; NUFIP2; CHD8; TSEN34; TBL1XR1; ESRP2; RUFY1; DUSP16; ING5; HIST1H2BK; CREB3L4
negative regulation of macromolecule metabolic process	BP	0.3696	129/349	4.5345	A2M; ADAR; PARP1; AR; ATF4; BCL6; CAST; CDKN1C; DDX5; DVL1; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HNRNPC; HSPA1A; IGF1R; RBPJ; ILF3; JUN; LDLR; LIMS1; CAPRIN1; MAGEA1; KMT2A; GADD45B; RPL10A; NFIB; NFIC; NOTCH2; NPM1; PHF2; PSMA6; PSMB7; PSMD1; PSMD4; RAN; RANBP2; UPF1; RNH1; RPL3; RPL5; RPL6; RPL7; RPL12; RPL13; RPL17; RPL22; RPL23A; RPL26; RPL27; RPL29; RPL32; RPL37; RPL39; RPL36A; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SEC13; SRSF4; SMARCC2; SOX4; SP100; SSB; SSX1; TAF7; TDG; UBA52; YWHAG; SF1; KAT6A; FXR1; HIST1H4C; HIST1H4L; KHSRP; EIF3A; EED; RPL14; TRIP12; H2AFY; NCOR2; HERPUD1; BCLAF1; THRAP3; DNAJB6; CTDSP2; ZMPSTE24; SYNCRIP; HEXIM1; SRSF10; PKP3; RPL35; CBX3; TCF25; SMG1; PHF8; ARL6IP1; RYBP; RPL13A; GTPBP4; NUP62; RPL36; PABPC1; PDCD4; NOP53; LEF1; CXXC5; SRRT; UIMC1; XRN1; YEATS2; CHD8; WNK1; SECISBP2; TBL1XR1; DUSP16; NIBAN1
negative regulation of metabolic process	BP	0.3656	136/372	4.5306	A2M; ADAR; PARP1; AR; ATF4; ATP1A1; BCL6; CAST; CDKN1C; DDX5; DVL1; EZH2; XRCC6; MSH6; H1F0; HIST1H1C; HDAC1; HMGB1; HMGB2; HNRNPC; HSPA1A; IGF1R; RBPJ; ILF3; JUN; LDLR; LIMS1; CAPRIN1; MAGEA1; KMT2A; GADD45B; RPL10A; NFIB; NFIC; NOTCH2; NPM1; PHF2; PSMA6; PSMB7; PSMD1; PSMD4; RAN; RANBP2; UPF1; RNH1; RPL3; RPL5; RPL6; RPL7; RPL12; RPL13; RPL17; RPL22; RPL23A; RPL26; RPL27; RPL29; RPL32; RPL37; RPL39; RPL36A; RPS3A; RPS6; RPS7; RPS8; RPS16; RPS18; RPS24; RPS25; RPS26; RPS27A; SEC13; SRSF4; SMARCC2; SOX4; SP100; SSB; SSX1; TAF7; TDG; UBA52; VDAC1; YWHAG; SF1; KAT6A; FXR1; HIST1H4C; HIST1H4L; KHSRP; EIF3A; EED; RPL14; TRIP12; H2AFY; NCOR2; HERPUD1; BCLAF1; THRAP3; DNAJB6; CTDSP2; NAMPT; ZMPSTE24; SYNCRIP; HEXIM1; RRAGA; SRSF10; PKP3; RPL35; CBX3; TCF25; SMG1; PHF8; ARL6IP1; RYBP; RPL13A; GTPBP4; NUP62; RPL36; NUPR1; PABPC1; PDCD4; NOP53; LEF1; CXXC5; SRRT; UIMC1; XRN1; YEATS2; ARFGEF3; CHD8; WNK1; SECISBP2; TBL1XR1; DUSP16; NIBAN1; NRBP2

We could compare these enriched gene-sets with the results in Figures 4 and 5 of the main manuscript.

5.3 Comparing Healthy Subjects vs Cancer Patients

Here, we repeat the linear contrast method on these 698 peptides to determine if fluorescence levels of these peptides are significantly higher among cancer patients than among healthy subjects.

There are 26 peptides that are associated with higher antibody response in cancer patients than in normal subjects. We tabulate these peptides with their differences (of average $\log_2$ fluorescence levels between cancer and healthy subjects) and p-values of their t-statistics.

(continued)

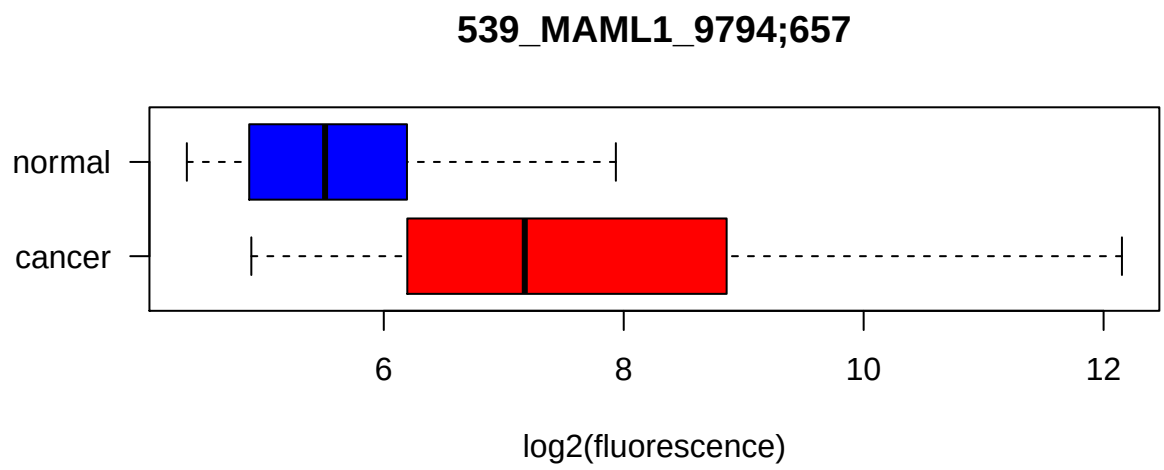
peptide_id	difference	tstat_pvalues

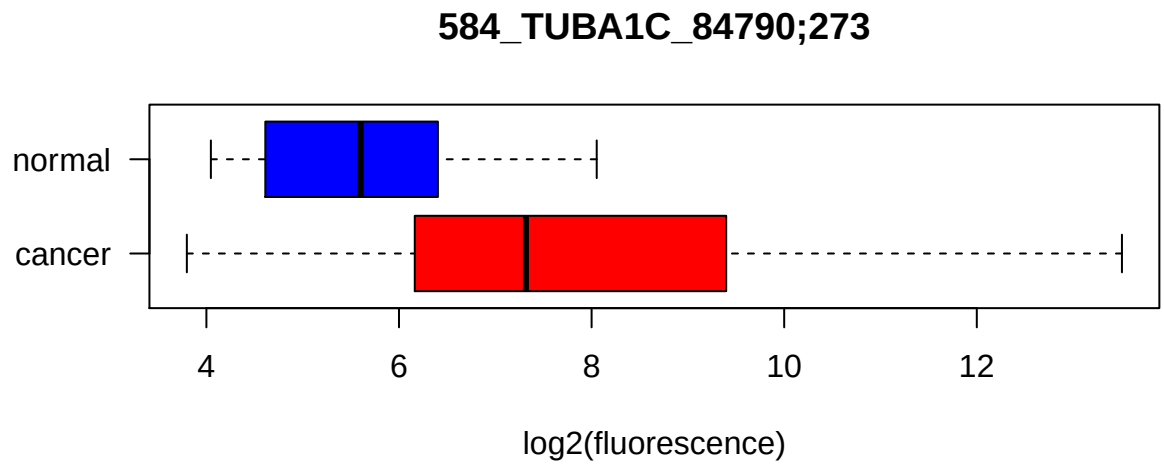
peptide_id	difference	tstat_pvalues
539_MAML1_9794;657	1.8619	0.0071
584_TUBA1C_84790;273	2.0708	0.0108
1098_ARGLU1_55082;197	1.7184	0.0140
1046_SRRT_51593;9	1.4562	0.0159
655_CCT2_10576;517	1.9199	0.0207
65_ZKSCAN1_7586;473	1.4761	0.0207
608_CCDC6_8030;301	1.1939	0.0248
1133_SMARCA1_6594;109	1.0842	0.0260
974_EIF3L_51386;349	0.7739	0.0284
1196_ZFP62_643836;381	1.3175	0.0290
267_ACTB_60;113	1.2223	0.0327
431_ACTG1_71;113	1.2223	0.0327
491_KDELR2_11014;145	0.1176	0.0355
591_ECHDC2_55268;1	0.1945	0.0359
478_VDAC1_7416;37	1.2779	0.0377
611_KIAA1429_25962;1717	0.7533	0.0388
PRO14;2497	1.0935	0.0394
743_AK4_205;41	1.4307	0.0394
141_CAMKK2_10645;21	1.1280	0.0411
CTA1;69	0.7611	0.0418
1004_TRIP12_9320;1145	0.6571	0.0441
150_YWHAG_7532;1	1.1831	0.0443
1406_CCDC85C_317762;109	1.2628	0.0445
633_USP54_159195;1001	1.0586	0.0457
1364_DHCR7_1717;337	0.1251	0.0464
985_CALD1_800;549	1.1548	0.0486

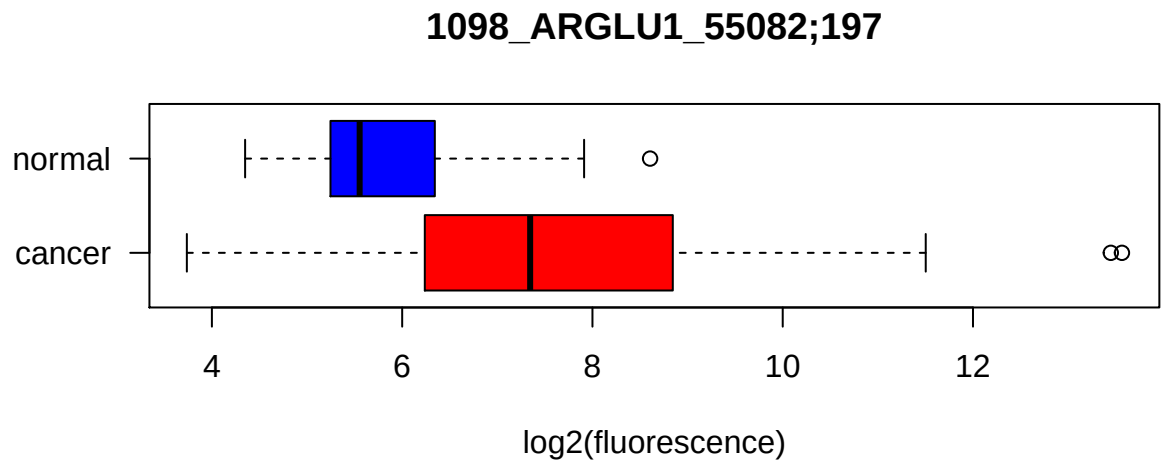
Another form of verification using the calls data would be as follows: we check the sum of positive calls for each cancer stage for these 26 peptides.

stage	n	total_calls
normal	15	1
new_dx	15	6
nmCSPC	35	16
nmCRPC	14	2
mCRPC	15	4

As expected, almost all of the positive calls for these peptides come from cancer patients instead of normal subjects. For purpose of visualization, we also produce the boxplots of $\log_2$ fluorescence levels of subjects from these two groups for some of these peptides.


539_MAML1_9794;657
normal
cancer
6
8
10
12
log2(fluorescence)


584_TUBA1C_84790;273
normal
cancer
4
6
8
10
12
log2(fluorescence)


1098_ARGLU1_55082;197
normal
cancer
4
6
8
10
12
log2(fluorescence)

6 Concluding Remarks

- Calls are conservative – most of the calls are zero anyway. No cancer-stage signals are observable from the calls data.
- There appears to be cancer stage effects present in the $\log_2$ fluorescence data with or without filtering peptides based on calls. The clustering patterns of data points based on cancer stages seem to be slightly more obvious without calls-filtering.