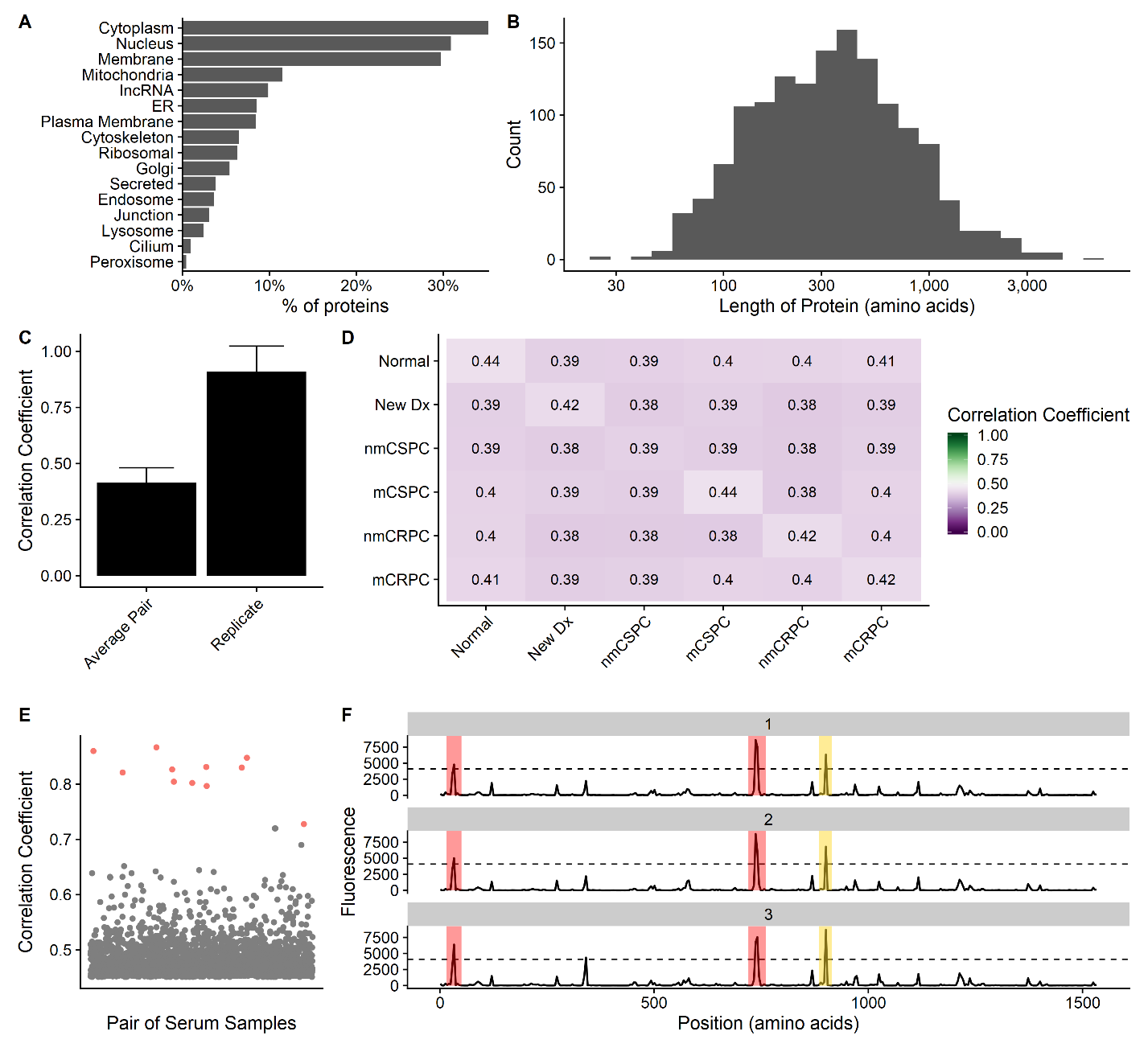
To characterize antibody responses to a wide variety of proteins in prostate cancer patients, we designed a peptide microarray to incorporate a large diversity of prostate-cancer associated probes. Our targets included peptides corresponding to 1381 of the most abundantly expressed gene products in metastatic prostate cancer [citation] in addition to 62 proteins identified in previous studies examining serum antibody responses in prostate cancer patients1. We also included a set of 158 potential open reading frames (ORFs) from long non-coding RNAs (lncRNAs) that have been shown to be highly expressed in prostate cancer[citation].

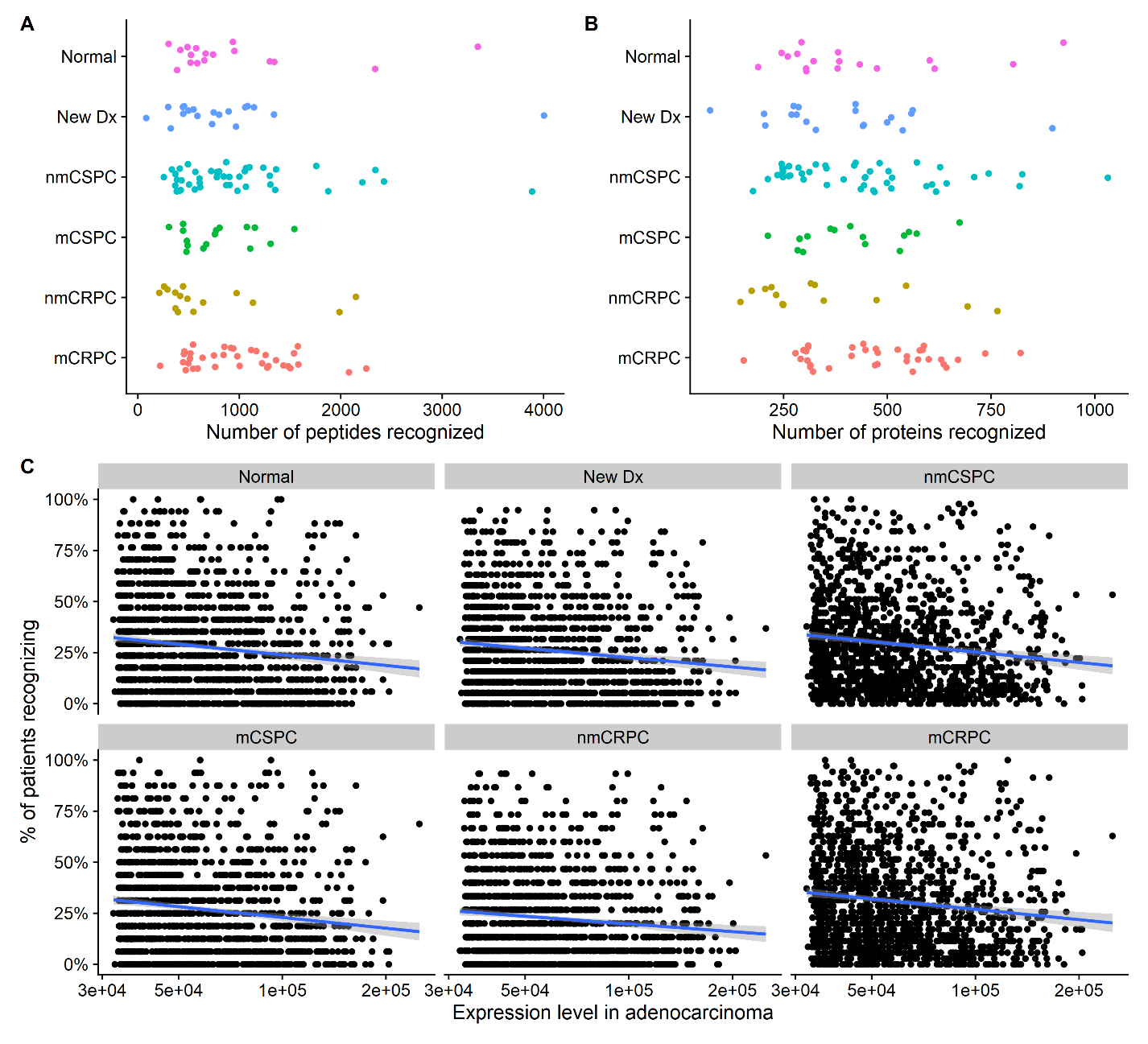
16-mer peptides spanning the amino acid sequences of these 1601 gene products, and overlapping by 12 amino acids, were used to generate a microarray comprising 177,604 peptides. The complete list of probes and corresponding proteins is available in Supplementary Table 1. The manufacture of the array and synthesis of peptides was performed as previously described2. The characteristics of the proteins included in the array are summarized in Figure 1, using data retrieved from UniProt [citation]. The majority of proteins were localized within the cytoplasm or nucleus, or traffick between the two compartments (Fig 1A). Approximately 16% of the proteins were localized to the mitochondria or ribosomes. The median protein length was 332 amino acids (Fig 1B).

We obtained serum samples from 17 normal male blood donors (Normal), 19 patients with newly diagnosed prostate cancer (New Dx), 45 patients with castration-sensitive non-metastatic prostate cancer (nmCSPC), 16 patients with castration-sensitive metastatic prostate cancer (mCSPC), 15 patients with castration-resistant non-metastatic prostate cancer (nmCRPC), and 35 patients with castration-resistant metastatic disease (mCRPC). Each patient’s serum was assayed in triplicate for peptide-specific IgG responses using the microarray. To assess the reproducibility of the assay, we calculated Pearson (?) correlation coefficents between each technical replicate and found high correlation between replicates (Fig 1C). To determine the degree of variability between serum samples, we calculated the mean correlation coefficient across all pairs of distinct serum samples (Fig 1C). To determine if the results of the assay were in accordance with published data on antibody responses in prostate cancer, we looked at PSA, PAP, AR, and HER2/Neu. 11.4% of mCRPC patients assayed on the array displayed antibody responses against PSA, while 5.9% of normals had PSA responses, in agreement with the literature.3 13.9% of prostate cancer patients and 0% of normals had responses to PAP, which is slightly higher than previously reported. We hypothesized that there would be greater similarity between antibody responses in two patients with the same clinical state of disease compared to two patients with different states of disease. When we calculated mean correlation coefficients between patients in each combination of clinical states, we saw no significant increase in correlation coefficients among patients within the same clinical state (Fig 1D). Included in this analysis were 11 patients who had serum collected at different time points: when they had an early stage of disease and again when they had a later stage of disease. Notably, these serum samples from the same patients had especially high correlation coefficients (Fig 1E). This indicates that each patient has a unique antibody signature that is relatively stable over time. In contrast, patients exhibit highly heterogenous patterns of antibody responses when compared to other patients with similar levels of disease burden.



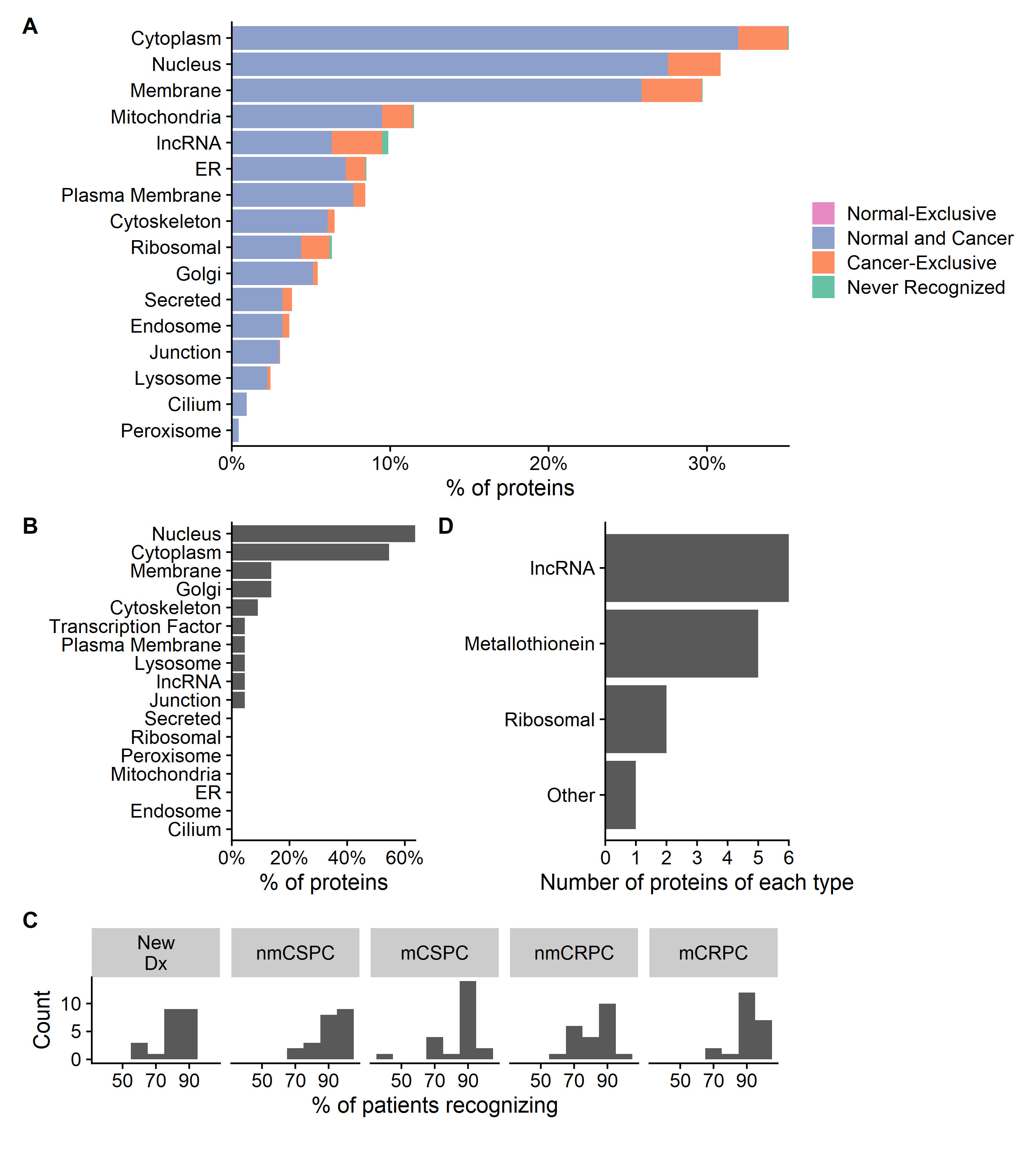
**Figure 1.** A prostate cancer-specific microarray is able to reproducibly measure antibody signatures from serum of healthy individuals and prostate cancer patients. **a)** Summary of the subcellular localization and **b)** length in amino acids of all 1601 unique proteins on the array according to UniProt. **c)** The mean correlation coefficient between all pairs of different individuals (Average Pair) compared to the average correlation coefficient between all technical replicates (Replicate). Error bars represent standard deviation. **d)** Heatmap of the correlation coefficient between individuals and members of different stages of disease. **e)** Each dot represents the correlation coefficient between antibody responses in two different serum samples. Dots marked in red are instances when the same individual had serum collected at two different stages of disease. **f)** Example microarray data for technical replicates of a single protein (ADT14) with the 212 signal threshold indicated by the dashed line. Positive calls are marked in red. In yellow is a negative call that did not meet the sliding window criterion.

We hypothesized that patients with higher disease burden would recognize more peptides because of increased presentation of cancer-associated epitopes [citation]. To assess this, we needed to define a positive antibody response. We considered probes with fluorescence intensity of at least 212 and a sliding window p value less than 0.054 in at least 2 of the 3 technical replicates to be positive. Two examples of positive calls are shown in Fig 1F. We found no correlation between stage of disease and the number of probes recognized at either the peptide level or the protein level (Fig 2A, 2B). We also expected that proteins that were expressed at higher levels in metastatic prostate cancer would be recognized at greater levels in patients with metastatic prostate cancer and that this correlation would not be present in normal controls. We observed no such correlation in any stage of disease (Fig 1C). There is a substantial amount of heterogeneity in antibody responses among patients, which appears to dominate over any potential trends at the stage level. For instance, the number of proteins recognized by normal controls ranged from 189 to 924.



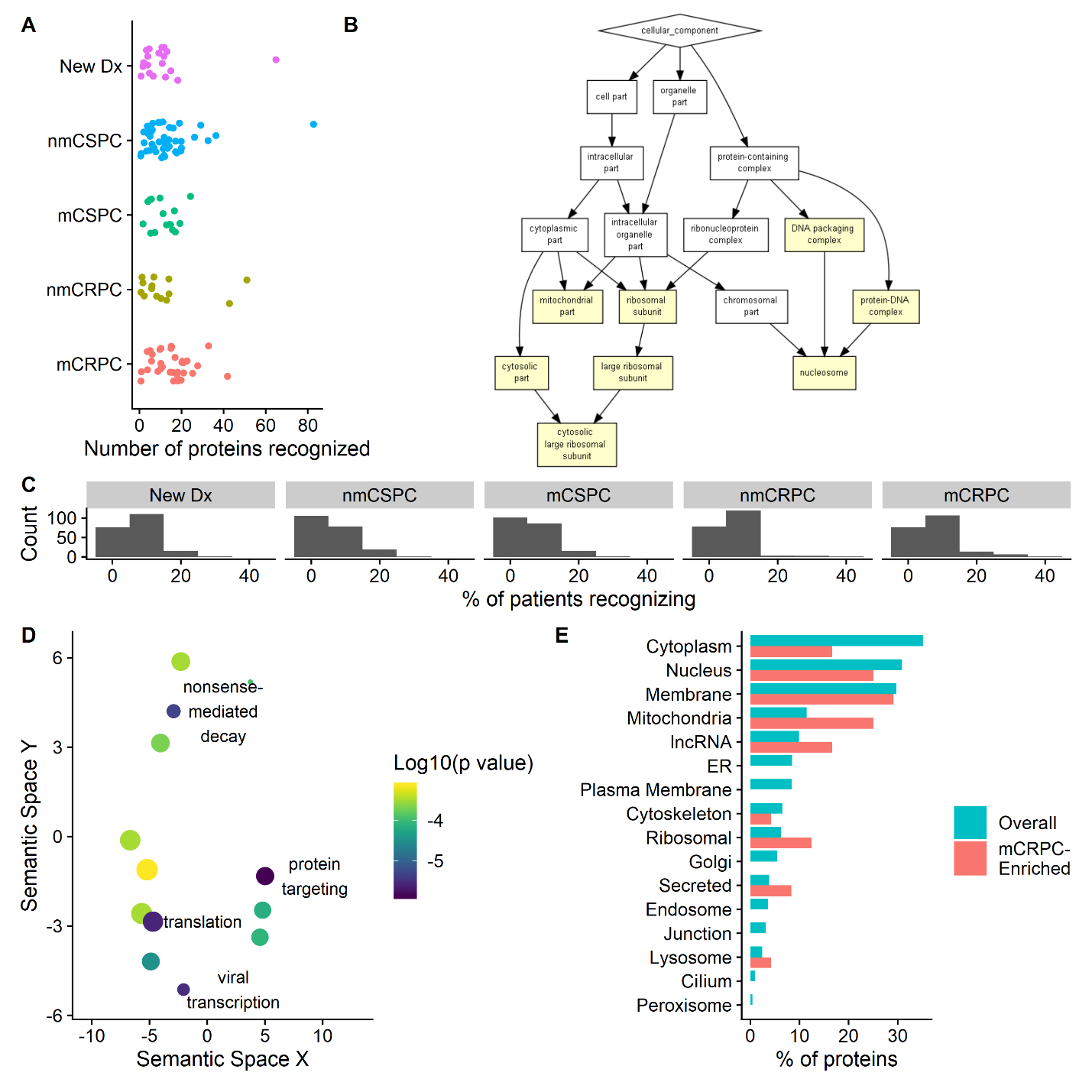
**Figure 2**. Frequency of protein recognition does not correlate with stage of disease or expression level of each protein in prostate cancer. The number of **a)** peptides and **b)** proteins recognized by each patient, categorized by stage of disease. **c)** Scatterplot of the expression level of each protein as measured by RNA-seq compared to the percent of patients recognizing the protein, categorized by stage of disease. The blue line represents a linear best fit.

Having established that there was a large diversity in antibody responses among our patients, we examined whether there were any broad trends in the types of proteins that were recognized. Nearly all proteins were recognized by at least patient. Conversely, there were no proteins that were recognized by all patients. Most proteins were recognized by both normals and cancer patients (Fig 3A). Contrary to our expectations, the majority of lncRNAs were recognized by at least one patient. We identified 22 proteins that were recognized by at least 90% of normal controls (Normal Proteins, Supplementary Table 2), hypothesizing that most of these heavily recognized proteins would be cytoplasmic or associated with the plasma membrane. We found that many of these Normal Proteins were actually associated with the nucleus, but found no specific features that were enriched via gene ontology analysis (Fig 3B). These Normal Proteins were also recognized at high levels by patients with each stage of cancer (Fig 3C). As expected, the majority of the proteins that were never recognized were ORFs from lncRNAs, as these may not be translated into gene products. (Fig 3D).

****

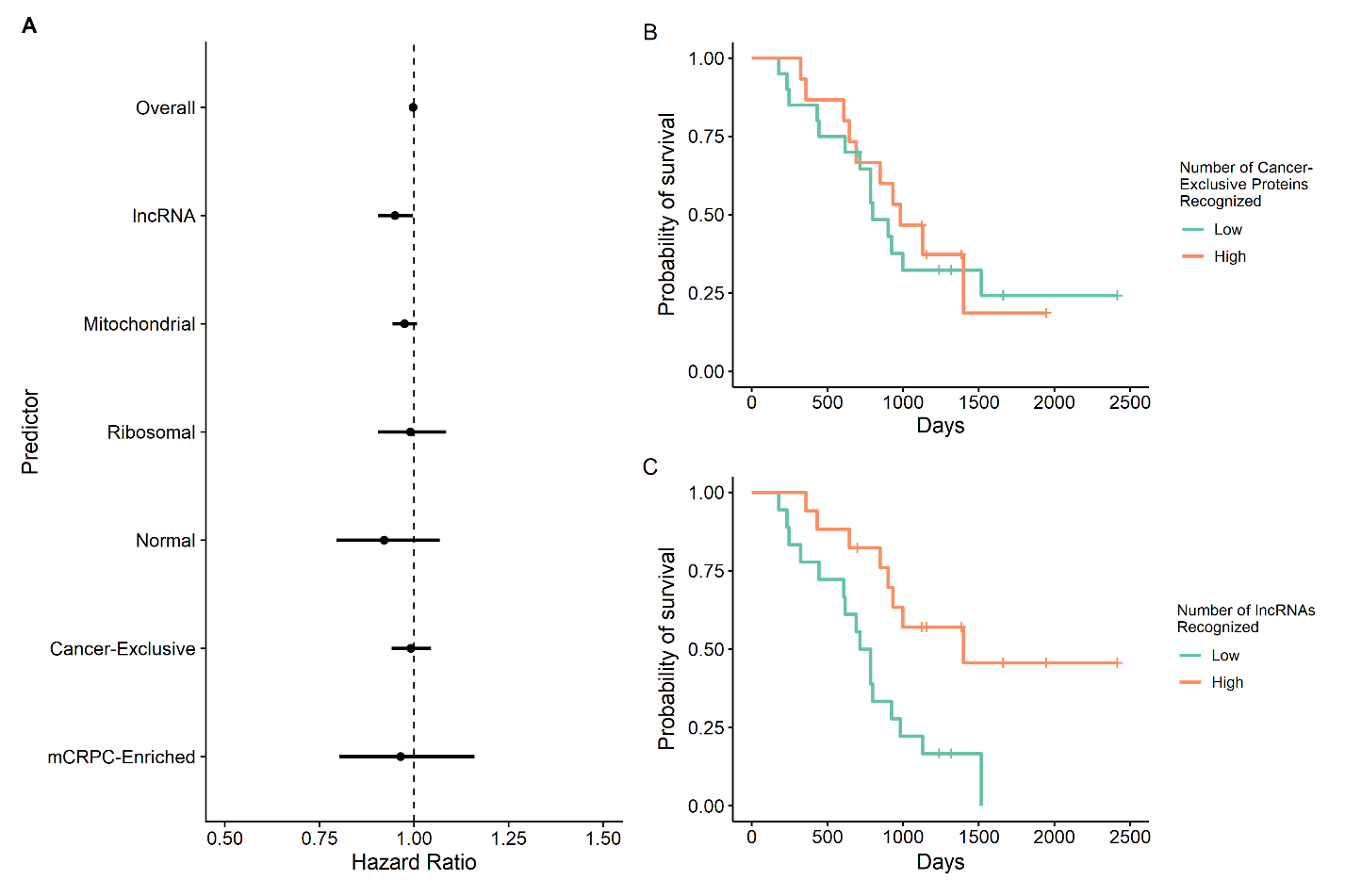
**Figure 3.** Nearly all proteins are recognized by at least one patient and proteins that are recognized by most normal patients are also recognized in all stages of cancer. **a)** Percentage of proteins that are recognized by only normal patients (*Normal-Exclusive*), at least one normal patient and one cancer patient (*Normal and Cancer*), percentage of proteins not recognized by any normal patients but recognized by at least one cancer patient (*Cancer-Exclusive*), and percentage not recognized at all (*Not Recognized*), categorized by subcellular localization. **b)** Characteristics of proteins that are recognized by at least 90% of normal patients. **c)** Histogram indicating the frequency with which proteins from **b** are recognized by patients with each stage of cancer. **d)** Characteristics of proteins that are never recognized.

We found that there was only one protein (RPS28) that was recognized by a normal patient but not by any cancer patients. However, there was a group of 205 proteins that were recognized exclusively in cancer patients (cancer-exclusive) as described in Fig 3A and Supplementary Table 3. We hypothesized that they would be more highly recognized in patients with later stage disease, but did not observe this correlation (Fig 4A). These cancer-exclusive proteins were significantly enriched for ribosomal and mitochondrial proteins as determined by gene ontology (GO) analysis (Fig 4B). Many of these proteins were associated with regulation of translation, nonsense-mediated decay, and protein targeting (Fig 4D). These cancer-exclusive proteins were recognized at varying levels across different stages of disease, with some recognized by up to 40% of patients in a given stage (Fig 4C). We also examined proteins that were either exclusively expressed in patients with castration resistant metastatic disease or were recognized at least two-fold more often in these patients than in patients with other stages of disease (mCRPC-enriched proteins). We identified 24 of these proteins (Supplementary Table 4). Nearly half of these proteins were mitochondrial, ribosomal, or lncRNAs (Fig 4E). No gene ontology terms were significantly enriched in this group of proteins.



**Figure 4.** Patients with castration resistant metastatic disease recognize mitochondrial proteins. **a)** The number of cancer-exclusive proteins that are recognized by patients in each stage. **b)** GO cellular component results using cancer-exclusive proteins as the target set and the full list of proteins on the array as the background set. **c)** Histogram indicating the frequency with which proteins from **a** are recognized by patients with each stage of cancer. **d)** Scatterplot of most significant GO Process terms for the cancer-exclusive proteins. More similar GO terms are grouped together. Color indicates the log10 of the p value. The size of the bubble is based on the frequency of the GO term in the UniProt Database. **e)** Characteristics of proteins recognized at higher rates by patients with castration resistant metastatic disease (*mCRPC-Enriched*) compared to characteristics of all proteins on the array (*Overall*).

There were data available on the overall survival of the patients with castration resistant metastatic disease, with a median follow-up of 901 days. Given mounting evidence that lncRNAs play an important role in prostate cancer [citation/more specific statement] and the unexpectedly high level of antibody responses against them observed in this study, we examined whether lncRNA recognition was associated with survival. We observed a correlation between recognition of a greater number of lncRNAs proteins and increased survival (p = 0.037, Fig 5A and 5C). There was no association between overall number of proteins recognized, number of cancer-exclusive proteins recognized, or number of mitochondrial proteins and survival (Fig 5A and B).



**Figure 5.**The number of lncRNAs recognized correlates with survival. **a)** Forest plot of Cox regressions using either the overall number of proteins recognized by each patient or the number of proteins from the indicated category recognized by each patient as the predictor of survival. Kaplan-Meier curve using number of **b)** cancer-exclusive or **c)** lncRNAs recognized as the predictor. Patients are divided into *Low* and *High* categories based on whether they recognized more or fewer cancer-exclusive or lncRNAs than the median.

1. Ha, S. *et al.* IgG responses to tissue-associated antigens as biomarkers of immunological treatment efficacy. *J Biomed Biotechnol* **2011**, 454861–454861 (2010).

2. Yan, Y. *et al.* Whole Genome-Derived Tiled Peptide Arrays Detect Prediagnostic Autoantibody Signatures in Non-Small-Cell Lung Cancer. *Cancer Res.* **79**, 1549–1557 (2019).

3. McNEEL DOUGLAS G. *et al.* Antibody immunity to prostate cancer associated antigens can be detected in the serum of patients with prostate cancer. *Journal of Urology* **164**, 1825–1829 (2000).

4. Lo, K. C. *et al.* Comprehensive Profiling of the Rheumatoid Arthritis Antibody Repertoire. *Arthritis & Rheumatology* **72**, 242–250 (2020).

Things for discussion:

All proteins are fairly highly expressed in both benign, metastatic adenocarcinoma, and neuroendocrine cancer because we chose the most highly expressed prostate cancer associated proteins

There are many more cancer proteins than normal proteins

How do we interpret the fact that supposedly non-coding RNAs have protein products that are recognized?

Questions for statistician:

Is averaging correlation coefficients using fisher’s Z valid? Is there a different way of summarizing parts of the correlation matrix?

Is correlation the best metric of reproducibility?

Did I do the Cox Regression correctly? Do we need to adjust for multiple testing?

Is there a way to statistically test if a cancer-exclusive protein appears more often than expected?

Is there a way to statistically test if there are more mitochondrial proteins in the mCRPC enriched proteins than overall?