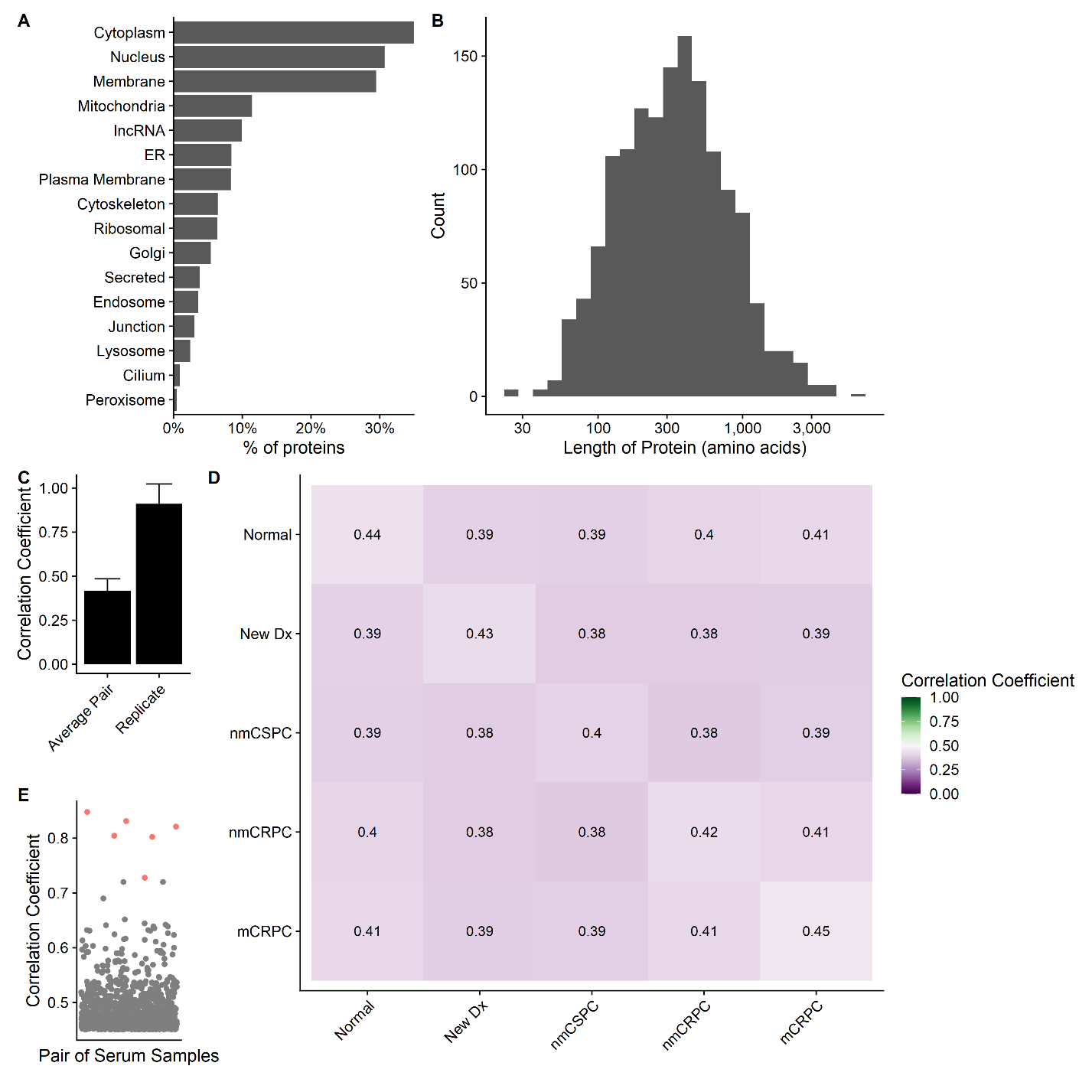
To characterize antibody responses to a wide variety of proteins in prostate cancer patients, we designed a peptide microarray able to be screened with patient sera. This array included peptides spanning the amino acid sequences of 1451 of the most abundantly expressed gene products in metastatic prostate cancer [citation], including 125 proteins identified in previous studies examining serum antibody responses in prostate cancer patients1,2. We also included a set of peptides spanning the predicted amino acid sequences of 160 potential open reading frames (ORFs) from long non-coding RNAs (lncRNAs) that have been shown to be highly expressed in prostate cancer[citation]. We included these given their strong association with prostate cancer, and while most would likely serve as negative controls as they would not be expected to encode gene products, other groups have shown that some lncRNAs may be translated into unstable peptides or even functional proteins especially with the dysregulation induced by cancer 3–5

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. 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 described6. The characteristics of the proteins included in the array are summarized in Figure 1, using data retrieved from UniProt7. The majority of proteins were localized within the cytoplasm or nucleus, or traffic between the two compartments (Fig 1A). Approximately 16% of the proteins were localized to the mitochondria or ribosomes. The median protein length was 483 amino acids (Fig 1B).

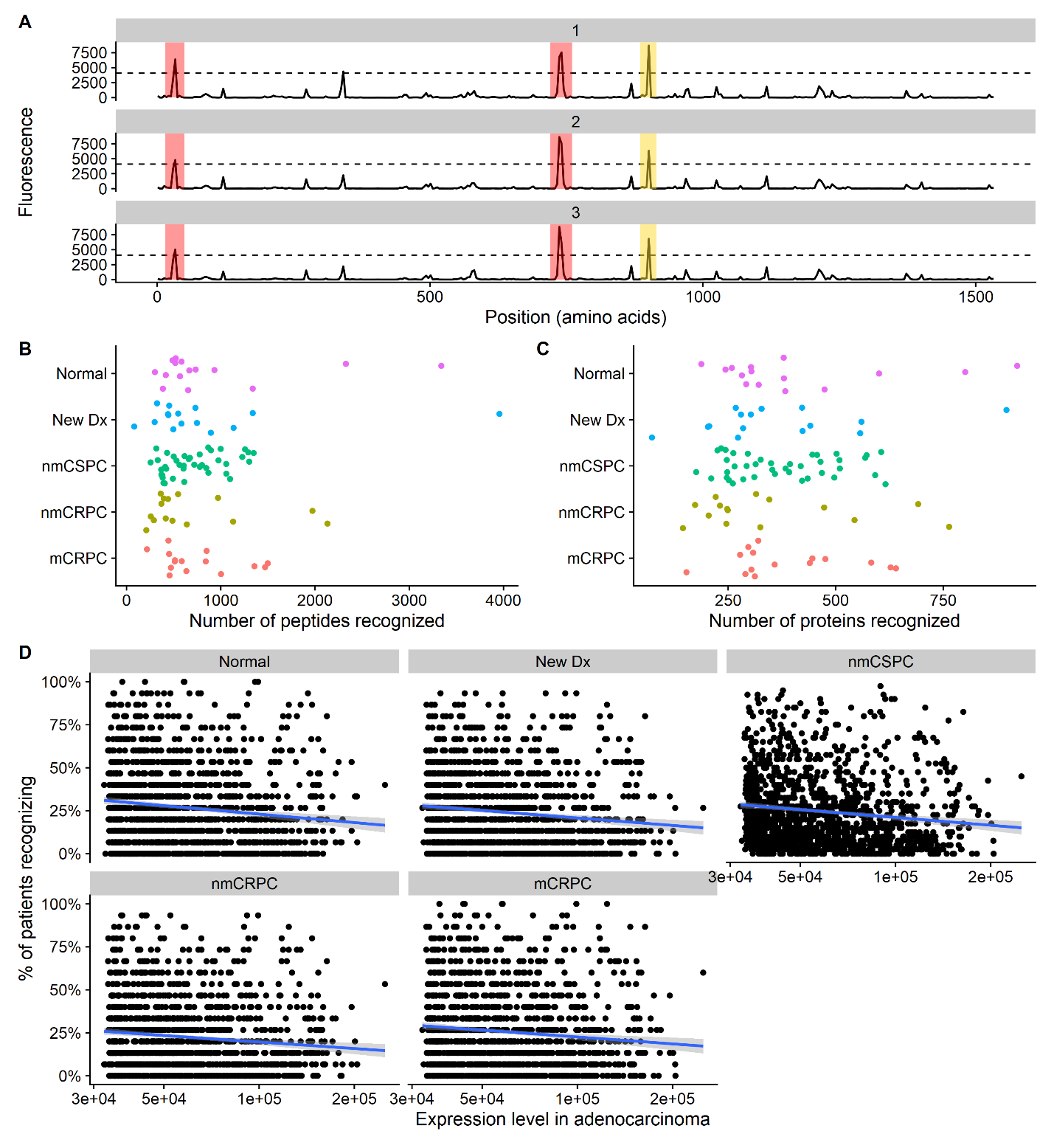
We obtained serum samples from 15 normal male blood donors (Normal), 15 patients with newly diagnosed prostate cancer (New Dx), 40 patients with castration-sensitive non-metastatic prostate cancer (nmCSPC), 15 patients with castration-resistant non-metastatic prostate cancer (nmCRPC), and 15 patients 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 pair of technical replicates and found high correlation on average among replicates (Fig 1C). To determine the degree of variability among 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 data8,9 on antibody responses in prostate cancer, we looked at Prostate Specific Antigen (PSA), Prostatic Acid Phosphatase (PAP), and the Androgen Receptor (AR). 13.3% of mCRPC patients assayed on the array displayed antibody responses against PSA, while 6.7% of normals had PSA responses. 8.2% of prostate cancer patients and 0% of normals had responses to PAP. Finally, 77.6% of prostate cancer patients and 86.6% of normals had responses to AR.

We first tested whether there was greater similarity between antibody responses in two patients with the same clinical state of disease compared to two patients with different states of disease. However, when we calculated mean correlation coefficients among 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 study were 6 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 stages of disease.



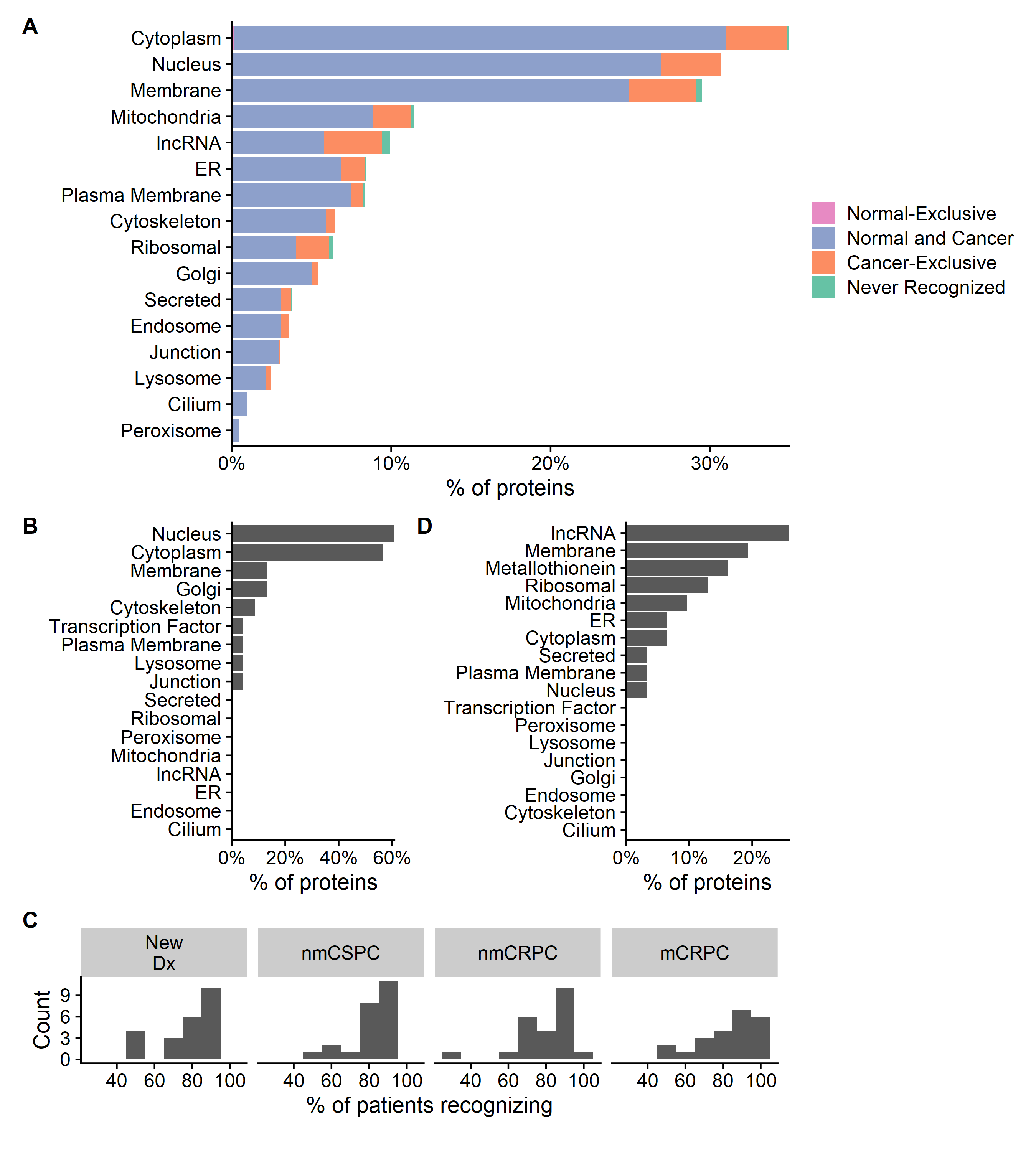
**Figure 1.** A prostate cancer-specific microarray is able to reproducibly measure antibody signatures from serum of healthy individuals and prostate cancer patients. Summary of the **a)** subcellular localization and **b)** length in amino acids of all 1611 unique proteins on the array according to UniProt. **c)** The mean correlation coefficient among all pairs of different individuals (Average Pair) compared to the average correlation coefficient among all technical replicates (Replicate). Error bars represent standard deviation. **d)** Heatmap of the correlation coefficient among 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.

We next tested the hypothesis 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 sliding window p value less than 0.05 (indicating high signal in adjacent peptides) in at least 2 of the 3 technical replicates to be positive as previously described10. Two examples of positive calls are shown in Fig 2A. We found no correlation between stage of disease and the number of probes recognized at either the peptide level or the protein level. The median number of proteins recognized were 321 for normals, 303 for new dx, 353 for nmCSPC, 249 for nmCRPC, and 320 for mCRPC (Fig 2B, 2C). 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 2D). There was a substantial amount of heterogeneity in antibody responses among patients, which appeared to dominate over any potential trends at the stage level. For instance, the number of proteins recognized by normal controls ranged from 188 to 922.



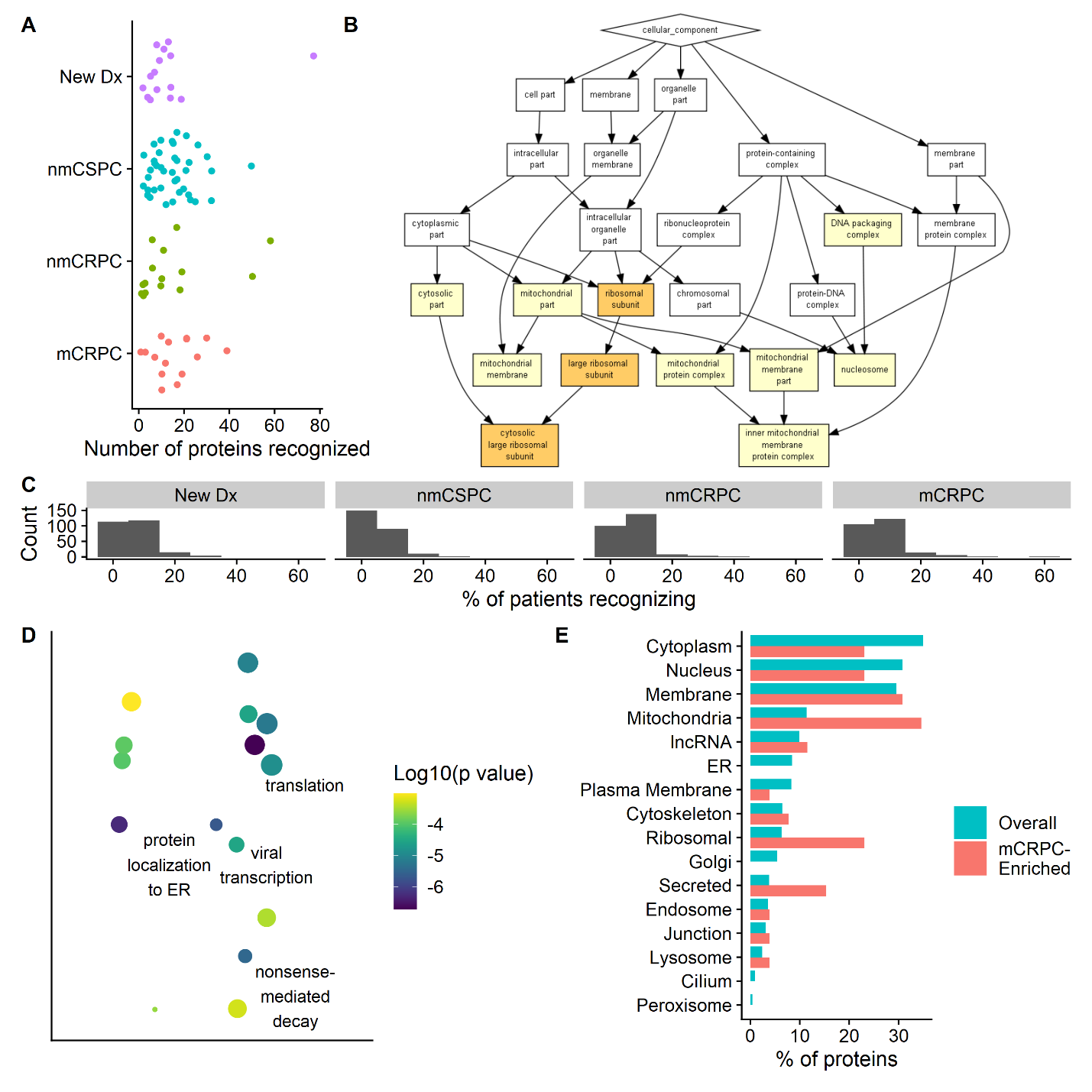
**Figure 2**. Frequency of protein recognition does not correlate with stage of disease or expression level of each protein in prostate cancer. **a)** Example microarray data for technical replicates of a single protein (ADT14) with the 212 signal threshold indicated by the dashed line. Positive calls are marked in red. In yellow is a negative call that did not meet the sliding window criterion. The number of **b)** peptides and **c)** proteins recognized by each patient, categorized by stage of disease. **d)** 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 patients, we next examined whether there were any broad trends in the types of proteins that were recognized. Nearly all proteins (1570 of 1611, 97%) were recognized by at least patient. Conversely, there were no proteins that were recognized by all patients. Most proteins (1326 of 1611, 82%) were recognized by both normals and cancer patients (Fig 3A). Contrary to our expectations, the majority of lncRNAs (145 of 160, 91%) were recognized by at least one patient. We identified 23 proteins that were recognized by at least 90% of normal controls (Normal Proteins, Supplementary Table 2). We found that many of these were nuclear proteins, but found no specific features that were enriched via gene ontology analysis (Fig 3B). These proteins commonly recognized by healthy volunteer blood donors were also recognized at high levels by patients with each stage of cancer (Fig 3C). As expected, the largest category of proteins that were never recognized were ORFs from lncRNAs, as these would not normally be expected to 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 controls are also recognized in all stages of cancer. **a)** Percentage of proteins that are recognized by only normal patients (*Normal-Exclusive*), percentage of proteins recognized by 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.

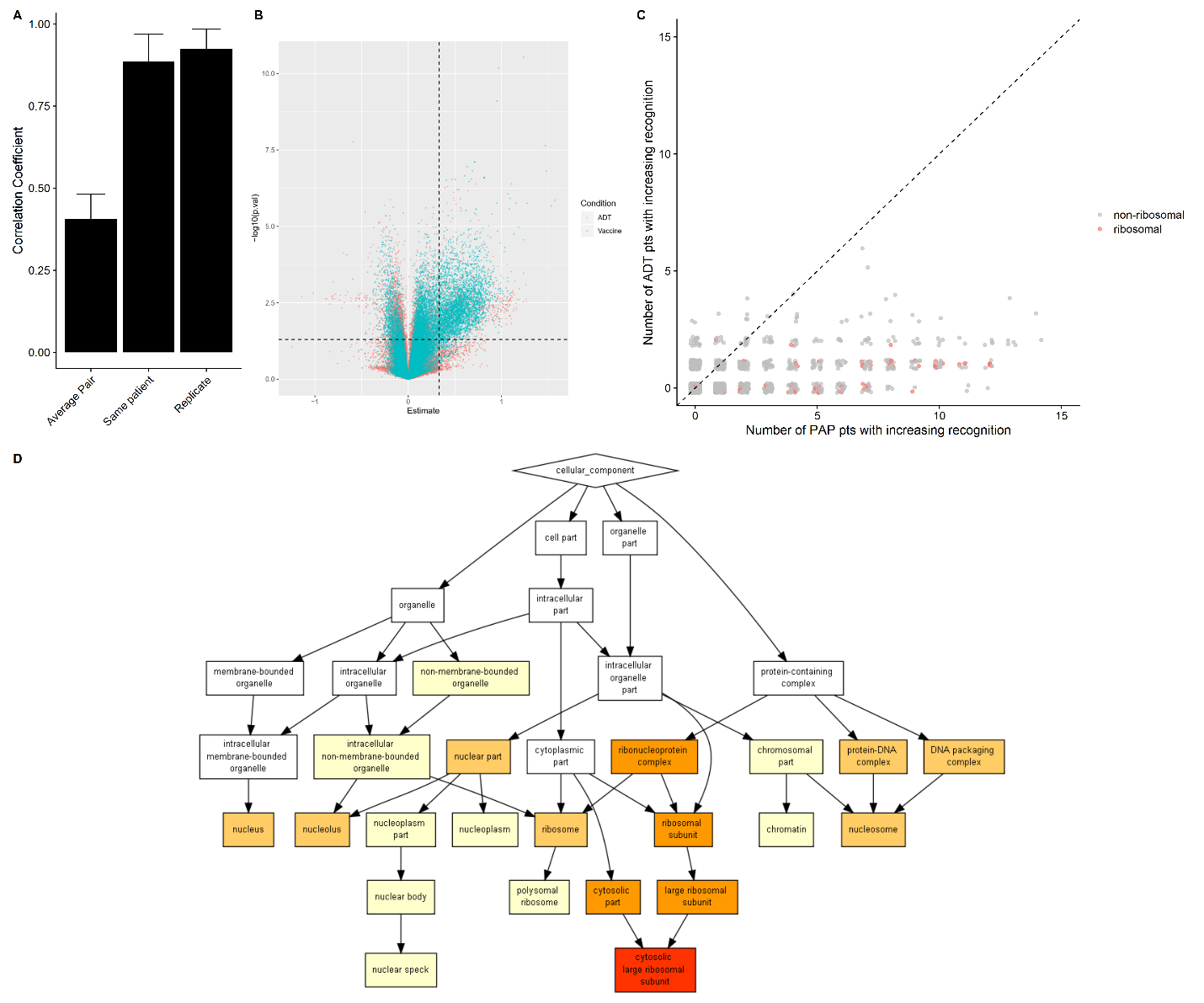
We found that there were only three proteins that were recognized by a normal patient but not by any cancer patients (Supplementary Table 3). However, there was a group of 241 proteins that were recognized exclusively in cancer patients (cancer-exclusive) as described in Fig 3A and Supplementary Table 4. We hypothesized that they would be more frequently 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). While many of these proteins were recognized at relatively low levels, some were recognized by nearly 60% of patients in a given stage (Fig 4C). GO process analysis revealed that cancer-exclusive proteins were associated with regulation of translation, nonsense-mediated mRNA decay, and protein targeting (Fig 4D). We also examined proteins that were either exclusively recognized by patients with mCRPC 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 31 of these proteins (Supplementary Table 5). Over half (16 of 31, 51%) of these proteins were mitochondrial, ribosomal, or lncRNAs (Fig 4E). GO analysis revealed that a significant number of proteins involved were associated with nuclear division.



**Figure 4.** Patients with mCRPC preferentially recognize mitochondrial, ribosomal, and lncRNA gene products. **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. X and Y axes are arbitrary, but more similar GO terms are plotted closer 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*).

Based on our finding that patients tend to have unique antibody response profiles that are largely stable over time, we next investigated whether treatment induces changes in antibody repertoires in individuals over time. To address this question, we used serum samples available from 40 patients with nmCSPC treated with one of two therapies. 20 patients received standard androgen deprivation therapy (ADT; GnRh analogue given every 3 months)11, and 20 patients received a DNA vaccine encoding prostatic-acid phosphatase (PAP; pTVG-HP given every 14 days for 6 administrations)12. Samples were collected from these patients at baseline, 3 months, and 6 months following initiation of treatment. Consistent with our observations in Fig 1E, we found high correlation between samples from an individual patient over time (Fig 5a). We next fit a linear model to each epitope for each patient. We found that using this peptide microarray approach we were able to track changes in antibody responses over time in individuals and found that antibody responses to many epitopes increased following treatment (Fig 5b). Furthermore, we found that many vaccine-treated patients developed increasing responses over time to the same proteins (Fig 5c). There were 177 proteins to which significantly more (by Fisher’s exact test) vaccine-treated patients than ADT patients developed increasing antibody responses, while no proteins had increasing recognition over time by significantly more ADT patients than vaccine-treated patients (Supplementary Table 6). In the most extreme case, 14 of 20 PAP patients had increasing responses to the HMG box transcription factor while only 2 of 20 ADT patients had an increasing response to this same protein (p=0.0002, Fisher’s exact test).

To further characterize which types of proteins vaccine-treated patients were developing increased responses to following vaccination, we performed GO analysis on this set of 177 proteins. We found that a significantly enriched set of these antibodies were specific to DNA binding proteins and histone binding proteins. There were also more antibodies against ribosomal proteins, nuclear proteins, and nucleosomes than would be expected by chance (Fig 5d). Many of the proteins were involved in regulation of translation and chromatin organization.

****

**Figure 5.** Antigen-specific vaccination elicits greater increases in antibody responses over time than hormonal therapy does. **a)** The mean correlation coefficient among all pairs of different individuals (*Average Pair*) compared to the average correlation coefficient among all technical replicates (*Replicate*) and the average correlation among samples collected from the same patient at different time points (*Same patient*). Error bars represent standard deviation. **b)** Volcano plot of the results of a linear model fit to each epitope for each patient. The vertical line represents an estimate of 0.33 (peptides that have a two-fold increase in signal every 3 months) and the horizontal line represents a p value of 0.05. The red dots correspond to ADT patient response to a given peptide, while blue dots represent vaccine-treated patients. **c)** Scatterplot in which each point represents a protein. The x-axis indicates how many PAP patients developed increasing response to a given protein over time, while the y-axis indicates how many ADT patients developed increasing response to the same protein. The dashed line corresponds to y = x. **d)** GO cellular component results using proteins that significantly more PAP patients than ADT patients developed increasing responses to as the target set and the full list of proteins on the array as the background set.

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. Maricque, B. B., Eickhoff, J. C. & McNeel, D. G. Antibody responses to prostate-associated antigens in patients with prostatitis and prostate cancer. *Prostate* **71**, 134–146 (2011).

3. Guo, Z.-W. *et al.* Translated Long Non-Coding Ribonucleic Acid ZFAS1 Promotes Cancer Cell Migration by Elevating Reactive Oxygen Species Production in Hepatocellular Carcinoma. *Front Genet* **10**, (2019).

4. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5’UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **4**,.

5. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity of Mammalian Proteomes. *Cell* **147**, 789–802 (2011).

6. 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).

7. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506–D515 (2019).

8. McNeel, D. G. *et al.* Antibody immunity to prostate cancer associated antigens can be detected in the serum of patients with prostate cancer. *J. Urol.* **164**, 1825–1829 (2000).

9. Olson, B. M. & McNeel, D. G. Antibody and T-cell responses specific for the androgen receptor in patients with prostate cancer. *The Prostate* **67**, 1729–1739 (2007).

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

11. Lang, J. M. *et al.* A Randomized Phase II Trial Evaluating Different Schedules of Zoledronic Acid on Bone Mineral Density in Patients With Prostate Cancer Beginning Androgen Deprivation Therapy. *Clin Genitourin Cancer* **11**, (2013).

12. McNeel, D. G. *et al.* Safety and Immunological Efficacy of a DNA Vaccine Encoding Prostatic Acid Phosphatase in Patients With Stage D0 Prostate Cancer. *J Clin Oncol* **27**, 4047–4054 (2009).

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?

Ribosomal proteins are recognized at high rates: maybe since cancer upregulates translational machinery

PAP/PSA/AR data agrees with literature

Specific proteins recognized by PAP patients could be biomarkers of response

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?

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?