Supplemental text

Quality control of RNA and RNA-seq

To perform the quality control of our samples (registered on EV-TRACK[1] waiting approval code), we first evaluated, by Nanoparticle tracking analysis (NTA), the sizes distribution of uEVs. The analysis showed classical EVs size distribution, with a peak at 120 nm, suggesting that we got exosomes and microvesicles (Extended Data Fig. 1a,b). Before sequencing, we examine the quantity and quality of extracted RNA from uEVs and FFPE PCa samples. By Qubit and NanoDrop RNA quantification we found that uEVs RNA yields were heterogenous and depend on the patient from 200 ng to 3 µg RNA. The ratios 260/280 and 260/230 were on average 1.9 and 1.1 respectively (Extended Data Fig. 1d and Extended Data Table1). uEVs and FFPE RNA samples were checked by Bioanalyzer and found to have different RNA profiles (Extended Data Table1). All uEVs RNAs shown similar profiles to those previously published with a peak at 120 nt (Extended Data Fig.1e). FFPE RNA profiles shown RNA degradation but except for one sample, DV200 were >70% (Extended Data Table1).

The RNA-seq reads quality and alignment were controlled by using MultiQC and shown an average of 22.7 million of total paired reads (17.6 M to 29.5 M) and 20 million (18 M to 22 M) for respectively, 6 tumor prostate FFPE biopsies and 6 uEVs. The percentage of aligned reads on human genome hg38 were on average 92% (86% to 96%) and 80% (69% to 88%) of unique reads. Because of ribosomal RNAs depletion during the library's preparations, only around 1.8% of aligned reads correspond to ribosomal RNAs (Extended Data Table 1).

The workflow of the bioinformatics analysis is presented in Extended Data Fig.1f. Read counting was performed for each sample, on the human gene annotation (Gencode v32) and on the human repeats, using Kallisto. In parallel, CIRI2 was used to discover the location of circRNAs (back-spliced reads) and to quantify them. Annotation and counts from human genes, repeats and circRNAs were combined to establish a count table with 61 053 official annotated genes, 15 352 repeats and 38 793

circRNAs. Then, the conditions FFPE and uEVs were compared using DESeq2. Only the features with adjusted p-value <= 0.05 and fold change >= 1.5 were retained as differentially expressed (Extended Data Fig.1f).

Selection of genes expressed in urinary EVs, specific to the prostate tissue

To ascertain the presence and select of prostate specific genes expressed in uEVs, we took advantage of public RNAseq data from a study that included 7 prostate normal tissues that we compared to two other urological tissues, that are in relation with urine, 4 kidney and 6 urinary bladder normal tissues, to eliminate transcripts coming from these last two tissues[2], accession numbers ERR315340, ERR315468 and ERR315453. We subsequently did differential expression analysis between prostate and kidney and, between prostate and urinary bladder using the same criteria as for FFPE versus uEVs analysis (Gencode 32, fold change >=1.5 and padj <=0.05) (Extended Data Fig. 2a).

Overlapping genes derived from each comparison: expressed genes in prostate cancer urinary EVs (n=10937, Gencode32 annotation, mean counts >=20), down regulated genes in healthy kidney (n=3644) and in urinary bladder (n=3982) compared to healthy prostate, were shown using Venn diagram. At least 1248 uEVs expressed genes are the subset of down regulated genes in kidney and urinary bladder compared to prostate tissue indicating that at least 1248 robust RNAs of uEVs come from prostate (*SuperExactTest*, $p<10^{-320}$)[3] . Because the sequences came from polyA-RNAs, we could not access to circRNAs, unpolyadenylated lncRNAs and small RNAs from the public data and probably much more uEVS transcripts derived from prostate exclusively (Extended data Fig. 2b and Extended Table 3). To verify the origin of uEVs RNAs, we checked the expression, in uEVs and FFPE, of several known RNAs markers of normal or tumor prostate tissues, KLK2, KLK3, PCA3 and PMEPA1. Normal and tumor prostate markers were confirmed highly expressed in FFPE and uEVs. On the contrary, UMOD and UPK2 RNAs highly expressed in kidney and urinary bladder respectively were almost not

50 expressed in FFPE tissues and present at low level in uEVs showing the enrichment of EVs derived
51 from prostate (tumor) tissue in collected urine (Extended Data Fig. 2c).

52

Extended figures legends

Extended Data Fig. 1. **Full transcriptome of paired liquid and solid biopsies of prostate cancer patients.** **a**. Characterization of uEVs by nanoparticle tracking analysis (NTA) showing average distribution of vesicle size and number. **b.** Visualization screen shot from video of light scatter of vesicles obtained by NTA. Microscope magnification: x10. **c.** The pellets recovered from 3 ultracentrifuged urines (lines 1,2,3) and HEK293 cell lysate (line 4) were analyzed by Western blot for the indicated proteins side by side. Stain free gel images of total amounts of proteins were used to quantify the relative level of analyzed proteins. **d.** Optical density profiles of RNAs giving their purity. Position of 230 and 260 nm absorbance are shown by black vertical lines. **e.** RNA Profiles analyzed by capillary electrophoresis giving their quality. Electropherograms show in the y-axis fluorescence units (FU) and in the x-axis the nucleotide length (nt) of the RNA. Peaks at 25 nucleotides represent internal standards and peaks at 2,000 nt and 4000 nt represent 18S and 28 S ribosomal RNAs, respectively. **f.** Bioinformatic procedure for differential RNA expression analysis between Tumor biopsies and uEVs. The number of total official annotated RNAs, repeats and circRNAs found in Tumor and uEvs are indicated. **g.** Number of expressed RNA features in Tumor and uEVs: 14,027 and 31,132 circRNAs (orange), 19,508 and 17,599 mRNA (violet), 12,647 and 7,763 lncRNAs (green), 6,655 and 4,082 pseudogenes (yellow), 3,224 and 409 repeats (pink), 718 and 132 snRNAs (blue), 680 and 74 Pre-miRNAs (dark pink), 530 and 180 snoRNAs (red), 1,378 and 739 others (miscRNA, ribozyme, rRNA, scaRNA, scRNA, sRNA, tRNA; grey), respectively.

Extended Data Fig. 2. **Selection of genes expressed in uEVs, specific to prostate tissue**. **a**. Workflow of analysis to select prostate specific expressed genes in uEVs biopsies. **b**. Venn diagram showing expressed genes in prostate cancer uEVs (white n= 10,937), down regulated genes in healthy kidney (blue, n=3644) and in urinary bladder (grey, n=3982) compared to healthy prostate. **c**. Box-plot of Log10 TPM normalized expression of prostate (KLK2, KLK3, PCA3 and PMEPA1), kidney UMOD) and urinary bladder (UPK2) specific RNAs in Tumor biopsies (blue) and uEVs (red).

Extended Data Fig. 3. **uEVs are enriched in circRNAs and lncRNAs. a.** Mean gene expression of paired Tumor against uEVs. DEseq2 normalized counts, for each type of RNA are plotted; circRNAs (orange), lncRNAs (green), all others type of RNAs (grey). Each dot represents all transcripts for each gene. From 49,553 RNAs of patient 1, R^2=0.0395 for 7,087 circRNAs and R^2=0.2486 for 9,617 lncRNAs. From 41,909 RNAs of patient 2, R^2=0.0013 for 13,482 circRNAs and R^2=0.3009 for 10,169 lncRNAs. From 46,773 RNAs of patient 3, R^2=10^{-5} for 11,350 circRNAs and R^2=0.3086 for 9,999 lncRNAs. From 45,638 RNAs of patient 4, R^2=0.0005 for 11,056 circRNAs and R^2=0.2748 for 9,696 lncRNAs. From 45,850 RNAs of patient 5, R^2=0.0029 for 10,982 circRNAs and R^2=0.2878 for 9,777 lncRNAs. From 47,812 RNAs of patient 6, R^2=0.0148 for 13,821 circRNAs and R^2=0.3975 for 9,452 lncRNAs. **b.** Number of over-expressed RNA species in Tumor (top) and in uEVs (bottom). **c.** Heatmap display unsupervised hierarchical clustering euclidean distance (CED) of all differentially expressed transcripts (n = 13,261, fold change >= 1,5) in each of the individual samples from 6 Tumor biopsies (turquoise) and paired 6 uEVs (magenta). Color scales represent log10(normalized counts +1).

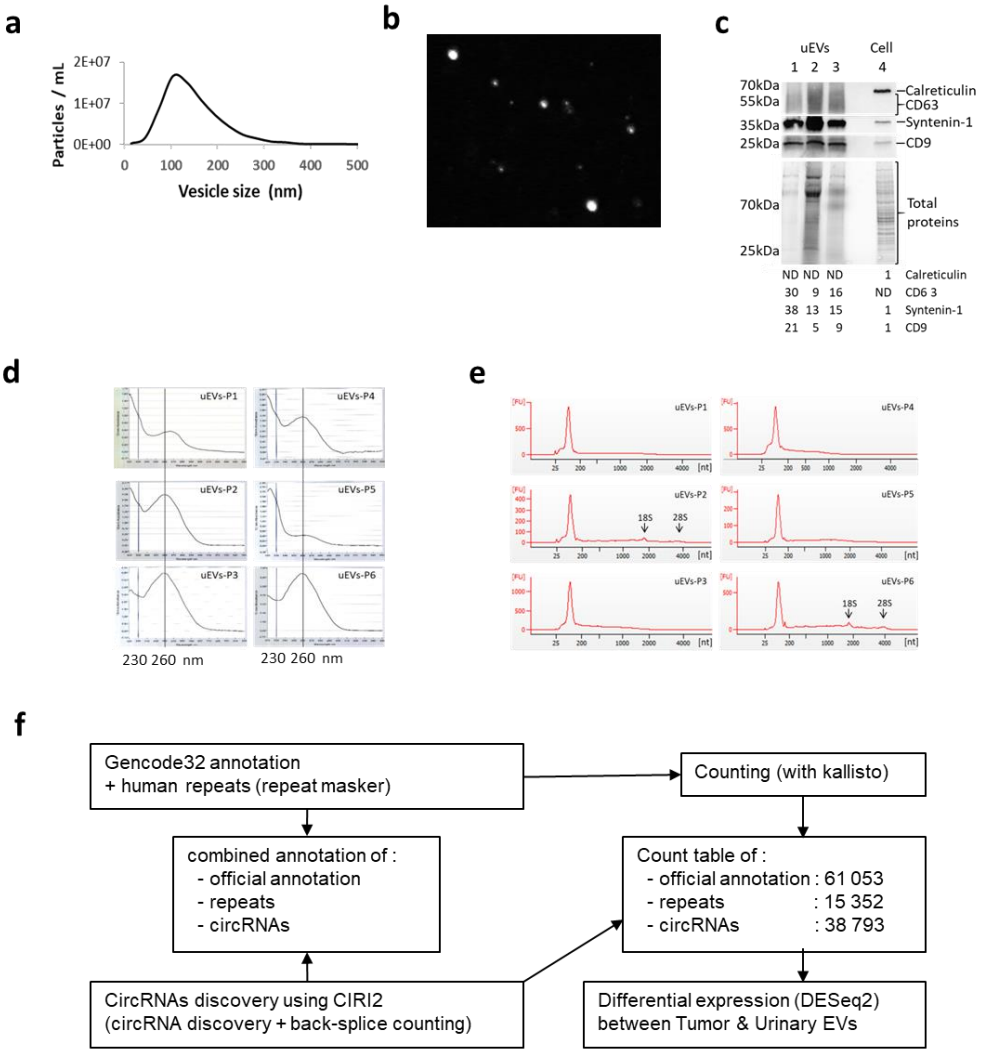
Extended Data Fig. 4. **uEVs lncRNAs can be predicted to form strong binding neoantigens.** Predicted binding score rank for 65,190 Tumor-neoantigens coming from 255 lncRNA transcripts and for 3,298 uEVs-neoantigens from 16 lncRNA transcripts. 15,677 and 768 strong score lncRNA-neopeptides (score<0.5) were predicted from Tumor (blue, median score= 0.253) and uEVs (red, median score=0.274) samples, respectively. 49,513 and 2,530 weak score lncRNA-neopeptides (score>0.5) were predicted from Tumor (median score=1.258) and uEVs (median score=1.283) samples, respectively. The difference between Tumor and uEVs rank of strong neoantigens is significant ($p<0.05$) but not for weak neoantigens (NS).

References

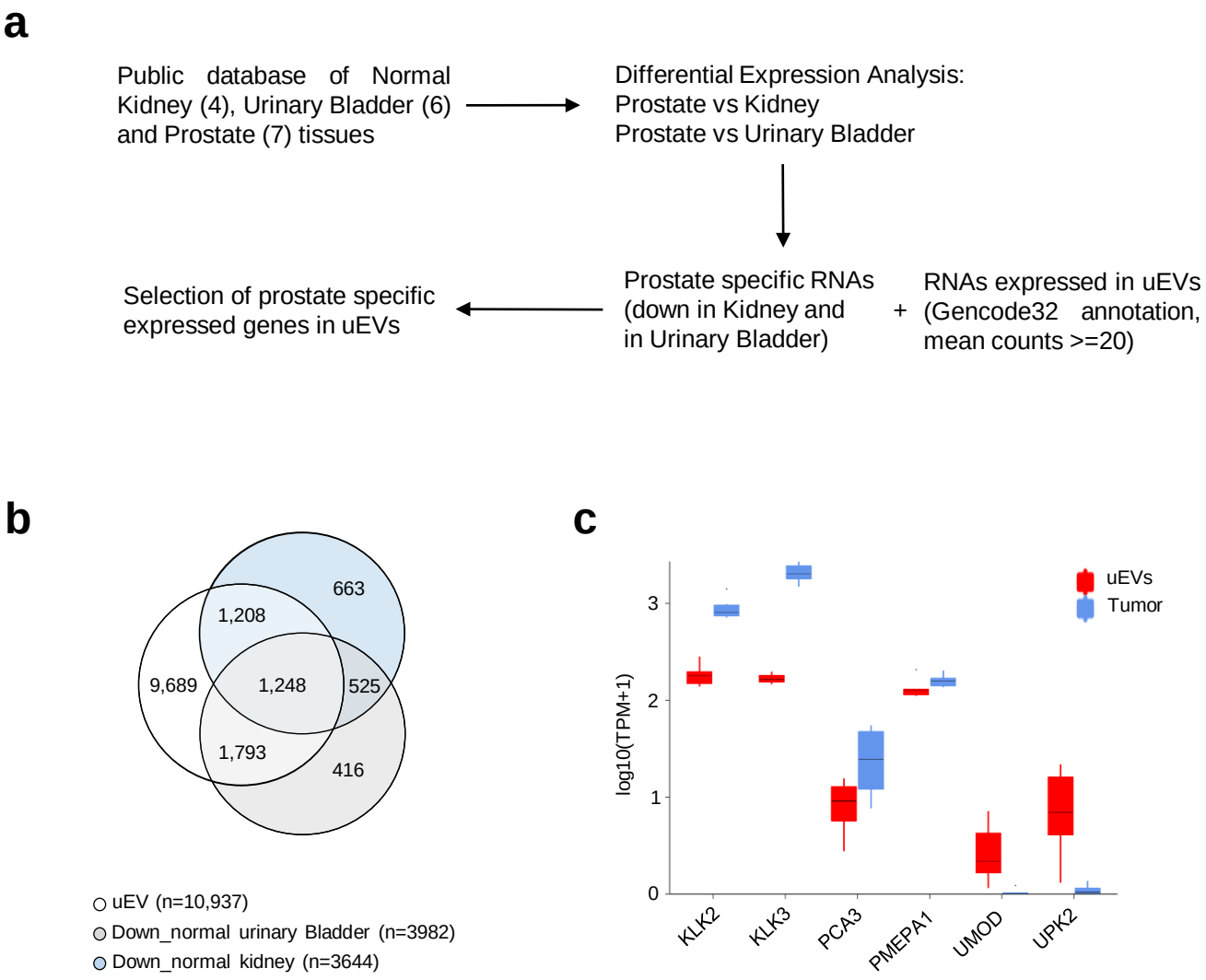
1 Van Deun, J. *et al.* EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research. *Nature methods* **14**, 228-232, doi:10.1038/nmeth.4185 (2017).

2 Fagerberg, L. *et al.* Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. *Molecular & cellular proteomics : MCP* **13**, 397-406, doi:10.1074/mcp.M113.035600 (2014).

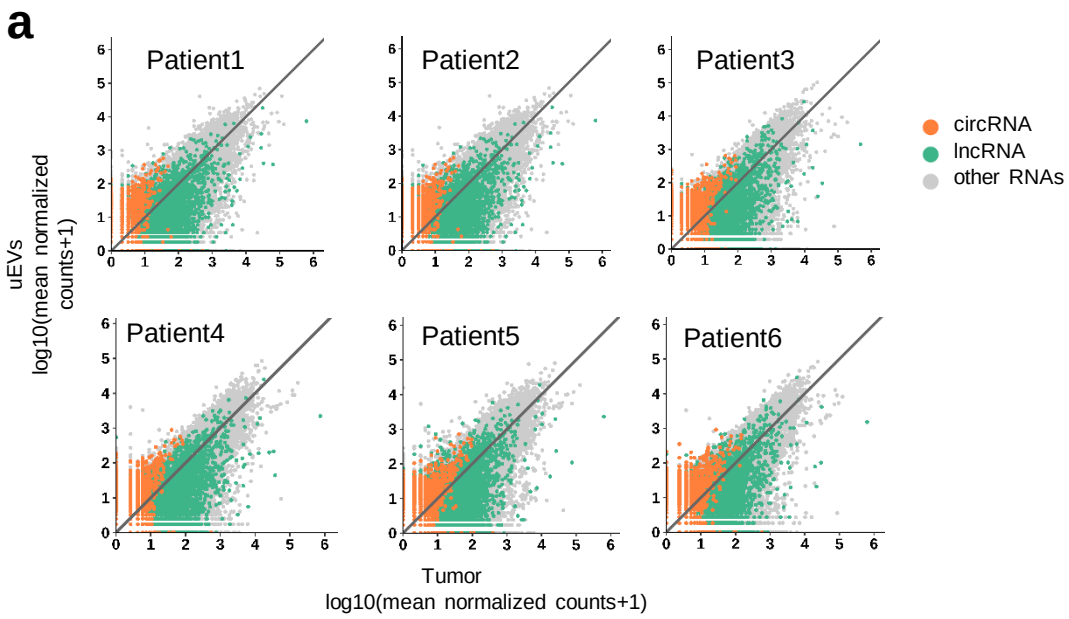
3 Wang, M., Zhao, Y. & Zhang, B. Efficient Test and Visualization of Multi-Set Intersections. *Scientific reports* **5**, 16923, doi:10.1038/srep16923 (2015).

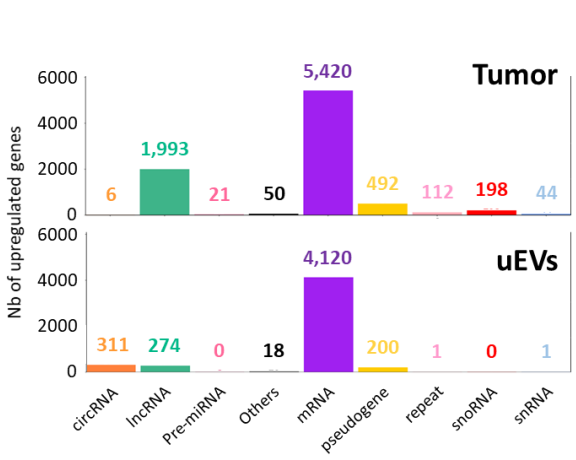


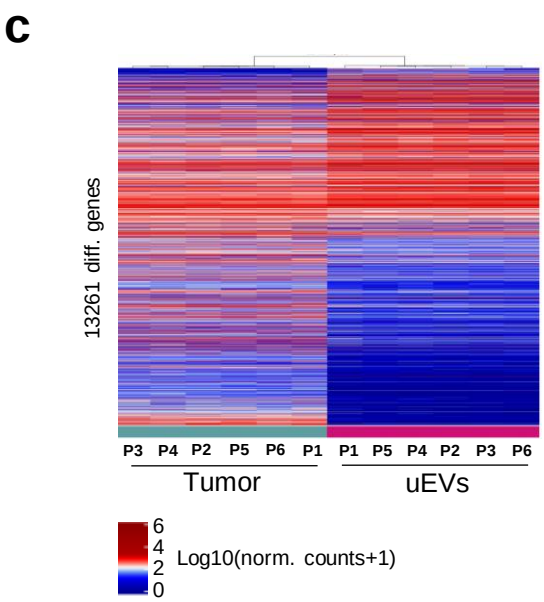
Extended Data Fig. 1. **Full transcriptome of paired liquid and solid biopsies of prostate cancer patients.** **a**. Characterization of uEVs by nanoparticle tracking analysis (NTA) showing average distribution of vesicle size and number. **b.** Visualization screen shot from video of light scatter of vesicles obtained by NTA. Microscope magnification: x10. **c.** The pellets recovered from 3 ultracentrifuged urines (lines 1,2,3) and HEK293 cell lysate (line 4) were analyzed by Western blot for the indicated proteins side by side. Stain free gel images of total amounts of proteins were used to quantify the relative level of analyzed proteins. **d.** Optical density profiles of RNAs giving their purity. Position of 230 and 260 nm absorbance are shown by black vertical lines. **e.** RNA Profiles analyzed by capillary electrophoresis giving their quality. Electropherograms show in the y-axis fluorescence units (FU) and in the x-axis the nucleotide length (nt) of the RNA. Peaks at 25 nucleotides represent internal standards and peaks at 2,000 nt and 4000 nt represent 18S and 28 S ribosomal RNAs, respectively. **f.** Bioinformatic procedure for differential RNA expression analysis between Tumor biopsies and uEVs. The number of total official annotated RNAs, repeats and circRNAs found in Tumor and uEvs are indicated. **g.** Number of expressed RNA features in Tumor and uEVs: 14,027 and 31,132 circRNAs (orange) including 4947 and 10723 present in circBase (hatched orange), 19,508 and 17,599 mRNA (violet), 12,647 and 7,763 lncRNAs (green), 6,655 and 4,082 pseudogenes (yellow), 3,224 and 409 repeats (pink), 718 and 132 snRNAs (blue), 680 and 74 Pre-miRNAs (dark pink), 530 and 180 snoRNAs (red), 1,378 and 739 others (miscRNA, ribozyme, rRNA, scaRNA, scRNA, sRNA, tRNA; grey), respectively.



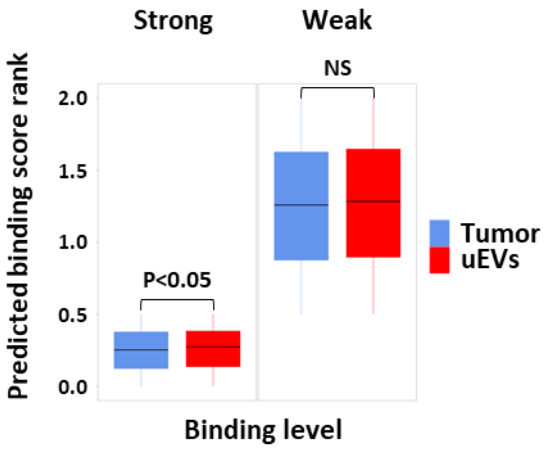
Extended Data Fig. 2. **Selection of genes expressed in uEVs, specific to prostate tissue**. **a**. Workflow of analysis to select prostate specific expressed genes in uEVs biopsies. **b**. Venn diagram showing expressed genes in prostate cancer uEVs (white n= 10,937), down regulated genes in healthy kidney (blue, n=3644) and in urinary bladder (grey, n=3982) compared to healthy prostate. **c**. Box-plot of Log10 TPM normalized expression of prostate (KLK2, KLK3, PCA3 and PMEPA1), kidney UMOD) and urinary bladder (UPK2) specific RNAs in Tumor biopsies (blue) and uEVs (red).







Extended Data Fig. 3. **uEVs are enriched in circRNAs and lncRNAs. a.** Mean gene expression of paired Tumor against uEVs. DEseq2 normalized counts, for each type of RNA are plotted; circRNAs (orange), lncRNAs (green), all others type of RNAs (grey). Each dot represents all transcripts for each gene. From 49,553 RNAs of patient 1, R^2=0.0395 for 7,087 circRNAs and R^2=0.2486 for 9,617 lncRNAs. From 41,909 RNAs of patient 2, R^2=0.0013 for 13,482 circRNAs and R^2=0.3009 for 10,169 lncRNAs. From 46,773 RNAs of patient 3, R^2=10^{-5} for 11,350 circRNAs and R^2=0.3086 for 9,999 lncRNAs. From 45,638 RNAs of patient 4, R^2=0.0005 for 11,056 circRNAs and R^2=0.2748 for 9,696 lncRNAs. From 45,850 RNAs of patient 5, R^2=0.0029 for 10,982 circRNAs and R^2=0.2878 for 9,777 lncRNAs. From 47,812 RNAs of patient 6, R^2=0.0148 for 13,821 circRNAs and R^2=0.3975 for 9,452 lncRNAs. **b.** Number of over-expressed RNA species in Tumor (top) and in uEVs (bottom). **c.** Heatmap display unsupervised hierarchical clustering euclidean distance (CED) of all differentially expressed transcripts (n = 13,261, fold change >= 1,5) in each of the individual samples from 6 Tumor biopsies (turquoise) and paired 6 uEVs (magenta). Color scales represent log10(normalized counts +1).



Extended Data Fig. 4. **uEVs lncRNAs can be predicted to form strong binding neoantigens.** Predicted binding score rank for 65,190 Tumor-neoantigens coming from 255 lncRNA transcripts and for 3,298 uEVs-neoantigens from 16 lncRNA transcripts. 15,677 and 768 strong score lncRNA-neopeptides (score<0.5) were predicted from Tumor (blue, median score= 0.253) and uEVs (red, median score=0.274) samples, respectively. 49,513 and 2,530 weak score lncRNA-neopeptides (score>0.5) were predicted from Tumor (median score=1.258) and uEVs (median score=1.283) samples, respectively. The difference between Tumor and uEVs rank of strong neoantigens is significant (p<0.05) but not for weak neoantigens (NS).