Supplemental text

Quality control of RNA and RNA-seq

To perform the quality control of our samples (registered on EV-TRACK platform[1] with the EV-TRACK ID EV210255), we first evaluated, by Nanoparticle tracking analysis (NTA), the sizes distribution of uEVs and cEVs. The analysis showed classical EVs size distribution ranging from 50 to 300nm in diameter, with a peak between 100 and 150 nm, suggesting that we got exosomes and microvesicles (Extended Data Fig. 1a). Transmission Electron Microscopy images obtained for uEVs, DU145-, LNCaP- and PC3-EVs showed the presence, in each sample, of various size EVs. We also noticed that the amount of EVs present in urine and produced by DU145 and PC3 was high unlike the EVs produced by LNCaP cells. UEVs preparation seems clean without contamination by polymers of the Tam Horsfall protein (THP; uromodulin) (Extended Data Fig. 1b). Before sequencing, we examine the quantity and quality of extracted RNA from uEVs, cEVs, cells and FFPE PCa samples. By Qubit and NanoDrop RNA quantification we found that uEVs RNA yields were heterogenous and depend on the patient from 150 ng to 3 µg RNA. The ratios 260/280 and 260/230 were on average 1.9 and 1.1 respectively for the 6 paired uEVs samples, 1.8 and 0.9 for the 20 independent uEVs samples, 1.7 and 0.8 for cell-EVs and 2 and 1.8 for cells (Extended Data Fig. 1eand Extended Data Table1). uEVs, cEVs and FFPE RNA samples were checked by Bioanalyzer and found to have different RNA profiles (Extended Data Fig.1f, g, Extended Data Table1). All 6 paired uEVs RNAs shown similar profiles to those previously published with a peak at 120 nt (Extended Data Fig.1f). FFPE RNA profiles shown RNA degradation but except for one sample, DV200 were >70% (Extended Data Fig. 1g, Extended Data Table1). cEVs RNAs and independent uEVs RNAs presented similar profiles to the 6 paired uEVs ones (data not shown).

The RNA-seq reads quality and alignment were controlled by using MultiQC and shown an average of 45million of total reads (35 M to 59M) and 45 million (36M to 44 for respectively, 6 tumor prostate FFPE biopsies and 6 uEVs. The percentage of aligned reads on human genome hg38 were on average

94% (93% to 96%) and 90% (86% to 94%) of 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 quality control for the 20 independent uEVs and cell-EVs gave similar data (Extended Data Table 1).

The workflow of the bioinformatics analysis is presented in Extended Data Fig.2a. 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.2a).

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

To ascertain the presence and select of prostate enriched 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. 3a).

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. 3b and Extended Data 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 expressed in FFPE tissues and present at low level in uEVs showing the enrichment of EVs derived from prostate (tumor) tissue in collected urine (Extended Data Fig. 3c).

Extended Data Figures legends

Extended Data Fig. 1. Characterization of EVs isolated from urine and culture supernatant. a. Characterization of EVs by nanoparticle tracking analysis (NTA) showing average distribution of vesicle size coming from urine (black), PC3 (orange), LNCaP (bleu) and DU145 (pink) cell lines. **b.** Transmission electron microscopy images of EVs from urine, PC3, LNCaP and DU145 cell lines. Scale bar sizes are 100 and 200 nm for urinary EVs and cell EVs respectively. Yellow arrows show some EVs with variable size**. 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.** Western blot analysis, for the indicated proteins side by side, of PC3 and DU145 cell lysates (C) and enriched cell-EVs (V). Total amounts of proteins were shown by stain free gel images. **e.** Optical density profiles of RNAs giving their purity. Position of 230 and 260 nm absorbance are shown by black vertical lines. **f and g.** 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 10 ng of RNA extracted from uEVs (f) and 0.5 ng from FFPE (g) paired samples. Peaks at 25 nucleotides represent internal standards and peaks at 2,000 nt and 4000 nt represent 18S and 28 S ribosomal RNAs, respectively.

Extended Data Fig. 2. Full transcriptome of paired liquid and solid biopsies of prostate cancer patients. a. 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. **b**. Number of expressed RNA features in Tumor and uEVs: 1,240 and 8,796 circRNAs (orange) including respectively 819 (66%) and 4,530 (52%) present in circBase (hatched orange), 15,490 and 13,614 mRNA (violet), 3,776 and 1,907 lncRNAs (green), 1,225 and 823 pseudogenes (yellow), 214 and 43 repeats (pink), 63 and 35 snRNAs (blue), 53 and 7 Pre-miRNAs (dark pink), 245 and 31 snoRNAs (red), 136 and 100 others (miscRNA, ribozyme, rRNA, scaRNA, scRNA, sRNA, tRNA; grey), respectively. **c.** Principal component analysis from expression of 311 circRNAs (left) and

274 lncRNAs (right) upregulated in uEVs (FC>= 1.5, at least 20 counts), in 6 paired tumors (green) and uEVs (black) and 20 independent uEVs (pink), from prostate cancer patients.

Extended Data Fig. 3. Selection of genes expressed in uEVs, enriched in prostate tissue. a. Workflow of analysis to select prostate enriched 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=3,644) and in urinary bladder (grey, n=3,982) 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. 4. 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. Following values are given for at least 5 counts for circRNA expression and 20 counts for all other RNA expression. From 21,633 RNAs of patient 1, R2=0.007 for 2,310 circRNAs and R2=0.099 for 2,979 lncRNAs. From 21,147 RNAs of patient 2, R2=0.095 for 2,334 circRNAs and R2=0.01 for 2,811 lncRNAs. From 21,487 RNAs of patient 3, R2=0.02 for 2,369 circRNAs and R2=0.14 for 3,008 lncRNAs. From 21,737 RNAs of patient 4, R2=0.027 for 2,296 circRNAs and R2=0.15 for 3,213 lncRNAs. From 20,565 RNAs of patient 5, R2=0.03 for 1,714 circRNAs and R2=0.11 for 2,857 lncRNAs. From 21,156 RNAs of patient 6, R2=0.033 for 2,633 circRNAs and R2=0.17 for 2,883 lncRNAs. **b.** Number of over-expressed RNA species in Tumor (top) and in uEVs (bottom). **c**. Heatmap display unsupervised hierarchical clustering with 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 (green) and paired 6 uEVs (black). Color scales represent log10(normalized counts +1).

Extended Data Fig. 5. cell-EVs are enriched in circRNAs and lncRNAs and are depleted in introns. a. Number of expressed RNA features in DU145, LNCaP and PC3 Cell lines and related cell-Evs. Hatched orange indicate the number of circRNAs found in circBase. **b.** Mean gene expression of paired Cell against cell-EVs. 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. Following values are given for at least 5 counts for circRNA expression and 20 counts for all other RNA expression. From 17,838 RNAs of DU145 cell line, R2=0.33 for 2,009 circRNAs and R2=0.34 for 1,761 lncRNAs. From 17,666 RNAs of LNCaP cell line, R2=0.2 for 2,384 circRNAs and R2=0.3 for 1,804 lncRNAs. From 19,742 RNAs ofPC3 cell line, R2=0.46 for 4,396 circRNAs and R2=0.45 for 1,546 lncRNAs. **c.** Genomic read counts distribution by percentage across exon, intron, 3'UTR, 5'UTR, intergenic and promoter in DU145, LNCaP, PC3 cell lines and associated cell-EVs.

Extended Data Fig. 6. cell-EVs are depleted in nuclear lncRNAs. a. Stacked barplot distribution, by percentage, of cytoplasmic (blue), nuclear (red), both (yellow) or non-polyA (grey) RNAs of expressed genes in cell and in cell-EVs. Experimental procedure to get the values are presented in fig. 3a. Percentage of expressed RNAs are indicated in each box. Above each barplot are noted the number of total RNAs **b.** Density distribution of log2 (fold change cytoplasmic/nuclear ratio) per RNA types in DU145, LNCaP and PC3 cells and related cEVs (4,048; 2,366; 4,079 RNAs from cells and 3,112; 2,135; 3,157 RNAs from cell-EVs, respectively), mRNA (purple), pseudogene (yellow), lncRNA (green) in Tumor (top) and uEVs (bottom). The left side of dotted line in both graphs corresponds to the nuclear RNAs, the right side corresponds to cytoplasmic RNAs.

Extended Data Fig. 7. uEVs lncRNAs can be predicted to form strong binding neoantigens. a. 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). **b.** Number and lengths of strong predicted neoantigens in Tumor (blue, n=15,677) and uEVs (red, n=768).

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).