Urinary Extracellular Vesicles contain simplified transcriptomes enriched in circular & long noncoding RNAs with functional significance in prostate cancer

anna Almeida
Institut Curie
Marc Gabriel
Institut Curie
Virginie Firlej
Université Paris-Est Créteil Val de Marne
Lorena Martin-Jaular
Institut Curie https://orcid.org/0000-0002-1511-8576
Matthieu Lejars
Institut Curie
Rocco Cippola
Institut Curie
Nicolas Vogt
Institut Curie
Damien Destouches
Université Paris-Est Créteil Val de Marne
Francis Vacherot
Université Paris-Est Créteil Val de Marne
Alexandre De la Taille
Hopital Mondor-APHP
Clotilde Théry
Institut Curie, PSL Research University, INSERM U932 https://orcid.org/0000-0001-8294-6884
Antonin Morillon (✉ antonin.morillon@curie.fr)
Institut Curie https://orcid.org/0000-0002-0575-5264

Research Article

Keywords: EV, lncRNA, CircRNA, neoantigen, TSA

Posted Date: September 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-871568/v2

Urinary Extracellular Vesicles contain simplified transcriptomes enriched in circular & long noncoding RNAs with functional significance in prostate cancer

Almeida*[1,2] Anna, Gabriel*[1] Marc, Firlej[3,4] Virginie, Martin-Jaular[5,6] Lorena, Lejars[1] Matthieu, Cipolla[1] Rocco, Vogt[1] Nicolas, Destouches[4] Damien, Vacherot[4] Francis, de la Taille[7] Alexandre, Théry[5,6] Clotilde, Morillon[1] Antonin

***Running title**: EVs contain functional circular and long noncoding RNAs*

Keywords : EV, lncRNA, CircRNA, neoantigen, TSA

Affiliation

[1] ncRNA, Epigenetic and Genome Fluidity, CNRS UMR3244, Sorbonne Université, PSL University, Institut Curie, Centre de Recherche, Paris, France

[2] Departement de Recherche Translationnelle, PSL University, Institut Curie, Centre de Recherche, Paris, France

[3] AP-HP, Hôpital H. Mondor, Plateforme de Ressources Biologiques, F-94000, Créteil, France

[4] Univ Paris Est Creteil, UR TRePCa, F-94010 Créteil, France

[5] INSERM U932, PSL Research University, Institut Curie, Centre de Recherche, Paris, France

[6] Extracellular Vesicles Platform, Institut Curie, Centre de Recherche, Paris, France

[7] AP-HP, Hôpital H. Mondor, Service d'urologie, F-94010 Créteil, France

* These authors contributed equally to this work.

Corresponding author. Tel: +33 1 56 24 65 15; E-mail: antonin.morillon@curie.fr

Abstract

Long noncoding (lnc)RNAs modulate gene expression alongside presenting unexpected source of neoantigens. Despite their immense interest, their ability to be transferred and control adjacent cells is unknown. Extracellular Vesicles (EVs) offer a protective environment for nucleic acids, with pro and anti-tumorigenic functions by controlling the immune response. In contrast to extracellular non-vesicular RNA, few studies have addressed the full RNA content within human fluids' EVs and none have compared them with their tissue of origin. Here, we performed Total RNA-Sequencing on 6 Formaldehyde-Fixed-Parafilm-Embedded (FFPE) prostate cancer (PCa) tumor tissues and their paired urinary (u)EVs to provide the first whole transcriptome comparison from the same patients. UEVs contain simplified transcriptome with intron-free cytoplasmic transcripts and specific lnc/circular (circ)RNAs, strikingly common to all patients. Our full cellular and EVs transcriptome comparison within 3 common PCa cell lines identified a set of overlapping 14 uEV-circRNAs characterized as essential for prostate cell proliferation in vitro and 15 uEV-lncRNAs that we predicted to encode 768 high-affinity neoantigens. Our dual analysis of EVs-lnc/circRNAs both in urines' and in vitro's EVs provides a fundamental resource for future uEV-lnc/circRNAs phenotypic characterization involved in PCa.

Introduction

Extracellular Vesicles (EVs) are secreted membrane-enclosed vesicles ranging from 50 to 1000 nm[1,2]. They are composed of a variety of proteins, lipids, metabolites and nucleic acids, and their cargo is controlled by specific molecular sorting machineries[3]. Recent efforts have been conducted towards a better and more rigorous classification of the multiple forms of EVs differing in size and composition[4], which led to identification of other particles containing proteins, lipids and nucleic acids, often co-isolated with EVs. These particles have been called exomeres, extracellular nanoparticles (ENPs)[5] or distinct nanoparticles[6]. EVs and probably ENPs protect biomolecules from potential degradation by nucleases, proteases or other environmental stress present in fluids. Since EVs work as a safe way to transport biological information through the whole body, they are now recognized as an important source of biomarkers in clinics but also as a mechanism of cell-cell communication within tumor cells and their distant environment[1,7]. Multiple studies in different cancer types have demonstrated that EVs play an important role in cell proliferation, migration, invasion, angiogenesis, epithelial-to-mesenchymal transition (EMT), metastasis and immune response[1,8-10] . EVs also act as a source of antigenic peptides for the activation of T and B cells[11,12]. Therefore, the role of EVs as potential therapeutic agents is actively considered in addition to their use as diagnostics and prognostics options[13-15] .

RNA studies in biological fluids were mostly focused on small ncRNAs[16-18] or captured mRNA/circRNA exomes[19,20], whenever performed on non-vesicular extracellular (ex)RNAs or EVs/ENPs. Following the early work on urinary microvesicules revealing extensive forms of every types of long RNAs[21], it is only recently that systematic exRNA whole transcriptome approaches started to emerge[22-27]. Whole transcriptome analyses assess the numerous types of RNA molecules including protein-coding RNAs (mRNA), lncRNAs, fusion transcripts, splice variants and RNA modifications. LncRNAs constitute the most prevalent family in human transcriptome and can exert a plethora of functions, such as epigenetic regulation, chromatin remodeling, regulation of proteins' activity and stability, mRNA stability and translation. They make an intimate part of mechanisms controlling cell/tissue homeostasis as well as various diseases including cancers[28]. Although lncRNAs were initially defined as non-coding, several contain open reading frames (ORFs)

that are translated into functional peptides and subjected to translation-dependent decay[29,30]. Among lncRNAs, circRNAs are circular molecules with a covalently closed loop structure lacking 5'cap and 3'polyadenylated tail, formed through noncanonical splicing[31]. Due to high stability, circRNAs can be easily found in circulating fluids as exRNAs, especially in cancer and proposed as potent biomarkers[19,25,26]. CircRNAs were shown to function as miRNA sponges to prevent mRNA translation or as part of ribonucleoprotein complexes regulating splicing or transcription[30,32,33]. Despite their regulatory potential, expression of lncRNAs in circulating EVs/ENPs isolated from cancer patients remains poorly explored. Determination of their identity, origin, and features through comparative transcriptomic studies of liquid and solid biopsies together with existing *in vitro* cell systems is a challenge for understanding lncRNAs and circRNAs functional significance.

Urine is an excellent source of biomarkers containing EVs secreted from different tissues/organs alongside the urogenital tract and is of high value for extravesicular now cargos identification reaching a consensus on the best methodological practices[34]. Prostate cancer (PCa) is the second leading cause of death from cancer in men. During medical digital rectal examination, prostate material, and particularly prostate cancer cells and EVs, can be released from prostatic ducts into the urethra and end up in urine. The identification of two well-known prostate-associated RNAs, PCA3, and TMPRSS2:ERG in urinary EVs[35] was the starting signal for intense efforts to define novel EVs-RNA markers in urines for urological-localized cancers[34]. To date, only 2 extensive transcriptome profiling have targeted PCa patient's vesicular or non-vesicular exRNAs in urines and they are excellent bases for future biomarker discovery on larger cohorts. In an elegant benchmarking study, Everaert et al. were pioneers in performing deep whole transcriptome on carefully isolated urinary vesicular exRNA for prostate cancer in one patient[25] but lacked the statistics of a larger cohort. Vo et al. provided the first exome capture in 3 PCa patients' urinary samples but limit their study on non-vesicular circRNA[20] for biomarkers discovery. Both studies lack key information on the originating tissue RNAs, fundamental for future functional studies. Thus, there is now a need to extend the urine cohort of PCa patients to robustly assess the enrichment of vesicular long exRNA content in comparison with their paired originating tumors, and more importantly to identify putative regulatory lncRNA enrichment within the EVs/ENPs.

Here, we performed the first whole transcriptome analysis of prostate tumor tissues and paired urine-derived EV-enriched ultracentrifugation pellets, called thereafter uEVs, to provide an exhaustive comparison of RNA components between circulating EVs and their tissues of origin from the same patients. We showed that most of the RNAs enriched in the uEVs are intron-free cytoplasmic RNA but also contain previously identified functional circRNAs and specific lncRNAs encoding potential neoantigens. Finally, to orientate subsequent functional studies, we generated and explored the transcriptomes of 3 common PCa cell lines and their corresponding EVs. In conclusion, in this study, most of the urines and EVs collection methodologies (urine collection, storage, RNA extraction and quality controls, see methods) were in lines with the gold standard very recently described by the urine EVs task force[34]. Thus, our study provided an extensive resource for methodologies, tools and datasets to determine the lncRNA content of circulating urinary EVs paving a way for future studies of their role in prostate cancer.

RESULTS

CircRNAs are enriched in uEVs

We performed total RNA extraction from matched prostate tumor FFPE tissue and uEVs from 6 PCa patients, RNA-sequencing and differential expression analysis. Briefly, after prostatic massage on the patients, urines were collected, depleted from living cells, cell debris and large EVs with a series of low-speed centrifugations, followed by ultracentrifugation for 2.5 h, to pellet mostly microvesicles and exosomes with limited presence of ENPs (recovered only after overnight ultracentrifugation[36]). Although we know that these preparations can contain some non-vesicular components[34], we choose to call these pellets uEVs, to simplify reading for the rest of the article (Fig. 1a). The absence of oncosomes and larger vesicles were further suggested by monitoring vesicle size distribution (supplemental text and Extended Data Fig. 1a, 1b). The presence of EVs' markers CD63, CD9 and Syntenin-1 and absence of the cellular endoplasmic reticulum protein Calreticulin confirmed the absence of cellular debris in the preparation (Extended Data Fig. 1c)[4]. Then, RNA from FFPE biopsies and from isolated uEVs were extracted, quantified, and used for total RNA library preparation using

an ultra-low input protocol for RNA-sequencing (see supplemental text and Extended Data Fig. 1d, e, f). The uniquely mapped reads corresponding to Gencode v32 annotation of snoRNA, snRNA, Repeat, Pseudogene, mRNA, lncRNA and circRNA were quantified and compared between uEVs and tissue samples. Quantification of each annotated genomic feature revealed similar RNA content in both sample types but striking difference in distribution for some RNA families (Extended data fig. 1g). Remarkably, mRNAs and pseudogenes count numbers were comparable in both sets respectively with 19,508 in FFPE tissues and 17,599 in uEVs, and 6,655 in FFPE tissues and 4,082 in uEVs, showing that EVs transcriptome are as much as complex as tumoral transcriptome (Extended data fig. 1g, Extended Data Table 2). Furthermore, a large majority of reads came from mRNAs confirming previous results[25]. The specificity of the prostate enriched RNAs was further explored by comparing the uEVs transcriptome with bladder and kidney tissues transcriptome (see supplemental text and Extended Data Figures 2a, b and c) showing robust strictly prostatic RNAs (*SuperExactTest*, $p<10^{-320}$), supporting urine as an excellent source of prostate EVs after prostatic massage. Among the ncRNA, pre-miRNAs, repeats and snRNA gene counts were under-represented in uEVs' transcriptome, whereas circRNAs were over-represented in uEVs with 31,132 versus 14,027 gene counts in tumors. LncRNAs and pseudogenes showed a slight reduction in uEVs with 12,647 gene counts in tumor and 7,763 in uEVs for lncRNA and 6655 in tumors and 4082 in uEVs for pseudogenes. Among circRNAs, 34% of uEVs and 35% of tumor tissues enriched transcripts were found in circBase, with the rest representing novel circRNAs (Extended data fig. 1g). To further confirm this observation, we compared the number and the level of expression of circRNAs and lncRNAs in pooled tumor tissues versus pooled uEVs (Fig.1b). The global correlation between tumor and uEVs indicate a larger disproportion of circRNA in uEV (R^2=0.277) in comparison with lncRNAs mostly enriched in tumor (R^2=0.427). Similar correlations were found in each patient comparing tumors RNA expression and uEVs (Extended Data Fig. 3a) supporting a general tendency for enrichment of circRNA in uEVs observed in previous analyses[25].

We next identified the most differentially overrepresented ncRNAs in uEVs and tumors. Differential expression analyses on the whole transcriptome (extended data fig 3b) and hierarchical clustering by Euclidean distance between pooled uEVs and tumor tissues datasets (extended data fig3c) revealed 4925 enriched RNAs in EVs and 8336 enriched in tumors (Extended Data Table 4). From these differential analyses,

very few or none pre-miRNAs, snoRNAs, snRNAs and repeats were defined as significantly overrepresented in urines. In contrast, a subgroup of circRNAs (N=311), mRNAs (N=4,120), pseudogenes (N=200) and lncRNAs (N=274) were defined as significantly enriched in uEVs compared to tumors (Fig. 1c; Extended Data Fig. 3b). To explore further to which extent each patient' urines contained specific noncoding RNAs, we analyzed the number of normalized counts of lncRNAs and circRNAs by plotting each value in uEVs and tumors and calculated for each linked pair the enrichment value (Fig.1d). Our result showed that most of the differential circRNA/lncRNA were reproducibly differential in all six patients (with a minimum of 88% common circRNAs in patient 5 and 63% lncRNA in patient 4, Fig.1d), pointing to a general rule for the enrichment the 2 families of ncRNA in each specimen.

We conclude that uEVs cargo all sort of RNAs in urines with specific and robust subgroups systematically enriched comparing to respective tumors. Among noncoding RNAs, 2 subgroups of lncRNAs and circRNAs present the most important differential enrichment in vesicular urines.

Mature lncRNAs and mRNA are enriched in uEVs

Next, to further characterize the uEVs RNA content, we compared the genic structures of transcripts in both types of samples, uEVs and tumors, based on read counts across six gencode v32 features: exon, intron, 5'-UTR, 3'-UTR, promoter, and intergenic region in tumors and uEVs specimens (Fig.2a, Extended Data Table 5). We found that intergenic and promoter mapped reads were extremely low in uEVs datasets (less than 1% in uEVs versus 4% in tumors). Although both fractions were sequenced using the same total RNA-seq protocol, intronic counts were very low in uEVs (2.6%) in comparison to tumors (46.3%). This is in striking contrast with 5', 3' untranslated regions and exonic regions all enriched in uEVs (3.2% and 8.3% for 3'UTR, 14.2% and 26.3% for 5'UTR, and 32% to 61.1% in tumors and EVs, respectively). To get a precise and quantitative assessment of the intron loss in uEVs, we then analyzed the density distribution of read counts of exonic/intronic ratios, normalized by the lengths of each feature in lncRNAs and mRNAs (Fig. 2b). The density plot showed that uEVs have a significantly larger proportion of exonic-enriched mRNAs (shift of the Exon-rich peak from log10FC(1)

to log10FC(>2.5)). The lncRNA distribution was less homogenous than for mRNA, probably reflecting the less accurate annotation for lncRNA exon/intron features[37]. Nevertheless, it also showed a shift in the exonic/intronic ratio in uEVs when compared to tumors. We conclude that the higher exon/intron ratios observed in uEVs corresponded to a group of lncRNAs and mRNA harboring essentially mature spliced transcripts. Metagene profiling of reads across the two first exons, the first intron and the last exon and intron of all transcripts with a minimum of 3 exons (N=25,166 genes), revealed higher coverage of exons in uEVs than in tumors and higher coverage of introns in tumors than in uEVs (Fig. 2c). These results suggested that uEVs contain preferentially fully processed mRNAs and lncRNAs, as illustrated for the mRNA GDAP1 and JPH1 (fig.2d), both harbouring intronic reads in tumors and none in uEVs. We hypothesized that uEVs carry mature RNAs, most probably originating from the cytoplasmic compartment, but less or none from the nucleus.

Nuclear lncRNAs are under-represented in uEVs

To directly prove that uEVs are devoid of nuclear transcripts, we analysed public strand-specific RNA-seq data from cytoplasmic and nuclear fractions of 22Rv1 prostate cancer cells[38]. Nuclear and cytoplasmic transcripts were defined by differential expression analysis of nuclear versus cytoplasmic datasets (see method). In total, 4,518 transcripts were considered as highly enriched in nuclear and 6,419 transcripts as enriched in cytoplasm. It is important to note that the public dataset was produced from poly(A) RNA-seq, thus missing circRNAs and non-polyadenylated lncRNAs. Nevertheless, the list of 22Rv1 cytoplasmic and nuclear-enriched RNAs was intersected with those enriched in tumors (n=8336) and uEVs (n=4925) (Fig.3a, Extended data Table 6). The assumed cytoplasmic RNAs represented 58% of uEVs-enriched RNAs, whereas the nuclear transcripts were mostly depleted (representing only 3%) from uEVs (Fig 3b). In contrast, tumors mainly contained transcripts from both compartments (57%) and nuclear (24%) whereas cytoplasmic RNA fraction was of only 15% (Fig 3b). To distinguish what type of transcripts were specifically depleted or enriched in the cytoplasmic and nuclear compartments in uEVs and tumors, we inferred the subcellular localization of mRNAs, pseudogenes and lncRNAs from 22Rv1 datasets. The density distribution of the ratio

cytoplasmic/nuclear reads showed that tumor libraries preferentially contained nuclear lncRNAs and pseudogene transcripts. As expected, the mRNAs were both nuclear and cytoplasmic. On the contrary, in uEVs the majority of enriched lncRNAs, pseudogenes and mRNA were all detected within the cytoplasmic fraction (Fig.3c).

In conclusion, transcriptome analyses of subcellular fractions confirmed that uEVs are enriched in cytoplasmic RNAs, in agreement with the presence of mostly mature intron-less RNAs.

Most uEVs-circRNAs are present in EVs released by prostate cell lines

EVs-RNA have a key role in intercellular communication. While a lot of attention has been focused on miRNA in the past, little is known on the potential functionality of lncRNA and circRNA in this intercellular signaling, raising the need to properly characterize cellular models of EVs cargo. We performed total RNA extraction and sequencing of PC3, DU145, and LNCaP human prostate cancer cell lines and resulting EVs collected from the cell culture media (cEVs). We focused our attention only on circRNAs species as the most abundant type of uEVs-lncRNAs (Extended Data Table 7). Strikingly, the intersection of circRNAs present in patients uEVs, cell lines and cEVs (Fig. 4a and Extended Data Table 7) showed 99% (308) cellular and 99,3% (309) cEVs-circRNAs in common with the 311 uEVs-enriched circRNAs. Hence, enriched-circRNAs from urines of PCa patients can also be detected in *in vitro* model systems, making these cell lines excellent models for functional studies. Among these circRNAs, we found in uEVs the androgen-responsive circ-SMARCA5 presenting an oncogenic activity[39], also detected within the 3 cell lines and in their cEVs. This finding supported the attractive hypothesis that circ-SMARCA5, shown to control cellular proliferation, could also act within the tumor context and the vesicular-mediated signalling. Moreover, we also identified several circRNAs already named “essential-circRNA” required for high proliferation rate in PCa cell lines[40]. Out of these 171 essential circRNAs, 154 (90%) and 165 (96.5%) were detected in PCa cell lines and cEVs, respectively (Fig. 4a and extended data table 7). Among those, 14 were enriched in uEVs as well (Fig. 4b and extended data table 7) with a systematic enrichment in cEVs or uEVs as illustrated for circ-GOLPH3 (Fig 4c). We note that circ-ELK4, the most-enriched circRNAs in uEVs and present in cEVs, has been previously shown to be the most abundant

PCa tumoral circRNA, which expression is negatively regulated with proliferation markers for prostate cancer[20,41]. Furthermore, in uEVs and cEVs, we found circ-ARIAD1A, circ-FAM13B, circ-MAN1A2 and circ-RHOBTB3, also detected by rtqPCR independently in non-vesicular urinary samples of 10 PCa patients[20]. Such overlap between different studies additionally reinforced the quality and the robustness of our work. We conclude that PC3, DU145 and LNCaP PCa cell lines and associated cEVs could be used as excellent cellular models for functional studies of circulating uEV-circRNA.

uEVS and cEVs are a source of lncRNAs that can be translated into neoantigens

Although lncRNAs are originally defined as unable to code for proteins, some were reported to contain ORF that encode functional peptides[42]. Here, we investigated the coding potential of uEV-enriched lncRNAs, with a specific focus on neoantigenic properties, in comparison with those enriched in tumor tissues. First, to avoid any ambiguity with existing ORFs or proteins, all lncRNA transcripts overlapping annotated protein-coding genes or pseudogenes from Gencode V32 were filtered out from the list, to obtain 1,404 and 228 lncRNAs, respectively in tumor and uEVs (Fig. 5a). We then defined all potential coding sequences starting from AUG codons and finishing with a STOP codon, in the 3 reading frames. For the genes transcribed into several isoforms, the most prevalent was considered for further analysis (see method). We retrieved all peptide sequences over 8 aminoacids, resulting in 11,707 tumors and 862 uEVs lncRNA-encoded peptides, with some isoforms being the source of several peptides. Neoantigens were then predicted with NetMHCpan-4, using MHC typing computed by seq2HLA from the RNA-seq file of each patient[43]. To best separate ligand from non-ligand peptides, only neopeptides with an elution ligand index lower than 0.5 were kept, corresponding to 15,677 (5941 unique) for tumors and 768 (351 unique) for uEVs, derived from 244 and 15 transcripts, respectively (see method, Extended Data Table 8, Fig. 5a and Extended Data Fig. 4). From the top 2% predictions, we estimated that the majority of potential lncRNA-derived neoantigens were 9-mer peptides (Fig. 5b), the gold standard size for neoantigen recognition[44]. This result showed that both EVs-lncRNAs and tumor lncRNA have similar capacities (and proportion) to encode high affinity binding to MHC. In conclusion, among the uEVs-enriched lncRNAs, we identified 15 transcripts (called EVs-neoLncRNA

hereafter) that could potentially serve as templates for the translation of 768 peptides with potent neoantigens properties, if internalized and translated into the host cells.

To define the translation capacities of the 15 uEVs-neolncRNA, we first quantified their expression in PC3, DU145 and LNCaP PCa cell lines and corresponding cEVs. As shown in the heatmap, all the 15 EVs-neolncRNA are expressed, at least in one of the 3 cell lines, at lower but comparable levels as the POLR2A mRNA used here as a moderate expressed control (Fig 5c). We then re-analyzed a public ribosome profiling dataset for the PC3 cell line, to determine ribosome occupancy on the 15 EVs-neolncRNAs[45]. We observed that 5, out of 11 neolncRNAs expressed in PC3, were associated with actively translated ribosomes (red names in Fig5c). As illustrated, the POLR2A mRNA contained similar levels (Fig.5d) of Ribosome Protected Fragments (RPFs) as the predicted ZNF503-AS2 EVs-neoLncRNA encoding a predicted high affinity peptide of 9 amino acids (Fig.5e, in red).

In conclusion, our results showed that uEVs can cargo lncRNAs containing ORFs with a high potential to be translated. *In vitro* cellular models can be used as a source of high-affinity neoantigens, also exported as RNA templates by EVs. Among the 3 tested PCa cell lines, androgen responsive LNCaP showed all the neoLncRNAs being expressed and looked like the best model for functional studies.

DISCUSSION

We performed here the first systematic comparison of whole RNA sequencing libraries from paired urinary EV-enriched pellets and prostate tumor FFPE samples. Our results show that uEV pellets are enriched in intron-free, processed cytoplasmic RNAs. Remarkably, this extracellular material is specifically enriched in some lncRNAs and circRNAs. Our whole transcriptome analyses confirmed the presence of these lncRNAs in EV pellets obtained from PCa cell lines and hence the utility of *in vitro* systems for functional molecular and cellular studies. Recent works suggest that lncRNA are a largely ignored source of neoantigens[46] and our results point out several of those as circulating templates for neoantigen production within a recipient cell. Historically, a challenge with cell-based phenotypic screens is the difficulty in gaining molecular insight that

are shared with real tumor expressing tissues. Our results in the present study, however, demonstrate in 6 paired tumors and urinary EVs, the robustness of such candidates that can be found in *in vitro* models. The data generated herein should thus be an important resource to further investigate loss or gain of function of the prostate circulating RNAs using cellular models and develop genetic tools to address EVs functional signaling pathways. The power of genome-scale CRISPR/Cas-9 or Cas-13 loss-of-function and gain of-function screens will provide mechanistic clues for such circulating RNAs.

It is conceivable that some of these lncRNAs and circRNAs can be brought in clinics as novel therapeutic strategies to limit communication with the immune response system or the oncogenic transformation of adjacent cells. EVs secreted by PCa cells can alter the transcription of infiltrating T cells[47] and their RNA content might be part of their signaling activity controlling key elements in the recipient hosting cells driving their transcription.

We identified here a subset of circRNAs that were previously defined and named “essential”-circRNA to increase proliferation of prostate cancer cells[40]. One such oncogenic mechanism could be transferred from tumor to adjacent cells and tissues to initiate perturbation. Such perturbations could provoke uncontrolled expanding proliferation in proximal cells, thus overwhelming antitumoral regulators while being beneficial to the initial tumor in a competitive environment. Since we found that those circRNAs are enriched in circulating EVs, and present in *in vitro* EVs, it is tempting to further address their role in intercellular communication using epithelial recipient cells in contact with such EVs containing or not these essential circRNAs. In contrast, one of the most abundant EV-circRNA, circ-ELK4 has been shown to negatively correlate with the expression of cell cycle progression genes in prostate tumors[20] similarly to other circRNA found to be anticorrelated with proliferative genes in ovarian normal and cancer cell lines[48]. This observation raised the exciting hypothesis that such amount of circulating vesicular circ-ELK4 could impact proximal normal tissues by contributing to cell cycle arrest and perturbing again the competitive environment of the tumor. An alternative scenario would follow previous hypothesis where acquisition of metastatic traits by tumors could be achieved by eliminating a negative regulator such as miR23 through exosomes[49]. Hence, similarly, the disposal of the cell cycle arrest factor circ-ELK4 in the uEVs, would promote proliferation of the PCa EV-secreting cells.

For lncRNA, in addition to their well-known role in controlling epigenetic landscape, we propose here that the traveling lncRNA characterized in this study might be more susceptible for translation since being originally found in the cytoplasm and more efficiently processed. We provide a first list of lncRNA candidates that can be translated into high affinity neoantigens. EV-RNAs can in principle serve as a source of novel proteins in recipient cells, since mRNAs transported by EVs can be actively translated into the recipient cells[50,51] even only as short as 1 hour after EV uptake during coculture cells[52]. Thus, one may hypothesize these circulating neo-lncRNAs could also be translated in the host cells and act as a decoy to the immune system. An attractive scenario is that some tumoral cells not only could silence their own HLA system using epigenetic regulation as recently shown[53], as some sort of invisibility blanket, but would transfer information and tumor specific neoantigens lncRNA templates to non-tumoral adjacent cells. With their fully operational antigen presenting machinery, these cells would then divert the anti-tumoral immune response away from the true tumor.

In addition to their regulatory roles, circRNAs are receiving considerable attention as potential liquid biopsy biomarkers (reviewed in[54,55]) which are more stable than their linear mRNAs isoforms[56] (48h versus 20h), and are released to the cytoplasm during mitosis, where they show extraordinary stability. Here we showed that circRNAs are an important fraction of urinary vesicular-enriched lncRNA, independently of their parental RNA enrichment. Indeed, by comparing the list of the 4,614 enriched RNAs in urines (excluding circRNAs) with the 311 enriched circRNAs, we observed that only 130 corresponded to parental linear RNAs leaving 61% enriched circRNAs without a corresponding enrichment of their parental RNA. This suggest that the mechanism of circRNA enrichment in uEVs could be independent to their parental linear RNA expression and instead may be due to the regulation of splicing mechanisms or to the circRNA lifetime itself. Further experiments will determine if specific circRNA processing are linked to their externalization.

Several causes explain RNA release into the extracellular stream, such as (1) cell death induced by stresses like hypoxia, (2) inflammation, (3) anti-tumor therapies, (4) tumor invasion and (5) metastasis process, all resulting in the increase of circRNA or lncRNA concentration in urines. Altogether, these might be the reason for the disproportion between circRNA and some lncRNA levels in urines vs tumors. EV heterogeneity and

subfamily classification are currently the target of large research efforts[4]. Our work did not address the specific contents of the different members of the EV family, nor did we formally demonstrate that the RNA analyzed here were specifically incorporated inside EVs: we cannot exclude that some of the sequences RNA may be non-specifically associated to EVs secondarily after EV secretion. Further works properly distinguishing EVs from non-vesicular circulating factors by size exclusion chromatography for instance[57], will address whether some of these extracellular entities contain distinct types of noncoding RNAs. Nevertheless, Our results come as further support of considering these noncoding RNAs as surrogate circulating markers and regulatory elements for prostate cancer progression[58].

MATERIAL AND METHODS

Sample collection and processing

Urine and paired Formalin-Fixed Paraffin-Embedded (FFPE) tissue samples were collected by Henri Mondor hospital, Créteil, France from 6 prostate cancer patients with written consent and approval by ethical Comité de Protection des Personnes (CPP) Ile-de-France V, N° ID-RCB:2016-A00789-42. All patients were newly diagnosed, had not received treatment for prostate cancer before biospecimen collection.

Following digital rectal exam (DRE) performed by the attending oncologist, first-catch urine samples were collected in 50 mL Falcon tubes without adding protease, nuclease inhibitors. Urines were placed at 4°C, centrifuged within 72 h of collection at 2000 g for 12 min followed by 3500 g for 17 min at 4°C to remove cell debris and large EVs while preserving the small and medium-sized EVs in the sample. A total of 30 mL of supernatant was used for EVs enrichment by ultracentrifugation at 160,000 g for 2.5 h at 4°C using a SW32-Ti swing bucket rotor in an Optima L-80 XP ultracentrifuge (Beckman Coulter, USA). The resultant pellets were resuspended in 100 µL cold PBS and then directly stored at -80°C for EVs characterization or vortexed 1 min with 700 µL of Qiazol (Qiagen) then keep 5 min at room temperature before storage at -80°C, for RNA extraction.

Absence of bacteria in urine supernatant was confirmed by inoculation on Lennox Broth media plates.

The prostate biopsies were fixed with 10% neutral buffered formalin solution (Sigma-Aldrich, Merck) at room temperature for 8 h but no longer than 24 h. Tissues were dehydrated through ascending grades of alcohol and cleared in xylene and embedded into paraffin blocks. Biopsy specimens were stored at room temperature at Henri Mondor hospital.

PC3, DU145 and LNCaP human prostate cancer cell lines, were obtained from American Type Culture Collection (Manassas, VA, USA), and cultured, in triplicate, in RPMI 1640 media supplemented with 10% Fetal

Bovine Serum (FBS), GlutaMAX (Gibco-Invitrogen Corporation, Carlsbad, CA, USA). Cells were maintained at 37°C in a humidified atmosphere containing 5% CO2. When cells reached 70% confluence, they were washed twice in PBS and cultured for an additional 48h reaching a maximum of 80-90% confluence, in the same media composition without FBS, to eliminate vesicles coming from the serum. LNCaP growth media were supplemented in 2nM dihydrotestosterone. Conditioned medium was then processed for EV enrichment as described for urine samples. Briefly, medium was centrifuged at 2000 g for 12 min followed by 3500 g for 17 min at 4°C and an ultracentrifugation at 160,000 g for 2.5 h at 4°C using a SW32-Ti swing bucket rotor in an Optima L-80 XP ultracentrifuge (Beckman Coulter, USA). The resultant pellets were resuspended in 100 µL cold PBS and then directly stored at -80°C for EVs characterization or vortexed 1 min with 700 µL of Qiazol (Qiagen) then keep 5 min at room temperature before storage at -80°C, for RNA extraction.

Extracellular vesicles characterization

Characterization, particle number and size, of the EVs-enriched pellets was performed using nanoparticle tracking analysis (NTA) using a ZetaView PMX-120 video microscope (Particle Metrix) equipped with a 488 nm laser and the software Zeta View version 8.05.10. For optimal measurements, samples were diluted with PBS until particle concentration was within the optimal concentration range for particle analysis. Experiments were performed briefly as follows: the instrument was set at 25°C, sensitivity of 70 and shutter of 75. Measurements were done at 11 different positions (5 cycles per position) and frame rate of 30 frames per second. 55 seconds videos were recorded with a resolution of 0.714µm/px). Then, the software tracks the brownian motion of individual vesicles, visualized by light scattering, and calculates their size and total concentration with corresponding standard error.

Western blot to measure proteins content in EVs and cellular extracts.

Cell lysates for Western blot were obtained by incubating HEK293 cell pellets at a concentration of 1×10^6 cells in 25 uL of lysis buffer (50 mM Tris, pH 7.5, 0.15 M NaCl, 1% Triton X-100) with 2% complete protease inhibitor (Roche) for 20 min on ice, followed by a 18,516×*g* centrifugation for 15 min at 4 °C to recover the

supernatant. Both cell lysates and EV pellets from 5ml urines were mixed with Laemmli sample buffer (BioRad), without reducing agent. After boiling for 5 min at 95 °C, samples were loaded on a 4–15% Mini-protean TGX stain-free gels (BioRad). Total proteins were imaged from the stain-free gels with the ChemiDoc Touch Imager (BioRad). Transfer was performed on Immuno-Blot PVDF membranes (BioRad), with the Trans-Blot Turbo Transfer System (BioRad) during 7 min. Blocking was performed during 30 min with Blocking Reagent (Roche) in TBS 0.1% Tween. Primary antibodies were incubated overnight at 4 °C and secondary antibodies during 1 h at room temperature (RT). Development was performed using either the Clarity Western ECL Substrate (BioRad) or the Immobilon Forte Western HRP substrate (Millipore), and the ChemiDoc Touch Imager (BioRad). Intensity of the bands was quantified using ImageJ.

Antibodies and reagents

Primary antibodies for Western blot were mouse anti-human CD63 (BD Bioscience, clone H5C6, 1/1000), -human CD9 (Millipore, clone MM2/57, 1/1000), rabbit anti-human syntenin (Abcam, EPR8102, 1/2000). Mouse anti human Calreticulin (Abcam, clone FMC 75, 1/1000).

Secondary antibodies HRP-conjugated goat anti-rabbit IgG (H + L) and HRP conjugated goat anti-mouse IgG (H + L) were purchased from Jackson Immuno-Research and used 1/10000.

RNA extraction

Extraction of total RNA including miRNA from urinary and cell samples was performed by using the miRNeasy micro kit (Qiagen) with modified on-column digestion of DNA recommended by Qiagen for less than 1 µg of RNA. The Modified procedure consists of reapplying to the membrane the flow-through after digestion and washing and to prepare RWT buffer with isopropanol instead of ethanol. RNA was eluted with 14 µL nuclease free water.

RNA from FFPE tissues was purified using RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE (Life technologies) from 10 slices of 5 µm. RNA concentration has been evaluated by with Qubit RNA HS Assay Kit

(Life technologies). The extracted RNA was checked for size distribution on the Agilent Bioanalyzer using RNA 6000 Pico kit (Agilent Technologies, Santa Clara, CA).

The Extracted RNA was checked for concentration on Qubit fluorometer using Qubit RNA HS Assay Kit (Thermo Fisher Scientific) and for size distribution on the Agilent Bioanalyzer using RNA 6000 Pico kit (Agilent Technologies, Santa Clara, CA). RNA from EVs and cells were checked for contaminants on NanoDrop spectrophotometer (Thermo Fisher Scientific).

Preparation of libraries and Illumina sequencing

For each sample, 10 ng of total RNA was used to construct a strand-specific library using SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian kit which incorporates a technology that enables removal of ribosomal cDNA following cDNA synthesis (Takara Bio, Europe). The libraries were sequenced on HiSeq 2500 instrument (NovaSeq2 sequencing system) from Next-Generation Sequencing platform of the Institut Curie as 2 x 100 nucleotides paired-end reads to obtain about 50 million reads per sample.

Bioinformatics analysis

- RNA-seq reads quality control and alignment for FFPE and uEVs:

The Curie bioinformatic platform performed the quality control of the RNA-seq data using MultiQC, and the alignement of the reads on the human genome hg38 was done using STAR 2.6.1a, with the following parameters : --outMultimapperOrder Random --outSAMtype BAM Unsorted --outSAMattributes All --outSAMprimaryFlagOneBestScore --outSAMmultNmax 1 --outFilterTypeBySJout --outFilterMultimapNmax 20 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1[37][38].

- Differential expression analysis of FFPE and uEVs:

Read counting was performed for each sample on the human gene annotation gencode v32 (https://www.gencodegenes.org/human/release_32.html) and on the human repeats (http://www.repeatmasker.org/species/hg.html), using Kallisto 0.46.1 with the parameter --rf-stranded[39]. The Kallisto index was built with the extracted sequences from the human genome hg38, using the annotations and the getfasta command of BEDTools 2.29 (for gencode 32, before the extraction, the exons of all the transcripts of each gene were merged by location, in order to obtain directly the counts at the gene level)[40].

The tool CIRIquant 1.1 was used on the samples with default parameters to discover the location of circRNAs and to quantify them [59]. The results (counts from human genes, repeats and circRNAs) were concatenated. The conditions FFPE and uEVs were compared using DESeq2, with the following parameters: betaPrior=FALSE, independentFiltering=F, cooksCutoff=F. Only the features with adjusted p-value <= 0.05,abs(log2FoldChange)>= 0.585, and normalized counts >= 20 in at least one sample were retained as differentially expressed. Heatmaps of expression were obained using the R package ComplexHeatmap from bioconductor.

- Read count genomic distribution:

We created from the human gene annotation gencode32 (https://www.gencodegenes.org/human/release_32.html) 6 classes of genomic features (exon, intron, 5'-UTR, 3'-UTR, promoter, intergenic), as follows: The exons from all the transcripts of each gene were merged by location, in order to have non-redundant segments of exons, using the merge command of BEDTools 2.29. Introns were inferred from these exons using the R package GenomicFeatures from bioconductor. The 5' and 3' UTRs from gencode32 were merged and were subtracted from the previous features. Promoters were inferred by using a distance of 1kb upstream of the first exon (when possible), in respect of the strand. Intergenic parts were constructed after the concatenation of all the previous features, using the complement

command of bedtools. Strand-specific counting from the alignment files was performed on these features using featureCounts of the Subread package (https://sourceforge.net/projects/subread/files/)[42]. The priority order when the counts were on several features was the following : exon > UTRs > intron > promoter > intergenic.

- Exon-intron read count ratio:

The exons from all the transcripts of each gene of the human gene annotation gencode32 (https://www.gencodegenes.org/human/release_32.html) were merged by location using the merge command of BEDTools 2.29, in ordrer to have non-redundant segments of exons. Introns were inferred from these exons using the R package GenomicFeatures from bioconductor. Genes with no introns were discarded.

The alignment files were converted to BEDPE format and were intersected with the new formed annotation to determine their status (exonic or intronic), using the count command of BEDTools 2.29. The number of exonic and intronic counts for each gene was then determined, using the same command. The counts were normalized using the length of each feature, and the ratio exonic count over intronic count was computed (genes that had 0 count on numerator and denominator before calculation were discarded. When just a part is equal to 0, +1 is added to both parts to avoid division by 0). The average ratio for each gene was computed for each condition, and the results were plotted using the R package ggplot2.

- Metagenes:

To select the exons and introns features, we have created firstly metatrancripts as follows:

The exons from all the transcripts of each gene were merged by location using the merge command of BEDTools 2.29, to have non-redundant segments of exons, and these exons were numbered; Introns were inferred from these exons.

The first exon, first intron, second exon, last intron and last exon were selected for each gene having at least 1 raw read count (read count from Kallisto).

The alignment files were converted in BigWig files with RPM normalization (reads per million of mapped reads), using UCSC tools (http://hgdownload.cse.ucsc.edu/admin/exe/linux.x86_64/), and for each part and each gene, the read coverage was extracted.

The obtained signal was scaled on 100 positions using the R base approx function (allows to compute interpolations), with the following parameters: method="linear",ties="ordered".

The average, minimum, and maximum values were computed at each of the 100 positions, and the result was plotted using the R package ggplot2.

- Neoantigen analysis :

HLA allotypes from FFPE and uEVs samples were determined using the tool seq2HLA. Upregulated lncRNAs from FFPE &uEVs were selected, and a research of the ORFs (start codon = AUG) of the most abundant transcript for each gene was performed using TransDecoder (https://transdecoder.github.io/) and the R bioconductor package ORFik https://github.com/Roleren/ORFik. The translation into peptides from the 3 frames was performed using the SeqinR R package[60], and only the ones with a length >= 8 amino acids were kept. The list of peptides for FFPE and uEVs was given to the tool netMHCpan-4.1 in order to find neoantigens. Only the peptides with an elution ligand rank <= 2 were kept for downstream analysis.

-Ribosome profiling analysis:

Raw data were extracted from[45]. The Ribosome protect Fragments were defined by counting reads on the defined putative ORF of the 15 Evs-lncRNA containing, encoding high affinity binding scores neoantigens, and on the ORF of POLR2A, normalized with the size (TPM).

Figure legends

Fig.1. uEVs are enriched in circRNAs and some lncRNAs. a. Full transcriptome of paired liquid and solid biopsies of prostate cancer patients. Experimental procedure from prostate tumor biopsies and urines collections to RNA sequencing, through FFPE biopsies and uEVs isolation. **b.** Mean gene expression of paired Tumor against uEVs (n=83,980 RNAs). 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. R^2=0.2770 and R^2=0.4267 for 38.793 circRNAs and 13,704 lncRNAs respectively. **c**. Density plot showing the distribution of log2 fold change uEVs/Tumor ratio per gene types. The right side of dotted line correspond to enriched genes in uEVs compared to Tumor tissues. The left side of dotted line correspond to the enriched genes in Tumors compared to uEVs. Each color represents an RNA type. Number of total differentially enriched genes in tumor and uEVs are indicated on the right with the number of enriched RNA in EVs in bracket. **d.** Interaction between the normalized number of counts of the 311 upregulated circRNAs (top) and 274 upregulated lncRNAs (bottom) in paired Tumor and uEVs for each patient (P1 to P6). Each dot corresponds to a circRNA or lncRNA upregulated in uEVs compared to Tumor with Log2 FC=0,5 to 1 (blue), Log2 FC<0,5 (grey), Log2 FC>1 (red). The number of counts for the same RNA, in Tumor and in uEV are linked together with a line as shown for 4 circRNAs in the zoom window of patient 6. For each patient are indicated the numbers of circRNA and lncRNA with a log2(FC)>1.

Fig.2. intronic RNAs are depleted in uEVs compared to Tumor. **a**. Genomic read counts distribution by percentage across exon, Intron, 3'UTR, 5'UTR, intergenic and promoter. **b**. Distribution of log10(exonic read counts/intronic read counts) normalized by length from Tumor (blue) and uEVs (red) samples for lncRNA (top) and mRNA (bottom) annotations. **c**. Metagene of mean coverage for 2 first exons, last exon, first intron and last intron of 25,166 mRNAs and lncRNAs from Tumor (blue) and uEVs (red) samples. **d**. GGbio-generated RNA reads profiling along minus (−; pink) and plus (+; bleu) strands of chr8:74198516-74398516 in Tumor and uEVs specimens. Arrow lines represent introns and rectangles represent exons of GENCODE-annotated protein-coding gene JPH1 (pink) and part of GDAP1. The maximum value of coverage, read count is shown in the left panel of read mapping. Some intronic reads are indicated for the 2 genes.

Fig.3. Depleted lncRNAs in uEVs are nuclear. **a.** Experimental procedure, starting from 22Rv1 cell line fractionation polyA RNA-seq, to propose cytoplasmic or nuclear localization of up-regulated genes in uEVs and up regulated genes in Tumor biopsies. **b.** Stacked barplot distribution, by percentage, of cytoplasmic (blue), nuclear (red), both (yellow) or non-polyA RNAs (grey) of up-regulated genes in Tumor (8,336) and up regulated genes in uEVs (4,925). 15.4% and 57.8% upregulated RNAs respectively in tumors and uEVs are cytoplasmic; 23.6% and 2.9% are nuclear; 57.2% and 32% are both. **c.** Density distribution of log2 (fold change cytoplasmic/nuclear ratio) per RNA types (7,732 RNAs from Tumor and 4,556 RNAs from uEVs), 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.

Fig. 4. uEVs-enriched circRNAs contain essential circRNAs and are common to PCa cell lines EVs. a. Venn diagram showing number of over-represented circRNA in prostate cancer uEVs (n=311), 171 essential circRNAs defined by Chen et al., circRNAs expressed in PC3, LNCaP and DU145 PCa cell lines (cell circRNAs, n=25832) and circRNA expressed in cell EVs (cEVs circRNAs, n=58320). **b**. List of the 14 essential circRNAs up-regulated in uEVs **c.** Sequencing read coverage from back splicing of GOLPH3 circRNA, from chromosome 5:32124716-32174319, is shown using GGplot2 in Tumor, uEVs, PC3, LNCaP and DU145 cEVs and in cells. The maximum value of coverage read count is shown in the left panel of read mapping. Parental transcript ENST000000265070.7 is schematized by blue rectangles representing exons and black arrow lines representing introns (shrunk to 100 nt). Junction of back splicing is indicated in light blue.

Fig. 5. uEVs-enriched RNAs show lncRNA-encoding neoantigens shared by prostate cell lines. a. Workflow of ORFs, peptides and neoantigens prediction from lncRNAs enriched in Tumor (n=1,993) and uEVs (n=274). **b.** Number and lengths of strong predicted neoantigens in Tumor (blue, n=15677) and uEVs (red, n=768). **c**. Heatmap representing the relative expression by log10(TPM+1), of POLR2A mRNA and the 15 uEVs-neoLncRNA, encoding the 351 strong unique neoantigens within uEVs, Tumor, PC3-, LNCaP- and DU145-cells

and their respective EVs. RPF from PC3 ribosome profiling dataset. **d.** Example of EV-neoLncRNA. IGV-generated public PC3 prostate cancer cell line ribosome occupancy[45] and uEVs RNA-seq profiling along plus (+, blue) and minus (−, pink) strands of POLR2A mRNA and **e.** of ZNF503-AS2 EV-neoLncRNA. Blue arrow-lines and rectangles represent introns and exons of metatranscripts, respectively. RPM, reads per million mapped reads. Open reading frames (ORFs, red rectangles), starting from AUG codon, of the most abundant transcripts for POLR2A and ZNF503-AS2, are indicated. The sequence of one ZNF503-AS2 ORF from the frame 1, generated with GGplot2, is presented and the sequences of the 26 translated neopeptides are underlined. Red sequence is the strongest 9-mere.

AVAILABILITY AND IMPLEMENTATION

The bioinformatic codes, intermediate data and Extended Data Tables are available at: https://github.com/MorillonLab/Prostate_6_tumor_FFPE_vs_6_tumor_uEVs_analysis. The raw data are available under GEO accession GSE183070.

ACKNOWLEDGEMENTS

We deeply thank all the members of our labs for discussions and critical reading of the manuscript. Daniel Gautheret for his constant encouragement and critical readings. Pascale Soyeux for technical help on FFPE manipulation. The platform of biological resources (PRB) of Henri Mondor hospital for tissues collection. We also would like to thank Benoît Albaud, Sonia Lameiras, Patricia Legoix-Né, Virginie Raynal, and Sylvain Baulande from the Next Generation Sequencing (NGS) platform of Institut Curie for discussions and expertise. RNA-seq sequencing efforts were supported by the grant from the ICGex program of Institut Curie attributed to A.M. Marc Gabriel and Matthieu Lejars and this project were supported by the European Research Council (875532-PROSTATOR-ERC-2019-PoC), CARNOT-2018-PROSTATOR, 17 MECENAT, attributed to A.M. Fondation ARC (PGA1 RF20180206962) to CT. This work was also supported by the Ligue contre le cancer 94 / Val-de-Marne (grant #A.P. 2020) and the Groupement d'Entreprises Françaises dans la Lutte contre le Cancer, GEFLUC (grant #A.P. 2020).

Authors Contributions

- Anna Almeida: supervision, validation, investigation, methodology, project administration, funding and article writing—original draft, review, and editing.
- Marc Gabriel: resources, data curation, software, formal analysis, validation, investigation, and visualization.
- Virginie Firlej: resources, methodologies and data curation.

- Lorena Martin-Jaular: formal analysis, methodology, and investigation.
- Mathieu Lejars: methodology, investigation.
- Rocco Cippola: Ressources, investigation.
- Nicolas Vogt: Investigation.
- Damien Destouches: Resources
- Francis Vacherot: Resources
- Alexandre De La Taille: Resources and data curation.
- Clotilde Théry: supervision, methodology, validation, formal analysis
- Antonin Morillon: conceptualization, supervision, funding acquisition, project administration, and writing—original draft, review, and editing.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

REFERENCES

1 Tkach, M. & Thery, C. Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. *Cell* **164**, 1226-1232, doi:10.1016/j.cell.2016.01.043 (2016).

2 Mathieu, M., Martin-Jaular, L., Lavieu, G. & Thery, C. Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. *Nature cell biology* **21**, 9-17, doi:10.1038/s41556-018-0250-9 (2019).

3 van Niel, G., D'Angelo, G. & Raposo, G. Shedding light on the cell biology of extracellular vesicles. *Nature reviews. Molecular cell biology* **19**, 213-228, doi:10.1038/nrm.2017.125 (2018).

4 Thery, C. *et al.* Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles* **7**, 1535750, doi:10.1080/20013078.2018.1535750 (2018).

5 Hoshino, A. *et al.* Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers. *Cell* **182**, 1044-1061 e1018, doi:10.1016/j.cell.2020.07.009 (2020).

6 Jeppesen, D. K. *et al.* Reassessment of Exosome Composition. *Cell* **177**, 428-445 e418, doi:10.1016/j.cell.2019.02.029 (2019).

7 Becker, A. *et al.* Extracellular Vesicles in Cancer: Cell-to-Cell Mediators of Metastasis. *Cancer cell* **30**, 836-848, doi:10.1016/j.ccell.2016.10.009 (2016).

8 Kalluri, R. & LeBleu, V. S. The biology, function, and biomedical applications of exosomes. *Science* **367**, doi:10.1126/science.aau6977 (2020).

9 Pelissier Vatter, F. A. *et al.* Extracellular vesicle- and particle-mediated communication shapes innate and adaptive immune responses. *The Journal of experimental medicine* **218**, doi:10.1084/jem.20202579 (2021).

10 Spinelli, C., Tawil, N., Adnani, L., Rak, J. & Choi, D. Extracellular Vesicle Mediated Vascular Pathology in Glioblastoma. *Subcell Biochem* **97**, 247-273, doi:10.1007/978-3-030-67171-6_10 (2021).

11 Wolfers, J. *et al.* Tumor-derived exosomes are a source of shared tumor rejection antigens for CTL cross-priming. *Nature medicine* **7**, 297-303, doi:10.1038/85438 (2001).

12 Chaput, N. *et al.* Exosomes as potent cell-free peptide-based vaccine. II. Exosomes in CpG adjuvants efficiently prime naive Tc1 lymphocytes leading to tumor rejection. *J Immunol* **172**, 2137-2146, doi:10.4049/jimmunol.172.4.2137 (2004).

13 LeBleu, V. S. & Kalluri, R. Exosomes as a Multicomponent Biomarker Platform in Cancer. *Trends Cancer* **6**, 767-774, doi:10.1016/j.trecan.2020.03.007 (2020).

14 Wiklander, O. P. B., Brennan, M. A., Lotvall, J., Breakefield, X. O. & El Andaloussi, S. Advances in therapeutic applications of extracellular vesicles. *Science translational medicine* **11**, doi:10.1126/scitranslmed.aav8521 (2019).

15 Reiner, A. T. *et al.* Concise Review: Developing Best-Practice Models for the Therapeutic Use of Extracellular Vesicles. *Stem Cells Transl Med* **6**, 1730-1739, doi:10.1002/sctm.17-0055 (2017).

16 Buschmann, D. *et al.* Evaluation of serum extracellular vesicle isolation methods for profiling miRNAs by next-generation sequencing. *J Extracell Vesicles* **7**, 1481321, doi:10.1080/20013078.2018.1481321 (2018).

17 Murillo, O. D. *et al.* exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. *Cell* **177**, 463-477 e415, doi:10.1016/j.cell.2019.02.018 (2019).

18 Tosar, J. P. *et al.* Fragmentation of extracellular ribosomes and tRNAs shapes the extracellular RNAome. *Nucleic Acids Res* **48**, 12874-12888, doi:10.1093/nar/gkaa674 (2020).

19 Hulstaert, E. *et al.* Charting Extracellular Transcriptomes in The Human Biofluid RNA Atlas. *Cell reports* **33**, 108552, doi:10.1016/j.celrep.2020.108552 (2020).

20 Vo, J. N. *et al.* The Landscape of Circular RNA in Cancer. *Cell* **176**, 869-881 e813, doi:10.1016/j.cell.2018.12.021 (2019).
21 Miranda, K. C. *et al.* Massively parallel sequencing of human urinary exosome/microvesicle RNA reveals a predominance of non-coding RNA. *PloS one* **9**, e96094, doi:10.1371/journal.pone.0096094 (2014).
22 Amorim, M. G. *et al.* A total transcriptome profiling method for plasma-derived extracellular vesicles: applications for liquid biopsies. *Scientific reports* **7**, 14395, doi:10.1038/s41598-017-14264-5 (2017).
23 Galvanin, A. *et al.* Diversity and heterogeneity of extracellular RNA in human plasma. *Biochimie* **164**, 22-36, doi:10.1016/j.biochi.2019.05.011 (2019).
24 Giraldez, M. D. *et al.* Phospho-RNA-seq: a modified small RNA-seq method that reveals circulating mRNA and lncRNA fragments as potential biomarkers in human plasma. *EMBO J* **38**, doi:10.15252/embj.2019101695 (2019).
25 Everaert, C. *et al.* Performance assessment of total RNA sequencing of human biofluids and extracellular vesicles. *Scientific reports* **9**, 17574, doi:10.1038/s41598-019-53892-x (2019).
26 Li, Y. *et al.* Extracellular Vesicles Long RNA Sequencing Reveals Abundant mRNA, circRNA, and lncRNA in Human Blood as Potential Biomarkers for Cancer Diagnosis. *Clinical chemistry* **65**, 798-808, doi:10.1373/clinchem.2018.301291 (2019).
27 Barreiro, K. *et al.* Comparison of urinary extracellular vesicle isolation methods for transcriptomic biomarker research in diabetic kidney disease. *J Extracell Vesicles* **10**, e12038, doi:10.1002/jev2.12038 (2020).
28 Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81**, 145-166, doi:10.1146/annurev-biochem-051410-092902 (2012).
29 Andjus, S., Morillon, A. & Wery, M. From Yeast to Mammals, the Nonsense-Mediated mRNA Decay as a Master Regulator of Long Non-Coding RNAs Functional Trajectory. *non-Coding RNA* **3**, 44-53, doi:10.3390/ncrna7030044 (2021).
30 Chen, L. L. The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nature reviews. Molecular cell biology* **21**, 475-490, doi:10.1038/s41580-020-0243-y (2020).
31 Kristensen, L. S. *et al.* The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* **20**, 675-691, doi:10.1038/s41576-019-0158-7 (2019).
32 Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384-388, doi:10.1038/nature11993 (2013).
33 Okholm, T. L. H. *et al.* Transcriptome-wide profiles of circular RNA and RNA-binding protein interactions reveal effects on circular RNA biogenesis and cancer pathway expression. *Genome medicine* **12**, 112, doi:10.1186/s13073-020-00812-8 (2020).
34 Erdbrugger, U. *et al.* Urinary extracellular vesicles: A position paper by the Urine Task Force of the International Society for Extracellular Vesicles. *J Extracell Vesicles* **10**, e12093, doi:10.1002/jev2.12093 (2021).
35 Nilsson, J. *et al.* Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer. *British journal of cancer* **100**, 1603-1607, doi:10.1038/sj.bjc.6605058 (2009).
36 Zhang, Q. *et al.* Transfer of Functional Cargo in Exomeres. *Cell reports* **27**, 940-954 e946, doi:10.1016/j.celrep.2019.01.009 (2019).
37 Hezroni, H. *et al.* Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell reports* **11**, 1110-1122, doi:10.1016/j.celrep.2015.04.023 (2015).
38 Kaul, T., Morales, M. E., Sartor, A. O., Belancio, V. P. & Deininger, P. Comparative analysis on the expression of L1 loci using various RNA-Seq preparations. *Mob DNA* **11**, 2, doi:10.1186/s13100-019-0194-z (2020).
39 Kong, Z. *et al.* Androgen-responsive circular RNA circSMARCA5 is up-regulated and promotes cell proliferation in prostate cancer. *Biochemical and biophysical research communications* **493**, 1217-1223, doi:10.1016/j.bbrc.2017.07.162 (2017).
40 Chen, S. *et al.* Widespread and Functional RNA Circularization in Localized Prostate Cancer. *Cell* **176**, 831-843 e822, doi:10.1016/j.cell.2019.01.025 (2019).

41 Cuzick, J. *et al.* Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *The Lancet. Oncology* **12**, 245-255, doi:10.1016/S1470-2045(10)70295-3 (2011).

42 Chen, J. *et al.* Pervasive functional translation of noncanonical human open reading frames. *Science* **367**, 1140-1146, doi:10.1126/science.aay0262 (2020).

43 Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* **199**, 3360-3368, doi:10.4049/jimmunol.1700893 (2017).

44 Dash, P. *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**, 89-93, doi:10.1038/nature22383 (2017).

45 Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55-61, doi:10.1038/nature10912 (2012).

46 Laumont, C. M. *et al.* Noncoding regions are the main source of targetable tumor-specific antigens. *Science translational medicine* **10**, doi:10.1126/scitranslmed.aau5516 (2018).

47 Chen, S. *et al.* Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. *Nature cell biology* **23**, 87-98, doi:10.1038/s41556-020-00613-6 (2021).

48 Bachmayr-Heyda, A. *et al.* Correlation of circular RNA abundance with proliferation--exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Scientific reports* **5**, 8057, doi:10.1038/srep08057 (2015).

49 Ostenfeld, M. S. *et al.* Cellular disposal of miR23b by RAB27-dependent exosome release is linked to acquisition of metastatic properties. *Cancer Res* **74**, 5758-5771, doi:10.1158/0008-5472.CAN-13-3512 (2014).

50 Skog, J. *et al.* Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nature cell biology* **10**, 1470-1476, doi:10.1038/ncb1800 (2008).

51 Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature cell biology* **9**, 654-659, doi:10.1038/ncb1596 (2007).

52 Lai, C. P. *et al.* Visualization and tracking of tumour extracellular vesicle delivery and RNA translation using multiplexed reporters. *Nat Commun* **6**, 7029, doi:10.1038/ncomms8029 (2015).

53 Griffin, G. K. *et al.* Epigenetic silencing by SETDB1 suppresses tumour intrinsic immunogenicity. *Nature* **595**, 309-314, doi:10.1038/s41586-021-03520-4 (2021).

54 Li, S. & Han, L. Circular RNAs as promising biomarkers in cancer: detection, function, and beyond. *Genome medicine* **11**, 15, doi:10.1186/s13073-019-0629-7 (2019).

55 Bach, D. H., Lee, S. K. & Sood, A. K. Circular RNAs in Cancer. *Molecular therapy. Nucleic acids* **16**, 118-129, doi:10.1016/j.omtn.2019.02.005 (2019).

56 Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *Rna* **19**, 141-157, doi:10.1261/rna.035667.112 (2013).

57 Royo, F., Thery, C., Falcon-Perez, J. M., Nieuwland, R. & Witwer, K. W. Methods for Separation and Characterization of Extracellular Vesicles: Results of a Worldwide Survey Performed by the ISEV Rigor and Standardization Subcommittee. *Cells* **9**, doi:10.3390/cells9091955 (2020).

58 Tucker, D., Zheng, W., Zhang, D. H. & Dong, X. Circular RNA and its potential as prostate cancer biomarkers. *World journal of clinical oncology* **11**, 563-572, doi:10.5306/wjco.v11.i8.563 (2020).

59 Zhang, J., Chen, S., Yang, J. & Zhao, F. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat Commun* **11**, 90, doi:10.1038/s41467-019-13840-9 (2020).

60 Charif, D. & J., L. *SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis*. 207-232 (Springer Verlag, 2007).

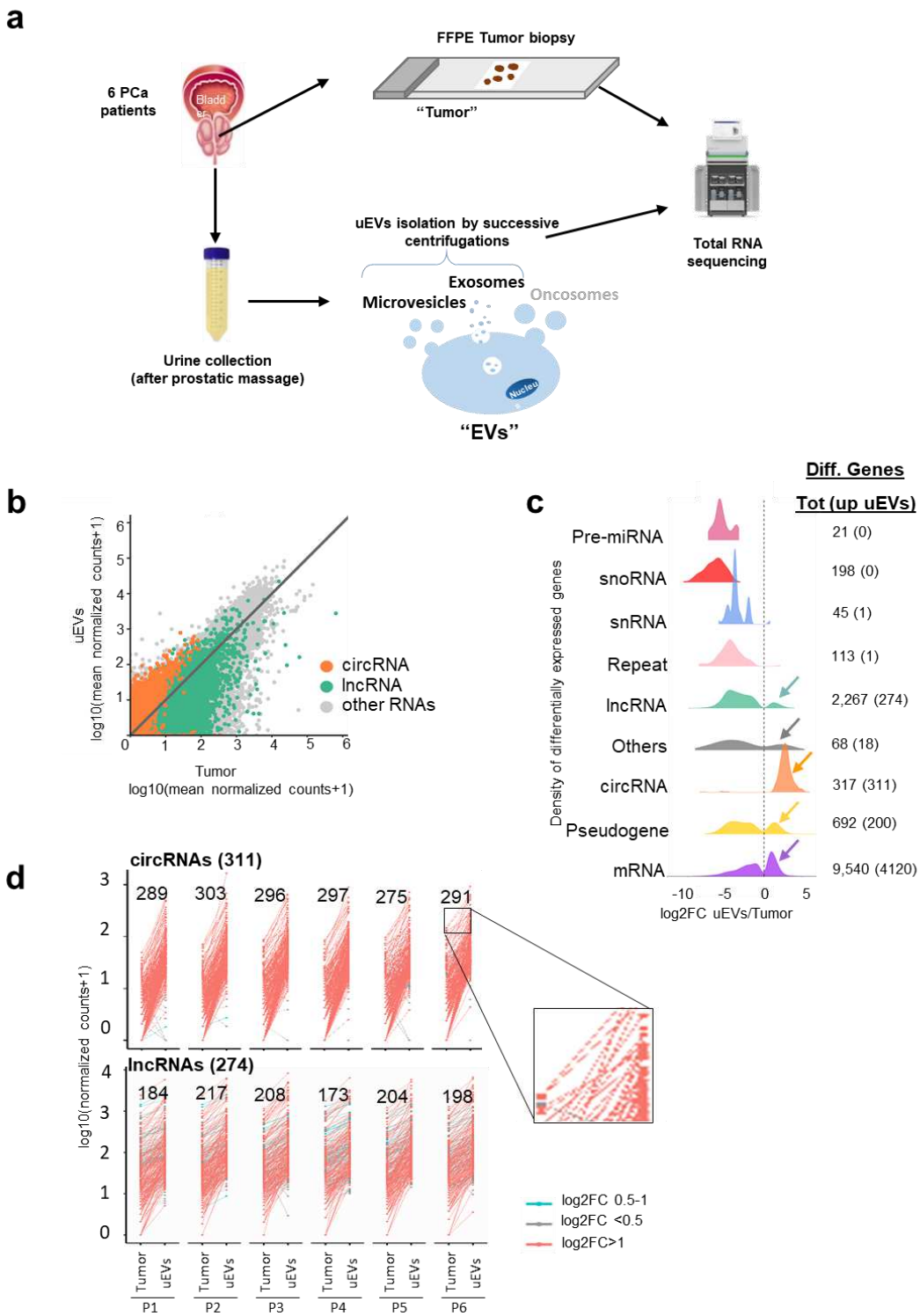


Fig.1. uEVs are enriched in circRNAs and some lncRNAs. a. Full transcriptome of paired liquid and solid biopsies of prostate cancer patients. Experimental procedure from prostate tumor biopsies and urines collections to RNA sequencing, through FFPE biopsies and uEVs isolation. **b.** Mean gene expression of paired Tumor against uEVs (n=83,980 RNAs). 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. R^2=0.2770 and R^2=0.4267 for 38.793 circRNAs and 13,704 lncRNAs respectively. **c**. Density plot showing the distribution of log2 fold change uEVs/Tumor ratio per gene types. The right side of dotted line correspond to enriched genes in uEVs compared to Tumor tissues. The left side of dotted line correspond to the enriched genes in Tumors compared to uEVs. Each color represents an RNA type. Number of total differentially enriched genes in tumor and uEVs are indicated on the right with the number of enriched RNA in EVs in bracket. **d.** Interaction between the normalized number of counts of the 311 upregulated circRNAs (top) and 274 upregulated lncRNAs (bottom) in paired Tumor and uEVs for each patient (P1 to P6). Each dot corresponds to a circRNA or lncRNA upregulated in uEVs compared to Tumor with Log2 FC=0,5 to 1 (blue), Log2 FC<0,5 (grey), Log2 FC>1 (red). The number of counts for the same RNA, in Tumor and in uEV are linked together with a line as shown for 4 circRNAs in the zoom window of patient 6. For each patient are indicated the numbers of circRNA and lncRNA with a log2(FC)>1.

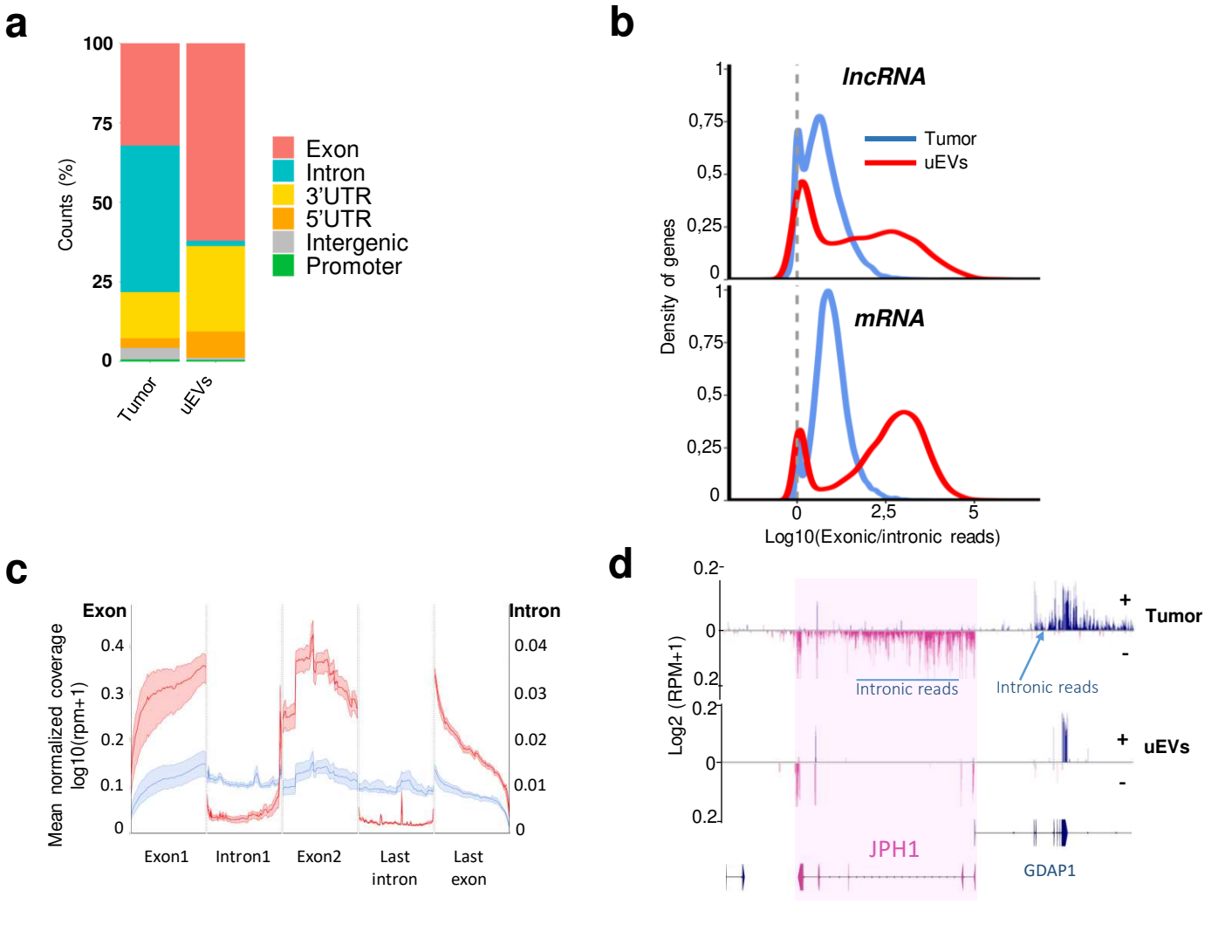


Fig.2. intronic RNAs are depleted in uEVs compared to Tumor. **a**. Genomic read counts distribution by percentage across exon, Intron, 3'UTR, 5'UTR, intergenic and promoter. **b**. Distribution of log10(exonic read counts/intronic read counts) normalized by length from Tumor (blue) and uEVs (red) samples for lncRNA (top) and mRNA (bottom) annotations. **c**. Metagene of mean coverage for 2 first exons, last exon, first intron and last intron of 25,166 mRNAs and lncRNAs from Tumor (blue) and uEVs (red) samples. **d**. GGbio-generated RNA reads profiling along minus (−; pink) and plus (+; bleu) strands of chr8:74198516-74398516 in Tumor and uEVs specimens. Arrow lines represent introns and rectangles represent exons of GENCODE-annotated protein-coding gene JPH1 (pink) and part of GDAP1. The maximum value of coverage, read count is shown in the left panel of read mapping. Some intronic reads are indicated for the 2 genes.

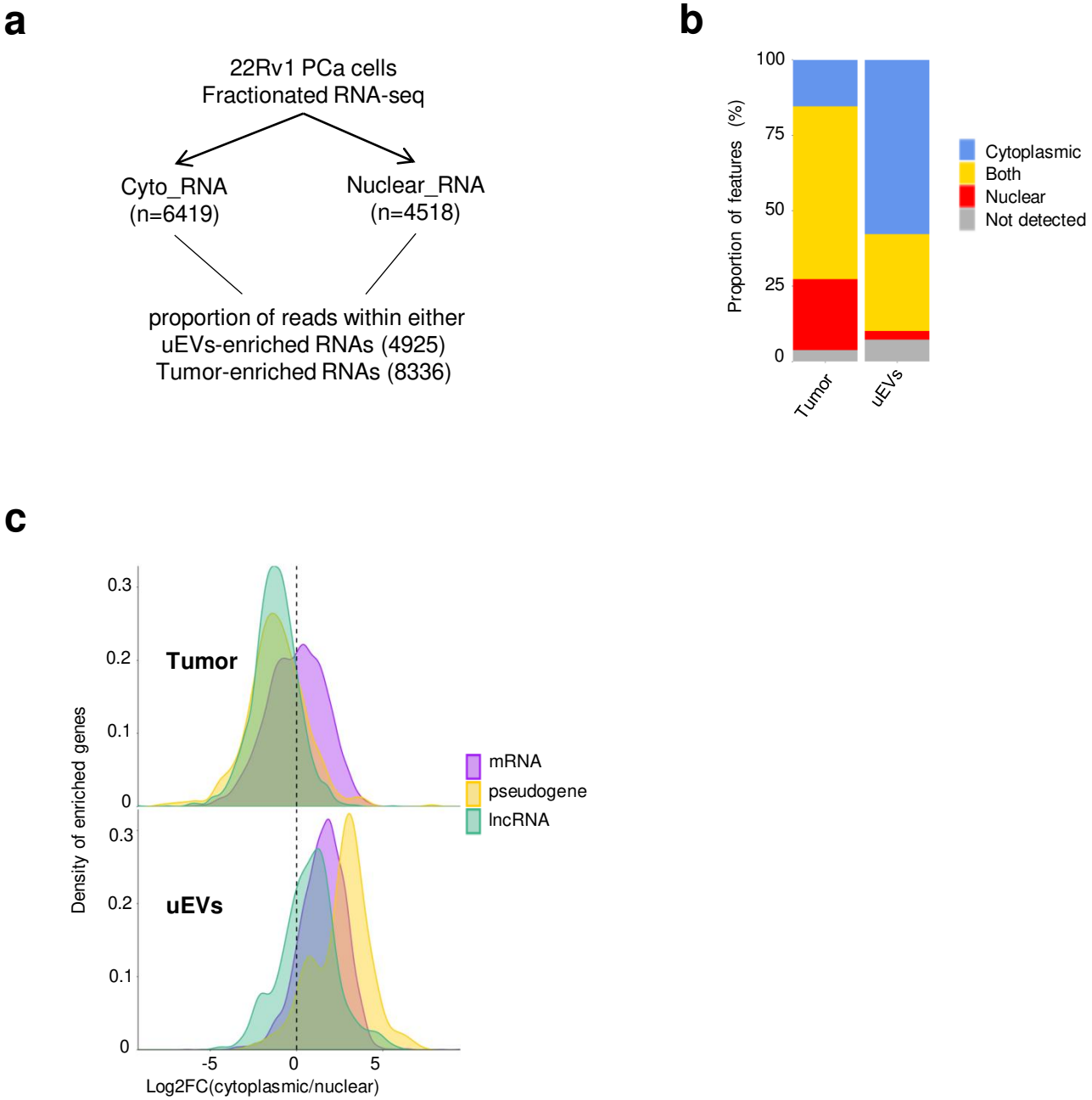
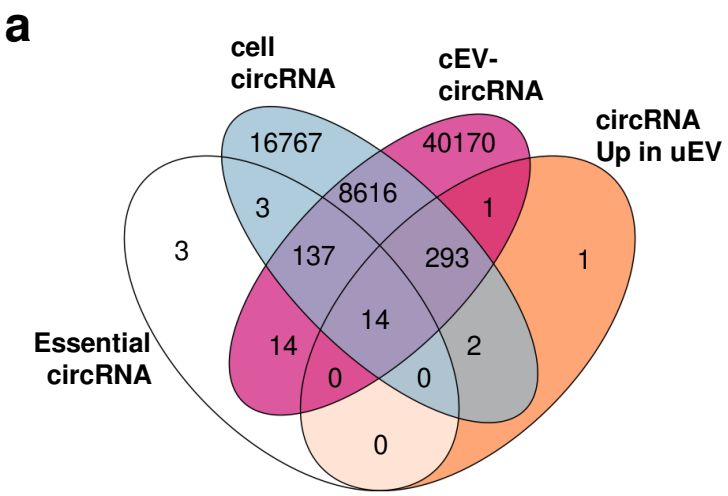


Fig.3. Depleted lncRNAs in uEVs are nuclear. **a.** Experimental procedure, starting from 22Rv1 cell line fractionation polyA RNA-seq, to propose cytoplasmic or nuclear localization of up-regulated genes in uEVs and up regulated genes in Tumor biopsies. **b.** Stacked barplot distribution, by percentage, of cytoplasmic (blue), nuclear (red), both (yellow) or non-polyA RNAs (grey) of up-regulated genes in Tumor (8,336) and up regulated genes in uEVs (4,925). 15.4% and 57.8% upregulated RNAs respectively in tumors and uEVs are cytoplasmic; 23.6% and 2.9% are nuclear; 57.2% and 32% are both. **c.** Density distribution of log2 (fold change cytoplasmic/nuclear ratio) per RNA types (7,732 RNAs from Tumor and 4,556 RNAs from uEVs), 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.



b

circRNA name	Strand	Genomic location
circVAMP3	+	chr1:7777160-7778169
circLZIC	-	chr1:9931891-9934860
circRPS6KC1	+	chr1:213077696-213117410
circBMPR2	+	chr2:202464809-202467689
circZNF148	-	chr3:125313308-125331238
circSLC4A4	+	chr4:71236576-71255399
circGOLPH3	-	chr5:32135572-32143880
circEMC2	+	chr8:108449823-108455930
circOSTCP8,UBAP2	-	chr9:33960826-33989126
circPRUNE2	-	chr9:76629192-76636557
circRAB6A	-	chr11:73707420-73720899
circCLNS1A	-	chr11:77624963-77625818
circRMDN3	-	chr15:40744047-40745259
circRN7SL233P, SNORA73,MIB1	+	chr18:21765772-21819646

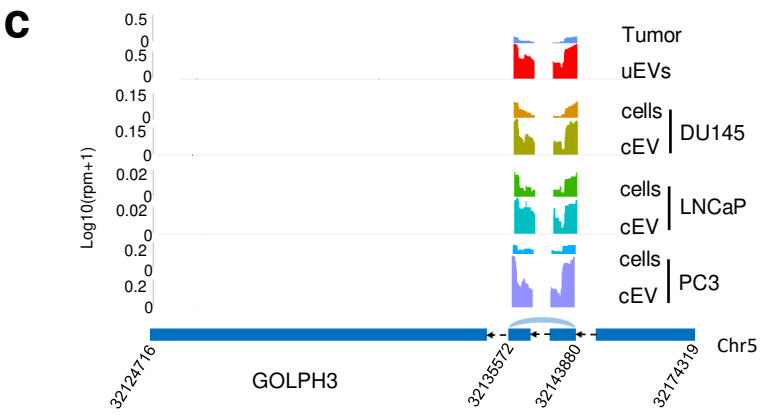


Fig. 4. uEVs-enriched circRNAs contain essential circRNAs and are common to PCa cell lines EVs. a. Venn diagram showing number of over-represented circRNA in prostate cancer uEVs (n=311), 171 essential circRNAs defined by Chen et al., circRNAs expressed in PC3, LNCaP and DU145 PCa cell lines (cell circRNAs, n=25832) and circRNA expressed in cell EVs (cEVs circRNAs, n=58320). **b.** List of the 14 essential circRNAs up-regulated in uEVs **c.** Sequencing read coverage from back splicing of GOLPH3 circRNA, from chromosome 5:32124716-32174319, is shown using GGplot2 in Tumor, uEVs, PC3, LNCaP and DU145 cEVs and in cells. The maximum value of coverage read count is shown in the left panel of read mapping. Parental transcript ENST000000265070.7 is schematized by blue rectangles representing exons and black arrow lines representing introns (shrunk to 100 nt). Junction of back splicing is indicated in light blue.

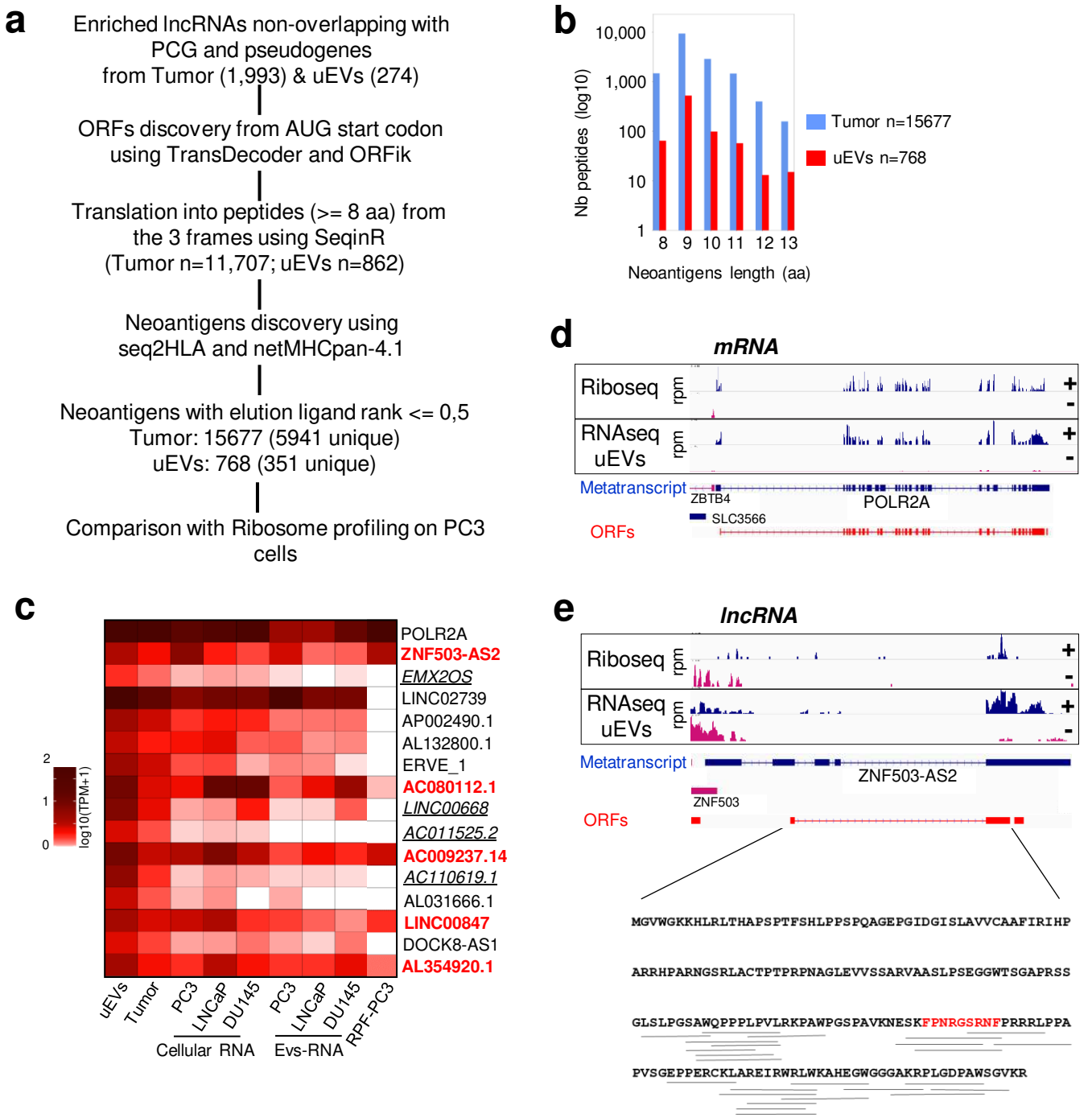


Fig. 5. uEVs-enriched RNAs show lncRNA-encoding neoantigens shared by prostate cell lines. a. Workflow of ORFs, peptides and neoantigens prediction from lncRNAs enriched in Tumor (n=1,993) and uEVs (n=274). **b.** Number and lengths of strong predicted neoantigens in Tumor (blue, n=15677) and uEVs (red, n=768). **c.** Heatmap representing the relative expression by log10(TPM+1), of POLR2A mRNA and the 15 uEVs-neoLncRNA, encoding the 351 strong unique neoantigens within uEVs, Tumor, PC3-, LNCaP- and DU145-cells and their respective EVs. RPF from PC3 ribosome profiling dataset. **d.** Example of EV-neoLncRNA. IGV-generated public PC3 prostate cancer cell line ribosome occupancy{Hsieh, 2012 #9115} and uEVs RNA-seq profiling along plus (+, blue) and minus (−, pink) strands of POLR2A mRNA and **e.** of ZNF503-AS2 EV-neoLncRNA. Blue arrow-lines and rectangles represent introns and exons of metatranscripts, respectively. RPM, reads per million mapped reads. Open reading frames (ORFs, red rectangles), starting from AUG codon, of the most abundant transcripts for POLR2A and ZNF503-AS2, are indicated. The sequence of one ZNF503-AS2 ORF from the frame 1, generated with GGplot2, is presented and the sequences of the 26 translated neopeptides are underlined. Red sequence is the strongest 9-mere.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AlmeidaSupplTextFig.pdf