Urine extracellular vesicles capture kidney transcriptome and hyperglycemia linked mRNA signatures for type 1 diabetic kidney disease

Om Prakash Dwivedi^{1*}, Karina Barreiro^{1,2}, Annemari Käräjämäki³, Erkka Valo^{4,5,6}, Rashmi B Prasad⁷, Harry Holthöfer^{1,8}, Per-Henrik Groop^{4,5,6,9}, Tiinamaija Tuomi^{1,4,7,10}, Carol Forsblom^{1,4,5}, Leif Groop^{1,7}, Maija Puhka^{1,2*}

1. Institute for Molecular Medicine Finland FIMM, HiLIFE, University of Helsinki, Finland.
2. Institute for Molecular Medicine Finland FIMM, EV Core, University of Helsinki, Finland.
3. Department of Primary Health Care, Vaasa Central Hospital, Hietalahdenkatu 2-4, 65130 Vaasa, Finland & Diabetes Center, Vaasa Health Care Center, Sepänkyläntie 14-16, 65100 Vaasa, Finland.
4. Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland.
5. Department of Nephrology, University of Helsinki and Helsinki University Hospital, Finland.
6. Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Finland.
7. Lund University Diabetes Centre, Department of Clinical Sciences Malmö, Lund University, Skåne University Hospital, Malmö, SE-20513, Sweden.
8. Department of Medicine, University Medical Center, Hamburg-Eppendorf, Hamburg, Germany.
9. Department of Diabetes, Central Clinical School Monash University Melbourne VIC Australia.
10. Endocrinology, Abdominal Centre, Helsinki University Hospital, Helsinki, Finland.

*Correspondence to Om Prakash Dwivedi (om.dwivedi@helsinki.fi) and Maija Puhka (maija.puhka@helsinki.fi)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Abstract

Diabetic kidney disease (DKD) is a severe complication of type 1 diabetes (T1D), which lacks non-invasive early biomarkers. Although less explored, mRNAs in urinary extracellular vesicles (uEV) could reflect changes in the kidney transcriptome during DKD development. We performed genome-wide mRNA sequencing of >100 uEV samples from two T1D cohorts with 24-hour and overnight urine collections. Our uEV pipeline allowed reproducible detection of >10,000 mRNAs bearing overall similarity to kidney transcriptome. uEV from T1D DKD groups showed significant upregulation of 13 genes, prevalently expressed by proximal tubular cells within the kidney. Strikingly, six genes involved in cellular stress responses including protection against oxidative stress (*GPX3*, *NOX4*, *MSRB*, *MSRA*, *HRSP12* and *CRYAB*) correlated with hyperglycemia and long-term changes in kidney function independent of albuminuria status. The study identified genes associated with glycemic stress in T1D DKD and confirmed the utility of uEV in capturing pathological gene expression signatures from kidney.

Introduction

Diabetes is a multifactorial disease with diverse complications causing a global health burden and >1.3 million annual deaths (1). One of the most severe complications is diabetic kidney disease (DKD) - a leading cause of end-stage renal disease (2). At tissue level, DKD is characterized by histological features such as glomerular basement membrane thickening, mesangial matrix expansion, nodular glomerulosclerosis, and arteriolar hyalinosis – more often detected in type 1 diabetes (T1D) compared to type 2 diabetes (T2D) (2-4). Kidney biopsy is considered as the gold standard to observe the DKD linked structural changes and progression stages (5). However, in practice, the diagnosis is usually based on albuminuria and decline in kidney function measured as glomerular filtration rate (6, 7). Despite their wide clinical applications, these traditional markers are non-specific and poorly detect early stages of DKD, for example hyperfiltration (8).

Transcriptomics of the kidney tissue has suggested association of kidney mRNA markers with DKD related structural changes like fibrosis (9, 10), and also led to development of sensitive biomarkers like epidermal growth factor (11). However, the invasiveness and rarity of biopsies has limited kidney transcriptome-based approaches in DKD precision medicine (12). This has turned the interest to extracellular RNAs in urine, which might mirror or provide novel insight into the pathological changes of kidney transcriptome and an alternative non-invasive way for diagnosing, prognostication and monitoring DKD (13).

Extracellular vesicles (EV) are lipid bilayer contained particles secreted by cells in varying sizes (~30-1000 nm). They carry active biomolecules (proteins, lipids, DNA and RNA) and contribute to many biological processes in health and disease e.g. through their role in cell-to-cell signaling (14-16). EV are found in all body fluids, where their membrane helps to protect the cargo from proteases and RNases (17). Thus, EV provide a potentially stable biomarker reservoir that reflects the status of the cell of origin (18).

Due to the rich small RNA content of urinary EV (uEV), DKD research has focused mainly on uEV miRNAs (19-22). In contrast, although uEV seem to carry a significant part of the cell free mRNA, studies on mRNA are rare, particularly using non-targeted genome-wide next-generation sequencing (NGS) in any significant number of urine samples (23-28). Reasons include missing standards for urine collections and EV isolation methods (29), and the need for optimized low-input NGS protocols. Recently, we successfully tested such methods for uEV mRNA profiling (28), but information of their reproducibility and applicability in larger cohorts of T1D DKD patients is still lacking.

In this study, we utilized low input mRNA sequencing (mRNAseq) for uEV and analyzed the whole uEV mRNA transcriptome in >100 samples including T1D individuals with varying stages of DKD (i.e. normo-, micro- or macroalbuminuria). We show 1) technical reproducibility of the uEV mRNA sequencing pipeline and feasibility in clinical setting by combining cohorts with different urine collection protocols, 2) comprehensive characterization of uEV mRNA transcriptomes with respect to other human tissues suggesting its similarity with kidney, and 3) identification of novel candidate marker genes that may have potential role in maintaining cellular stress response in kidney during DKD pathogenesis in T1D individuals.

Results

Study design

The uEV mRNAseq study of 102 samples focused on three main aims: 1) to assess the feasibility and reproducibility of low input uEV mRNAseq for different types of clinical urine collections, 2) to assess how uEV capture gene expression signatures of the kidney or other tissues, and 3) to discover novel candidate genes for kidney disease in individuals with T1D (**Fig. 1**). The tested technical variables were: 1) overnight (ON) vs. 24-hour (24-h) urine collection (n=6 donors, 12 pairs), 2) urine samples processed with or without centrifugation prior to freezing (n=4 donors, 4

pairs), and 3) technical replicates of urine passing through the whole pipeline from uEV isolation to mRNAseq at 1-5 months intervals (n=8 donors, 6 duplicates and 2 triplicates) (**Fig. 1**- technical comparisons). To find novel DKD candidate genes, we performed mRNAseq for uEV samples from 72 T1D individuals collected from two different T1D cohorts (FinnDiane, Finnish Diabetic Nephropathy Study; n=34 and DIREVA, Diabetes Registry in Vaasa; n=38) with different urine collection types (24-h and ON urine collection, respectively) (**Fig. 1**). The included samples differed regarding the degree of urine albuminuria (normoalbuminuria, n=38; microalbuminuria, n=15 and macroalbuminuria, n=19) as well as estimated glomerular filtration rate (eGFR), HbA_{1c} (glycated hemoglobin A1c) and diabetes duration (table S1, S2, S3). We also utilized the long-term (15 years) clinical laboratory data (eGFR, HbA_{1c}) for association analysis. As the study included only men to reduce heterogeneity (e.g. due to differences in the genitourinary system), we assessed the impact of sex on the expression of all robustly detected uEV mRNAs by clustering analysis (women, n=3, men, n=3 or 67) (**Fig. 1**).

Additionally, we combined the mRNA expression profiles of uEV and various tissues obtained from the Genotype-Tissue Expression (GTEx) project (30, 31) (17 tissues, n=7,466 tissue samples) to assess the similarity of uEV mRNAs to mRNAs from kidney cortex and other tissues (**Fig. 1**, tissue map).

Quality control of uEV

Urinary EV isolates obtained by ultracentrifugation from the different study groups produced acceptable uEV and RNA quality, as shown by electron microscopy (EM), Western blotting, nanoparticle tracking analysis (NTA) and EV-RNA profiling by Bioanalyzer Pico assay (**Fig. 2**). We detected typical round EVs showing different staining intensities, sizes and morphologies by EM (**Fig. 2, A and B**). The isolates contained also some Tamm-Horsfall protein filaments. The EVs were roughly up to 1 μ m in diameter, although the main population consisted of smaller (exosomal

range) sizes (30-150 nm). NTA indicated a similar size range of particles (**Fig. 2C**). Western blotting showed variable quantities of two uEV-enriched proteins, CD9 and podocalyxin, in the study samples (**Fig. 2D**). Despite the variable EV marker quantities, sufficient amounts of total RNA could be extracted from all EV samples (mean 22 ng, range 2-140 ng). Bioanalyzer showed typical RNA profiles for the uEV from all study groups: a relatively large amount of small RNA and no or minor peaks of rRNA (**Fig. 2E**). We have already previously found the uEV quality from this ultracentrifugation protocol to be good after a thorough characterization of a set of 10 samples comprising 24-h urines from healthy controls and individuals with T1D (28)- these samples were included in both this study and in Barreiro et al. (28).

Quality, robustness and reproducibility of uEV mRNA profiling by NGS

We performed NGS of uEV mRNAs using low RNA input protocol for both T1D cohorts (total=72, 24-h or ON urine collection) and three different technical comparisons (**Fig. 1**). In the T1D cohorts, we observed a median (25% - 75% percentile) of 11.5 million (7.4-16.3) raw sequencing reads (fig. S1A). However, the read distributions differed between the cohorts - the ON tended to yield higher read numbers than the 24-h collections. Most raw sequencing reads (median% [25%-75% percentile], 90.4% [87.6-91.9]) mapped uniquely to the human genome with <7% of unmapped or multi-mapped reads (4.6% [3.7-5.9] and 4.7% [3.7-6.5], respectively) (fig. S2). Furthermore, as expected, most uniquely mapped reads aligned to exonic genomic regions (73.6% [72.3-74.6] for CDS exons, 14.5% [13.7-15.7] for 3'UTR exons and 11.6% [10.9-12.8] for 5'UTR exons), and <0.1% mapped to intronic or intergenic regions (10 Kb upstream or downstream of coding regions) (fig. S3).

We detected robust expression of 10,596 protein coding genes in uEV from the T1D cohorts that were expressed in >90% of samples (with count ≥ 1). The distribution of gene expression of all the genes appeared similar for both T1D cohorts despite the differences in urine collection protocols (**Fig. 3A**). To further test for differences in uEV mRNA profiles, we performed principal component

analysis (PCA) of all (n=10,596) robustly expressed genes. We observed no systematic differences as 67 out of 72 samples clustered together (**Fig. 3B**). However, the five outliers were removed from downstream case-control analysis.

133

We further tested the robustness of uEV mRNA profiles using pairs of samples for three different technical variables, 1) urine collection protocol: 24-h vs. ON, 2) effect of centrifuging the urine samples prior to freezing ($\pm$ centrifugation), and 3) reproducibility by processing aliquoted urine samples through the whole uEV mRNAseq pipeline - from uEV isolation to sequencing - at different time points (replicates) (**Fig. 1** and table S4). In these three technical comparisons, similar to the main T1D study cohorts, we observed (median-IQR) 11.5 (8.6-13.5) million raw sequencing reads that were mostly [91.6% (88.2-93.9)] uniquely mapped to human genome, particularly to the exonic gene regions compared to intronic or non-coding regions (fig. S1B, S4A and S4B). We detected the expression of 10,449 protein coding genes (with count ≥ 1) present in $\geq 90\%$ of samples used for the three technical comparisons. The read count data for these commonly expressed genes (10,449) showed clustering (based on correlation) for most technical replicates (fig. S4C). We performed hierarchical clustering of the sample pairs based on correlation of gene expression using all common detected protein coding genes (n=10,449) or “kidney enriched genes” (n=233, present in uEV out of 413) – a subset of protein coding genes known to be enriched in kidney (see Methods) and expressed in our technical uEV sample sets. Pairs of replicate urine samples and those with or without centrifugation showed high correlation (r=0.81 to 0.95) (fig. S5) and most of these technical pairs clustered together both when all (**Fig. 3, C, E and G**) or when kidney enriched genes were considered (**Fig. 3, D, F and H**). All 24-h and ON pairs did not cluster together but no systematic difference was observed for the urine sample pairs either from T1D individuals or non-diabetic individuals regarding all commonly expressed genes (**Fig. 3, I and K**) or kidney enriched genes (**Fig. 3, J and L**). Finally, we also tested sex-dependent differences in uEV mRNA profiles by hierarchical clustering of three

samples from women with either three samples from men matched for the stage of albuminuria (n=3) or including all the men from the T1D cohorts (n=67). Men and women clustered separately based on global gene expression (n=10,596 genes) and sex chromosome genes (n=329), but not based on kidney enriched genes (n=247 present in uEV out of 413, see methods) (fig. S6 and S7).

Together, these results showed the expression of >10,000 protein coding genes in uEV from different clinical cohorts. It is noteworthy that this is the first study that successfully combined a relatively large number of samples from different urine collections highlighting the comparability and reproducibility of the uEV mRNA profile and the used technologies.

Similarity of the uEV mRNA profile with kidney transcriptome

Gene expression profiles of uEV could be derived from several tissues – thus, to study the origin of the uEV mRNAs, we compared the uEV transcriptome to individual organ -specific gene expression signatures. We obtained sets of known “kidney enriched genes” (413 genes with higher expression in kidney than in other tissues, see the Methods) and “kidney depleted genes” (4,631 genes not detected in kidney) from the human protein atlas database (32, 33), and checked their expression in T1D uEV (with the total of n=10,596 genes). We detected the expression of 59.8% (247 out of 413) of kidney enriched genes compared to only 4.3% (201 out of 4,631) of kidney depleted genes in the T1D uEV. We further looked for more “specific kidney enriched genes”- a subset of kidney enriched genes that are reported to be detected in kidney and only in some human tissues (see Methods)- and detected the expression of 40.5% of such genes (77 out of 190) (**Fig. 4A**).

We next compared the expression profile of uEV mRNAs with the profiles of 17 other human tissues obtained from the GTEx database (30, 31) (all male, **Fig. 4**). To infer proximity between the profiles, we performed PCA analysis using expression levels of three sets of genes; 1) all protein coding genes expressed in both uEV and the 17 reference human tissues (n=10,350), 2) a subset of “kidney depleted genes” as negative control and 3) a subset of “specific kidney enriched genes” (n=77, as shown in **Fig. 4A**). PCA analysis using gene expression data of all protein coding genes showed that uEV samples clustered together with the kidney cortex tissue samples and away from other tissues suggesting close similarity of the mRNA expression patterns between uEV and kidney (**Fig. 4B**). The PCA analysis of kidney depleted genes showed proximity of prostate samples with uEV samples (**Fig. 4C**). We observed a similar pattern, clustering of kidney cortex and uEV samples, when using specific kidney enriched genes in the PCA analysis (**Fig. 4D**).

In summary, we detected expression of a high number of kidney enriched genes in uEV and generally a lack of expression of genes not known to be expressed in kidney. The overall expression pattern of protein coding genes in uEV also showed close similarity with kidney suggesting that kidney is the main contributor of mRNAs detected in uEV.

Genome-wide uEV mRNA analysis revealed novel candidate markers for T1D DKD

We performed hypothesis-free global differential gene expression analysis using mRNA expression profiles of uEV (n=10,596 genes) and comparing macroalbuminuric and normoalbuminuric T1D individuals (n=17 and n=37, respectively, mRNAseq PCA quality passed) (table S2 and **Fig. 5**). The analysis revealed differential expression of 13 genes (*MAP7*, *MSRB1*, *GPX3*, *IL32*, *NOX4*, *HRSP12*, *TINAG*, *CAPN3*, *CXCL14*, *MSRA*, *CRYAB*, *RBP5* and *TMEM9*) with robust significance ($P \leq 3.85 \times 10^{-6}$) after accounting for multiple correction of all genes (**Fig. 5A**). Interestingly, the expression level of all the differentially expressed genes was upregulated in the uEV of macroalbuminuric T1D individuals (**Fig. 5, A and B**).

Having shown the difference between normo- and macroalbuminuric individuals, we then included also the microalbuminuric group (n=13, mRNAseq PCA quality passed) to study expression of the identified 13 genes. The gene expression tended to increase from normo- to micro- to macroalbuminuria (**Fig. 5C**). While the difference between normo- and microalbuminuric individuals was not statistically significant for all the genes, the combined micro- and macroalbuminuria group differed significantly from the normoalbuminuric group (table S5).

Finally, we reanalyzed the uEV mRNA profile of T1D individuals based on CKD stage, irrespective of albuminuria, using retrospective eGFR data (> 5 years) from routine laboratory follow-up testing for CKD classification (table S3). This confirmed significant ($p < 0.004$) upregulation of all 13 differentially expressed genes also in T1D individuals with CKD stage ≥ 3 compared with CKD stage ≤ 2 (table S5).

To technically validate the differential expression results from NGS, we studied the expression of five of the top differentially expressed genes (*MAP7*, *GPX3*, *IL32*, *NOX4*, *HRSP12*) using Taqman qPCR assays in the same set of samples. The qPCR- and mRNAseq-based expression values were mostly highly correlated (**Fig. 6A**). Similar to the mRNAseq results, the qPCR-based expression levels were significantly ($p \leq 0.005$) upregulated in macroalbuminuric T1D individuals and less so in microalbuminuric subjects compared to those with normoalbuminuria (**Fig. 6B**).

T1D DKD differential uEV mRNAs correlation with hyperglycemia and co-expression similarity with kidney

We studied whether the expression level of the 13 differentially expressed genes, considered as candidate marker genes for T1D DKD, correlated with clinical variables contributing to DKD progression (hyperglycemia and hypertension) at or close to the time of urine collection (**Fig. 7**).

Most of the differentially expressed candidate genes (11 out of 13), except *MAP7* and *CAPN3*, showed a positive correlation with HbA_{1c} (**Fig. 7, A and B**). Of note, the strongest correlation

($p \leq 0.001$) was observed for the candidate genes known to be highly expressed in kidney (*TINAG*, *GPX3* and *RBP5*) compared to other human tissues (fig. S8). In a subanalysis that included samples from normoalbuminuric individuals only (from both cohorts, $n=37$), we also saw a trend of positive correlation between expression levels of 10 candidate genes and HbA_{1c} suggesting, that the observed correlations in the whole T1D cohort were not only driven by albuminuria or kidney disease (**Fig. 7C**).

We next explored the origins of the differential uEV transcripts using co-expression pattern and by positioning the candidate signature on a tissue map that we created by using an external human tissue expression reference database (GTEx) (30, 31). Considering the urogenital organs previously established to be important contributors of uEV (**Fig. 4**), the expression level of 8 out of 12 candidate marker genes (present in GTEx dataset) of T1D DKD were higher in kidney cortex compared to prostate, bladder and testis samples in the GTEx datasets (male only, **Fig. 8A**). By pairwise correlation analysis, the co-expression pattern of the candidate genes in uEV resembled the pattern in kidney cortex when focusing on most of the kidney enriched genes (*MSRA*, *RBP5*, *GPX3*, *CYCL14*, *TINAG*, *NOX4*, *IL32* and *CRYAB*), but also in case of many of the non-kidney enriched candidate genes (*MSRB1*, *CAPN3*, *TMEM9* and *MAP7*) (**Fig. 8, B to E**). Further, we utilized single nuclei RNAseq data ($>23,000$ nuclei) from human diabetic kidney tissue (34), to look up the expression levels of the T1D DKD candidate genes in different cell types of the kidney. Most of the candidate genes (6 out of 9 genes that had cell type specific expression data available) showed highest expression in the proximal convoluted tubule (PCT) cells of kidney (**Fig. 8F**).

In summary, expression analysis of candidate marker genes for T1D DKD suggested that hyperglycemia may contribute to their higher expression in uEV samples. The main contributors of the candidate marker genes to uEV could be the kidney proximal convoluted tubule cells.

T1D DKD differential uEV mRNAs-based stress score and long-term eGFR change

All or many of the 13 differentially expressed candidate marker genes for T1D DKD showed common characteristics: 1) upregulation in DKD individuals (**Fig. 5**), 2) positive correlation with hyperglycemia (**Fig. 7**), and 3) higher expression level in kidney cortex and proximal tubular cells (**Fig. 8**). This suggested that also the underlying disease mechanism responsible for these common characteristics might be common. Based on literature regarding the 13 DKD candidate genes, we found that 6 out of 13 of the genes executed functions in cellular stress responses, namely 1) oxidative stress related responses (*GPX3*, *NOX4*, *MSRB1*, *MSRA*) and 2) protection of cellular proteins during stress (*HRSP12* and *CRYAB*) (table S6). In addition, the expression level of these 6 genes showed a positive correlation with HbA_{1c} (**Fig. 7**) and co-expression with each other (**Fig. 8D**). Hyperglycemia induced stress in kidney cells including the PCT is considered to be an important disease mechanism contributing to DKD progression (35). Thus, we speculated that these 6 genes may be involved in hyperglycemia induced stress responses of the kidney and that their expression level in uEV could reflect the long-term change in kidney function of diabetic individuals. To test this hypothesis, we derived a stress score based on average normalized expression level of the 6 stress response related differentially expressed genes in uEV (*GPX3*, *NOX4*, *MSRB1*, *MSRA*, *HRSP12* and *CRYAB*, see methods statistics). As expected, the stress score had a clear increasing trend from normo- to macroalbuminuria and showed a positive correlation with HbA_{1c} measured at the time point of urine sample collection (**Fig. 9A and B**). The stress-score did not correlate with eGFR, SBP and DBP measured at the time of urine collection (**Fig. 9B**).

We then stratified the T1D cohorts into quartiles (0-25, 25-50, 50-75, 75-100) based on the stress score and compared annual change in clinical parameters between them over time. To compare the annual eGFR change, we collected long-term eGFR data retrospectively (15 years, from 2004 to 2018) for most of the T1D individuals (available for 51 of 67) from clinical laboratory databases. The lowest stress score quartile (0-25) showed only minimal decline in the annual eGFR (slope_(mL/min/1.73 m²)±SE; -0.58±0.19, p=0.008) (**Fig. 9C**). In contrast, the annual eGFR decline was >2-fold steeper in

the highest quartile (75-100; slope_(mL/min/1.73m²)±SE; -1.55±0.27, p<0.0001). As the stress score was defined based on the change in macroalbuminuria group, we performed a subanalysis by excluding this group. Hence, we compared the normoalbuminuric T1D individuals within the lowest stress score quartile and within the three higher quartiles (0-25 and >25) to each other and to the group of all microalbuminuric T1D individuals (regardless of the stress score) for the decline of long-term eGFR. The normoalbuminuric group with the lowest stress score (0-25) showed only a minimal decline (slope_(mL/min/1.73m²)±SE; -0.59±0.19, p=0.007) in the eGFR compared with normoalbuminuria group with mid-to-high stress score (>25; slope_(mL/min/1.73m²)±SE; -1.34±0.20, p<0.0001), which did not markedly differ from the microalbuminuria group (slope_(mL/min/1.73m²)±SE; -1.08±0.16, p<0.0001) (**Fig. 9D**). The overall change in the longitudinal HbA_{1c}% (13 years) did not differ between the normoalbuminuria groups having low or mid-to-high stress score (0-25 and >25) (**Fig. 9E**), but the most recent HbA_{1c} level measured at the time point of uEV collection was higher for the mid-to-high (>25) stress score group similar to as observed for microalbuminuria group (**Fig. 9F**).

Overall, the analysis showed that a lower uEV stress score – a lower level of candidate gene mRNA in uEV - associates with a minimal long-term decline of kidney function. Conversely, a higher stress score associates with a faster decline of kidney function in T1D individuals regardless of their albuminuria status. This may reflect that these genes could be putative candidate genes involved in the response of the kidney to the stress caused by hyperglycemic conditions.

Discussion

Extracellular RNAs originating from the kidney and passed to urine inside EVs could serve as a non-invasive proxy for the biopsies, a liquid kidney biopsy. However, as most EV studies have focused on small RNAs (36), we miss systematic characterization of whole mRNA transcriptomes in uEV, particularly with a focus on DKD. Here we characterized the uEV mRNA transcriptomes obtained from different urine collections, showed their similarity to human kidney mRNA profile and utility in discovering candidate markers for T1D DKD. This proof-of-concept study supports

the idea that uEV mRNA transcriptomes capture kidney-derived gene expression signatures suitable for DKD precision medicine.

Overall, we detected expression of >10,000 genes despite focusing on the most commonly expressed genes (>90% samples) defining the most stable EV-mRNA transcriptome in population scale. The expression of large gene numbers in uEV is in line with results from some recent studies with few individuals (23, 27, 28). Previous studies have also confirmed the occurrence of some mRNAs in urine by qPCR even in population scale (37). However, why such a high number of mRNAs occur in uEV is still not known.

EVs in urine are reported to originate from the urogenital system (38, 39), but also from distant organs beyond the glomerular filtration barrier, including the brain (40-42). We detected a high proportion (>50%) of kidney enriched genes in uEV and through placing the total uEV mRNA transcriptome on a tissue map, i.e. comparing it with other human tissues, we confirmed the overall similarity with the kidney. Interestingly, irrespective of the albuminuria status, the uEV showed only a small subset of transcripts (<5%) not known to be expressed in the kidney. This suggests that kidney is the main contributor of uEV and that the glomerular filtration barrier limits the access of EV from non-urinary tract organs into the urine. Nevertheless, as we did detect some non-kidney expressed genes, we cannot totally exclude the possibility of a low degree EV transfer from circulation to urine (43). Out of the other urogenital system tissues of men, prostate tissue showed the closest similarity with uEV (**Fig. 4**). This finding is in line with application of uEV for screening of prostate cancer (the Exosome Diagnostic's ExoDx Prostate cancer test) (44).

EV field is still in its infancy with non-standardized workflows for urine transcriptome, particularly when considering EV mRNAs as compared to miRNAs in DKD (45). Thus, to be able to evaluate the potential of the uEV mRNA transcriptome as a DKD biomarker source, we needed to test several pre-analytical and analytical factors. Specifically, we assessed the impact of the urine collection methods used in the clinic, 24-h or ON, which can affect the uEV concentration or cargo

and thereby the detection sensitivity. We confirmed that ON urine – easier to collect than 24-h – produced almost similar mRNA profile and expression level of kidney enriched genes as the 24-h urine. Additionally, our whole uEV pipeline was reproducible, as it gave similar mRNA transcriptomes from the technical replicates processed at 1-5 months intervals through the pipeline. This suggests that urine storage at -80°C preserves uEV mRNAs well. Further, the degree of variation due to technical steps that are often assumed to cause problems, such as uEV isolation by ultracentrifugation (46) or sequencing in different batches (47), was not significant here. We also validated the mRNAseq results with qPCR for the top differential genes. We finally showed the difference in the whole uEV mRNA transcriptomes between men and women. However, the difference was minimal when focusing on kidney enriched genes. This technical information should assist in establishing a standardized workflow from biobanking of urine to future large scale uEV mRNA research.

The study revealed putative marker genes for DKD in T1D individuals. Interestingly, we found that the majority of the top differentially expressed genes between different albuminuria or eGFR status groups shared a common pattern: upregulation in T1D individuals with DKD, positive correlation with hyperglycemia, high expression in PCT and implication as cellular stress responsive genes (table S6). These top genes appear to follow the main path of the development of DKD, starting with hyperglycemia altering the metabolic flux of glucose, which causes oxidative stress mainly in PCT and further leads to inflammation, fibrosis and decline of the renal function (48, 49). The important roles of PCT in kidney dysfunction (DKD and CKD) is also supported by kidney tissue bulk (9) and single cell RNA sequencing data (50, 34). Remarkably, now uEV provide an opportunity to assess the expression of some PCT enriched genes (e.g. *TINAG*, *CXCL14* and *RBP5*) highly specific for PCT, **Fig. 8**) non-invasively in T1D individuals progressing towards DKD. Based on the known functions of the candidate marker genes, it appears that many associate with kidney diseases, and our stress score genes, particularly *GPX3*, *HRSP12*, *MSRA*, *MSRB1* and

354 *CRYAB*, protect the cell and/or cellular proteins during stress (table S6). They thus may have a
355 reno-protective role by mitigating hyperglycemia-induced stress in the kidney. *GPX3* is one such
356 antioxidant enzyme mainly produced in PCT and involved in relieving oxidative stress by
357 neutralizing hydrogen peroxide (51). The *GPX3* and *MSRBI* belong to the selenoprotein class (52).
358 Selenium deficiency induces oxidative stress in murine kidney cells (53). Recently, enhancing the
359 *GPX3* activity by selenium supplementation was shown to rescue kidney function in T2D human
360 subjects (54). *NOX4*, also one of the stress score genes, has many roles in DKD ranging from
361 detrimental to protective and its protein level was recently shown to be decreased in PCT during
362 CKD (55, 56). It is tempting to speculate that secretion of these reno-protective transcripts by PCT
363 to urine via EVs may contribute to kidney dysfunction following hyperglycemia induced stress.
364 Indeed, based on our clinical data, the T1D individuals with the lowest level of the stress score
365 transcripts in uEV showed only minimal decline of the kidney function over 15 years. Thus,
366 retaining the transcripts in the kidney may be associated with protection. Increased EV secretion
367 during various types of stress, including oxidative stress (mechanical stretch, hypoxia,
368 hyperglycemia, inflammation, acidosis) could contribute to the “leaking” of kidney transcripts (57-
369 59). One mechanistic explanation could be that the impaired intracellular EV degradation via
370 defective autophagy in DKD would lead to increased EV secretion (60, 61), as the two systems
371 appear closely linked, which is supported by accumulating data from cell culture studies (62-64).
372 Alternatively, or additionally, the observed increased transcript levels in uEV could be caused by
373 lack of uptake and disturbed intra-renal signaling (15, 22). Overall, experimentation, including
374 using EV secretion inhibitors, would help to confirm whether increased transcript secretion via uEV
375 in T1D individuals causes a decrease in the level of reno-protective proteins in the kidney or
376 whether it is just a consequence of an ongoing compensatory mechanism in a hyperglycemic
377 kidney.

378 Among the rest of the candidate genes with no known role in cellular stress, *TINAG* associated
379 strongly with hyperglycemia. The changes in *IL32* and *CYXCL14* suggest participation of
380 inflammatory pathways in DKD. The expression level of *IL32* in kidney (11) and a genetic variant
381 in the *TINAG* gene region have been shown to correlate with eGFR levels (rs142516820, $P=$
382 0.95×10^{-13}), suggesting roles in kidney function (65). In summary, our hypothesis-free strategy
383 identified DKD candidate genes, which are mostly kidney-derived and have previously established
384 roles in kidney function.

385 The main limitations of our study were 1) low sample size of individual groups, 2) focusing on
386 male subjects only (due to expected higher heterogeneity in women, which could have had an
387 impact considering the low sample size), and 3) lack of replication of the findings in independent
388 cohorts. However, given the novelty of the whole transcriptome EV mRNAseq methods, this is the
389 largest study till now with inclusion of two different T1D cohorts and detailed long term
390 retrospective clinical data. Additionally, we define a robust uEV transcriptome of >10,000 genes,
391 show the position of uEV on tissue map, and provide evidence that uEV pipelines can be
392 reproducible and applicable for high-impact large scale biomarker studies. Of note, better
393 exploration of these topics has recently been warranted by both the extracellular RNA
394 communication consortium (36) and Urine Task Force of the International Society for Extracellular
395 Vesicles (29).

396 In summary, the study comprehensively shows the utility of uEV mRNA transcriptomics in
397 capturing the gene expression patterns of the kidney and reveals type 1 DKD candidate marker
398 genes including putative reno-protective genes. These genes could be targeted for development of
399 novel biomarkers or therapeutic targets for DKD. Thus, we conclude that uEV offer a promising
400 liquid kidney biopsy of DKD.

401
402

Materials and Methods

Study groups and ethics

Urine samples and clinical data were collected in the nation-wide prospective FinnDiane (Finnish Diabetic Nephropathy) Study targeting risk factors for DKD in type 1 diabetes (T1D), and the DIREVA study (Diabetes Registry in Vaasa), inviting all diabetic individuals living in the Vaasa Central Hospital District to participate since 2007. The studies followed the principles of the Declaration of Helsinki and all participants gave an informed consent prior to participation. The study protocol for this substudy of the FinnDiane and DIREVA studies was approved by the Ethics Committee of the Turku University Hospital. The study protocol of the FinnDiane Study has also been approved by the Ethical Committee of the Helsinki and Uusimaa Hospital District (FinnDiane 163/E5/04, 199/E5/05, 491/E5/2006; 238/13/03/00/15), and that of DIREVA, by the Ethical Committees of the Vaasa Central Hospital (6/2007) and Turku University Hospital (48/1801/2014; 116/1805/2016).

In the FinnDiane Study, type 1 diabetic men aged 18-70 years were recruited. Type 1 diabetes was defined as an onset of diabetes before the age of 40 and insulin treatment initiated within the first year after diagnosis. In spurious cases, C-peptide levels (<0.2 mmol/l) were required and medical files were reviewed. In DIREVA, male individuals with clinical type 1 diabetes aged 30-60 years with different degrees of albuminuria and no history of pancreatitis were invited. Information on diabetes heredity, history of pancreatitis or gestational diabetes was also collected to verify the diagnosis of type 1 diabetes. Study subjects were investigated at the local study centers, where trained research nurses or physicians collected information using standardized questionnaires on, among others, sex, age, age at onset of diabetes, duration of diabetes, micro- and macrovascular complication profile and current medication. The nurses measured height, weight, waist and hip circumference in light clothing, and BMI and waist- to-hip ratio (WHR) was calculated. The blood

pressure was measured twice with a 5 min interval from the right arm of a sitting person after an initial rest. The mean of the two measurements was calculated and used in the analyses. Venous blood samples were taken for the assessment of HbA_{1c}, blood lipids and lipoproteins (total and HDL-cholesterol and triglycerides), creatinine, fS-C-peptide, fP-Glucose, S-GAD-antibodies (DIREVA) and DNA. The samples were taken either fasting or a minimum of two hours after a light meal. In addition, retrograde data on HbA_{1c}, serum and plasma creatinine for estimation of glomerular filtration rate (eGFR using the CKD-EPI formula) (66), urine albumin excretion rate (AER) and urine albumin-creatinine ratio (ACR) were obtained from hospital laboratory databases for the validation of the albuminuria status as well as for the longitudinal analyses.

Altogether, our study included 72 men with T1D from the two cohorts. The participants were stratified into normo-, micro- and macroalbuminuria groups. In the FinnDiane cohort, the stratification was done according to the AER (based on 24-h urine collection) in two out of three consecutive urine collections: macroalbuminuria with AER >300 mg/24-h, microalbuminuria with AER 30-300 mg/24-h, and normoalbuminuria with AER <30 mg/24-h (table S1 and S2). In the DIREVA cohort, macroalbuminuria was defined as AER (based on ON urine collection) >200 µg/min or ACR (spot urine albumin/creatinine ratio) >35 mg/mmol, and microalbuminuria as AER 20-200 µg/min or ACR >3.5 mg/mmol at the last clinical visit. Normoalbuminuria was defined as AER <20 µg/min and ACR <3.5 (mg/mmol) at all of the clinical visits (table S1 and S2). For trajectory analysis (**Fig. 9**), we used available retrospective data on HbA_{1c} and serum or plasma creatinine from both cohorts: 1) for the last 13-15 years (available for 51 individuals), and 2) for the minimum of last five years (available for 66 individuals). For eGFR slope calculations using 15 years data, we used the average of all eGFR values per year for each individual. The classification into CKD stages (table S3) was based on retrospective data of at least > 5 years eGFR measurements for each individual (n=66), where at least two annual average values falling into the lowest eGFR range determined the CKD stage; stage 1-2 (eGFR >60), or stage 3-5 (eGFR <60). For

the technical comparisons of uEV mRNAseq (**Fig. 1, 3** and fig. S5-6), we also included age and sex matched non-diabetic controls (n=4) and women with T1D and normo- or microalbuminuria (n=3) from the FinnDiane Study (table S4).

Urine collection and isolation of uEV

The participants collected whole voids of either 24-h (FinnDiane) or ON (DIREVA) urine into containers containing Protease Inhibitor Cocktail Set III (Calbiochem, final dilution 1/1000-1/2000). The urine samples were kept in temperature ranging from ambient to +4°C during collection and delivery (the same or next day) to laboratory, where they were processed with (DIREVA) or without (FinnDiane) centrifugation at 1800g for 10 min at +4°C and frozen at -80°C. In addition, for technical comparisons, we collected similarly paired 24-h and ON urines from the same day and individuals (without centrifugation), and, 24-h urine samples from which aliquots were processed in parallel with and without centrifugation (1800g, 10 min +4°C) prior freezing (FinnDiane). EVs were isolated from 30 ml of urine by differential centrifugation as reported previously (67, 28). Briefly, the urine samples were vortexed for 90s, centrifuged at 8000g for 15 min at +4°C and the resulting supernatant filtered (1.2 µm cellulose acetate, Whatman) to remove cell debris. EVs were pelleted by ultracentrifugation with a SW-28 rotor at 27500 rpm (136367g_{max}, k-factor 254.5, Beckmann-Coulter) for 1h 30 min at +4°C, washed with PBS and the ultracentrifugation step repeated. EV pellets were suspended in PBS and stored in protein or DNA Low-bind tubes (Eppendorf, Hamburg, Germany) at -80°C.

Western blotting, electron microscopy and nanoparticle tracking analysis of uEV

Urinary EV samples were analyzed as explained previously (67, 28). Briefly, Western blotting of uEV marker proteins was performed using antibodies against CD9 (SC-13118, Santa Cruz) and podocalyxin (clone 3D3, Novus Biologicals). EVs from equal volumes of urine (EVs derived from 1.7 ml of urine samples) and 20 µg of protein from LNCaP cell lysate control were loaded to gels.

478 For detection, x-ray film (Ultra Cruz™ Autoradiography Film, Santa Cruz Biotechnology (SC),
479 Dallas, TX, USA) was used.

480 Electron microscopy (EM) samples were prepared with a negative staining protocol. The samples
481 were viewed with Jeol JEM-1400 (Jeol Ltd., Tokyo, Japan) at 80 kV and imaged with Gatan Orius
482 SC 1000B CCD-camera (Gatan Inc., USA) with 4008 × 2672 px image size and no binning.

483 NTA was done with Nanosight LM14 (Malvern Instruments Ltd, Malvern, UK) with blue (404 nm,
484 70 mW) laser and SCMOS camera (Hamamatsu photonics K.K., Hamamatsu, Japan). Samples were
485 diluted in filtered (0.1μm, Millex VV, Millipore) DPBS to obtain 40-100 particles per view. Five 30
486 s videos were recorded with camera level 13. The videos were analyzed using software version 3.0
487 (Malvern Instruments Ltd), detection threshold 5 and screen gain 10.

488

489 **RNA preparation, mRNA sequencing and data analysis**

490 Total RNAs from the uEV were extracted using miRNEasy mini kit according to manufacturer's
491 instructions (Qiagen, Hilden, Germany). RNA profiles and concentrations were measured with
492 RNA 6000 Pico Total RNA Kit run on Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA).

493 For mRNAseq, the cDNA was prepared with SMART-Seq® v4 Ultra™ Low Input RNA Kit for
494 Sequencing (Takara Bio Inc., Mountain View, CA) using 1 ng of total RNA as the input. The
495 cDNA libraries were prepared with Nextera XT DNA Library Preparation and Index Kits (Illumina
496 Inc., San Diego, CA, USA). The cDNAs and libraries were measured with High Sensitivity D5000
497 ScreenTape system of TapeStation (Agilent). Paired end (150 or 300 cycles) sequencing was
498 performed on NextSeq 550 (Illumina). RNA sequencing reads were aligned to hg38 using STAR
499 (STAR-2.6.1d) (68), genome annotations were obtained from the GENCODE (Encyclopedia of
500 Genes and Gene Variants) v227 program (69), and read counting was done using featureCounts
501 (Version 1.6.4) software package (70) Read counts were normalized using TMM (trimmed mean of
502 M-values) method (71).

503 ***Kidney enriched gene list and GTEx analysis***

504 We used human protein atlas to obtain the lists of 1) “kidney enriched genes” (n=413, detected in
505 kidney and also many other tissues) with an elevated expression in the kidney compared to other
506 tissue types, 2) “specific kidney enriched genes” (n=190) – a subset of “kidney enriched genes”
507 (n=413) that are detected in kidney and only in some (more than one but less than one third of 37
508 tissues) other tissues and 3) “kidney depleted genes” (n=4631) not detected in kidney (32, 33). In
509 the group of uEV samples used for technical comparisons, we detected 233 “kidney enriched
510 genes” (≥ 1 count in all samples). In the two T1D cohorts combined we detected 247 “kidney
511 enriched genes” (≥ 1 count in all samples) and 77 “specific kidney enriched genes”. Additionally, in
512 both T1D cohorts we detected only 201 “kidney depleted genes” (≥ 1 count in all samples) in uEV.
513 We used raw gene count (GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_reads) and
514 TPM values (GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_tpm) from GTEx databases
515 (GTEx analysis V8) for analysis (30, 31). For analysis purpose to focus on the main tissues, various
516 levels of the single tissue data were merged and presented as one (e.g. sun exposed and no sun
517 exposed skin merged and presented as “skin” only). The downstream processing of read count data
518 used for PCA analysis were done using edgeR software (72).

519 **Taqman qPCR**

520 Selected differentially expressed genes from mRNAseq were validated using Taqman qPCR in
521 Fluidigm Biomark™ HD 96.96 platform (Fluidigm, San Francisco, CA, USA) according to
522 manufacturer's instructions. Briefly, 1,25 μ l of cDNAs (see preparation from RNA sequencing)
523 were preamplified with 14 PCR-cycles according to the instructions in Fluidigm’s Real-Time PCR
524 Analysis User Guide (PN 68000088 K1, Appendix D); the conditions were 95°C for 2 min,
525 followed by 14 cycles of 95°C for 15 s and 60°C for 4 min. The number of preamplification cycles,
526 14 against 20, was tested separately with a part of the assays and samples, where 14 cycles gave
527 better Ct values and amplification curves (data not shown). We applied one to two different assays

for each gene, and 4-6 technical replicates, including two reverse transcription replicates, of each sample/assay. Water instead of sample served as a negative control. The assays with FAM-MGB labelled probes were: MAP7 (Hs01010631_m1, Hs01009609_m1), GPX3 (Hs00173566_m1, Hs01078668_m1), IL32 (Hs00170403_m1, Hs00992441_m1), NOX4 (Hs04980924_m1, Hs01379108_m1), HRSP12 (Hs01066170_m1), GAPDH (Hs99999905_m1) (all Applied Biosystems; Thermo Fisher Scientific, Inc., Waltham, MA, USA). The qPCR was performed using the online protocol “Gene Expression with the 96.96 IFC Using Fast TaqMan Assays (Biomark HD Only) (PN 100-2638 D1)” with conditions of 95°C for 60 s and 35 repeats of 97°C for 5 s and 60°C for 20 s. Data was analyzed with the method according to Livak and Schmittgen (73) with normalization using GAPDH. Samples with less than two successful replicas or high deltaCt (>19) were excluded from the analysis.

Statistics

Details of statistical analysis used for each section have been described in figure legends and manuscript text. Briefly, differential expression analysis was performed based on count data using generalized linear models, as implemented in edgeR software (72), adjusting for age, BMI, diabetes duration and urine collection protocols (ON and 24-h) or other covariates (described in figures) as appropriate. For qPCR data analysis, deltaCt values were transformed using inverse normal transformation. Correlation analysis were performed using non-parametric Spearman’s correlation (two tailed). The transcriptional risk score was derived by averaging the inverse rank normalized expression values of the candidate genes. Slopes in the longitudinal data analysis were estimated using linear regression analysis. Data is presented as Mean $\pm$ SEM or box and whiskers plot (interquartile range, median and summary minimum/maximum values) as described in figures.

References:

1. X. Lin, Y. Xu, X. Pan, J. Xu, Y. Ding, X. Sun, X. Song, Y. Ren, P. F. Shan, Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. *Sci. Rep.* **10**, 14790 (2020).
2. M. C. Thomas, M. Brownlee, K. Susztak, K. Sharma, K. A. Jandeleit-Dahm, S. Zoungas, P. Rossing, P. H. Groop, M. E. Cooper, Diabetic kidney disease. *Nat. Rev. Dis. Primers* **1**, 15018 (2015).
3. T. W. Tervaert, A. L. Mooyaart, K. Amann, A. H. Cohen, H. T. Cook, C. B. Drachenberg, F. Ferrario, A. B. Fogo, M. Haas, E. de Heer, K. Joh, L. H. Noël, J. Radhakrishnan, S. V. Seshan, I. M. Bajema, J. A. Bruijn, Pathologic classification of diabetic nephropathy. *J. Am. Soc. Nephrol.* **21**, 556-563 (2010).
4. H. J. Anders, T. B. Huber, B. Isermann, M. Schiffer, CKD in diabetes: diabetic kidney disease versus nondiabetic kidney disease. *Nat. Rev. Nephrol.* **14**, 361-377 (2018).
5. M. L. Gonzalez Suarez, D. B. Thomas, L. Barisoni, A. Fornoni, Diabetic nephropathy: Is it time yet for routine kidney biopsy?. *World J. Diabetes.* **4**, 245-255 (2013).
6. J. L. Gross, M. J. de Azevedo, S. P. Silveiro, L. H. Canani, M. L. Caramori, T. Zelmanovitz, Diabetic nephropathy: diagnosis, prevention, and treatment. *Diabetes Care.* **28**, 164-176 (2005).
7. N. M. Selby, M. W. Taal, An updated overview of diabetic nephropathy: Diagnosis, prognosis, treatment goals and latest guidelines. *Diabetes Obes. Metab. Suppl* **1**, 3-15 (2020).
8. C. G. Champion, O. Sanchez-Ferras, S. N. Batchu, Potential Role of Serum and Urinary Biomarkers in Diagnosis and Prognosis of Diabetic Nephropathy. *Can J. Kidney Health Dis.* **4**, 2054358117705371 (2017).
9. Y. Fan, Z. Yi, V. D. D'Agati, Z. Sun, F. Zhong, W. Zhang, J. Wen, T. Zhou, Z. Li, L. He, Q. Zhang, K. Lee, J. C. He, N. Wang, Comparison of Kidney Transcriptomic Profiles of Early and Advanced Diabetic Nephropathy Reveals Potential New Mechanisms for Disease Progression. *Diabetes* **68**, 2301-2314 (2019).
10. P. Beckerman, C. Qiu, J. Park, N. Ledo, Y. A. Ko, A. D. Park, S. Y. Han, P. Choi, M. Palmer, K. Susztak, Human Kidney Tubule-Specific Gene Expression Based Dissection of Chronic Kidney Disease Traits. *EBioMedicine* **24**, 267-276 (2017).
11. W. Ju, V. Nair, S. Smith, L. Zhu, K. Shedden, P. X. K Song, L. H. Mariani, F. H. Eichinger, C. C. Berthier, A. Randolph, J. Y. Lai, Y. Zhou, J. J. Hawkins, M. Bitzer, M. G. Sampson, M. Thier, C. Solier, G. C. Duran-Pacheco, G. Duchateau-Nguyen, L. Essioux, B. Schott, I. Formentini, M. C. Magnone, M. Bobadilla, C. D. Cohen, S. M. Bagnasco, L. Barisoni, J. Lv, H. Zhang, H. Y. Wang, F. C. Brosius, C. A. Gadegbeku, M. Kretzler, Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci. Transl. Med.* **7**, 316ra193 (2015).
12. S. Eddy, L. H. Mariani, M. Kretzler, Integrated multi-omics approaches to improve classification of chronic kidney disease. *Nat. Rev. Nephrol.* **16**, 657-668 (2020).
13. Y. Zhou, J. Yang, Extracellular RNA in renal diseases. *ExRNA* **1**, 5 (2019). <https://doi.org/10.1186/s41544-018-0001-x>.
14. M. Yáñez-Mó, P. R. Siljander, Z. Andreu, A. B. Zavec, F. E. Borràs, E. I. Buzas, K. Buzas, E. Casal, F. Cappello, J. Carvalho, E. Colás, A. Cordeiro-da Silva, S. Fais, J. M. Falcon-Perez, I. M. Ghobrial, B. Giebel, M. Gimona, M. Graner, I. Gursel, M. Gursel, N. H. Heegaard, A. Hendrix, P. Kierulf, K. Kokubun, M. Kusanovic, V. Kralj-Iglic, E. M. Krämer-Albers, S. Laitinen, C. Lässer, T. Lener, E. Ligeti, A. Linē, G. Lipps, A. Llorente, J. Lötval, M. Manček-Keber, A. Marcilla, M. Mittelbrunn, I. Nazarenko, E. N. Nolte-t Hoen, T. A. Nyman, L. O'Driscoll, M. Oliván, C. Oliveira, É. Pállinger, H. A. Del Portillo, J. Reventós, M. Rigau, E. Rohde, M. Sammar, F. Sánchez-Madrid, N. Santarém, K. Schallmoser, M. S. Ostfeld, W. Stoorvogel, R. Stukelj, S. G. Van der Grein, M. H. Vasconcelos, M. H. Wauben, O. De Wever, Biological properties of extracellular vesicles and their physiological functions. *J. Extracell. Vesicles* **4**, 27066 (2015).
15. M. A. C. Pomatto, C. Gai, B. Bussolati, G. Camussi, Extracellular Vesicles in Renal Pathophysiology. *Front. Mol. Biosci.* **4**, 37 (2017).
16. K. Barreiro, H. Holthofer, Urinary extracellular vesicles. A promising shortcut to novel biomarker discoveries. *Cell Tissue Res.* **369**, 217-227 (2017).
17. K. M. Kim, K. Abdelmohsen, M. Mustapic, D. Kapogiannis, M. Gorospe, RNA in extracellular vesicles. *Wiley Interdiscip. Rev. RNA* **8**, 10.1002/wrna.1413 (2017).
18. O. Pös, O. Biró, T. Szemes, B. Nagy, Circulating cell-free nucleic acids: characteristics and applications. *Eur. J. Hum. Genet.* **26**, 937-945 (2018).
19. F. Barutta, M. Tricarico, A. Corbelli, L. Annaratone, S. Pinach, S. Grimaldi, G. Bruno, D. Cimino, D. Taverna, M. C. Deregibus, M. P. Rastaldi, P. C. Perin, G. Gruden, Urinary exosomal microRNAs in incipient diabetic nephropathy. *PLoS One* **8**, e73798 (2013).
20. D. Delić, C. Eisele, R. Schmid, P. Baum, F. Wiech, M. Gerl, H. Zimdahl, S. S. Pullen, R. Urquhart, Urinary Exosomal miRNA Signature in Type II Diabetic Nephropathy Patients. *PLoS One* **11**, e0150154 (2016).
21. V. Ghai, X. Wu, A. Bheda-Malge, C. P. Argyropoulos, J. F. Bernardo, T. Orchard, D. Galas, K. Wang, Genome-wide Profiling of Urinary Extracellular Vesicle microRNAs Associated With Diabetic Nephropathy in Type 1 Diabetes. *Kidney Int. Rep.* **3**, 555-572 (2017).

22. T. Medeiros, R. L. Myette, J. R. Almeida, A. A. Silva, D. Burger, Extracellular Vesicles: Cell-Derived Biomarkers of Glomerular and Tubular Injury. *Cell Physiol. Biochem.* **54**, 88-109 (2020).
23. K. C. Miranda, D. T. Bond, M. McKee, J. Skog, T. G. Păunescu, N. Da Silva, D. Brown, L. M. Russo, Nucleic acids within urinary exosomes/microvesicles are potential biomarkers for renal disease. *Kidney Int.* **78**, 191-199 (2010).
24. T. Murakami, M. Oakes, M. Ogura, V. Tovar, C. Yamamoto, M. Mitsuhashi, Development of glomerulus-, tubule-, and collecting duct-specific mRNA assay in human urinary exosomes and microvesicles. *PLoS One* **9**, e109074 (2014).
25. D. Enderle, A. Spiel, C. M. Coticchia, E. Berghoff, R. Mueller, M. Schlumpberger, M. Sprenger-Haussels, J. M. Shaffer, E. Lader, J. Skog, M. Noerholm, Characterization of RNA from Exosomes and Other Extracellular Vesicles Isolated by a Novel Spin Column-Based Method. *PLoS One* **10**, e0136133 (2015).
26. B. G. Bazzell, W. E. Rainey, R. J. Auchus, D. Zocco, M. Bruttini, S. L. Hummel, J. B. Byrd, Human Urinary mRNA as a Biomarker of Cardiovascular Disease. *Circ. Genom. Precis. Med.* **11**, e002213 (2018).
27. C. Everaert, H. Helsmoortel, A. Decock, E. Hulstaert, R. Van Paemel, K. Verniers, J. Nuytens, J. Anckaert, N. Nijs, J. Tulkens, B. Dhondt, A. Hendrix, P. Mestdagh, J. Vandesompele, Performance assessment of total RNA sequencing of human biofluids and extracellular vesicles. *Sci. Rep.* **9**, 17574 (2019).
28. K. Barreiro, O. P. Dwivedi, G. Lepar, M. Rolser, D. Delic, C. Forsblom, P. H. Groop, L. Groop, T. B. Huber, M. Puhka, H. Holthofer, Comparison of urinary extracellular vesicle isolation methods for transcriptomic biomarker research in diabetic kidney disease. *J. Extracell. Vesicles* **10**, e12038 (2020).
29. U. Erdbrügger, C. J. Blijdorp, I. V. Bijnsdorp, F. E. Borràs, D. Burger, B. Bussolati, J. B. Byrd, A. Clayton, J. W. Dear, J. M. Falcón-Pérez, C. Grange, A. F. Hill, H. Holthöfer, E. J. Hoorn, G. Jenster, C. R. Jimenez, K. Junker, J. Klein, M. A. Knepper, E. H. Koritzinsky, J. M. Luther, M. Lenassi, J. Leivo, I. Mertens, L. Musante, E. Oeyen, M. Puhka, M. E. van Royen, C. Sánchez, C. Soekmadji, V. Thongboonkerd, V. van Steijn, G. Verhaegh, J. P. Webber, K. Witwer, P. S. T. Yuen, L. Zheng, A. Llorente, E. S. Martens-Uzunova, Urinary extracellular vesicles: A position paper by the Urine Task Force of the International Society for Extracellular Vesicles. *J. Extracell. Vesicles* **10**, e12093 (2021).
30. A. Battle, C. D. Brown, B. E. Engelhardt, S. B. Montgomery; GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration & Visualization—EBI; Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis & Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
31. <https://www.gtexportal.org/home/datasets>
32. M. Habuka, L. Fagerberg, B. M. Hallström, C. Kampf, K. Edlund, Å. Sivertsson, T. Yamamoto, F. Pontén, M. Uhlén, J. Odeberg, The kidney transcriptome and proteome defined by transcriptomics and antibody-based profiling. *PLoS One* **9**, e116125 (2014).
33. <https://www.proteinatlas.org/humanproteome/tissue/kidney>
34. P. C. Wilson, H. Wu, Y. Kirita, K. Uchimura, N. Ledru, H. G. Rennke, P. A. Welling, S. S. Waikar, B. D. Humphreys, The single-cell transcriptomic landscape of early human diabetic nephropathy. *Proc. Natl. Acad. Sci.* **116**, 19619-19625 (2019).
35. V. Vallon, The proximal tubule in the pathophysiology of the diabetic kidney. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **300**, R1009-22 (2011).
36. S. Das, Extracellular RNA Communication Consortium, K. M. Ansel, M. Bitzer, X. O. Breakefield, A. Charest, D. J. Galas, M. B. Gerstein, M. Gupta, A. Milosavljevic, M. T. McManus, T. Patel, R. L. Raffai, J. Rozowsky, M. E. Roth, J. A. Saugstad, K. Van Keuren-Jensen, A. M. Weaver, L. C. Laurent, The Extracellular RNA Communication Consortium: Establishing Foundational Knowledge and Technologies for Extracellular RNA Research. *Cell* **177**, 231-242 (2019).
37. L. Antoury, N. Hu, L. Balaj, S. Das, S. Georghiou, B. Darras, T. Clark, X. O. Breakefield, T. M. Wheeler, Analysis of extracellular mRNA in human urine reveals splice variant biomarkers of muscular dystrophies. *Nat. Commun.* **9**, 3906 (2018).
38. Y. T. Wang, T. Shi, S. Srivastava, J. Kagan, T. Liu, K. D. Rodland, Proteomic Analysis of Exosomes for Discovery of Protein Biomarkers for Prostate and Bladder Cancer. *Cancers (Basel)* **12**, 2335 (2020).
39. J. Linxweiler, K. Junker, Extracellular vesicles in urological malignancies: an update. *Nat. Rev. Urol.* **17**, 11–27 (2020).
40. Y. Cheng, X. Wang, J. Yang, X. Duan, Y. Yao, X. Shi, Z. Chen, Z. Fan, X. Liu, S. Qin, X. Tang, C. Zhang, A translational study of urine miRNAs in acute myocardial infarction. *J. Mol. Cell Cardiol.* **53**, 668-676 (2012).

41. K. B. Fraser, A. B. Rawlins, R. G. Clark, R. N. Alcalay, D. G. Standaert, N. Liu, Ser(P)-1292 LRRK2 in urinary exosomes is elevated in idiopathic Parkinson's disease. *Mov Disord.* **31**, 1543-1550 (2016).
42. S. Wang, K. Kojima, J. A. Mobley, A. B. West, Proteomic analysis of urinary extracellular vesicles reveal biomarkers for neurologic disease. *EbioMedicine* **45**, 351-361(2019).
43. P. Simeone, G. Bologna, P. Lanuti, L. Pierdomenico, M. T. Guagnano, D. Pieragostino, P. Del Boccio, D. Vergara, M. Marchisio, S. Miscia, R. Mariani-Costantini, Extracellular Vesicles as Signaling Mediators and Disease Biomarkers across Biological Barriers. *Int. J. Mol. Sci.* **21**, 2514 (2020).
44. J. McKiernan, M. Noerholm, V. Tadiogola, S. Kumar, P. Torkler, G. Sant, J. Alter, M. J. Donovan, J. Skog, A urine-based Exosomal gene expression test stratifies risk of high-grade prostate Cancer in men with prior negative prostate biopsy undergoing repeat biopsy. *BMC Urol.* **20**, 138 (2020).
45. Y. Lu, D. Liu, Q. Feng, Z. Liu, Diabetic Nephropathy: Perspective on Extracellular Vesicles. *Front. Immunol.* **11**, 943 (2020).
46. A. H. Gheinani, M. Vögeli, U. Baumgartner, E. Vassella, A. Draeger, F. C. Burkhard, K. Monastyrskaya, Improved isolation strategies to increase the yield and purity of human urinary exosomes for biomarker discovery. *Sci. Rep.* **8**, 3945 (2018).
47. J. T. Leek, R. B. Scharpf, H. C. Bravo, D. Simcha, B. Langmead, W. E. Johnson, D. Geman, K. Baggerly, R. A. Irizarry, Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **11**, 733-9 (2010).
48. R. Z. Alicic, M. T. Rooney, K. R. Tuttle, Diabetic Kidney Disease: Challenges, Progress, and Possibilities. *Clin. J. Am. Soc. Nephrol.* **12**, 2032-2045 (2017).
49. R. A. DeFronzo, W. B. Reeves, A. S. Awad, Pathophysiology of diabetic kidney disease: impact of SGLT2 inhibitors. *Nat. Rev. Nephrol.* **17**, 319-334 (2021).
50. J. Park, R. Shrestha, C. Qiu, A. Kondo, S. Huang, M. Werth, M. Li, J. Barasch, K. Suszták, Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* **360**, 758-763 (2018).
51. J. R. Drevet, The antioxidant glutathione peroxidase family and spermatozoa: a complex story. *Mol. Cell Endocrinol.* **250**, 70-9 (2006).
52. G. V. Kryukov, S. Castellano, S. V. Novoselov, A. V. Lobanov, O. Zehtab, R. Guigó, V. N. Gladyshev, Characterization of mammalian selenoproteomes. *Science* **300**, 1439-43 (2003).
53. H. Lai, T. Nie, Y. Zhang, Y. Chen, J. Tao, T. Lin, T. Ge, F. Li, H. Li, Selenium Deficiency-Induced Damage and Altered Expression of Mitochondrial Biogenesis Markers in the Kidneys of Mice. *Biol. Trace. Elem. Res.* **199**, 185-196 (2021).
54. K. Zitouni, M. R. C. P. Steyn, E. Lyka, F. J. Kelly, P. Cook, I. C. Ster, K. A. Earle, Derepression of glomerular filtration, renal blood flow and antioxidant defence in patients with type 2 diabetes at high-risk of cardiorenal disease. *Free Radic. Biol. Med.* **161**, 283-289 (2020).
55. Q. Yang, F. R. Wu, J. N. Wang, L. Gao, L. Jiang, H. D. Li, Q. Ma, X. Q. Liu, B. Wei, L. Zhou, J. Wen, T. T. Ma, J. Li, X. M. Meng, Nox4 in renal diseases: An update. *Free Radic. Biol. Med.* **124**, 466-472 (2018).
56. R. D. Rajaram, R. Dissard, A. Faivre, F. Ino, V. Delitsikou, V. Jaquet, T. Cagarelli, M. Lindenmeyer, P. Jansen-Duerr, C. Cohen, S. Moll, S. de Seigneux, Tubular NOX4 expression decreases in chronic kidney disease but does not modify fibrosis evolution. *Redox Biol.* **26**, 101234 (2019).
57. H. Jing, S. Tang, S. Lin, M. Liao, H. Chen, J. Zhou, The role of extracellular vesicles in renal fibrosis. *Cell Death Dis.* **10**, 367 (2019).
58. D. Burger, J. F. Thibodeau, C. E. Holterman, K. D. Burns, R. M. Touyz, C. R. Kennedy, Urinary podocyte microparticles identify prealbuminuric diabetic glomerular injury. *J. Am. Soc. Nephrol.* **25**, 1401-7 (2014).
59. X. M. Wu, Y. B. Gao, F. Q. Cui, N. Zhang, Exosomes from high glucose-treated glomerular endothelial cells activate mesangial cells to promote renal fibrosis. *Biol. Open.* **5**, 484-91 (2016).
60. D. Yang, M. J. Livingston, Z. Liu, G. Dong, M. Zhang, J. K. Chen, Z. Dong, Autophagy in diabetic kidney disease: regulation, pathological role and therapeutic potential. *Cell Mol. Life. Sci.* **75**, 669-688 (2018).
61. Y. Ding, M. E. Choi, Autophagy in diabetic nephropathy. *J. Endocrinol.* **224**, R15-30 (2015).
62. N. P. Hessvik, A. Øverbye, A. Brech, M. L. Torgersen, I. S. Jakobsen, K. Sandvig, A. Llorente, PIKfyve inhibition increases exosome release and induces secretory autophagy. *Cell Mol. Life Sci.* **73**, 4717-4737 (2016).
63. J. M. Gudbergsson, K. B. Johnsen, Exosomes and autophagy: rekindling the vesicular waste hypothesis. *J. Cell Commun. Signal* **13**, 443-450 (2019).
64. A. M. Leidal, H. H. Huang, T. Marsh, T. Solvik, D. Zhang, J. Ye, F. Kai, J. Goldsmith, J. Y. Liu, Y. H. Huang, T. Monkkonen, A. Vlahakis, E. J. Huang, H. Goodarzi, L. Yu, A. P. Wiita, J. Debnath, The LC3-conjugation machinery specifies the loading of RNA-binding proteins into extracellular vesicles. *Nat. Cell. Biol.* **22**, 187-199 (2020).

65. M. Kanai, M. Akiyama, A. Takahashi, N. Matoba, Y. Momozawa, M. Ikeda, N. Iwata, S. Ikegawa, M. Hirata, K. Matsuda, M. Kubo, Y. Okada, Y. Kamatani, Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390-400 (2018).
66. A. S. Levey, K. U. Eckardt, Y. Tsukamoto, A. Levin, J. Coresh, J. Rossert, D. De Zeeuw, T. H. Hostetter, N. Lameire, G. Eknoyan, Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int.* **67**, 2089-100 (2005).
67. M. Puhka, M. Takatalo, M. E. Nordberg, S. Valkonen, J. Nandania, M. Aatonen, M. Yliperttula, S. Laitinen, V. Velagapudi, T. Mirtti, O. Kallioniemi, A. Rannikko, P. R. Siljander, T. M. Af Hällström, Metabolomic Profiling of Extracellular Vesicles and Alternative Normalization Methods Reveal Enriched Metabolites and Strategies to Study Prostate Cancer-Related Changes. *Theranostics* **7**, 3824-3841 (2017).
68. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
69. J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J. M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigó, T. J. Hubbard, GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760-74 (2012).
70. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
71. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
72. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).
73. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8 (2001).

Acknowledgements: We acknowledge the Sequencing Units of FIMM Tech Center of the University of Helsinki, Finland, and the Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmö, Sweden, for RNAseq, and Biomedicum Functional Genomics Unit of the University of Helsinki, Finland, for the qPCR service, the EV Core facility of the University of Helsinki, Finland, for performing the NTA. The skillful assistance of Carola Pähls, Paula Kokko, Anna Sandelin and Jaana Tuomikangas is gratefully acknowledged. We also thank Olli Veikkola for technical assistance and Olli Kallioniemi group for laboratory facilities.

Funding: This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115974 (BEAt-DKD). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA with JDRF. The Direva Study (A.K.) has been supported by the Vasa Hospital District, State Research Funding via the Turku University Hospital, Vasa Central Hospital, Jakobstadsnejdens Heart Foundation, the Medical Foundation of Vaasa and the Finnish Medical Foundation. This

substudy has also been financially supported (LG, TT) by grants from Folkhälsan Research Foundation, the Sigrid Juselius Foundation, the Academy of Finland (grants no. 263401, 267882, 312063, 336822 to LG; 312072, 336826 to TT, 317599 to OPD), University of Helsinki, Ollqvist Foundation, and the Health Care Center in Vasa. The FinnDiane study was supported by grants from Folkhälsan Research Foundation, Wilhelm and Else Stockmann Foundation, Liv och Hälsa Society, Helsinki University Central Hospital Research Funds (EVO), Novo Nordisk Foundation (NNFOC0013659), and Academy of Finland (#299200 and #316664).

Author contributions: Conceptualization: O.P.D., M.P., H.H., P-H.G., C.F., T.T., and L.G.; methodology (sample and data collection management): T.T., C.F., O.P.D., M.P., A.K., E.V. and R.B.P.; methodology (EV sample processing): M.P. and K.B.; methodology (data analysis): O.P.D., M.P., K.B., C.F., and T.T.; supervision: M.P., O.P.D., C.F., T.T. and L.G.; writing (original draft): O.P.D. and M.P.; writing (review and editing): T.T., A.K., C.F., K.B., O.P.D. and M.P.

Competing interests: L.G. has received research funding from Pfizer Inc, Regeneron Pharmaceuticals, Eli Lilly and Astra Zeneca. P-H.G. has received research grants from Eli Lilly and Roche; is an advisory board member for AbbVie, AstraZeneca, Boehringer Ingelheim, Cebix, Eli Lilly, Jansen, MSD, Novartis, NovoNordisk and Sanofi; and has received lecture fees from Boehringer Ingelheim, Eli Lilly, Elo Water, Genzyme, MSD, Novartis, Novo Nordisk and Sanofi. The funding sources were not involved in the design or conduct of the study.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Fig. 1: Study design.

Figure depicting study design and work flow for next generation sequencing of uEV mRNAs and discovery of candidate markers for T1D DKD.

24 hour (24h); albumin excretion rate (AER); blood pressure (BP); body mass index (BMI); diabetic kidney disease (DKD); estimated glomerular filtration rate (eGFR); Genotype-Tissue Expression (GTEx) project (30); glycated hemoglobin (HbA_{1c}); macroalbuminuria (macro); microalbuminuria (micro); normoalbuminuria (normo); messenger RNA sequencing (mRNAseq); principal component analysis (PCA); proximal convoluted tubule (PCT); quantitative polymerase chain reaction (qPCR); type 1 diabetes (T1D); urinary extracellular vesicles (uEV); waist-to-hip ratio (WHR).

Fig. 2: Quality control of the uEV preparations.

(A and B), Transmission electron micrographs showing uEV of typical morphology and variable sizes; filamentous structures compatible with Tamm–Horsfall protein filaments are present in the in-set A1. (C), Representative NTA profiles showing enrichment of vesicle sizes between 50–300 nm. (D), Western blotting showing the presence of uEV-enriched markers CD9 and PDX in uEV preparations. (E), Representative electropherograms of RNA isolated from uEV and analyzed with Bioanalyzer using Agilent Pico kit. A typical RNA peak between 25–200 nt was detected in all samples.

Lymph Node Carcinoma of the prostate (LnCaP), macroalbuminuria (macro); microalbuminuria (micro); nanoparticle tracking analysis (NTA); non-diabetic control (control); normoalbuminuria (normo); Podocalyxin (PDX), urinary extracellular vesicles (uEV).

Fig. 3: Sequencing of uEV mRNAs showed consistent gene expression distribution and reproducibility across different sample sets.

(A), Density plot depicting gene expression distribution (log2CPM) for all commonly expressed protein coding genes (count ≥ 1 in $\geq 90\%$ samples, $n=10,596$) in urine samples from the T1D cohorts (total $n=72$; 24h (24 hour) collections in blue, $n=34$; ON (overnight) collections in red, $n=38$). (B), Principal component (PC) analysis using gene expression data for all commonly expressed genes ($n=10,596$) in the T1D cohorts ($n=72$).

(C to L), Hierarchical clustering based on gene expression correlation matrix (Spearman's) using all commonly expressed genes ($n=10,449$, in C, E, G, I and K) or kidney enriched genes ($n=233$, in D, F, H, J and L) for the sample sets included in technical comparisons: (C to F), Technical replicates (C-D, T1D, and E-F, non-diabetic donors). Aliquots of individual samples (24h urine collection) were processed through the uEV mRNA sequencing pipeline twice (R1 and R2, $n=6$ donors, T1D: Samples 1, 2, 5 and 6; Non-diabetic: Samples 10 and 11) or thrice (R1, R2 and R3, $n=2$ non-diabetic donors: Samples 12 and 13) at different time points. (G and H), Centrifugation test samples (T1D donors). Urine samples (24h collection) were processed with (_C) and without centrifugation before freezing ($n=4$ T1D donors: Samples 2, 6, 8 and 9). (I to L), Samples from ON

vs 24h urine collections (**I-J**, T1D, and **K-L**, non-diabetic donors). Samples from individual donors were collected from the same day (n=12 donors, T1D: Samples 1, 3-9; Non-diabetic: Samples 10-13). The gene expression values (log2CPM) were inverse normally transformed and converted to z score unit (sd unit) before analysis.

Counts per million (CPM); principal component (PC); replicate (R); sample (Sam); type 1 diabetes (T1D); urinary extracellular vesicles (uEV).

Fig. 4: Genome-wide mRNA profile of uEV showed similarity with the kidney mRNA profile. (A), The boxplot depicts the interquartile range, median and minimum/maximum summary values of expression levels of the 77 known specific kidney enriched genes that are reported to be detected in kidney and only in some human tissues (see Methods). (B to D), PC analysis using gene expression levels (log2CPM) in uEV from the T1D cohorts (n=72), and the respective gene expression levels (log2CPM) in 17 human tissues obtained from the GTEx reference database (30, 31) (tissues from men only; Adipose=692, Bladder=9, Blood=426, Brain=1612, Colon=475, Esophagus=866, Heart=515, Kidney Cortex=49, Liver=138, Lung=327, Muscle=474, Pancreas=189, Prostate=140, Skin=1088, Small Intestine=118, Spleen=132, Testis=216). (B), all commonly expressed protein coding genes as detected in uEV and GTEx tissues (n=10,350); (C), a subset of “kidney depleted genes”, i.e. genes normally not expressed in kidney (n=200 present in uEV+GTEx combined dataset out of 4,631, details in Methods); (D), “specific kidney enriched genes”, a subset of kidney enriched genes (n=77, as shown in A) that are reported to be detected in kidney and only in some other human tissues (see Methods). Counts per million (CPM); genotype-Tissue Expression (GTEx); principal component (PC); type 1 diabetes (T1D); urinary extracellular vesicles (uEV).

Fig. 5: Global differential gene expression analysis revealed upregulation of 13 transcripts in the uEV of type 1 diabetic individuals with macroalbuminuria.

(A and B), Global DGE analysis (n=10,596 genes) comparing macro- (n=17) and normoalbuminuric T1D (n=37) individuals, depicts significantly ($p \leq 3.85 \times 10^{-6}$) differentially expressed transcripts (red points in A; names and respective logFC in B). (C), The box plots illustrate the gene expression distribution (sd unit) of 13 differentially expressed genes in macro- and normoalbuminuric T1D subjects (as in A) along with the group of microalbuminuric T1D patients (n=13).

The DGE analysis (Normo vs Macro) in A and B was performed based on count data using generalized linear models adjusting for age, body-mass index, diabetes duration and urine collection protocols (overnight and 24 hour) and multiple correction using the Bonferroni method. In C, the three groups comparisons were performed using one-way ANOVA. The gene expression values (log2CPM) were inverse normally transformed and converted to z score unit (sd unit). The boxplots depict the interquartile range, median and minimum/maximum summary values.

Differential gene expression (DGE); fold change (FC); log2 normalized counts per million (log2CPM); macroalbuminuria (Macro); microalbuminuria (Micro); normoalbuminuria (Normo); type 1 diabetes (T1D); urinary extracellular vesicles (uEV).

Fig. 6. Validation of mRNA sequencing -based differentially expressed genes using quantitative PCR.

Differential expression of five of the top 13 DKD associated transcripts of uEV (*MAP7*, *GPX3*, *IL32*, *NOX4*, *HRSP12*; Fig. 5) was validated by qPCR. (A), Correlation of the gene expression level by next generation mRNA sequencing (x-axis) vs. qPCR (y-axis) in the same samples ($n=57-66$). DeltaCt values from the qPCR experiments and log2 counts per million values from RNA sequencing experiments were inverse normally transformed and converted to z score unit (sd unit); r =non-parametric Spearman's (two-tailed) method; * $p<0.006$, ** $p<0.00001$. (B), The boxplots (interquartile range, median and minimum/maximum summary values) show the gene expression level derived from qPCR assays using 1-2 Taqman assays (assay1 and assay 2) per gene in T1D subjects stratified for the degree of albuminuria (normo- ($n=29-36$), micro- ($n=13$), macroalbuminuria ($n=15-17$). The gene expression values (deltaCt values from qPCR) were inverse normally transformed and converted to z score unit (sd unit) and statistical comparisons of the three groups were performed using one-way ANOVA. Diabetic kidney disease (DKD); macroalbuminuria (Macro); microalbuminuria (Micro); normoalbuminuria (Normo); quantitative PCR (qPCR); type 1 diabetes (T1D); uEV; urinary extracellular vesicles (uEV).

Fig. 7: Expression level of the top differentially expressed uEV transcripts correlated with hyperglycemia.

(A), Heatmap depicting the strength of the correlation between expression level of the 13 differential genes in uEV with eight continuous clinical variables in the T1D individuals (n=65-67 samples, including 35-37 with normo-, 13 with micro- and 17 with macroalbuminuria). The 13 genes were obtained from differential expression analysis of uEV between macro- and normoalbuminuric T1D groups (Fig. 5). (B and C), Correlation between expression level of the 13 genes and HbA1c level (%) in all T1D subjects (B, n=67), and in normoalbuminuric T1D subjects only (C, n=37). HbA1c was measured close to the time of urine collection. The gene expression values (log2CPM) were inverse normally transformed and converted to z score units (sd unit). All data were transformed using inverse rank transformation before Spearman's correlation analysis (two tailed).

Body mass index (kg/m^2) (BMI); diabetes duration (Duration); diastolic blood pressure (DBP); estimated glomerular filtration rate (eGFR); glycated hemoglobin (HbA1c); log2 normalized counts per million (log2CPM); systolic blood pressure (SBP); type 1 diabetes (T1D); urinary extracellular vesicles (uEV); waist-to-hip ratio (WHR).

Fig. 8: Resemblance in gene expression signature of the top DKD candidate genes between uEV and human kidney tissue.

(A), Expression levels (interquartile range, median and minimum/maximum summary values) of the genes- shown to be differentially expressed in uEV from macro- vs normoalbuminuric patients (12 out of 13 genes depicted)- in male urogenital organs obtained from the GTEx reference database^{30,31} (tissues from men only; kidney-cortex, n=49; prostate, n=140; bladder, n=9; testis, n=216). (B to E), Pairwise gene expression correlation (Spearman's correlation analysis) among the differentially expressed genes (n=12) in male urogenital organs (from GTEx, (B)-kidney-cortex, n=49; (D)-prostate, n=140; (E)-testis, n=216) and uEV (C, n=72, both T1D cohorts). (F), Heatmap showing fraction of specified kidney cells expressing the nine differentially expressed genes. Data is from single cell sequencing of human diabetic kidney and obtained from Wilson PC et al.(34). The assessed kidney cell types were: PCT, proximal convoluted tubule; CFH, complement factor H; LOH, loop of Henle; DCT, distal convoluted tubule; CT, connecting tubule; CD, collecting duct; PC, principal cell; IC, intercalated cell; PODO, podocyte; ENDO, endothelium; MES, mesangial cell; LEUK, leukocyte.

Correlation (Corr.); counts per million (CPM); diabetic kidney disease (DKD); genotype-Tissue Expression (GTEx); principal component (PC); transcripts per million (TPM); urinary extracellular vesicles (uEV).

Fig. 9: Stress score based on differential uEV mRNAs stratified T1D individuals differing in decline rate of kidney function.

(A), Transcriptomic stress scores in uEV from T1D individuals stratified for their albuminuria status: normo- (n=37), micro- (n=13) and macroalbuminuria (n=17). (B), Correlation of HbA1c, eGFR and blood pressure (SBP and DBP) level with the transcriptomic stress score in all T1D individuals (n=67; clinical data and gene expression measured at urine collection time). (C and D), Sequential laboratory data were retrospectively obtained from clinical databases for T1D individuals (n=51) who were stratified into quartiles based on the uEV transcriptomic stress score (n=12-13 per quartile); (C), change of eGFR over 15 years (mean/year and slope) according to the transcriptomic stress score quartiles; and (D), change of eGFR over 15 years in normoalbuminuric individuals within the lowest quartile ($\leq 25\%$, n=11) or higher quartiles ($> 25\%$, n=19) of the stress score. For comparison, data for all microalbuminuric individuals (n=10,) with available retrospective eGFR data and without any stress score stratification are shown as a positive control; (E), change of HbA1c over 13 years (mean/year and slope) as described in D. (F), distribution of clinical data measured at the time of uEV collection among the microalbuminuric individuals (n=13) and normoalbuminuric individuals stratified for the stress score (lowest quartile, n=15; two middle quartiles, n=22).

Data were analysed by the Kruskal-Wallis test (A), Spearman's correlation analysis (two tailed, B), linear regression analysis (C to E) or Mann-Whitney test (F). Data are in Mean $\pm$ SEM (A), mean (C to E) or boxplot with interquartile range, median and summary minimum/maximum values (F).

Albumin excretion rate (AER); estimated glomerular filtration rate (eGFR); glycated hemoglobin (HbA1c); diabetes duration (Duration); diastolic blood pressure (DBP); macroalbuminuria (Macro); microalbuminuria (Micro); normoalbuminuria (Normo); systolic blood pressure (SBP); type 1 diabetes (T1D); urinary extracellular vesicles (uEV).