Circulating small non-coding RNAs associate with kidney function in people with type 2 diabetes

LM 't Hart[1,2,3,4], GA Bouland[1], JA de Klerk[1], M.T. Blom[3,5], SJ Cramer[1], JWJ Beulens[2,3], RC Slieker[1,2,3]

[1] Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, The Netherlands

[2] Department of Epidemiology and Data Sciences, Amsterdam University Medical Center, VU University Amsterdam, Amsterdam, The Netherlands

[3] Amsterdam Public Health, Amsterdam, The Netherlands

[4] Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[5] Department of General Practice and Elderly Care Medicine, Amsterdam University Medical Center location VUmc, Amsterdam, The Netherlands

Author for correspondence:
LM 't Hart, PhD
Leiden University Medical Center
Department of Cell and Chemical Biology
Albinusdreef 2
2333ZA Leiden, The Netherlands
+31 715269796
lmthart@lumc.nl

Word count:
Abstract: 248 words
Body: 3725 words (excl references)

Abstract

Aims/hypothesis Micro- and macrovascular complications are common among persons with type 2 diabetes. Interest into the potential of circulating small non-coding RNAs (sRNAs) as biomarkers or drivers of the development of diabetic complications is growing. In this study we investigated if circulating sRNAs levels associate with prevalent chronic kidney disease (CKD) in persons with type 2 diabetes.

Methods Plasma sRNAs levels were determined using sRNA-seq, allowing detection of miRNAs, snoRNAs, piRNAs, tRNA-fragments and various other sRNA classes, in persons with type 2 diabetes, with CKD (n=69) or without CKD (n=405). Multiple regression analyses were used to test for associations between the sRNAs and primary endpoint CKD. In secondary analyses, we also tested the association with eGFR and albuminuria (UACR).

Results In total, twelve sRNA were associated with prevalent CKD (logFC = -0.32 to -1.28, all $P \leq 6.24 \times 10^{-4}$). Although miRNAs represent the majority of the sRNAs measured (64%) only four miRNA were significantly associated with prevalent CKD. Interestingly, the majority of the significant sRNA belonged to the snoRNAs (58%). Similar results were observed for eGFR and UACR. Only one of the twelve sRNAs showed its highest expression in kidney whereas several others showed high expression in liver or colon.

Conclusions/interpretation Small RNAs present in the circulation associate with CKD in persons with type 2 diabetes. Further studies are warranted to elucidate the biological role of sRNA in diabetic CKD and in particular snoRNAs. High expression of these circulating sRNAs in tissues other than kidney suggest a role in inter-organ communication.

Research in context

- *What is already known about this subject? (maximum of 3 bullet points)*
  miRNAs present in the circulation have been associated with CKD in diabetes.
  The role of other small non-coding RNAs, for example snoRNAs, have not been studied.
- *What is the key question? (one bullet point only; formatted as a question)*
  Are circulating small RNAs, such as miRNA, snoRNAs, tRNA fragments and piRNAs associated with CKD in persons with type 2 diabetes (n=475)?
- *What are the new findings? (maximum of 3 bullet points)*
  Twelve small RNAs associate with prevalent CKD in type 2 diabetes.
  Especially snoRNAs are more frequently associated with CKD.
  Identified circulating small RNAs are often expressed in tissues other than kidney, which suggest a role for these small non-coding RNAs in inter-organ communication.
- *How might this impact on clinical practice in the foreseeable future? (one bullet point only)*
  Circulating small RNAs and especially snoRNAs might represent novel mechanisms and therapeutic targets for CKD in diabetes.

Introduction

Chronic kidney disease (CKD) in diabetes is one of the most common complications in type 2 diabetes. Nonetheless, while some develop complications early in the disease, others will never. The mechanisms underlying CKD are not fully understood (1). Many groups including our own (2), have investigated biomarkers such as genomics, metabolomics, transcriptomics and proteomics data (3; 4). This research has improved our knowledge about the mechanisms underlying CKD development in T2D and have yielded novel biomarkers and drug targets. Some of these biomarkers such as a proteomic score are currently tested for clinical use (3; 4). A novel class of potential biomarkers are the circulating small non-coding RNAs (sRNA) which are mostly contained in extracellular vesicles or exosomes or are bound to lipoproteins and originate from various tissues including tissues important in kidney function (5-7). Whereas it was previously thought that sRNAs merely reflect degradation by-products, it has now been shown that they exert important functions in cell-to-cell communication, the regulation of gene expression and ribosomal RNA maturation (5-7). The composition of sRNAs in the circulation is dependent on tissue homeostasis and can be altered by disease development. Irrespective of the mechanism involved, sRNAs may have an impact on the development and progression of CKD when deregulated. Importantly, each tissue releases its own combination of tissue-specific and more ubiquitous sRNAs into the bloodstream. These circulating sRNAs can be taken up by other tissues and affect target tissue function (8).

Previous studies, focused only on miRNAs, identified several miRNAs associated with kidney functions (9-11). In this study we examined, for the first time, all circulating sRNAs, including miRNAs, piRNAs, tRNA fragments (tRFs) and snoRNAs in relation to prevalent CKD. We use next-generation sequencing to capture the whole sRNA transcriptome allowing for a hypothesis-free, comprehensive analysis to study the role of sRNAs in diabetic kidney disease (12).

Materials and Methods

The study has been approved by the ethics committee of the Amsterdam University Medical Center, location VUmc. All participants provided informed consent prior to this study. A schematic overview of the study design can be found in Supplemental Figure 1.

Study sample

For this project we used data from the Hoorn Diabetes Care System cohort (DCS), a large prospective study comprised of over 14,000 individuals with type 2 diabetes who receive structed diabetes care. (13) Participants visited the Hoorn diabetes research center annually for diabetes care including routine biochemistry, medication review and clinical care. Patients were also invited to participate in DCS biobanks in which we combine the annual phenotypic data with biobanking of liquid biopsies which includes serum, plasma and urine. Blood samples were collected during two phases of biobanking in 2008/2009 and 2012-2014. For the current study we used data collected for a previous case-cohort study

102 investigating sRNAs in relation to insulin initiation ('t Hart et al. in preparation). In short the study
103 included 145 cases of incident insulin use and 330 random controls all with a plasma sample collected
104 within four years after diagnosis of diabetes and no-insulin use at the time of blood sampling
105 All laboratory measurements were on samples taken in a fasted state on the same day as the
106 sample for sRNA-seq. Routine anthropometric, biochemistry methods and definitions have been
107 described previously in detail in van der Heijden et al. (13). Kidney function was assessed based on
108 eGFR (CKD-EPI) and urinary albumin to creatinine ratio (UACR) and included as continuous variables
109 in our analyses. Stages of chronic kidney disease (CKD) were defined according the KDIGO criteria as
110 previously described (14; 15). Cases were defined as having CKD stage ≥ 1 at the time of blood sampling
111 for sRNA analysis which resulted in 69 cases.
112
113 *Circulating small RNA sequencing*
114 Circulating sRNAs were isolated from 900ul of citrate plasma using a commercial cell-free RNA
115 isolation kit (Quick-cfRNA Serum & Plasma Kit, Zymo Research, Irvine, CA, USA). After quality
116 control 475 samples were submitted for sRNA sequencing. Library preps were constructed using the
117 NEBNext® Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs, Ipswich, MA,
118 USA). A Pippin Prep instrument (Sage Science, Beverly, MA,USA) was used for size selection of
119 fragments ranging from 120 to 200 nucleotides to further enrich for sRNAs with a size of ~20 to 100
120 nucleotides. Sequencing was performed using the Illumina NextSeq500 sequencer with v2.5 sequencing
121 reagent kits and 35 cycle paired-end reads (PE35). On average 11.3±2.8 million reads per sample were
122 generated. The exceRpt pipeline (16), developed by the NIH Extracellular RNA Communication
123 Consortium (ERCC), was used to map the obtained reads, for quality control, to process the data and
124 generate RNA abundance estimates for miRNA, piRNA, lncRNA, tRNA fragments, Y_RNA
125 (fragments), snRNA, snoRNA, scaRNA and various other sRNA species. Using respectively miRBase
126 v22(17), piRNABank v1(18), Gencode (version 38)(19), circBase(20) and GtRNAdb(21). One sample
127 did not pass the QC thresholds.
128
129 *Statistical analysis*
130 The R-package edgeR was used to identify differentially expressed sRNAs (22). A negative binomial
131 generalized log-linear model was applied using edgeR's function *estimateGLMRobustDisp*. The base
132 model was adjusted for age and sex. A fully adjusted model was additionally adjusted for BMI, diabetes
133 duration, HbA1c, HDL, SBP and smoking unless otherwise stated. UACR was log10-transformed
134 before analysis. In sensitivity analyses we additionally included use of antihypertensive medication,
135 eGFR or UACR and case-control status. Where possible missing data were imputed from previous
136 measurements taken within one year of the blood sample for RNA isolation (n_{max}=17). The Benjamini-
137 Hochberg procedure was used to adjust for multiple testing and an FDR below 0.05 was considered
138 significant.

139

140 *Expression patterns of small RNAs in tissues*

141 Publicly available raw sRNA-seq data for various metabolic tissues were downloaded from the Gene
142 Expression Omnibus (GEO) (23). The tissues for which sRNA-seq data was used included kidney,
143 thyroid, brain, liver, muscle, heart, colon, pancreas, subcutaneous white adipose tissue and whole blood
144 and urine. Accesion numbers are given in Supplemental Table 1. Data were processed using the exceRpt
145 pipeline as described above. The tissue data were used to estimate the abundance of the sRNAs in these
146 tissues. Data was z-scaled in figures.

147

148 **Results**

149 The main clinical characteristics of our study group are shown in Table 1. The characteristics resemble
150 those of a typical T2D population with slightly more males than females, a mean age-at-diagnosis of
151 59.3 ± 9.0 years, BMI of 30.7 ± 6.0 kg/m2 and an average duration of diabetes at time of blood sampling
152 of 2.2 ± 1.1 years. Glycemic control was good with an average HbA1c of 46 (42, 51) mmol/mol.
153 Prevalence of CKD at the time of blood sampling was 15%.

154 After quality control 938 sRNAs were expressed above the threshold of a median ≥5 copies in
155 our data, which included miRNAs, piRNA, Y_RNA fragments, lncRNAs, tRNA_fragments, snoRNAs,
156 snRNAs and various other sRNA species. (Table 2). Twelve sRNAs were significantly associated with
157 prevalent CKD (P_{FDR}≤0.05, Table 3, Figure 1A). All identified sRNAs were down-regulated in CKD
158 cases. The top hit was SNORD118 (log2 fold change=-0.873, P_{FDR}=2.35x10^{-3}, Figure 1B). Respectively,
159 73 and 209 sRNAs were significantly associated with eGFR and UACR, on which the CKD
160 classification is based (Table 3, Supplemental Table 2, Supplemental Figures 2 and 3). The overlap
161 between the sRNAs associated with CKD and its components eGFR and UACR resulted both in unique
162 and overlapping sRNAs (Table 3, Figure 1C). Three sRNAs overlapped between all three endpoints
163 used (U8, SNORD107 and SNORD25), whereas one overlapped between eGFR and CKD (miR-145-
164 5p) and six between UACR and CKD (SNORD118, SNORD24, SNORD87, miR-143-3p, SNORD15B,
165 miR-4685-3p).

166 In the full model, including adjustment for age, sex, BMI, diabetes duration, HbA1c, HDL, SBP
167 and smoking, respectively 3, 53 and 183 sRNAs remained significantly associated with CKD, eGFR
168 and UACR (Supplemental Table 3, Supplemental Figure 4; 100%; 91% and 86% overlap for CKD,
169 eGFR and UACR respectively). In a sensitivity analysis, additional adjustment for use of
170 antihypertensive medication and baseline eGFR or UACR (only for eGFR and UACR) did not have a
171 major impact on the effect sizes but reduced the number of significant associations (r^2>=0.94,
172 Supplemental Figure 5). In addition adding case-control status for incident insulin use only marginally
173 affected the results (data not shown).

174 Comparing the frequency of different types of sRNAs identified showed that while snoRNAs
175 represent 10% of the total sRNA transcriptome in our RNA-seq study they are the most abundant class

176 associated with kidney function (≥31%, Table 2). A similar picture is observed in the full models
177 (Supplemental Table 4).
178 The pairwise correlation between the 12 sRNAs associated with diabetic CKD was very modest
179 (Figure 1D). The only exception were U8 and SNORD118, which is likely due to the fact that they are
180 encoded in the same host gene (TMEM107) on chromosome 17. Results of these analyses for eGFR and
181 UACR are shown in Supplemental Figure 6.
182 In a next step we examined the tissue-specific expression patterns of the identified sRNAs using
183 publicly available expression data in 11 tissues or body fluids (Figure 1E). This illustrated that many
184 sRNAs show tissue-specific expression patterns. For the 12 sRNAs associated with CKD only mir-143-
185 3p showed its highest expression in kidney tissue suggestive of a direct role in kidney function. piRNA-
186 019825 was the only sRNA that showed high expression in urine. Interestingly, several sRNAs were
187 highly expressed in other tissues such as liver and colon. This suggest that these sRNAs might be
188 involved in inter-organ communication in the regulation of kidney function. It is further interesting that
189 several sRNA show their highest expression in liver or colon. Similar observations were made for eGFR
190 and UACR (Supplemental Figure 7).
191
192 **Discussion**
193 In this study we show that circulating sRNAs are associated with prevalent CKD in people with type 2
194 diabetes. In addition, we show that circulating sRNAs are associated with the two components of CKD:
195 eGFR and UACR. Importantly these sRNAs do not only represent miRNAs but also other classes of
196 small non-coding RNAs, particularly snoRNAs. Using publicly available data we show that several of
197 these are highly expressed in colon or liver but not kidney. This suggests that circulating sRNAs may
198 be involved inter-organ crosstalk to regulate kidney function or reflect differences in clearance of
199 sRNAs/vesicles from the circulation.
200
201 **sRNA tissue specific expression**
202 Interestingly, only one of the significant sRNAs was found at high levels in urine showing that the
203 circulating sRNA transcriptome and the sRNA urine transcriptome are not similar or interchangeable.
204 Besides one sRNA highly expressed in kidney (miRNA-143-3p) we found that many significant sRNAs
205 are overexpressed in one or a few other tissues we tested, where especially liver and colon seem to be
206 overrepresented among significant sRNAs. We speculate that these sRNAs represent interesting
207 candidates for further studies as they are potentially involved in organ crosstalk in the regulation of and
208 or response to kidney dysfunction. On the other hand it cannot be excluded that clearance of (specific)
209 sRNAs or vesicles is affected by the disease.
210
211 **Associations between kidney function and sRNAs**

212 Although a substantial number of studies have examined circulating sRNAs associated with kidney
213 function in persons with diabetes, most studies are relatively small (n<100) and focused only on
214 microRNAs. From these studies a large number of miRNAs have emerged that are putatively associated
215 with kidney function. However, probably due to small sample sizes, heterogeneity in endpoints and
216 methodology used there is limited consistency between studies as shown in various meta-analyses (11;
217 25; 26). We have performed a large untargeted sRNA-seq study in a large well defined sample of persons
218 with T2D with and without CKD from the Hoorn DCS study. Although miRNAs represent 64% of all
219 sRNAs detected in our study we found that for the outcomes studied only between 33 and 44% of the
220 significant sRNAs belong to the miRNAs. Highlighting the potential of the other classes of sRNAs in
221 the etiology of CKD in diabetes. Of the twelve sRNAs significantly associated with CKD four were
222 miRNAs, seven belonged to the snoRNAs and we also identified a single piRNA.
223
224 *miRNAs*
225 MicroRNAs are involved in the post-transcriptional regulation of gene expression and are known to be
226 involved in crosstalk between (distant) cells and tissues (5; 6; 10). As such, there is an ever growing
227 interest to study their role in disease development and as biomarkers of disease and disease progression.
228 In our study one miRNA, miR-145-5p, was associated with both CKD and eGFR. Levels of this
229 miRNA are lower in both CKD and persons with low eGFR. Analysis of the expression pattern of this
230 miRNA showed its highest expression in colon and moderate expression in thyroid but much less in
231 other tissues, including kidney or urine. In a previous meta-analyses of this miR it was downregulated
232 in those with T2D and diabetic nephropathy (26). In addition, it was shown that this miRNA was lower
233 in those with faster eGFR decline (27). Functional studies support a role for this miR in CKD in persons
234 with diabetes and in animal studies (28; 29). miR-143-3p, associated with CKD and UACR, showed its
235 highest expression in kidney suggesting a direct role in kidney function. In previous miRNA studies it
236 was shown that this miR was down-regulated in renal tissue of animal models of (diabetic nephropathy)
237 DN (26). This was further supported by functional studies in an animal model showing impairment of
238 the glomerular filtration barrier (30). miR-143-5p was the only miRNA that remained significant after
239 further adjustment for baseline BMI, diabetes duration, HbA1c, HDL, SBP, smoking, eGFR and UACR
240 in the full models, showing its robustness as a biomarker for CKD in diabetes. To the best of our
241 knowledge this miRNA has not been linked to CKD previously. These three miRNAs are all encoded
242 in the same region on chromosome 5q32. In a GWAS aiming to identify blood miRNA QTLs both miR-
243 143-3p and miR-145-5p but not miR-143-5p or miR-145-3p were shown to be under control of the same
244 SNPs (24). It was further shown that they exert a similar function in regulation of genes involved in
245 thrombosis and angiogenesis. However, we observed quite a different expression pattern in kidney, and
246 other tissues, suggestive that their function in kidney might not be identical. The fourth significant
247 miRNA, miR-4685-3p also associated with UACR and was expressed at very low levels in all tissues
248 studied. In addition, we could not find any other studies investigating this miR or its role in kidney

249 function. Further studies are thus warranted to elucidate its function. In addition to the association with
250 CKD we also found many miRNAs associated with eGFR and or UACR. A large number of which
251 remained significant in the full models. This includes miRNAs that were also found in previous studies,
252 such as let-7a-5p, miR-192-5p, miR-194-5p, miR-214-3p, miR-374a-3p, miR-451a (11; 25; 26), as well
253 as novel ones. Interestingly, of the 123 miRNAs found in the base models for eGFR and UACR none of
254 them were associated with both phenotypes suggesting that the associated miRNAs are specific to either
255 defect in kidney function but not both. If confirmed in independent studies this provides interesting
256 starting points for further functional studies investigating their role in either condition.
257
258 *piRNAs*
259 PIWI-interacting RNAs are involved in silencing of transposons but they can also be involved in the
260 regulation of gene expression in a miRNA-like manner (31). Only one piRNA, piRNA-019825
261 (DQ597218), was significantly associated with CKD. Previously it was shown that this piRNA is one
262 of the most predominant piRNA species in urine and plasma (32). It was the only of the twelve
263 significant sRNA that showed its highest expression in urine. Whether piRNA-019825 is also associated
264 with CKD in urine requires further investigation, if so, it may prove useful as a non-invasive biomarker
265 of CKD in diabetes. To the best of our knowledge this piRNA has not been studied previously in relation
266 to kidney function. Interestingly, we also find various other piRNAs associated with either eGFR and
267 or UACR (11-14%). Several of which remained significant in the full models. Suggesting that they may
268 be useful early markers of CKD in diabetes. Interestingly, a recent study showed that DNA methylation
269 near the *PIWIL1* gene was associated with diabetic CKD (33). PIWI proteins, like *PIWIL1*, use piRNAs
270 to form the piRNA-induced silencing complex, resulting in RNA degradation and epigenetic silencing.
271 Based on these observations it will be interesting to elucidate whether the piRNAs are causally involved
272 in CKD or merely biomarkers of kidney dysfunction.
273
274 *snoRNAs*
275 Small Nucleolar RNAs (snoRNA) are predominantly localized to the nucleolus where they have a
276 function in rRNA biogenesis, structure and activity (34). There are two major types of snoRNAs
277 containing either box C/D or box H/ACA motifs. In our study we only detected box C/D snoRNAs
278 associated with CKD. Box C/D snoRNAs are involved in 2'-O-methylation of specific sequences in
279 target RNA sequences, mainly rRNA. In addition other functions have been ascribed to the snoRNAs
280 such as mRNA splicing and editing. SnoRNA derived fragments have also been described and it has
281 been speculated that they exert miRNA like functions (34-37).
282 Although snoRNAs represent only ten percent of the total circulating sRNA transcriptome in
283 our study they represent fifty-eight percent of the sRNAs found associated with CKD suggesting that
284 they may function as early biomarkers of CKD in diabetes. Whether they are causally involved in
285 regulation of kidney function needs to be shown. Of the seven snoRNAs that were associated with CKD

286 several showed their highest expression in liver or colon whilst in general kidney and urine levels of
287 these snoRNAs are low. Three snoRNAs, U8, SNORD107 and SNORD25 were associated with all three
288 phenotypes but neither had previously been linked to kidney function. SnoRNAs SNORD118,
289 SNORD24, SNORD87 and SNORD15B where associated with CKD and UACR but not eGFR. At
290 present there are no studies that have investigated the role of snoRNAs in diabetic CKD. There are,
291 however, several studies showing the association of snoRNAs with cardiometabolic and other traits (34;
292 38). This confirms that functional defects in snoRNA function can have a causal role in diseases.
293 snoRNAs U8 and SNORD118 both remained significant in the full models. These snoRNAs are encoded
294 within the *TMEM107* gene on chromosome 17p13.1, where U8 is transcribed from SNORD118 locus.
295 Mutations in SNORD118 cause a rare genetic cerebral microangiopathy called leukoencephalopathy
296 with brain calcifications and cysts (LCC; also known as Labrune syndrome) (39). However, based on
297 current evidence this a strictly neurological disease with no evidence for involvement of other tissues or
298 the kidneys. Mutations in *TMEM107* though have been linked to various syndromes that often also
299 involve the kidney, mostly through cyst development (40; 41). Further replication and functional studies
300 are thus warranted to elucidate the role of U8/SNORD118 and other snoRNAs in diabetic CKD and the
301 putative mechanism behind our observations. Given their expression pattern in tissues other than kidney
302 it will also be of interest to elucidate the origin of the circulating snoRNAs and their potential role in
303 communication between organs. In addition to the snoRNAs described above various other snoRNAs
304 are associated with eGFR and UACR (31-40% of total number of significant sRNA). Almost all
305 remained significant in the full models and thus represent interesting starting points for further
306 replication and functional studies.
307
308 *Members of other sRNA classes*
309 Although other classes of sRNAs, such as Y_RNA derived fragments, tRNA fragments, lncRNAs and
310 snRNAs represent roughly twenty percent of all circulating sRNA detected we did not find any
311 associations with CKD. We did find a small number of associations with eGFR and UACR though. The
312 lncRNA GAS5 was associated with both eGFR and UACR which persisted in the full model. GAS5-
313 lncRNA is a snoRNA host gene which encodes a large number of snoRNAs. In addition, a number of
314 studies have linked this lncRNA to kidney function (42-44). Interestingly, this lncRNA is also a binding
315 partner for SNORD118 described above (45).
316
317 Because there are to our best knowledge no GWASes for sRNAs other than for miRNAs we performed
318 a GWAS on our own samples. However, we did not identify any genome wide significant cis QTLs
319 (data not shown). Larger, better powered, studies will be needed to elucidate the role of genetics in the
320 regulation of sRNA expression levels.
321
322 *Strengths and limitations*

Our study has several strengths. It is one of the largest circulating sRNA profiling studies studying the role of circulating sRNAs in diabetic CKD. This improves power and reduces the chance of false positives. In addition, it uses untargeted RNA sequencing allowing not only the study of miRNAs but also several other classes of sRNA showing that especially a large number of these other sRNAs associated with CKD and its components eGFR and UACR. The use of early, mild cases of CKD facilitates the identification of early markers of disease and may help to introduce early interventions aiming to preserve kidney function in those at risk. The use of cases with mild CKD is on the other hand also a limitation as it has to be established if our results can also be extrapolated to those with more advanced stages of CKD. We used a total RNA isolation procedure and as such we were not able to investigate if the sRNAs are carried by extracellular vesicles or are associated with carriers such as HDL and Ago-2. Such detailed studies are also needed to establish the cells/tissue of origin of the sRNAs and elucidation of the mechanism and potential crosstalk between tissues via the sRNAs as has already been well documented for miRNAs (5). The use of plasma for RNA isolation is another limitation. Although from a biomarker perspective this isn't a limitation, from a functional perspective the role of the sRNAs we identified needs to be established in the relevant cells and tissues. Another limitation of our study is its cross-sectional design. Prospective studies are needed to investigate if these circulating sRNAs can be used in the early prediction of those at risk of diabetic CKD before onset.

In this study we have shown that several circulating sRNAs are associated with CKD in subjects with diabetes. Interestingly, only a relatively small fraction belong to the well-studied class of non-coding RNAs, the microRNAs. This highlights the potential of other classes of non-coding sRNAs and especially the snoRNAs in diabetic kidney disease development. Interestingly, many of the sRNAs identified are not highly expressed in kidney and may thus be used in between tissue crosstalk in the regulation of kidney function. Future prospective and functional studies are needed to establish the role of circulating sRNAs in subjects with diabetes and chronic kidney disease.

Acknowledgements

We would like to thank all staff and participants of the Hoorn Diabetes Care System for their support and participation.

Funding

This work was supported by a diabetes breakthrough grant from the Dutch Diabetes Research Foundation and ZonMw (grant number 459001015) and a grant from the EFSD/Boehringer Ingelheim European Research Programme on "Multi-System Challenges in Diabetes" 2021. The study sponsors/funders were not involved in the design of the study; the collection, analysis, and interpretation of data; writing the report; and did not impose any restrictions regarding the publication of the report.

360 **Contribution statement**

361 LMtH, JWJB and RCS designed the study. JAdK and SJC isolated RNA and prepared the samples for
362 RNAseq. MTB and JWJB were responsible for phenotypic data acquisition and preparation. LMtH,
363 GAB and RCS prepared, analyzed and interpreted the data. LMtH and RCS wrote the first draft of the
364 manuscript. LMtH and RCS are the guarantors of this work and, as such, had full access to all the data
365 in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All
366 authors critically read and revised the manuscript and approved the final version of the manuscript.

367

368 **References**

369 1. Persson F, Rossing P: Diagnosis of diabetic kidney disease: state of the art and future perspective. Kidney Int
370 Suppl (2011) 2018;8:2-7

371 2. Tofte N, Vogelzangs N, Mook-Kanamori D, Brahimaj A, Nano J, Ahmadizar F, van Dijk KW, Frimodt-
372 Møller M, Arts I, Beulens JWJ, Rutters F, van der Heijden AA, Kavousi M, Stehouwer CDA, Nijpels G, van
373 Greevenbroek MMJ, van der Kallen CJH, Rossing P, Ahluwalia TS, t Hart LM: Plasma Metabolomics Identifies
374 Markers of Impaired Renal Function: A Meta-analysis of 3089 Persons with Type 2 Diabetes. J Clin Endocrinol
375 Metab 2020;105

376 3. Tofte N, Persson F, Rossing P: Omics research in diabetic kidney disease: new biomarker dimensions and new
377 understandings? Journal of nephrology 2020;

378 4. Barutta F, Bellini S, Canepa S, Durazzo M, Gruden G: Novel biomarkers of diabetic kidney disease: current
379 status and potential clinical application. Acta Diabetol 2021;58:819-830

380 5. Guay C, Regazzi R: Exosomes as new players in metabolic organ cross-talk. Diabetes Obes Metab 2017;19
381 Suppl 1:137-146

382 6. Thomou T, Mori MA, Dreyfuss JM, Konishi M, Sakaguchi M, Wolfrum C, Rao TN, Winnay JN, Garcia-
383 Martin R, Grinspoon SK, Gorden P, Kahn CR: Adipose-derived circulating miRNAs regulate gene expression in
384 other tissues. Nature 2017;542:450-455

385 7. Umu SU, Langseth H, Bucher-Johannessen C, Fromm B, Keller A, Meese E, Lauritzen M, Leithaug M, Lyle
386 R, Rounge TB: A comprehensive profile of circulating RNAs in human serum. RNA Biol 2018;15:242-250

387 8. Fritz JV, Heintz-Buschart A, Ghosal A, Wampach L, Etheridge A, Galas D, Wilmes P: Sources and Functions
388 of Extracellular Small RNAs in Human Circulation. Annu Rev Nutr 2016;36:301-336

389 9. Florijn BW, Duijs J, Levels JH, Dallinga-Thie GM, Wang Y, Boing AN, Yuana Y, Stam W, Limpens R, Au
390 YW, Nieuwland R, Rabelink TJ, Reinders MEJ, van Zonneveld AJ, Bijkerk R: Diabetic Nephropathy Alters the
391 Distribution of Circulating Angiogenic MicroRNAs Among Extracellular Vesicles, HDL, and Ago-2. Diabetes
392 2019;68:2287-2300

393 10. Trionfini P, Benigni A: MicroRNAs as Master Regulators of Glomerular Function in Health and Disease. J
394 Am Soc Nephrol 2017;28:1686-1696

395 11. Assmann TS, Recamonde-Mendoza M, de Souza BM, Bauer AC, Crispim D: MicroRNAs and diabetic
396 kidney disease: Systematic review and bioinformatic analysis. Mol Cell Endocrinol 2018;477:90-102

397 12. Freedman JE, Gerstein M, Mick E, Rozowsky J, Levy D, Kitchen R, Das S, Shah R, Danielson K, Beaulieu
398 L, Navarro FC, Wang Y, Galeev TR, Holman A, Kwong RY, Murthy V, Tanriverdi SE, Koupenova-Zamor M,

399 Mikhalev E, Tanriverdi K: Diverse human extracellular RNAs are widely detected in human plasma. Nat
400 Commun 2016;7:11106
401 13. van der Heijden AA, Rauh SP, Dekker JM, Beulens JW, Elders P, 't Hart LM, Rutters F, van Leeuwen N,
402 Nijpels G: The Hoorn Diabetes Care System (DCS) cohort. A prospective cohort of persons with type 2 diabetes
403 treated in primary care in the Netherlands. BMJ Open 2017;7:e015599
404 14. Slieker RC, van der Heijden A, Nijpels G, Elders PJM, 't Hart LM, Beulens JWJ: Visit-to-visit variability of
405 glycemia and vascular complications: the Hoorn Diabetes Care System cohort. Cardiovasc Diabetol 2019;18:170
406 15. KDIGO: KDIGO clinical practice guideline for the diagnosis, evaluation, prevention, and treatment of
407 Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). Kidney international Supplement 2009:S1-
408 130
409 16. Rozowsky J, Kitchen RR, Park JJ, Galeev TR, Diao J, Warrell J, Thistlethwaite W, Subramanian SL,
410 Milosavljevic A, Gerstein M: exceRpt: A Comprehensive Analytic Platform for Extracellular RNA Profiling.
411 Cell Syst 2019;8:352-357.e353
412 17. Griffiths-Jones S: The microRNA Registry. Nucleic Acids Res 2004;32:D109-111
413 18. Sai Lakshmi S, Agrawal S: piRNABank: a web resource on classified and clustered Piwi-interacting RNAs.
414 Nucleic Acids Res 2008;36:D173-177
415 19. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa
416 A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G,
417 Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian
418 S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A,
419 Reymond A, Gerstein M, Guigó R, Hubbard TJ: GENCODE: the reference human genome annotation for The
420 ENCODE Project. Genome Res 2012;22:1760-1774
421 20. Glažar P, Papavasileiou P, Rajewsky N: circBase: a database for circular RNAs. Rna 2014;20:1666-1670
422 21. Chan PP, Lowe TM: GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic
423 Acids Res 2009;37:D93-97
424 22. Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential expression analysis
425 of digital gene expression data. Bioinformatics 2009;26:139-140
426 23. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH,
427 Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A: NCBI
428 GEO: archive for functional genomics data sets—update. Nucleic Acids Research 2012;41:D991-D995
429 24. Nikpay M, Beehler K, Valsesia A, Hager J, Harper ME, Dent R, McPherson R: Genome-wide identification
430 of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation of
431 cardiometabolic phenotypes. Cardiovasc Res 2019;115:1629-1645
432 25. Loganathan TS, Sulaiman SA, Abdul Murad NA, Shah SA, Abdul Gafor AH, Jamal R, Abdullah N:
433 Interactions Among Non-Coding RNAs in Diabetic Nephropathy. Front Pharmacol 2020;11:191
434 26. Gholaminejad A, Abdul Tehrani H, Gholami Fesharaki M: Identification of candidate microRNA biomarkers
435 in diabetic nephropathy: a meta-analysis of profiling studies. J Nephrol 2018;
436 27. Chen NX, Kiattisunthorn K, O'Neill KD, Chen X, Moorthi RN, Gattone VH, 2nd, Allen MR, Moe SM:
437 Decreased microRNA is involved in the vascular remodeling abnormalities in chronic kidney disease (CKD).
438 PLoS One 2013;8:e64558

439 28. Barutta F, Tricarico M, Corbelli A, Annaratone L, Pinach S, Grimaldi S, Bruno G, Cimino D, Taverna D,
440 Deregibus MC, Rastaldi MP, Perin PC, Gruden G: Urinary exosomal microRNAs in incipient diabetic
441 nephropathy. PLoS One 2013;8:e73798
442 29. Wei B, Liu YS, Guan HX: MicroRNA-145-5p attenuates high glucose-induced apoptosis by targeting the
443 Notch signaling pathway in podocytes. Exp Ther Med 2020;19:1915-1924
444 30. Müller-Deile J, Gellrich F, Schenk H, Schroder P, Nyström J, Lorenzen J, Haller H, Schiffer M:
445 Overexpression of TGF-β Inducible microRNA-143 in Zebrafish Leads to Impairment of the Glomerular
446 Filtration Barrier by Targeting Proteoglycans. Cell Physiol Biochem 2016;40:819-830
447 31. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD: PIWI-interacting RNAs: small RNAs with big
448 functions. Nat Rev Genet 2019;20:89-108
449 32. El-Mogy M, Lam B, Haj-Ahmad TA, McGowan S, Yu D, Nosal L, Rghei N, Roberts P, Haj-Ahmad Y:
450 Diversity and signature of small RNA in different bodily fluids using next generation sequencing. BMC
451 Genomics 2018;19:408
452 33. Lecamwasam A, Novakovic B, Meyer B, Ekinci EI, Dwyer KM, Saffery R: DNA methylation profiling
453 identifies epigenetic differences between early versus late stages of diabetic chronic kidney disease. Nephrol
454 Dial Transplant 2021;36:2027-2038
455 34. Deogharia M, Majumder M: Guide snoRNAs: Drivers or Passengers in Human Disease? Biology (Basel)
456 2018;8
457 35. Bergeron D, Fafard-Couture É, Scott MS: Small nucleolar RNAs: continuing identification of novel
458 members and increasing diversity of their molecular mechanisms of action. Biochem Soc Trans 2020;48:645-
459 656
460 36. Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G:
461 A human snoRNA with microRNA-like functions. Mol Cell 2008;32:519-528
462 37. Scott MS, Ono M, Yamada K, Endo A, Barton GJ, Lamond AI: Human box C/D snoRNA processing
463 conservation across multiple cell types. Nucleic Acids Res 2012;40:3676-3688
464 38. Ingen E, Homberg DAL, Bent ML, Mei H, Papac-Milicevic N, Kremer V, Boon RA, Quax PHA, Wojta J,
465 Nossent AY: C/D box snoRNA SNORD113-6/AF357425 plays a dual role in integrin signalling and arterial
466 fibroblast function via pre-mRNA processing and 2'O-ribose methylation. Hum Mol Genet 2021;
467 39. Crow YJ, Marshall H, Rice GI, Seabra L, Jenkinson EM, Baranano K, Battini R, Berger A, Blair E,
468 Blauwblomme T, Bolduc F, Boddaert N, Buckard J, Burnett H, Calvert S, Caumes R, Ng AC, Chiang D,
469 Clifford DB, Cordelli DM, de Burca A, Demic N, Desguerre I, De Waele L, Di Fonzo A, Dunham SR, Dyack S,
470 Elmslie F, Ferrand M, Fisher G, Karimiani EG, Ghoumid J, Gibbon F, Goel H, Hilmarsen HT, Hughes I, Jacob
471 A, Jones EA, Kumar R, Leventer RJ, MacDonald S, Maroofian R, Mehta SG, Metz I, Monfrini E, Neumann D,
472 Noetzel M, O'Driscoll M, Õunap K, Panzer A, Parikh S, Prabhakar P, Ramond F, Sandford R, Saneto R, Soh C,
473 Stutterd CA, Subramanian GM, Talbot K, Thomas RH, Toro C, Touraine R, Wakeling E, Wassmer E, Whitney
474 A, Livingston JH, O'Keefe RT, Badrock AP: Leukoencephalopathy with calcifications and cysts: Genetic and
475 phenotypic spectrum. Am J Med Genet A 2021;185:15-25
476 40. Shaheen R, Almoisheer A, Faqeih E, Babay Z, Monies D, Tassan N, Abouelhoda M, Kurdi W, Al Mardawi
477 E, Khalil MM, Seidahmed MZ, Alnemer M, Alsahan N, Sogaty S, Alhashem A, Singh A, Goyal M, Kapoor S,

478 Alomar R, Ibrahim N, Alkuraya FS: Identification of a novel MKS locus defined by TMEM107 mutation. Hum
479 Mol Genet 2015;24:5211-5218
480 41. Lambacher NJ, Bruel AL, van Dam TJ, Szymańska K, Slaats GG, Kuhns S, McManus GJ, Kennedy JE, Gaff
481 K, Wu KM, van der Lee R, Burglen L, Doummar D, Rivière JB, Faivre L, Attié-Bitach T, Saunier S, Curd A,
482 Peckham M, Giles RH, Johnson CA, Huynen MA, Thauvin-Robinet C, Blacque OE: TMEM107 recruits
483 ciliopathy proteins to subdomains of the ciliary transition zone and causes Joubert syndrome. Nat Cell Biol
484 2016;18:122-131
485 42. Tong H, Qiuling F, Jingbin C, Li X, Xu W, Dongfang L, Lining W: The efficiency of serum lncRNA
486 GAS5/miR-21 as biomarkers in patients with diabetes and diabetic nephropathy. Chinese Journal of Nephrology
487 2017;33:906-913
488 43. Chu PM, Yu CC, Tsai KL, Hsieh PL: Regulation of Oxidative Stress by Long Non-Coding RNAs in
489 Vascular Complications of Diabetes. Life (Basel) 2022;12
490 44. Coellar JD, Long J, Danesh FR: Long Noncoding RNAs and Their Therapeutic Promise in Diabetic
491 Nephropathy. Nephron 2021;145:404-414
492 45. Bouchard-Bourelle P, Desjardins-Henri C, Mathurin-St-Pierre D, Deschamps-Francoeur G, Fafard-Couture
493 É, Garant JM, Elela SA, Scott MS: snoDB: an interactive database of human snoRNA sequences, abundance and
494 interactions. Nucleic Acids Res 2020;48:D220-d225
495
496
497

498 **Table 1. Clinical characteristics of the study population.**

	all	No CKD	CKD
n	475	406	69
Gender (%Female)	45	46	42
Age (years)	61.4 ± 9.1	61.0 ± 8.8	63.8 ± 10.1*
BMI (Kg/m^2)	30.7 ± 6.0	30.3 ± 5.6	32.7 ± 7.3*
Age-at-Diagnosis (years)	59.3 ± 9.0	58.8 ± 8.7	61.8 ± 9.8*
Diabetes duration (years)	2.2 ± 1.1	2.2 ± 1.1	2.0 ± 1.1
Fasting glucose (mmol/L)	8.0 ± 1.6	8.0 ± 1.7	8.1 ± 1.4
HbA1c (mmol/mol)	46 (42, 51)	46 (42, 51)	46 (43, 54)
HDL (mmol/L)	1.19 ± 0.32	1.20 ± 0.33	1.11 ± 0.25
eGFR (ml/min/1,73 m^2)	84 ± 16	87 ± 13	70 ± 22*
eGFR ≤ 60 (%)	34 (7)	0	34 (50)
Urine Albumin-to-Creatinine ratio (mg/mmol)	0.48 (0.20, 1.00)	0.41 (0.18, 0.79)	3.99 (0.79, 10.5)*
Albuminuria (%)	43 (9)	0	43 (62)
Chronic Kidney Disease (%))	69 (15)	0	69 (100)
Diastolic blood pressure (mm Hg)	78 ± 9	78 ± 9	79 ± 10
Systolic blood pressure (mm Hg)	139 ± 18	139 ± 18	142 ± 20
Smoking (current/ever/never)	92/130/250	76/107/220	16/23/30
Medication use			
Metformin (%)	340 (72)	287 (71)	53 (77)
SU (%)	130 (27)	105 (26)	25 (36)
Other glucose-lowering (%)	13 (3)	12 (3)	1 (1)
Antihypertensive medication (%)	319 (67)	260 (64)	59 (86)*
Lipid lowering medication (%)	320 (67)	271 (67)	49 (71)

499 Data are mean ± SD, n (%) or median (IQR). All measures taken at the time of blood sampling.

500 * P≤0.05 versus those without CKD.

501

502

503 **Table 2. Frequency of different types of sRNAs associated with the three endpoints.**

	Total # of sRNA detected			CKD			eGFR			UACR	
	n	%		n	%		n	%		n	%
Total # sRNAs	**938**	**100**		**12**			**73**			**209**	
miRNA	602	64		4	33		31	43		92	44
snoRNA	96	10		7	58		29	40		65	31
lincRNA	93	10					2	3		18	9
piRNA	45	5		1	8		10	14		23	11
Y_RNA fragments	33	4									
others	26	3								5	2
tRNA fragments	24	3					1	1			
snRNA	12	1								3	1
scaRNA	7	1								3	1

504

505

506 **Table 3. Overview of twelve sRNAs associated with prevalent CKD**

sRNA	Prevalent CKD			eGFR			UACR	
	LogFC	**P_{FDR}**		**logFC**	**P_{FDR}**		**logFC**	**P_{FDR}**
SNORD118	-0.873	$2.35x10^{-3}$		0.013	$7.72x10^{-2}$		-0,861	$1.76x10^{-3}$
U8	-0.851	$2.35x10^{-3}$		0.014	$4.21x10^{-2}$		-0,733	$6.26x10^{-3}$
hsa-miR-143-5p	-0.479	$8.04x10^{-3}$		0.007	$1.05x10^{-1}$		-0,339	$5.57x10^{-2}$
SNORD107	-1.283	$2.59x10^{-2}$		0.028	$2.11x10^{-2}$		-2,192	$3.02x10^{-5}$
SNORD24	-0.794	$2.59x10^{-2}$		0.015	$5.96x10^{-2}$		-0,952	$1.45x10^{-3}$
SNORD87	-1.188	$3.03x10^{-2}$		0.018	$1.41x10^{-1}$		-2,281	$1.07x10^{-5}$
hsa-miR-145-5p	-0.321	$3.03x10^{-2}$		0.008	$2.10x10^{-2}$		-0,169	$2.13x10^{-1}$
hsa-miR-143-3p	-0.328	$3.40x10^{-2}$		0.002	$6.38x10^{-1}$		-0,375	$6.65x10^{-3}$
SNORD15B	-0.756	$4.88x10^{-2}$		0.014	$1.04x10^{-1}$		-1,028	$1.59x10^{-3}$
hsa_piR_019825|gb|DQ597218	-0.463	$4.88x10^{-2}$		0.009	$1.05x10^{-1}$		-0,188	$3.86x10^{-1}$
hsa-miR-4685-3p	-0.430	$4.88x10^{-2}$		0.003	$6.84x10^{-1}$		-0,527	$4.29x10^{-3}$
SNORD25	-0.905	$4.88x10^{-2}$		0.025	$1.25x10^{-2}$		-1,458	$2.25x10^{-4}$

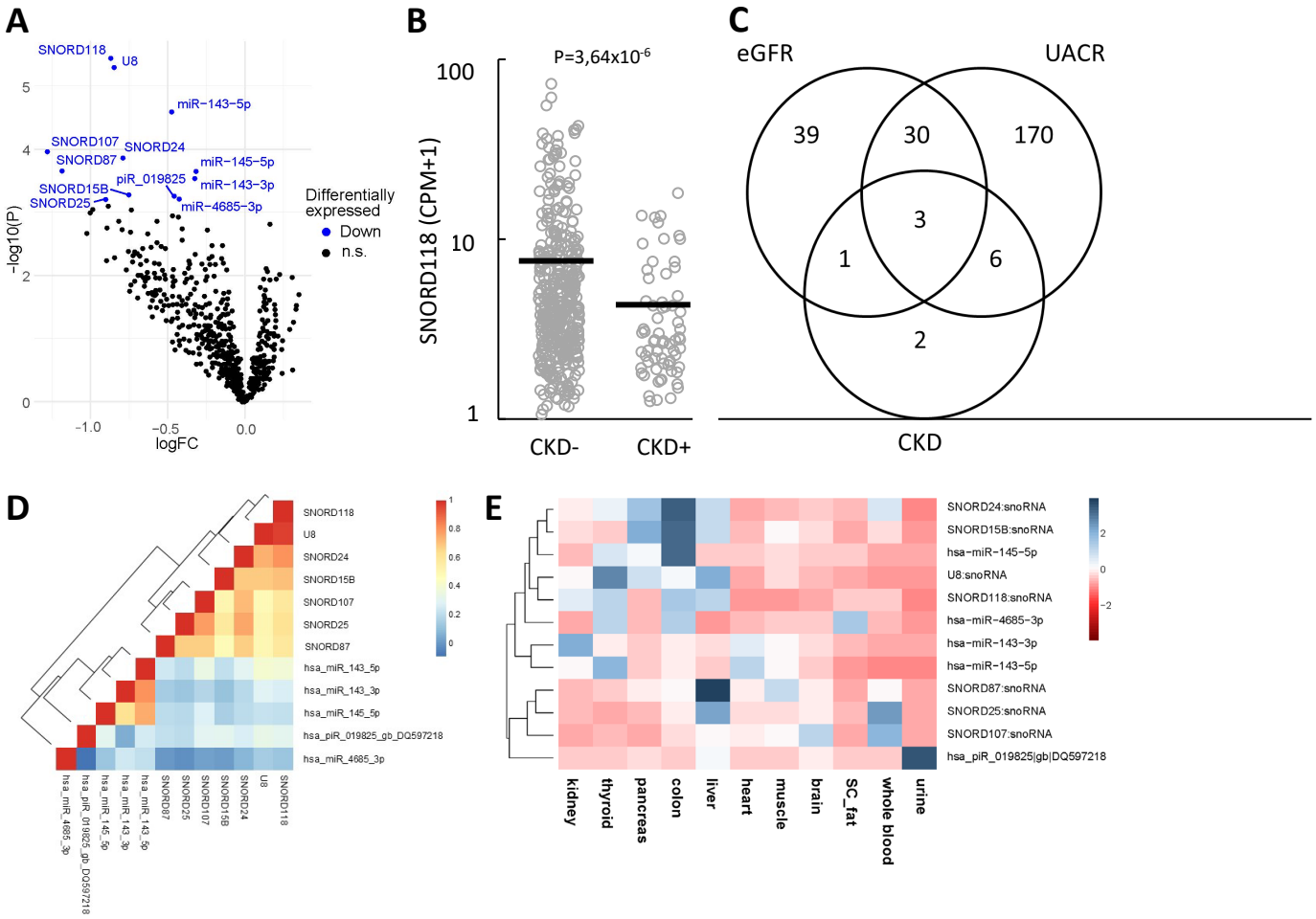
507 LogFC is the log2 of the fold change. P_{FDR} represents the P value after correction for multiple hypothesis testing using the
508 Benjamini-Hochberg procedure.

509

510

511

512 **Figure 1. sRNAs significantly associated with CKD.**



513

514 Panel A. Volcano plot showing the association between sRNAs and CKD in the base model. sRNAs in blue are significantly
515 lower is cases with CKD after multiple hypothesis testing (FDR≤0.05). logFC is the natural logarithm of the fold change
516 between cases and controls. Panel B. SNORD118 levels in those with or without CKD. CPM = counts per million reads. Panel
517 C. Overlap between sRNAs significantly associated with CKD, eGFR or UACR in the base model. Panel D. Pairwise
518 correlations for the twelve sRNA associated with CKD in the base model. Colors represent Spearman's rho. Panel E. Expression
519 pattern for sRNAs associated with CKD in publicly available datasets for 11 tissues of interest. Colors represent the Z-scaled
520 expression of a sRNA where blue indicates relative high expression and red relative low expression.

521