

1 **Small RNA sequencing reveals snoRNAs and piRNA-019825 as novel players in diabetic kidney**
2 **disease**

3

4 LM 't Hart^{1,2,3,4} (ORCID 0000-0003-4401-2938), JA de Klerk^{1,5,6}, GA Bouland¹, JHD Peerlings¹, M.T.
5 Blom^{3,7}, SJ Cramer¹, R Bijkerk^{5,6}, JWJ Beulens^{2,3}, RC Slieker¹

6

7 ¹ Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, The
8 Netherlands

9 ² Department of Epidemiology and Data Sciences, Amsterdam University Medical Center, Location
10 Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

11 ³ Health Behaviors & Chronic Diseases Research Program & Personalised Medicine Research
12 Program, Amsterdam Public Health, Amsterdam, The Netherlands

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

15 ⁵ Department of Internal Medicine, Division of Nephrology, Leiden University Medical Center,
16 Leiden, The Netherlands

17 ⁶ Eindhoven Laboratory for Vascular and Regenerative Medicine, Leiden University Medical Center,
18 Leiden, The Netherlands

19 ⁷ Department of General Practice, Amsterdam University Medical Center, Location Vrije Universiteit
20 Amsterdam, Amsterdam, The Netherlands

21

22 Author for correspondence:

23 LM 't Hart, PhD

24 Leiden University Medical Center

25 Department of Cell and Chemical Biology

26 Albinusdreef 2

27 2333ZA Leiden, The Netherlands

28 +31 715269796

29 lmthart@lumc.nl

30

31

32 **Abstract**

33 **Introduction** Micro- and macrovascular complications are common among persons with type 2
34 diabetes. Recently there has been growing interest to investigate the potential of circulating small non-
35 coding RNAs (sncRNAs) as contributors to the development of diabetic complications. In this study
36 we investigate to what extent circulating sncRNAs levels associate with prevalent diabetic kidney
37 disease (DKD) in persons with type 2 diabetes.

38 **Methods** Plasma sncRNAs levels were determined using small RNA-seq, allowing detection of
39 miRNAs, snoRNAs, piRNAs, tRNA-fragments and various other sncRNA classes. We tested for
40 differentially expressed sncRNAs in persons with type 2 diabetes, with DKD (n=69) or without DKD
41 (n=405). In secondary analyses, we also tested the association with eGFR, albuminuria (UACR) and
42 the plasma proteome.

43 **Results** In total seven sncRNAs were negatively associated with prevalent DKD (all $P_{FDR} \leq 0.05$).
44 Including one microRNA (miR-143-5p), five snoRNAs (U8, SNORD118, SNORD24, SNORD107,
45 SNORD87) and a piRNA (piR-019825|DQ597218). Proteomic analyses showed that the seven
46 sncRNAs, and especially the piRNA piR-019825, were associated with plasma levels of 24 proteins of
47 which several have known associations with kidney function including TNF sR-I (TNFRFS1A), DAN
48 (NBL1) and cystatin C (CST3).

49 **Conclusion** We have identified novel small non-coding RNAs, primarily from classes other than
50 microRNAs, that are associated with diabetic kidney disease. Our results shows that the involvement
51 of small non-coding RNAs in DKD goes beyond the already known microRNAs and also involves
52 other classes of sncRNA, in particular snoRNAs and the piRNA piR-019825, that have never been
53 studied before in relation to kidney function.

54

55 **Keywords:** Type 2 diabetes, small RNA, microRNA, snoRNA, piRNA, Chronic kidney disease,
56 Nephropathy

57

58 **Introduction**

59 Diabetic kidney disease (DKD) is one of the most common complications in type 2 diabetes.
60 Nonetheless, while some develop complications early in the disease, others will never. The
61 mechanisms underlying DKD are not fully understood though [1].

62 Circulating small non-coding RNAs (sncRNA) are mostly contained in extracellular vesicles
63 or are bound to RNA-binding proteins and lipoproteins and originate from various tissues including
64 tissues important in kidney function [2]. Whereas previously it was thought that sncRNAs merely
65 reflect degradation by-products, it has now been shown that they exert important functions in cell-to-
66 cell communication, the regulation of gene expression, splicing and ribosomal RNA maturation [2].

67 Although a substantial number of studies have examined circulating sncRNAs associated with
68 kidney function in persons with diabetes, most studies are relatively small ($n < 100$) and focused on
69 microRNAs. From these studies a large number of miRNAs have emerged that are putatively
70 associated with kidney function. However, probably due to small sample sizes, heterogeneity in
71 endpoints and methodology used there is limited consistency between studies as shown in various
72 meta-analyses [3-5].

73 To the best of our knowledge there are currently no studies that have systematically examined
74 the other classes (biotypes) of small ncRNA. We thus go beyond the current state-of-the-art by not
75 only investigating microRNAs but also other sncRNAs classes, including besides miRNAs, piRNAs,
76 tRNA fragments (tRFs), lncRNA fragments, Y-RNAs and snoRNAs. For this we applied small RNA
77 sequencing in a large well-defined cohort ($n=474$) of persons with and without DKD to capture the
78 whole plasma sncRNA transcriptome allowing for a hypothesis-free, comprehensive analysis of the
79 role of sncRNAs in diabetic kidney disease [6].

80

81 **Materials and Methods**

82 A schematic overview of the study design can be found in Supplemental Figure 1.

83

84 *Study sample*

85 For this project we used data from the Hoorn Diabetes Care System cohort (DCS), a large prospective
86 study comprised of 14,000+ individuals with type 2 diabetes who receive structured diabetes care [7].
87 Participants visit the Hoorn diabetes research center annually for diabetes care including routine
88 biochemistry, medication review and clinical care. Patients were also invited to participate in DCS
89 biobanks in which we combine the annual phenotypic data with biobanking of liquid biopsies which
90 includes serum, plasma and urine. Blood samples were collected during two phases of biobanking in
91 2008/2009 and 2012-2014. For the current study we used a subsample from the whole DCS biobank
92 ($n=475$) using the following inclusion criteria: a plasma sample collected within four years after
93 diagnosis of diabetes and no-insulin use at the time of blood sampling.

94 All laboratory measurements were on samples taken in a fasted state on the same day as the
95 sample for RNA-seq. Routine anthropometric, biochemistry methods and definitions have been
96 described previously in detail in van der Heijden et al. [7]. Kidney function was assessed based on
97 eGFR (CKD-EPI) and urinary albumin to creatinine ratio (UACR) and included as continuous
98 variables in our analyses. Stages of chronic kidney disease (CKD) were defined according the KDIGO
99 criteria as previously described [8]. In short, this definition uses a combination of eGFR and UACR to
100 categorize persons based on kidney function. Persons in stage 0 do not show signs of kidney
101 dysfunction whereas the stages 1 to 3 have increasing kidney dysfunction as shown in Supplemental
102 Figure 2. Cases were defined as having CKD stage ≥ 1 (i.e. eGFR < 60 ml/min/1.73 m² and/or UACR
103 > 3 mg/mmol) at the time of blood sampling for sncRNA analysis which resulted in 69 cases.

104

105 *Circulating small RNA sequencing*

106 Circulating sncRNAs were isolated from 900ul of citrate plasma using a commercial cell-free RNA
107 isolation kit (Quick-cfRNA Serum & Plasma Kit, Zymo Research, Irvine, CA, USA). After quality
108 control 475 samples were submitted for RNA sequencing. Library preps were constructed using the
109 NEBNext® Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs, Ipswich, MA,
110 USA). A Pippin Prep instrument (Sage Science, Beverly, MA, USA) was used for size selection to
111 further enrich for sncRNAs with a size of ~20 to 100 nucleotides. Sequencing was performed using the
112 Illumina NextSeq500 sequencer with v2.5 sequencing reagent kits and 35 cycle paired-end reads
113 (PE35). On average 11.3 ± 2.8 million reads per sample were generated. The exceRpt pipeline [9],
114 developed by the NIH Extracellular RNA Communication Consortium (ERCC), was used to map the
115 obtained reads, for quality control, to process the data and generate RNA abundance estimates for
116 miRNA, piRNA, lncRNA, tRNA fragments, Y_RNA (fragments), snRNA, snoRNA, scaRNA and
117 various other sncRNA species. Using respectively miRBase v22 [10], piRNABank v1 [11], Gencode
118 (version 38) [12], circBase [13] and GtRNAdb [14]. One sample did not pass the QC thresholds.

119

120 *Taqman validation of the RNA-seq results*

121 We used Taqman qPCR assays (ThermoFisher Scientific, NL) to verify the validity and
122 reproducibility of the small RNA-seq results. First cDNA was generated from 5ng of RNA isolated
123 from plasma using the TaqMan™ Advanced miRNA cDNA Synthesis Kit (ThermoFisher Scientific,
124 NL) according to the instructions of the manufacturer. We chose four miRNA assays with a range
125 from low to relative high expression in our study sample (miR-542-3p, miR-323b-3p, miR-186-5p and
126 miR-423-5p) which were measured in 12 randomly selected plasma RNA samples for which sncRNA-
127 seq data were available. Clinical characteristics are shown in Supplemental Table 1. The Taqman
128 qPCR assays were run according to the manufacturer's instructions using TaqMan Fast Advanced
129 Master mix and TaqMan Advanced miRNA Assays (ThermoFisher Scientific, NL). miR-191-5p

130 expression was used as the reference (housekeeping gene) in both the Taqman and small RNA-seq
131 datasets as suggested by the manufacturer [15].

132

133 *Expression patterns of small RNAs in tissues*

134 Publicly available raw sncRNA-seq data for various metabolic tissues were downloaded from the
135 Gene Expression Omnibus (GEO) [16]. The tissues for which sncRNA-seq data was used included
136 kidney, thyroid, brain, liver, muscle, heart, colon, pancreas, subcutaneous white adipose tissue and
137 whole blood and urine. Accession numbers are given in Supplemental Table 2. Data were processed
138 using the exceRpt pipeline as described above. The tissue data were used to estimate the abundance of
139 the sncRNAs in these tissues. Data was z-scaled in figures.

140

141 *Plasma proteomics.*

142 Plasma proteomic data was available for 589 participants of the Hoorn DCS study and generated using
143 the SomaLogic SOMAscan platform (Boulder, CO) as previously described [17]. After quality control
144 1195 protein levels were available for analysis. For 199 subjects we also had the corresponding
145 sncRNA-seq data from the same sampling date. Clinical characteristics of the participants in this study
146 are shown in Supplemental Table 1.

147

148 *Statistical analysis*

149 The R-package edgeR (v3.42.4) was used to identify differentially expressed sncRNAs [18]. A
150 negative binomial generalized log-linear model was applied using edgeR's function
151 *estimateGLMRobustDisp*. For the determination of differential expression and p-values we used the
152 quasi-likelihood (QL) F-test using edgeR's *glmQLFTest* function according to the recommendations
153 of the developers. The base model was unadjusted. A fully adjusted model was adjusted for BMI,
154 diabetes duration, HbA1c, HDL, SBP and smoking unless otherwise stated. Age and sex were
155 excluded from the model as these are part of the definition of eGFR. UACR was log₁₀-transformed
156 before analysis. In sensitivity analyses, we additionally included use of antihypertensive medication
157 and or eGFR or UACR. The correlation between the sncRNA transcriptome and plasma proteome was
158 tested using a linear model with the log transformed proteomic measurements as outcome and the
159 sncRNA transcriptome as predictors with adjustment for age, BMI, sex and technical covariates.
160 Associations between the proteome and eGFR or UACR were tested using the same model. The
161 Benjamini-Hochberg procedure was used to adjust for multiple hypothesis testing and an FDR below
162 0.05 was considered significant.

163

164 **Results**

165 *Clinical characteristics*

166 The clinical characteristics of our study group are shown in Table 1. The characteristics resemble
167 those of a typical T2D population with slightly more males than females, a mean age-at-diagnosis of
168 59.3 ± 9.0 years, BMI of 30.7 ± 6.0 kg/m² and an average duration of diabetes at time of blood
169 sampling of 2.2 ± 1.1 years. Glycemic control was good with an average HbA1c of 46 (42, 51)
170 mmol/mol. Prevalence of DKD at the time of blood sampling was 15%.

171

172 *sncRNA biotype abundance in our samples*

173 After quality control, 938 sncRNAs were expressed above the threshold of a mean ≥ 5 copies
174 in our data, which included miRNAs, piRNA, Y_RNA fragments, lncRNA fragments,
175 tRNA_fragments, snoRNAs, snRNAs and various other sncRNA species (Table 2). miRNAs, Y-
176 RNAs and tRFs were the most abundant sncRNAs biotypes comprising respectively 57.6, 34.9 and 4.8
177 percent of all reads. Other sncRNA biotypes were much less abundant ($\leq 1.4\%$). There were no
178 significant differences in the total number of reads per biotype between cases and controls (all $P > 0.4$).

179

180 *Taqman validation shows good reproducibility*

181 We next sought to verify validity and reproducibility of the sncRNA-seq. Taqman qPCR
182 assays for sncRNAs with high, medium and low expression were used and these showed a good
183 correlation with the RNA-seq results ($r^2 = 0.86$, Supplemental Figure 3).

184

185 *sncRNAs associate with DKD, eGFR and UACR*

186 Seven sncRNAs were significantly associated with prevalent DKD in the unadjusted model
187 ($P_{FDR} \leq 0.05$, Table 3, Figure 1A). All identified sncRNAs were downregulated in DKD cases. The top
188 hit was snoRNA U8 (\log_2 fold change = -0.859, $P_{FDR} = 1.90 \times 10^{-3}$, Figure 1B). Five of these sncRNAs
189 also showed a significant trend for lower levels in those with more advanced stages of CKD
190 ($P_{FDR} \leq 0.05$, Supplemental Figure 4).

191 Respectively, 107 and 220 sncRNAs were significantly associated with eGFR and UACR,
192 which are used for the DKD classification (Table 3, Supplemental Table 3, Supplemental Figures 5
193 and 6). The overlap between the sncRNAs associated with DKD and its components eGFR and UACR
194 is shown in figure 1C and Table 3. snoRNA U8 overlapped between all three endpoints used and
195 piR_019825 overlapped between eGFR and DKD. SNORD118, SNORD24, SNORD107, SNORD87
196 and miR-143-5p all overlapped between UACR and DKD.

197 In the full model respectively 7, 153 and 205 sncRNAs remained significantly associated with
198 DKD, eGFR and UACR (Supplemental Table 4, Supplemental Figure 7; 57%; 95% and 84% overlap
199 with the unadjusted model). U8, SNORD118, piR_019825 and miR-143-5p were significant in both
200 the base and full models. In a sensitivity analysis, additional adjustment for use of antihypertensive
201 medication and baseline eGFR or UACR (only for eGFR and UACR) did not have a major impact on
202 the effect sizes but reduced the number of significant associations ($r^2 > 0.92$, Supplemental Figure 8).

203

204 *Certain types of sncRNAs are enriched in DKD*

205 Comparing the frequency of different types of sncRNAs identified showed that while
206 snoRNAs represent 10% of the total small RNA transcriptome in our RNA-seq study they are the most
207 abundant class associated with prevalent DKD (71%, OR=21.9, $P=2.4 \times 10^{-4}$, Table 2). Enrichment for
208 snoRNAs is also seen for UACR (OR=3.8, $p=8.4 \times 10^{-13}$). Associations with miRNAs on the other hand
209 are less frequent whilst piRNAs also seem enriched (Table 2). A similar pattern is again seen for
210 UACR and to a lesser extent for eGFR. In the full model a largely similar picture is observed
211 (Supplemental Table 5).

212 The pairwise correlation in expression between the seven sncRNAs associated with diabetic
213 DKD was very modest (Figure 1D). The only exceptions were U8 and SNORD118, which is likely
214 due to the fact that they are encoded in the same host gene (TMEM107) on chromosome 17. Results of
215 these analyses for eGFR and UACR are shown in Supplemental Figure 9.

216

217 *sncRNAs are tissue-enriched and correlate with plasma protein levels*

218 In a next step, we examined the tissue expression patterns of the identified sncRNAs. We used
219 publicly available expression data in 11 tissues or body fluids (Fig. 1E). This illustrated that some
220 sncRNAs show tissue-enriched expression patterns. For the seven sncRNAs associated with DKD
221 none had its highest expression in kidney tissue. piRNA-019825 was the only sncRNA that showed
222 high expression in urine. Similar observations were made for eGFR and UACR associated sncRNAs
223 (Supplemental Fig. 10).

224 Finally, since the function of most of the sncRNAs we found in our study is unknown we
225 compared the expression levels of the sncRNAs with expression of 1195 proteins measured in plasma
226 from the same date to identify potential targets or (co-)regulated proteins or pathways. As shown in
227 Supplemental Table 6 we found for the seven sncRNAs 31 FDR significant pairwise correlations
228 involving 24 proteins. The strongest association was observed between piR-019825 and TNF sR-1
229 (TNFRSF1A, Beta=-0.080 (-0.112 - -0.040), $P_{FDR}=3.04 \times 10^{-2}$). piRNA-019825 also showed the largest
230 number of correlations (n=11, Figure 2A). Interestingly, among the proteins associated with piR-
231 019825 was Cystatin C, a known marker of kidney function. Next we assessed whether the 24 proteins
232 were also significantly associated with eGFR or UACR ($FDR \leq 0.05$). Six were associated with eGFR
233 and none with UACR. Four out of the six proteins that were negatively associated with piR-019825
234 expression were also negatively associated with eGFR (all $P_{FDR} \leq 3.78 \times 10^{-11}$) (Figure 2A, Supplemental
235 Table 7). Two proteins were associated with miR-143-5p expression and another eleven proteins were
236 associated with snoRNA expression. Five proteins were correlated with two or more snoRNAs
237 (NAMPT, UFM1, PDGFB, CCL27 and CCL28) and two were also significantly associated with eGFR
238 (CCL27, RARRES2, Figure 2B). Pairwise correlations between the piR-019825 or snoRNA associated
239 proteins is shown in figures 2C, D

240

241 **Discussion**

242 In this study we show that circulating sncRNAs are associated with prevalent DKD in people with
243 type 2 diabetes. Importantly these sncRNAs do not only represent miRNAs but also other classes of
244 small non-coding RNAs. In addition, these sncRNAs correlate with proteins that function in pathways
245 known to be involved in DKD/kidney function.

246

247 *Associations between kidney function and sncRNAs*

248 *miRNAs*

249 MicroRNAs are involved in the post-transcriptional regulation of gene expression and are known to be
250 involved in crosstalk between (distant) cells and tissues [2]. As such, there is an ever growing interest
251 to study their role in disease development and as biomarkers of disease and disease progression.

252 In our study one miRNA, miR-143-5p, was associated with both DKD and UACR. Levels of
253 this miRNA are lower in both DKD and persons with high UACR. miR-143-5p remained significant
254 in the full models, suggesting it might have biomarker potential. Previous in vitro studies showed that
255 this miRNA is associated with vascular pathology [19], an impaired glomerular filtration barrier [20]
256 and with mesangial cell proliferation and fibrosis [21, 22]. miR-143-5p plasma levels associate with
257 two proteins but neither of those have a clear relationship with kidney function nor were they
258 predicted targets for miR-143-5p (miRDB, miRTarBase, MiRBase). In the full model three additional
259 microRNAs were significant, miR-455-3p, miR-122-3p and miR-193b-3p. Previously it was shown
260 that miR-455-3p was down regulated in human mesangial cells (HMC) and human proximal tubule
261 epithelial cells (HK-2) under DKD stimulating conditions [23]. Interestingly, it was also shown that
262 overexpression of this miR could alleviate the pathological changes such as renal fibrosis in a rat DN
263 model [23]. In a mouse mesangial cell line miR-455-3p improved cell proliferation, inflammation and
264 ECM accumulation induced by high glucose treatment [24]. Involvement of this miR in fibrosis in
265 other tissues like liver and lung has also been reported. A previous report indicated that miR-122, a
266 liver specific miR associated with liver injury, was reduced in patients with end stage renal disease but
267 not in CKD when compared to healthy controls [25]. miR-193b-3p was identified as a potential
268 biomarker for pre-diabetes and also for CKD after nephrectomy for renal cell carcinoma [26, 27].

269 In addition to the association with DKD we also found many miRNAs associated with eGFR
270 and or UACR. A large number of which remained significant in the full models. Interestingly, of the
271 154 miRNAs found in the base models for eGFR and UACR only two of them were associated with
272 both phenotypes suggesting that the associated miRNAs are specific to either defect in kidney function
273 but not both.

274

275 *piRNAs*

276 PIWI-interacting RNAs are involved in silencing of transposons but they can also be involved
277 in the regulation of gene expression in a miRNA-like manner [28]. One piRNA, piRNA-019825
278 (DQ597218), was significantly associated with DKD. Previously, it was shown that this piRNA is one
279 of the most abundant piRNA species in urine and plasma [29]. In our proteomics analysis to identify
280 potential targets or pathways we found eleven proteins whose expression was correlated with the
281 levels of piR-019825 including cystatin C, a known marker for kidney function. Four of the
282 negatively associated proteins are also strongly associated with eGFR. This is in line with a putative
283 miRNA like function for piR-019825 in which lower piRNA expression would result in higher protein
284 levels, lower eGFR and increased risk of DKD. This is further corroborated by the fact that increased
285 levels of two of these proteins were previously found to increase risk of CKD or end stage renal
286 disease (ESRD) in prospective studies in persons with diabetes (TNF sR-I (TNFRSF1A), NBL1
287 (DAN) [30-33]. Whereas the other negatively associated proteins we found associated with this
288 piRNA (CST3, sCXCL16, CFD, PLA2G2A) have almost all also been implicated previously in kidney
289 dysfunction [34-36]. Interestingly, we also find various other piRNAs associated with either eGFR and
290 or UACR (10-14%). Most of which remained significant in the full models. Suggesting that they may
291 be useful early markers of DKD. Of note, eGFR-associated piRNAs often show their highest
292 expression in urine, whereas UACR associated piRNAs show low expression in urine, but often high
293 levels in tissues like pancreas, liver and colon.

294

295 *snoRNAs*

296 Small Nucleolar RNAs (snoRNA) are predominantly localized to the nucleolus where they
297 have a function in rRNA biogenesis, structure and activity [37]. There are two major types of
298 snoRNAs containing either box C/D or box H/ACA motifs. In our study we only detected box C/D
299 snoRNAs associated with DKD. Box C/D snoRNAs are involved in 2'-O-methylation of specific
300 sequences in target RNA sequences, mainly rRNA, thereby affecting RNA structure, stability and
301 interactions. In addition, other functions have been ascribed to the snoRNAs such as mRNA splicing
302 and editing. SnoRNA derived fragments have also been described and it has been speculated that they
303 exert piRNA or miRNA like functions [38].

304 Although snoRNAs represent only ten percent of the total circulating sncRNA transcriptome in our
305 study they represent seventy-one percent of the sncRNAs found associated with DKD suggesting that
306 snoRNAs may function as early biomarkers of DKD. SnoRNA U8 was associated with all three
307 phenotypes but has not been linked to kidney function previously. SnoRNAs SNORD118, SNORD24,
308 SNORD107 and SNORD87 were associated with DKD and UACR but not eGFR. SnoRNAs U8 and
309 SNORD118 both remained significant in the full models. SNORD87 and SNORD107 associate with
310 the same set of proteins, including UFM1, NAMPT, PDGFB, and the two chemokines CCL27 and
311 CCL28, which bind to the same receptor (CCR10). A recent report by Chen et al. showed that post-
312 transcriptional stabilization of the CCL28 mRNA by N6-methyladenosine (m6A) modification

313 increases protein levels of this chemokine and reduces kidney damage after ischemia reperfusion
314 injury [39]. Given the known role of snoRNAs in post-transcriptional RNA modification and
315 stabilization via 2'-O-methylation this might provide a putative mechanism and explanation for the
316 observed snoRNA-protein associations. The other proteins associated with these snoRNAs (UFM1,
317 NAMPT, PDGFB) have also been implicated in kidney function previously [40-43]. Further studies
318 are needed to investigate the role of these snoRNAs in DKD. In addition to the snoRNAs described
319 above various other snoRNAs are associated with eGFR and UACR (13-31% of total number of
320 significant sncRNA). Almost all remained significant in the full models. At present there are no other
321 studies that have investigated the role of snoRNAs in diabetic DKD.

322

323 *Members of other sncRNA classes*

324 Although other classes of sncRNAs, such as Y_RNA derived fragments, tRNA fragments (tRF),
325 lncRNA fragments and snRNAs represent roughly twenty percent of all circulating sncRNA detected
326 we did not find any significant associations with DKD. We did find a small number of associations
327 with eGFR and UACR though.

328

329 *Strengths and limitations*

330 Our study has several strengths. It is one of the largest circulating sncRNA profiling studies studying
331 the role of circulating sncRNAs in DKD. This improves power and reduces the chance of false
332 positives. In addition, it uses untargeted RNA sequencing allowing not only the study of miRNAs but
333 also several other classes of sncRNA showing that especially these other types of sncRNAs associated
334 with DKD and its components eGFR and UACR. The use of early, mainly mild cases of DKD
335 facilitates the identification of early markers of disease and may help to introduce early interventions
336 aiming to preserve kidney function in those at risk.

337 A limitation of our study is the use of a total RNA isolation procedure and as such we were not able to
338 investigate if the sncRNAs are carried by extracellular vesicles or are associated with carriers such as
339 HDL and Ago-2 [2]. The use of plasma for RNA isolation is another limitation. Although from a
340 biomarker perspective this is not a limitation, from a functional perspective the role of the sncRNAs
341 we identified needs to be established in the relevant cells and tissues. Another limitation of our study
342 is its cross-sectional, observational design. As such we cannot exclude that (part of) the observed
343 changes and associations are a reflection of the reduced kidney function in the participants.

344

345 In conclusion, in this study we have shown that several circulating sncRNAs are associated with DKD.
346 Interestingly, only a relatively small fraction belong to the well-studied microRNAs. This highlights
347 the potential involvement of other classes of sncRNAs, such as snoRNAs and piRNAs, that have so far
348 not been investigated in relation to kidney function before. Future prospective and functional studies

349 are needed to establish the role of these novel circulating sncRNAs in subjects with diabetes and
350 chronic kidney disease.

351

352

353 **Statements and Declarations**

354 **Conflict of Interest**

355 The authors declare that they have no conflict of interest.

356

357 **Acknowledgements**

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

360

361 **Funding**

362 This work was supported by a diabetes breakthrough grant from the Dutch Diabetes Research
363 Foundation and ZonMw (grant number 459001015) and a grant from the EFSD/Boehringer Ingelheim
364 European Research Programme on “Multi-System Challenges in Diabetes” 2021. The proteomic
365 analyses were funded by a grant from the Innovative Medicines Initiative 2 Joint Undertaking, under
366 grant agreement no. 115881 (RHAPSODY). This Joint Undertaking receives support from the
367 European Union’s Horizon 2020 research and innovation program and EFPIA and is supported by the
368 Swiss State Secretariat for Education, Research and Innovation (SERI), under contract no. 16.0097.
369 The study sponsors/funders were not involved in the design of the study; the collection, analysis, and
370 interpretation of data; writing the report; and did not impose any restrictions regarding the publication
371 of the report.

372

373 **Competing interests**

374 The authors have no relevant financial or non-financial interests to disclose.

375

376 **Author Contributions**

377 LMtH, JWJB and RCS designed the study. JAdK and SJC isolated RNA and prepared the samples for
378 RNAseq. MTB and JWJB were responsible for phenotypic data acquisition and preparation. LMtH,
379 GAB and RCS prepared, analyzed and interpreted the data. LMtH and RCS wrote the first draft of the
380 manuscript. RB contributed to the critical interpretation of the results. LMtH and RCS are the
381 guarantors of this work and, as such, had full access to all the data in the study and take responsibility
382 for the integrity of the data and the accuracy of the data analysis. All authors critically read and revised
383 the manuscript and approved the final version of the manuscript.

384

385 **Ethics approval**

386 This study was performed in line with the principles of the Declaration of Helsinki. The study has
387 been approved by the ethics committee of the Amsterdam University Medical Center, location VUmc
388 (approval number 2007/57; NL14692.029.07).

389

390 **Consent to participate**

391 All participants provided written informed consent prior to participating in this study.

392

393 **Data availability**

394 The datasets generated during and/or analyzed during the current study are not publicly available due
395 restrictions in the informed consent but are available from the corresponding author on reasonable
396 request.

397

398 **References**

- 399 [1] Persson F, Rossing P (2018) Diagnosis of diabetic kidney disease: state of the art and future
400 perspective. *Kidney Int Suppl* (2011) 8(1): 2-7. 10.1016/j.kisu.2017.10.003
- 401 [2] van Zonneveld AJ, Zhao Q, Rotmans JI, Bijkerk R (2023) Circulating non-coding RNAs in
402 chronic kidney disease and its complications. *Nat Rev Nephrol* 19(9): 573-586. 10.1038/s41581-023-
403 00725-w
- 404 [3] Loganathan TS, Sulaiman SA, Abdul Murad NA, et al. (2020) Interactions Among Non-
405 Coding RNAs in Diabetic Nephropathy. *Front Pharmacol* 11: 191. 10.3389/fphar.2020.00191
- 406 [4] Gholaminejad A, Abdul Tehrani H, Gholami Fesharaki M (2018) Identification of candidate
407 microRNA biomarkers in diabetic nephropathy: a meta-analysis of profiling studies. *J Nephrol*.
408 10.1007/s40620-018-0511-5
- 409 [5] Assmann TS, Recamonde-Mendoza M, de Souza BM, Bauer AC, Crispim D (2018)
410 MicroRNAs and diabetic kidney disease: Systematic review and bioinformatic analysis. *Mol Cell*
411 *Endocrinol* 477: 90-102. 10.1016/j.mce.2018.06.005
- 412 [6] Freedman JE, Gerstein M, Mick E, et al. (2016) Diverse human extracellular RNAs are widely
413 detected in human plasma. *Nat Commun* 7: 11106. 10.1038/ncomms11106
- 414 [7] van der Heijden AA, Rauh SP, Dekker JM, et al. (2017) The Hoorn Diabetes Care System
415 (DCS) cohort. A prospective cohort of persons with type 2 diabetes treated in primary care in the
416 Netherlands. *BMJ Open* 7(5): e015599. 10.1136/bmjopen-2016-015599
- 417 [8] KDIGO (2013) Kidney Disease: Improving Global Outcomes (KDIGO). CKD work group.
418 KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney
419 Disease. *Kidney Int* 3: 1-150
- 420 [9] Rozowsky J, Kitchen RR, Park JJ, et al. (2019) exceRpt: A Comprehensive Analytic Platform
421 for Extracellular RNA Profiling. *Cell Syst* 8(4): 352-357.e353. 10.1016/j.cels.2019.03.004
- 422 [10] Griffiths-Jones S (2004) The microRNA Registry. *Nucleic Acids Res* 32(Database issue):
423 D109-111. 10.1093/nar/gkh023
- 424 [11] Sai Lakshmi S, Agrawal S (2008) piRNABank: a web resource on classified and clustered
425 Piwi-interacting RNAs. *Nucleic Acids Res* 36(Database issue): D173-177. 10.1093/nar/gkm696
- 426 [12] Harrow J, Frankish A, Gonzalez JM, et al. (2012) GENCODE: the reference human genome
427 annotation for The ENCODE Project. *Genome Res* 22(9): 1760-1774. 10.1101/gr.135350.111
- 428 [13] Glažar P, Papavasileiou P, Rajewsky N (2014) circBase: a database for circular RNAs. *Rna*
429 20(11): 1666-1670. 10.1261/rna.043687.113
- 430 [14] Chan PP, Lowe TM (2009) GtRNADB: a database of transfer RNA genes detected in genomic
431 sequence. *Nucleic Acids Res* 37(Database issue): D93-97. 10.1093/nar/gkn787
- 432 [15] ThermoFisherScientific (2016) A technical guide to identifying miRNA normalizers using
433 TaqMan Advanced miRNA Assays. Available from [https://assets.thermofisher.com/TFS-
434 Assets/GSD/Reference-Materials/identifying-mirna-normalizers-white-paper.pdf](https://assets.thermofisher.com/TFS-Assets/GSD/Reference-Materials/identifying-mirna-normalizers-white-paper.pdf)
- 435 [16] Barrett T, Wilhite SE, Ledoux P, et al. (2012) NCBI GEO: archive for functional genomics
436 data sets—update. *Nucleic Acids Research* 41(D1): D991-D995. 10.1093/nar/gks1193
- 437 [17] Slieker RC, Donnelly LA, Fitipaldi H, et al. (2021) Distinct Molecular Signatures of Clinical
438 Clusters in People with Type 2 Diabetes: an IMIRHAPSODY Study. *Diabetes* 70(11): 2683-2693.
439 10.2337/db20-1281

- 440 [18] Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: a Bioconductor package for
441 differential expression analysis of digital gene expression data. *Bioinformatics* 26(1): 139-140.
442 10.1093/bioinformatics/btp616
- 443 [19] Taïbi F, Metzinger-Le Meuth V, M'Baya-Moutoula E, et al. (2014) Possible involvement of
444 microRNAs in vascular damage in experimental chronic kidney disease. *Biochim Biophys Acta*
445 1842(1): 88-98. 10.1016/j.bbadis.2013.10.005
- 446 [20] Müller-Deile J, Gellrich F, Schenk H, et al. (2016) Overexpression of TGF- β Inducible
447 microRNA-143 in Zebrafish Leads to Impairment of the Glomerular Filtration Barrier by Targeting
448 Proteoglycans. *Cell Physiol Biochem* 40(5): 819-830. 10.1159/000453142
- 449 [21] Bai S, Xiong X, Tang B, et al. (2020) Exosomal circ_DLGAP4 promotes diabetic kidney
450 disease progression by sponging miR-143 and targeting ERBB3/NF- κ B/MMP-2 axis. *Cell Death Dis*
451 11(11): 1008. 10.1038/s41419-020-03169-3
- 452 [22] Ge X, Xi L, Wang Q, et al. (2020) Circular RNA Circ_0000064 promotes the proliferation and
453 fibrosis of mesangial cells via miR-143 in diabetic nephropathy. *Gene* 758: 144952.
454 10.1016/j.gene.2020.144952
- 455 [23] Wu J, Liu J, Ding Y, et al. (2018) MiR-455-3p suppresses renal fibrosis through repression of
456 ROCK2 expression in diabetic nephropathy. *Biochem Biophys Res Commun* 503(2): 977-983.
457 10.1016/j.bbrc.2018.06.105
- 458 [24] Zhu XJ, Gong Z, Li SJ, Jia HP, Li DL (2019) Long non-coding RNA Hottip modulates high-
459 glucose-induced inflammation and ECM accumulation through miR-455-3p/WNT2B in mouse
460 mesangial cells. *Int J Clin Exp Pathol* 12(7): 2435-2445
- 461 [25] Rivoli L, Vliegthart AD, de Potter CM, et al. (2017) The effect of renal dysfunction and
462 haemodialysis on circulating liver specific miR-122. *Br J Clin Pharmacol* 83(3): 584-592.
463 10.1111/bcp.13136
- 464 [26] Mollet IG, Macedo MP (2023) Pre-Diabetes-Linked miRNA miR-193b-3p Targets
465 PPARGC1A, Disrupts Metabolic Gene Expression Profile and Increases Lipid Accumulation in
466 Hepatocytes: Relevance for MAFLD. *Int J Mol Sci* 24(4). 10.3390/ijms24043875
- 467 [27] Trevisani F, Ghidini M, Larcher A, et al. (2016) MicroRNA 193b-3p as a predictive biomarker
468 of chronic kidney disease in patients undergoing radical nephrectomy for renal cell carcinoma. *Br J*
469 *Cancer* 115(11): 1343-1350. 10.1038/bjc.2016.329
- 470 [28] Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD (2019) PIWI-interacting RNAs:
471 small RNAs with big functions. *Nat Rev Genet* 20(2): 89-108. 10.1038/s41576-018-0073-3
- 472 [29] El-Mogy M, Lam B, Haj-Ahmad TA, et al. (2018) Diversity and signature of small RNA in
473 different bodily fluids using next generation sequencing. *BMC Genomics* 19(1): 408. 10.1186/s12864-
474 018-4785-8
- 475 [30] Gohda T, Niewczas MA, Ficociello LH, et al. (2012) Circulating TNF receptors 1 and 2
476 predict stage 3 CKD in type 1 diabetes. *J Am Soc Nephrol* 23(3): 516-524. 10.1681/asn.2011060628
- 477 [31] Niewczas MA, Gohda T, Skupien J, et al. (2012) Circulating TNF receptors 1 and 2 predict
478 ESRD in type 2 diabetes. *J Am Soc Nephrol* 23(3): 507-515. 10.1681/asn.2011060627
- 479 [32] Kobayashi H, Looker HC, Satake E, et al. (2022) Neuroblastoma suppressor of tumorigenicity
480 1 is a circulating protein associated with progression to end-stage kidney disease in diabetes. *Sci*
481 *Transl Med* 14(657): eabj2109. 10.1126/scitranslmed.abj2109
- 482 [33] Niewczas MA, Pavkov ME, Skupien J, et al. (2019) A signature of circulating inflammatory
483 proteins and development of end-stage renal disease in diabetes. *Nat Med* 25(5): 805-813.
484 10.1038/s41591-019-0415-5
- 485 [34] Christensson A, Ash JA, DeLisle RK, et al. (2018) The Impact of the Glomerular Filtration
486 Rate on the Human Plasma Proteome. *Proteomics Clin Appl* 12(3): e1700067.
487 10.1002/prca.201700067
- 488 [35] Jalal D, Sanford B, Renner B, et al. (2021) Detection of pro angiogenic and inflammatory
489 biomarkers in patients with CKD. *Sci Rep* 11(1): 8786. 10.1038/s41598-021-87710-0
- 490 [36] Ngo D, Wen D, Gao Y, et al. (2020) Circulating testican-2 is a podocyte-derived marker of
491 kidney health. *Proc Natl Acad Sci U S A* 117(40): 25026-25035. 10.1073/pnas.2009606117
- 492 [37] Deogharia M, Majumder M (2018) Guide snoRNAs: Drivers or Passengers in Human
493 Disease? *Biology (Basel)* 8(1). 10.3390/biology8010001

- 494 [38] Bergeron D, Fafard-Couture É, Scott MS (2020) Small nucleolar RNAs: continuing
495 identification of novel members and increasing diversity of their molecular mechanisms of action.
496 *Biochem Soc Trans* 48(2): 645-656. 10.1042/bst20191046
- 497 [39] Chen J, Xu C, Yang K, et al. (2023) Inhibition of ALKBH5 attenuates I/R-induced renal
498 injury in male mice by promoting Ccl28 m6A modification and increasing Treg recruitment. *Nat*
499 *Commun* 14(1): 1161. 10.1038/s41467-023-36747-y
- 500 [40] Ostendorf T, Boor P, van Roeyen CR, Floege J (2014) Platelet-derived growth factors
501 (PDGFs) in glomerular and tubulointerstitial fibrosis. *Kidney Int Suppl* (2011) 4(1): 65-69.
502 10.1038/kisup.2014.12
- 503 [41] Li H, Li M, Liu C, et al. (2023) Causal effects of systemic inflammatory regulators on chronic
504 kidney diseases and renal function: a bidirectional Mendelian randomization study. *Front Immunol* 14:
505 1229636. 10.3389/fimmu.2023.1229636
- 506 [42] Hu X, Zhang H, Zhuang L, et al. (2020) Ubiquitin-Fold Modifier-1 Participates in the Diabetic
507 Inflammatory Response by Regulating NF- κ B p65 Nuclear Translocation and the Ubiquitination and
508 Degradation of I κ B α . *Drug Des Devel Ther* 14: 795-810. 10.2147/dddt.S238695
- 509 [43] Muraoka H, Hasegawa K, Sakamaki Y, et al. (2019) Role of Nampt-Sirt6 Axis in Renal
510 Proximal Tubules in Extracellular Matrix Deposition in Diabetic Nephropathy. *Cell Rep* 27(1): 199-
511 212.e195. 10.1016/j.celrep.2019.03.024
- 512
- 513
- 514

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

| | all | No DKD | DKD |
|---|-------------------|-------------------|--------------------|
| 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) |

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

517 * P≤0.05 versus those without DKD.

518

519 **Table 2. Frequency of different types of sncRNAs associated with the three endpoints.**

| | Total # of sncRNA detected | | DKD | | | eGFR | | | UACR | | |
|-----------------|----------------------------|------------|----------|----|----------------------|------------|----|-------|------------|-----|-----------------------|
| | n | % | n | % | p* | n | % | p* | n | % | p* |
| Total # sncRNAs | 938 | 100 | 7 | | | 107 | | | 220 | | |
| miRNA | 602 | 64 | 1 | 14 | 0.011 | 63 | 59 | 0.29 | 93 | 42 | 4.5x10 ⁻⁹ |
| snoRNA | 96 | 10 | 5 | 71 | 2.2x10 ⁻⁴ | 14 | 13 | 0.40 | 67 | 31 | 8.4x10 ⁻¹³ |
| lncRNA | 93 | 10 | | | | 13 | 12 | 0.50 | 20 | 9 | 0.80 |
| piRNA | 45 | 5 | 1 | 14 | 0.30 | 11 | 10 | 0.037 | 23 | 11 | 4.0x10 ⁻³ |
| Y_RNA fragments | 33 | 4 | | | | | | | 1 | 0.5 | 0.013 |
| others | 26 | 3 | | | | | | | 6 | 3 | 1.00 |
| tRNA fragments | 24 | 3 | | | | 4 | 4 | 0.52 | 1 | 0.5 | 0.067 |
| snRNA | 12 | 1 | | | | 1 | 1 | 1.00 | 6 | 3 | 0.13 |
| scaRNA | 7 | 1 | | | | 1 | 1 | 0.58 | 3 | 1 | 0.41 |

* P-value for comparison against the distribution of all sncRNAs detected.

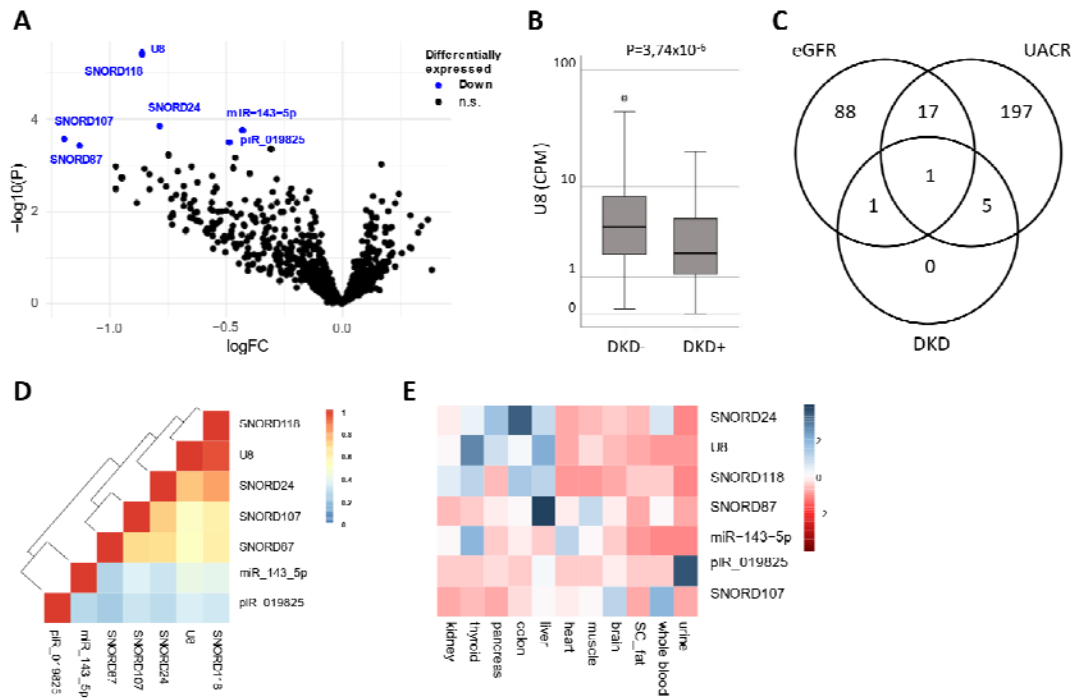
520
521
522

523 **Table 3. Overview of seven sncRNAs associated with prevalent DKD**

| sncRNA | Prevalent DKD | | eGFR | | UACR | |
|---------------------|---------------|------------------------|-------|------------------------|--------|------------------------|
| | LogFC | P _{FDR} | logFC | P _{FDR} | logFC | P _{FDR} |
| U8 | -0,859 | 1,90x10 ⁻⁰³ | 0,012 | 2,79x10 ⁻⁰² | -0,724 | 6,74x10 ⁻⁰³ |
| SNORD118 | -0,861 | 1,90x10 ⁻⁰³ | 0,009 | 8,94x10 ⁻⁰² | -0,858 | 1,73x10 ⁻⁰³ |
| miR-143-5p | -0,429 | 4,01x10 ⁻⁰² | 0,000 | 9,86x10 ⁻⁰¹ | -0,376 | 3,30x10 ⁻⁰² |
| SNORD24 | -0,785 | 4,01x10 ⁻⁰² | 0,011 | 6,42x10 ⁻⁰² | -0,919 | 1,73x10 ⁻⁰³ |
| piR_019825 DQ597218 | -0,485 | 4,91x10 ⁻⁰² | 0,012 | 3,04x10 ⁻⁰³ | -0,129 | 5,71x10 ⁻⁰¹ |
| SNORD107 | -1,194 | 4,91x10 ⁻⁰² | 0,013 | 1,43x10 ⁻⁰¹ | -2,214 | 1,58x10 ⁻⁰⁵ |
| SNORD87 | -1,129 | 4,91x10 ⁻⁰² | 0,012 | 2,17x10 ⁻⁰¹ | -2,253 | 1,18x10 ⁻⁰⁵ |

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

525 **Figure 1. sncRNAs significantly associated with DKD.**



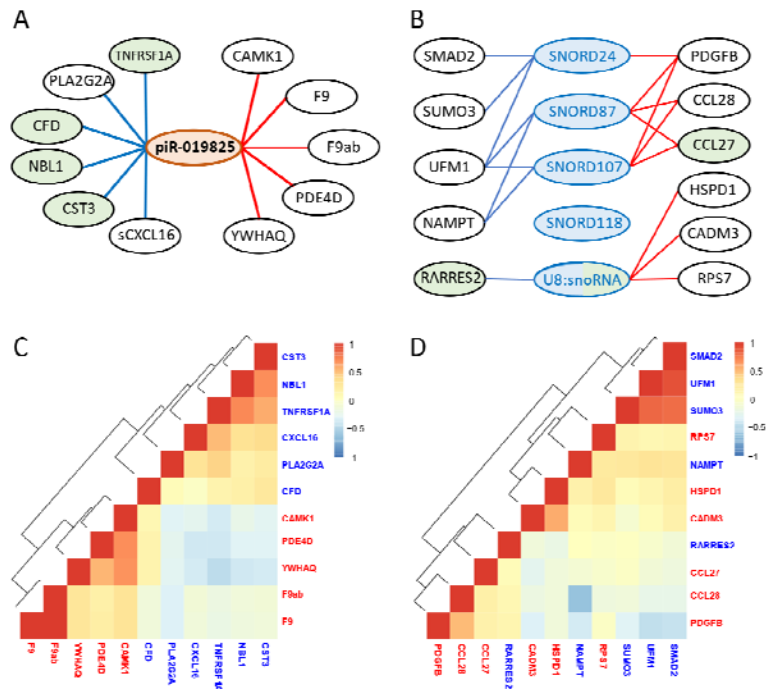
526

527 Panel A. Volcano plot showing the association between sncRNAs and DKD in the base model. sncRNAs in blue are
 528 significantly lower in cases with DKD after multiple hypothesis testing ($FDR \leq 0.05$). $\log_{2}FC$ is the \log_{2} of the fold change
 529 between cases and controls. Panel B. U8 levels in those with or without DKD. CPM = counts per million reads. Panel C.
 530 Overlap between sncRNAs significantly associated with DKD, eGFR or UACR in the base model. Panel D. Pairwise
 531 correlations for the seven sncRNAs associated with DKD. Colors represent Spearman's rho. Panel E. Expression pattern for
 532 sncRNAs associated with DKD in publicly available datasets for 11 tissues of interest. Colors represent the Z-scaled
 533 expression of a sncRNA where blue indicates relative high expression and red relative low expression.

534

535

536 **Figure 2. sncRNAs associated proteins.**



537

538

539 Overview of the proteins that associate with sncRNA expression levels in plasma from persons with T2D. sncRNAs and
 540 proteins in green are also significantly associated with eGFR and those in blue are associated with UACR ($P_{FDR} \leq 0.05$). Blue
 541 line and text indicate a negative association and a red line and text indicates a positive association. Panel A. piR-019825
 542 associated proteins ($P_{FDR} \leq 0.05$). Panel B. snoRNA associated proteins ($P_{FDR} \leq 0.05$). Panel C. Pairwise correlations between
 543 the 11 piR-019825 associated proteins. Panel D. Pairwise correlations between the 11 snoRNA associated proteins.

544

545