Research Article

Various types of circulating small non-coding RNAs associate with kidney function in people with type 2 diabetes

Running title: Small non-coding RNA and kidney function in type 2 diabetes

LM 't Hart[a, b, c, d] (ORCID 0000-0003-4401-2938), JA de Klerk[a, e, f], GA Bouland[a], JHD Peerlings[a], M.T. Blom[c, g], SJ Cramer[a], R Bijkerk[e, f], JWJ Beulens[b, c], RC Slieker[a, b, c]

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

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

[c] Amsterdam Public Health, Amsterdam, The Netherlands

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

[e] Department of Internal Medicine, Division of Nephrology, Leiden University Medical Center, Leiden, The Netherlands

[f] Einthoven Laboratory for Vascular and Regenerative Medicine, Leiden University Medical Center, Leiden, The Netherlands

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

Author for correspondence:

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

Keywords: chronic kidney disease, type 2 diabetes, small non-coding RNA, microRNA, snoRNA

Abstract

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

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

Results In total seven sncRNAs were associated with prevalent DKD ($P_{FDR} \leq 0.05$). Although miRNAs represent the majority of the sncRNAs measured (64%) only one miRNA was significantly associated while the majority of the significant sncRNA belonged to the snoRNA class (71%). Similar results were observed for eGFR and UACR. In addition, the seven sncRNAs, and especially the piRNA piR-019825, were associated with plasma levels of 80 proteins of which several have known associations with kidney function.

Conclusion We have identified novel small non-coding RNAs, primarily from classes other than microRNAs, that are associated with DKD in persons with type 2 diabetes. Further studies, in particular for snoRNAs and the piRNA piR-019825 that have never been studied before, are warranted to explore their biomarker potential and elucidate their biological role in kidney function.

Introduction

Diabetic kidney disease (DKD) 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 DKD 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 DKD development and have yielded novel biomarkers and drug targets.

Circulating small non-coding RNAs (sncRNA) are mostly contained in extracellular vesicles or are bound to RNA-binding proteins and lipoproteins and originate from various tissues including tissues important in kidney function [5]. The composition of sncRNAs in the circulation is dependent on tissue homeostasis and can be altered by disease development. Whereas previously it was thought that sncRNAs 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, splicing and ribosomal RNA maturation [5]. Irrespective of the mechanism involved, sncRNAs may have an impact on the development and progression of DKD when deregulated. Importantly, each tissue releases its own combination of tissue-specific and more ubiquitous sncRNAs into the bloodstream. These circulating sncRNAs can be taken up by other tissues and affect target tissue function [5].

Although a substantial number of studies have examined circulating microRNAs associated with kidney function in persons with diabetes, most studies are relatively small (n<100). From these studies a large number of miRNAs have emerged that are putatively associated with kidney function. However, probably due to small sample sizes, heterogeneity in endpoints and methodology used there is limited consistency between studies as shown in various meta-analyses [6-8].

To the best of our knowledge there are currently no studies that have systematically examined the other classes (biotypes) of small ncRNA. Therefore, we have performed a large untargeted small non-coding RNA-seq study in a large well defined sample of persons with T2D with and without DKD from the Hoorn DCS study. Importantly, we go beyond the current state-of-the-art by not only investigating microRNAs but also other sncRNAs classes, including besides miRNAs, piRNAs, tRNA fragments (tRFs), lncRNA fragments, Y-RNAs and snoRNAs. For this we used next-generation sequencing to capture the whole sncRNA transcriptome allowing for a hypothesis-free, comprehensive analysis to study the role of all types of sncRNAs in diabetic kidney disease [9].

55 **Methods**

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

57

58 *Study sample*

59 For this project we used data from the Hoorn Diabetes Care System cohort (DCS), a large prospective
60 study comprised of 14,000+ individuals with type 2 diabetes who receive structed diabetes care [10].
61 Participants visit the Hoorn diabetes research center annually for diabetes care including routine
62 biochemistry, medication review and clinical care. Patients were also invited to participate in DCS
63 biobanks in which we combine the annual phenotypic data with biobanking of liquid biopsies which
64 includes serum, plasma and urine. Blood samples were collected during two phases of biobanking in
65 2008/2009 and 2012-2014. For the current study we used a subsample from the whole DCS biobank
66 (n=475) using the following inclusion criteria: a plasma sample collected within four years after
67 diagnosis of diabetes and no-insulin use at the time of blood sampling.
68 All laboratory measurements were on samples taken in a fasted state on the same day as the sample
69 for sRNA-seq. Routine anthropometric, biochemistry methods and definitions have been described
70 previously in detail in van der Heijden et al. [10]. Kidney function was assessed based on eGFR (CKD-
71 EPI) and urinary albumin to creatinine ratio (UACR) and included as continuous variables in our
72 analyses. Stages of chronic kidney disease (CKD) were defined according the KDIGO criteria as
73 previously described [11]. Cases were defined as having CKD stage ≥ 1 at the time of blood sampling
74 for sRNA analysis which resulted in 69 cases.

75

76 *Circulating small RNA sequencing*

77 Circulating sncRNAs were isolated from 900ul of citrate plasma using a commercial cell-free RNA
78 isolation kit (Quick-cfRNA Serum & Plasma Kit, Zymo Research, Irvine, CA, USA). After quality control
79 475 samples were submitted for sRNA sequencing. Library preps were constructed using the
80 NEBNext® Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs, Ipswich, MA,
81 USA). A Pippin Prep instrument (Sage Science, Beverly, MA,USA) was used for size selection to further
82 enrich for sncRNAs with a size of ~20 to 100 nucleotides. Sequencing was performed using the
83 Illumina NextSeq500 sequencer with v2.5 sequencing reagent kits and 35 cycle paired-end reads
84 (PE35). On average 11.3±2.8 million reads per sample were generated. The exceRpt pipeline [12],
85 developed by the NIH Extracellular RNA Communication Consortium (ERCC), was used to map the
86 obtained reads, for quality control, to process the data and generate RNA abundance estimates for
87 miRNA, piRNA, lncRNA, tRNA fragments, Y_RNA (fragments), snRNA, snoRNA, scaRNA and various

88 other sncRNA species. Using respectively miRBase v22 [13], piRNABank v1 [14], Gencode (version 38)
89 [15], circBase [16] and GtRNAdb [17]. One sample did not pass the QC thresholds.
90
91 *Taqman validation of the RNA-seq results*
92 We used Taqman assays (Fisher Scientific, NL) to verify the validity and reproducibility of the small
93 RNA-seq results. We choose four miRNAs with a range from low to relative high expression in our
94 study sample (miR-542-3p, miR-323b-3p, miR-186-5p and miR-423-5p) which were measured in 12
95 plasma samples for which sncRNA-seq data were available. The Taqman assays were run according to
96 the manufacturer's instructions. miR-191-3p expression was used as the reference (housekeeping
97 gene) in both the Taqman and small RNA-seq datasets as suggested by the manufacturer.
98
99 *Expression patterns of small RNAs in tissues*
100 Publicly available raw sncRNA-seq data for various metabolic tissues were downloaded from the
101 Gene Expression Omnibus (GEO) [18]. The tissues for which sncRNA-seq data was used included
102 kidney, thyroid, brain, liver, muscle, heart, colon, pancreas, subcutaneous white adipose tissue and
103 whole blood and urine. Accession numbers are given in Supplemental Table 1. Data were processed
104 using the exceRpt pipeline as described above. The tissue data were used to estimate the abundance
105 of the sncRNAs in these tissues. Data was z-scaled in figures.
106
107 *Plasma proteomics.*
108 Plasma proteomic data was available for 589 participants of the Hoorn DCS study and generated
109 using the SomaLogic SOMAscan platform (Boulder, CO) as previously described [19]. After quality
110 control 1195 protein levels were available for analysis. For 199 subjects we also had the
111 corresponding sncRNA-seq data from the same sampling date.
112
113 *Statistical analysis*
114 The R-package edgeR (v3.42.4) was used to identify differentially expressed sncRNAs [20]. A negative
115 binomial generalized log-linear model was applied using edgeR's function *estimateGLMRobustDisp*.
116 The base model was unadjusted. A fully adjusted model was adjusted for BMI, diabetes duration,
117 HbA1c, HDL, SBP and smoking unless otherwise stated. Age and sex were excluded from the model as
118 these are part of the definition of eGFR. UACR was log10-transformed before analysis. In sensitivity
119 analyses, we additionally included use of antihypertensive medication and or eGFR or UACR. The
120 Benjamini-Hochberg procedure was used to adjust for multiple testing and an FDR below 0.05 was
121 considered significant. The correlation between the sncRNA transcriptome and plasma proteome

122 was tested using a linear model with the log transformed proteomic measurements as outcome and
123 the sncRNA transcriptome as predictors with adjustment for age, BMI, sex and technical covariates.
124 In this exploratory analysis we used an arbitrary significance threshold of $p \leq 1.00 \times 10^{-3}$.
125

126 **Results**

127 *Clinical characteristics*
128 The clinical characteristics of our study group are shown in Supplemental Table 2. The characteristics
129 resemble those of a typical T2D population with slightly more males than females, a mean age-at-
130 diagnosis of 59.3 ± 9.0 years, BMI of 30.7 ± 6.0 kg/m2 and an average duration of diabetes at time of
131 blood sampling of 2.2 ± 1.1 years. Glycemic control was good with an average HbA1c of 46 (42, 51)
132 mmol/mol. Prevalence of DKD at the time of blood sampling was 15%.
133

134 *sncRNA biotype abundance in our samples*
135 After quality control, 938 sncRNAs were expressed above the threshold of a median ≥5 copies in our
136 data, which included miRNAs, piRNA, Y_RNA fragments, lncRNA fragments, tRNA_fragments,
137 snoRNAs, snRNAs and various other sRNA species. miRNAs, Y-RNAs and tRFs were the most abundant
138 sncRNAs biotypes comprising respectively 57.6, 34.9 and 4.8 percent of all reads (Table 1). Other
139 sncRNA biotypes were much less abundant (≤1.4%). There were no significant differences in the total
140 number of reads per biotype between cases and controls (all P>0.4).
141

142 *Taqman validation shows good reproducibility*
143 We next sought to verify validity and reproducibility of the sncRNA-seq. Taqman assays for sncRNAs
144 with high, medium and low expression were used and these showed a good correlation with the
145 RNA-seq results (r^2= 0.86, Supplemental Fig. 2).
146

147 *sncRNAs associate with DKD, eGFR and UACR*
148 Seven sncRNAs were significantly associated with prevalent DKD in the unadjusted model ($P_{FDR} \leq 0.05$,
149 Table 2, Fig. 1A). All identified sncRNAs were downregulated in DKD cases. The top hit was snoRNA
150 U8 (log2 fold change=-0,859, $P_{FDR}=1.90 \times 10^{-3}$, Fig. 1B). Five of these sncRNAs also showed a significant
151 trend for lower levels in those with more advanced stages of CKD ($P_{FDR} \leq 0.05$, Supplemental Fig. 3).
152 However, due to the small sample size this was non-significant when analyzed only in the cases.
153 Respectively, 107 and 220 sncRNAs were significantly associated with eGFR and UACR, which are
154 used for the DKD classification (Table 3, Supplemental Table 3, Supplemental Fig. 4 and 5). The

155 overlap between the sncRNAs associated with DKD and its components eGFR and UACR is shown in
156 Fig. 1C and Table 2. snoRNA U8 overlapped between all three endpoints used and piR_019825
157 overlapped between eGFR and DKD. SNORD118, SNORD24, SNORD107, SNORD87 and miR-143-5p all
158 overlapped between UACR and DKD.
159 In the full model, including adjustment for BMI, diabetes duration, HbA1c, HDL, SBP and smoking,
160 respectively 7, 153 and 205 sncRNAs remained significantly associated with DKD, eGFR and UACR
161 (Supplemental Table 4, Supplemental Fig. 6; 57%; 95% and 84% overlap with the unadjusted model
162 for DKD, eGFR and UACR respectively). U8, SNORD118, piR_019825 and miR-143-5p were significant
163 in both the base and full models. In a sensitivity analysis, additional adjustment for use of
164 antihypertensive medication and baseline eGFR or UACR (only for eGFR and UACR) did not have a
165 major impact on the effect sizes but reduced the number of significant associations (r^2>=0.92,
166 Supplemental Fig. 7).
167
168 *Certain types of sncRNAs are enriched in DKD*
169 Comparing the frequency of different types of sncRNAs identified showed that while snoRNAs
170 represent 10% of the total sncRNA transcriptome in our RNA-seq study they are the most abundant
171 class associated with prevalent DKD (71%, OR=21.9, P=2.4×10^{-4}, Table 1). Enrichment for snoRNAs is
172 also seen for UACR (OR=3.8, p=8.4×10^{-13}). Associations with miRNAs on the other hand are less
173 frequent whilst piRNAs also seem enriched (Table 1). A similar pattern is again seen for UACR and to
174 a lesser extend for eGFR. In the full model a largely similar picture is observed (Supplemental Table
175 5).
176 The pairwise correlation in expression between the seven sncRNAs associated with diabetic DKD was
177 very modest (Fig. 1D). The only exception were U8 and SNORD118, which is likely due to the fact that
178 they are encoded in the same host gene (TMEM107) on chromosome 17. Results of these analyses
179 for eGFR and UACR are shown in Supplemental Fig. 8.
180
181 *sncRNAs are tissue-enriched and correlate with plasma protein levels*
182 In a next step, to elucidate the putative tissues of origin, we examined the tissue expression patterns
183 of the identified sncRNAs. We used publicly available expression data in 11 tissues or body fluids (Fig.
184 1E). This illustrated that some sncRNAs show tissue-enriched expression patterns. For the seven
185 sncRNAs associated with DKD none had its highest expression in kidney tissue. piRNA-019825 was
186 the only sncRNA that showed high expression in urine. Similar observations were made for eGFR and
187 UACR associated sncRNAs (Supplemental Fig. 9).

188 Finally, since the function of most of the sncRNAs we found in our study is unknown we compared in
189 an exploratory study the expression levels of the sncRNAs with expression of 1195 proteins
190 measured in plasma from the same date to identify potential targets or (co-)regulated proteins or
191 pathways. As shown in Supplemental Table 6 we found for the seven sncRNAs 105 nominal
192 significant pairwise correlations ($p \leq 1x10^{-3}$) involving 80 proteins. The strongest association was
193 observed between the piRNA piR-019825 and TNF sR-1 (TNFRSF1A, Beta=-0.080 (-0.112-0.040),
194 $p=4.29x10^{-6}$). piRNA-019825 also showed the largest number of correlations (n=32, Fig. 2A) whereas
195 miR-143-5p showed only four. Interestingly, among the proteins associated with piR-019825 was
196 Cystatin C, a known marker of kidney function. Next we assessed whether the 80 proteins were also
197 significantly associated with eGFR or UACR (FDR≤0.05). Fourteen were associated with eGFR and two
198 with UACR. Interestingly, ten of the fourteen eGFR associated and both UACR associated proteins
199 were also associated with piR-019825 expression (Fig. 2A). Protein-protein interactions (PPI) and
200 pathway enrichment was subsequently used for each set of sncRNA associated proteins. This showed
201 that the 32 proteins associated with piR-019825 show a strong PPI enrichment ($P=7.0x10^{-4}$, Fig. 2B).
202 Pairwise correlation between the proteins is shown in Fig. 2C. Enrichment analysis showed
203 enrichment for KEGG pathways such as *Cytokine-cytokine receptor interaction* ($P=6.7x10^{-6}$, *NF-kappa*
204 *B signaling pathway* ($P=4.6x10^{-3}$) and *PI3K-Akt signaling pathway* ($P=3.2x10^{-2}$). For the other smaller
205 sncRNA associated protein sets we note that several proteins have previously been implicated in
206 DKD. Fifteen proteins were correlated with two or more sncRNAs. Especially the snoRNAs seem to
207 have multiple overlapping protein associations (Supplemental Fig. 10). Several of these proteins have
208 already previously been implicated in kidney function (including CTACK (CCL27), Chemerin
209 (RARRES2), visfatin (NAMPT), PDE4D, PDGFB, SDF-1 (CXCL12), sICAM-3, SMAD2) whereas the first
210 two were also in our study significantly associated with eGFR.
211
212

213 **Discussion**

214 In this study we show that circulating sncRNAs are associated with prevalent DKD in people with type
215 2 diabetes. Importantly these sncRNAs do not only represent miRNAs but also other classes of small
216 non-coding RNAs. In addition the sncRNAs correlate with proteins that function in pathways known
217 to be involved in DKD/kidney function.
218
219 *Associations between kidney function and sncRNAs*
220 *miRNAs*

221 MicroRNAs are involved in the post-transcriptional regulation of gene expression and are known to
222 be involved in crosstalk between (distant) cells and tissues [5]. As such, there is an ever growing
223 interest to study their role in disease development and as biomarkers of disease and disease
224 progression.
225 In our study one miRNA, miR-143-5p, was associated with both DKD and UACR. Levels of this miRNA
226 are lower in both DKD and persons with high UACR. Analysis of the expression pattern of this miRNA
227 showed its highest expression in thyroid and heart but less in other tissues, including kidney or urine.
228 miR-143-5p remained significant after further adjustment for baseline BMI, diabetes duration,
229 HbA1c, HDL, SBP and smoking in the full models, suggesting it might have biomarker potential.
230 Analysis of the co-expressed proteins showed enrichment for the *ovarian steriodogenesis* and *GnRH*
231 *signalling pathway* but no clear relationship to DKD. However, there is quite some functional
232 evidence from in vitro studies showing that this miRNA is associated with vascular pathology [21], an
233 impaired glomerular filtration barrier [22] and with mesangial cell proliferation and fibrosis [23, 24].
234 In the full model three additional microRNAs were significant, miR-455-3p, miR-122-3p and miR-
235 193b-3p. Previously it was shown that miR-455-3p was down regulated in human mesangial cells
236 (HMC) and human proximal tubule epithelial cells (HK-2) under DKD stimulating conditions [25].
237 Interestingly, it was also shown that overexpression of this miR could alleviate the pathological
238 changes such as renal fibrosis in a rat DN model [25]. In a mouse mesangial cell line miR-455-3p
239 improved cell proliferation, inflammation and ECM accumulation induced by high glucose treatment
240 [26]. Involvement of this miR in fibrosis in other tissues like liver and lung has also been reported. A
241 previous report indicated that miR-122, a liver specific miR associated with liver injury, was reduced
242 in patients with end stage renal disease but not in CKD when compared to healthy controls [27]. miR-
243 193b-3p was identified as a potential biomarker for pre-diabetes and also for CKD after nephrectomy
244 for renal cell carcinoma [28, 29].
245 In addition to the association with DKD we also found many miRNAs associated with eGFR and or
246 UACR. A large number of which remained significant in the full models. Interestingly, of the 154
247 miRNAs found in the base models for eGFR and UACR only two of them were associated with both
248 phenotypes suggesting that the associated miRNAs are specific to either defect in kidney function but
249 not both. If confirmed in independent studies this provides interesting starting points for further
250 functional studies investigating their role in either condition.
251
252 *piRNAs*
253 PIWI-interacting RNAs are involved in silencing of transposons but they can also be involved in the
254 regulation of gene expression in a miRNA-like manner [30]. One piRNA, piRNA-019825 (DQ597218),

255 was significantly associated with DKD. Previously it was shown that this piRNA is one of the most
256 abundant piRNA species in urine and plasma [31]. In our exploratory proteomics analysis we found
257 32 proteins whose expression was correlated with the levels of piR-019825 including cystatin C, a
258 known marker for kidney function. Twelve of the proteins showed an association with eGFR or UACR.
259 In addition, we found that many of the proteins have previously been found in studies of kidney
260 function or function in disease relevant pathways such as the enriched *cytokine-cytokine receptor*
261 pathway. This includes among the most strong associations proteins such as TNF sR-I (TNFRSF1A),
262 DAN (NBL1), CXCL16 and CFD which have all previously been associated DKD or with progression to
263 ESRD [32-34]. Interestingly, we also find various other piRNAs associated with either eGFR and or
264 UACR (10-14%). Most of which remained significant in the full models. Suggesting that they may be
265 useful early markers of DKD. Of note, eGFR-associated piRNAs often show their highest expression in
266 urine, whereas UACR associated piRNAs show low expression in urine, but often high levels in tissues
267 like pancreas, liver and colon.
268
269 *snoRNAs*
270 Small Nucleolar RNAs (snoRNA) are predominantly localized to the nucleolus where they have a
271 function in rRNA biogenesis, structure and activity [35]. There are two major types of snoRNAs
272 containing either box C/D or box H/ACA motifs. In our study we only detected box C/D snoRNAs
273 associated with DKD. Box C/D snoRNAs are involved in 2'-O-methylation of specific sequences in
274 target RNA sequences, mainly rRNA. In addition other functions have been ascribed to the snoRNAs
275 such as mRNA splicing and editing. SnoRNA derived fragments have also been described and it has
276 been speculated that they exert piRNA or miRNA like functions [36].
277 Although snoRNAs represent only ten percent of the total circulating sRNA transcriptome in our
278 study they represent seventy-one percent of the sncRNAs found associated with DKD suggesting that
279 they may function as early biomarkers of DKD. The five snoRNAs that were associated with DKD are
280 expressed in most tissues but two showed their highest expression in liver or colon whilst in general
281 kidney and urine levels of these snoRNAs are much lower. SnoRNA U8 was associated with all three
282 phenotypes but has not been linked to kidney function previously. SnoRNAs SNORD118, SNORD24,
283 SNORD107 and SNORD87 where associated with DKD and UACR but not eGFR. Interestingly,
284 especially the snoRNAs seem to be associated with the same set of proteins suggestive of
285 coregulation of the snoRNAs and these proteins. Two of them (CTACK, RARRES2) are also associated
286 with eGFR in our study. Several other proteins associated with these snoRNAs have been implicated
287 in kidney function previously. SnoRNAs U8 and SNORD118 both remained significant in the full
288 models. In addition to the snoRNAs described above various other snoRNAs are associated with eGFR

289 and UACR (13-31% of total number of significant sRNA). Almost all remained significant in the full
290 models.
291 At present there are no studies that have investigated the role of snoRNAs in diabetic DKD.
292 Recently, it was shown that four snoRNAs were differentially expressed in both plasma and urine
293 samples from 48 hypertensive persons with or without albuminuria [37]. SNORD116 and SNORD127,
294 two highly correlated snoRNAs (R^2=0.81), were also measured in our study and both were
295 significantly and negatively associated with UACR which is in contrast to this small previous study
296 which showed an upregulation in plasma but down-regulation in urine.
297
298 *Members of other sncRNA classes*
299 Although other classes of sncRNAs, such as Y_RNA derived fragments, tRNA fragments (tRF), lncRNA
300 fragments and snRNAs represent roughly twenty percent of all circulating sRNA detected we did not
301 find any significant associations with DKD. We did find a small number of associations with eGFR and
302 UACR though.
303
304 *Strengths and limitations*
305 Our study has several strengths. It is one of the largest circulating sncRNA profiling studies studying
306 the role of circulating sncRNAs in DKD. This improves power and reduces the chance of false
307 positives. In addition, it uses untargeted RNA sequencing allowing not only the study of miRNAs but
308 also several other classes of sncRNA showing that especially these other types of sncRNAs associated
309 with DKD and its components eGFR and UACR. The use of early, mainly mild cases of DKD facilitates
310 the identification of early markers of disease and may help to introduce early interventions aiming to
311 preserve kidney function in those at risk.
312 A limitation of our study is the use of a total RNA isolation procedure and as such we were not able
313 to investigate if the sncRNAs are carried by extracellular vesicles or are associated with carriers such
314 as HDL and Ago-2 [5]. The use of plasma for RNA isolation is another limitation. Although from a
315 biomarker perspective this is not a limitation, from a functional perspective the role of the sncRNAs
316 we identified needs to be established in the relevant cells and tissues. Another limitation of our study
317 is its cross-sectional, observational design. As such we cannot exclude that (part of) the observed
318 changes and associations are a reflection of the reduced kidney function in the participants.
319
320 In conclusion, in this study we have shown that several circulating sncRNAs are associated with DKD.
321 Interestingly, only a relatively small fraction belong to the well-studied microRNAs. This highlights the
322 potential of other classes of sncRNAs, such as snoRNAs and piRNAs, that have so far not been

323 investigated in relation to kidney function before. Future prospective and functional studies are
324 needed to establish the role of these circulating sncRNAs in subjects with diabetes and chronic
325 kidney disease.
326
327

Statements

Acknowledgement (optional)

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

Statement of Ethics

Study approval statement: The study has been approved by the ethics committee of the Amsterdam University Medical Center, location VUmc (approval number 2007/57; NL14692.029.07)

Consent to participate statement:. All participants provided informed consent prior to participating in this study.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

Funding Sources

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 proteomic analyses were funded by a grant from the Innovative Medicines Initiative 2 Joint Undertaking, under grant agreement no. 115881 (RHAPSODY). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA and is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI), under contract no. 16.0097. The funders had no role in the design, data collection, data analysis, and reporting of this study.

Author Contributions

LMtH, JWJB and RCS designed the study. JAdK and SJC isolated RNA and prepared the samples for RNAseq. MTB and JWJB were responsible for phenotypic data acquisition and preparation. LMtH, GAB and RCS prepared, analyzed and interpreted the data. LMtH and RCS wrote the first draft of the manuscript. RB contributed to the critical interpretation of the results. LMtH and RCS are the

guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors critically read and revised the manuscript and approved the final version of the manuscript.

Data Availability Statement

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

References

1. Persson F, Rossing P. Diagnosis of diabetic kidney disease: state of the art and future perspective. Kidney Int Suppl (2011). 2018 Jan;8(1):2-7.
2. Tofte N, Vogelzangs N, Mook-Kanamori D, Brahimaj A, Nano J, Ahmadizar F, et al. Plasma Metabolomics Identifies Markers of Impaired Renal Function: A Meta-analysis of 3089 Persons with Type 2 Diabetes. J Clin Endocrinol Metab. 2020 Jul 1;105(7).
3. Tofte N, Persson F, Rossing P. Omics research in diabetic kidney disease: new biomarker dimensions and new understandings? J Nephrol. 2020 Oct;33(5):931-48.
4. Barutta F, Bellini S, Canepa S, Durazzo M, Gruden G. Novel biomarkers of diabetic kidney disease: current status and potential clinical application. Acta Diabetol. 2021 Jul;58(7):819-30.
5. van Zonneveld AJ, Zhao Q, Rotmans JI, Bijkerk R. Circulating non-coding RNAs in chronic kidney disease and its complications. Nat Rev Nephrol. 2023 Sep;19(9):573-86.
6. Assmann TS, Recamonde-Mendoza M, de Souza BM, Bauer AC, Crispim D. MicroRNAs and diabetic kidney disease: Systematic review and bioinformatic analysis. Mol Cell Endocrinol. 2018 Dec 5;477:90-102.
7. Gholaminejad A, Abdul Tehrani H, Gholami Fesharaki M. Identification of candidate microRNA biomarkers in diabetic nephropathy: a meta-analysis of profiling studies. J Nephrol. 2018 Jul 17.
8. Loganathan TS, Sulaiman SA, Abdul Murad NA, Shah SA, Abdul Gafor AH, Jamal R, et al. Interactions Among Non-Coding RNAs in Diabetic Nephropathy. Front Pharmacol. 2020;11:191.
9. Freedman JE, Gerstein M, Mick E, Rozowsky J, Levy D, Kitchen R, et al. Diverse human extracellular RNAs are widely detected in human plasma. Nat Commun. 2016 Apr 26;7:11106.
10. van der Heijden AA, Rauh SP, Dekker JM, Beulens JW, Elders P, 't Hart LM, et al. The Hoorn Diabetes Care System (DCS) cohort. A prospective cohort of persons with type 2 diabetes treated in primary care in the Netherlands. BMJ Open. 2017 Jun 06;7(5):e015599.
11. KDIGO. KDIGO clinical practice guideline for the diagnosis, evaluation, prevention, and treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). Kidney Int Suppl. 2009 Aug(113):S1-130.
12. Rozowsky J, Kitchen RR, Park JJ, Galeev TR, Diao J, Warrell J, et al. exceRpt: A Comprehensive Analytic Platform for Extracellular RNA Profiling. Cell Syst. 2019 Apr 24;8(4):352-57.e3.
13. Griffiths-Jones S. The microRNA Registry. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D109-11.
14. Sai Lakshmi S, Agrawal S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. Nucleic Acids Res. 2008 Jan;36(Database issue):D173-7.
15. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012 Sep;22(9):1760-74.
16. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. Rna. 2014 Nov;20(11):1666-70.
17. Chan PP, Lowe TM. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. 2009 Jan;37(Database issue):D93-7.
18. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Research. 2012;41(D1):D991-D95.

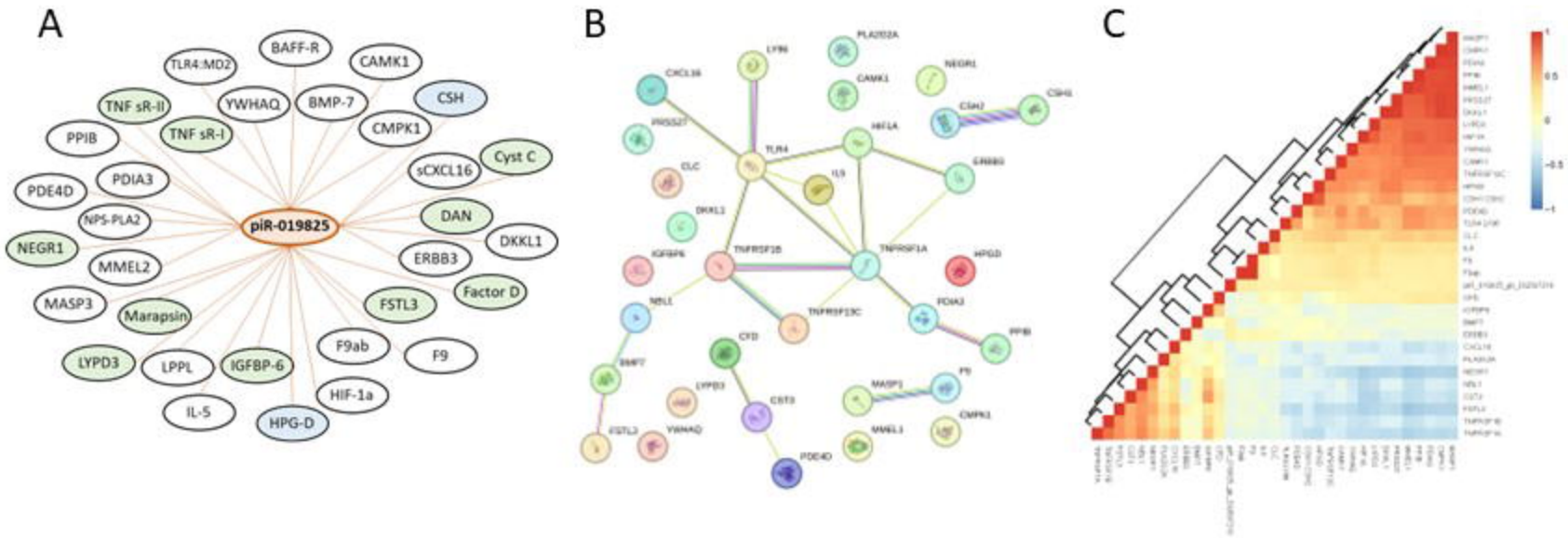
19. Slieker RC, Donnelly LA, Fitipaldi H, Bouland GA, Giordano GN, Åkerlund M, et al. Distinct Molecular Signatures of Clinical Clusters in People with Type 2 Diabetes: an IMIRHAPSODY Study. Diabetes. 2021 Aug 10;70(11):2683-93.
20. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139-40.
21. Taïbi F, Metzinger-Le Meuth V, M'Baya-Moutoula E, Djelouat M, Louvet L, Bugnicourt JM, et al. Possible involvement of microRNAs in vascular damage in experimental chronic kidney disease. Biochim Biophys Acta. 2014 Jan;1842(1):88-98.
22. Müller-Deile J, Gellrich F, Schenk H, Schroder P, Nyström J, Lorenzen J, et al. Overexpression of TGF-β Inducible microRNA-143 in Zebrafish Leads to Impairment of the Glomerular Filtration Barrier by Targeting Proteoglycans. Cell Physiol Biochem. 2016;40(5):819-30.
23. Bai S, Xiong X, Tang B, Ji T, Li X, Qu X, et al. Exosomal circ_DLGAP4 promotes diabetic kidney disease progression by sponging miR-143 and targeting ERBB3/NF-κB/MMP-2 axis. Cell Death Dis. 2020 Nov 23;11(11):1008.
24. Ge X, Xi L, Wang Q, Li H, Xia L, Cang Z, et al. Circular RNA Circ_0000064 promotes the proliferation and fibrosis of mesangial cells via miR-143 in diabetic nephropathy. Gene. 2020 Oct 20;758:144952.
25. Wu J, Liu J, Ding Y, Zhu M, Lu K, Zhou J, et al. MiR-455-3p suppresses renal fibrosis through repression of ROCK2 expression in diabetic nephropathy. Biochem Biophys Res Commun. 2018 Sep 5;503(2):977-83.
26. Zhu XJ, Gong Z, Li SJ, Jia HP, Li DL. Long non-coding RNA Hottip modulates high-glucose-induced inflammation and ECM accumulation through miR-455-3p/WNT2B in mouse mesangial cells. Int J Clin Exp Pathol. 2019;12(7):2435-45.
27. Rivoli L, Vliegenthart AD, de Potter CM, van Bragt JJ, Tzoumas N, Gallacher P, et al. The effect of renal dysfunction and haemodialysis on circulating liver specific miR-122. Br J Clin Pharmacol. 2017 Mar;83(3):584-92.
28. Trevisani F, Ghidini M, Larcher A, Lampis A, Lote H, Manunta P, et al. MicroRNA 193b-3p as a predictive biomarker of chronic kidney disease in patients undergoing radical nephrectomy for renal cell carcinoma. Br J Cancer. 2016 Nov 22;115(11):1343-50.
29. Mollet IG, Macedo MP. Pre-Diabetes-Linked miRNA miR-193b-3p Targets PPARGC1A, Disrupts Metabolic Gene Expression Profile and Increases Lipid Accumulation in Hepatocytes: Relevance for MAFLD. Int J Mol Sci. 2023 Feb 15;24(4).
30. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet. 2019 Feb;20(2):89-108.
31. El-Mogy M, Lam B, Haj-Ahmad TA, McGowan S, Yu D, Nosal L, et al. Diversity and signature of small RNA in different bodily fluids using next generation sequencing. BMC Genomics. 2018 May 29;19(1):408.
32. Gohda T, Niewczas MA, Ficociello LH, Walker WH, Skupien J, Rosetti F, et al. Circulating TNF receptors 1 and 2 predict stage 3 CKD in type 1 diabetes. J Am Soc Nephrol. 2012 Mar;23(3):516-24.
33. Jalal D, Sanford B, Renner B, Ten Eyck P, Laskowski J, Cooper J, et al. Detection of pro angiogenic and inflammatory biomarkers in patients with CKD. Sci Rep. 2021 Apr 22;11(1):8786.
34. Kobayashi H, Looker HC, Satake E, D'Addio F, Wilson JM, Saulnier PJ, et al. Neuroblastoma suppressor of tumorigenicity 1 is a circulating protein associated with progression to end-stage kidney disease in diabetes. Sci Transl Med. 2022 Aug 10;14(657):eabj2109.
35. Deogharia M, Majumder M. Guide snoRNAs: Drivers or Passengers in Human Disease? Biology (Basel). 2018 Dec 20;8(1).

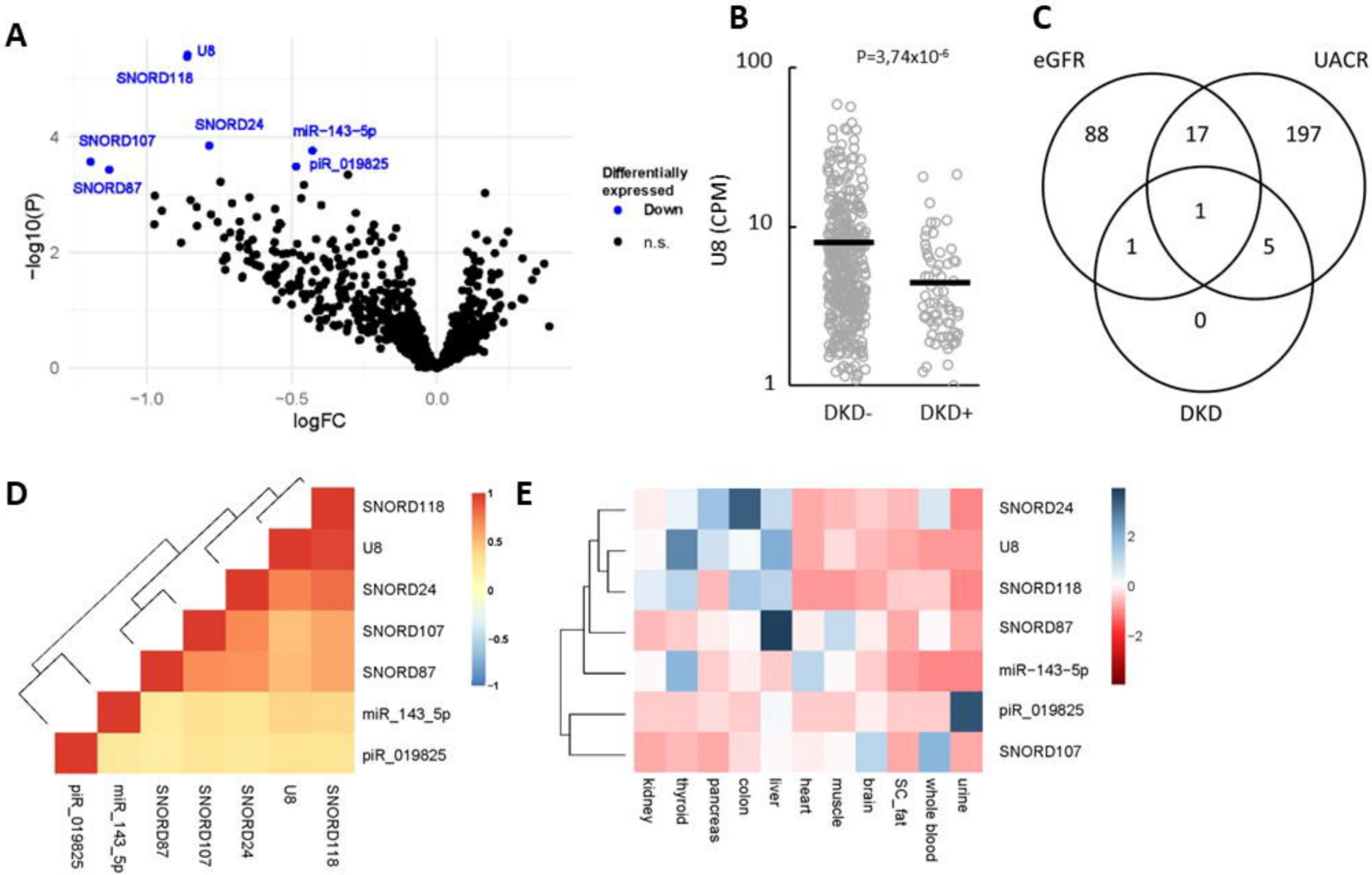
36. Bergeron D, Fafard-Couture É, Scott MS. Small nucleolar RNAs: continuing identification of novel members and increasing diversity of their molecular mechanisms of action. Biochem Soc Trans. 2020 Apr 29;48(2):645-56.
37. Riffo-Campos AL, Perez-Hernandez J, Ortega A, Martinez-Arroyo O, Flores-Chova A, Redon J, et al. Exosomal and Plasma Non-Coding RNA Signature Associated with Urinary Albumin Excretion in Hypertension. Int J Mol Sci. 2022 Jan 13;23(2).

Figure Legends

Fig. 1. Panel A. Volcano plot showing the association between sncRNAs and DKD in the base model. sncRNAs in blue are significantly lower is cases with DKD after multiple hypothesis testing (FDR≤0.05). logFC is the Log2 of the fold change between cases and controls. Panel B. SNORD118 levels in those with or without DKD. CPM = counts per million reads. Panel C. Overlap between sncRNAs significantly associated with DKD, eGFR or UACR in the base model. Panel D. Pairwise correlations for the seven sncRNAs associated with DKD. Colors represent Spearman's rho. Panel E. Expression pattern for sncRNAs associated with DKD in publicly available datasets for 11 tissues of interest. Colors represent the Z-scaled expression of a sncRNA where blue indicates relative high expression and red relative low expression.

Fig. 2. Panel A. Overview of the 32 proteins that associate with piR-019825 expression levels in plasma from persons with T2D (P≤1x10-3). Proteins in green are also significantly associated with eGFR and those in blue are associated with UACR (PFDR≤0.05). Panel B. Protein-protein interaction network from the proteins correlated with piR-019825 levels as predicted by String-db.org (accessed Nov-2023). Panel C. Pairwise correlations between the 32 piR-019825 associated proteins.


A
piR-019825
BAFF-R
CAMK1
TLR4:MD2
YWHAQ
BMP-7
CSH
TNF sR-II
TNF sR-I
CMPK1
PPIB
Cyst C
sCXCL16
PDIA3
PDE4D
DAN
NPS-PLA2
DKKL1
NEGR1
ERBB3
MMEL2
Factor D
MASP3
FSTL3
Marapsin
F9ab
F9
LYPD3
LPPL
IGFBP-6
HIF-1a
IL-5
HPG-D
B
C


A
U8
SNORD118
SNORD24
miR−143−5p
SNORD107
piR_019825
SNORD87
−log10(P)
logFC
Differentially expressed
Down
n.s.
B
P=3,74x10-6
U8 (CPM)
DKD-
DKD+
C
eGFR
UACR
DKD
88
17
197
1
1
5
0
D
SNORD118
U8
SNORD24
SNORD107
SNORD87
miR_143_5p
piR_019825
E
SNORD24
U8
SNORD118
SNORD87
miR−143−5p
piR_019825
SNORD107
kidney
thyroid
pancreas
colon
liver
heart
muscle
brain
SC_fat
whole blood
urine