

# SCIENTIFIC DATA

## OPEN Data Descriptor: Small non-coding RNA expression in mouse nephrogenic mesenchymal progenitors

Received: 29 March 2018

Accepted: 15 August 2018

Published: 13 November 2018

Yu Leng Phua<sup>1,2,\*</sup>, Andrew Clugston<sup>1,2,3,\*</sup>, Kevin Hong Chen<sup>1,2</sup>, Dennis Kostka<sup>1,3,4</sup> & Jacqueline Ho<sup>1,2</sup>

MicroRNAs (miRNAs) are small non-coding RNAs that are essential for the regulation of gene expression and play critical roles in human health and disease. Here we present comprehensive miRNA profiling data for mouse nephrogenic mesenchymal progenitors, a population of cells enriched for nephron progenitors that give rise to most cell-types of the nephron, the functional unit of the kidney. We describe a miRNA expression in nephrogenic mesenchymal progenitors, with 162 miRNAs differentially expressed in progenitors when compared to whole kidney. We also annotated 49 novel miRNAs in the developing kidney and experimentally validated 4 of them. Our data are available as a public resource, so that it can be integrated into future studies and analyzed in the context of other functional and epigenomic data in kidney development. Specifically, it will be useful in the effort to shed light on molecular mechanisms underlying processes essential for normal kidney development, like nephron progenitor specification, self-renewal and differentiation.

<b>Design Type(s)</b>	replicate design • microRNA profiling by array design • epigenetic modification identification objective • sequence analysis objective
<b>Measurement Type(s)</b>	microRNA profiling assay
<b>Technology Type(s)</b>	microRNA profiling by high throughput sequencing
<b>Factor Type(s)</b>	Replicate
<b>Sample Characteristic(s)</b>	Mus musculus • nephron progenitor • kidney

<sup>1</sup>Rangos Research Center, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA, USA. <sup>2</sup>Department of Pediatrics, Division of Nephrology, University of Pittsburgh School of Medicine, PA, USA. <sup>3</sup>Department of Developmental Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. <sup>4</sup>Department of Computational & Systems Biology and Pittsburgh Center for Evolutionary Biology and Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to D.K. (email: kostka@pitt.edu) or J.H. (email: jacqueline.ho2@chp.edu)

## Background & Summary

MicroRNAs (miRNAs) are small, endogenously synthesized non-coding RNAs (~22 nucleotides long) that are critical in a wide variety of biological processes, where they primarily act to fine-tune gene expression at the post-transcriptional and translational level<sup>1</sup>. Biosynthesis of miRNAs begins in the nucleus, where a primary miRNA (pri-miRNA) transcript is first transcribed by RNA polymerase II<sup>1,2</sup>. Subsequently, the endoribonuclease enzyme Drosha-Dgcr8 complex processes pri-miRNAs into individual hairpin-shaped stem loop precursor miRNAs (pre-miRNAs), which are then exported into the cytoplasm via Exportin5<sup>3–5</sup>. In the final steps of miRNA biogenesis, Dicer processes pre-miRNA into mature miRNA, with the miRNA loaded into the RNA-induced silencing complex (RISC)<sup>6–8</sup>. mRNA 3'UTR target recognition by miRNA is primarily dependent on the miRNA seed sequence located within nucleotide position 2–8 of the miRNA, and complementary Watson-Crick base pairing between the miRNA and mRNA allows the RISC complex to catalyse the process of mRNA degradation and/or translational inhibition<sup>9,10</sup>.

Nephron progenitors are multipotent cells that undergo a mesenchymal to epithelial transition to subsequently differentiate into glomerular podocytes, proximal tubules, loops of Henle and distal tubule in the developing kidney<sup>11</sup>. They also self-renew throughout kidney development, and the number of nephron progenitors is one of the primary determinants of nephron endowment at birth<sup>12</sup>. Since no new nephrons are formed postnatally in humans, low nephron number increases an individual's risk to develop chronic kidney disease in adulthood<sup>13</sup>. While the role of many protein-coding genes in nephron progenitor self-renewal, maintenance and differentiation is well established, the precise role of small non-coding RNAs remains ill-defined. There are several lines of evidence that point to the importance of miRNAs in kidney development and disease. Mutations in *DROSHA* and *DICER1* have been identified and implicated in several human kidney disorders, including paediatric Wilms Tumour and multilocular cystic renal tumours<sup>14</sup>. Conditional ablation of Dicer-dependent miRNAs in mouse nephron progenitor or ureteric bud lineages during renal development resulted in severe renal hypodysplasia (small kidneys) and collecting duct cyst formation, due to aberrant progenitor apoptosis and attenuated cilium length respectively<sup>15,16</sup>. Also, conditional ablation of Dicer or Drosha in glomerular podocytes and renal stroma results in early postnatal death and a wide variety of renal anomalies including podocyte foot process effacement, marked proteinuria, and glomerular aneurysms<sup>17–21</sup>.

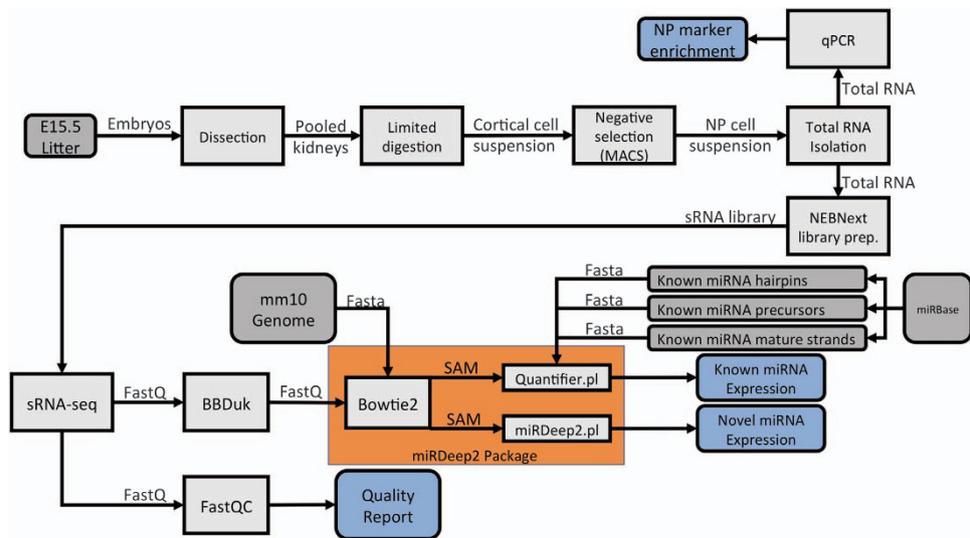
However, more precise characterisation of the role of small RNAs in nephron progenitors during kidney development has been hampered by the lack of comprehensive miRNA expression datasets in this context. Given the biomedical relevance of this system, we have performed high throughput small RNA Sequencing (sRNA-Seq) in three biological replicates of embryonic day 15.5 (E15.5) nephrogenic mesenchymal cells enriched for nephron progenitors and whole kidney (Fig. 1). Using an adjusted p-value cutoff of 0.05, we identified a total of 162 miRNAs (5p and 3p strand inclusive) out of 792 detectable miRNAs to be differentially expressed in this population when compared to whole kidney. Among the top differentially expressed miRNAs are members of the epithelial-specific *miR-200* family, consisting of *miRs-200a*, *200b*, *200c*, *141* and *429*<sup>22</sup>. Levels of *miR-200* family miRNAs<sup>22</sup> were significantly lower in nephron progenitors, as might be predicted given that nephron progenitors undergo a mesenchymal to epithelial transition upon differentiation. Furthermore, we uncovered 49 novel miRNA species expressed in the developing kidney. Of these miRNAs, 4 were validated via quantitative real-time PCR (qPCR), with 3 via section *in situ* hybridization. In general, the data resource will be useful for researchers studying miRNA related biology in kidney/nephron development.

## Methods

### Nephron progenitor isolation and total RNA preparation

Nephrogenic mesenchymal cells enriched for nephron progenitors and whole kidney samples were isolated from 3 litters of E15.5 CD1 mouse embryos (Charles River Laboratories) in accordance to a published protocol using a negative selection approach<sup>23</sup>. Briefly, intact embryonic kidneys were subjected to limited digestion, followed by incubation with a cocktail of monoclonal biotinylated antibodies (eBioscience; CD140a #13-1401-82, CD105 #13-1051-82, Epcam #13-5791-82 and Ter119 #13-5921-82), and magnetic activated cell sorted using Dynabeads MyOne Streptavidin C1 magnetic beads (Thermo; #65001) to deplete unwanted cell types (Fig. 1). To minimise undesired gene expression changes, total RNA from nephron progenitors were immediately processed in QIAzol Lysis Reagent (Qiagen; #79306) and purified using a miRNeasy Micro Kit (Qiagen; #217084). For whole kidney samples, remnant kidneys left from the limited digestion step were homogenised in QIAzol Lysis Reagent, clarified by centrifugation at 13,000 rpm for 5 min, and subsequently purified using a miRNeasy Mini Kit (Qiagen; #217004). Purified total RNA samples were stored at -80 °C until further processing, and freeze thawing of samples was limited to no more than 2 cycles. Quantitative PCR (qPCR) was carried out using standard SYBR Green detection on a BioRad CFX96 Real Time PCR Instrument to determine enrichment of nephron progenitors from the isolation. Primers used include *Six2*, *Cited1* (nephron progenitors), *Pdgfrβ* (renal stroma), *VE-Cad* (endothelial), *Calb*, *Epcam* (epithelial tubules) and *Synpo* (podocytes) (Table 1). Statistical analysis was performed using an unpaired Student's t-test, and genes with a p-value of < 0.05 were considered statistically significant.

To account for biological heterogeneity, a total of 3 biological replicates were collected from 3 litters of wildtype CD1 embryos. All experimental procedures were performed in accordance with the University



**Figure 1.** Schematic pipeline illustrating the workflow of nephron progenitor isolation and bioinformatics analysis of the small RNA-Seq dataset. For nephron progenitor isolation, embryonic day 15.5 (E15.5) CD1 mouse kidneys were subjected to limited digestion, followed by negative cell selection through Magnetic Activated Cell Sorting (MACS). Total RNA from the isolate was extracted and subjected to qPCR analysis for enrichment of nephron progenitor markers. Following verification of nephron progenitor enrichment, total RNA was used as an input for the NEBNextMultiplex Small RNA Library Prep to generate libraries for the sRNA-Seq. The sRNA-Seq dataset was analysed in accordance with the pipeline, with the fastq files first analysed by FastQC to determine the quality of the sequencing reads, followed by adaptor removal using the *BBDuk* package, and finally aligned, quantified and annotated to the *mus musculus* mm10 genome using the miRDeep2 package.

of Pittsburgh Institutional Animal Care and Use Committee guidelines, which adheres to the NIH Guide for the Care and Use of Laboratory Animals. sRNA-Seq experimental workflow and quality control standards were carried out in accordance with ENCODE guidelines (<https://www.encodeproject.org/rna-seq/small-rnas/>).

### Small RNA library preparation and sequencing

Because miRNAs can be stably bound to mRNAs<sup>24</sup>, total RNA instead of size-selected purified small RNA was used as the initial input for cDNA library synthesis. The NEBNextMultiplex Small RNA Library Prep Set for Illumina (NEB; #E7300S) was used to synthesise barcoded cDNA libraries from 100 ng total RNA in accordance with the manufacturer's protocol and size selection performed using the Agencourt AMPure XP beads (Beckman Coulter; #A63881). The purified cDNA libraries were then pooled, normalised and multiplex sequenced at 1 X 50 bp with the use of a NextSeq 500/550 Mid Output v2 Kit (150 cycles) (Illumina; FC-404-2001) on an Illumina NextSeq 550 System at the Rangos Research Center, yielding approximately 18 million reads per sample.

### Small RNA-Seq data analysis and quality control

Adapter trimming was performed using the Bbmap software (<https://sourceforge.net/projects/bbmap/>) and reads were aligned to the mouse genome (mm10) using Bowtie2<sup>25</sup>. miRNA specific analysis was performed with miRDeep2 software<sup>26</sup>. Expression of known miRNAs (obtained from miRBase v21<sup>27</sup>) was quantified, and miRNAs with more than 10 counts across all sample conditions were subsequently analysed using DESeq2<sup>28</sup> to identify differential expression between nephron progenitor and whole kidney samples. miRNAs with an Benjamini-Hochberg adjusted p-value of 5% or less are reported. Novel miRNAs were annotated with the standard miRDeep2 parameters, and read alignment data was visualized with the Integrative Genomics Viewer software available from the Broad Institute<sup>29–32</sup>.

### Data Records

The sRNA-Seq fastq files were deposited at NCBI Sequence Read Archive (Data Citation 1). This accession contains all 6 fastq files from the sRNA-Seq. The raw fastq files were subsequently processed and deposited into NCBI Gene Expression Omnibus (Data Citation 2).

Gene	Forward (5'-3')	Reverse (5'-3')
Gapdh	AGGTCGGTGTGAACGGATTG	TGTAGACCATGTAGTTGAGGTC
Six2	TCTGCTCGGTATCCITTTGGG	TTAAAAATCGGGTGGTGGTG
Cited1	AACCTTGGAGTGAAGGATCGC	GTAGGAGAGCCTATTGGAGATGT
Ve-Cad	CACTGCTTTGGGAGCCTTC	GGGGCAGCGATTCAITTTTCT
Pdgrf $\beta$	TTCCAGGAGTGATACCAGCTT	AGGGGGCGTGTACTAGG
Calb	ATGATCAGGATGGCAACGGA	GTTCCGTACAGCTCCCTCC
Epcam	GCGGCTCAGAGAGACTGTG	CCAAGCATTAGAGCCAGTTT
Synpo	CCTGCCGTAACITTCGTG	GAGCGGGGTAGGAAAAG

**Table 1. Primer sequences used for quantitative PCR.**

miRNA	Mature forward sequence (5'-3')	Star forward sequence (5'-3')
U6	GTGCTCGCTTCGGCAGC	NA
chr1_100950463-100950527	GCGGAAAGCTGAAACTTAGAGG	TGAGTCTCCAGCCCTACCCTAA
chr8_66396466-66396533	GCGGTTCTTAGTTGGTGGAGC	GCGCTGATTGACTGAGAATGT
chr11_108839740-108839780	GCGTGAGAAGAACTTTGAAGAGC	GCGTCTCAGAGITAGTGAGA
chr13_23436971-23437058	GTCTAGGGGTATGATTCTCGCAA	GCGTTTTATCGTITTCCTGT

**Table 2. Primer sequences used for novel miRNA quantitative PCR.**

### miRNA Quantitative PCR and Locked Nucleic Acid *in situ* hybridization validation

For the validation of differentially expressed miRNAs, cDNA synthesis was carried out with the use of TaqMan Advanced miRNA cDNA Synthesis Kit (Thermo; #A28007) in accordance to the manufacturer's protocol using equivalent amounts of total RNA. qPCR was carried out with the use of TaqMan Advanced miRNA Assay probes (Thermo; #A25576; 477952\_miR, 477970\_miR, 477885\_miR, 478008\_miR) and TaqMan Universal Master Mix II, no UNG (Thermo; #4440040) on a BioRad CFX96 Real Time PCR Instrument using FAM detection.

For the validation of novel miRNAs, cDNA synthesis from total RNA was performed with the use of NCode VILO miRNA cDNA Synthesis Kit (Thermo; #A11193050) in accordance to the manufacturer's protocol using equivalent amounts of total RNA, and qPCR was performed using standard SYBR Green detection on a BioRad CFX96 Real Time PCR Instrument. The universal reverse primer is supplied with the NCode VILO miRNA cDNA Synthesis Kit, and the forward primer sequence is listed in Table 2. Locked Nucleic Acid miRNA *in situ* hybridisation on E15.5 embryonic kidney cryosections was performed as previously described with the use of custom-designed LNA detection probes<sup>19</sup> (Table 3) (Exiqon).

### Code Availability

Software and settings used in the processing of this data are listed as follows: 1) BBDuk: from the BBDuk version 35.59 package, adapter references for Illumina indices and the Nextera smRNA library prep set, quality trimming right to left (qtrim = 30), kmer trimming right to left (ktrim = r) then left to right (ktrim = l), minimum kmers of 11 (mink = 11), hamming distance of 1 bp (hdist = 1), minimum final read length of 18 bp (minlen = 18). 2) Bowtie 2: version 2.2.3, parameters of -threads 8, -q, -U, -S, --sensitive-local. 3) miRDeep2 version 2.0.0.8, using the bwa\_sam\_converter.pl script to convert Bowtie2's SAM output formats, and the miRDeep2.pl script using default parameters for the mouse species (-t Mouse) and miRbase v18 identifiers (-P). Reference data included a pre-built Bowtie2 index for the mm10 genome and miRNA mature and hairpin sequences available in Illumina's mm10 iGenomes package. A FASTA with miRNA "other" sequences were provided from miRbase version 21. 4) R: version 3.4.3, data analysis was performed using the R programming language, including packages mirbase.db, DESeq2, and GenomicRanges.

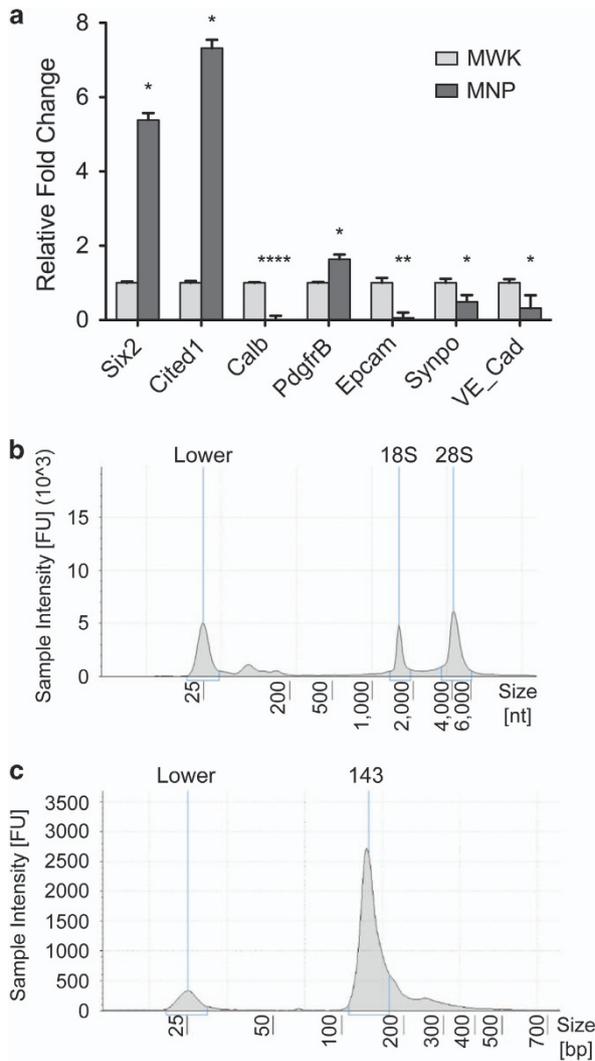
### Technical Validation

#### Cellular enrichment and total RNA quality control

Our overall goal was to quantify small RNA expression in mouse nephron progenitors and whole kidney. We used a previously published protocol<sup>23</sup> to obtain an enriched fraction of nephron progenitor cells from whole embryonic kidney samples. Confirming that this approach was successful, qPCR analysis of the nephron progenitor-enriched isolate showed high enrichment of nephron progenitor transcripts (*Six2*, *Cited1*) in comparison to whole kidney samples, with evidence of minimal cellular contamination from other key renal cell lineages including the endothelium (*VE-Cad*), epithelial tubules (*Calb*, *Epcam*) and podocytes (*Synpo*) (Fig. 2a). An inherent limitation of this protocol<sup>23</sup> is minor contamination from

LNA Probe	Product No.	Product sequence 5'-3'
Scramble-miR	699003-300	Scrambled sequence
miR-125b-5p	611756-300	CACAAGTTAGGGTCTCAGGGA
miR-615-3p	615721-300	AAGAGGGAGACCCAGGCTC
miR-10a-5p	YD00612528	ACAAATTCGGATCTACAGGGA
chr1_100950463- 100950527	Custom	CCTCTAAGTTTCAGCTTCC
chr9_113756831- 113756880	Custom	GTCTGTATTGTTATTTT
chr11_108839740- 108839780	Custom	GCTCTCAAAGTCTCTCA

**Table 3.** Exiqon miRCURY LNA detection probes used for miRNA section *in situ* hybridisation.



**Figure 2.** Quality control of the input samples demonstrated an enrichment of nephron progenitors in the cellular isolate with good quality RNA, and successful cDNA library synthesis. (a) qPCR analysis showed that the cells in the isolate are enriched for nephron progenitors (*Six2*, *Cited1*) with minimal cellular contamination from ureteric bud (*Calb*), podocytes (*Sympo*), endothelial (*VE-Cad*) and epithelial cells (*Epcam*), but minor contamination from the renal stroma lineage (*Pdgfrβ*). (b) TapeStation analysis of the total RNA showed good quality RNA traces with distinct 18S and 28S rRNA peaks with retention of small RNAs as evident by the broad peak to the right of the lower molecular marker. (c) TapeStation analysis showed that the purified cDNA libraries exhibited a distinct peak at ~140–150 nucleotides (adaptor ligated small RNA products), indicating that both the cDNA library construction and cleanup are successful. N = 3, \*p-value < 0.05, \*\*p-value < 0.01, \*\*\*\*p-value < 0.0001.

Sample	RIN	Fastq file name
MNP1	10	MNP1.fastq
MNP2	10	MNP2.fastq
MNP3	10	MNP3.fastq
MWK1	9.9	MWK1.fastq
MWK2	9.8	MWK2.fastq
MWK3	10	MWK3.fastq

**Table 4. Sample IDs and RIN score of RNA samples used for small RNA-Sequencing.**

the renal stroma (*Pdgfrβ*), which was indeed detected in our isolate. Although the Six2-TGC transgenic mouse<sup>33</sup> provides the opportunity for isolating a more pure fraction of nephron progenitor cells, these mice are known to have renal hypoplasia<sup>34</sup>, suggesting that their nephron progenitors may not be entirely normal. Taken together, our small RNA sequencing dataset represents a highly-enriched fraction of wild-type nephron progenitors, with minor contamination from renal stroma.

Following cellular isolation and total RNA extraction, the quality of the total RNA was assessed on an Agilent 4200 TapeStation System (Agilent; #G2991AA) using a High Sensitivity RNA ScreenTape (Agilent; #5067-5579) to determine the RNA Integrity Number (RIN) score, which is generated based on the electrophoretic profile of 18S and 28S ribosomal RNA (rRNA)<sup>35</sup> (Fig. 2b). All RNA samples used for the sRNA-Seq had a RIN score of above 9 (Table 4) and were therefore considered to be good quality intact RNA with minimal degradation. Retention of small RNA in the total RNA isolation was also evident by a distinct broad peak to the right of the lower molecular marker (Fig. 2b). To ensure equivalent amounts of total RNA input for the cDNA library construction, the Qubit RNA BR Assay Kit (Thermo; #Q10210) was used for RNA quantitation.

#### Quality control of the cDNA libraries

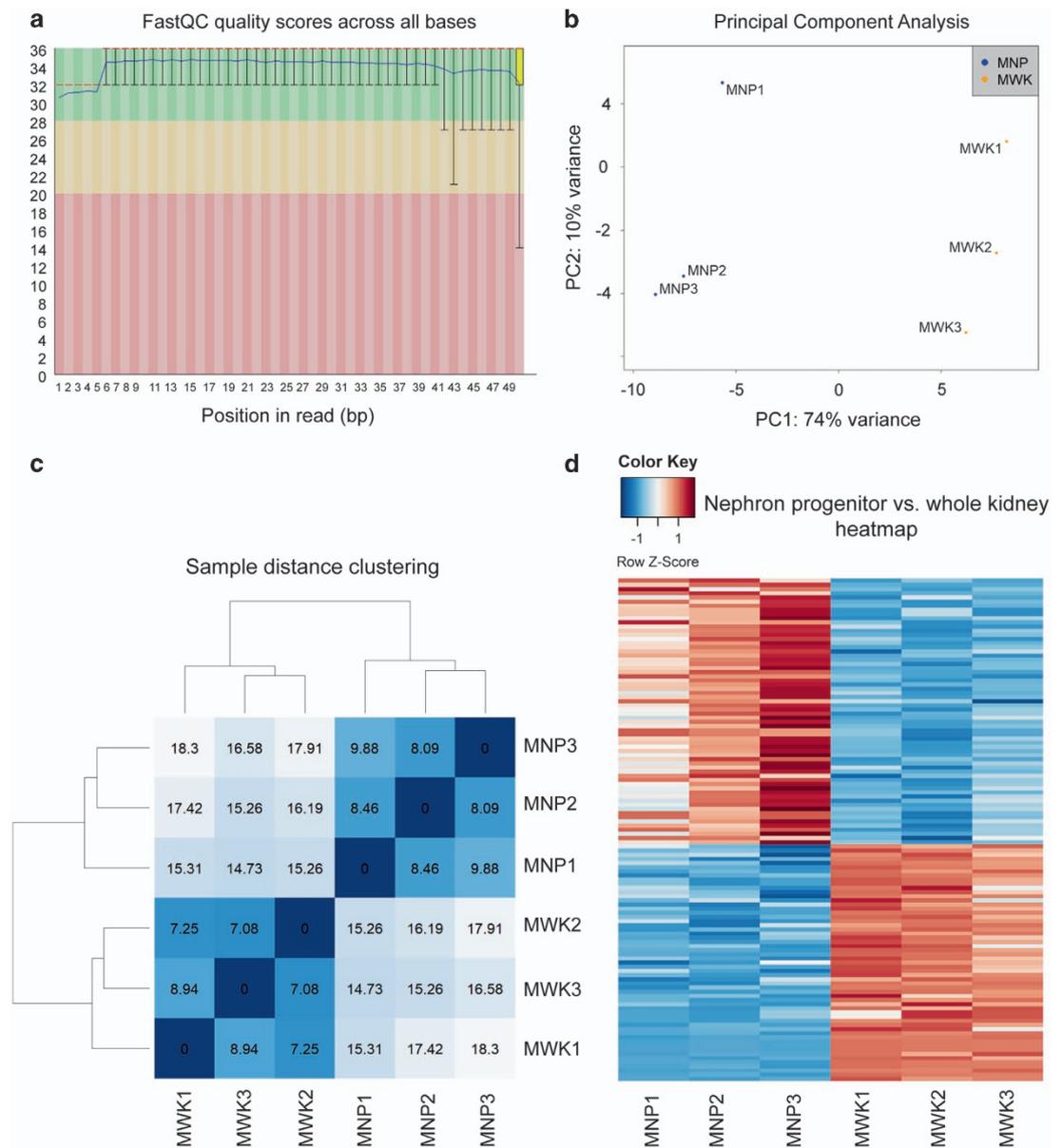
cDNA libraries were assayed on an Agilent 4200 TapeStation System (Agilent; #G2991AA) using a High Sensitivity D1000 ScreenTape (Agilent; #5067-5584) and were verified to exhibit a distinct peak at ~140-150 nucleotides that corresponds to adapter-ligated constructs derived from small RNAs (Fig. 2c). Moreover, TapeStation analysis verified that the final cDNA libraries for sequencing were depleted of large molecular weight products that one would expect to be indicative of larger RNA molecules like messenger or ribosomal RNA.

#### Quality control of sRNA-Seq data

Purified cDNA libraries for all samples were normalised and multiplex sequenced using a single flow cell on the Illumina NextSeq550 System to minimise technical variability from the sequencing. Following sequencing, reads for each sample was evaluated using the FastQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and all sequenced libraries had an average quality score of above 30, indicating high quality base of the sequenced product (Fig. 3a). Percentage of aligned reads was consistently above 70%, with 13 million reads mapped per sample on average. For miRDeep2 analysis, aligned reads shorter than 18 bp were discarded in keeping with the software's requirements<sup>26</sup>. Principal component analysis (PCA) revealed a clear distinction between and separation of the nephron progenitor (MNP) and whole kidney (MWK) samples (Fig. 3b), as did hierarchical clustering of samples based on their miRNA abundances (Fig. 3c). PCA and hierarchical clustering analysis both indicate that samples enriched for nephron progenitors are transcriptionally distinct from whole kidney samples.

We identified a total of 162 miRNAs (5p and 3p strand inclusive) out of 792 detectable miRNAs to be statistically differentially expressed in nephrogenic mesenchymal cells enriched for nephron progenitors relative to whole kidney (Fig. 3d and Differentially expressed miRNAs in nephron progenitors & miRNAs expressed in kidney, Data Citation 3), and validated the differential profile of *miR-210*, *miR-125b* and *miR-30c* by qPCR (Fig. 4a). Among the top differentially expressed miRNAs are members of the epithelial-specific *miR-200* family, consisting of *miRs-200a*, *200b*, *200c*, *141* and *429*<sup>22</sup>. Considering that nephron progenitors are mesenchymal in nature and that they undergo a mesenchymal to epithelial transition upon differentiation<sup>33</sup>, the under-expression of the *miR-200* cluster in nephron progenitors not only serves as an excellent internal validation of the phenotypic characteristics of these cells, but also supports the overall integrity of the miRNA profiling dataset.

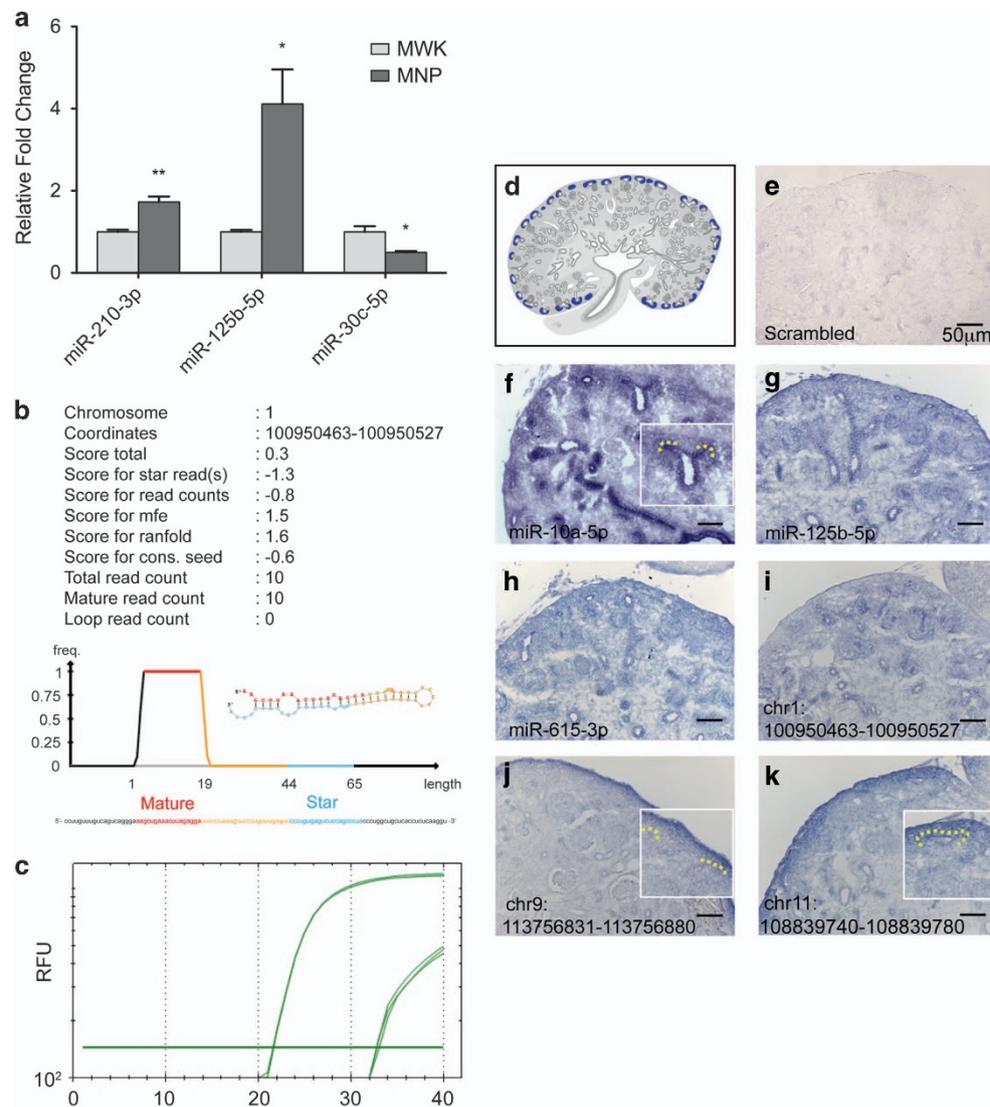
We explored the use of alternative bioinformatics approaches and obtained comparable results with choices such as *Rsubread* for read mapping to the mm10 genome (parameters: maximum allowed for 3 nucleotides mismatch and 1 insertion-deletion)<sup>36</sup>, *featureCounts* which outputs number of reads assigned to genomic features<sup>37</sup> and *limma-voom* for differential expression statistical analysis<sup>38</sup>.



**Figure 3.** Quality control of the small RNA-Seq dataset showed good sequencing quality reads and congruence of the biological samples. (a) FastQC analysis of a sequenced library (MNP1) showing that each sequenced base had a mean value score of >30, indicating good quality sequencing. (b) Principal Component Analysis of the sRNA-Seq dataset showed clear separation of the nephron progenitor (MNP) and whole kidney (MWK) samples. (c) Hierarchical clustering correctly grouped samples based on their tissue of origin. (d) Heatmap representation of differentially expressed miRNAs in nephron progenitors vs. whole kidney samples with a false discovery rate of 0.05.

### Locked Nucleic Acid section in situ hybridisation and qPCR validation of the small RNA-Seq data

The additional novelty of this dataset stems from its inherent ability to identify novel unannotated miRNA transcripts without *a priori* sequence knowledge. A major challenge in novel miRNA identification is that presumed novel miRNA transcripts are in fact byproducts resulting from degraded mRNA transcripts, and hence not necessarily representative as *bona fide* novel miRNAs. To overcome this, the miRDeep2 package<sup>26</sup> was used because it includes a series of stringent core analysis modules including the RNAfold tool (<http://rna.tbi.univie.ac.at/>) and the miRDeep2 core algorithm. The RNAfold tool determines and predicts whether the novel miRNA sequence would exhibit a typical hairpin secondary structure, and the miRDeep2 core algorithm evaluates both the miRNA's secondary structure and the expected read mapping signature of its hairpin precursor. The core algorithm also checks that the

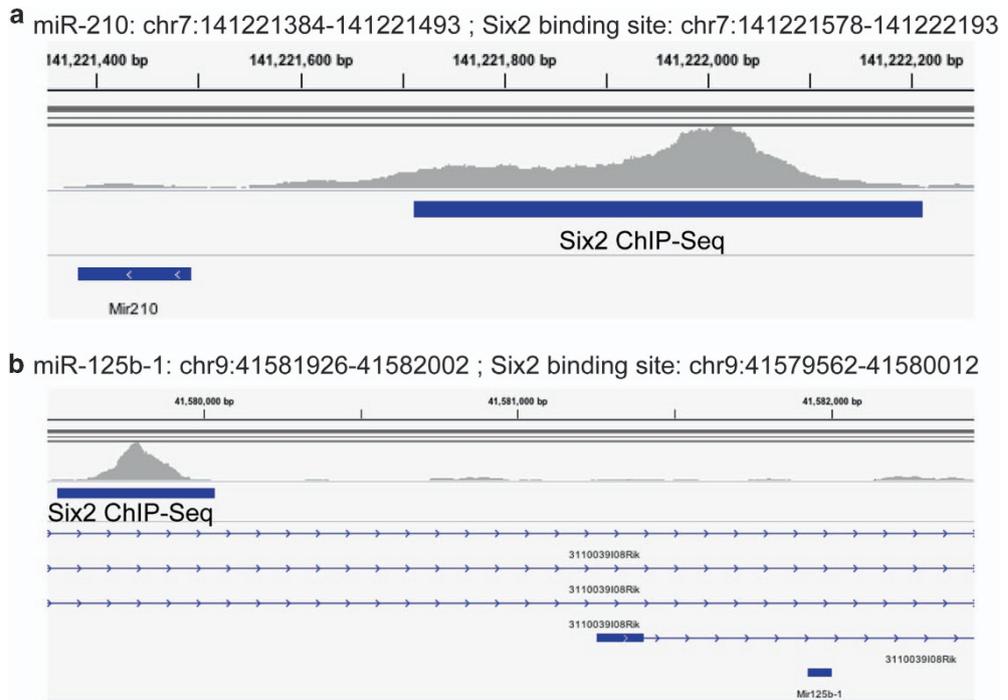


**Figure 4.** Small RNA-Seq and miRDeep2 analysis revealed novel unannotated miRNAs expressed in the developing kidney. (a) qPCR analysis validated the differential expression profile of *miRs-210*, *125b* and *30c* in nephron progenitor when compared to whole kidney.  $N = 3$ , \* $p$ -value  $< 0.05$ , \*\* $p$ -value  $< 0.01$ . (b) miRDeep2 output for chr1\_100950463-100950527, a novel miRNA discovered in our sRNA-Seq data; high frequency (freq) of reads mapped to the mature region of the predicted pre-miRNA structure. (c) qPCR validation of chr1\_100950463-100950527 showed a clear distinction between and enrichment of mature (low  $C_T$ ) over star strand (High  $C_T$ ) transcripts. RFU: Relative Fluorescence Unit. (d) For spatial orientation purposes, schematic representative image of nephron progenitors is highlighted in blue. (e-k) Locked Nucleic Acid section *in situ* hybridisation was used to validate the spatial expression pattern of miRNAs in the developing kidney. *miRs-125b*, *615* and chr1\_100950463-100950527 exhibited a ubiquitous expression pattern in the developing kidney. *miR-10a*, chr9\_113756831-113756880 and chr11\_108839740-108839780 appear spatially enriched in the nephrogenic zone, comprising nephron progenitors, renal stroma and the ureteric bud. Schematic image in (d) was designed by Kylie Georgas (University of Queensland) and are publicly available for use from the GUDMAP database (<http://www.gudmap.org/Schematics>).

sequencing reads which span the predicted hairpin structure contains specific recognition sites for subsequent processing by Dicer into mature miRNAs. Novel miRNAs were considered high confidence when both mature and star strands were detected in at least 2 independent samples, in conjunction with sequencing reads spanning across the putative chromosome coordinates independently verified on IGV Viewer. In this manner, we identified a total of 49 novel miRNAs, exemplified by the 20 nucleotides long novel miRNA (GGA AAG CTG AAA CTT AGA GG) depicted with a characteristic hairpin precursor

miRNA	Mature $C_T$	Star $C_T$
chr1_100950463-100950527	20.25	29.01
chr8_66396466-66396533	7.04	37.65
chr11_108839740-108839780	26.59	34.48
chr13_23436971-23437058	20.60	30.65

**Table 5.** Quantitative PCR profile of novel miRNA assayed.



**Figure 5.** Integrative Genome Viewer visualisation of Six2 binding upstream of expressed miRNAs in nephron progenitors. By combining a publicly available Six2 ChIP-Seq dataset (SRA Study: SRP064623; GEO: GSE73867) with the current sRNA-Seq results, evidence for Six2 binding was observed in the promoter region of *miRs-210* and *125b*, miRNAs that are enriched in nephron progenitors.

secondary structure located within chr1: 100950463-100950527 (Fig. 4b and Novel miRNAs expressed in kidney, Data Citation 3). These novel miRNAs will be submitted for names through miRBase in accordance to their criteria.

As a confirmation that the novel miRNAs were not an artefact from sRNA sequencing, qPCR was performed on 4 candidate novel miRNAs. The qPCR analysis depicted distinct detection of mature (low  $C_T$ ) over star strand (high  $C_T$ ) transcripts (Fig. 4c and Table 5), thereby validating the overall approach in novel miRNA discovery through sRNA-Seq and miRDeep2 identification. Using miRNA Locked Nucleic Acid *in situ* hybridization for spatial localisation of miRNA expression sites, both annotated (Fig. 4f–h) (*miRs-10a*, *125b*, *615*) and novel miRNAs (Fig. 4i–k) (chr1\_100950463-100950527, chr9\_113756831-113756880, chr11\_108839740-108839780) were found to be expressed in the developing mouse kidney. *miR-10a*, chr9\_113756831-113756880 and chr11\_108839740-108839780 exhibited a distinct expression pattern within the nephrogenic zone, encompassing nephron progenitors, renal stroma and the ureteric bud epithelium. Together, these assays validate both the sRNA-Seq dataset and the approach for novel miRNA discovery in the developing mouse kidney.

### Usage Notes

The sRNA-Seq data presented here represents a comprehensive resource for miRNA expression in developing nephrogenic mesenchymal cells enriched for nephron progenitors and whole kidney. This resource can be utilised to predict regulatory networks downstream of these expressed miRNAs. For example, this dataset can be integrated with other published RNA-Seq data containing transcriptomic information (mRNA and long non-coding RNA) from nephron progenitors and whole kidney for the analysis of miRNA-mRNA interactions using algorithms such as TargetScan<sup>39</sup> or DIANA-Tools<sup>40</sup>. Such

analysis would allow one to uncover potential candidate miRNAs and/or downstream target mRNAs for future studies. For example, our sRNA-Seq dataset supports the expression of the highly conserved *let-7* family members (*let-7a*, *7b*, *7c*, *7d*, *7e*, *7f*, *7g*, *7h*, *7i*, *7j*, *7k* and *miR-98*) in the developing kidney with *let-7c*, *7d*, *7e* being differentially enriched in nephron progenitors compared to whole kidney during development. The pathogenesis of Wilms tumour has been linked to overexpression of Lin28, a negative regulator of *let-7* microprocessing<sup>41</sup>. Wilms tumour is the most common paediatric kidney cancer and arises from a failure of embryonic kidney tissues to terminally differentiate. Thus, de-repression of *let-7* downstream targets including *Hmga2*<sup>42,43</sup>, *Ras*<sup>44</sup> and *Myc*<sup>45</sup> oncogenes may be implicated in the malignant transformation of nephron progenitors in Wilms Tumour<sup>41</sup>. This integrative analysis approach allows for future hypothesis testing studies to elucidate the biological function of key miRNAs in nephron progenitors.

Finally, the sRNA-Seq dataset can be complemented with published CHIP-Seq datasets to infer potential relationships between expressed miRNAs and *cis*-regulatory regions in nephron progenitors. The transcription factor Six2 is known to be critical in maintaining self-renewal and has been shown to regulate target genes that are indispensable for maintaining multipotency of these cells during nephrogenesis<sup>46</sup>. By analysing available Six2 CHIP-Seq data (SRA Study: SRP064623; GEO: GSE73867)<sup>47</sup> with our current sRNA-Seq dataset, it is now possible to curate candidate miRNAs that are potentially regulated by Six2. Two examples are Six2 binding within the promoter regions of *miR-210* and *125b*, miRNAs that are enriched in nephron progenitors based on this sRNA-Seq dataset (Fig. 5). This is an example of how to use this resource to identify putative Six2-regulated miRNAs that may play important roles in nephron progenitor maintenance and differentiation during kidney development.

## References

1. Ho, J. & Kreidberg, J. A. The long and short of microRNAs in the kidney. *J Am Soc Nephrol* **23**, 400–404 (2012).
2. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051–4060 (2004).
3. Han, J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016–3027 (2004).
4. Kim, Y. K., Kim, B. & Kim, V. N. Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis. *Proc Natl Acad Sci USA* **113**, E1881–E1889 (2016).
5. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
6. Bernstein, E. *et al.* Dicer is essential for mouse development. *Nat Genet* **35**, 215–217 (2003).
7. MacRae, I. J., Ma, E., Zhou, M., Robinson, C. V. & Doudna, J. A. In vitro reconstitution of the human RISC-loading complex. *Proc Natl Acad Sci USA* **105**, 512–517 (2008).
8. Rivas, F. V. *et al.* Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol* **12**, 340–349 (2005).
9. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91–105 (2007).
10. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
11. Little, M. H. & McMahon, A. P. Mammalian kidney development: principles, progress, and projections. *Cold Spring Harb Perspect Biol* **4**:a008300 (2012).
12. Cebrian, C., Asai, N., D'Agati, V. & Costantini, F. The number of fetal nephron progenitor cells limits ureteric branching and adult nephron endowment. *Cell Rep* **7**, 127–137 (2014).
13. Brenner, B. M., Garcia, D. L. & Anderson, S. Glomeruli and blood pressure. *Less of one, more the other?* *Am J Hypertens* **1**, 335–347 (1988).
14. Foulkes, W. D., Priest, J. R. & Duchaine, T. F. DICER1: mutations, microRNAs and mechanisms. *Nat Rev Cancer* **14**, 662–672 (2014).
15. Ho, J. *et al.* The pro-apoptotic protein Bim is a microRNA target in kidney progenitors. *J Am Soc Nephrol* **22**, 1053–1063 (2011).
16. Nagalakshmi, V. K. *et al.* Dicer regulates the development of nephrogenic and ureteric compartments in the mammalian kidney. *Kidney Int* **79**, 317–330 (2011).
17. Harvey, S. J. *et al.* Podocyte-specific deletion of dicer alters cytoskeletal dynamics and causes glomerular disease. *J Am Soc Nephrol* **19**, 2150–2158 (2008).
18. Ho, J. *et al.* Podocyte-specific loss of functional microRNAs leads to rapid glomerular and tubular injury. *J Am Soc Nephrol* **19**, 2069–2075 (2008).
19. Phua, Y. L. *et al.* Renal stromal miRNAs are required for normal nephrogenesis and glomerular mesangial survival. *Physiol Rep* **3**, e12537 (2015).
20. Shi, S. *et al.* Podocyte-selective deletion of dicer induces proteinuria and glomerulosclerosis. *J Am Soc Nephrol* **19**, 2159–2169 (2008).
21. Zhdanova, O. *et al.* The inducible deletion of Drosha and microRNAs in mature podocytes results in a collapsing glomerulopathy. *Kidney Int* **80**, 719–730 (2011).
22. Park, S. M., Gaur, A. B., Lengyel, E. & Peter, M. E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* **22**, 894–907 (2008).
23. Brown, A. C., Muthukrishnan, S. D. & Oxburgh, L. A synthetic niche for nephron progenitor cells. *Dev Cell* **34**, 229–241 (2015).
24. Cloonan, N. *et al.* MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol* **12**, R126 (2011).
25. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
26. Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* **40**, 37–52 (2012).
27. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140–D144 (2006).
28. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
29. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).
30. Tyner, C. *et al.* The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res* **45**, D626–D634 (2017).
31. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034–1050 (2005).
32. Siepel, A. & Haussler, D. Combining phylogenetic and hidden Markov models in biosequence analysis. *J Comput Biol* **11**, 413–428 (2004).

33. Kobayashi, A. *et al.* Six2 defines and regulates a multipotent self-renewing nephron progenitor population throughout mammalian kidney development. *Cell Stem Cell* **3**, 169–181 (2008).
34. Park, J. S., Valerius, M. T. & McMahon, A. P. Wnt/beta-catenin signaling regulates nephron induction during mouse kidney development. *Development* **134**, 2533–2539 (2007).
35. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* **7**, 3 (2006).
36. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**, e108 (2013).
37. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
38. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
39. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
40. Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I. & Hatzigeorgiou, A. G. Functional microRNA targets in protein coding sequences. *Bioinformatics* **28**, 771–776 (2012).
41. Urbach, A. *et al.* Lin28 sustains early renal progenitors and induces Wilms tumor. *Genes Dev* **28**, 971–982 (2014).
42. Lee, Y. S. & Dutta, A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev* **21**, 1025–1030 (2007).
43. Mayr, C., Hemann, M. T. & Bartel, D. P. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science* **315**, 1576–1579 (2007).
44. Johnson, S. M. *et al.* RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635–647 (2005).
45. Sampson, V. B. *et al.* MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. *Cancer Res* **67**, 9762–9770 (2007).
46. Self, M. *et al.* Six2 is required for suppression of nephrogenesis and progenitor renewal in the developing kidney. *EMBO J* **25**, 5214–5228 (2006).
47. O'Brien, L. L. *et al.* Differential regulation of mouse and human nephron progenitors by the Six family of transcriptional regulators. *Development* **143**, 595–608 (2016).

### Data Citations

1. *NCBI Sequence Read Archive* SRP134975 (2018).
2. *NCBI Gene Expression Omnibus* GSE111729 (2018).
3. Phua, Y. L. *et al.* *Figshare* <http://dx.doi.org/10.6084/m9.figshare.c.4156250> (2018).

### Acknowledgements

Dr. Ho's laboratory is supported by funding by an NIDDK R00DK087922, NIDDK R01DK103776 and a March of Dimes Basil O'Connor Starter Scholar Award. Dr. Yu Leng Phua was a George B. Rathmann Research Fellow supported by the *American Society of Nephrology* Ben J. Lipps Research Fellowship Program. Kevin Chen was supported by the Research Advisory Committee of Children's Hospital of Pittsburgh. Dr. Kostka's laboratory is supported by an NIGMS GM115836. We thank the supercomputing resources provided by the University of Pittsburgh Center for Research Computing.

### Author Contributions

Y.L.P. and K.H.C. performed the experiments. A.C., Y.L.P. and K.H.C. analysed the data with supervision from D.K. and J.H. Y.L.P., A.C., D.K. and J.H. wrote the manuscript, with the final version approved by all authors.

### Additional Information

**Competing interests:** The authors declare no competing interests.

**How to cite this article:** Phua, Y.L. *et al.* Small non-coding RNA expression in mouse nephrogenic mesenchymal progenitors. *Sci. Data*. 5:180218 doi: 10.1038/sdata.2018.218 (2018).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files made available in this article.

© The Author(s) 2018