

Keywords: ccRCC; miRNA; tumorigenesis; recurrence; obesity

MicroRNA profiling in clear cell renal cell carcinoma tissues potentially links tumorigenesis and recurrence with obesity

X Shu¹, M A Hildebrandt¹, J Gu¹, N M Tannir², S F Matin³, J A Karam³, C G Wood³ and X Wu^{*1}

¹Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA; ²Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA and ³Department of Urology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Background: Twenty to 40% localised RCC patients may experience recurrence after curative surgery. Limited miRNA predictors have been identified for ccRCC recurrence.

Methods: Through a multi-phase study design, we analysed miRNAs in tissues obtained from 203 ccRCC patients. Paired t-test was used for tumour–normal comparisons and Cox regression model was performed to compute hazard ratios (HRs) and corresponding 95% CIs.

Results: A 17-miRNA signature was identified that can concordantly classify > 95% of tumour/adjacent normal samples. Significant enrichment was found as 6 out of 17 miRNAs were associated with obesity (binomial probability = 0.001). Decreased levels of miR-204-5p and miR-139-5p were each associated with an approximately three-fold increased risk of recurrence ($P < 0.01$). Risk score was generated based on expressions of miR-204-5p and miR-139-5p, and the trend test was significant in both discovery and validation sets ($P_{\text{for trend}} < 0.05$). Striking MST reduction was observed for patients with a high-risk score (high vs low: discovery, 9.4 vs >97.7 months; validation, 20.8 vs >70.3 months). Expressions of miR-204-5p were also associated with body mass index ($\beta = 5.64$, $P < 0.001$). Significant inverse correlations were observed and validated between miR-204-5p and 13 obesity-related genes ($r < 0$, $P < 0.01$).

Conclusions: We identified 17 miRNAs dysregulated in ccRCC tissues and showed that low expressions of miR-204-5p and miR-139-5p were associated with the higher risk of recurrence. The link between miR-204-5p and ccRCC recurrence may be partially mediated by regulating the expression of targeted obesity-related genes.

Kidney cancer remains one of the top 10 most common cancers for both males and females in the United States (Siegel *et al*, 2016). Over 80% of kidney cancers are composed of renal cell carcinoma (RCC) and the major histological type is clear cell RCC (ccRCC). Smoking, history of hypertension, and obesity are the three established risk factors for RCC (Chow *et al*, 2010). It has been estimated that ~20–40% of localised patients may experience disease recurrence after curative therapy (Janzen *et al*, 2003) and five-year survival rate is only 11.7% for the patients with distant metastasis (Howlader *et al*, 2013). Previous studies have reported

that clinical stage (Fergany *et al*, 2000; Leibovich *et al*, 2003; Russo *et al*, 2008), grade (Leibovich *et al*, 2003; Russo *et al*, 2008), tumour size (Leibovich *et al*, 2003), and microvascular invasion (VanPoppel *et al*, 1997) are associated with the recurrence or progression. Interestingly, in contrast to its association with an increased risk for RCC development, being obese may reduce the risk of recurrence and increase the overall survival (Yu *et al*, 1991; Parker *et al*, 2006). However, the complete set of predictors for RCC recurrence has not been well established, underscoring the need for new biomarkers to identify those at high risk.

*Correspondence: Dr X Wu; E-mail: xwu@mdanderson.org

Received 15 June 2016; revised 15 September 2016; accepted 31 October 2016; published online 1 December 2016

© 2017 Cancer Research UK. All rights reserved 0007–0920/17

MiRNAs are a class of noncoding RNAs of 18–25 nucleotides in length. MiRNAs bind to the 3'-untranslated region (UTR) of their target genes, typically resulting in gene silencing by triggering the degradation of the target mRNA. MiRNAs are promising biomarkers for the cancer risk and prognosis due to its stability and functionality (Bartel, 2004). Multiple miRNA signatures for RCC tumorigenesis have been generated by performing quantitative real-time PCR, microarray, or next-generation sequencing (Jung *et al*, 2009; White *et al*, 2011; Osanto *et al*, 2012). Recently, The Cancer Genome Atlas (TCGA) has completed genome-wide profiling of miRNAs in samples of various types of cancer, including ccRCC. A signature based on levels of miR-10b, miR-21, miR-204, miR-30a, miR-143, and let-7a has been reported by TCGA, which may distinguish ccRCC patients with favourable survival (Cancer Genome Atlas Research N, 2013). High levels of miR-21 were correlated with worse survival in another study (Faragalla *et al*, 2012). Knowledge regarding the role of miRNAs in relation to RCC recurrence is still limited. One study reported that levels of miR-143, miR-26a, miR-145, miR-10b, miR-195, miR-126, and miR-127 were decreased in RCC patients with recurrence (Slaby *et al*, 2012). In another study, a 4-miRNA signature was generated consisting of miR-10b, miR-139-5p, miR-130b, and miR-199b-5p (Wu *et al*, 2012). However, the sample sizes of these previous studies have been relatively small.

In this study, we set out to accomplish two aims. First, identify a miRNA signature for tumorigenesis through a multi-stage design to provide a better insight of miRNAs dysregulated in ccRCC. Second, identify miRNAs associated with ccRCC recurrence and explore their associations with obesity.

MATERIALS AND METHODS

Study population and sample collection. A total of 135 MD Anderson ccRCC patients were involved in the present study. The details of the study population have been reported previously (Clague *et al*, 2009). There were no age, sex, ethnicity, or cancer-stage restrictions on recruitment. Patient demographic variables, tobacco and alcohol use history, weight and height to calculate body mass index (BMI), and medical history were obtained by in-person interview. For smoking history, a never smoker was defined as an individual who had never smoked or had smoked <100 cigarettes. Those subjects who had quit smoking >12 months before diagnosis/recruitment were considered former smokers. Clinical information was abstracted from patient medical records, including clinical stage, grade, comorbidities, tumour size, pathological stage, histology, treatment, and clinical outcomes. We also collected tumour and adjacent normal tissues from a subset of patients (Hildebrandt *et al*, 2012). We utilised 32 tumour–normal pairs and 32 unpaired tumour samples from 64 patients in the discovery. Tumour–normal pairs were collected from all 71 patients in the validation. External independent data set consisted of 68 ccRCC tumour–normal pairs was downloaded from the TCGA portal (<https://tcga-data.nci.nih.gov/tcga/>).

The study flowchart is shown in Figure 1. For tumour–normal comparisons, three data sets consisting of data generated from 32, 68, and 71 pairs of ccRCC tumour and adjacent normal tissues were used. For testing the association with ccRCC recurrence, 64 (31 recurrence vs 33 non-recurrence) and 71 tumour samples

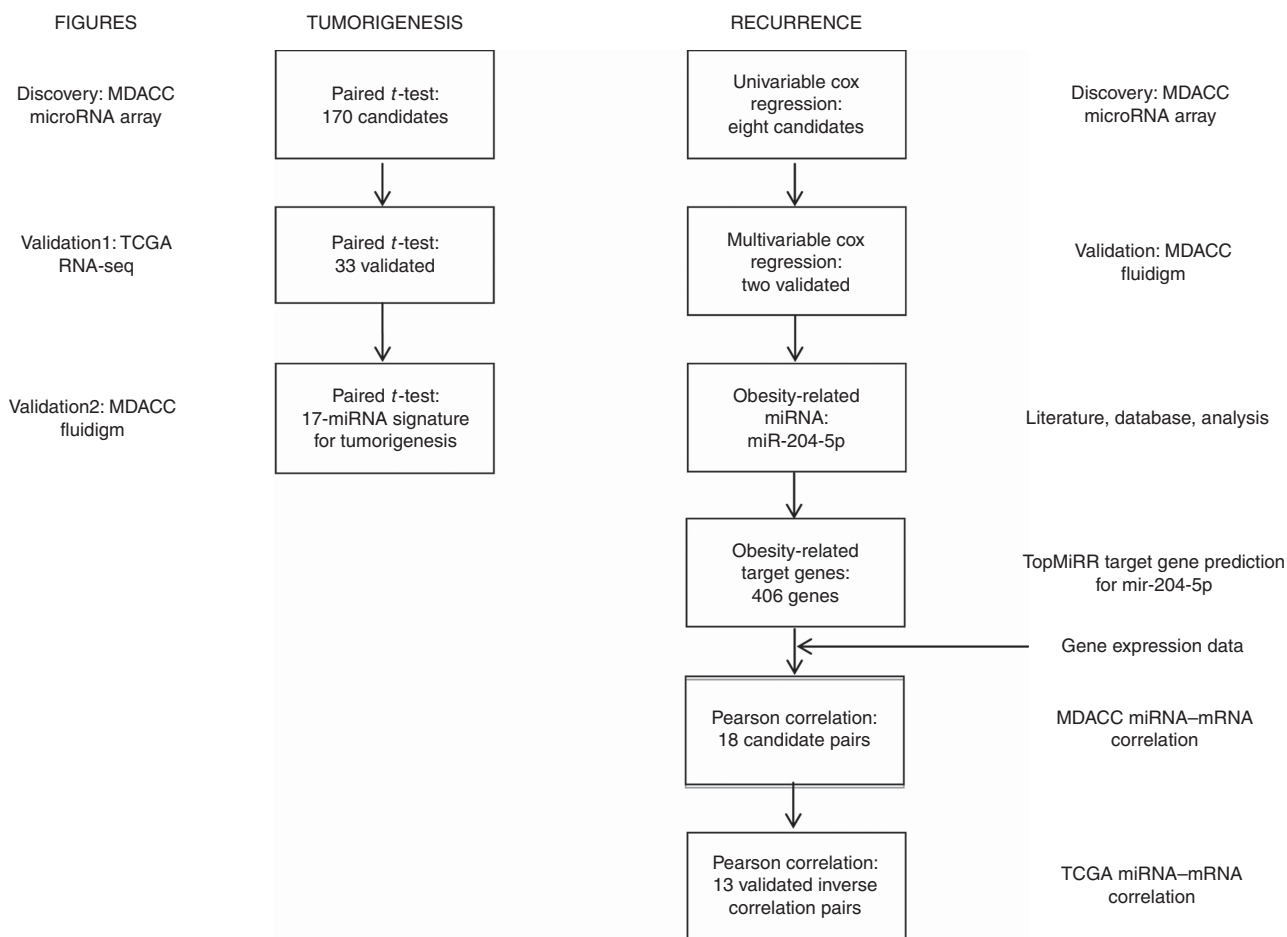


Figure 1. Study flowchart.

(29 recurrence vs 42 non-recurrence) were utilised in the discovery and validation phases, respectively. ccRCC patients without recurrence were frequency matched to patients with recurrence by age, sex, and clinical stage (I and II/III and IV) in the validation. The study protocol was approved by the MD Anderson Cancer Center Institutional Review Board. All participated patients provided written informed consent.

Lab procedures

Discovery phase. Total RNA was extracted from 32 pairs of ccRCC tumour and adjacent normal tissues, and 32 unpaired ccRCC tumour tissues. Total RNA was isolated using the *mirVana* RNA Isolation Kit (Life Technologies, Gaithersburg, MD, USA). Labelled cDNA was synthesised, amplified, and purified from 300 ng total RNA using the TotalPrep RNA Amplification Kit (Life Technologies). Each sample was then hybridised to Human-6 v2 Expression BeadChips and read using the BeadStation 500 scanner (Illumina, San Diego, CA, USA) to generate gene expression data (Hildebrandt *et al*, 2012). MiRNAs were profiled using Illumina Sentrix Array Matrix 96-well MicroRNA Expression Profiling Assays. After quality control (calling rate <80% and excluding non-mature miRNAs), 300 mature human miRNAs remained for the final analyses.

Validation phase. Same protocols for total RNA isolation, reverse transcription, and PreAmp were used for the 71 pairs of ccRCC tumour and adjacent normal tissues in the validation.

Selection of miRNAs for validation. There were two parts of validation for our study. First, for tumour–normal comparisons, candidate miRNAs with significant differential expressions between 32 paired ccRCC tumour–normal tissues from the discovery ($P < 0.01$), and the association was also significant in the same direction in the TCGA data set (> 200 reads per million miRNA mapped, $P < 0.01$) were selected. For the association of recurrence, significant miRNAs in univariable Cox regression ($P < 0.01$) were selected for the analysis in the validation set.

MiRNA expression by Fluidigm Microfluidics Dynamic Arrays. Selected miRNAs were measured using high-throughput BioMark HD Real-Time PCR system (Fluidigm, South San Francisco, CA, USA). Briefly, reverse transcription was carried out as described above using pooled miRNA primers with 150 ng of total input RNA. Pre-amplification was performed with pooled Taqman assay. PCR products were cleaned up using an enzymatic digestion approach by Exonuclease I (New England Biolabs, Ipswich, MA, USA). After pre-amplification, a 5 μ l sample mixture was prepared for each sample. The IFC controller HX (Fluidigm) was used to distribute the sample mix and assay mix from the loading inlets into the 96.96 Dynamic array reaction chips. After loading, the chip was placed in the BioMark instrument for real-time PCR at 95 °C for 10 min, followed by 40 cycles at 95 °C for 15 s and 60 °C for 1 min. Data were analysed with Real-Time PCR Analysis Software in the BioMark instrument (Fluidigm).

Each PCR reaction was done in duplicate and the mean of cycle threshold (Ct) was calculated. Small nuclear RNAs U44 were used as internal control for input normalisation. The mean Ct value of each sample was normalised to the averaged expression of U44 snRNA and then subjected to analysis with $2^{-\Delta\Delta Ct}$ method. Data were set to missing from further analysis if one of the following criteria were fulfilled: (1) the generated duplicated Ct values with over one cycle variance; (2) samples with a Ct value > 35 ; and (3) miRNAs with a detection rate $< 80\%$.

Compilation of obesity-related miRNAs and genes. We compiled a list of obesity-related miRNAs ($N = 75$) from an online database (Kunej *et al*, 2013). Sixty nine of 75 miRNAs were covered in our miRNA microarray. We compiled a list of obesity-related

genes according to four sources: (1) bioinformatics tool Text-mined Hypertension, Obesity, and Diabetes Candidate Gene Database (Dai *et al*, 2013). We restricted genes to those reported by three or more studies and further reviewed these genes in details. Two-hundred and sixteen candidate genes remained in further analysis. (2) Online database of obesity-related genes: integratome TIME (Kunej *et al*, 2013). (3) Obesity-relevant pathways selected from Biocarta, KEGG and Reactome pathway databases. Fifteen pathways included in our list were: adipocytokine signalling pathway (KEGG, 67 genes), type II diabetes mellitus (KEGG, 47 genes), insulin signalling pathway (KEGG, 137 genes), insulin signalling pathway (BioCarta, 22 genes), IGF-1 signalling pathway (BioCarta, 21 genes), leptin pathway (BioCarta, 11 genes), PPAR signalling pathway (KEGG, 69 genes), metabolism of lipids and lipoproteins (Reactome, 478 genes), peroxisomal lipid metabolism (Reactome, 21 genes), P53-hypoxia pathway (BioCarta, 23 genes), mTOR signalling pathway (KEGG, 52 genes), mTOR pathway (BioCarta, 23 genes), energy-dependent regulation of mTOR by LKB1-AMPK (Reactome, 18 genes), oxidative phosphorylation (KEGG, 135 genes), and mitochondrial fatty acid beta oxidation (Reactome, 14 genes). (4) Genes close to GWAS confirmed loci for BMI or obesity. We downloaded the list consisted of 43 studies from a Catalog of Published Genome-Wide Association Studies (<http://www.genome.gov/gwastudies/>). The keywords used for searching were: BMI, obesity, obesity (early-onset extreme), obesity (extreme), BMI (interaction), adiposity, fat body mass, and weight. We included both upstream and downstream genes closest to the SNP if it is located in the intergenic regions. Loci with genome-wide significant SNPs ($P < 5 \times 10^{-8}$) were eligible to be studied. One hundred genes were included after duplicates being removed. In total, we compiled a list consisted of 2051 obesity-related genes.

Prediction of target genes for miRNA. We used a web-based analytical tool: ToppMiR (Wu *et al*, 2014) (<https://toppmir.cchmc.org/>) to search putative miRNA targets for obesity-related miRNAs. It searches for any evidence of putative target genes for miRNAs through multiple prediction tools, including: mirSVR, miRTarbase, MsiGDB, TargetScan, miRecords, PicTar and PITA.

Statistical analysis. MiRNAs that had been detected in $< 80\%$ of samples were excluded. Continuous host characteristics were analysed using Student's *t*-test, whereas categorical variables were analysed using Pearson's χ^2 -test. miRNA array data were quantile normalised (Bolstad *et al*, 2003) and \log_2 transformed in the discovery set, whereas reads per million miRNA-mapped values were obtained from TCGA miRNA-seq data and CT values generated from Fluidigm platform were used. For tumour–normal comparisons, paired Student's *t*-test was performed to compare levels of normalised miRNAs between ccRCC tumour and adjacent normal pairs. Fold change was calculated as normalised miRNA in tumour samples divided by normalised miRNA in paired adjacent normal tissues for both the TCGA and our miRNA microarray data. In the validation, fold change was calculated using $2^{-\Delta\Delta Ct}$ method ($2^{-\Delta\Delta Ct} = 2^{-((Ct_{miRNA} \text{ in tumour} - Ct_{RNU44} \text{ in tumour}) - (Ct_{miRNA} \text{ in normal} - Ct_{RNU44} \text{ in normal}))}$). For discovery phase analysis of recurrence, the normalised levels of miRNAs were directly compared between the patients with and without recurrence. In the validation, relative quantification (RQ) was calculated as $2^{-((Ct_{miRNA} - Ct_{miRNA} \text{ mean}) - (Ct_{RNU44} - Ct_{RNU44} \text{ mean}))}$ in tumour tissues using Ct value of each miRNA. Levels (RQ) of miRNAs were dichotomised according to the median level in patients without recurrence. The associations between 300 miRNAs and obesity were tested in the adjacent normal tissues in the discovery phase, using univariable and multivariable linear regression model. The adjustment included age, sex, smoking status, and history of hypertension. Pearson's correlations were conducted to compare the correlations of fold changes of miRNAs identified for the tumorigenesis signature between data sets. Kaplan–Meier survival

curves and Cox regression models were performed to test dichotomised miRNAs in association with recurrence-free survival.

To test whether the identified miRNAs were enriched to be obesity related, we calculated the binomial probability for observing the exact number or more miRNAs that are obesity related in the present study, given the assumption that 23 miRNA are obesity related ($P < 0.05$ in the multivariable linear regression model and listed in the Integratomics TIME database; Supplementary Table S5) among 300 miRNAs tested (probability of event is $23/300 = 0.077$). Pearson's correlations were conducted for the predicted miRNA-mRNA pairs using our ccRCC miRNA and mRNA microarray data from our discovery phase ($N_{\text{sample}} = 64$) and the TCGA ccRCC level 3 tumour miRNA and mRNA-seq data ($N_{\text{sample}} = 236$, both Pearson's correlation $P < 0.01$). To maintain consistency, the TCGA mRNA data were also \log_2 transformed. All tests were performed using STATA 13.0 (College Station, TX, USA) and R 3.01. The heatmaps were generated using GenePattern (v3.1, Broad Institute, Cambridge, MA, USA). The hierarchical clustering algorithm with Pearson's correlation as column distance measure and pairwise average linkage as clustering method were used.

RESULTS

The study flowchart is shown in Figure 1. Host characteristics of 135 ccRCC patients recruited at MD Anderson are shown (Supplementary Table S1). The proportions of male (80.6% vs 39.4%), current smoker (22.6% vs 12.1%), and late-stage patients (III and IV, 71.0% vs 21.2%) are higher in the 31 patients with recurrence, compared with 33 patients without recurrence in the discovery. We frequency matched 29 patients with recurrence to 42 patients without by age, sex, and clinical stage in the validation.

MiRNAs differentiating ccRCC tumour-normal pairs

Discovery. One-hundred seventy miRNAs were dysregulated in the 32 ccRCC tumour samples compared with their adjacent normal samples ($P < 0.01$; Supplementary Table S2). The fold change ranged from 0.14 to 10.4 with miR-514a-3p being most downregulated and miR-122a-5p being most upregulated.

First validation. Significant miRNAs identified in the discovery phase were tested using data from 68 TCGA ccRCC tumour-normal pairs. Thirty three of 170 miRNAs were significantly dysregulated with the same up/downregulated trend in tumours ($P < 0.01$; Supplementary

Table S2). The fold change ranged from 0.03 to 13.6 with miR-141-3p being most downregulated and miR-155-5p being most upregulated. The scatter plot showed the correlations between fold changes of expressions of the 33 candidate miRNAs in the discovery and TCGA validation data sets (Supplementary Figure S1A). The correlation coefficient was 0.96 ($P < 0.001$).

Second validation. We further tested these 33 miRNAs in a third independent data set that consisted of 71 ccRCC tumour-normal pairs. The final signature consisted of 13 downregulated and 4 upregulated miRNAs across the three independent data sets (Figure 2; Supplementary Table S2). The correlation between fold changes of expressions of these 17 validated miRNAs between discovery and second validation data sets was highly significant ($r = 0.90$, $P < 0.001$; Supplementary Figure S1B). In clustering analysis using the 17-miRNA signature, it concordantly classified $> 95\%$ of samples (Figure 2; Supplementary Figures S2 and S3). Interestingly, six of these miRNAs were identified to be obesity related that was significant in the enrichment test (6 out of 17, binomial probability = 0.001; Supplementary Table S5).

MiRNAs associated with ccRCC recurrence

Discovery. The results of univariable Cox regression identified three upregulated and five downregulated miRNAs in patients with recurrence (dichotomised, $P < 0.01$; Supplementary Table S3). The most significant upregulated and downregulated miRNA in patients with recurrence was miR-365a-3p (hazard ratio (HR) = 0.32, 95% CI = 0.15–0.68, $P = 0.003$) and miR-204-5p (HR = 3.15, 95% CI = 1.47–6.76, $P = 0.003$), respectively. The miRNA expressions were also modelled as continuous predictors which showed persistent associations.

Validation. Two of the eight miRNAs were significantly associated with ccRCC recurrence in the validation set (dichotomised, $P < 0.01$; Table 1). In the univariable model, low levels of miR-204-5p was associated with a significantly increased three-fold risk of ccRCC recurrence (HR = 3.01, 95% CI = 1.34–6.80, $P = 0.008$). Similar results were found for miR-139-5p (HR = 2.79, 95% CI = 1.29–6.03, $P = 0.009$). The associations remained significant and the strengths subtly changed with the adjustment of covariates (Supplementary Table S4). Consistent associations were observed when modelling miRNAs levels in continuous form (Table 1; Supplementary Table S4). Striking reduction in recurrence-free median survival time, from > 107.2 months to 46.2.0 months ($P_{\text{log-rank}} = 0.002$) and 62.9 months to 25.0 months ($P_{\text{log-rank}} = 0.006$)

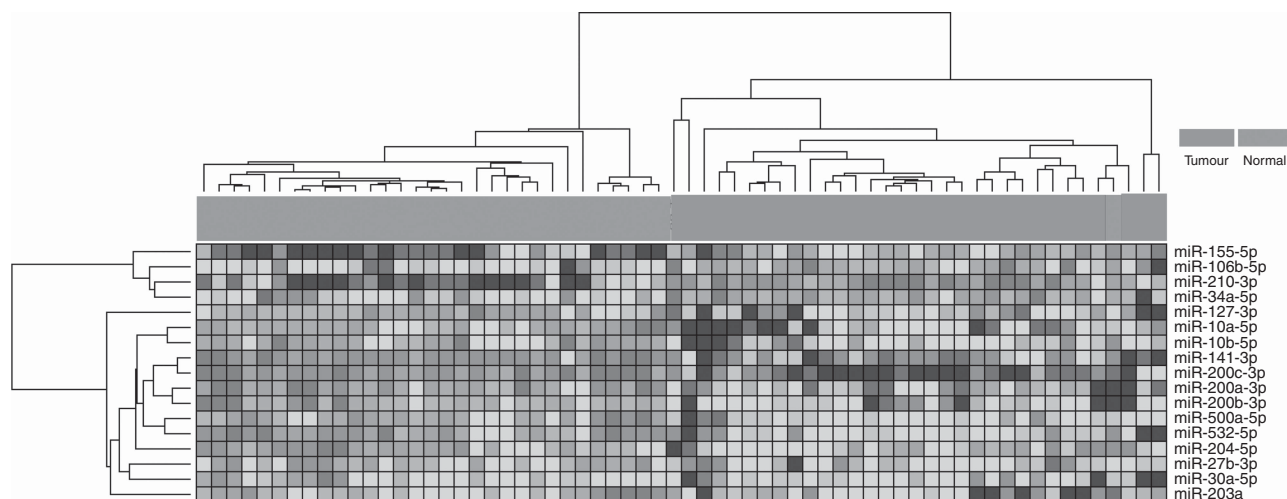


Figure 2. Heatmap for MDACC discovery set of miRNA tumour-normal comparisons. A full colour version of this figure is available at *British Journal of Cancer* online.

Table 1. Univariable Cox regression of miRNAs and ccRCC recurrence in the discovery and validation phases

miRNA	Up/down in patients with recurrence	Discovery			Validation			Obesity related
		HR	95% CI	P-value	HR	95% CI	P-value	
miR-204-5p ^a	Down	0.55	0.35–0.86	0.009	0.85	0.77–0.94	0.001	Yes
High expression ^b Low expression ^b		Ref. 3.15	 1.47–6.76	 0.003	 3.01	 1.34–6.80	 0.008	
miR-139-5p ^a	Down	0.39	0.23–0.68	<0.001	0.74	0.62–0.88	<0.001	No
High expression ^b Low expression ^b		Ref. 3.10	 1.45–6.62	 0.003	 2.79	 1.29–6.03	 0.009	

^aUnivariable Cox regression with miRNAs ($-\Delta\Delta Ct$) modelled as continuous variables.

^bLevels were dichotomised according to median level in non-recurrence patients.

was observed for low miR-204-5p levels in the discovery and validation set, respectively. Similarly, the decrease was observed from >113.5 months to 39.5 months ($P_{\log\text{-rank}}=0.002$) and 70.3 months to 25.0 months ($P_{\log\text{-rank}}=0.006$) for low miR-139-5p levels in the discovery and validation set, respectively (Supplementary Figure S4). Risk score derived from miR-139-5p and miR-204-5p was able to stratify our study population into high-risk, intermediate and low-risk groups (Figure 3). The increasing risk of recurrence with higher-risk score was consistently observed in the discovery and validation sets in the multivariable Cox regressions ($P_{\text{for trend}}<0.05$; Table 2).

Target gene prediction for miR-204-5p and test for the correlations. Obesity-related miR-204-5p ($\beta=5.64$, $P<0.001$; Supplementary Table S5) was associated with both ccRCC tumorigenesis (Figure 2) and recurrence. The test of enrichment was not significant (one out of two, binomial probability = 0.148). Of 2051 obesity-related genes (Supplementary Table S6), 406 were predicted to be regulated by miR-204-5p. We further tested the correlation between miR-204-5p expression and gene expression levels for each of the 395 genes that were available in the same samples (Supplementary Table S7). Eighteen pairs exhibited significant inverse Pearson correlations ($P<0.01$) with receptor tyrosine kinase-like orphan receptor 2 (*ROR2*) being most significant ($r=-0.53$, $P=1.46\times 10^{-6}$). We conducted the same correlation analysis for the genes in the TCGA data and 13 miRNA-gene pairs remained highly statistically significant (Table 3). The P -value reached 10^{-22} for insulin-like growth factor 2 (*IGF2*) mRNA-binding protein 2 (*IGF2BP2*) that was most significant in TCGA data set.

DISCUSSION

In the present study, several miRNAs identified were obesity related. Most interestingly, obesity-related miR-204-5p was associated with both ccRCC tumorigenesis and recurrence. Furthermore, miR-204-5p was consistently inversely correlated with 13 obesity-related genes in two independent data sets.

Many miRNAs in our signature for tumorigenesis overlap with findings from the previous studies. For example, overexpressed miR-34a, miR-155, miR-210, and under-expressed miR-10a/b, miR-30a, miR-141, miR-200a/b/c, miR-204, miR-500a, and miR-532 were reported by Jung *et al* (2009) and Juan *et al* (2010). Our findings support the role of these miRNAs played in the development of the disease. Moreover, the highly significant correlation coefficient between fold changes of expressions of validated dysregulated miRNAs across different data sets demonstrated the robustness of our findings. With the incorporation of a three-stage design that strengthens our findings, we further refined miRNA profiles of ccRCC tumorigenesis.

Both miR-204-5p and miR-139-5p were identified as most influential miRNAs for ccRCC pathogenesis in the network analysis (Butz *et al*, 2014). Interestingly, low expression of miR-204 was the key feature (together with high levels of miR-21) of discriminatory miRNA group 2 identified by original TCGA analysis, in which the patients had worst prognosis compared with other groups (Cancer Genome Atlas Research N, 2013). However, limited studies have focused on the role of these miRNAs in the development of ccRCC recurrence. Our observation of association between decreased level of miR-204-5p and shorter RCC recurrence-free survival is novel. In a previous study, lower levels of miR-204-5p were observed in RCC patients who progressed to metastatic disease compared with those without progression. However, it is not clear what covariates were adjusted in their analyses (Gowrishankar *et al*, 2014). It has been suggested that miR-204-5p may function as a tumour suppressor. Higher expressions of miR-204-5p have been observed in breast and gastric cancer tissues obtained from patients free from disease metastasis (Li *et al*, 2014). *In vitro* studies have shown that overexpression of miR-204-5p could markedly suppress cell migration and invasion in different cell lines (Chung *et al*, 2012; Qiu *et al*, 2013; Ying *et al*, 2013). Previous studies also have reported the tumour suppressive function of miR-139-5p in cancer recurrence or metastasis. In one study, reduced level of miR-139-5p was found in tissues obtained from ccRCC patients with recurrence after nephrectomy (Slaby *et al*, 2012). However, miR-139-5p was not selected for validation. In other studies, it was consistently downregulated in ccRCC metastatic samples and oesophageal squamous cell carcinoma tissues obtained from patients with lymph node metastasis (Wu *et al*, 2012; Liu *et al*, 2013). miR-139-5p was also found to be associated with ccRCC survival in some studies but not all (Osanto *et al*, 2012; Wu *et al*, 2012). Our findings further support the role of miR-139-5p as a tumour suppressor in cancers, including ccRCC, although additional studies are required to validate our findings.

Interestingly, several miRNAs identified in this study were obesity related. One caveat is that our study was not ideal to identify or confirm whether the associated miRNAs are obesity related. The obesity-related miRNAs were defined based on association tests in our own samples with relatively small sample size and large number of comparisons. Therefore, we can only make a suggestive inference. Studies using samples collected from healthy subjects and with larger sample size would be more appropriate to achieve this goal. However, we further used a database to consolidate our observations.

Obesity has been related to later recurrence and favourable survival in RCC patients (Yu *et al*, 1991; Kamat *et al*, 2004; Parker *et al*, 2006). Studies showed that comparing with those with normal BMI, obese RCC patients had >50% reduced risk of recurrence and longer survival (Yu *et al*, 1991; Parker *et al*, 2006). However, the exact mechanisms involved in these processes have been elusive. Therefore, our exploratory analysis for miR-204-5p

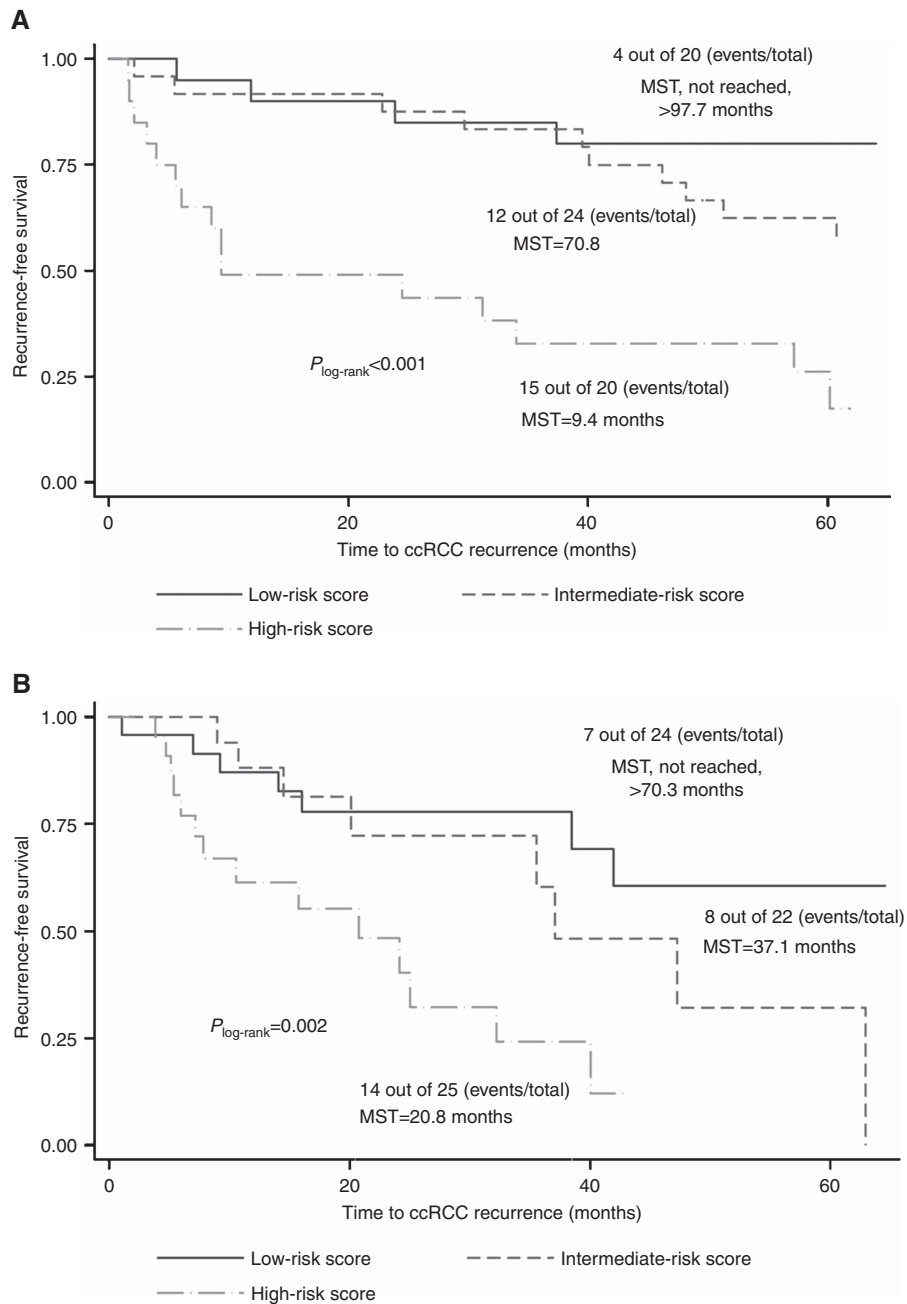


Figure 3. Five-year recurrence-free survival curves for miR-204-5p and miR-139-5p. (A, B) Kaplan–Meier curves of miR-204-5p in the discovery/validation. (C, D) Kaplan–Meier curves of miR-139-5p in the discovery/validation.

Table 2. Risk group stratification using risk score generated from miR-204 and miR-139								
Risk group ^a	Discovery				Validation			
	HR ^b	95% CI	P-value	P _{trend}	HR ^b	95% CI	P-value	P _{trend}
Univariable								
Low RS	Ref.				Ref.			
Intermediate RS	2.69	0.87–8.34	0.087		1.78	0.64–4.92	0.271	
High RS	8.00	2.60–8.34	<0.001	<0.001	4.67	1.80–12.1	0.002	0.001
Multivariable								
Low RS	Ref.				Ref.			
Intermediate RS	1.15	0.31–4.32	0.834		2.89	0.87–9.57	0.082	
High RS	3.45	0.81–14.7	0.094	0.035	6.91	1.94–24.6	0.003	0.003

^aAbbreviations: Low RS = low-risk score consists of high levels of miR-204 and miR-139; Intermediate RS = intermediate-risk score consists of high level of either miR-204 or miR-139; High RS = high-risk score consists of low levels of both miRNAs.

^bMultivariable model 1: adjusted for age, gender, clinical stage, smoking, hypertension, and obesity (BMI < 30 vs BMI ≥ 30).

Table 3. Pearson correlation test for obesity-related genes potentially regulated by miR-204-5p

Gene	MDACC		TCGA	
	r	P-value	r	P-value
ROR2	-0.53	1.46×10^{-6}	-0.48	6.84×10^{-15}
ADAM12	-0.49	9.50×10^{-6}	-0.50	5.01×10^{-16}
PTX3	-0.41	3.02×10^{-4}	-0.34	5.52×10^{-8}
IGF2BP2	-0.40	4.14×10^{-4}	-0.58	1.40×10^{-22}
IL2RA	-0.37	1.22×10^{-4}	-0.31	1.37×10^{-6}
GAD1	-0.35	2.43×10^{-3}	-0.45	1.35×10^{-12}
IL1R1	-0.34	2.90×10^{-3}	-0.37	3.59×10^{-9}
MTMR4	-0.34	3.37×10^{-3}	-0.47	1.43×10^{-14}
GFAP	-0.32	4.86×10^{-3}	-0.39	7.21×10^{-10}
TFAP2A	-0.32	5.96×10^{-3}	-0.55	4.27×10^{-20}
DGKG	-0.31	6.51×10^{-3}	-0.24	2.52×10^{-4}
GPC3	-0.30	8.32×10^{-3}	-0.36	9.27×10^{-9}
CIDEA	-0.30	8.86×10^{-3}	-0.34	4.23×10^{-7}

and obesity deserves further investigations. Overexpressed miR-204-5p was shown to promote adipocyte differentiation and increase lipid droplet accumulation in mesenchymal stem cell lines (Huang *et al*, 2010; Alexander *et al*, 2011). In this study, the expression of miR-204-5p was also positively associated with BMI. Thus, miR-204-5p may contribute to ccRCC recurrence through its link with obesity. In addition, we observed several putative target genes regulated by miR-204-5p, which includes *IGF2BP2*. *IGF2BP2* binds to the 5'-UTR of the *IGF2* mRNA transcripts and subsequently represses the translational process (Nielsen *et al*, 1999). Increasing evidence supports its association with obesity and cancer risk (Sandhu *et al*, 2003; Livingstone, 2013). Several GWAS have also linked a common variant located in *IGF2BP2* (rs4402960) to risk of type 2 diabetes (Saxena *et al*, 2007; Scott *et al*, 2007; Zeggini *et al*, 2007).

Another gene of interest is *ADAM12*, which is also involved in the IGF receptor signalling pathway (Kveiborg *et al*, 2008). The overexpression of *ADAM12* has been reported in various cancers (Wewer *et al*, 2005). Its protease and adhesion activities, stimulation on cell proliferation, and increased resistance to apoptosis may contribute to the progression of tumours (Kveiborg *et al*, 2005, 2008). Therefore, we hypothesised that miR-204-5p may serve as an intermediate between obesity and recurrence, potentially through IGF signalling (Supplementary Figure S5).

Our study has several strengths. The sample size of the present study is relatively large in comparison to other studies. Importantly, independent data sets were used to validate our findings. The miRNAs remained significant in the multivariable model, which indicates their independent prognostic value. In addition, to increase the likelihood that the predicted miRNA-mRNA relationships are plausible, we used a prediction tool that integrates multiple prediction algorithms and evaluated the correlations in two independent data sets. We also recognised several limitations of our study. Although the findings were validated in independent internal/external data sets, the possibility of false positives still exists. In addition, the correlation tests were exploratory that laboratory-based experiments are required to validate the putative miRNA-mRNA relationships. Another limitation is that our data set is not ideal to investigate the relationship between miRNAs and obesity. Finally, the curated obesity-related gene set includes genes having various biological functions that the genes are not 'obesity related' only. However, there is no well-defined obesity-related gene set that could be found in any commonly used databases, including BioCarta, KEGG, Reactome, and GO.

Our findings may have clinical implications in predicting ccRCC patients who are at higher risk of recurrence and provide new insights of mechanisms involved in the link between obesity and ccRCC recurrence. However, more efforts are warranted to establish the exact biological mechanisms for the interplay of obesity, miRNAs and their targeted genes, and ccRCC recurrence.

ACKNOWLEDGEMENTS

This work was supported in part by the National Institutes of Health (grant R01 CA170298), and the Center for Translational and Public Health Genomics, Duncan Family Institute for Cancer Prevention, The University of Texas MD Anderson Cancer Center.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Alexander R, Lodish H, Sun L (2011) MicroRNAs in adipogenesis and as therapeutic targets for obesity. *Expert Opin Ther Targets* **15**(5): 623–636.
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**(2): 281–297.
- Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**(2): 185–193.
- Butz H, Szabo PM, Nofech-Mozes R, Rotondo F, Kovacs K, Mirham L, Girgis H, Boles D, Patocs A, Yousef GM (2014) Integrative bioinformatics analysis reveals new prognostic biomarkers of clear cell renal cell carcinoma. *Clin Chem* **60**(10): 1314–1326.
- Cancer Genome Atlas Research N (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**(7456): 43–49.
- Chow WH, Dong LM, Devesa SS (2010) Epidemiology and risk factors for kidney cancer. *Nat Rev Urol* **7**(5): 245–257.
- Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, Chan LK, Yu MY, Kwong J, Doran G, Barroilhet LM, Ng AS, Wong RR, Wang VW, Mok SC, Smith DI, Berkowitz RS, Wong YF (2012) Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXO1. *Int J Cancer* **130**(5): 1036–1045.
- Clague J, Lin J, Cassidy A, Matin S, Tannir NM, Tamboli P, Wood CG, Wu X (2009) Family history and risk of renal cell carcinoma: results from a case-control study and systematic meta-analysis. *Cancer Epidemiol Biomarkers Prev* **18**(3): 801–807.
- Dai HJ, Wu JCY, Tsai RTH, Pan WH, Hsu WL (2013) T-HOD: a literature-based candidate gene database for hypertension, obesity and diabetes. *Database (Oxford)* **2013**: bas061.
- Faragalla H, Youssef YM, Scorilas A, Khalil B, White NMA, Mejia-Guerrero S, Khella H, Jewett MAS, Evans A, Lichner Z, Bjarnason G, Sugar L, Attalah MI, Yousef GM (2012) The clinical utility of miR-21 as a diagnostic and prognostic marker for renal cell carcinoma. *J Mol Diagn* **14**(4): 385–392.
- Fergany AF, Hafez KS, Novick AC (2000) Long-term results of nephron sparing surgery for localized renal cell carcinoma: 10-year followup. *J Urol* **163**(2): 442–445.
- Gowrishankar B, Ibragimova I, Zhou Y, Sliker MJ, Devarajan K, Al-Saleem T, Uzzo RG, Cairns P (2014) MicroRNA expression signatures of stage, grade, and progression in clear cell RCC. *Cancer Biol Ther* **15**(3): 329–341.
- Hildebrandt MAT, Tan WQ, Tamboli P, Huang MS, Ye YQ, Lin J, Lee JS, Wood CG, Wu XF (2012) Kinome expression profiling identifies IKBKE as a predictor of overall survival in clear cell renal cell carcinoma patients. *Carcinogenesis* **33**(4): 799–803.
- Howlader N, Noone A, Krapcho M (2013) *SEER Cancer Statistics Review, 1975-2010*. National Cancer Institute: Bethesda, MD.
- Huang J, Zhao L, Xing L, Chen D (2010) MicroRNA-204 regulates Runx2 protein expression and mesenchymal progenitor cell differentiation. *Stem Cells* **28**(2): 357–364.

- Janzen NK, Kim HL, Figlin RA, Beldegrun AS (2003) Surveillance after radical or partial nephrectomy for localized renal cell carcinoma and management of recurrent disease. *Urol Clin North Am* **30**(4): 843–852.
- Juan D, Alexe G, Antes T, Liu H, Madabhushi A, Delisi C, Ganesan S, Bhanot G, Liou LS (2010) Identification of a microRNA panel for clear-cell kidney cancer. *Urology* **75**(4): 835–841.
- Jung M, Mollenkopf HJ, Grimm C, Wagner I, Albrecht M, Waller T, Pilarsky C, Johannsen M, Stephan C, Lehrach H, Nietfeld W, Rudel T, Jung K, Kristiansen G (2009) MicroRNA profiling of clear cell renal cell cancer identifies a robust signature to define renal malignancy. *J Cell Mol Med* **13**(9B): 3918–3928.
- Kamat AM, Shock RP, Naya Y, Rosser CJ, Slaton JW, Pisters LL (2004) Prognostic value of body mass index in patients undergoing nephrectomy for localized renal tumors. *Urology* **63**(1): 46–50.
- Kunje T, Jevsinek Skok D, Zorc M, Ogrinc A, Michal JJ, Kovac M, Jiang Z (2013) Obesity gene atlas in mammals. *J Genomics* **1**: 45–55.
- Kveiborg M, Albrechtsen R, Couchman JR, Wewer UM (2008) Cellular roles of ADAM12 in health and disease. *Int J Biochem Cell Biol* **40**(9): 1685–1702.
- Kveiborg M, Frohlich C, Albrechtsen R, Tischler V, Dietrich N, Holck P, Kronqvist P, Rank F, Mercurio AM, Wewer UM (2005) A role for ADAM12 in breast tumor progression and stromal cell apoptosis. *Cancer Res* **65**(11): 4754–4761.
- Leibovich BC, Blute ML, Chevillie JC, Lohse CM, Frank I, Kwon ED, Weaver AL, Parker AS, Zincke H (2003) Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma - A stratification tool for prospective clinical trials. *Cancer* **97**(7): 1663–1671.
- Li WD, Jin XJ, Zhang QB, Zhang G, Deng XB, Ma L (2014) Decreased expression of miR-204 is associated with poor prognosis in patients with breast cancer. *Int J Clin Exp Pathol* **7**(6): 3287–3292.
- Liu R, Yang M, Meng Y, Liao J, Sheng J, Pu Y, Yin L, Kim SJ (2013) Tumor-suppressive function of miR-139-5p in esophageal squamous cell carcinoma. *PLoS One* **8**(10): e77068.
- Livingstone C (2013) IGF2 and cancer. *Endocr Relat Cancer* **20**(6): R321–R339.
- Nielsen J, Christiansen J, Lykke-Andersen J, Johnsen AH, Wewer UM, Nielsen FC (1999) A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. *Mol Cell Biol* **19**(2): 1262–1270.
- Osanto S, Qin Y, Buermans HP, Berkers J, Lerut E, Goeman JJ, van Poppel H (2012) Genome-wide microRNA expression analysis of clear cell renal cell carcinoma by next generation deep sequencing. *PLoS One* **7**(6): e38298.
- Parker AS, Lohse CM, Chevillie JC, Thiel DD, Leibovich BC, Blute ML (2006) Greater body mass index is associated with better pathologic features and improved outcome among patients treated surgically for clear cell renal cell carcinoma. *Urology* **68**(4): 741–746.
- Qiu YH, Wei YP, Shen NJ, Wang ZC, Kan T, Yu WL, Yi B, Zhang YJ (2013) miR-204 inhibits epithelial to mesenchymal transition by targeting slug in intrahepatic cholangiocarcinoma cells. *Cell Physiol Biochem* **32**(5): 1331–1341.
- Russo P, Jang TL, Pettus JA, Huang WC, Eggner SE, O'Brien MF, Karellas ME, Karanikolas NT, Kagiwada MA (2008) Survival rates after resection for localized kidney cancer: 1989 to 2004. *Cancer* **113**(1): 84–96.
- Sandhu MS, Gibson JM, Heald AH, Dunger DB, Wareham NJ (2003) Low circulating IGF-II concentrations predict weight gain and obesity in humans. *Diabetes* **52**(6): 1403–1408.
- Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Alshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Rickett D, Purcell S (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**(5829): 1331–1336.
- Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**(5829): 1341–1345.
- Siegel RL, Miller KD, Jemal A (2016) Cancer statistics, 2016. *CA Cancer J Clin* **66**(1): 7–30.
- Slaby O, Redova M, Poprach A, Nekvindova J, Iliev R, Radova L, Lakomy R, Svoboda M, Vyzula R (2012) Identification of MicroRNAs associated with early relapse after nephrectomy in renal cell carcinoma patients. *Genes Chromosomes Cancer* **51**(7): 707–716.
- VanPoppel H, Vandendriessche H, Boel K, Mertens V, Goethuys H, Haustermans K, VanDamme B, Baert L (1997) Microscopic vascular invasion is the most relevant prognosticator after radical nephrectomy for clinically nonmetastatic renal cell carcinoma. *J Urology* **158**(1): 45–49.
- Wewer UM, Albrechtsen R, Engvall E (2005) *ADAM12: The long and the short of it*. Springer: The Netherlands.
- White NM, Bao TT, Grigull J, Youssef YM, Girgis A, Diamandis M, Fatoohi E, Metias M, Honey RJ, Stewart R, Pace KT, Bjarnason GA, Yousef GM (2011) miRNA profiling for clear cell renal cell carcinoma: biomarker discovery and identification of potential controls and consequences of miRNA dysregulation. *J Urol* **186**(3): 1077–1083.
- Wu C, Bardes EE, Jegga AG, Aronow BJ (2014) ToppMiR: ranking microRNAs and their mRNA targets based on biological functions and context. *Nucleic Acids Res* **42**: W107–W113.
- Wu X, Weng L, Li X, Guo C, Pal SK, Jin JM, Li Y, Nelson RA, Mu B, Onami SH, Wu JJ, Ruel NH, Wilczynski SP, Gao H, Covarrubias M, Figlin RA, Weiss LM, Wu H (2012) Identification of a 4-microRNA signature for clear cell renal cell carcinoma metastasis and prognosis. *PLoS One* **7**(5): e35661.
- Ying Z, Li Y, Wu JH, Zhu X, Yang Y, Tian H, Li W, Hu B, Cheng SY, Li MF (2013) Loss of miR-204 expression enhances glioma migration and stem cell-like phenotype. *Cancer Res* **73**(2): 990–999.
- Yu ML, Asal NR, Geyer JR (1991) Later recurrence and longer survival among obese patients with renal-cell carcinoma. *Cancer* **68**(7): 1648–1655.
- Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**(5829): 1336–1341.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)