

CLEAR CELL RENAL CELL CARCINOMA BIOMARKERS

CROSS-REFERENCE TO RELATED APPLICATIONS

[001] This application claims the benefit of priority of Singapore application No.
5 10201707218R, filed 05 September 2017, the contents of it being hereby incorporated by reference in its entirety for all purposes.

FIELD OF INVENTION

[002] The present invention relates to molecular biology in particular biomarkers. In
10 particular, the present invention relates to biomarkers associated with clear cell renal cell carcinoma (ccRCC) and methods and uses thereof.

BACKGROUND OF THE INVENTION

[003] Renal cell carcinoma (RCC) is one of the most deadly cancers due to frequent late
15 diagnosis and poor treatment options. Success in curing the disease relies on early detection of RCC and complete resection of the malignant cells. Since the kidney lies deeply in the retroperitoneal space, renal cell carcinoma is primarily asymptomatic in the early phase and upon diagnosis the tumour is large and/or metastasized. The three most common subtypes of renal cell carcinoma are clear cell renal cell carcinoma, papillary renal cell carcinoma and chromophobe
20 renal cell carcinoma. Clear cell renal cell carcinoma is the most common subtype, accounting for 75-90% of all renal cell carcinomas and with 338,000 new cases in 2012 worldwide.

[004] With most clear cell renal cell carcinomas being resistant to chemotherapy and
radiotherapy, patients with metastatic clear cell renal cell carcinomas exhibit a dismal 8% five-year overall survival. Even early stage tumours remain at risk of metastatic progression after surgery,
25 with 20-40% of patients having recurrence. Therefore, identification of this high-risk group of renal cell carcinoma patients remains a challenge. Furthermore, different subtypes of renal cell carcinoma have variable prognoses and treatment response rates. Therefore, it is crucial to be able to differentiate between different subtypes of renal cell carcinoma.

[005] Previous methods for diagnosing patients with clear cell renal cell carcinoma involve
30 invasive methods, including tumour biopsy, or imaging methods including ultrasound imaging or magnetic resonance imaging. However, based on known methods, it is difficult to determine whether a renal mass of less than 4 cm is a tumour, and/or whether the tumour is benign or malignant based on imaging studies alone. Around 50% to 60% of renal masses are less than 4 cm, of which 25% to 30% are benign tumours. The risk of overtreatment for small renal masses ranges
35 from 40% for lesions less than 1 cm to 17% for masses 3 to 4 cm in diameter. In addition to initial

diagnosis, the main tools for post-treatment follow up or active surveillance also include only imaging studies. Surgery or ablation would result in tissue change and scar formation, causing the detection of local recurrence to be challenging. Lastly, there is also a lack of methods to assess the efficiency of systemic treatments in advanced clear cell renal cell carcinoma patients. In view of the above, there is an unmet need for a method of identifying clear cell renal cell carcinoma, differentiation of benign lesions from malignant tumours, for detection of recurrence after local treatments, and for assessment of systemic treatments.

SUMMARY OF INVENTION

- 10 [006] In one aspect, the present invention refers to a clear cell renal cell carcinoma (ccRCC) biomarker set, wherein the biomarker set comprises at least two biomarkers selected from the group consisting of ZNF395, SMPDL3A, SLC28A1, SLC6A3, VEGFA, EGLN3, wherein one of the at least two biomarkers is SMPDL3A or SLC28A1; wherein the biomarkers are proteins, or nucleic acids encoding the same, or variants thereof.
- 15 [007] In another aspect, the present invention refers to a detection system comprising a) a receiving section to receive a sample from a subject suspected to suffer from clear cell renal cell carcinoma, and wherein the sample is suspected to comprise the biomarker set as disclosed herein; and b) a detection section comprising a substance or substances capable of detecting the biomarker set as disclosed herein.
- 20 [008] In one aspect, the present invention refers to a method of determining whether a subject has or shows recurrence of clear cell renal cell carcinoma, wherein the method comprises obtaining a sample from the subject; detecting the presence of the biomarker set as disclosed herein in the sample using a detection system as disclosed herein, wherein the presence of the biomarker set determines that the subject has or shows recurrence of clear cell renal cell carcinoma.
- 25 [009] In a further aspect, the present invention refers to a method of detecting response of a subject to systemic treatment, the method comprising a) obtaining a sample from the subject; and b) determining the levels of the biomarker set as defined herein in the sample; wherein a decrease in levels or an absence of the biomarker set indicates that the subject is responsive to treatment.
- 30 [0010] In yet another aspect, the present invention refers to a method of determining whether a renal mass sample is benign or malignant, the method comprising obtaining a sample from the renal mass of a subject; determining the levels or the presence or absence of the biomarker set as disclosed herein in the sample; wherein the increase in levels of the biomarker set compared to a benign sample indicates that the sample is malignant.

[0011] In another aspect, the present invention refers to a kit for carrying out the method as disclosed herein, wherein the kit comprises a detection buffer, a lysis buffer, and a substance or substances as defined herein suitable for the detection of the biomarker set as disclosed herein.

[0012] In a further aspect, the present invention refers to a kit as disclosed herein and a
5 detection system as disclosed herein for detecting the biomarker set as defined herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in
10 which:

[0014] **Figure 1** illustrates that von Hippel-Lindau (*VHL*) deficient clear cell renal cell carcinoma tumours exhibit an aberrant cis-regulatory landscape. **Figure 1A** shows a graph that illustrates the percentage of overlap of the histone chromatin immunoprecipitation sequencing (ChIP-seq) (H3K27ac, H3K4me3 and H3K4me1) peaks of normal kidney tissues with peaks from
15 adult kidney tissues in the Epigenomics Roadmap dataset. **Figure 1B** shows a graph illustrating the percentage of overlap of histone ChIP-seq (H3K27ac, H3K4me3 and H3K4me1) peaks between five primary clear cell renal cell carcinoma tumours and cell lines derived from these tumours. **Figure 1C** shows a table and a diagram illustrating that the putative active promoters are defined by co-occurrence of H3K4me3, H3K27ac within 2 kilo bases (kb) proximity to transcription start sites (TSS); and putative active enhancers are defined by the presence of H3K4me1, H3K27ac and the exclusivity with promoters. The table further illustrates the total number of identified putative promoters and putative enhancers; and the total number of gained promoters, lost promoters,
20 gained enhancers and lost enhancers identified in this study. **Figure 1D** shows a graph illustrating principal component analysis (PCA) using all 17,497 promoters and 66,448 enhancers to classify normal samples and tumour samples into distinct clusters. The numbers in the graph depicts patient IDs which are the following 1-12364284; 2-17621953; 3-20431713; 4-40911432; 5-57398667; 6-70528835; 7-74575859; 8-77972083; 9-86049102 and 10-75416923. **Figure 1E** shows graphs illustrating saturation analysis of the total number of predicted promoters or enhancers across increasing number of primary samples. The total number of predicted promoters saturates at 4 or
25 more samples while the total number of predicted enhancers saturates at 16 or more samples. The dotted line indicates the total number of predicted regulatory elements by integrating all 10 normal-tumour pairs (n=20). The whiskers indicate standard deviations. **Figure 1F** shows graphs and tables describing the variances captured by each principle component from normalized H3K27ac signals at promoters or enhancers. The cumulative percentages of variance are indicated in the
30 tables. **Figure 1G** shows a graph of the number of altered promoters and enhancers per patient.

Figure 1H shows a graph of the fraction of altered regions that meet statistical significance defined by paired t-tests with Benjamini-Hochberg correction (q value < 0.10) at different cut-offs of recurrence. **Figure 1I** shows a graph of the differences in the fractions of regions meeting statistical significance. Promoters reach saddle point at $n \geq 5$ while enhancers reach saddle point at $n \geq 6$. **Figure 1J** shows a heatmap with H3K27ac levels of altered promoters and enhancers in a paired patient tissue (patient 40911432). High signal levels are reflected in white and low signal levels are reflected in black. **Figure 1Ki and Figure 1Kii** shows a plot referring to examples of H3K27ac chromatin immunoprecipitation sequencing (ChIP-Seq) signals in 10 normal-tumour pairs shown for gained promoter, lost promoter, unaltered promoter, gained enhancer, lost enhancer and unaltered enhancer. N refers to signal from normal tissue and T refers to signal from tumour tissue. **Figure 1L** shows box plots of H3K27ac levels, chromatin accessibility and DNA methylation of gained promoters and gained enhancers. Gene expression of the nearest clear cell renal cell carcinoma long coding RNA (lncRNA) is compared between normal and tumour tissues. *** p -value < 0.001 , two-sided t -test. **Figure 1M** shows a plot referring to histone ChIP-seq signals (H3K27ac, H3K4me1, H3K4me3), RNA-Seq signals and FAIRE-Seq signals at the *CCND1* locus in a tumour-normal tissue sample pair of patient 40911432. For comparison, the histone ChIP-seq profiles of normal adult kidney tissue from the Epigenome Roadmap are displayed above the normal tissue profiles generated by Nano-ChIP-seq. The histone profile of a cell line derived from tumour tissue is displayed below the profile of the normal tissue.

[0015] **Figure 2** shows that enhancer aberration is a signature of clear cell renal cell carcinoma. **Figure 2A** shows bar graphs of enriched pathways associated with gained promoters and gained enhancers revealed by GREAT algorithm that is ranked by binomial FDR q -value. **Figure 2B** illustrates a plot that refers to a histone chromatin immunoprecipitation sequencing (ChIP-Seq) profile of *VEGFA*. *De novo* enhancers are acquired in a clear cell renal cell carcinoma tumour tissue upstream of *VEGFA*. Capture-C confirmed interactions of this *VEGFA* enhancer (E) with its promoter (P) in 786-O cells. The arcs represent significant interactions detected by r3Cseq ($q < 0.05$). The input-subtracted H3K27ac signals of this enhancer are highly correlated with *VEGFA* gene expression (Spearman's correlation). **Figure 2C** illustrates a plot that refers to a histone ChIP-Seq profile of *SLC2A1/GLUT1*. A *de novo* tumour enhancer interacts with the *SLC2A1/GLUT1* promoter. **Figure 2D** and **Figure 2E** illustrates a plot that refers to a histone ChIP-Seq profile of (D) *PLIN2* and (E) *SLC38A1*, with gain of promoters and enhancers near overexpressed respective clear cell renal cell carcinoma oncogenes *PLIN2* and *SLC38A1*. **Figure 2F** shows a graph of the top 5 gene ontology Molecular Functions of tumour promoters and enhancers. **Figure 2G** shows a dot plot referring to Spearman's correlation between gene expression of *VEGFA* and *SLC2A1* and the input subtracted H3K27ac levels of their predicted

enhancers in 10 tumour samples and their matched normals. **Figure 2H** shows a graph referring to the cumulative distribution of distance spanned by significant Capture C interactions.

[0016] **Figure 3** illustrates the identification of key oncogenic drivers by tumour super-enhancers. **Figure 3A** shows a graph referring to a total of 1,451 super-enhancers that are identified by ROSE and ranked by their differential H3K27ac intensity between normal and tumour tissues. Genes associated with the top gained and lost super-enhancers are listed. **Figure 3B** shows a plot that refers to RNA sequencing (RNA seq), histone chromatin immunoprecipitation sequencing (ChIP-seq) and Capture C profiles of *PVT1/MYC* gene. Capture C shows chromosomal interactions between the c-Myc promoter and the super-enhancer. **Figure 3C** shows a plot that refers to RNA seq and histone ChIP-seq of *EPAS1* gene. Histone ChIP-Seq validated gained super-enhancers at *PVT1/MYC* (Figure 3B) and *EPAS1* (Figure 3C) loci overlapping with a renal cell carcinoma risk allele respectively. **Figure 3D** shows a heatmap of The Cancer Genome Atlas (TCGA) RNA-seq data indicating that genes associated with top 10 gained enhancers are upregulated in tumours while genes associated with top 10 lost enhancers are downregulated. This tumour-specificity is restricted to clear cell renal cell carcinoma, but not the other two renal cell carcinoma subtypes, papillary and chromophobe. Without being bound by theory, it is thought that *ZNF395*, *SLC28A1*, *SMPLD3A*, *VEGFA* and *EGLN3* are able to distinguish clear cell renal cell carcinoma from other major renal cell carcinoma subtypes, based on p-value and tumour/normal ratio (T/N) shown. **Figure 3E** refers to a graph with expression of *ZNF395* and *SMPLD3A* measured in a panel of normal kidney cell lines and clear cell renal cell carcinoma cell lines by real-time PCR (RT-qPCR). **Figure 3F** shows an immunoblot comparing protein expression of *ZNF395* in a panel of clear cell renal cell carcinoma cell lines. **Figure 3G** refers to images that indicate pooled siRNA against *ZNF395* inhibits colony formation of clear cell renal cell carcinoma cell lines A-498 and 786-O but not HK-2 normal immortalized kidney cells. Pooled siRNA against *SMPLD3A* inhibits colony formation of A-498 but not 786-O. **Figure 3H** shows a graph of siRNA knockdown efficiency of *SMPLD3A* and *ZNF395* as measured by real-time PCR (RT-qPCR) in HK-2, 786-O and A498 cells. **Figure 3I** refers to a plot of histone ChIP profile with H3K27ac ChIP-seq showing an active *ZNF395* super-enhancer only in clear cell renal cell carcinoma cells (A-498 and 786-O) but not normal kidney cells (PCS-400, HK-2). **Figure 3J** refers to plots of histone ChIP profile with H3K27ac and H3K4me3 ChIP-seq of tumour/normal pair showing that *SMPLD3A* or *SLC28A1* is associated with a clear cell renal cell carcinoma specific super enhancer. **Figure 3K (Figure 3Ki, 3Kii and 3Kiii)** refers to a plot of histone ChIP-seq profile with H3K27ac, H3K4me3 and H3K4me1 ChIP-seq of tumour/normal pair showing the gain of promoters and enhancers in tumours for genes *SLC6A3*, *EGLN3* or *VEGFA*. **Figure 3L to Figure 3P** refers to graphs showing expression data obtained from The Cancer Genome Atlas (TCGA).

Figure 3L shows expression of *SMPDL3A* in a panel of cancers, with the highest expression being present in clear cell renal cell carcinoma (KIRC) (TCGA symbol for clear cell renal cell carcinoma). **Figure 3M** shows expression of *SLC28A1* in a panel of cancers, with the highest expression being present in clear cell renal cell carcinoma (KIRC). **Figure 3N** shows expression of *SLC6A3* in a panel of cancers, with the highest expression being present in clear cell renal cell carcinoma (KIRC). **Figure 3O** shows expression of *VEGFA* in a panel of cancers, with the highest expression being present in clear cell renal cell carcinoma. **Figure 3P** shows expression of *EGLN3* in a panel of cancers, with the highest expression being present in clear cell renal cell carcinoma (KIRC). **Figure 3Q** refers to a box plot with The Cancer Genome Atlas (TCGA) RNA-seq data that shows exclusive overexpression of *ZNF395* amongst 12 cancer types. **Figure 3R** shows a graph referring to shRNA knockdown efficiency of *ZNF395* levels measured by reverse transcription polymerase chain reaction (RT-PCR) in 786-O and A-498 cells. **Figure 3S** illustrates an immunoblot referring to shRNA knockdown efficiency of *ZNF395* levels measured by immunoblotting in 786-O and A-498 cells. **Figure 3T** shows images of *ZNF395* inhibition by two shRNA clones that decrease colony formation in 786-O and A-498 cells. **Figure 3U** refers to graphs of *ZNF395* inhibition by two shRNA clones that decrease *in vitro* proliferation. **Figure 3V** shows a graph of *ZNF395* inhibition by two shRNA clones that increases apoptosis measured by cleavage of Caspase3/7 substrate. **p*-value < 0.05, two-sided *t*-test. **Figure 3W** shows histograms of Annexin V staining analyzed by flow cytometry after *ZNF395* shRNA knockdown in 786-O and A-498 cells. **Figure 3X** refers to line graphs showing *ZNF395* inhibition by shRNA that leads to total elimination of A-498 tumours *in vivo* and delayed tumour growth in 786-O cells. Negative control (NC): n=7, shZNF395-1: n=7, shZNF395-2: n=6

[0017] **Figure 4** illustrates that *VHL* deficiency remodels clear cell renal cell carcinoma enhancers. **Figure 4A** show graphs referring to *in vitro* proliferation of 786-O, A-498 and 12364284 cell lines with and without *VHL* restoration. Proliferation rates were measured with CellTiterGlo, and normalized to day of seeding. EV refers to empty vector control and VHL refers to wild-type VHL restored. **Figure 4B** refers to images of *in vitro* colony formation of 786-O, A-498 and 12364284 cell lines with and without *VHL* restoration. Rates of colony formation were measured by seeding 10,000 cells in the well and allowing colonies to form until the wells are confluent. EV refers to empty vector control and VHL refers to wild-type VHL restored. **Figure 4C** shows graphs referring to apoptosis measured by cleavage of caspase3/7 substrates and normalized to empty vector controls of 786-O, A-498 and 12364284 cell lines, with and without VHL restoration. EV refers to empty vector control and VHL refers to wild-type VHL restored. **Figure 4D** illustrates a graph where *in vivo* growth of 786-O subcutaneous tumours in nude mice is compared for isogenic cells with and without VHL restoration. EV refers to empty vector control

and VHL refers to wild-type VHL restored. **Figure 4E** shows dot plots with log fold changes of H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) signals at gained promoters, gained enhancers and gained super-enhancers as defined in the primary clear cell renal cell carcinoma dataset after *VHL* restoration in 786-O cells. Dots represent *cis*-regulatory elements with significant changes (p -value < 0.05 , negative binomial) in H3K27ac levels after *VHL* restoration. The number and percentage of altered regions (p -value < 0.05 , negative binomial) are shown at the upper and lower right corners. **Figure 4F, Figure 4G and Figure 4H** shows dot plots with log fold changes of H3K27ac ChIP-seq signals at gained promoters, gained enhancers and gained super-enhancers after *VHL* restoration in (Figure 4F) A-498 cells, (Figure 4G) 12364284 cells and (Figure 4H) 40911432 cells. Dots represent p -value < 0.05 . The number and percentage of altered regions are shown at the upper and lower right corners. EV refers to empty vector control and VHL refers to wild-type VHL restored. **Figure 4I** shows box plots with read coverage of H3K27ac ChIP-seq at *VHL*-responsive enhancers in *VHL*-mutant clear cell renal cell carcinoma cell lines compared to *VHL*-wild-type clear cell renal cell carcinoma, normal kidney cell lines and 31 other cancer cell lines. Enhancers with H3K27ac depletion or enrichment after *VHL* restoration are shown. **Figure 4J** refers to a box plot that shows changes in expression of genes linked to *VHL*-responsive tumour enhancers after *VHL* restoration in 786-O cells. $*p$ -value < 0.05 , two-sided t -test. **Figure 4K** shows a box plot with changes of gene expression linked to *VHL*-responsive tumour enhancers in 12364284 cells. $*p$ -value value < 0.05 , two-sided t -test. **Figure 4L** refers to a bar graph with frequency of gained enhancers showing H3K27ac depletion after *VHL* restoration in patients. **Figure 4M** shows a heatmap with unsupervised hierarchical clustering of differential H3K27ac ChIP-seq signals at gained enhancers showing H3K27ac depletion after *VHL* restoration. **Figure 4N** shows a plot with H3K27ac ChIP-seq signals of all 10 tumour/normal pairs at the *ZNF395* super-enhancer. **Figure 4O, Figure 4P and Figure 4Q** show plots of histone ChIP seq signals with examples of lost *VHL*-responsive enhancer/super-enhancers associated with *EGFR* (Figure 4O), *CCND1* (Figure 4P) and *ITGB3* (Figure 4Q) in 786-O cells. **Figure 4R** show plot of histone ChIP signals with examples of lost *VHL*-responsive enhancer associated with *SLC2A1* in 786-O cells. **Figure 4S** show plot of histone ChIP seq signals with examples of lost *VHL*-responsive enhancer/super-enhancers associated with *VEGFA* in 786-O cells. **Figure 4T** show plot of histone ChIP signals with examples of lost *VHL*-responsive enhancer associated with *HK2* in 786-O cells. **Figure 4U** shows dot plot of Pearson's correlation of log fold changes of H3K27ac and H3K4me1 in 786-O and 12364284 after *VHL* restoration. (removed as there is no colour in figure) **Figure 4V** shows dot plot of Pearson's correlation of log fold changes of H3K27ac and H3K27me3 in 786-O and 12364284 after *VHL* restoration. **Figure 4W** refers to heatmap with log

fold changes of H3K27ac, H3K4me1 and H3K27me3 signals at gained enhancers showing H3K27ac depletion after *VHL* restoration in 786-O cells.

[0018] **Figure 5** illustrates that HIF2 α is enriched at enhancers of *VHL*-responsive tumour tumours. **Figure 5A** refers to a table with motif analysis of gained enhancers using HOMER, revealing significant enrichment of AP-1 family, ETS family, NF κ B and HIF1 α /2 α (hypergeometric test). Lost enhancers were used as background in the motif search to identify tumour-specific transcription factors. **Figure 5B** shows an immunoblot with protein expressions of putative transcription factors enriched at gained enhancers in 9 tumour cell lines (4 commercial cell lines and 5 patient-derived cell lines) and 2 normal cell lines. ACHN is a papillary renal cell carcinoma cell line. **Figure 5C** shows scatter plots with gene expression of selected transcription factors in 73 pairs of normal kidney and clear cell renal cell carcinoma tumours of the The Cancer Genome Atlas (TCGA) cohort (RNA-Seq dataset). ***p-value < 0.001, **p-value < 0.01, n.s. (not significant), paired t-test. **Figure 5D** illustrates graphs showing that chromatin immunoprecipitation sequencing (ChIP-seq) validated the enrichment of transcription factors in gained enhancers over lost enhancers. **Figure 5E** refers to an immunoblot with protein expression of transcription factors shown in 786-O and 12364284 cells with and without wild-type *VHL*. **Figure 5F** illustrates line graphs of transcription factor binding at *VHL*-responsive gained enhancers that shows enrichment of HIF2 α and HIF1 β at enhancers with H3K27ac depletion over regions with H3K27ac enrichment after *VHL* restoration. **Figure 5G** refers to pie charts with distribution of exogenous HIF1 α and endogenous HIF2 α ChIP-seq binding sites in 786-O cells annotated using ChIPseeker. **Figure 5H** shows pie charts of ChIP-Seq data that shows distribution of exogenous HIF1 α and endogenous HIF2 α binding at altered promoters and enhancers in 786-O cells that have been genetically engineered to overexpress HIF1 α . **Figure 5I** shows pie charts of distribution of endogenous HIF1 α and HIF2 α ChIP-seq binding sites in 40911432 cells annotated using ChIPseeker. **Figure 5J** shows pie chart of ChIP-Seq that shows distribution of endogenous HIF1 α and HIF2 α binding at altered promoters and enhancers in 40911432 cells. **Figure 5K** refers to graph that shows transcription factor binding at *VHL*-responsive enhancers showing higher enrichment of HIF2 α than HIF1 α at enhancers with H3K27ac depletion after *VHL* restoration over regions with H3K27ac enrichment after *VHL* restoration. **Figure 5L** refers to plot of ChIP Seq with example of a *VHL*-responsive enhancer near *UBR4* with only HIF2 α binding but not HIF1 α binding. **Figure 5M** refers to plot of ChIP Seq with example of a *VHL*-responsive super-enhancer near *CMIP* with only HIF2 α binding but not HIF1 α binding.

[0019] **Figure 6** illustrates that HIF2 α -HIF1 β bound enhancers modulate gene expression. **Figure 6A** refers to dot plots with Pearson's correlation of gene expression changes after either *VHL* restoration or *HIF2* α siRNA knockdown at all genes or genes adjacent to HIF2 α binding sites.

Figure 6B refers to dot plots with Pearson's correlation of H3K27ac changes after *VHL* restoration and *HIF2α* siRNA knockdown at either all gained enhancers or HIF2α-bound enhancers adjacent to binding sites. **Figure 6C** refers to dot plots with Pearson's correlation of H3K27ac changes after *VHL* restoration and *HIF2α* siRNA knockdown at either all gained super-enhancers or HIF2α-bound super-enhancers adjacent to binding sites. **Figure 6D** refers to plots of histone chromatin immunoprecipitation sequencing (ChIP-Seq) profiles showing changes in RNAseq and H3K27ac ChIP-Seq signals after *VHL* restoration or *HIF2α* siRNA knockdown at *ZNF395* super-enhancer (SE), together with binding profiles of transcription factors enriched at enhancers. **Figure 6E** refers to graphs showing both *VHL* restoration and *HIF2α* siRNA knockdown decreases expression of genes with HIF2α-bound enhancers in 786-O cells. *p-value < 0.05, two-sided *t*-test. **Figure 6F** refers to graphs showing both *VHL* restoration and *HIF2α* siRNA knockdown decrease enhancer activities measured by luciferase reporter assay in 786-O cells. *p-value < 0.05, two-sided *t*-test. **Figure 6G** refers to a column graph showing RT-qPCR measurement of *ZNF395* expression in four wild-type clones and four clones with *ZNF395* enhancer depleted by CRISPR. Depleted region has the highest HIF2α binding at *ZNF395* super-enhancer in 786-O cells indicated by the back bar above the scissors (deleted region indicated in Fig 6D). *p-value < 0.05, two-sided *t*-test.

[0020] **Figure 7** shows that *VHL* restoration reduces p300 recruitment but preserves promoter-enhancer interactions. **Figure 7A** shows a line graph of the enrichment of p300 binding at gained and lost enhancers based on chromatin immunoprecipitation sequencing (ChIP-Seq). **Figure 7B** refers to a column graph showing the percentage of overlap between HIF2α and other transcription factors. **Figure 7C** refers to a heatmap of ChIP-Seq binding profiles of HIF2α and p300. **Figure 7D** shows an immunoblot of the protein expression of p300 with and without *VHL* in 786-O and 12364284 cells as measured by immunoblotting. **Figure 7E** shows a graph referring to ChIP-qPCR of p300 binding at enhancers with and without *VHL* restoration in 786-O cells. NC refers to negative control regions. **Figure 7F** refers to a graph showing the ChIP-qPCR of p300 binding at enhancers with and without *HIF2α* siRNA knockdown in 786-O cells. NC refers to negative control regions. **Figure 7G** shows scatter plots referring to correlation of enhancer interactions measured by Capture-C (RPM – reads per million) between 786-O cells with and without *VHL* restoration at both *VHL*-responsive and non-*VHL*-responsive enhancers. **Figure 7H** shows a plot referring to Capture-C that shows that *VEGFA* enhancer-promoter interactions are maintained even after *VHL* restoration. E refers to enhancer and P refers to promoter. **Figure 7I** shows a diagram with the schematics of *VHL*-driven enhancer aberration in clear cell renal cell carcinoma. **Figure 7J** refers to a plot with histone ChIP Seq profile of H3K27ac ChIP-seq and Capture C of 786-O (clear cell renal cell carcinoma cell line) and KATO III (gastric cancer cell line) showing that the SLC2A1 enhancer is specific to clear cell renal cell carcinoma.

[0021] **Figure 8 (Figures 8A to 8C)** shows graphs of analysis from exosomes obtained from A498 (clear cell renal cell carcinoma cell line) and HK-2 (normal kidney cell line). Each graph shows the level of the respective marker (as denoted by the title of each graph) that was detected based on exosomes secreted from the respective cell lines based on the graph legend. The y-axis of each graph shows gene expression of these markers as measured by quantitative polymerase chain reaction (qPCR).

[0022] **Figure 9** shows a heatmap with microarray data analysis from patient cohorts with clear cell renal cell carcinoma (ccRCC) or benign oncocytoma (B). *VEGFA*, *EGLN3*, *ZNF395*, *SMPDL3A*, *SLC6A3* and *SLC28A1* expression levels were compared between benign oncocytoma and clear cell renal carcinoma. A higher Z score indicates higher expression levels.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0023] Renal cell carcinoma (RCC) is the most common type of kidney cancer in adults, with clear cell renal cell carcinoma (ccRCC) being one of the common subtypes of renal cell carcinoma. Other subtypes of renal cell carcinoma include, for example, papillary renal cell carcinoma (pRCC) and chromophobe renal cell carcinoma (crRCC). One particular challenge in the treatment of kidney tumours is the range of histologies and tumour phenotypes that a renal mass can represent. A kidney tumour range from benign to clinically indolent malignancy to aggressive disease. Examples of aggressive disease include, but are not limited to, clear cell renal cell carcinoma (ccRCC). Kidney cancers of various subtypes have diverse treatment response rates and variable prognoses. Therefore, the key to proper treatment is accurate diagnosis of the different subtypes of renal cell carcinomas.

[0024] Current methods of diagnosis include, but are not limited to, invasive methods involving histologies from tumour biopsies, such as, staining of glycogens by special stains, contrast-enhanced computed tomography (CT) scan (which demonstrates high vascularity of the tumours), ultrasound imaging or magnetic resonance imaging (MRI). However, invasive procedures causes patient discomfort, require local or general anaesthesia and can be, relatively expensive. Histologies from biopsies and imaging methods including CT scan, MRI and ultrasound imaging, are also unable to detect tumour stages accurately, resulting in frequent misdiagnosis. Another option to identify clear cell renal cell carcinoma patients is through detection of biomarkers. However, there is currently no clinically validated biomarker for diagnosis of clear cell renal cell carcinoma.

[0025] In view of the above problems, the inventors of the present disclosure have provided biomarker(s) for identifying clear cell renal cell carcinoma. Accordingly, in one example, there is disclosed one or more clear cell renal cell carcinoma biomarkers.

- [0026] As used herein, the term “clear cell renal cell carcinoma” or “ccRCC” refers to the most common subtype of a kidney cancer, namely renal cell carcinoma (RCC). Kidney cancer refers to cancer that forms in tissues of the kidney, which is the organ that filters waste products from the blood. Kidney is also involved in regulating blood pressure, electrolyte balance and red blood cell production in the body. Each kidney is attached to a ureter, a tube that carries excreted urine to the bladder. Renal cell carcinoma is a kidney cancer that originates in the lining of the proximal convoluted tubule, a part of the very small tubes in the kidney that transport primary urine. Renal cell carcinoma is classified as an adenocarcinoma. The symptoms and implications accompanying renal cell carcinoma are well known in the art, for example hematuria (blood in the urine), low back pain or pain in the flank and / or noticeable lump in the flank. Clear cell renal cell carcinoma, which is the most common subtype of renal cell carcinoma, accounts for 75% to 85% of all renal cell carcinoma, and is also the most aggressive type of renal cell carcinoma. Clear cell renal cell carcinoma is associated with genetic lesions in chromosome 3p, encompassing the von Hippel-Lindau gene. On gross examination, clear cell renal cell carcinomas are typically golden yellow and often develop hemorrhage and infarction with formation of cysts within the tumour. Clear cell renal cell carcinoma is typically characterized by malignant epithelial cells with clear cytoplasm and a compact alveolar or acinar growth pattern interspersed with intricate, arborizing vasculature.
- [0027] As used herein, the term “tumour” refers to a group of abnormal cells that form lumps or growth. A tumour can be cancerous (malignant), non-cancerous (benign) or pre-cancerous. Benign tumour usually consists of angiomyolipoma and oncocytoma. Angiomyolipoma is easy to differentiate based on imaging, while being difficult to differentiate oncocytoma with imaging studies.
- [0028] As used herein, the term “carcinoma” refers to a type of cancer that starts in cells that make up the skin (also known as epithelial cells) of the tissue lining organs, for example, but not limited to, the liver or kidneys. Common types of carcinoma include, but are not limited to, basal cell carcinoma, squamous cell carcinoma and renal cell carcinoma. In one example, clear cell renal cell carcinoma refers to malignant tumours, while oncocytoma is or represents benign tumours. In another example, benign oncytomas are low in SLC28A1, VEGFA, ZNF395, EGLN3 and SLC6A3.
- [0029] When confronted with a renal mass, it is difficult to differentiate between malignant or benign tumour. Benign tumours usually consist of angiomyolipoma and oncocytoma. Differentiation of angiomyolipoma can be performed with imaging studies, as performed in the art. However, it is difficult to differentiate oncocytoma based on imaging studies alone. Thus, in one example, there is disclosed a method for determining whether a renal mass sample is benign or

malignant. In another example, the method of determining whether a renal mass sample is benign or malignant, the method comprising obtaining a sample from the renal mass of a subject; determining the levels or the presence or absence of the biomarker set as disclosed herein in the sample. In another example, the increase in levels of the biomarker set in such a renal mass sample, compared to a benign sample, indicates that the sample is malignant. In yet another example, the method of determining whether a renal mass sample is benign or malignant comprises obtaining a sample from the renal mass of a subject; determining the levels or the presence or absence of the biomarker set as disclosed herein in the sample; wherein the increase in levels of the biomarker set compared to a benign sample indicates that the sample is malignant.

- 5
- 10 [0030] As used herein, the term “biomarker” refers to molecular indicators of a specific biological property, a biochemical feature or facet that can be used to determine the presence or absence and/or severity of a particular disease or condition. In the present disclosure, the term “biomarker” refers to a polypeptide or a nucleic acid sequence encoding the polypeptide, a fragment or variant of a polypeptide being associated with clear cell renal cell carcinoma. In addition to a polypeptide or a nucleic acid sequence encoding the polypeptide, a fragment or variant of such a polypeptide being associated with clear cell renal cell carcinoma peptides as disclosed herein, the biomarker also refers to metabolites or metabolized fragments of the expressed polypeptide. A person skilled in the art would understand that a metabolite of one of the biomarkers referred to herein can still retain the capability of being used as biomarker for the methods described herein. It is also noted that some of the biomarkers in the biomarker set can be present in their variant form or metabolized form while others are still intact. The term “variant” as used herein includes a reference to substantially similar sequences. Generally, nucleic acid sequence variants of the invention encode a polypeptide which retains qualitative biological activity in common with the polypeptide encoded by the “non-variant” nucleic acid sequence.
- 20
- 25 Variants of said polypeptide include polypeptides that differ in their amino acid sequence due to the presence of conservative amino acid substitutions. For example, such variants have an amino acid sequence being at least 80%, at least 90%, at least 95%, at least 98%, or at least 99% identical over the entire sequence region to the amino acid sequences of the “non-variant” polypeptides. Variants can be allelic variants, splice variants or any other species-specific homologs, paralogs, or orthologs. In one example, the percentage of identity can be determined by known in the art algorithms. The sequence identity values recited above in percent (%) are to be determined, preferably, using programs known in the art, for example, BLASTp and the like. Variants can be made using, for example, the methods of protein engineering and site-directed mutagenesis as is well known in the art.
- 30

[0031] Conservative amino acid substitution tables providing functionally similar amino acids are well known to one of ordinary skill in the art. The following six groups are examples of amino acids that are considered to be conservative substitutions for one another: i) Alanine (A), Serine (S), Threonine (T); ii) Aspartic acid (D), Glutamic acid (E); iii) Asparagine (N), Glutamine (Q); iv) Arginine (R), Lysine (K); v) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and vi) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[0032] A non-conservative amino acid substitution can result from changes in: i) the structure of the amino acid backbone in the area of the substitution; ii) the charge or hydrophobicity of the amino acid; or iii) the bulk of an amino acid side chain. Substitutions generally expected to produce the greatest change in protein properties are those in which i) a hydrophilic residue is substituted for (or by) a hydrophobic residue ii) a proline is substituted for (or by) any other residue; iii) a residue having a bulky side chain, for example, phenylalanine, is substituted for (or by) one not having a side chain, for example, glycine; or iv) a residue having an electropositive side chain, for example, lysyl, arginyl, or histadyl, is substituted for (or by) an electronegative residue, for example, glutamyl or aspartyl.

[0033] As defined herein, the terms “peptide”, “protein”, “polypeptide”, and “amino acid sequence” are used interchangeably herein to refer to polymers of amino acid residues of any length. The polymer can be linear or branched, it can comprise modified amino acids or amino acid analogues, and it can be interrupted by chemical moieties other than amino acids. The terms also encompass an amino acid polymer that has been modified naturally or by intervention; for example disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification, such as conjugation with a labelling or bioactive component. The term peptide encompasses two or more naturally occurring or synthetic amino acids linked by a covalent bond (for example, an amide bond). The amino acid residues are joined together through amide bonds. When the amino acids are alpha-amino acids, either the L-optical isomer or the D-optical isomer can be used, the L-isomers being preferred in nature. The term polypeptide or protein as used herein encompasses any amino acid sequence and includes, but is not be limited to, modified sequences such as glycoproteins. The term polypeptide is specifically intended to cover naturally occurring proteins, as well as those that are recombinantly or synthetically produced.

[0034] In one example, the clear cell renal cell carcinoma protein biomarker set includes at least two selected from the group consisting of ZNF395, SMPDL3A, SLC28A1, SLC6A3, VEGFA, EGLN3, and variants thereof, wherein one of the at least two biomarkers is SMPDL3A or SLC28A1. In another example, the clear cell renal cell carcinoma (ccRCC) biomarker set comprises at least two biomarkers selected from the group consisting of ZNF395, SMPDL3A, SLC28A1, SLC6A3, VEGFA, and EGLN3. In another example, one of the at least two biomarkers

is SMPDL3A or SLC28A1. In yet another example, the biomarkers are proteins, or nucleic acids encoding the same, or variants thereof.

[0035] Also disclosed herein is a composition comprising the biomarker set as disclosed herein.

5 [0036] As used herein, the term “ZNF395”, also known as HDBP-2, HDRF-2, PBF or PRF-1, refers to both a gene and the expressed polypeptide thereof, both of which are associated with Huntington Disease. This gene is a hypoxia-inducible transcription factor that is controlled by I_κB signalling and activates genes involved in innate immune response and cancer. It has been found to be overexpressed in various human cancers, particularly in response to hypoxia. ZNF395 has also
10 been shown to play a role in papillomavirus gene transcription.

[0037] As used herein “SMPDL3A”, also known as sphingomyelin phosphodiesterase acid like 3A, or ASML3a, refers to a gene and the expressed polypeptide thereof, both of which has *in vitro* nucleotide phosphodiesterase activity with nucleoside triphosphates, such as for example, ATP. This protein has no activity with nucleoside diphosphates, and no activity with nucleoside
15 monophosphates. SMPDL3A has *in vitro* activity with CDP-choline and CDP-ethanolamine, with no spingomyelin phosphodiesterase activity. As mentioned in the experimental section below, SMPDL3A is a target of a master regulator of cholesterol metabolism.

[0038] As used herein, the term “SLC28A1”, also known as concentrative nucleoside transporter 1 (CNT1), HCNT1 or solute carrier family 28 member 1, refers to a gene and the
20 expressed polypeptide thereof, both of which is sodium dependent and pyrimidine selective. SLC28A1 exhibits transport characteristics of the nucleoside transport system cit or N2 subtype (N2/cit). SLC28A1 also transports the antiviral pyrimidine nucleoside 3'-azido-3'-deoxythymidine (AZT) and 2',3'-dideoxycytidine (ddC). SLC28A1 is involved in the uptake of nucleoside-derived drugs using antiviral and chemical therapies.

25 [0039] As used herein, the term “SLC6A3”, also known as solute carrier family 6 member 3, DAT, dopamine transporter, sodium dependent dopamine transporter, or PKDYS, refers to a gene or the expressed polypeptide thereof, both of which encodes a dopamine transporter, which is a member of the sodium and chloride dependent neurotransmitter transporter family. SLC6A3 terminates the action of dopamine by its high affinity sodium-dependent re-uptake into presynaptic
30 terminals. Variation of the number of repeats of this gene or the expressed polypeptide thereof is associated with idiopathic epilepsy, attention-deficit hyperactivity disorder, dependence on alcohol and cocaine, susceptibility to Parkinson disease and protection against nicotine dependence.

[0040] As used herein, the term “VEGFA”, also known as vascular endothelial growth factor A, vascular permeability factor, VEGF, VPF or MVCD1, refers to a gene or the expressed
35 polypeptide thereof that is a member of the PDGF/VEGF growth factor family. VEGFA encodes a

heparin-binding protein. It is a growth factor that induces proliferation and migration of vascular endothelial cells and is essential for both physiological and pathological angiogenesis. Disruption of this gene in mice resulted in abnormal embryonic blood vessel formation. This gene is up-regulated in many known tumours and its expression is correlated with tumour stage and progression. Variants of this gene has been reported, including, but not limited to, allelic variants associated with microvascular complications of diabetes 1 (MVCD1) and atherosclerosis, alternatively spliced transcript variants encoding different isoforms, alternative translation initiation from upstream non-AUG (CUG) codons (resulting in additional isoforms), and C-terminally extended isoforms produced by use of an alternative in-frame translation termination codon via a stop codon read-through mechanism (with this isoform being antiangiogenic).

[0041] As used herein, the term “EGLN3”, also known as Egl-9 family hypoxia inducible factor 3, prolyl hydroxylase domain-containing protein 3, hypoxia-inducible factor prolyl hydroxylase 3, HIF-prolyl hydroxylase 3, HIF-PH3, HPH-1, HPH-3, PHD3, HIFP4H3, Egl Nine-like protein 3 isoform, refers to a gene or the expressed polypeptide thereof associated with diseases including hypoxia and chronic mountain sickness. Related pathways of this protein include HIF-1 signaling pathway and HIF repressor pathways. EGLN3 functions as a cellular oxygen sensor that catalyzes the post-translational formation of 4-hydroxyproline in hypoxia-inducible factor (HIF) alpha proteins under normoxic conditions.

[0042] The clear cell renal cell carcinoma (ccRCC) biomarker set disclosed herein can be used in combination with two or more further biomarkers, wherein one of the at least two biomarkers is SMPDL3A or SLC28A1. Thus, in one example, the clear cell renal cell carcinoma protein biomarker set comprises any two, three, four, five or all six biomarkers. In another example, the biomarker set as disclosed herein comprises at least three biomarkers. In another example, the biomarker set as disclosed herein comprises at least four biomarkers. In another example, the biomarker set as disclosed herein comprises at least five biomarkers. In another example, the biomarker set as disclosed herein comprises at least six biomarkers.

[0043] In another example, the biomarker set or the biomarkers are, but are not limited to, ZNF395, SMPDL3A, SLC28A1, SLC6A3, VEGFA and EGLN3, and combinations thereof. In one example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A and SLC28A1. In another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A and ZNF395. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A and SLC6A3. In still another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A and VEGFA. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A and EGLN3. In one example, the

clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1 and ZNF395. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1 and SLC6A3. In still another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1 and VEGFA. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1 and EGLN3. In one example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPL3A, SLC28A1 and ZNF395. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPL3A, SLC28A1 and SLC6A3. In still another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPL3A, SLC28A1 and VEGFA. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPL3A, SLC28A1 and EGLN3. In one example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, ZNF395 and SLC6A3. In another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, ZNF395 and VEGFA. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, ZNF395 and EGLN3. In still another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, SLC6A3 and VEGFA. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, SLC6A3 and EGLN3. In another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, VEGFA and EGLN3. In still another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, ZNF395 and SLC6A3. In another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, ZNF395 and VEGFA. In another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, ZNF395 and EGLN3. In one example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, SLC6A3 and VEGFA. In another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, SLC6A3 and EGLN3. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, VEGFA and EGLN3. In one example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, SLC28A1, ZNF395 and SLC6A3. In another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, SLC28A1, ZNF395 and VEGFA. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, SLC28A1, ZNF395 and EGLN3. In still another example, the clear cell renal cell carcinoma biomarker set or biomarkers

comprise or consist of SMPDL3A, ZNF395, VEGFA and EGLN3. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, ZNF395, SLC6A3 and VEGFA. In another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, ZNF395, SLC6A3 and EGLN3. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, SLC6A3, VEGFA and EGLN3. In one example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, ZNF395, SLC6A3 and VEGFA. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, ZNF395, SLC6A3 and EGLN3. In still another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, ZNF395, SLC6A3 and VEGFA. In another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, SLC28A1, ZNF395, SLC6A3 and VEGFA. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, SLC28A1, ZNF395, SLC6A3 and EGLN3. In still another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SMPDL3A, ZNF395, SLC6A3, VEGFA and EGLN3. In a further example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, SLC6A3, VEGFA and EGLN3. In yet another example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of SLC28A1, ZNF395, SLC6A3, VEGFA and EGLN3. In one example, the clear cell renal cell carcinoma biomarker set or biomarkers comprise or consist of ZNF395, SMPDL3A, SLC28A1, SLC6A3, VEGFA and EGLN3.

[0044] The biomarkers of the present invention can be combined with one another, for example, as a biomarker set, thereby providing sensitive and specific determination of clear cell renal cell carcinoma subjects thought to suffer from clear cell renal cell carcinoma. The option of combining the biomarkers of the present disclosure provides a statistically reliable detection of clear cell renal cell carcinomas. In one example, the presence of the two or more biomarkers in a sample is indicative of the presence of clear cell renal cell carcinoma. In another example, the presence of the biomarker set as disclosed herein is indicative of the presence of clear cell renal cell carcinoma. In another example, the upregulation of the biomarker set in a sample is indicative of the presence of clear cell renal cell carcinoma. This can be seen from the data provided in Figure 3D, where statistical relevance based on a p-value is provided for a list of markers.

[0045] As used herein, the term “upregulation” refers to an increased level of expression of a nucleic acid or a protein in a sample obtained from a disease subject, whereby the increase is

compared to expression of the same nucleic acid or protein in a control sample. In one example, the subject is suffering from clear cell renal cell carcinoma. As disclosed herein, this upregulation can be depicted as, for example, but not limited to, high tumour-normal ratios, low p-values or high levels of mRNA expression. In one example, the expression levels of nucleic acids can be measured, for example, by polymerase chain reaction (PCR). In one example, the polymerase chain reaction (PCR) is reverse transcription polymerase chain reaction (RT-PCR), or real-time polymerase chain reaction (qPCR) or combinations thereof. It is noted that RT-PCR is used to qualitatively detect gene expression through the creation of complementary DNA (cDNA) transcripts from RNA, while qPCR is used to quantitatively measure the amplification of DNA using fluorescent dyes. qPCR is also referred to in the art as quantitative PCR, quantitative real-time PCR, and real-time quantitative PCR.

[0046] As illustrated in the experimental section, the biomarkers of the present disclosure in particular ZNF395, SMPL3A and SLC28A1, have been shown to be specific to clear cell renal cell carcinoma, and were shown to not be overexpressed in papillary and chromophobe renal cell carcinomas, two other distinct renal cell carcinoma subtypes (Figure 3D). ZNF395 exhibited tumour-normal ratio of about 7 in clear cell renal cell carcinoma with p-value of 1×10^{-22} , while showing little overexpression in papillary and chromophobe renal cell carcinomas with tumour-normal ratio of 1.2 ($p=0.02$) and 1.3 ($p=0.06$) respectively.

[0047] Furthermore, the experimental section, and for example, Figure 3, also show that among the 12 types of cancer profiled by The Cancer Genome Atlas (TCGA), ZNF395, SMPL3A, SLC28A1, SLC6A3, VEGFA and EGFA were exclusively overexpressed in clear cell renal cell carcinoma tumours (KIRC). ZNF395 depletion *in vivo* further validates that ZNF395 plays an important role in clear cell renal cell carcinoma tumourigenesis. ZNF395 depletion significantly slowed *in vivo* tumour growth of 786-O clear cell renal cell carcinoma cells. In addition, SMPDL3A and SLC28A1 are shown to be associated with clear cell renal cell carcinoma-specific super-enhancer.

[0048] As described in the experimental section of the present disclosure, analysis of exosomes in culture medium of cell cultures can be performed, for example, by measuring levels of gene expression. This data can be found, for example, in Figure 8. In one example, this can be done by qPCR, which was performed on clear cell renal cell carcinoma cell line (A498) and normal kidney cell lines (HK2). Results from exosome analysis in the present disclosure shows higher expression of ZNF395, SMPL3A, VEGFA and EGLN3 in clear cell renal cell carcinoma cell lines compared to normal kidney cell lines.

[0049] As used herein, the term “exosome” refers to a type of small extracellular vesicle (EV), ranging from 30 to 200 nm in diameter. These exosomes can be isolated from cell culture media, as

well as an array of eukaryotic fluids. These fluids include, but are not limited to, blood, urine and sputum samples. Therefore, exosomes can be used to identify biomarkers from liquid samples using non-invasive methods. Exosomes are either released from the cell when multi-vesicular bodies fuse with the plasma membrane, or are released directly from the plasma membrane. These vesicles carry nucleic acid (for example, RNA and DNA) and proteins from, for example, the tumour in tumour-bearing subjects. Exosomes have been implicated in driving malignant cell behaviour including, but not limited to stimulation of tumour cell growth and suppression of a host immune response. Therefore, exosomes are a viable source for identifying biomarkers for cancer. Isolating exosomes from specific tissues also allows identification of tissue-specific or disease-specific biomarkers. Briefly, in one example, exosome analysis can be performed by ultracentrifugation from cells. After tryptic digestion, proteomic analysis can be performed, and candidate biomarkers are validated by, for example, methods known in the art, including but not limited to, western blotting and immunohistochemistry. In one example, the exosomes are detected using the detection system and/or the methods disclosed herein. In another example, the exosomes are detected and/or analysed using quantitative polymerase chain reaction (qPCR).

[0050] The terms “isolated” or “isolating” as used herein relates to a biological component (such as a nucleic acid molecule, protein or organelle) that has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, i.e., other chromosomal and extra-chromosomal DNA and RNA, proteins and organelles. Nucleic acids that have been “isolated” include nucleic acids purified by standard purification methods.

[0051] As illustrated in the experimental section of the present disclosure, the inventors of the present disclosure found that the biomarkers of the present disclosure can be detected in various sample types as described herein. The term “sample”, as used herein, refers to single cells, multiple cells, fragments of cells, tissue, or body fluid, which has been obtained from, removed from, or isolated from a subject. An example of a sample includes, but is not limited to, blood, stool, serum, saliva, urine, sputum, cerebrospinal fluid, bone marrow fluid, frozen fresh tissue of a tumour sample, or frozen fresh tissue of a non-diseased tissue harvested from sites distant from the tumour. For example, the biomarkers were clearly detected in solid samples, which include, but are not limited to, solid tumour biopsy from suitable organs, such as the kidney. Fresh-frozen normal-tumour tissues were obtained from nephrectomy cases and normal tissues were harvested from sites distant from the tumour. Normal-tumour tissues as described herein or normal-tumour pair as described in the experimental section in other words means a tumour sample and a non-diseased sample that is obtained from the same subject. The sample can include, but is not limited to, tissue obtained from the lung, the muscle, brain, liver, skin, pancreas, stomach, bladder, and other organs.

In another example, the sample includes, but is not limited to, fluid samples derived from or comprising bodily fluids, such as whole blood, serum, plasma, tears, saliva, nasal fluid, sputum, gastrointestinal fluid, exudate, transudate, fluid harvested from a site of an immune response, fluid harvested from a pooled collection site, bronchial lavage, a nucleated cell sample, a fluid associated with a mucosal surface, hair, or skin, and urine. In one example, the fluid sample is liquid tumour biopsy, urine sample, blood sample, sputum sample or cell culture medium. In another example, the fluid sample contains exosomes suspected to comprise the biomarkers or biomarker set as disclosed herein. In one example, the detection of the biomarkers in urine sample, blood sample or sputum sample is desirable as it allows for non-invasive detection of clear cell renal cell carcinoma in subjects.

[0052] In another example, there is provided a detection system for detecting the biomarker as disclosed herein. In one example, the detection system of the present disclosure comprises a receiving section to receive a sample from a patient suspected to suffer from clear cell renal cell carcinoma. In another example, the sample is suspected to comprise the two or more biomarkers of the present disclosure. In yet another example, the detection system comprises a substance or substances capable of detecting the two or more biomarkers of the present disclosure. In a further example, one of the at least two biomarkers is SMPDL3A or SLC28A1. In one example, the detection system of the present disclosure comprises a) a receiving section to receive a sample from a patient suspected to suffer from clear cell renal cell carcinoma, and wherein the sample is suspected to comprise the two or more biomarkers of the present disclosure, and b) a detection section comprising a substance or substances capable of detecting the two or more biomarkers, or the biomarker set, of the present disclosure, wherein one of the at least two biomarkers is SMPDL3A or SLC28A1.

[0053] In one example, the detection system comprises a receiving section to receive a sample from a patient suspected to suffer from clear cell renal cell carcinoma, and wherein the sample is suspected to comprise two, three, four, five or all six biomarkers, or a biomarker set, of the present disclosure. In one example, the receiving section can be a biochip, test strip, a real time polymerase chain reaction (qPCR) apparatus or microtiter plate. In one example, the sample can be fluid samples, as described herein.

[0054] In one example, the detection system or method of the present disclosure can require a fluid sample volume, such as, but not limited to, a sample volume of between about 1 µl to about 30 ml, 1µl to 5µl, 4µl to 10µl, 9µl to 15µl, 14µl to 20µl, 19µl to 25µl, 24µl to 30µl, 29µl to 35µl, 34µl to 40µl, 39µl to 45µl, 44µl to 50µl, 49µl to 60µl, 59µl to 80µl, 79µl to 100µl, 99µl to 150µl, 149µl to 200µl, 199µl to 250µl, 249µl to 300µl, 299µl to 500µl, 499µl to 1ml, 999µl to 5ml, 4.99ml to 10ml, 9.99ml to 20ml and 19.99ml to 30ml. In one example, the fluid or sample volume can be

about 1 µl, about 5 µl, about 10 µl, about 15 µl, about 20 µl, about 25 µl, about 30 µl, about 35 µl, about 40 µl, about 45 µl, about 50 µl, about 100 µl, about 150 µl, about 200 µl, about 250 µl, about 300 µl, about 350 µl, about 400 µl, about 450 µl, about 500 µl, about 550 µl, about 600 µl, about 650 µl, about 700 µl, about 750 µl, about 800 µl, about 850 µl, about 900 µl, about 950 µl, about 1ml, about 2ml, about 3ml, about 4ml, about 5ml, about 6ml, about 7ml, about 8ml, about 9ml, about 10ml, about 11ml, about 12ml, about 13ml, about 14ml, about 15ml, about 16ml, about 17ml, about 18ml, about 19ml, about 20ml, about 21ml, about 22ml, about 23ml, about 24ml, about 25ml, about 26ml, about 27ml, about 28ml, about 29ml, to about 30ml, or any values there between.

5
10 **[0055]** To assist in detecting the biomarkers of the present disclosure, the detection system of the present disclosure can comprise a substance capable of binding or specifically binding to two, three, four, five or all six biomarkers of the present disclosure. In one example, the substance is a biospecific capture reagent, such as, but not limited to, antibodies (or antigen-binding fragments thereof), interacting fusion proteins, aptamers or affibodies (which are non-immunoglobulin-derived affinity proteins based on a three-helical bundle protein domain), all of which can be
15 chosen for their ability to recognize the biomarker and/or variants thereof. Antibodies can include, but are not limited to primary antibodies, secondary antibodies or horseradish peroxidase (HRP)-tagged secondary antibodies and the like. In one example, the substance includes antibodies known in the art to specifically recognise ZNF395, SMPDL3A, SLC28A1, SLC6A3, VEGFA or EGLN3.

20 **[0056]** In one example, the substance can be bound to a solid phase, wherein the biomarkers can then be detected by mass spectrometry, or by eluting the biomarkers from the biospecific capture reagents and detecting the eluted biomarkers by traditional matrix-assisted laser desorption/ionisation (MALDI) or by surface-enhanced laser desorption/ionization (SELDI). In another example, the detection system can be a biochip, test strip, qPCR apparatus, or microtiter
25 plate.

[0057] In another example, the detection system comprising a receiving section and a detection section can be configured to detect one, two, three, four, five or all six biomarkers, or the biomarker set, as described herein, individually, or in combination with one another. The detection system, as disclosed herein, can also be configured to detect the one or more biomarkers
30 simultaneously or in any given sequence (in other words, one at time).

[0058] In another example, there is disclosed a method of determining whether a subject has or shows recurrence of clear cell renal cell carcinoma (ccRCC). In one example, the method comprises obtaining a sample from a subject. In another example, the method comprises detecting the presence of the biomarker set of the present invention using a detection system of the present
35 invention, wherein the presence of the biomarker set determines that a subject has or shows

recurrence of clear cell renal cell carcinoma. In yet another example, the method comprises a) obtaining a sample from a subject; and b) detecting the presence of the biomarker set of the present invention using a detection system of the present invention, wherein the presence of the biomarker set determines that a subject has or shows recurrence of clear cell renal cell carcinoma.

5 [0059] In yet another example, the method comprises detecting the presence of the biomarker set of the present invention obtained from a sample from a subject, using a detection system of the present invention, wherein the presence of the biomarker set determines that a subject has or shows recurrence of clear cell renal cell carcinoma.

10 [0060] As used herein, the term “patient” or “subject” or “individual”, which can be used interchangeably, relates to animals, for example mammals, including but not limited to, cows, horses, non-human primates, dogs, cats and humans. The subject or the patient of the present disclosure can be suspected of suffering from, or have previously suffered from, clear cell renal cell carcinoma. In one example, the method of the present invention can be applied to a subject suspected of suffering from clear cell renal cell carcinoma. In another example, the method of the present disclosure can be applied to a subject suspected of having a recurrence of clear cell renal cell carcinoma. The term “recurrence” as used herein refers to the return of or re-detection of clear cell renal cell carcinoma in a patient who has been deemed to be free of renal carcinomas or, specifically, free of clear cell renal carcinoma.

15 [0061] The biomarkers, or the biomarker set, disclosed herein can be detected in samples using methods known in the art. It is appreciated that the person skilled in the art would understand which assays known in the art would be suitable in detecting the biomarkers of the present disclosure. For example, detection of the biomarkers of the present disclosure relates to the observance of presence or absence of the biomarkers.

20 [0062] Detection can be done directly or indirectly. Direct detection relates to detection of the polypeptide based on a signal which is obtained from the polypeptide itself and the intensity of which directly correlates with the number of molecules of the polypeptide present in the sample. Such a signal - sometimes referred to as intensity signal, can be obtained, for example, by measuring an intensity value of a specific physical or chemical property of the polypeptide. Indirect measuring includes measuring of a signal obtained from a secondary component (i.e. a component not being the polypeptide itself) or a biological read out system, for example, measurable cellular responses, ligands, labels, or enzymatic reaction products. The concept outlined above can also be applied to genes, whereby the determination of the level of the gene can be determined by measuring gene expression, either global or targeted gene expression, using methods known in the art. For example, the detection can be carried out using molecular biological methods.

25

30

35

[0063] The molecular biological methods can include, but are not limited to, polymerase chain reaction (PCR), such as reverse transcription polymerase chain reaction (RT-PCR), or real-time polymerase chain reaction (qPCR, also known as quantitative PCR); Western Blot, Dot Blot; mass spectrometry; nucleic acid sequencing; immunological methods, such as enzyme-linked immunosorbent assay (ELISA) using antibodies; and the like. For example, in the experimental section of the present disclosure, Western Blot, real-time quantitative PCR (qPCR) and nucleic acid sequencing are used.

[0064] In one example, the detection system detects exosomes using quantitative polymerase chain reaction.

10 [0065] In one example, the indication as to whether the two or more biomarkers, or the biomarker set as disclosed herein, are present in a sample obtained from the patient or whether the subject has or shows recurrence of clear cell renal cell carcinoma can be made based on comparison of the two or more biomarkers, or the biomarker set as disclosed herein, with the same biomarkers in a control group. As use herein, a control group includes disease-free subjects and/or
15 samples from non-diseased areas of the same or different subjects suffering from clear cell renal cell carcinoma. Control samples from the same or different subjects can also be known as matched or unmatched pairs, respectively. The control group can also be a non-cancerous sample obtained from a different subject with clear cell renal cell carcinoma; a subject that has a different renal cell carcinoma subtype; a subject with another type of cancer; or a sample obtained from non-
20 cancerous kidney cell lines. A non-cancerous kidney cell line is, but is not limited to, HK-2 and PCS-400 cell lines. In one example, the clear cell renal cell carcinoma (ccRCC) sample and control sample are obtained from the same subject (normal-tumour pair or tumour-normal pair, as described herein). Therefore, the method also includes differentiation of clear cell renal cell carcinoma from other types of renal cell carcinoma or from another type of cancer.

25 [0066] In one example, there is disclosed a kit comprising a detection system as described herein, and substances needed to carry out the method as described herein. In one example, the kit comprises a detection buffer, a lysis buffer, and substance or substances as described herein.

[0067] In one example, the biomarkers, methods, detection system or kit of the present disclosure are used to identify clear cell renal cell carcinoma in patients. The biomarkers, methods,
30 detection system or kit of the present disclosure can be used for detecting or predicting recurrence in clear cell renal cell carcinoma patients who may or may not be undergoing treatment or had received treatment for clear cell renal cell carcinoma.

[0068] In another example, there is disclosed a method of treating clear cell renal cell carcinoma in a subject, wherein the method comprises detecting the biomarker set as described

herein, and treating the subject determined to suffer from clear cell renal cell carcinoma with an anti-clear cell renal cell carcinoma compound and/or treatment.

[0069] It will be appreciated that the biomarker set as disclosed herein can be used to detect whether a treatment being performed or which had been performed on a subject was successful or not. This is because there is a difference in biomarker expression level and/or presence or absence of the biomarker in diseased tissue when compared to non-diseased tissue. Thus, in one example, there is a method of detecting response of a subject to systemic treatment, the method comprising obtaining a sample from the subject; and determining the levels of the biomarker set as defined herein, wherein a decrease in levels or an absence of the biomarker set indicates that the subject is responsive to treatment.

[0070] Also disclosed herein is a method for detecting susceptibility of a subject to an anti-clear cell renal cell carcinoma treatment. This method comprises determining the response of a sample from a diseases subject when subjected to one or more anti-clear cell renal cell carcinoma treatments based on the expression and/or presence or absence of the biomarker set as disclosed herein. In one example, the anti-clear cell renal cell carcinoma treatment comprises anti-cancer treatment, antibodies and the like.

[0071] The data shown herein examines somatically altered super-enhancers, which enabled the identification of a master regulator thought to play a key role in the pathogenesis of clear cell renal cell carcinoma, *ZNF395*. This disclosure describes specific von Hippel-Lindau-dependent enhancer required for *ZNF395* expression and shows the role of *ZNF395* in clear cell renal cell carcinoma tumourigenesis *in vitro* and *in vivo*.

[0072] Epigenetic maps of this study reveal targets that contribute to clear cell renal cell carcinoma tumourigenesis. Extensive enhancer gains were found around well-characterized hypoxia-related targets (*VEGFA*, *CXCR4*, *HK2*), SLC-mediated membrane transporters (*SLC2A1*, *SLC2A2*, *SLC38A1*), *SLC16A* family, and adipogenesis (*PLIN2*). Targets revealed in this study include *SMPDL3A*, *SLC28A1*, *SLC6A3*, *VEGFA* and *EGLN3*. *SMPDL3A* is a clear cell renal cell carcinoma-specific oncogene with a role in lipid and cholesterol metabolism. One finding from this epigenomic study is the tumourigenic requirement of *ZNF395* in clear cell renal cell carcinoma. *ZNF395* is also known as HDBP2 or papillomavirus binding factor (PBF). *ZNF395* is required for the differentiation of mesenchymal stem cells to adipocytes, by partnering with PPAR γ 2 to promote adipogenesis. *ZNF395* has been shown to bind to the promoters of Huntington gene and interferon-induced genes, and to cause upregulation of cancer-related genes (*MACC1*, *PEG10*, *CALCOCO1*, and *MEF2C*) and proangiogenic chemokines including *IL6* and *IL8* under hypoxia.

[0073] The invention illustratively described herein can suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for

example, the terms "comprising", "including", "containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed can be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

5
10 **[0074]** As used in this application, the singular form "a," "an," and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a genetic marker" includes a plurality of genetic markers, including mixtures and combinations thereof.

[0075] As used herein, the term "about", in the context of concentrations of components of the formulations, typically means +/- 5% of the stated value, more typically +/- 4% of the stated value, more typically +/- 3% of the stated value, more typically, +/- 2% of the stated value, even more typically +/- 1% of the stated value, and even more typically +/- 0.5% of the stated value.

[0076] Throughout this disclosure, certain embodiments can be disclosed in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the disclosed ranges. Accordingly, the description of a range should be considered to have specifically disclosed all the possible sub-ranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed sub-ranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

20
25
30 **[0077]** Certain embodiments can also be described broadly and generically herein. Each of the narrower species and sub-generic groupings falling within the generic disclosure also form part of the disclosure. This includes the generic description of the embodiments with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[0078] The invention has been described broadly and generically herein. Each of the narrower species and sub-generic groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[0079] Other embodiments are within the following claims and non-limiting examples. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

5

EXPERIMENTAL SECTION

[0080] The following examples illustrate methods by which aspects of the invention may be practiced or materials that may be prepared which is suitable for the practice of certain embodiments of the invention.

10 Example 1 – Materials and Methods

Patient Information

[0081] Fresh-frozen normal-tumour tissues were obtained from nephrectomy cases under approvals from institutional research ethics review committees and patient consent. Normal tissues were harvested from sites distant from the tumour. Table 1 refers to detailed patient information of this study.

15

ID	Age at the time of diagnosis	Race	Sex	Histo Type	Fuhrman Nuclear Grade
74575859	80	Chinese	Female	Clear cell carcinoma	Grade II
17621953	48	Chinese	Male	Clear cell carcinoma	Grade II
57398667	56	Chinese	Male	Clear cell carcinoma	Grade III
70528835	52	Chinese	Male	Clear cell carcinoma	Grade III
77972083	50	Chinese	Male	Clear cell carcinoma	Grade IV
75416923	49	Unknown	Male	Clear cell carcinoma	Grade II
20431713	65	Indian	Male	Clear cell carcinoma with papillary features	Grade IV
40911432	56	Chinese	Female	Clear cell carcinoma	Grade II
12364284	Unknown	Unknown	Unknown	Clear cell carcinoma	Grade III
86049102	51	Chinese	Male	Clear cell carcinoma	Grade II

Table 1: Patient information

Cell Lines

[0082] Commercial cell lines (786-O, A-498, HK2, and PCS-400) were purchased from ATCC. Cell lines were maintained in RPMI (Invitrogen) with 10% FBS with the exception of primary renal proximal tubule epithelial cells, PCS-400, which were maintained in Renal Epithelial Cell Basal Medium (ATCC). Cell line authentication was performed by short tandem repeat (STR) analysis against publicly available STR profiles. *Mycoplasma* testing was performed using the MycoSensor PCR assay kit (Stratagene).

Establishment of Tumour-Derived Cell Lines from Primary Tumours

[0083] Tumour cells were disassociated from primary tumours by collagenase, seeded, and maintained in RPMI with 10% FBS. At 80% to 90% confluency, the cells were passaged at a ratio of 1:3. Cultured cells were considered to be successfully immortalized after 60 passages. Correct pairing of tumour tissues and cell lines was achieved by comparing the percentage identity of single nucleotide polymorphisms (SNP) based on targeted sequencing. All tumour–cell line pairs showed identities of > 90% whereas shuffling of pairing showed identities < 80%. Tumours and cell lines from 12364284 and 40911432 showed the same von Hippel-Lindau (VHL) mutations, but tissue from 86049102 (86049102T) is *VHL*-mutant, whereas the cognate 86049102 cell line (86049102L) is *VHL*-wild-type.

Stable von Hippel-Lindau Restoration in clear cell renal cell carcinoma lines

[0084] 786-O cells (WT2, *VHL*+) and 786-O cells (RC3, *VHL*–) were used. Stable transduction of von Hippel-Lindau (*VHL*) was performed in A-498, 12364284, and 40911432 cells as follows: HA-*VHL* wt-pBabe-puro plasmid was transfected into PlatA cells (RV-102, Cell Biolabs) at 2 µg DNA/well of a 6-well plate using Lipofectamine 3000 (LifeTechnologies). A medium change was performed 10 to 16 hours after transfection. The supernatant from PlatA cells containing retroviruses was harvested 48 hours later and added to clear cell renal cell carcinoma cells, which were then selected with puromycin for 3 days after transduction.

Histone Nano-chromatin immunoprecipitation sequencing (Nano-ChIP-seq)

[0085] Nano-ChIP-seq was performed as previously described with slight modifications. Fresh-frozen cancer and normal tissues were dissected using a razor blade to obtain about 5 mg of tissue. The tissues were fixed in 1% formaldehyde for 10 minutes at room temperature. Fixation was stopped by addition of glycine to a final concentration of 125 nmol/L. Tissue pieces were washed three times with TBSE buffer. Pulverized tissues were lysed in 100 µL lysis buffer and sonicated for 16 cycles (30s on, 30s off) using a Bioruptor (Diagenode). The following antibodies were used: H3K27ac (ab4729, Abcam), H3K4me3 (07-473, Millipore), H3K4me1 (ab8895, Abcam), and H3K27me3 (07-449, Millipore). The total volume of immunoprecipitation was 1 mL and the amount of antibody used was 2 µg. The input DNA was precleared with protein G

Dynabeads (Life Technologies) for 1 hour at 4°C and then incubated with antibodies conjugated protein G beads overnight at 4°C. The beads were washed 3 times with cold wash buffer. After recovery of chromatin immunoprecipitation (ChIP) and input DNA, whole-genome amplification was performed using the WGA4 kit (Sigma-Aldrich) and BpmI-WGA primers. Amplified DNA was digested with BpmI [New England Biolabs (NEB)]. After that, 30 ng of the amplified DNA was used with the NEBNext ChIP-seq library prep reagent set (NEB). Chromatin immunoprecipitation sequencing (ChIP-seq) in cell lines was performed using the same Nano-ChIP-seq protocol described above but with 1×10^6 cells. Each library was sequenced to an average depth of 20 to 30 million raw reads on HiSeq2500 using 101-bp single end reads.

10 **Histone chromatin immunoprecipitation sequencing (ChIP-seq) Analysis**

[0086] Sequencing tags were mapped against the human reference genome (hg19) using Burrows-Wheeler Aligner (BWA-mem; version 0.7.10). Reads were trimmed 10 bp from the front and the back to produce 81 bp. Only reads with mapQ >10 and with duplicates removed by rmdup were used for subsequent analysis. Significant peaks were called using CCAT ($P < 0.05$). The strength and quality of immunoprecipitation were assessed using CHANCE.

15 **Transcription Factor chromatin immunoprecipitation sequencing (ChIP-seq)**

[0087] For each transcription factor, 3×10^7 cells were cross-linked with 1% formaldehyde for 10 minutes at room temperature and stopped by adding glycine to a final concentration of 125 nmol/L. Chromatin was extracted and sonicated to about 500 bp (Vibra cell, SONICS). The following antibodies were used for chromatin immunoprecipitation: c-Jun (sc-1694, Santa Cruz Biotechnology), NF- κ B p65 (sc-372, Santa Cruz Biotechnology), ETS1 (sc-350, Santa Cruz Biotechnology), HIF1 α (610959, BD Biosciences), HIF2 α (NB100-122, Novus Bio), and p300 (sc-585, Santa Cruz Biotechnology). The total volume of immunoprecipitation was 1.5 mL and the amount of antibody used was 15 μ g. Input DNAs were precleared with protein G Dynabeads (Life Technologies) for 2 hours at 4°C and then incubated with antibody-conjugated protein G beads overnight at 4°C. The beads were washed 6 times with wash buffer at room temperature. At least 10 ng of the DNA was used with the NEBNext ChIP-seq library prep reagent set (NEB). Each library was sequenced to an average depth of 30 to 50 million reads on a HiSeq2500 using 101-bp single-end reads.

30 **Transcription factor chromatin immunoprecipitation sequencing (ChIP-Seq) analysis**

[0088] Sequencing tags were mapped against the human reference genome (hg19) using Burrows-Wheeler Aligner (BWA-mem) (version 0.7.10). Only reads with mapQ >10 and with duplicates removed by rmdup were used in the subsequent analysis. Significant peaks were called using MACS2 (q -value < 0.01). Fastq files of HIF2 α ChIP-seq (GSM856790), HIF1 β ChIP-seq (GSM856790) and HIF1 α ChIP-Seq (GSM1642764) were downloaded from GEO database. Peaks

were called using MACS2 using the same settings as above.

RNA-Seq

[0089] Ten pairs of normal-tumour tissue matching the chromatin immunoprecipitation sequencing (ChIP-seq) tissues were prepared for RNA-seq. Total RNA was extracted using the
 5 Qiagen RNeasy Mini kit. RNAseq libraries were prepared using the Illumina Tru-Seq RNA Sample Preparation v2 protocol, according to the manufacturer's instructions. Briefly, poly-A RNAs were recovered from 1 µg of input total RNA using poly-T oligo conjugated magnetic beads. The recovered poly-A RNA was chemically fragmented and converted to SuperScript II and random primers. The second strand was synthesized using the Second Strand Master Mix.
 10 Libraries were validated with an Agilent Bioanalyzer (Agilent Technologies, Palo Alto, CA), diluted to 11 pM and applied to an Illumina flow cell using the Illumina Cluster Station. Sequencing was performed on a HiSeq2000 with 74bp or 76 base pair paired-end reads.

RNA-Seq analysis

[0090] RNA-seq reads were aligned to the human genome (hg19) using TopHat2-2.0.12
 15 (default parameter and --library-type fr-firststrand). Only uniquely mapped reads were analysed. Gene counts were obtained using HTSeq against the GENCODE v19 reference gene models and subsequent differential analysis was performed using DESeq2.

Capture-C

[0091] Capture-C was performed as previously described. Briefly, 1×10^7 cells were cross-
 20 linked by 2% formaldehyde, followed by lysis, homogenization, DpnII digestion, ligation, and de-cross-linking. DNA was sonicated using a Covaris to 150 to 200 bp to produce DNA suitable for oligo capture. A total of 3 µg of sheared DNA was used for sequencing library preparation (NEB). Enhancer sequences were double captured by hybridization to customized biotinylated oligos (IDT) and enriched with Dynabeads (Life Technologies). Captured DNA was sequenced to an
 25 average depth of 2 million reads per probe on the HiSeq Illumina platform using 150-bp paired-end reads.

Capture-C Analysis and Gene Assignment

[0092] Preprocessing of raw reads was performed to remove adaptor sequences (trim_galore), and overlapping reads were merged using FLASH. In order to achieve short read mapping to the
 30 hg19 reference genome, the resulting preprocessed reads were then *in silico* digested with DpnII and aligned using Bowtie (using p1, m2, best, and strata settings). Aligned reads were processed using Capture-C analyzer to (i) remove PCR duplicates; (ii) classify subfragments as "capture" if they were contained within the capture fragment, "proximity exclusion" if they were within 1 Kb on either side of the capture fragment, or "reporter" if they were outside of the "capture" and
 35 "proximity exclusion" regions; and (iii) normalize read counts per 100,000 interactions in bigwig

format. r3Cseq package was used on the capture and reporter fragments to identify significant interactions of the viewpoint against a scaled background (q -value < 0.05). Gene assignment is defined by the overlap of significant Capture-C peaks with genes with start and end defined by GENCODE v19. Interactions were plotted using Epigenome Gateway v40.0.

5 Identification of Differentially Enriched Regions

[0093] Significant H3K27ac peaks called by CCAT were merged across all normal–tumour samples. The same was performed with H3K4me1 and H3K4me3 chromatin immunoprecipitation sequencing (ChIP-seq) data. Transcription start sites (TSS) were based on GENCODE v19. Promoters were defined as regions of overlap between H3K27ac and H3K4me3 and also
 10 overlapping with ± 2.0 Kb around the TSS. Enhancers were defined as regions of overlap between H3K27ac and H3K4me1 but not overlapping with promoters. To minimize stromal contamination, we performed further filtering using cell line data, where enhancers and promoters not overlapping with H3K27ac peaks in any of the cell lines were discarded. Wiggle files of window size 50 bp were generated using MEDIPs from bam files. The input subtracted signal for each promoter or
 15 enhancer region was computed using bigWigAverageOverBed to yield reads per kilobase per million (RPKM). The RPKM of H3K27ac, H3K4me1, and H3K4me3 chromatin immunoprecipitation sequencing (ChIP-seq) from promoters and enhancers were corrected for batch effects using Combat. Tumour-specific regions were defined as regions that have a fold difference of ≥ 2 and a difference of 0.5 RPKM from patientmatched normal tissue. Normal regions
 20 were defined as regions that have a fold difference of ≤ 0.5 , and a difference of -0.5 RPKM from the corresponding regions in patient-matched tumours. Recurrently gained regions were defined as gain in $\geq 5/10$ patients and no loss in any patients. Recurrently lost regions were defined as loss in $\geq 5/10$ patients and no gain in any patients. Statistical testing for each *cis* regulatory region was performed using paired t tests with Benjamini–Hochberg correction. The differential regions were
 25 visualized using NGSplot.

Identification of Superenhancer Regions

[0094] Superenhancer regions were identified using ROSE (with promoter excluded), using H3K27ac peak regions merged from all patients (both normal and tumour tissue). Wiggle files of window size 50 bp were generated using MEDIPs from bam files. The input-subtracted signal for
 30 each superenhancer was computed using bigWigAverageOverBed (sum of reads over covered bases). The superenhancer regions were ranked by the average difference of normal–tumour H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) signals. Gained superenhancers were defined as regions that have average differential H3K27ac ChIPseq signals > 0 . Lost superenhancers were defined as regions that have average differential H3K27ac ChIP-seq
 35 signals < 0 .

Targeted sequencing

[0095] Ten pairs of normal-tumour tissue matching the chromatin immunoprecipitation sequencing (ChIP-seq) tissues were prepared for targeted mutation sequencing. Genomic DNA was extracted using the QIAamp DNA Mini Kit. Genomic DNA libraries were prepared using KAPA
 5 Hyper Prep Kit, according to the manufacturer's instructions. Briefly, genomic DNA was fragmented to 150–200 bp by sonication using a Covaris E-220 Focused Ultrasonicator (Duty Factor: 10%, Cycles per Burst: 200, Treatment Time: 360; Covaris Inc.). After the fragmentation process, end-repair, A-tailing, adapter ligation, and PCR reactions before target enrichment was performed, following the manufacturer's recommended protocols. After each step, the purification
 10 step was performed with AMPure XP beads to remove short fragments such as adapter dimers. Enrichment was performed using SureSelect XT2 Xplora RNA Bait (Custom, 5.9 Mb). Sequencing was performed on a Hiseq2500 with the paired-end 100bp option.

Principal component analysis (PCA)

[0096] RPKM values of H3K27ac intensities of all the *cis*-regulatory elements were first
 15 corrected for batch effects using COMBAT. PCA was performed on the entire 17,497 promoters or entire 66,448 enhancers. Variances and the cumulative proportion of each principal component were computed using R.

Saturation analyses

[0097] Saturation analyses were performed independently for enhancers and promoters.
 20 Specifically, subsets of the H3K27ac profiles from 20 primary samples (consisting of 10 primary tumours and matched normal samples) were selected. All combinations in each subset size were tested except those subsets with >10,000 possible combinations (n=5-15 samples), in which case 10,000 randomly selected combinations were tested. Then, H3K27ac enriched regions from each subset were combined, and overlapping regions were merged. These unique regions were then
 25 further classified as promoters and enhancers using the definitions reported in "Identification of differentially enriched regions".

GREAT analysis

[0098] Altered promoters were assigned using GREAT v3.0 by the nearest single gene. Altered enhancers were assigned to the genes with a proximal 5.0 Kb upstream, 1.0 Kb
 30 downstream extension and a distal extension up to 1000 Kb using default GREAT settings. The top pathways enriched in the MSigDB Pathways and Gene Ontology (GO) Molecular Functions were ranked by their hypergeometric *q*-values.

Epigenome Roadmap datasets

[0099] The bed files from H3K27ac, H3K4me1 and H3K4me3 chromatin immunoprecipitation
 35 sequencing (ChIP-Seq) of two normal kidneys were generated by the Epigenome Road. Peaks were

identified using CCAT. Similarities between the Epigenome Roadmap and our ChIP-Seq data were computed by the percentage of overlap between peaks.

DNA methylation analysis

[00100] In total, 160 tumour-normal matched pairs were obtained from The Cancer Genome Atlas (TCGA) database. Quantile normalization was performed across all the samples. Probes were
5 assigned to the nearest promoter or enhancer with a maximal cutoff of 10kb.

Chromatin accessibility analysis

[00101] Bigwig-formatted files of 7 clear cell renal cell carcinoma matched normal-tumour FAIRE-Seq datasets obtained from EMBL-EBI ArrayExpress under accession number E-MTAB-
10 1936. FAIRE-Seq signals for each promoter or enhancer region were computed using bigWigAverageOverBed with the promoters and enhancer regions as the input bed file. FAIRE-Seq data was normalized for batch effects using Combat.

lncRNA analysis

[00102] A list of differentially expressed lncRNA in kidney cancer was downloaded from a
15 previous study. RPKM values of each lncRNA were computed across the same ten pairs of normal-matched tissue where chromatin immunoprecipitation sequencing (ChIPseq) was performed, using bigWigAverageOverBed with chromosome positions defined by a previous study. These differentially expressed lncRNA were assigned to the nearest promoter and enhancer but with a maximum distance cut off of 10Kb. In total, around 200 lncRNAs were assigned to a promoter or
20 an enhancer.

Motif analysis

[00103] Motif analysis was performed using HOMER using the gained promoters and enhancers as the input regions and lost promoters and enhancers as the background. The input regions covered the entire span of promoters and enhancers. For von Hippel-Lindau (*VHL*)-
25 responsive regions, input regions were gained enhancers with H3K27ac depletion after *VHL* restoration and background regions were gained enhancers with H3K27ac enrichment after *VHL* restoration. Only known motifs were considered.

Histone chromatin immunoprecipitation sequencing (ChIP-seq) with von Hippel-Lindau restoration

[00104] H3K27ac, H3K4me1 and H3K27me3 ChIP-seq were performed using histone ChIP-seq. Sonicated DNA was normalized for each pair of cells with and without wild-type von Hippel-Lindau (*VHL*) before immuno-precipitation. Differential analysis of H3K27ac was performed using Deseq2 using raw counts of H3K27ac ChIP-seq with p -value < 0.05.

The Cancer Genome Atlas (TCGA) RNA-Seq

[00105] Preprocessed RNA-seq v2 data level 3 of clear cell renal cell carcinoma, papillary and
35

chromophobe renal cell carcinoma was downloaded from TCGA. Only patients with matched normal-tumour pairs (72 clear cell renal cell carcinoma pairs, 32 papillary renal cell carcinoma pairs and 25 chromophobe renal cell carcinoma pairs) were considered. The overall tumour-normal ratio of a given gene was computed from averaging individual tumour-normal ratios, and *p*-values computed by paired *t*-test. Pan-cancer compilation of TCGA data was obtained from pancan12.

Immunoblotting

[00106] Cell lines were harvested with cold RIPA lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 0.1% Triton X-100, 0.5% Sodium deoxycholate, 0.1% SDS) with protease inhibitors (Roche) on ice. Cells were mechanically lysed by passing through a 25 Gauge needle and centrifuged at 13,000 rpm for 15 min at 4°C. Protein concentrations were measured by the Pierce BCA protein assay (Life Technologies). Cell lysates were heated at 70°C for 10 min in sample buffer. Per well, 15 µg of cell lysate was loaded and gel electrophoresis was run at 130V constant for 90 minutes. Proteins were transferred to nitrocellulose membranes by transferring at 100 V for 100 minutes in ice. Western blotting was performed by incubating membranes overnight at 4°C with the following antibodies and dilutions: ZNF395 (1 µg/ml), von Hippel-Lindau (VHL) (1:250 dilution, Cell Signaling 2738), HIF1A (1:500 dilution, BD #610959), HIF1B (1:2000 dilution, Novus Bio NB100-110), HIF2A (1:1000 dilution, Novus Bio NB100-122), ETS1 (1:1000 dilution, Santa Cruz sc-350), c- Fos (1:500 dilution, Santa Cruz sc-7202), c-Jun (1:500 dilution, Santa Cruz sc-1694), NFκB p65 (ab7970, AbCAM) and β-actin (1:2000, Santa Cruz sc-47779). Membranes were incubated in secondary antibodies at 1:10,000 dilution for 1 hr at room temperature and developed with SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific).

siRNA knockdown

[00107] ON-TARGETplus SMARTpool siRNA (Dharmacon, UK) were used with Non-Targeting Control Pool as negative control and GAPDH Control Pool as positive control. The sequences of the SMARTpool siRNAs were as follows:

HIF2α (*EPAS1*) (SEQ ID NO: 1 GGCAGCACCUCACAUUUGA, SEQ ID NO: 2 GAGCGCAA AUGUACCCAAU, SEQ ID NO: 3 GACAAGGUCUGCAAAGGGU, SEQ ID NO: 4 GCAAAGACAUGUCCACAGA)

30 *SMDPL3A* (SEQ ID NO: 5 CAGUAUGAUCCUCGUGAUU, SEQ ID NO: 6 GAAGAUUUGCAGCCGGA, SEQ ID NO: 7 GACAGUAAGCAGUUUAUAA, SEQ ID NO: 8 CGGCCCAAUAUAAUGACA)

ZNF395 (SEQ ID NO: 9 CCAAACUGAUCAUGGCUUU, SEQ ID NO: 10 UCAGGCAGAUCAUGCAUAC, SEQ ID NO: 11 GUUCUGCGCUCCAUGUGG, SEQ ID NO: 12 GGACGAACCAGCUCCACGA)

35

[00108] A-498, 786-O and 12364284 and cells were trypsinized and diluted to appropriate concentrations. Lipofectamine RNAiMAX (Life Technologies) and SMARTpool siRNAs were diluted in Opti-MEM to a final siRNA concentration of 50 nM. The diluted Lipofectamine RNAiMAX was added to the diluted siRNA and incubated for 15 min at room temperature to allow complex formation to occur. The siRNA mixtures were aliquoted to wells in a 6-well plate. 48 hours after transfection, cells were re-seeded into 6-well plates for colony formation assays and 96-well plates for cell viability assay.

shRNA knockdown

[00109] Lentiviral plasmids were transfected into HEK293T cells. MISSION shRNA clones against *ZNF395* were purchased from Sigma Aldrich. The sequences of the clones are as follows:

TRCN0000233231 SEQ ID NO: 13

CCGGGCATCAAACGACACGTCAAAGCTCGAGCTTTGACGTGTCGTTTGATGCTTTTTG

TRCN0000233234 SEQ ID NO: 14

CCGGCAGAAGCCTTTACTGATTAACTCGAGTTTAATCAGTAAAGGCTTCTGTTTTTG

[00110] Cells were transduced with lentiviral particles for 48 hours and selected with puromycin (2 µg/ml) for four days before being analyzed for gene and protein expression and other functional assays.

Quantitative PCR Analysis (qPCR)

[00111] Total RNA was extracted from cell lines using Trizol (ThermoFisher) and purified with the RNeasy Mini Kit (Qiagen). Reverse transcription was performed using iScript Reverse Transcription Supermix for RT-qPCR (Biorad). qPCR was performed using Taqman probes (*ZNF395* Assay ID: Hs00608626_m1, *SMPDL3A* Assay ID: Hs00378308_m1) with TaqMan Gene Expression Master Mix (ThermoFisher). Gene expression changes were normalized to *GAPDH* (Assay ID: Hs00699446_m1).

Chromatin immunoprecipitation quantitative polymerase chain reaction (ChIP-qPCR)

[00112] ChIP DNA was probed with the following primers using the SYBR qPCR master mix (ThermoFisher).

ZNF395-E1 (hg19 chr8: 28221378-28221459)

ZNF395-E1-F: GCAACCTTCCAGGCCTGCCG (SEQ ID NO: 15)

ZNF395-E1-R: AGGAGAAAGGGGACAGGAGGGC (SEQ ID NO: 16)

ZNF395-E2 (hg19 chr8: 28222803-28222908)

ZNF395-E2-F: TGGGCCGCCCGTGACTTTTC (SEQ ID NO: 17)

ZNF395-E2-R: GGTTGGAAGGAGGCCACCGC (SEQ ID NO: 18)

ZNF395-E3 (hg19 chr8: 28223142-28223230)

ZNF395-E3-F: TCGTGCTGAAGGCTTCTCAGGAAA (SEQ ID NO: 19)

ZNF395-E3-R: CCCCTCCTGTTGGTGACGGC (SEQ ID NO: 20)

ZNF395-E4 (hg19 chr8: 28269095-28269211)

ZNF395-E4-F: AAGCGGCGGGAGGAGGTTGA (SEQ ID NO: 21)

ZNF395-E4-R: GGGCTGCGTCACCTGCAGAA (SEQ ID NO: 22)

5 **Luciferase assay**

[00113] Genomic DNA from where 786-O cells were extracted using DNeasy Blood & Tissue Kit (Qiagen). Regions corresponding to putative enhancers were amplified using CloneAmp HiFi PCR Premix (Clontech) and cloned into the pGL3 luciferase reporter vector with a minimal FOS promoter.

10 Forward primer: (SEQ ID NO: 23)

GTAGCTGCATAGATCTGCGCGCCACCCCTCTGGCGCCACCGT

Reverse primer: (SEQ ID NO: 24)

GTAGCTGCATCAAGCTTGCCGGCTCAGTCTTGGCTTCTC

[00114] The day prior to transfection, 1×10^4 cells were seeded into each well of a 96-well plate.

15 Cells were transfected with 100 ng of pGL3-Fos-enhancer and 20 ng of pRL-SV40 (Renilla luciferase vector, Promega). Cells were lysed and analyzed using the Dual-Luciferase Reporter System (Promega). Primer sequences used to amplify genomic regions for luciferase reporter assays are as follows:

VEGFA-E1 (hg19 chr6: 43635485-43636708)

20 VEGFA-E1-F_MluI: GCTCTTACGCGT TGGGGGTGCCTCTCCCACTG (SEQ ID NO: 25)

VEGFA-E1-R_NheI: GCCCGGGCTAGC GGGTGGGGGTCCAACAGGACA (SEQ ID NO: 26)

VEGFA-E2 (hg19 chr6: 43692413-43693560)

VEGFA-E2-F_MluI: GCTCTTACGCGT CCCATCCCCTGCCTCCTGCT (SEQ ID NO: 27)

VEGFA-E2-R_NheI: GCCCGGGCTAGC TGGGCTGGCTGCAAAGTGGC (SEQ ID NO: 28)

25 SLC2A1-E1 (hg19 chr1: 43523259-43525686)

SLC2A1-E1-F_MluI: GCTCTTACGCGT TGGTGACCGTGTTGGGGGTGA (SEQ ID NO: 29)

SLC2A1-E1-R_NheI: GCCCGGGCTAGC TCCCCGCCCTCTGTTGCAT (SEQ ID NO: 30)

ZNF395-E1 (hg19 chr8: 28220788-28221483)

ZNF395-E1-F_MluI: GCTCTTACGCGT ACAGGTGTGCGCTACCACGC (SEQ ID NO: 31)

30 ZNF395-E1-R_NheI: GCCCGGGCTAGCTGGTGTGGAATTCTGGCCAGTTAAAGG (SEQ ID NO: 32)

ZNF395-E2 (hg19 chr8: 28221957-28222965)

ZNF395-E2-F_MluI: GCTCTTACGCGT TCGGGAGGTTCAAGACCAGCCT (SEQ ID NO: 33)

ZNF395-E2-R_NheI: GCCCGGGCTAGCGCTCCCAAGAAAGAACTTACCAGAGG (SEQ ID

35 NO: 34)

ZNF395-E3 (hg19 chr8: 28222984-28224154)

ZNF395-E3-F_MluI: GCTCTTACGCGT ACCAGCCATCCCCTAGTTTGCC (SEQ ID NO: 35)

ZNF395-E3-R_NheI: GCCCGGGCTAGC GGCATTTGTCAGCAGAGATGTTGGC (SEQ ID NO: 36)

5 Colony formation and cell viability assays

[00115] For colony formation assays, 5000 cells per condition were seeded into 6 well dishes and were allowed to grow for 12 days. Colonies were stained with 0.05% Crystal Violet. For cell viability assay, 1000 cells per condition were seeded into 96-well plate and the cell viability was measured by CellTiter-Glo Luminescent Cell Viability Assay (Promega) for 5 days.

10 Apoptosis assay

[00116] For each condition, 1×10^3 cells were seeded into each well of a 96-well plate. Caspase3/7 activity was measured with the cleavage of proluminescent caspase-3/7 substrate after 1 hour incubation using Caspase-Glo® 3/7 Assay (Promega). Alternatively, cells were stained with FITC Annexin V Apoptosis Detection Kit (BD Biosciences) and Calcein AM (ThermoFisher) and analyzed on a flow cytometer.

15

In vivo studies

[00117] All animal studies were conducted in compliance with animal protocols approved by Institutional Animal Care and Use Committee (IACUC) of Singapore. Female NOD/SCID mice (6-8 week old) were implanted with 1×10^6 A-498 or 1×10^6 786-O cells transduced with either empty vector control or shRNA clones subcutaneously in the flank. Tumour volume was monitored every 2 to 3 days. Tumour volume was calculated as $(\text{length} \times \text{width} \times \text{width}) \times \pi/6$. Animals were sacrificed when the tumour volume exceeded 1000 mm^3 .

20

CRISPR-mediated enhancer deletion

[00118] To delete enhancer regions, 2 gRNAs (left and right) were used to cleave targeted regions as previously described. gRNAs were designed with ATUM gRNA Design Tool. Briefly, phosphorylated and annealed sense and antisense oligos were ligated into Bpil digested vectors. Left gRNAs were cloned into the Bpil digested pX330A-2A-GFP-1X2 backbone (Addgene #58766) whereas the right gRNAs into Bpil digested pX330S backbone (Addgene #58778). Golden gate assembly was performed to assemble the 2 gRNA protospacers into the pX330A-2A-GFP-1X2 plasmid backbone using a one-step digestion and ligation with slight modifications. After transfection using Lipofectamine 3000 (Life Technologies), GFP-positive single 786-O cells were sorted and cultured. Individual clones were validated for enhancer deletion by PCR of genomic DNA and the resulting gene expression was measured using qPCR and Taqman probes. Clones that were transfected with gRNAs but failed to have enhancer deletions were used as negative controls. The gRNAs used for deletion of enhancers are as follows:

30

35

ZNF395_E3 (hg19 chr8: 28223203-28224208)

ZNF395_E3_L_F_gRNA: CACCGTCCCTACTGCCGTCACCAAC (SEQ ID NO: 37)

ZNF395_E3_L_R_gRNA: AAACGTTGGTGACGGCAGTAGGGAC (SEQ ID NO: 38)

ZNF395_E3_R_F_gRNA: CACCGAAATATGTTTATGGTCCTCC (SEQ ID NO: 39)

5 ZNF395_E3_R_R_gRNA: AAACGGAGGACCATAAACATATTTC (SEQ ID NO: 40)

Validation primers for deletion of enhancers:

ZNF395-E3 (Product size after deletion: 293bp; WT: 1299bp)

ZNF395-E3-F: ACCAGCCATCCCCTAGTTTGCCA (SEQ ID NO: 41)

ZNF395-E3-R: GCCACCAGGTAGCAGTTGGGT (SEQ ID NO: 42)

10 **Date Accession**

[00119] Chromatin immunoprecipitation sequencing (ChIP-seq) and RNAseq data are available at Gene Expression Omnibus (GSE86095).

Example 2 - Cis-regulatory landscapes in clear cell renal cell carcinoma tumours are aberrant

- 15 **[00120]** To explore whether clear cell renal cell carcinoma tumours display alterations in their *cis*-regulatory landscapes *in vivo*, histone chromatin immunoprecipitation sequencing (ChIP-seq) profiles (3 marks: H3K27ac, H3K4me3, and H3K4me1) were generated in 10 primary tumour/normal pairs, 5 patient-matched tumour-derived cell lines, 2 commercially available clear cell renal cell carcinoma lines (786-O and A-498), and 2 normal kidney cell lines (HK2 and PCS-
20 400. Table 1 in example 1 shows patient clinical information. Of the original 87 samples, 79 samples passed pre-sequencing quality-control filters and were subjected to ChIP-seq processing and downstream analysis. In total, 2,363,904,778 uniquely mapped reads were generated. On average, 89% of H3K27ac peaks, 98% of H3K4me3 peaks, and 76% of H3K4me1 peaks obtained in our normal kidney tissues overlapped with peaks from adult kidney tissues in the Epigenomics
25 Roadmap dataset (**Figure 1A**). Among the 10 primary clear cell renal cell carcinomas, 9 harbored von Hippel-Lindau (*VHL*) mutations, detected by targeted sequencing and confirmed by Sanger sequencing (Table 2). Cell lines 786-O and A-498 also harbor *VHL* truncating mutations (Table 2). The *VHL* mutations co-occurred with somatic mutations of other chromatin modifiers commonly found in clear cell renal cell carcinoma, including *PBRM1* (7/10), *SETD2* (1/10), *KDM5A* (1/10),
30 *KDM5C* (1/10), *ARID1A* (1/10), and *KMT2C* (1/10).

	Sample ID	Mutation	Chr	Position	Ref	Alt	amino acid change	% alt alleles
tissue	12364284	indel	chr3	10188261	T	TAA	Phe136Asn-fsTer24	42.15
tissue	17621953	indel	chr3	10183790	GTATGGCTCAAC	G	Trp88Arg-	14.29

	Sample ID	Mutation	Chr	Position	Ref	Alt	amino acid change	% alt alleles
							fsTer41	
tissue	20431713	indel	chr3	10191582	AA	A	Asn193Met- fsTer201	49.09
tissue	40911432	indel	chr3	10191500	G	GT	Val125Cys- fsTer8	33.05
tissue	57398667	indel	chr3	10183765	TCGCAGTC	T	Ser80Ala- fsTer36	36.11
		missense	chr3	10183754	A	G	Ile75Val	32.86
tissue	70528835	indel	chr3	10183699	C	CG	Arg58Ala- fsTer75	77.27
tissue	74575859	missense	chr3	10188320	G	A	Val155Met	37.9
tissue	77972083	splice	chr3	10188195	TAG	T	splice acceptor variant	36.92
tissue	86049102	missense	chr3	10183771	T	G	Ser80Arg	28.6
tissue	75416923	wt						
cell line	786-O	indel	chr3	10183840	GG	G	Gly104Ala- fsTer55	100
cell line	A498	indel	chr3	10188282	TTGAC	T	Gly144Ser- fsTer14	93.87
cell line	40911432	indel	chr3	10191500	G	GT	Val125Cys- fsTer8	92.36
cell line	86049102	missense	chr3	10191548	G	A	Val140Ile	2.61
cell line	12364284	indel	chr3	10188261	T	TAA	Phe136Asn- fsTer24	93.17

Table 2: Clear cell renal cell carcinoma tissue and cell lines von Hippel-Lindau (VHL) mutation confirmation by sequencing

[00121] Specific histone modifications can distinguish different categories of functional regulatory elements—H3K4me3 is generally associated with promoters, H3K4me1 with enhancers, and H3K27ac with active elements. Integrating signals from three histone marks and GENCODE v19 annotated transcription start sites (TSS), active promoters were defined as

H3K27ac⁺/H3K4me3⁺/±2.0 kb TSS regions, and distal enhancers as H3K27ac⁺/H3K4me1⁺ regions not overlapping with promoters. Focusing on epigenomic events specific to somatic cancer cells, cell lines were derived from five primary tumours and, combined with the commercial lines, excluded peaks not found in any of the cell lines to reduce confounding effects from stromal cells.

5 On average, 80% overlap of chromatin immunoprecipitation sequencing (ChIP-seq) peaks was observed between primary tumours and matched lines (**Figure 1B**). Using these criteria, 17,497 putative promoters and 66,448 putative enhancers (**Figure 1C**) were identified, with numbers comparable with previous studies in other tumour types. The numbers of defined promoters and enhancers reached saturation after 4 and 16 samples, respectively, suggesting that a sample size of
10 20 (10 tumour/normal pairs) is sufficiently powered to discover the majority of *cis*-regulatory elements in clear cell renal cell carcinoma (**Figure 1E**). Principal components analysis (PCA) using the first two components of global H3K27ac intensities at promoters or enhancers (representing 83% and 64% of total variance, respectively; **Figure 1F**) successfully separated normal and tumour samples, indicating that genome-wide pervasive alterations in *cis*-regulatory
15 elements are a salient feature of clear cell renal cell carcinoma (**Figure 1D**).

[00122] Differential analysis was performed to identify altered promoters and enhancers. To define gained or lost regions, a fold difference of H3K27ac RPKM ≥ 2 , an absolute difference ≥ 0.5 , and for greater stringency no alterations in the reverse direction in the remaining tumour/normal pairs was applied (**Figure 1G**). At the threshold of $\geq 5/10$ patients, 80% of the
20 altered regions achieved statistical significance (q -value < 0.1 , paired t test, with Benjamini–Hochberg correction; **Figure 1H**), and at this same threshold, the increase in the fraction of samples meeting statistical significance reached a saddle point (**Figure 1I**). Applying these criteria, a high-confidence and comprehensive set of 4,719 gained promoters, 592 lost promoters, 4,906 gained enhancers, and 5,654 lost enhancers was obtained (**Figure 1C**, **Figure 1J**). Representative
25 regions are presented in **Figure 1K** (**Figure 1Ki and 1Kii**).

[00123] Supporting these data, gained promoters and enhancers exhibited increased chromatin accessibility measured by higher FAIRE-seq signals in tumour tissues than normal tissues, respectively ($P < 0.0001$) and also decreased DNA methylation based on data from The Cancer Genome Atlas (TCGA), consistent with reciprocal relationships between active regulatory regions
30 and DNA methylation (**Figure 1L**). Interestingly, elevated expression of long noncoding RNAs adjacent to gained promoters and enhancers was noted in tumour tissues compared with normal tissues ($P < 0.0001$, respectively). Lastly, many of the *cis*-regulatory elements were confirmed to involve regions previously implicated in clear cell renal cell carcinoma; for example, gains of H3K27ac signals and enrichment of H3K4me1 at a distal enhancer of *CCND1* overlapping with a
35 renal cell carcinoma susceptibility locus (rs7105934; **Figure 1M**) was observed. The ability to

identify this previously known enhancer with unbiased profiling further supports the method of this study.

Example 3 - Tumour-specific enhancers are associated with hallmarks of clear cell renal cell carcinoma

[00124] To identify genes modulated by the tumour-specific regulatory elements, enhancers were assigned using three approaches. The first approach utilized predefined linear proximity rules involving a set of highly confident genes (GREAT algorithm). MSigDB pathway analysis using GREAT-assigned genes revealed that gained enhancers exhibit a highly significant renal cell carcinoma -specific signature compared with gained promoters (enhancer q -value = 3.2×10^{-26} ; promoter q -value = 1.5×10^{-1} , binomial FDR; **Figure 2A**). Although gained promoters were involved in general cancer processes (for example, cell cycle, transcription, and RNA metabolism) for a complete list of promoter pathways), gained enhancers were enriched in disease-specific features of clear cell renal cell carcinoma, including HIF1 α network activity, proangiogenic pathways (platelet activation and PDGFR β signaling), and SLC-mediated transmembrane transport (**Figure 2A**) for a complete list of enhancer pathways). Notably, HIF1 α network activity consistently emerged as one of the top five pathways, even with perturbations in the patient thresholds used to define gained enhancers (≥ 3 –8 patients; Table 3).

Cut-offs	Binomial Rank	FDR	Binomial FDR
≥ 3	1		2.50E-31
≥ 4	3		7.90E-28
≥ 5	1		9.20E-30
≥ 6	1		1.50E-25
≥ 7	1		2.70E-18
≥ 8	4		3.30E-18

Table 3: HIF pathway as the top pathway with patient thresholds. FDR stand for false discovery rate.

[00125] Individual genes associated with gained enhancers included well-known hypoxic targets (*VEGFA*, **Figure 2B**; *CXCR4*) and metabolic genes involved in glycolysis, glutamine intake, and lipid storage (*GLUT1/SLC2A1*, **Figure 2C**; *HK2*, *PFKFB3*, *PLIN2*, **Figure 2D**) and *SLC38A1* (**Figure 2E**). The presence of enhancers around metabolic enzymes and transporters is largely consistent with the metabolic contexture of clear cell renal cell carcinoma, which involves

increased glycolysis and glutaminolysis. Indeed, gene ontology (GO) analysis of gained enhancers strongly reflected hallmark metabolic changes associated with clear cell renal cell carcinoma, including monocarboxylic acid transmembrane transporter activity (binomial FDR q -value = 1.6×10^{-10} ; **Figure 2F**).

- 5 **[00126]** A second method of enhancer–gene assignment based on correlations between H3K27ac signals and expression of genes within the same topologic associated domain (TAD). Using a q -value of <0.05 based on Spearman correlation, 2,311 gained enhancers were assigned to 2,186 protein-coding targets. H3K27ac signals of many gained enhancers were highly correlated with gene expression of their putative target genes. For example, H3K27ac levels of a *VEGFA* enhancer exhibited high correlation with *VEGFA* gene expression ($r = 0.83$, Spearman correlation),
 10 whereas H3K27ac signals of an *SLC2AI* enhancer were highly correlated with *SLC2AI* gene expression ($r = 0.72$, Spearman correlation; **Figure 2B**; **Figure 2G**). Similar to the GREAT approach, the TAD correlation approach also highlighted hypoxia (Krieg_Hypoxia_not_via_KDM3A, FDR q -value = 7×10^{-120}) and metabolism
 15 (Chen_Metabolic_Syndrome_Network, FDR q -value = 2×10^{-91}) as highly enriched pathways (Table 4).

Gene Set Name	# Genes in Gene Set (K)	Description
KRIEG_HYPOXIA_NOT_VIA_KDM3A	770	Genes induced under hypoxia independently of KDM3A [GeneID=55818] in RCC4 cells (renal carcinoma) expressing VHL [GeneID=7428].
PILON_KLF1_TARGETS_DN	1972	Genes down-regulated in erythroid progenitor cells from fetal livers of E13.5 embryos with KLF1 [GeneID=10661] knockout compared to those from the wild type embryos.
PUJANA_ATM_PCC_NETWORK	1442	Genes constituting the ATM-PCC network of transcripts whose expression positively correlated (Pearson correlation coefficient, PCC ≥ 0.4) with that of ATM [GeneID=472] across a compendium of normal tissues.
CHEN_METABOLIC_SYNDROM_NETWORK	1210	Genes forming the macrophage-enriched metabolic network (MEMN) claimed to have a causal relationship with the metabolic syndrom traits.
BLALOCK_ALZHEIMERS_DISEASE_UP	1691	Genes up-regulated in brain from patients with Alzheimer's disease.

RODWELL_AGING_KIDNEY_UP	487	Genes whose expression increases with age in normal kidney.
PUJANA_BRCA1_PCC_NETWORK	1652	Genes constituting the BRCA1-PCC network of transcripts whose expression positively correlated (Pearson correlation coefficient, PCC ≥ 0.4) with that of BRCA1 [GeneID=672] across a compendium of normal tissues.
DODD_NASOPHARYNGEAL_CARCINOMA_DN	1375	Genes down-regulated in nasopharyngeal carcinoma (NPC) compared to the normal tissue.
MARSON_BOUND_BY_E2F4_UNSTIMULATED	728	Genes with promoters bound by E2F4 [GeneID=1874] in unstimulated hybridoma cells.
NUYTEN_EZH2_TARGETS_UP	1037	Genes up-regulated in PC3 cells (prostate cancer) after knockdown of EZH2 [GeneID=2146] by RNAi.

Table 4: Highly enriched pathways of TAD correlation approach

[00127] Third, to independently validate the GREAT and TAD approaches in the specific context of clear cell renal cell carcinoma, the interactome of clear cell renal cell carcinoma tumour-specific enhancers was studied by performing Capture-C assays. Compared with other chromatin capture techniques, Capture-C offers both high-resolution (down to single Kb resolution) and high-throughput interrogation of user-defined regions (a usual working range of 10–500 regions). Probes were designed against a subset of 56 gained enhancers and examined their interactions with protein-coding genes in 786-O cells. Each gene–enhancer pair revealed by Capture-C was further filtered by correlations between gene expression and H3K27ac levels (q -value < 0.05). The 56 gained enhancers were paired with 36 protein-coding genes. 58% of these were predicted by GREAT, and 80% by gene correlations within TADs. The median distance of interactions detected by Capture-C was 16 kb, and 83% of the interactions fell within a 100-kb window (**Figure 2H**). As a visual example, Capture-C confirmed interactions between *VEGFA* enhancer and the *VEGFA* TSS, spanning a distance of about 100 kb (**Figure 2B**), and interactions between the *SLC2A1* enhancer and its promoter (**Figure 2C**). Taken collectively, these findings highlight the disease-specific nature of enhancer elements and an important role for enhancer malfunction in modulating clear cell renal cell carcinoma pathology.

Example 4 - Tumour super-enhancers identify *ZNF395* as a master regulator of clear cell renal cell carcinoma tumorigenesis

[00128] The importance of enhancers in clear cell renal cell carcinoma led to this study to examine the landscape of “superenhancers” or “stretch-enhancers”—dense clusters of enhancers located near master regulators of cell identity and disease. Using ROSE, 1,451 superenhancers were identified in the clear cell renal cell carcinoma cohort, of which 1,157 were gained in tumours and 294 were lost in tumours.

[00129] Putative targets of top gained superenhancers validated well-known oncogenes including *MYC/PVT1*, *VEGFA*, and *HIF2A* (Figure 3A, 3B and 3C). In addition, several less-known genes were found including *ERGIC1*, *ZNF395*, *SLC28A1*, and *SMPDL3A* (Figure 3D). These genes were highly overexpressed in tumours compared with their matched normal tissues (Figure 3D). Furthermore, they were unique to clear cell renal cell carcinoma and were not overexpressed in papillary and chromophobe renal cell carcinomas, two other distinct clear cell renal cell carcinoma subtypes (Figure 3D). For instance, *ZNF395* exhibited a tumour–normal ratio of about 7 in clear cell renal cell carcinoma ($P = 1 \times 10^{-22}$, paired *t* test) but experienced little overexpression in papillary and chromophobe renal cell carcinoma with tumour–normal ratios of 1.2 and 1.3, respectively ($P = 0.02$ in papillary and $P = 0.06$ in chromophobe, paired *t* test).

[00130] Conversely, genes associated with lost super-enhancers were recurrently suppressed in clear cell renal cell carcinoma and included *EFHD1*, *EHF*, *MAL*, *GCOM1*, and *HOXB9* (Figure 3D). In contrast to the lineage-specific nature of tumour super-enhancers, genes associated with lost super-enhancers were common between clear cell renal cell carcinoma and papillary renal cell carcinoma, implying a more universal function of tumour suppressor genes. For example, *EHF/ESE2*, a tumour suppressor previously found in prostate cancer, exhibited reduced expression across all three renal cell carcinoma subtypes (clear cell renal cell carcinoma tumour/normal = 0.05, $P = 3 \times 10^{-15}$; papillary tumour/normal = 0.1, $P = 2 \times 10^{-6}$; chromophobe tumour/normal = 0.1, $P = 2 \times 10^{-6}$).

[00131] Since current therapeutic targets in kidney cancer are limited to angiogenesis and mTOR pathways, less-understood genes uncovered by superenhancer profiling were examined. *ZNF395* and *SMPDL3A* were chosen for their differential tumour expression (6–7 tumour–normal ratio; Figure 3D) and high abundance (average RPKM of *ZNF395* about 112; average RPKM of *SMPDL3A* about 58). Even though *ZNF395* was previously identified as a potential clear cell renal cell carcinoma biomarker, its functional role in clear cell renal cell carcinoma malignancy remains unexplored. *SMPDL3A* shares 31% amino acid identity with the acid sphingomyelinase *SMPDI* and is a target of a master regulator of cholesterol metabolism, liver X receptors (LXR).

[00132] Quantitative PCR (**Figure 3E**) and immunoblotting (**Figure 3F**) confirmed that A-498 and 786-O clear cell renal cell carcinoma cells exhibited high expression of *ZNF395* and *SMPDL3A*, whereas normal kidney proximal tubule cells, PCS-400 and HK2, exhibited low expression of both genes. siRNA mediated knockdown of *SMPDL3A* had a cell line-dependent effect on colony formation, inhibiting the growth of A-498 cells but having no observable effect on 786-O cells (**Figure 3G**). On the other hand, *ZNF395* consistently inhibited colony formation in both 786-O and A-498 cells but had minimal effect on normal kidney cells (**Figure 3G, Figure 3H**). Consistent with this phenotypic observation, the *ZNF395* super-enhancer was active only in clear cell renal cell carcinoma cells (786-O and A-498) but silent in normal kidney cells (HK2 and PCS-400; **Figure 3I**). *SMPDL3A* and *SLC28A1* (**Figure 3J**) are also shown to be associated with a clear cell renal cell carcinoma-specific super-enhancer. *SLC6A3*, *EGLN3* and *VEGFA* shows gain in promoters and enhancers in the tumour sample as compared to the normal (non-diseased) sample (**Figure 3Ki, 3Kii and 3Kiii**). Furthermore, among the 33 types of cancer profiled by The Cancer Genome Atlas (TCGA), *SMPDL3A* (**Figure 3L**), *SLC28A1* (**Figure 3M**), *SLC6A3* (**Figure 3N**), *VEGFA* (**Figure 3O**), *EGLN3* (**Figure 3P**), *ZNF395* (**Figure 3Q** – only 12 cancer types profiled) are also shown to be highly expressed in clear cell renal cell carcinoma tumours (KIRC) from The Cancer Genome Atlas (TCGA) data.

[00133] No study to date has functionally tested the tumourigenic requirement of *ZNF395* in clear cell renal cell carcinoma or any other cancer type. *ZNF395*'s tumour-promoting effect using individual shRNA clones was validated (**Figure 3R, Figure 3S**). Two independent *ZNF395* shRNA clones drastically decreased *in vitro* colony formation (**Figure 3T**) and cell viability (**Figure 3U**) in both A-498 and 786-O cells. *ZNF395* knockdown also resulted in increased apoptosis measured by cleavage of caspase 3/7 substrates (**Figure 3V**) and Annexin V staining (**Figure 3W**). *In vivo*, tumour formation studies in mouse xenograft models revealed marked tumour suppression by *ZNF395* depletion (**Figure 3X**). Knockdown of *ZNF395* led to elimination of A-498 tumours up to day 74, when tumours in the control group began to exceed the size limits imposed by institutional animal protocols. Similarly, *ZNF395* depletion significantly slowed *in vivo* tumour growth of 786-O cells (**Figure 3X**). Taken together, the role *ZNF395* plays in clear cell renal cell carcinoma tumourigenesis was shown.

30

Example 5 - von Hippel-Lindau (*VHL*) deficiency remodels clear cell renal cell carcinoma enhancer landscapes

[00134] To explore the extent to which epigenetic changes observed in primary clear cell renal cell carcinomas (**Figure 1**) are directly driven by von Hippel-Lindau (*VHL*) loss, chromatin changes in isogenic cell lines were examined with and without *VHL* restoration. Consistent with

35

earlier functional studies of *VHL*, *VHL* restoration in 786-O, A-498, and 12364284 cells had negligible effects on proliferation, colony formation, and apoptosis *in vitro*, but profoundly delayed tumour growth *in vivo* (**Figure 4A, 4B, 4C and 4D**), suggesting the importance of *VHL* in modulating processes required for *in vivo* tumourigenesis, including tumour–stroma cross-talk, angiogenesis, cell–matrix interactions, or tumour metabolism.

[00135] Focusing on the same regions defined in the primary tumours (4,719 gained promoters, 4,906 gained enhancers, and 1,157 gained super-enhancers; **Figure 1C**), von Hippel-Lindau (*VHL*)-driven H3K27ac changes in four different cell lines (two commercial cell lines: 786-O and A-498; and two patient-derived cell lines: 12364284 and 40911432) was examined. Consistently across all four cell lines, *VHL* restoration induced more pronounced changes on enhancers and super-enhancers than on promoters (**Figure 4E, 4F, 4G and 4H**). For example, in 786-O cells, after *VHL* restoration 12% of enhancers (549 enhancers) were significantly depleted, compared with 6.5% of promoters (321 promoters; **Figure 4E**). This confirmed that a greater fraction of enhancers were significantly altered by *VHL* restoration than promoters ($P < 2.2 \times 10^{-16}$, proportions test), and an even higher proportion involved gained superenhancers ($P < 2.2 \times 10^{-16}$, proportions test).

[00136] Even though gained enhancers were expected to show only depletion after von Hippel-Lindau (*VHL*) restoration, changes in H3K27ac levels were bidirectional (**Figure 4E**). However, only gained enhancers with H3K27ac depletion were uniquely active in *VHL*-mutated clear cell renal cell carcinoma cell lines (786-O, A-498, and 12364284) compared with *VHL*-wild-type clear cell renal cell carcinoma cells (86049102L), normal kidney cell lines (PCS-400, HK2, and HKC-8), and 31 other cell lines of various cancer types (**Figure 4I**). The lack of H3K27ac signals in normal kidney cell lines argues against tissue lineage as the dominant contributor to the high H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) signals seen in clear cell renal cell carcinoma cell lines. On the other hand, gained enhancers with H3K27ac enrichment after *VHL* restoration showed high activity across multiple cancer types, suggesting that these enhancers are not unique to clear cell renal cell carcinoma (**Figure 4I**).

[00137] Furthermore, only gained enhancers showing H3K27ac depletion after von Hippel-Lindau (*VHL*) restoration were significantly associated with a concomitant downregulation of gene expression of their putative targets in both 786-O and 12364284 cells, whereas enhancers gained in primary clear cell renal cell carcinomas and further H3K27ac enriched after *VHL* restoration did not lead to significant gene upregulation on a global level (**Figure 4J, Figure 4K**). These results suggest that the former enhancers (H3K27ac depletion) are likely to represent clear cell renal cell carcinoma - and *VHL* specific epigenomic alterations, whereas the latter enhancers (H3K27ac enrichment) are likely to represent signify generic, compensatory mechanisms in response to *VHL*

restoration. Combining data from multiple lines, a total of 1,564 enhancers were depleted by *VHL* restoration in ≥ 1 cell line, representing almost a third (32%) of all gained enhancers identified in primary clear cell renal cell carcinoma tumours. The proportion of *VHL* responsive enhancers increased with the level of patient recurrence—only 7.8% of nonrecurrent gained enhancers (1/10 patients) showed von Hippel-Lindau (*VHL*)-mediated H3K27ac depletion, whereas 18% of enhancers recurrently gained in 9 of 10 patients and 20% of enhancers gained in 10 of 10 patients showed H3K27ac depletion in 786-O cells (**Figure 4L**, $P = 0.0001$, proportions test), consistent with the high prevalence of *VHL* mutations (9/10 patients) in the studies. Interestingly, unsupervised clustering using the 1,564 *VHL*-responsive gained enhancers segregated the single *VHL*-wild-type tumour (ID 75416923) away from the remaining 9 *VHL*-mutant tumours (**Figure 4M**), with the *VHL*-wild-type tumour showing low H3K27ac signals at the *ZNF395* superenhancer comparable with its patient-matched normal (**Figure 4N**). Collectively, pathway analysis of enhancers depleted in ≥ 2 cell lines highlighted direct p53 effectors, integrin-linked kinase signaling, and HIF1 α transcription factor networks as the top five pathways, covering genes such as *EGFR* (**Figure 4O**), *CCND1* (**Figure 4P**), *ITGB3* (**Figure 4Q**), *VEGFA* (**Figure 4S**), *SLC2A1* (**Figure 4R**), and *HK2* (**Figure 4T**). These results support a role for *VHL* loss in clear cell renal cell carcinoma enhancer malfunction, even in the presence of other driver mutations.

[00138] It was also examined whether other histone marks were concomitantly altered with H3K27ac marks. A high degree of correlation was found between H3K27ac and H3K4me1 in response to von Hippel-Lindau (*VHL*) restoration in both 786-O cells ($r = 0.77$, Pearson correlation) and 12364284 cells ($r = 0.61$, Pearson correlation) in **Figure 4U**. Globally, enhancers exhibiting H3K27ac depletion also experienced concomitant H3K4me1 depletion (**Figure 4W**). It was next examined whether *VHL* restoration led to acquisition of the H3K27me3 repressive mark. Despite a moderate anticorrelation of H3K27ac and H3K27me3 (786-O cells: $r = -0.28$, Pearson correlation; 12364284 cells: $r = -0.22$, Pearson correlation, **Figure 4V**), H3K27me3 levels remained low at gained enhancers even after *VHL* restoration (**Figure 4W**). These findings suggest that *VHL* restoration may result in a loss of enhancer identity by codepletion of H3K27ac and H3K4me1, but not a formal transition to a poised enhancer state that would have retained H3K4me1 but acquired H3K27me3.

30

Example 6 - HIF2 α -HIF1 β heterodimer is enriched at von Hippel-Lindau (*VHL*)-responsive enhancers

[00139] It was investigated which transcription factors might mediate von Hippel-Lindau (*VHL*)-dependent chromatin remodeling at gained enhancers. Using the primary clear cell renal cell carcinoma dataset, enrichment of *trans*-regulators in gained enhancers over lost enhancers was

35

examined. Using HOMER, it was found that the top enriched motifs were the AP1 family, ETS family, and NF- κ B– p65–Rel and HIF1 α /2 α motifs (**Figure 5A**). For subsequent *in vitro* validation, c-Jun was chosen as a representative AP1 family member because of its activation in clear cell renal cell carcinoma and ETS1 as an ETS family representative because of its known interaction with HIF2 α , but acknowledge that other family AP1 and ETS family members may play a role in clear cell renal cell carcinoma. Immunoblotting of c-Jun, ETS1, and NF- κ B–p65 showed variable protein expression in both normal and tumour cell lines, but expression of HIF1 α and HIF2 α restricted to tumour cells only (**Figure 5B**). HIF2 α was expressed in a higher proportion of clear cell renal cell carcinoma cell lines than HIF1 α (**Figure 5B**). Gene expression of these transcription factors was further examined in the The Cancer Genome Atlas (TCGA) cohort and found that *ETS1*, *RELA* (subunit of NF- κ B–p65), and *HIF2a* were significantly overexpressed in tumours compared with normal tissues, with a range of tumour-association expression patterns similar to variations in clear cell renal cell carcinoma lines (**Figure 5C**).

[00140] To further investigate chromatin occupancy of these factors, chromatin immunoprecipitation sequencing (ChIP-seq) binding profiles of c-Jun, ETS1, and NF- κ B cells were generated and HIF2 α , HIF1 α , and HIF1 β binding profiles from the previous literature were examined in 786-O cells. As 786-O cells contain lost endogenous *HIF1a* expression through genomic deletion, the HIF1 α ChIP-seq was performed on 786-O cells genetically manipulated to reexpress HIF1 α protein. ChIP-seq results showed that all six transcription factors exhibited increased occupancy at gained enhancers compared with lost enhancers, validating the HOMER predictions (**Figure 5D**).

[00141] To determine which of these transcription factors might be directly dependent on von Hippel-Lindau (*VHL*), their protein expression was then compared in *VHL*-mutated isogenic cell lines with and without wild-type-*VHL* restoration. As shown in **Figure 5E**, *VHL* restoration consistently downregulated HIF2 α expression in both 786-O and 12364284 cell lines, but protein levels of other factors displayed contrasting trends between the two cell lines, implying that among the six factors examined, HIF2 α protein expression was the most *VHL*-dependent. Indeed, supporting an important role for HIF2 α in *VHL*-dependent enhancer remodeling, only HIF2 α and HIF1 β were significantly enriched at enhancers showing *VHL*-dependent H3K27ac depletion (**Figure 5F**). Moreover, among all known motifs in the HOMER database, HIF2 α was the most enriched motif at *VHL*-responsive enhancers exhibiting H3K27ac depletion ($P = 1 \times 10^{-11}$). In contrast, HIF1 α was not enriched at enhancers showing H3K27ac depletion (**Figure 5F**). Despite sharing many binding sites with HIF2 α , HIF1 α predominantly localized to promoter-proximal regions, whereas HIF2 α frequently occupied introns and intergenic regions in 786-O cells (**Figure 5G**), consistent with a promoter-centric occupancy of HIF1 α and an enhancer-centric occupancy of

HIF2 α (**Figure 5H**). Gained enhancers displayed a HIF2 α occupancy twice that of tumour-specific promoters ($P < 1 \times 10^{-16}$, proportions test) in 786-O cells, suggesting that HIF2 α may play a greater role in regulating enhancers than promoters.

[00142] To extend these HIF1 α and HIF2 α occupancy-pattern findings to a system that
 5 expresses endogenous levels of both factors, HIF1 α and HIF2 α chromatin immunoprecipitation
 sequencing (ChIP-seq) was performed in 40911432 clear cell renal cell carcinoma cells, which
 abundantly coexpress both HIF α subunits (**Figure 5B**). Similar to 786-O, in 40911432 cells, HIF1 α
 showed a preferential occupancy at promoter-proximal regions, whereas a large proportion of
 HIF2 α was found in distal regions (introns and distal intergenic regions; **Figure 5I**). A higher
 10 proportion of HIF1 α binding sites overlapped with gained promoters than HIF2 α (68% of HIF1 α
 vs. 41% of HIF2 α , $P = 0.002$, proportions test; **Figure 5J**). Conversely, a higher proportion of
 HIF2 α binding sites overlapped with gained enhancers than HIF1 α (29% of HIF1 α vs. 51% of
 HIF2 α , $P < 2.2 \times 10^{-16}$, proportions test). HIF2 α 's preferential occupancy at enhancers was further
 substantiated by its higher enrichment at enhancers showing H3K27ac depletion after von Hippel-
 15 Lindau (*VHL*) restoration than HIF1 α (**Figure 5K**). Specific examples of *VHL*-responsive
 enhancers bound exclusively by HIF2 α but not HIF1 α included an enhancer near *UBR4* (**Figure**
5L) and a superenhancer near *CMIP* (**Figure 5M**). Therefore, even in HIF1 α /HIF2 α coexpressing
 clear cell renal cell carcinoma cells, these results suggest that HIF2 α plays a greater role in *VHL*-
 mediated enhancer remodeling than HIF1 α .

20

Example 7 - HIF2 α -HIF1 β bound enhancers modulate gene expression

[00143] To investigate the extent to which HIF2 α silencing is sufficient to recapitulate the
 effects of von Hippel-Lindau (*VHL*) restoration, H3K27ac chromatin immunoprecipitation
 sequencing (ChIP-seq) and RNA sequencing (RNA-seq) was performed in 786-O cells with *HIF2a*
 25 siRNA-mediated knockdown and analyzed correlations between *HIF2a* siRNA knockdown and
VHL restoration. When assessed against all genes, there was a low correlation ($r = 0.1$, $P = 5.2 \times$
 10^{-31}) between *HIF2* α knockdown and *VHL* restoration. Importantly, however, this correlation
 increased to 0.23 ($P = 5.8 \times 10^{-14}$) for genes near HIF2 α binding sites (**Figure 6A**). Similar results
 were obtained at the epigenomic level, where for gained enhancers the correlation was low at 0.06
 30 across all gained enhancers ($P = 1.9 \times 10^{-5}$) but increased substantially to 0.37 ($P = 9.5 \times 10^{-8}$) at
 HIF2 α -bound enhancers (**Figure 6B**) and at super-enhancers increased from 0.089 ($P = 0.0025$) to
 0.25 ($P = 0.00054$) at HIF2 α -bound super-enhancers (**Figure 6C**). As a visual example, H3K27ac
 signals at the *ZNF395* super-enhancer were diminished after *VHL* restoration or *HIF2a*
 knockdown, concomitant with decreased *ZNF395* gene expression (**Figure 6D**). Validation by RT-
 35 qPCR showed that *HIF2a* siRNA knockdown downregulated *VEGFA*, *SLC2A1*, and *ZNF395*

expression to a comparable degree as *VHL* restoration (**Figure 6E**). Decreases in luciferase reporter activity of enhancer elements were also consistent between *HIF2a* siRNA knockdown and *VHL* restoration (**Figure 6F**).

[00144] It was aimed to establish a causal link between HIF2 α -bound enhancers and control of
5 gene expression. CRISPR-mediated genomic depletion of the *ZNF395* enhancer region with the
highest HIF2 α peak was performed (**Figure 6G**). All four clones with the homozygous deleted
ZNF395 enhancer consistently downregulated their *ZNF395* expression compared with clones with
the intact enhancer ($P < 0.05$), providing evidence that *ZNF395* expression is epigenetically
controlled by this HIF2 α -HIF1 β -bound enhancer (**Figure 6G**). Taken together, these results
10 indicate that that HIF2 α is an important mediator of von Hippel-Lindau (VHL)-driven enhancer
remodeling.

Example 8 - von Hippel-Lindau (*VHL*) restoration reduced P300 recruitment but preserved promoter-enhancer interactions

15 [00145] Finally, this study sought to investigate the reason von Hippel-Lindau (*VHL*)
restoration caused a decrease in H3K27ac levels. Previous pulldown assays have reported that both
HIF2 α and HIF1 β can interact with histone acetyltransferase p300. Indeed, p300 frequently marks
enhancers and is thought to be recruited by tissue-specific transcription factors. However,
chromatin profiles of p300 have not been previously established in kidney cancer cell lines, so the
20 contribution of p300 in shaping enhancers in clear cell renal cell carcinoma remains unclear.
Therefore, p300 chromatin immunoprecipitation sequencing (ChIP-seq) was performed in 786-O
cells and confirmed its enrichment at gained enhancers over lost enhancers (**Figure 7A**).
Comparing p300 ChIP-seq with HIF2 α ChIP-seq yielded a surprisingly high degree of overlap
between HIF2 α and p300 (96%), even more than that of HIF2 α and HIF1 β (89%; **Figure 7B and**
25 **7C**). In contrast, other transcription factors such as c-Jun, ETS1, and NF- κ B did not exhibit such a
high degree of overlap ($\leq 60\%$; **Figure 7B**).

[00146] p300 binding at tumour enhancers with and without VHL was compared. Despite
increased p300 protein levels in 786-O cells after *VHL* restoration (**Figure 7D**), binding of p300
decreased across all three enhancers examined (**Figure 7E**). *HIF2a* depletion by siRNA
30 knockdown also decreased p300 recruitment (**Figure 7F**), suggesting that loss of HIF2 α may
interfere with p300 recruitment.

[00147] It was investigated whether von Hippel-Lindau (*VHL*) restoration and the subsequent
loss of p300 binding disrupted promoter-enhancer interactions. Capture-C of enhancer regions in
paired 786-O cell lines with and without *VHL* restoration was performed. Capture-C interactions
35 showed a relatively high correlation between *VHL*-deficient and *VHL*-restored 786-O cells at VHL-

responsive regions ($r = 0.74$, Pearson correlation), even higher than correlations observed at non-VHL-responsive regions ($r = 0.57$, Pearson correlation; **Figure 7G**). As a visual example, interactions between the *VEGFA* promoter and enhancer were intact even after *VHL* restoration (**Figure 7H**), indicating that loss of enhancer activity is likely insufficient to dissociate promoter–enhancer interactions. Furthermore, many of these promoter–enhancers were lineage specific; for example, the interaction between *SLC2A1* enhancer with its promoter was not detected in KATOIII, a gastric cancer cell line (**Figure 7J**). Therefore, promoter–enhancer interactions often preexist in kidney cells, frequently in a tissue-specific manner.

5
10
[00148] Clear cell renal cell carcinoma biomarkers were analysed from exosomes obtained from the culture medium of clear cell renal cell carcinoma cell line (A498) and normal kidney cell line (HK2), by measuring gene expression of the biomarkers by quantitative polymerase chain reaction (qPCR) (**Figure 8**). *ERGIC*, *EGLN3*, *ETS1*, *PVT1*, *MYC*, *SMPDL3A*, *SNX10*, *VEGFA* and *ZNF395* (**Figure 8A, 8B and 8C**) showed higher expression in clear cell renal cell carcinoma cell line compared to normal kidney cell line.

15
[00149] Microarray data from patient cohorts with clear cell renal cell carcinoma or benign oncocytoma were compared. Expression levels of *VEGFA*, *EGLN3*, *ZNF395*, *SLC6A3* and *SLC28A1* are higher in clear cell renal cell carcinoma compared to benign oncocytoma as shown by higher Z score values (Figure 9).

	Accession number	Sequence
ZNF395 mRNA (SEQ ID NO: 43)	NM_0186 60.2	aagtgcgcat gtgcgcgagg agtcgctcgg gcacttattg agcgccgact gtctacggcgccgggggt gatgggcaga ggcttcagt tcccctcgc ctccgcagga gaggagaggcagcagcatgg cgagtgtcct gtcccgacgc cttggaagc ggctccctct gggagcccgggtgttgggac ccagtgcctc ggagggggccc tcggctgccc caccctcgga gccactgcta gaaggggccc ctccccagcc ttaccacc tctgatgaca cccctgcca ggagcagccc aagggaagtc ttaaggtcc cagcacctcg ggcttcagc aggtggcctt tcagctggg cagaaggttt atgtgtgta cgggggtcaa gactgcacag gactgtgga gcagcacagtggatggagg gtcagggtgac cgtctggctg ctggagcaga agctgcaggt ctgctgcagggtggaggagg tgtggctggc agagctgcag ggcccctgtc cccaggcacc acccctggagcccgagccc aggcctggc ctacaggccc gtctccagga acatcgatgt cccaaagaggaagtcggagc cagtggaaat ggatgagatg atggcgcca tggctgtgac gtccctgtctgcagccctg ttgtacagag tctcccggg accgaggcca acttctctgc ttccgtgcggcctgcgacc catggaagga gactgggtgac atctcggaca gcggcagcag cactaccagcggtcactgga gtgggagcag tgggtctcc acccctcgc cccccacc ccaggccagcccaagtatt tgggggatgc tttggttct cccaaactg atcatggtt tgagaccgatcctgaccctt tctgtctgga cgaaccagct ccacgaaaaa gaaagaactc tgtgaagggtgatgtacaagt gcctgtggcc aaactgtggc aaagtctgc gtcattgt gggcatcaaa cgacacgtca aagccctcca tctgggggac acagtggact ctgatcatt caagcgggag gaggattct actacacaga ggtgcagctg aaggaggat ctgctgtgc tgtgtgtgtgtgccgag gcacccagt cctgggact cccacctcg agccagctcc ccccccagcatgactggcc tgcctctgtc tgtcttcca ccactctgc acaaagccca gtctccggc ccagaacatc ctggcccga gtctccctg cctcagggg ctctagcaa gtcagctcctgggtcctt ggcacattca ggcagatcat gcataccagg ctctgccatc ctccagatc ccagtctac cacacatcta caccagtgtc agctgggtg ctgccccctc cgccgctgc tctctctc cggtccggag ccggtcgcta agcttcagcg agccccagca gccagcacctgcgatgaaat ctatctgat cgtcacttct ccacccggg cccagagtgg tgccaggaaa gcccagggg aggctaagaa gtccgcaag gtgtatggca tcgagcaccg ggaccagtgg tgcacggcct gccggtgga gaaggcctgc cagcgcttc tggactgagc tgtgtgcag gttctactct gttcctggcc ctccggcag cactgacaa gaggccagtg tgtcaccagc cctcagcaga aaccgaaaga gaaagaacgg aaacacggag ttgggctct gttggttaagggtgaacacttaagcaatttttccattgtgcgaacatttttttaaaaaaaga aacaaaaata ttttcccc taaaataggagagagccaaa actgaccaag gctattcagc agtgaaccag tgaccaaaga attaattacc ctccgttcc cacatccca ctctctaggg

	Accession number	Sequence
		gattagcttg tgcgtgtcaa aagaaggaa agctcgttct gcttcctgct gagtccgtga attctttgct ttctaaactc ttccagaaag gactgtgagc aagatgaatt tacttttctt aaaaaaaaaa aaaaaaaaaa aaaaaaagag ttctggctg atgggtgact cagagtgcaggactgcctgg ccgtggggca gaggggtttg ccttctcgg aggggtacac ctgttcctg tctgagcatc ctgcatggaa gtcaaaggaa atccctttct tgggtgacgac ttaaactctg gttccctcag acattgggtt gcacccaac aaatattaaa tggcttctc ttaaagcca gagaaagagg tttttaaaa gactgtcgcc aaatagctga gccaaaaggc tgatcagaat tcaacttttg gaatgtggca gttaaact accttgatca ttctctctc ttctcaggaactcctgg agggtttgag cgtctggaaa ctctctctc tgaccgagg aagcacctc ctgacgccgc cttctccgg ttattgaaag gacgcctcag aaatgctttg tttcttta cgtgtatc agaagcctt actgattaaa gtttctttt atttgggtgg ccgggagaga cccaggagg ttctggaggt tcctttctgt ctctggccc caccaggat tccccatt ctgttctg cctgaaagca ggatgaggaa ggcaaggag agtccttgca cccgtgagcg tcaggatgag gaaatgacag gaggaagacg tgggtttggg ttagtggtg ctggcgtttggcccttgg tttctggag cctccaggga tctaggggag cctgggctgc gtgcatgctg ataagcagag ctgttcttg ggagaaggag ggaggtctcg ggagtgtagc accatgcaa ccagccctgc gcgaagacag agtgaaccac gcccggatgg cagggcattg ttctgtttg gtgtctact ttctccag cgtgacttat ttggggattc ctcaggcct actggaatgt gactgccac tgcccagctg cctcgggtac aagtctggc cctatgtccc agctgtcaggggctcaggga atctacca gccacctgc ctgggatgga gtgtcagcat ccaccttg gttgtcatc agccgcct cccagctcg ggtgaagata ttgggccac cagggtctcc ttggccctt cactagga atagacacgt gcttttaat gcaggacact ttgagtgtta caaatctgt agacctggca gtagggtcat gatgttggga aggtgtagt gccctagggt ggtgacagaa gggacagaca ctgtgcaca ggtgtcttg gtgatgggt tttttttt ataactagt aaaaaaaaaa aaatgtatgt ggaattctgt ctcttgtaa agctcaaagc caggctagcc tgagtggtgc cagggtctc cttctgtcc ctctgctc ctgagaatt aagagctggc agctgctgat ggtgttccc aacccccctc acttccaag acaacccca gcttcaggtc ctcatgggga ggggagggca cgttctgac acatgggaac ttcgtcagg agggcctccc ctccccct cctcagagt tttactgcc gtctgtctt tagaaagctg ttgaattcc cccgccccc agtttgacc gtgtagatat aactggatat acggatttt ctcttctgc aggccttta tccgttgg atacaggga ggaagagag gaataaagg agagagcagt gtggaacca cgtgtgtt gcttctct tactaggtt tgggtccacc ttccctgct gccttctgc cccctctct ccttgccact ggcggcctc ttgctcct tccaccgtg ctgccatcc gtgcctgctg tgtgttct

	Accession number	Sequence
		tcacacgtgc tctgttctcg ggggtgttcc attcatgcct tcttgagggg tgagggtggc ttgggaaccg acccagtgat catgcctact ttcttctttg tatctccctc ctcccagcc caccgggca gcagactctg atggaaggaa ggtgccgtag gtgggctttt agaaactaac gggactggtt tcaaagcag ttatctggg aaactgttta ttccagcgat gtgactttt tcagaatatt tcttggaatc atattcagag tctggggctg tgtgttgagc agccttaagg atgctagaca ctcathtagtcccaggag tccagcgaat gacgtctgtg gccaaagcag gtctcagggtg caaagcaaaa ggaccattta aagtaaaata gcttgattc aatcatgtga cttttaaatt ggctcagaaa gcaattttgt aatttcagag agtgttttga gccatggcca cgtgtcatt gtgagtctat agcttgactc ctggagAAC aatattcatt tgggtgtgga gactgatttg ctgggagaaa tctgtcctgt tactttctgg tcattcccagg ttctgacttt taccaggggc aaaaaaaaaaagcaagag ggagataaat cccatctgtg agttgtctt attggcgcct tttctcag ctgtctcca agtattattt ttactgttaa aaaattttt aaaaatgtga aatgtaatgt ttttacagca acaatatgaa atatatttta taaggataaa aatggtacct tgtctgattt aaaaaa
ZNF395 protein (SEQ ID NO: 44)	NP_06113 0.1	masvlrrlg krsllgarvl gpsasegpsa appseplleg aapqpftsd dtpcqeqpkevlkapstsgl qqvafqpgqk vyvwyggqec tglveqhsww egqvtwille qklqvccrveevwlaelqgp cpqapplepg aqalayrpvs rnidvpkrks davemdemma amvltslscspvvqppgte anfsasraac dpwksgdis dsgssttsg hsgssgvstp spphpqaspylgdavgspq tdhgfetdpd pflldepapr krknskvmy kclwpnecgv lrsivgikrhvkalhlgtv dsdqfcreed fyytevlke esaaaaaaaa agtpvpgtpt sepaptsmtglplsalppp lhkaqssgpe hpgpesslps galsksapgs fwhiqadhay qalpsfqipvsphiytsvsw aaapsaacs slpvrslsf sepqqpapam kshlivtsp raqsgarkargeakkcrkvy giehrdqwet acrwkkacqr fld
SMPDL3A mRNA (isoform b) (SEQ ID NO: 45)	NM_0012 86138.1	accagtatgt cagtgttga catcaactgc accactgata cagagtcgg aattgagcttctacaagta cattccttcc taggccaac actgacgcta agaaatacga gaacagatcatcgtaaac gcagctgaag gtcaggcgaa ctgactcgct gcggaatctg cctttgcacgtgatcagtc gacgtctaca cccgcagccg tcttctgtct ccgcctcacc ctcaggcctgacggtcagag tggagctgcg ggacagcccc aacctccagg tcagccccgc ggccctcatggcgctggtg cgcgcactcg tctgctgctt gctgactgcc tggcactgcc gctccggcctcgggctgcc gtggcgcccc caggcggcag gaatctctct ccggcgatag ggatagcccacctcatgtt ctgtacctga actctcaaca gacactgtta taaatgtgat cactaatatgacaaccacca tccagagtct ctttccaaat ctccagggtt tccttcgcgt gggtaatcatgactattggc cacaggatca actgcctgta gtcaccagta aagtgtacaa

	Accession number	Sequence
		tgcagtagcaaacctctgga aaccatggct agatgaagaa gctattagta ctttaaggaa aggtgggttttattcacaga aagttacaac taatccaaac cttaggatca tcagtctaaa cacaactgtactacggcc caaatataat gacactgaac aagactgacc cagccaacca gtttgaatggctagaaagta cattgaacaa ctctcagcag aataaggaga aggtgtatat catagcacatgtccagtggt ggtatctgcc atcttcacag aacatcacag caatgagaga atactataatgaaaattga tagatatttt tcaaaaatac agtgaatgca ttgcaggaca atttatggacacactcaca gagacagcat tatggttctt tcagataaaa aaggaagtcc agtaaattctttgttgggtg ctctgctgtg tacaccagtg aagagtgttt tagaaaaaca gaccaacaatcctggatatca gactgtttca gtatgatcct cgtgattata aattattgga tatgttgagctattacttga atctgacaga ggcgaaatcta aaggagagagt ccatctggaa gctggagtatactcctgacct agacctacga cattgaagat ttgcagccgg aaagtattata tggattagctaaacaattta caatcctaga cagtaagcag ttataaaat actacaatta ctctttgtgagttatgaca gcagtgtaac atgtgataag acatgtaagg cctttcagat ttgtgcaattatgaatcttg ataattttc ctatgcagat tgcctcaaac agctttatat aaagcacaattactagtatt tcacagtttt tgctaataga aaatgctgat tctgattctg agatcaatttggggaattt tacataaatc ttgttaatt actgagtgagg caagtagact tcctgtctttgctttcttt ttttttctt ttgtatgcct taatgtagat atctttatca ttctgaattgtattatatat ttaaagtgtc cattaataga atgatggatg taaattggat gtaaatattcagtttatata attatatcta attgtacct ttgttgaat tgcatttat acaataaagcgaattcttta tctctaaaaa aaaaaaaaaa aaa
SMPDL3A protein (isoform b) (SEQ ID NO: 46)	NP_00127 3067.1	mtttiqslfp nlqvfpalgn hdywpqdqlp vvtskvynav anlwkpwldc eaistlrkggfysqkvtnp nlriisltn lyygpnimtl nktdpanqfe wlestlnnsq qnkekvyiihvpvgylpss qnitamreyy neklidifqk ysdviagqfy ghthrdsimv lsdkkgsppvnsfvpavtp vksvlekqtn npgirlfqyd prdyklldml qyylnltean lkgesiwkleyiltqtydie dlqpeslygl akqftildsk qfikyynyff vsydsstcd ktckafqicaimlnldnisya dclqlyikh ny
SMPDL3A protein (isoform a) (SEQ ID NO: 47)	NM_0067 14.4	accagtatgt cagtgtttga catcaactgc accactgata cacgagtcgg aattgagcttctacaagta cattccttcc taggccaac actgacgcta agaaatcga gaacagatcatcgctaaaca gcagctgaag gtcaggcgaa ctgactcgct gcggaatctg cctttgcacgtgatcagtcg gacgtctaca cccgcagccg tcttctgtct ccgcctcacc ctcaggcctgacggtccgag tggagctgcg ggacagcccg aacctccagg tcagccccgc ggccctccatggcgctggtg cgcgcactcg tctgctgcct gctgactgcc tggcactgcc getcggcctcgggctgccc gtggcgcccc caggcgccag gaatcctct cggcgatag

[illegible]

	Accession number	Sequence
SLC28A1 mRNA (SEQ ID NO: 49)	NM_0012 87761.1	acaacgatgt gaaggttata agctgcactg catggttgctgctggatgtgtgttcttgcttcctc tggatgctga cagaaacaag gctggaaggt ctgggacatg gagaacgacc cctcgagacg aagagagtcc atctctctca cacctgtggc caagggtctg gagaacatgg gggctgattt ctggaaagc ctggaggaag gccagctccc taggagtac ttgagccccg cagagatcag gagcagctgg agcgaggcgg cgccgaagcc ctctccaga tggaggaacc tgcagccagc cctgagagcc agaagcttct gcagggagca catgcagctg ttctgatgga tcggcacagg cctgctctgc actgggctct ctgccttctt gctggtggcc tgcctctgg atttcagag ggcctggct ctgtttgtcc tcacctgtgt ggtctcacc ttctgggcc accgcctgct gaaacggctt ctggggccaa agctgaggag gttctcaag cctcagggcc atccccgctt gctgctctgg tttaagaggg gtctagctct tctgcttctc ctggcctgg tctgtggtt gtctctggac acctccagc ggcctgagca actggtgtcc ttgcaggaa tctgcgtgtt cgtcgtctc ctctttgctt gctcaaagca tcattgcgca gtgtctgga gggccgtgtc ttggggactt ggactgcagt ttgtacttg actcctctc atcagaacag aaccaggatt cattgcgttc gagtggctgg gcgagcagat ccggaatttc ctgagtaca cgaaggctgg ctccagcttc gtgtttgggg aggcgctggt caaggatgtc ttgctttc aggttctccc catcattgtc ttttcagct gtgtcatac cgttcttac cacgtggcc tcattgcagt ggtgactctg aagattgcct ggctgatgca agtcaccatg ggcaccacag ccactgagac cctgagtggt gctggaaaca tctttgtgag ccagaccgag gctccattac tgatccggcc ctactggca gacatgacac tctctgaagt ccagttgtc atgaccggag gttacccac cattgtggc agcctgctgg gtgcctacat ctctttggg gtcagagctg aagtcctcac gacgtttgcc ctctgtggat ttgccaattt cagctccatt gggatcatgc tgggaggctt gacctcatg gtcccccaac ggaagagcga ctctccag atagtgtcc gggcgctctt cacgggagcc tgtgtgtccc tgggaacgc ctgtatggca gggatcctt acatgccag gggggctgaa gttgactgca tgcctctt gaacacgacc ctacgcagc gtagcttga gattaccag tgctgccgtg aggccttcca gacgtcaat ccagagtca gccagaggc cctggacaac tgctgtcggg tttaacca cagatctgt gcacagtga gacagaacat gctgtgctt ctgcgttct gagggctgtt ctccccggg aacctctgt cccaccttc ctttccag agccctctc agggagcca caggacttag accagctca atcccacaat tgggaagggt tcattgagtg agtgtgcaga gattgagtg ggacataagg aaggacatgt cccactccat ccccttctt gctccccat ttctaactc cccagtggt aatttcagg gtcactctg cctctccc ttcccccc acatccaaac agcacctgg tctctctat cccccctc ctggggctcc tcacatgcc ctctcttct gttgtgggt gcacacaaa gcctctccc ctccccact ctaggcact aggatctct tgtgcttc

	Accession number	Sequence
		cctgctgggt ggtgtcacct cttctctgc ttcagagaa acccttcccg cctttctca gagtgttcc caaactgagg tcccatggca cactgtcctg ggaggcgtc agagggtcc atgatggact aggtttgaa ccactgggtt aaataaact agagagggt gttta
SLC28A1 protein (SEQ ID NO: 50)	NP_00127 4690.1	mendpsrre sisltpvakg lenmgadfle sleegqlprs dlspacirss wseaapkpfs rwrnlqpalr arsfcrehm q lfrwigtgll ctglsafllv acldfqr al fvlvtcvl tflghrllkr llgpkllrfl kpqghprlll wfkr glalaa flglv lwlsl dtsqrpeqlv sfagicvfva llfacskhhe avswravswg lglqfvlgl virtepgfia fewlgeqiri flsytkagss fvfgealvkd vfafqlpii vffscvisvl yhvglmqwvi lkia wlmqvt mgttatetls vagnifvsqt eapllirpyl admtlsevhv vmtggyatia gslgayisf gvraevlttf alcgfanfss igimlgglts mvpqrksdfs qivlralfgt acvslvnacm agilymprga evdcmsllnt tlsssfeyi qccreafqsv npefspeald nccrfynhti caq
VEGFA mRNA (SEQ ID NO: 51)	NM_0010 25366.2	tcgcggaggc ttggggcagc cgggtagctc ggaggtcgtg gcgctggggg ctagcaccag cgctctgtcg ggaggcgagc cggttaggtg gaccggtcag cgactcacc ggccaggcg ctcgggtgctg gaatttgata ttcattgac cgggtttat cctcttctt ttttctaaa cattttttt taaaactgta ttgttctcg tttaattta ttttcttg ccattcccca cttgaatcgg gccgacggct tggggagatt gctctactt cccaaatcac tgttgattt ggaaaccagc agaaagagga aagaggtagc aagagctca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcgggct gcgagcagc aaagcgacag gggcaaagtgagtgcctgc tttgggggt gaccgccgga gcgcggcgtg agccctccc cttgggacccgcagctgac cagtcgcgt gacggacaga cagacagaca ccgccccag ccccagctaccacctctcc ccggccggcg gcggacagt gacgcggcgg cgagccgcgg gcaggggcccggagccgcgc ccggaggcgg ggtggagggg gtcggggctc gcggcgtcg actgaaacttttctccaac ttctgggtg ttctcgctt ggaggagccg tggctcgcgc gggggaagccgagccgagc gagcccgag aagtgttagc tcgggccggg aggagccga gccggaggagggggaggagg aagaagagaa ggaagaggag agggggccgc agtggcgact cggcgctcggaagccgggct catggacggg tgaggcggcg gtgtgcgcag acagtgtcc agccgcgcgcgtccccagg ccctggccc ggctcgggc cggggaggaa gtagtctcg ccgagcgccgaggagagc ggccgcccc cagcccgagc cggagaggga gcgcgagccg cgccggccccggtcgggcct ccgaaacat gaactttctg ctgtcttggg tgcattggag cttgccttgtgtctacc tccacatgc caagtgtcc cagggtcac ccatggcaga aggaggagggcagaatcatc acgaagtgt gaagttcatg gatgtctatc agcgcagcta ctccatccaatcgagacc ttgtggacat ctccaggag tacctgatg agatcgagta

	Accession number	Sequence
		catcttcaagccatcctgtg tgccctgat gcgatgcggg ggctgctgca atgacgaggg cctggagtgtgtgccactg aggagtccaa catcaccatg cagattatgc ggatcaaacc tcaccaaggccagcacatag gagagatgag cttctacag cacaacaaat gtgaatgcag accaaagaaa gatagagcaa gacaagaaaa aaaatcagtt cgaggaaagg gaaaggggca aaaacgaaagcgcaagaaat cccggfataa gtcctggagc gtgtacgttg gtgcccgtg ctgtctaattgccctggagcc tccctggccc ccatccctgt gggccttgct cagagcggag aaagcatttgtttgtacaag atccgcagac gtgtaaatgt tctgcaaaa acacagactc gcgttgcaaggcgaggcagc ttgagttaa cgaacgtact tgcagatgtg acaagccgag gcggtgagccgggcaggagg aaggagcctc cctcagggtt tcgggaacca gatctctac caggaaagactgatacagaa cgategatac agaaaccacg ctgcccac cacaccatca ccatgcacagaacagtcctt aatccagaaa cctgaaatga aggaagagga gactctgcgc agagcactttgggtccggag ggcgagactc cggcggaagc attcccgggc gggtagacca gcacggtccctcttgaatt ggattcgcca tttattttt cttgctgcta aatcacgag cccggaagattagagagttt tatttctggg attcctgtag acacaccac ccacatacat acatttatatatatatat tatatatata taaaaataaa tatctctatt ttatatatat aaaatatatatatttttt ttaaattaac agtgctaag ttattggtgt cttactgga tgtattgac tgctgtggac ttgagtggg aggggaatgt tccactcag atcctgacag ggaagaggaggagatgagag actctggcat gatcttttt ttgtcccact tgggtgggccc agggtcctctcccctgcccc ggaatgtgca aggccagggc atgggggcaa atatgacca gttttggaacaccgacaaa ccagccctg gcgctgagcc tcttaccac aggtcagacg gacagaaagacagatcacag gtacagggat gaggacaccg gctctgacca ggagtttggg gagcttcaggacattgtgt gctttgggga ttccctccac atgctgcacg cgcatctgc ccccaggggactgcctgga agattcagga gcctgggagg ccttcgctta ctctacctg cttctgagttgccaggaga ccaatggcag atgtcccggc gaagagaaga gacacattgt tggaagaagcagcccatgac agtcccctt cctgggactc gccctcatcc tcttctgct ccccttctggggtgcagcc taaaaggacc tatgtcctca caccattgaa accactagtt ctgtccccccaggagacctg gttgtgtgtg tgtgagtggt tgaccttct ccatccctg gtccttccctccctcccg aggcacagag agacagggca ggateccagt gccattgtg gaggcagagaaaaagaaaag tgtttatat acggtactta ttaatatcc cttttaatt agaaattaa acagttaatt taattaaaga gtagggtttt tttcagtat tcttggttaa tatttaattcaactatta tgagatgtat ctttgcctc ctctgctct cttatttga cgggttttgtatataaaat tcatgttcc aatctctctc tccctgatcg gtgacagtca ctgcttatcttgaacagat atttaatttt gctaacactc agctctgccc tccccgatcc cctggtccccagcacat tctttgaaa taaggttca

	Accession number	Sequence
		atatacatct acatactata tatatatttgcaacttgta tttgtgtgta tatatatata tatatgttta tgtatatatg tgattctgataaaatagaca ttgctattct gttttttata tgtaaaaca aaacaagaaa aaatagagaattctacatac taaatctctc tccttttta attttaatat ttgttatcat ttatttattggtgctactgt ttatccgtaa taattgtggg gaaaagatat taacatcacg tctttgtcttagtcagtt tttegagata ttccgtagta catatttatt tttaacaac gacaaagaaatacagatata tcttaaaaaa aaaaaagcat tttgtattaa agaatttaat tctgatctcaaaaaaaaaa aaaaaa
VEGFA protein (SEQ ID NO: 52)	NP_00102 0537.2	mtdrqtatp spsyhlpg rrtvdaaasr gqgpepapgg gvegvargv alkfvqllg csrfggavvr ageaepsgaa rsassgreep qpeegeeeee keeergpqwr lgarkpgswt geaavcadsa paarapqala rasgrggrva rrgaesgpp hsprrgsas ragpgraset mnflswvhw slalllylh akwsqaapma eggqnhhev vkfmdvyqrs ychpietlvdifqeypdeie yifkpscpl mrcggccnde glecvptees nitmqimrik phqqhigemsflqhnkcec rpkkdrarqe kksvrgkgkg qkrkrksry kswsvyvgar cclmpwslpgphpcgpcser rkhlfvdpq tkcsckntd srkarqlel nertercdkp rr
EGLN3 mRNA (isoform 1) (SEQ ID NO: 53)	NM_0013 08103.1	ggcttcgcgc tcgtgtagat cgtccctct ctggtgcac gctggggatc cggacctcg attctgcggg cgagatgccc ctgggacaca tcatgaggt ggacctggag aaaattgcc tggagtacat cgtgccctgt ctgcacgagg caatgggtgc ttgctatccg ggaaatggaa caggttatgt tcgccactg gacaaccca acggtgatgg tcgtgcac acctgcact actatctgaa caagaattgg gatgccaage tacatggtgg gatcctgcgg atatttcag aggggaaatc attcatagca gatgtggagc ccattttga cagactcctg ttctctggt cagatcgtag gaaccacac gaagtgcagc cctcttacgc aaccagatat gctatgactg tctggtactt tgatgctgaa gaaagggcag aagccaaaaa gaaattcagg aatttaacta ggaaaactga atctgcctc actgaagact gaccgtgctc tgaatctgc tggccttgt cattttagta acggttctg aattctctta aattcttga gatccaaaga tgacctctc agtgaaca atctcctgc tactcttgc atcttcaca tccctgtctt gtgtgtgta ctcatgtt tcttccaag actgtgtga tctcagata ctctcttgc cagatgaagt tacttgctaa ctccagaaat tctgcagac atctactcg gccagcgggt tacctgatag attcggtaat actatcaaga gaagagccta ggagcacagc gagggaatga accttacttg cactttatgt atactctcg attgaaagg aggaggttg aaaagaaaaa aatggaggtg gtagatgcca cagagaggca tcacggaagc ctaacagca ggaacagag aaatttgtgt catctgaaca attccagat gttcttaac caggctgtgt ggggtttctg gagaattatc acaacctaat gacattaata cctctagaaa gggtctgt catagtgaac aatttataag tgtccatgg ggcagacact cctttttcc cagtctgca acctggattt tctgctcag

	Accession number	Sequence
		ccccatttg ctgaaaataa tgactttctg aataaagatg gcaacacaat ttttctcca ttttcagttc ttacctggga acctaattcc ccagaagcta aaaaactaga cattagttgt tttggtgct ttgttggaat ggaatttaa tttaatgaa aggaaaaata taccctggt agttttgtgt taacctga taactgtgga aagagctagg tctactgata tacaataaac atgtgtgcat ctgaacaat ttgagagggg aggtggagtt ggaaatgtgg gtgttcctgt tttttttt tttttttt tagttttct tttaatgag ctcaccctt aacacaaaaa aagcaagggt atgtattta aaaaaggaag tggaaataaa aaaatctca agctattga gtctcgtct gtcctagca gtctttctc agtcacttg gctctctaga tccactgtgg ttggcagtat gaccagaatc atggaattg ctagaactgt ggaagcttct actcctgcag taagcacaga tcgactgcc tcaataact ggtattgagc acgtattttg caaaagctac tttctagt ttcagtatt actttcatgt tttaaaatc ctttaattt ctgcttgaa aatcccatga acattaaaga gccagaaata tttccttg ttatgtacgg atatatatat atatagtctt ccaagataga agttacttt ttcctctct ggtttgga aatttcaga taagacatgt caccattaat tctaacgac tgctctattt tgtgtacgg taatagtat caccttctaa attactatgt aattattca cttattatgt ttattgtct gtatccttc tctggagtgt aagcacaatg aagacaggaa tttgtatat tttaacaa tgcaacatac tctcagcacc taaatagtg ccgggaacat agtaagggt cagtaaacat ttgtgaata aactcagtct cctacattag cattctaaaa aaaaaaaaaa
EGLN3 protein(iso form 1) (SEQ ID NO: 54)	NP_00129 5032.1	mplghimrld lekialeiv pcleamvac ypgngtyvr hvdnpngdr citciylnk nwdaklhgi lrifpegksf iadvepifdr llffwsdrn phevqpsyat ryamtvwyfd aeeraeakkk frnltrktes altd
EGLN3 mRNA (isoform 2) (SEQ ID NO: 55)	NM_0220 73.3	gagtctggcc geagtcgagg cagtgggtgc tcccatccc caaaaggcgc cctccgactc cttgcgccgc actgctgcc gggccagtcc ggaaacgggt cgtggagctc cgcaccactc ccgtctgttc ccgaaggcag atccctctc ccgagagttg cgagaaactt tccctgttc ccgacgtgc agcggctcgg gtaccgtggc agccgcaggt ttctgaacc cgggccacgc tccccgcgc tcggtctgc gctcgtgtg atcgttcct ctctggtgc acgtgggga tcccgacct cgattctgc ggcgagatgc ccctgggaca catcatgagg ctggacctgg agaaaattgc cctggagtac atcgtgccct gtctgcacga ggtgggcttc tgctacctgg acaacttct gggcgagggtg gtgggcgact gcgtcctgga gcgcgtcaag cagctgcact gcaccggggc cctgcgggac ggccagtgg cggggccgcg cgcggcgctc tccaagcgac acctgcgggg cgaccagatc acgtggatcg ggggcaacga ggagggtgc gaggccatca

	Accession number	Sequence
		gcttctctct gtcctcatc gacaggctgg tectctactg cgggagccgg ctgggcaa actacgtcaa ggagaggctc aaggcaatgg tggcttgcta tccgggaaat ggaacagg atgttcgcca cgtggacaac cccaacggtg atggctgctg catcacctgc atctactatc tgaacaagaa ttgggatgcc aagctacatg gtgggacccg gcggatattt ccagagggga aatcattcat agcagatgtg gagcccattt ttgacagact cctgttcttc tggcagatc gtaggaaccc acacgaagtg cagccctctt acgcaaccag atatgctatg actgtctggt actttgatgc tgaagaaagg gcagaagcca aaaagaaatt caggaattta actaggaaaa ctgaatctgc cctcactgaa gactgaccgt gctctgaaat ctgctggcct tgttcatttt agtaacgggt cctgaattct cttaattct ttgagatcca aagatggcct cttcagtgc aacaatctcc ctgctacttc ttgcatcctt cacatccctg tcttgtgtgt ggtacttc gtttcttgcc caagactgtg ttgatctca gatactctct ttgccagatg aagtacttg ctaactccag aaattcctgc agacatccta ctggccagc ggtttacctg atagattcgg taatactatc aagagaagag cctaggagca cagcgaggga atgaacctta ctgcatctt atgtatactt cctgatttga aaggaggagg ttgaaaaga aaaaaatgga ggtgtagat gccacagaga ggcatcacgg aagccttaac agcaggaaac agagaaattt gtgtcatctg aacaatttcc agatgttctt aatccagggc tgttggggtt tctggagaat tatcacaacc taatgacatt aatacctcta gaaagggctg ctgtcatagt gaacaattta taagtgtccc atggggcaga cactcctttt ttcccagtc tgcaacctgg attttctgcc tcagcccat tttctgaaa ataagtactt tctgaataaa gatggcaaca caatttttc tccatttca gtttctacct gggaacctaa ttcccagaa gctaaaaaac tagacattag ttgtttggt tgctttgttg gaatggaatt taaatttaa tgaaaggaaa aatatatccc tggtagttt gtgttaacca ctgataactg tggaaagagc taggtctact gatatacaat aaacatgtgt gcatcttgaa caatttgaga ggggagggtg agttggaaat gtgggtgttc ctgtttttt tttttttt ttttagttt tcttttta tgagctcacc cttaacaca aaaaaagcaa ggtgatgtat ttaaaaaag gaagtggaaa taaaaaatc tcaagctat ttgagttctc gtctgtccct agcagctttt ctcagctca cttggctctc tagatccact gtggttgga gtatgaccag aatcatggaa ttgctagaa ctgtggaagc ttctactcct gcagtaagca cagatcgac tgcctcaata acttggtatt gagcacgtat ttgcaaaag ctactttcc tagttttcag tattacttct atgttttaa aatccctta atttctgtc tgaataccc atgaacatta aagagccaga aatatttcc ttgttatgt acgatatat atatatatg tctccaaga tagaagtta cttttctc ttctgggttt ggaaatttc cagataagac atgtcacat taattctcaa cgactgctct atttgtgtg acggtaatg ttatcacctt ctaaattact atgaattta ttcacttatt atgtttatg tctgtatcc tttctctgga

	Accession number	Sequence
		gtgtaagcac aatgaagaca ggaattttgt atatttttaa ccaatgcaac atactctcag cacctaaaat agtgccggga acatagtaag ggctcagtaa atactgttg aataaactca gtctcctaca ttagcattct aa
EGLN3 protein (isoform 2) (SEQ ID NO: 56)	NP_07135 6.1	mplghimrld lekialeiv pclhevgfey ldnflgevvg dcvlervkql hetgalrdgq lagpragvsk rhlrqdqitw iggneegcea isflslidr lvlycgsrlg kyyvkerska mvacypgngt gyvrhvdnpn gdgreitciy ylnknwdakl hggilrifpe gksfiadvep ifdrllffws drmphevqp syatryamtv wyfdaeerae akkkfnlnt ktesalted
SLC6A3 mRNA (SEQ ID NO: 57)	NM_0010 44.4	cgtgcggag cgggagggga ggcttcgcgg aacgctctcg gcgccaggac tcgctgcaaa agcccaggcc cgggaggcca gaccaagagg gaagaagcac agaattctc aactcccagt gtgccatga gtaagagcaa atgctcctg ggactcatgt ctccgtggt ggccccggct aaggagccca atgccgtggg ccgaaggag gtggagctca tcctgtcaa ggagcagaac ggagtgcagc tcaccagctc caccctacc aaccgcggc agagccccgt ggaggccag gatcgggaga cctggggcaa gaagatcgac ttctcctgt ccgtcattgg ctttgctgtg gacctggcca acgtctggcg gtccctac ctgtgtaca aaaatgggtg cggtgccttc ctgtccct acgtctctt catgtcatt gctggatgc cactttcta catggagctg gccctggcc agttcaacag ggaaggggcc gctggtgtct ggaagatctg cccatactg aaaggtgtg gttcacggt catctcctc tcatgtatg tcggttctt ctacaacgtc atcatgcctt gggcgctgca ctatctctc tctcctca ccacggagct ccctggatc cactgaaca actcctggaa cagcccaac tgctcgatg cccatcctgg tgactccagt ggagacagct cgggcctcaa cgacatttt gggaccacac ctgtgccga gtactttgaa cgtggcgtgc tgcacctca ccagagccat ggcatcgac acctggggcc tccgcgtgg cagctcacag cctgcctggt gctgtgcatc gtgtgtctct actcagcct ctggaagggc gtgaagacct cagggaaggt ggtatggatc acagccacca tgccatactg ggtctcact gccctgctc tgcgtgggt caccctcct ggagccatag acggcatcag agcatactg agcgttgact tctaccggct ctgcaggcg tctgttga ttgacggcg caccaggtg tgcttctcc tggcggtgg gttcgggtg ctgatecct tctcagcta caacaagttc accaacaact gctacaggga cgcgattgtc accacctca tcaactcct gacgagttc tctcgggt tctgtctt ctctctctg gggtacatg cacagaagca cagtgtccc atcggggacg tggccaagga cgggccagg ctgatctca tcatctacc ggaagccatc gccacgtcc ctctgtctc agcctgggc gtgtctct tcatcatgt gtcacctg ggatcgaca gcgcatggg tggatggag tcatgtatc cgggctcat cgtatgttc cagctgtgc acagacacc tgagctctc acgtctca tctcctgc gacctctc

	Accession number	Sequence
		ctgtccctgt tctgcgtcac caacgggtggc atctacgtct tcacgtctct ggaccatttt gcagccggca cgtccatcct ctttgagtg ctcacgaag ccatcggagt ggccctggtc tatggtgttg ggcagttcag cgacgacatc cagcagatga ccgggcagcg gccacgctg tactggcggc tgtgctggaa gctggtcagc cctgcttct tctgttctgt ggctggtgct agcattgtga ccttcagacc ccccactac ggagcctaca tctccccga ctgggccaac gcgtgggct gggtcatcgc cacatctcc atggccatgg tgcccatcta tgcggcctac aagttctgca gcctgcctgg gtccttctga gagaaactgg cctacgcat tgcacccgag aaggaccgtg agctggtgga cagaggggag gtgcgccagt tcacgtccg cactggctc aaggtgtaga gggagcagag acgaagacc caggaagtca tctgcaatg ggagagacac gaacaaacca aggaaatcta agtttcgaga gaaaggagg caactctac tctcaacct ctactgaaa cacaacaac aaagcagaag actcctctct tctgactgt tacaccttc cgtgccggga gcgcacctg ccgtgtcttg tttgctgta ataacacgt agatctgtc agcgaggctc acccggtgt gtccctgca gggcagaaaa acgtctaact tcactgtct tgtgtgaggc tcctccctc cctgctcct gctcccgct ctgaggtgc ccaggggga ctgtttctc aggcggggat cagatcctt gtagacgac ctgtgagaa tccccgtct cacagtagct tctagacca ttactttgc ccatattaaa aagccaagt tctgcttg tttagctgtg cagaaggta aatggaggaa accacaaatt catgcaaagt ctttccga tgcgtggctc ccagcagagg ccgtaattg agcgttcagt tgacacattg cacacacgt ctgttcagag gcattggagg atgggggtcc tggatgtct caccaggaaa ttctgttat gtttctcag cagagagaaa taaaactct tgaaccagc tcaggctact gccactcagg cagcctgtgg gtccttgagg ttaggggaac ggcctgagag gagcgtgtcc tatccccga cgcatgcagg gccccacag gagcgtgtcc tatccccga cgcatgcagg gccccacag gagcatgtcc taccctgga cgcatgcagg gccccacag gagcgtgtac taccagaa cgcatgcagg gccccacag gagcgtgtac taccagaa cgcatgcagg gccccactg gagcgtgtac taccagaa cgcatgcagg gccccacag gagcgtgtcc tatccccga ccggacgcat gcaggggccc cagaggagcg tgtactacc caggacgcat gcaggggccc cacaggagcg tgtactacc caggatgcat gcaggggccc cagaggagcg tgtactacc caggacgcat gcaggggccc catgcaggca gcctgcagac cacactctgc ctggcctga gccgtgacct ccaggaagg accccactgg aattttatt ctctcagtg cgtgccacat caataacaac agttttatg ttgcgaatg gctttttaa atcatattt cctgtgaatc aaaacaaat caagaatgca gtatccgca gcctgctgc tgatattgca gttttgtt acaagaataa ttagcaatac tagtggaagg atgttgcca aaagctgctt tcatggcac actgccctct gccactgaca ggaaagtga tgccatggt tgaattcat cctcaagtcg

	Accession number	Sequence
		gtgggcctgc ctacgtgctg cccgagggca ggggccgtgc agggccagtc atggctgtcc cctgcaagtg gacgtgggct ccagggactg gagtgtaatg ctcggtggga gccgtcagcc tgtgaactgc caggcagctg cagttagcac agaggatggc tccccattg cttctgggg agggacacag aggacggctt ccccatcgcc ttctggccgc tgcagtcagc acagagagcg gttcccat tgccttctgg ggaggacac agaggacagc tccccatg cttctggct gtgcagtc gcacagagag cggttcccc atgccttct ggggaggggc tccgtgtagc aaccaggtg ttgtccgtgt ctgttgacca atctctatc agcatcgtgt gggtccctaa gcacaataaa agacatccac aatggaaaaa ctgcaaaaaa aaaaaaaaaa aa
SLC6A3 protein (SEQ ID NO: 58)	NP_00103 5.1	msskscsvgl mssvvpake pnavgpkeve lilvkeqngv qltsstlnp rqpveaqdr etwgkkidfl lsvigfavdl anvwrpyle ykngggaflv pyllfmviag mplfymelal gqfnregaag vwkiqipilkg vgftvilisl yvgffynvii awalhyfss fttelpwihe nnswnspncs dahpgdssgd ssglndtfgt tpaaeyferg vlhlhqshgi ddlgpprwql taclvlvivi lyfslwkgvk tsgkvvita tmpyvvtal llrgvtlpga idgiraylsv dfyrlceasv widaatqvcf slgvvgfvli afssynkftn ncyrdaiytt sinsltsfss gfvvvsflgy maqkhsvpig dvakdgppli fiiypeaiat lplssawavv ffimlltgi dsamggmesv itglidefql lhrhrelftl fivlatflls lfcvtnggiy vftldhfaa gtsilfgvli caigvawfyg vgqfsddiqq mtgqrpslyw rlcwklvpsc flfvvvvsi vtfrpphyga yifpdwanal gwviatssma mvpiyaaykf csllpgsfrek layaiapekd relvdrgevr qftlrhwkv

Table 5: A list of genes and proteins, and their accession numbers.

[00150] The foregoing examples are presented for the purpose of illustrating the invention and should not be construed as imposing any limitation on the scope of the invention. It will readily be apparent that numerous modifications and alterations may be made to the specific embodiments of the invention described above and illustrated in the examples without departing from the principles underlying the invention. All such modifications and alterations are intended to be embraced by this application.