Journal of proteome research

Article

Metabolic Footprinting of a Clear Cell Renal Cell Carcinoma *in vitro* Model for Human Kidney Cancer Detection

María Elena Knott, Malena Manzi, Nicolás Zabalegui, Mario O. Salazar, Lydia I. Puricelli, and Maria Eugenia Monge

J. Proteome Res., **Just Accepted Manuscript** • DOI: 10.1021/acs.jproteome.8b00538 • Publication Date (Web): 27 Sep 2018

ACS Publications

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Metabolic Footprinting of a Clear Cell Renal Cell Carcinoma *in vitro* Model for Human Kidney Cancer Detection

María Elena Knott[1⊥], Malena Manzi[1⊥], Nicolás Zabalegui[1], Mario O. Salazar[2], Lydia I. Puricelli[3], María Eugenia Monge[1*]

[1]Centro de Investigaciones en Bionanociencias (CIBION), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2390, C1425FQD, Ciudad de Buenos Aires, Argentina

[2]Farmacognosia, Departamento de Química Orgánica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario S-2002LRK, Santa Fé, Argentina.

[3]Instituto de Oncología Ángel H. Roffo, Facultad de Medicina, Universidad de Buenos Aires, Av. San Martín 5481, C1417DTB, Ciudad de Buenos Aires, Argentina.

[⊥]These authors contributed equally to this work.

*Corresponding author.

E-mail: maria.monge@cibion.conicet.gov.ar

Ph: +54 11 4899 5500 (ext. 5614)

Abstract

A protocol for harvesting and extracting extracellular metabolites from an *in vitro* model of human renal cell lines was developed to profile the exometabolome by means of a discovery-based metabolomics approach using ultraperformance liquid chromatography coupled to quadrupole-time-of-flight mass spectrometry. Metabolic footprints provided by conditioned media (CM) samples (n=66) of two clear cell Renal Cell Carcinoma (ccRCC) cell lines with different genetic background and a non-tumor renal cell line, were compared with the human serum metabolic profile of a pilot cohort (n=10) comprised of stage IV ccRCC patients and healthy individuals. Using a cross-validated orthogonal projection to latent structures-discriminant analysis model, a panel of 21 discriminant features selected by iterative multivariate classification, allowed differentiating control from tumor cell lines with 100% specificity, sensitivity and accuracy. Isoleucine/leucine, phenylalanine, N-lactoyl-leucine, and N-acetyl-phenylalanine, and cysteinegluthatione disulfide (CYSSG) were identified by chemical standards, and hydroxyprolyl-valine was identified with MS and MS/MS experiments. A subset of 9 discriminant features, including the identified metabolites except for CYSSG, produced a fingerprint of classification value that enabled discerning ccRCC patients from healthy individuals. To our knowledge, this is the first time that N-lactoyl-leucine is associated to ccRCC. Results from this study provide a proof of concept that CM can be used as a serum proxy to obtain disease-related metabolic signatures.

Keywords: *in vitro* cell culture, conditioned media, metabolomics, ultraperformance liquid chromatography-mass spectrometry, clear cell Renal Cell Carcinoma, metabolic footprinting.

Introduction

Kidney cancer is fundamentally a disease of dysregulated cellular metabolism.[1] Renal cell carcinoma (RCC), originated from the renal epithelium, accounts for >90% of cancers in the kidney,[2] and is among the 10 most common cancers in both men and women worldwide.[3, 4] RCC patients are often incidentally diagnosed by imaging procedures. Only 10% of patients exhibit the classic triad of hematuria, low back pain, and flank mass symptoms, and nearly 40% of patients lack all of these and present with systemic symptoms, including weight loss, abdominal pain, anorexia, and fever.[5] More than 30% of the patients exhibit locally advanced or metastatic RCC at the time of diagnosis.[6, 7] Although the disease is inherently resistant to chemotherapy[8] and radiotherapy,[9] the survival of advanced RCC patients has improved significantly with the advent of tyrosine-kinase inhibitors as standard of care therapy.[10] However, the discovery of early detection biomarkers is an important priority area to give more opportunities for early intervention and improved outcome of ccRCC patients.

Clear cell RCC (ccRCC) is the most common (75%) histological subtype and accounts for most cancer-related deaths.[2, 11] At molecular level, 50-80% of all ccRCC patients show mutations in *Von Hippel Lindau* (VHL) gen which is involved in hypoxia inducible factor 1α (HIF 1α) expression.[3, 12, 13] Overexpression of HIF 1α triggers the transcription of genes involved in glucose metabolism.[14] In addition, ccRCC is a lipogenic tumor with abnormal cholesterol metabolism.[15, 16] In this context, and due to complex pathways contributing to kidney cancer progression, a single molecular marker might not be efficient enough as tumor biomarker, suggesting the need of a multiple biomarker panel to achieve sufficient clinical information.

Mammalian cell metabolomics[17, 18] has emerged as a promising tool for studying cellular biochemistry and investigate altered metabolic networks that contribute to cell proliferation,

dissemination and survival in RCC. Metabolomics uses a holistic approach to characterize and quantify the metabolome, comprised of all the small molecules (MW<1500) in biological systems.[19, 20] Different metabolic fingerprinting studies have shown alterations associated with RCC. Mass spectrometry (MS)-based urine untargeted metabolomics approaches have suggested alterations in metabolic pathways of tryptophan[21] and acylcarnitines.[22] Weiss and collaborators showed that serum is a more accurate proxy for tissue changes than urine, using a mouse xenograft model of kidney cancer, and suggested that tryptophan degradation is highly represented in RCC.[23] Untargeted serum MS-based metabolomics studies from RCC patients have also suggested disease-related alterations in the phospholipid catabolism, sphingolipid, cholesterol, phenylalanine, and arachidonic acid metabolisms in addition to fatty acid beta-oxidation.[23, 24] Recently, tumor progression and metastasis have been associated with metabolite increases in glutathione and the cysteine/methionine pathways by means of a metabolomic profiling study on 138 matched ccRCC/normal tissue pairs.[25]

In vitro cell models are of particular interest for understanding the metabolism of cellular processes, and allow the study of both intracellular (fingerprint) and extracellular (footprint) metabolic profiles,[17, 18, 26] being the latter a closer proxy of serum. Cell lines can be used as *in vitro* models for research including biomarker discovery studies and the evaluation of new drugs for treatment,[17] as recommended by the Food and Drug Administration.[27] Recent genomic studies have identified molecular differences between commonly used renal cancer cell lines and human tumor samples.[28] However, studies comparing the exometabolome of kidney cell lines with the human serum metabolome to evaluate the feasibility of using *in vitro* models for serum sample classification have not been reported up to date.

In this study, we have optimized a protocol for harvesting, extracting, lyophilizing and reconstituting conditioned media (CM) metabolites derived from two human ccRCC cell lines 786-O ($VHL^{-}/^{-}$) and Caki-1 ($VHL^{+}/^{+}$), and the non-tumor human renal cell line HEK-293 (n=22 for each cell line); and we have profiled the exometabolome using a discovery-based metabolomics approach by means of ultraperformance liquid chromatography coupled to quadrupole-time-of-flight mass spectrometry (UPLC-QTOF-MS). Metabolic features (Rt, *m/z* pairs) were analyzed using a cross-validated orthogonal projection to latent structures-discriminant analysis model, coupled to a genetic algorithm variable selection method. A panel of 21 discriminant features, obtained from the binary comparison of control and tumor cells, allowed sample classification with 100% specificity, sensitivity and accuracy. In addition, 9 of these compounds were present in human serum samples and enabled discriminating stage IV ccRCC patients from healthy individuals, which could potentially be relevant in kidney cancer diagnosis.

Materials and Methods

Chemicals

LC-MS grade acetonitrile, methanol, isopropanol, acetic acid and formic acid purchased from Fisher Chemical (NC, USA) and ultrapure water with 18.2 MΩ·cm resistivity (Thermo Scientific Barnstead Micropure UV ultrapure water system, USA) were used to prepare chromatographic mobile phases and solutions. Leucine enkephalin was purchased from Waters Corp. (Milford, MA, USA). Sodium hydroxide was purchased from EMSURE® ISO (Merck Millipore, Burlington, MA, USA). Roswell Park Memorial Institute (RPMI) 1640 and

Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F12) powder culture media were purchased from GIBCO™ (Thermo Fisher Scientific, MA USA). Fetal bovine serum (FBS) was purchased from Internegocios S.A. (Buenos Aires, Argentina), gentamicin from Laboratorio Drewer S.A. (Buenos Aires, Argentina), and L-glutamine from GIBCO™ (Thermo Fisher Scientific, MA USA). The analgesic mix comprised of the following standards: acetaminophen, 2-acetaminophen, acetanilide, acetylsalicylic acid, caffeine, phenacetin and salicylic acid was purchased from Waters (Waters Corporation, Manchester, UK). L-cysteine-glutathione disulfide (purity ≥95%) was purchased from Cayman Chemical (MI, USA). L-leucine (≥98%) and L-phenylalanine (≥98%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). L-isoleucine (≥98%) was purchased from Fluka (Steinheim, Germany). N-acetyl-phenylalanine and N-lactoyl-leucine reference compounds were chemically synthesized in our laboratory (Supporting Information, Figures S1 and S2).

In Vitro Model - Cell Culture

The *in vitro* model was comprised of three different kidney cell lines: HEK-293, 786-O and Caki-1, which were obtained from the American Type Culture Collection (ATCC). HEK-293, a non-tumor human embryonic kidney cell line, was used as control and was compared to two ccRCC cell lines with different genetic background, 786-O and Caki-1. The 786-O cell line derives from primary ccRCC tumor and has a deletion of a gene encoding the VHL protein ($VHL^{-/-}$). The Caki-1 cell line arises from ccRCC skin metastasis, and expresses wild type VHL protein ($VHL^{+/+}$). HEK-293 and 786-O cell lines were cultured in RPMI culture media and Caki-1 cell line was cultured in DMEM/F12 media; all supplemented with 10% FBS, 2 mM L-glutamine, and 8 μg mL^{-1} gentamicin, in a humidified atmosphere of 5% CO_2 at 37 ˚C. Cell cultures were routinely checked for mycoplasma contamination by DAPI (2-(4-Amidinophenyl)-

1H-indole-6-carboxamidine) staining. The protocol designed for cell culture and CM harvesting generated 24 samples for each cell line (Figure 1, Table S1). Briefly, two cryovials with similar cell passage number were thawed for each cell line. Cells from each cryovial were split into two flasks (A and B in Figure 1) and were treated independently during the experiments. After cell counting, two wells were plated for each cell line. Culture media blanks were obtained by incubation of cell culture media with the same protocol used for cell lines (Figure 1). Two CM samples from each cell line were used for analytical method development. Thus, 22 samples per cell line were used for the untargeted metabolic footprinting study.

In vitro Model - Conditioned Media Collection

In each well of 9.5 cm^2 area, 5.00×10^5, 1.00×10^5, and 1.25×10^5 cells were plated for HEK-293, 786-O, and Caki-1 cell lines, respectively. Once 80% confluence was reached, cells were gently washed 3 times during 10 minutes with the corresponding culture media, without FBS. Then, cell monolayers were incubated overnight with 800 µL of their respective cell culture media without FBS (starving conditions).[29] CM samples and culture media blanks were collected, centrifuged at 300 × *g* for 5 minutes to remove cell debris, and finally supernatants were collected and immediately frozen at -80 ºC.

Serum Samples

Serum samples and associated clinical data were provided by the public oncologic serum biobank "Biobanco Público de Muestras Séricas Oncológicas" (BPMSO) from "Instituto de Oncología A. H. Roffo" (IOAHR), Buenos Aires, Argentina. The patient cohort consisted of 5 healthy individuals (age range 40−64, mean (SD) age 56(9) years, 20% male) and 5 ccRCC patients at stage IV (SIV) (age range 53−72, mean age 64(6) years, 100% male). At the 0.05

level, the means of the age populations were not significantly different with the two-sample t-test. Blood samples were drawn from untreated cancer patients. Serum sample collection followed the guidelines approved by the IOAHR Institutional Review Board, and samples were drawn after signature of the corresponding informed consent. According to the BPMSO standard operating procedure, 20 mL of blood were collected in tubes without any anticoagulant and left 15 minutes at 25 °C to allow the clot formation and centrifuged at 600 × *g* for 10 minutes. Then, serum was split into aliquots and stored at -80 °C. Serum aliquots were used only once after thawing.

Conditioned Media and Serum Sample Preparation

All frozen samples were thawed at 0 ºC on a water-ice bath. Protein precipitation was performed by addition of cold (4 ºC) isopropanol to 500 µL of CM or 60 µL of serum samples, in a 3:1 solvent:sample volume ratio. Samples were vortex-mixed for 10 seconds and centrifuged at 16000 × *g* for 20 minutes and 4 ºC. Supernatants were frozen and lyophilized at -80 °C and 50 mTorr for 48 hours using a Telstar LYOQuest-85 freeze dryer (Telstar, Madrid, Spain). Sample residues were reconstituted in water/methanol (80:20 v/v) with a concentration factor of 7 for CM samples, and 1 for serum samples; and analyzed by UPLC-QTOF-MS. Sample preparation blanks containing ultrapure water and culture media blanks also went through the same sample preparation procedure. CM and serum samples were analyzed in different batches, with quality control (QC) and quality assurance (QA) samples to verify the stability of retention times, peak shapes and areas during the analysis.[30] QCs consisted in randomly pooled CM samples from the 3 cell lines studied and were processed in an identical approach as samples; i.e., a small aliquot of a subset of the CM samples were pooled into a single QC sample, followed by protein precipitation and further sample preparation steps. All CM samples were used to prepare QCs.

For each batch a new QC sample was reconstituted and analyzed every 14 sample injections. Samples were randomly analyzed by UPLC-QTOF-MS together with QCs, solvent blanks, sample preparation blanks and culture media blanks. Additionally, for serum samples, QA was verified with the periodical injection of a mix of standards comprised of acetaminophen, 2-acetaminophen, acetanilide, acetylsalicylic acid, caffeine, phenacetin and salicylic acid. CM samples were analyzed along 5 consecutive days and serum samples were analyzed in one day, after conditioning the analytical platform.

Ultraperformance Liquid Chromatography-Mass Spectrometry

Ultraperformance liquid chromatography-mass spectrometry (UPLC-MS) analyses were performed using a Waters ACQUITY UPLC I Class system fitted with a Waters ACQUITY UPLC BEH C_{18} column (2.1×100 mm, 1.7 μm particle size, Waters Corporation, Milford, MA, USA), coupled to a Xevo G2S QTOF mass spectrometer (Waters Corporation, Manchester, UK) with an electrospray ionization (ESI) source. The typical resolving power and mass accuracy of the Xevo G2S QTOF mass spectrometer were 32,000 FWHM and 0.2 ppm at *m/z* 554.2615, respectively. Gradient elution was utilized in the chromatographic separation method using water with 0.1% acetic acid (mobile phase A) and acetonitrile (mobile phase B), with the following program: 0-1 min 10% B; 1-2.5 min 10-15% B; 2.5-4 min 15-22% B; 4-6 min 22-38% B; 6-9 min 38-65% B; 9-12 min 65-80% B; 12-16 min 80-100% B; 16-19 min 100% B. The flow rate was constant at 0.25 mL min^{-1} for 12 min and was increased to 0.30 mL min^{-1} between 12 and 19 min. After each sample injection, the gradient was returned to its initial conditions in 11 min. The injection volume was 5 μL for CM and 2μL for serum samples. The column and autosampler tray temperatures were set at 35 and 5 °C, respectively. The mass spectrometer was operated in negative ion mode with a probe capillary voltage of 2.3 kV, and a sampling cone

voltage of 30.0 V. The source and desolvation gas temperatures were set to 120 and 300 °C, respectively. The nitrogen gas desolvation flow rate was 600 L h^{-1}, and the cone desolvation flow rate was 10 L h^{-1}. The mass spectrometer was daily calibrated before each batch analysis across the range of *m/z* 50-1200 using a 0.5 mM sodium formate solution prepared in isopropanol/water (90:10 v/v). Data were drift corrected during acquisition using a leucine enkephalin (*m/z* 554.2615) reference spray infused at 2 µL min^{-1}, every 45 seconds. Data were acquired in MS^E continuum mode[31] in the range of *m/z* 50-1200, and the scan time was set to 0.5 s. Technical duplicates were acquired in all cases. For UPLC-MS/MS experiments, the product ion mass spectra were acquired with collision cell voltages between 10 and 30 V, depending on the analyte. Ultra-high-purity argon (≥99.999%) was used as the collision gas. Data acquisition and processing were carried out using MassLynx version 4.1 (Waters Corp., Milford, MA, USA). The mass spectrometry data have been deposited to the MetaboLights public repository (https://www.ebi.ac.uk/metabolights/index) with the data set identifier MTBLS737.

Data Analysis

Spectral features (retention time (R_t), *m/z* pairs) were extracted from UPLC-QTOF-MS data using Progenesis QI version 2.1 (Nonlinear Dynamics, Waters Corp., Milford, MA, USA). The procedure included retention time alignment, peak picking, deisotoping, integration, and grouping together adducts derived from the same compound. Subsequently, if a feature had a peak area in a CM sample that was 3-fold or less than the mean peak area in the solvent and sample preparation blanks of the same feature, then its peak area was set to 0.[32] Otherwise, the mean peak area in those blanks was subtracted from the feature peak areas in the CM samples. After blank subtraction, feature abundances from technical duplicates were averaged, and only those that were present in at least 80% of one group class were retained. Since Caki-1 cell lines

were cultured with a different culture medium, only features from HEK-293 or 786-O were considered to build the feature matrix for multivariate statistical analysis to avoid influence of culture media on the detected exometabolome (cell secretome).[33] Chromatographic peak shape and signal intensity of each feature was further evaluated for data curation. Features with signal intensity $<10^3$ in the continuum mass spectra or with a mass difference larger than 10 mDa along the sample list were discarded. The feature matrix obtained after this procedure was normalized to the number of plated cells for each cell line and to the total peak area for each sample. This matrix (Data Set S1 in the Supporting Information) was then utilized to build unsupervised and supervised multivariate statistical analysis models using MATLAB R2012b (The MathWorks, Natick, MA, USA) with the PLS Toolbox version 8.1 (Eigenvector Research, Inc., Manson, WA, USA). Principal component analysis (PCA) was used to track data quality, reduce the dimensionality, identify and remove outliers in the dataset, as well as to identify sample clusters. Two samples were identified as outliers by PCA (data not shown), one from a HEK-293 CM and another from 786-O CM, and were not further considered for data analysis. Orthogonal projection to latent structures-discriminant analysis (oPLS-DA)[34, 35] coupled with a genetic algorithm (GA) variable selection method was applied to find a feature panel that maximized classification accuracy for the binary comparison of HEK-293 and 786-O. A panel of 23 discriminant features had the lowest root-mean-square error of cross-validation (RMSECV) at the conclusion of the GA variable selection process; however, only 21 features were retained for statistical analysis (see *Metabolite Identification Procedure)*. The parameters for genetic algorithm variable selection were as follows: population size: 112, variable window width: 1, % initial terms (variables): 10, target minimum # of variables: 1, target maximum # of variables: 30, penalty slope: 0.05, maximum generations: 100, % at convergence: 80, mutation rate: 0.005,

crossover: double, regression choice: PLS, # of latent variables: 2, cross-validation: random, # of splits: 4, # of iterations: 10, replicate runs: 20. The oPLS-DA model was cross-validated using venetian blinds with 10 data splits. Data were preprocessed by autoscaling prior to PCA or oPLS-DA analysis. PCA was also performed to inspect data before and after GA variable selection (*i.e.,* on the curated spectral feature matrix and on the discriminant feature panel). Feature abundances of QC samples for each batch were averaged before PCA.

Fold changes were calculated as the ratio of median peak areas between CM from 786-O and HEK-293, Caki-1 and HEK-293, and Caki-1 and 786-O samples. Mann-Whitney U tests were used to calculate statistical significance, and *p* values were corrected using the Benjamini-Hochberg[36] procedure for multiple comparisons with a FDR of 0.1.

Metabolite Identification Procedure

Metabolite identification was attempted for the 23 discriminant features resulting from the GA variable selection process. Mass spectral adduct ions and elemental formula generated based on accurate masses and isotopic patterns, were searched against the Human Metabolome Database (HMDB).[37] Based on the list of tentative candidates, metabolite identification was performed considering the accurate mass, isotopic pattern, and the fragmentation pattern obtained from tandem QTOF-MS experiments. Tandem MS spectra were compared to the Metlin database,[38] and MassBank,[39] and for cases where MS/MS spectra were not available in databases, fragmentation patterns were manually interpreted for metabolite annotation. Putative identifications were validated with chemical standards. Though Progenesis performed adduct grouping, the software did not account for NaCl adduct ions, which were evidenced by isotopic pattern analysis of the selected features. Out of the 23 features, 2 features were identified as different NaCl adduct ions of the same compound; i.e. different $[M+x\text{NaCl-H}]^-$ ionic species, and

one of the features was identified as paraben, which is a non-endogenous metabolite. Therefore, only 21 features were considered for further statistical analysis. Chemical standards were prepared in ultrapure water, and were analyzed under identical conditions as CM and serum samples to validate putative metabolite identities by chromatographic retention time and MS/MS fragmentation pattern matching. Spiking experiments were also conducted with the standards on CM and serum samples as well as on cultured media blanks, to address retention time differences caused by matrix effects.

Results and Discussion

Metabolic Footprint & CM Sample Classification

A protocol for harvesting and extracting extracellular metabolites from an *in vitro* model of 3 different human renal cell lines was developed (Figure 1, Table S1) and implemented to profile the cell secretome[33] with a UPLC-QTOF-MS-based method. In addition, serum samples from 5 healthy individuals and 5 SIV ccRCC patients were analyzed with the same optimized analytical method to evaluate the use of CM as a serum proxy to obtain a disease-related metabolic signature from the detected metabolomes.

A total of 6002 features (Rt, *m/z* pairs) were extracted by Progenesis software from the UPLC-MS negative ion mode data from the three studied cell lines. The number of features was reduced to 5030 after deconvolution, to group together adducts derived from the same compound. Following solvent blank deduction, feature abundances from technical replicates were averaged, and only features that were present in at least 80% of one group class were retained. Since Caki-1 cell line was cultured with a different culture media, features that were

only present in Caki-1 CM were excluded for further analysis to avoid bias in sample classification by the media composition.[33] This data set, comprised of 2358 features, was further manually filtered to keep only those features with signal intensity, peak shape and mass variance within the established thresholds that would allow accurate identification, leaving 755 spectral features. This matrix was normalized (Data Set S1 in the Supporting Information) and utilized to build a PCA model. The 2D score plot illustrated in Figure 2A shows distinguishable separation between the three classes, mainly achieved by the contribution of the first PC. In addition, QC data points clustering around the origin of the PCA score plot indicates reproducibility in the sample preparation method, high data quality and adequate performance of the analytical platform. The PCA model consisted of 5 PCs with 70.67% total captured variance, with the first two PCs accumulating 47.44% of the total variance. Figure 2B shows the score plot for CM samples from the non-tumor control cell line HEK-293 and the tumor cell line 786-O, which were cultured and incubated with the same culture media. The model, which consisted of 2 PCs with 53.38% total captured variance, provided a clear degree of class separation, mainly achieved by the contribution of the first PC. The loadings plot associated to PC1 showed similar contribution from all spectral features to sample separation in the score plot (figure not shown). Thus, sample discrimination was further analyzed by means of oPLS-DA coupled to a GA variable selection method to find a reduced metabolic feature set that would allow sample classification and class membership prediction. A panel of 23 metabolic features with the lowest RMSECV was selected through the GA process; although 2 of them were removed for further statistical analysis (see *Metabolite Identification Procedure*). Figure 2C shows the cross-validated prediction plot using the 21 discriminant metabolic features. The model consisted of 1 latent variable that interpreted 50.55% and 99.47% variance from the X- (feature peak areas) and

Y- (class membership) blocks, respectively. This oPLS-DA model resulted in 100% cross-validated accuracy, sensitivity, and specificity; therefore, no CM samples were misclassified.

Comparison of in vitro Exometabolome with Serum Metabolome

Figure S3 shows the different base peak intensity (BPI) chromatograms obtained for CM samples of each cell line, and for serum samples of a healthy individual and a SIV ccRCC patient. Differences observed were probably associated to metabolite concentration levels, and matrix effects. Additionally, human serum metabolic profiles are influenced by the presence of the disease, lifestyle, diet, environmental exposures, i.e. the exposome,[40, 41] and molecules derived from the interactions with associated microorganisms (the microbiome),[42] demonstrating the need for multivariate statistical analysis to extract a disease-related metabolic signature from the detected metabolomes.

To evaluate the ccRCC *in vitro* exometabolome as a good proxy to study metabolic changes in serum, the 755 features of the CM matrix were searched in the feature matrix extracted from human serum sample analysis. A total of 163 features were found to be common to both CM and sera. In addition, out of the 21 metabolic features that perfectly classified CM samples of the renal *in vitro* model, 9 were present in serum samples. These feature sets were further utilized to build unsupervised models to explore sample clustering in both types of biological samples, i.e. CM and sera. Figures 3A and 3B show the 2D PCA score plots of serum samples using the 163 common feature set and the smaller subset of 9 common features from the CM discriminant panel, respectively. Using the former set, the PCA model composed of 2 PCs with 57.97% total captured variance did not show sample separation (Figure 3A). However, the PCA model based on the subset of 9 common CM discriminant features utilized 2 PCs that captured a larger percentage of total variance (70.87%) after dimension reduction, and was able

to provide a good degree of sample separation between SIV ccRCC patients and healthy individuals (Figure 3B).The overlapping of two samples from both classes in the score plot was not associated to gender, even if the latter was not balanced between sample classes. Therefore, the metabolic footprint of the ccRCC *in vitro* model assisted in differentiating human individuals based on the presence of disease.

The PCA model that used the 163 common features provided similar class separation in CM samples (Figure 3C) as the one based on 755 features (Figure 2A). The first PC of the scores plot allowed sample separation between tumor (786-O and Caki-1) and non-tumor (HEK-293) cell lines, whereas the second PC provided separation between 786-O and Caki-1 cell lines. Interestingly, an improved CM class separation with larger captured variance (65.02%) was observed in the score plot of the PCA model using the subset of 9 discriminant features (Figure 3D) compared with 755 and 163 features (Figure 2A and 3C, respectively). Actually, this reduced panel provided an improved class separation for the binary comparison of HEK-293 vs. 786-O cell lines (Figure S4A) than with 755 features (Figure 2B) and also a clear separation between tumor cells with different genetic background (Figure S4B). Considering that these 9 features provide a good simplified ccRCC metabolic signature, and allow serum sample classification, their analysis could potentially offer useful information in studies of diagnosis purposes, and drug discovery for cancer treatment.

Discriminant Metabolite Identification

Metabolite identification was attempted for the 21 features of the discriminant panel, as illustrated in Figure 4 for N-lactoyl-leucine (N-Lac-Leu). First, extracted ion chromatograms (EICs) and the corresponding mass spectra were obtained for each feature in CM and serum samples. Mass spectral detected adduct ions and elemental formula of possible candidates were

searched in the HMDB[37] database. Subsequently, fragmentation patterns obtained from tandem MS experiments were compared to MS/MS spectra in the HMDB[37] or Metlin database,[38] or manually interpreted. Finally, the tentatively identified metabolites were confirmed by matching retention times and fragmentation patterns in CM and serum samples with chemical standards, whenever possible. Chemical standards also aided to discard non matches with putatively identified compounds, leaving several features with no ID (Table 1). The identity of some candidate molecules could not be confirmed due to i) co-elution with similar molecular weight compounds that interfered in the quadrupole selection process of the precursor ion, providing product ion overlap with the target feature in the tandem MS spectra, ii) insufficient precursor ion intensity for MS/MS experiments, or iii) limitations associated to metabolite databases.

The isotopic pattern analysis of the discriminant features suggested the presence of several ionic species being the product of non-specific binding caused by the matrix composition, which occurs when chemical species that are trapped in the same ionization droplets start to interact during solvent evaporation in the ESI mechanism.[43, 44] Out of the 21 features, 6 were identified as $[M+x\text{NaCl-H}]^-$ ionic species, with x between 1and 7, for different compounds (M), which were not accounted by Progenesis software. These type of adduct ions were detected for N-acetyl-phenylalanine (N-Ac-Phe); phenylalanine (Phe); isoleucine/leucine (Ile/Leu), and for hydroxyprolyl-valine (OHPro-Val) (Table 1, Figures S5-S8). Since the chemical composition of the culture media utilized to incubate the cell lines favored this type of nonspecific ion pairing, spiking experiments were conducted with the standards on CM and serum samples, as well as on cultured media blanks, to validate metabolite identity, and address retention time shifts caused by matrix effects (Figure 4, Figures S5-S7). The salt content of CM samples due to the culture media formulation and the concentration factor optimized for

metabolite detection (7 for CM samples vs. 1 for serum samples), was actually translated, in some cases, into 3-5 seconds shorter retention times compared to metabolites detected in serum samples or in standard solutions (Figure 4 and Figures S5-S7).

Matrix composition mainly affected Ile/Leu identification, since the chromatographic gradient implemented for sample analysis was not able to separate these isomers, but instead provided partial resolution for a Ile+Leu standard aqueous solution (Figure S7A, S7B). A different chromatographic initial gradient was utilized in combination with different spiking experiments to illustrate the presence of both Ile and Leu in CM samples, as expected from the culture media composition (Figure S7B). Since neither the chromatographic separation nor tandem MS experiments allowed identifying the discriminant amino acid (Figure S7D), for further discussion both isomers will be considered together as Ile+Leu.

Both $[M-H]^-$ and $[M+NaCl-H]^-$ adduct ions were detected for M = OHPro-Val. To distinguish between the two possible isomers of this dipeptide, i.e., OHPro-Val or valyl-hydroxyproline (Val-OHPro), tandem MS experiments were conducted both in negative and positive ESI modes, and the product ions detected in the corresponding mass spectra suggested OHPro-Val being the discriminant metabolite (Figure S8). Since no chemical standard was analyzed for this compound, its identification was given a different confidence level (Table 1). For cysteine glutathione disulfide (CYSSG), only the $[M-H]^-$ ion was detected, and both a spiked CM sample and a CYSSG aqueous standard solution provided identical retention times (Figure S9).

Overall, 6 of the 21 CM discriminant metabolic features were successfully identified by MS and MS/MS experiments, while 5 were further chromatographically confirmed by chemical

standards (Table 1). In addition, 5 out of 6 identified features were detected in human serum samples (Table 1).

Biological significance of Identified Discriminant Metabolites in ccRCC

Metabolic patterns are powerful tools for sample classification, though understanding their biological significance can become more challenging.[45] Changes in gene expression do not necessarily correlate with changes in metabolites of a given pathway during ccRCC tumor progression.[25] Table 1 summarizes the identification of the 6 discriminant endogenous metabolites, discussed below based on their level change in CM and serum samples and the affected pathways, many of which have been reported to be involved in RCC progression. Fold changes were calculated as the ratio of median peak areas between CM samples, and *p* values were corrected using the Benjamini-Hochberg[36] procedure for multiple comparisons with a FDR of 0.1. Though no significant fold changes were obtained for serum samples, probably due to the small cohort, trends between classes are shown in Table 1.

Isoleucine and leucine (Ile+Leu), Phe, N-Lac-Leu and N-Ac-Phe exhibited significant higher levels in CM of 786-0 cells (CM-7) than HEK-293 cells (CM-H) with fold changes between 1.6 and 7.3 (Table 1, Figure 5). In contrast, CYSSG showed decreased levels in CM-7 compared to CM-H with the highest significant fold change equal to 18. Regarding the binary comparison between CM from non-tumor cells (CM-H) and Caki-1 cells (CM-C), CYSSG and OHPro-Val presented significant decreased levels in CM-C, with fold changes of 270 and 10, respectively; whereas N-Lac-Leu levels were 1.7 higher (Table 1, Figure 5). Significantly decreased levels of Ile+Leu, Phe, N-Ac-Phe and OHPro-Val were detected in CM-C compared to CM-7 with fold changes between 1.7 and 200, while CYSSG levels were 1.8-fold higher in CM-C (Table 1, Figure 5). Ile and Leu concentration was equivalent in both culture media

formulations, but Phe concentration was doubled in DMEM/F12 compared to RPMI. Therefore, comparisons between CM-C and CM-H or/and CM-7 should be cautiously interpreted for Phe.

Cysteineglutathione disulfide (CYSSG) – Redox state

CYSSG can be endogenously produced *via* a thiol-disulfide exchange reaction between γ-glutamyl-cysteinyl-glycine (GSH) and L-cystine or gluthathione disulfide (GSSG) and Cys.[46] The balance between GSH and GSSG is crucial for regulating the redox potential of the cell.[47, 48] GSH counteracts the increased ROS production, thus minimizing oxidative damage to tissues and cells.[47, 49] GSH is synthesized from amino acid precursors (Cys, Glu, and Gly) in the cell cytosol, and it is primarily regulated by GCS (glutathione-S-transferase), Cys availability, and GSH feedback inhibition.[50, 51] The intracellular GSH concentration depends on a dynamic balance between its synthesis and consumption inside the cell, and its efflux.[52] Since oxidative stress occurs in kidney cancer,[53] as well as in RCC cell lines including 786-O and Caki-1 cells,[54] it is not surprising to find that CYSSG levels in CM-7 and CM-C were lower than those in CM-H. Even if we do not have information about the ratio GSH/GSSC, our results suggest an alteration in the cell redox state. Along these lines, higher GSH levels were detected in tumor tissue of ccRCC patients compared to normal tissue,[55] while levels of GSH, GSSG, Cys-Gly and α-hidroxibutyrate, all involved in the GSH biosynthetic pathway, were increased in late-stage tumors compared to early stages.[25] As well, higher levels of CYSSG were detected in tumor tissue of a mouse xenograft model characterized by implantation of Caki-1 cells in the kidney, compared to control, presenting the highest fold change among all measured metabolites.[23] We were not able to detect CYSSG in the patient cohort, probably due to the relatively lower GSH levels reported in human plasma, compared to intracellular GSH content.[47]

Isoleucine, Leucine, Phenylalanine, N-Lactoyl-Leucine, N-Acetyl-Phenylalanine and Hydroxyproline-Valine –Metabolism of Amino acids and Proteins

The identified amino acids, Ile+Leu and Phe, and the amino acid derivatives N-Lac-Leu and N-Ac-Phe, exhibited significantly increased levels in CM-7 compared to CM-H, while similar levels of Ile+Leu, Phe and N-Ac-Phe were detected between CM-H and CM-C (Table 1, Figure 5). However, as mentioned above, the biological interpretation of Phe fold changes in CM-C could be biased by the different culture media composition. A different trend was obtained for OHPro-Val, which showed decreased levels in CM-7 and CM-C, with significant differences between CM-H and CM-C (Table 1, Figure 5). These compounds are all involved in the metabolism of amino acids and proteins. In this regard, Hakimi *et al.* identified that most of the pathways significantly decreased in ccRCC tumor tissue were involved in amino acid metabolism, including Ile, Leu and Phe metabolism.[25]

Previous evidence showed that a large number of dipeptides were detected in higher levels in tumor tissue of patients with SIV ccRCC compared to early stages.[25] Our *in vitro* model analysis showed lower OHPro-Val levels in CM from tumor cell lines compared to the non-tumor cell line (with significant decrease for CM-C vs. CM-H), and serum samples exhibiting the same trend (Table 1, Figure 5). A possible explanation for this result is that Caki-1 cells may reuse dipeptides as a source of intracellular amino acids so that, even if they are exported to the extracellular milieu they could be internalized faster than in normal cells. In this sense, it is well known that peptides can penetrate into the cells by endocytosis.[56]

Lower serum amino acid levels can be expected if biosynthetic requirements are enhanced in tumor cells. Our results showed no change in Phe levels and reduced levels of Ile+Leu in serum samples from SIV ccRRC patients (Table 1, Figure 5). Nevertheless, there are

some contradictory results in the literature. Non-significant changes in the levels of Ile, Leu, and Phe were reported in a serum sample cohort with 65% ccRCC (out of RCC) American patients by means of ion exchange chromatography with post derivatization[57] whereas increased levels of Ile, Leu and Phe were detected in a cohort with 100% ccRCC Chinese patients using an LC-MS/MS-based method, demonstrating the possible influence of ethnicity, histology, and/or to the techniques used in the results obtained.[58]

Regarding N-Ac-Phe, its levels were increased in CM from 786-O cells compared to CM-H, and a similar trend was observed for SIV ccRCC serum samples in relation to healthy individuals (Table 1, Figure 5). Elevated N-Ac-Phe in ccRCC serum has already been reported coupled to a tyrosine decline,[24] associating this alteration either to the inhibition of the phenylalanine hidroxylase, the enzyme necessary to metabolize phenylalanine to tyrosine,[59] or to impaired glycine N-acyltransferase activity.[60] Interestingly, the loss of VHL in ccRCC, among other changes, leads to elevated levels of citrate and cytosolic AcetylCoA,[25, 61] that may favor Phe acetylation. Higher level of acetylated proteins and/or amino acids could lead to higher levels of N-Ac-Phe, among others, in the extracellular compartment. A similar process may explain the experimentally observed higher levels of N-Ac-Phe in CM of 786-O cells, which lack VHL.

There is evidence showing that N-lactoyl-amino acids are synthetized from lactic acid and amino acids, catalyzed by cytosolic non-specific dipeptidase 2, and then exported by ATP-binding cassette subfamily C member 5 to the extracellular compartment.[62] High levels of lactate and certain amino acids have been shown to correlate with high levels of N-lactoyl-amino acids in plasma.[62] This has been observed in healthy individuals subjected to exercise, and patients with phenylketonuria, with high levels of Phe and N-lactoyl-phenylalanine. ccRCC exhibits high

rates of glycolisis and activation of lactate dehydrogenase,[61] leading to high cytosolic lactate content. Based on this evidence, we hypothesized that elevated N-Lac-Leu levels in the CM from tumor cells could result from the intracellular reaction of lactate and Leu, followed by N-Lac-Leu release. N-lactoyl-amino acids might represent useful extracellular biomarkers for intracellular amino acid concentration because they are only formed inside cells.[62] Interestingly, N-Lac-Leu exhibited the same increasing trend in CM-7 and CM-C as well as in SIV ccRCC serum samples compared to CM-H and healthy individuals, respectively (Table 1, Figure 5). Thus, it could be a more robust potential biomarker to be evaluated in a larger human serum cohort including different stages of disease.

Conclusions

In the present study, we developed a protocol for harvesting and extracting extracellular metabolites from an *in vitro* model of human renal cell lines. The exometabolome was profiled using a discovery-based metabolomics approach *via* UPLC-QTOF-MS. The metabolic footprints of ccRCC cell lines and a non-tumor renal cell line were compared with the human serum metabolic profile of SIV ccCRR and healthy individuals. A panel of 21 discriminant features obtained from the binary comparison of control HEK-293 and tumor 786-O cells allowed differentiating kidney cell lines with 100% specificity, sensitivity and accuracy. A subset of 9 discriminant features from CM samples was detected in human serum, and produced a fingerprint that enabled discerning stage IV ccRCC patients from healthy individuals. Identified discriminant metabolites suggest alterations in amino acid metabolism, and the redox status of cells. To our knowledge, this is the first time that N-lactoyl-leucine is associated to ccRCC. Even if results in serum need to be validated in a larger cohort, our study highlights the utility of RCC

*in vit*ro models for finding disease-related discriminant metabolites capable of human serum classification.

Acknowledgements

We acknowledge CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina), and the National Agency of Scientific and Technological Promotion (PRH-PICT-2015-0022 project) for providing the funding; and the "Biobanco Público de Muestras Séricas Oncológicas" (BPMSO) from Instituto de Oncología A. H. Roffo for providing the samples. M.E.M., M.O.S. and L.I.P. are Research Staff members from CONICET. M.O.S. acknowledges the National Mass Spectrometry System (SNEM) for the travel grant received for his research stance at CIBION. We also would like to thank Dr. Ana Bellomo and Dr. Mariela Bollini for providing useful guidance regarding the synthesis of chemical standards.

Figure Captions

Figure 1. Scheme of conditioned media (CM) incubation and collection for the different cell lines studied. O/N: overnight.

Figure 2. (**A**) Principal Component Analysis (PCA) scores plot of conditioned media (CM) samples using the set of 755 spectral features for HEK-293 non-tumor control cell line (blue squares), 786-O cell line (red triangles), Caki-1 cell line (black circles); and quality controls (green diamonds), with a model that consisted of 5 PCs with 70.67% total captured variance. (**B**) PCA scores plot for HEK-293 (blue squares) and 786-O cell line (red triangles), with a model that consisted of 2 PCs with 53.38% total captured variance. (**C**) Cross-validated prediction plot

from the orthogonal projection to latent structures-discriminant analysis model of CM samples from a tumor cell line 786-O (red triangle) versus a non-tumor control cell line HEK 293 (blue squares) using the 21 discriminant metabolic feature panel obtained from genetic algorithm variable selection.

Figure 3. Principal Component Analysis (PCA) scores plot of serum (top panel) and CM (low panel) samples using the set of 163 common spectral features (left) and the set of 9 discriminant features of CM samples (right) that were also present in serum samples. The model consisted of (**A**) 2 PCs with 57.97% total captured variance; (**B**) 2 PCs with 70.87% total captured variance; (**C**) 4 PCs with 69.29% total captured variance, and (**D**) 2 PC with 65.02% total captured variance. Serum samples from healthy individuals (blue squares) and stage IV ccRCC patients (magenta diamonds); CM samples from the non-tumor control cell line HEK-293 (blue squares); and tumor cell lines 786-O (red triangles), and Caki-1(black circles).

Figure 4. (**A**) Extracted ion chromatograms for [N-lactoyl-leucine-H]$^-$ ion at *m/z* 202.1079 ± 0.0500 generated from conditioned media (CM, dotted line) and serum (solid line), non-spiked (blue), and 150 μM spiked (red) samples; and a 15 μM N-lactoyl-leucine (N-Lac-Leu) standard solution (red dashed-dotted line). (**B**) Mass spectrum for [N-Lac-Leu-H]$^-$ ion with *m/z* 202.1079 in CM sample (red), and its simulated isotopic pattern (black). (**C**) Product ion mass spectra of [N-Lac-Leu-H]$^-$ precursor ion for a CM sample (red), and for a 15 μM N-Lac-Leu standard (green) using a collision cell voltage of 20 V.

Figure 5. Box plots for discriminant metabolites in CM and serum samples. Comparisons are shown for CM of HEK-293 (n= 21), 786-O (n= 21) and Caki-1 (n=22) cell lines; and between stage IV ccRCC patients and healthy individuals. Mean values are represented by a filled circle in the box; median values are represented by a line in the box; the upper and lower edges of the box are the 25^{th} and 75^{th} percentiles; the whisker extends to the most extreme values in data, not including outliers defined as 1.5 of the interquartile range and represented by colored crosses. Fold changes are calculated as the ratio of median peak areas between CM from 786-O and HEK-293 (CM-7/CM-H), Caki-1 and HEK-293 (CM-C/CM-H), and Caki-1 and 786-O (CM-C/CM-7) samples. Mann-Whitney U tests were used to calculate statistical significance, and *p* values were corrected using the Benjamini-Hochberg procedure for multiple comparisons with a FDR of 0.1. Statistically significant differences after correction for multiple comparisons between classes are indicated on top of the boxes with i) * for CM-7 or CM-C vs CM-H; and ii) # for CM-C vs CM-7. Metabolite identities are indicated for each case.

Figure 1.

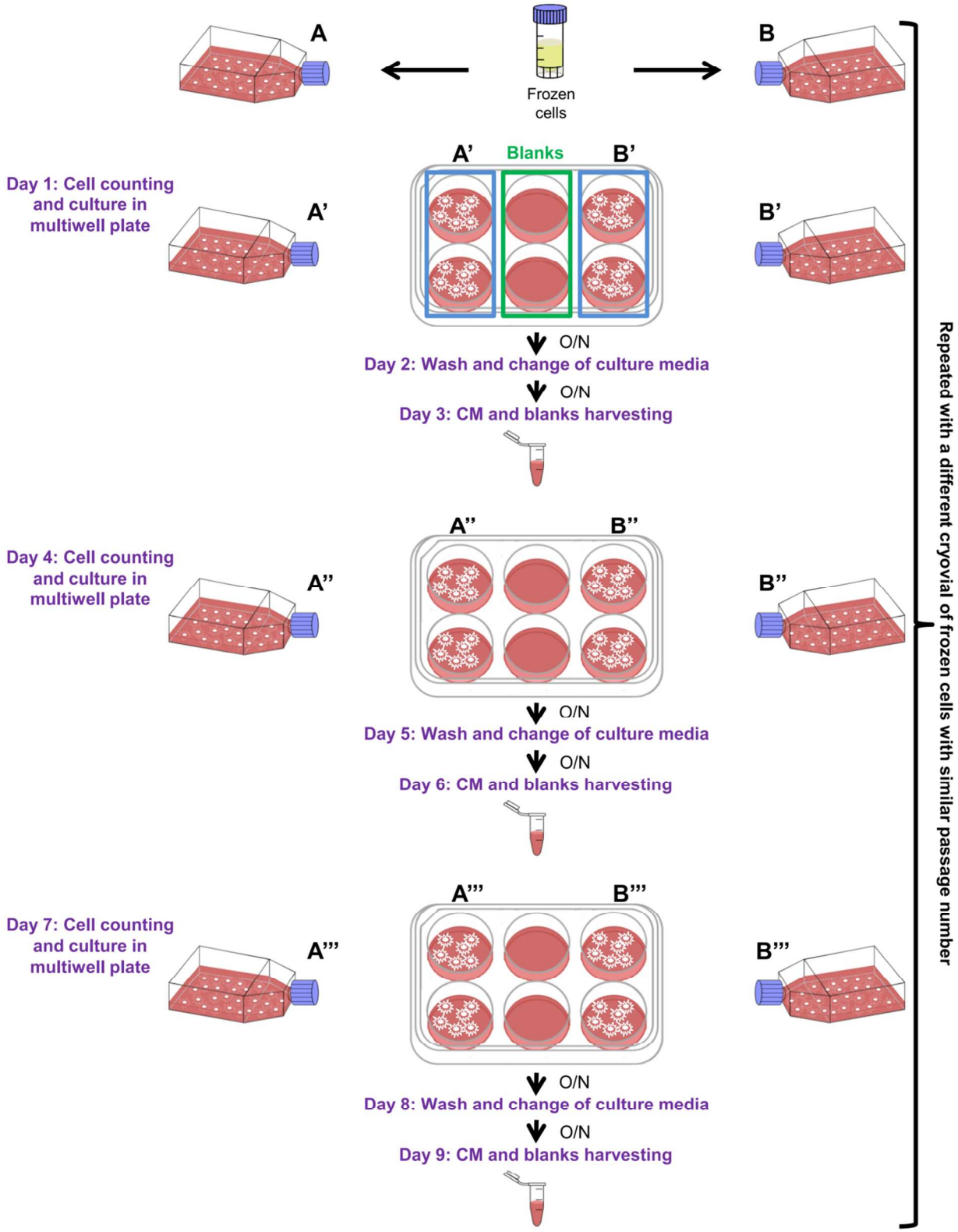

A
Frozen cells
B
A'
Blanks
B'
Day 1: Cell counting and culture in multiwell plate
A'
B'
O/N
Day 2: Wash and change of culture media
O/N
Day 3: CM and blanks harvesting
A''
B''
Day 4: Cell counting and culture in multiwell plate
A''
B''
O/N
Day 5: Wash and change of culture media
O/N
Day 6: CM and blanks harvesting
A'''
B'''
Day 7: Cell counting and culture in multiwell plate
A'''
B'''
O/N
Day 8: Wash and change of culture media
O/N
Day 9: CM and blanks harvesting
Repeated with a different cryovial of frozen cells with similar passage number

Figure 2.

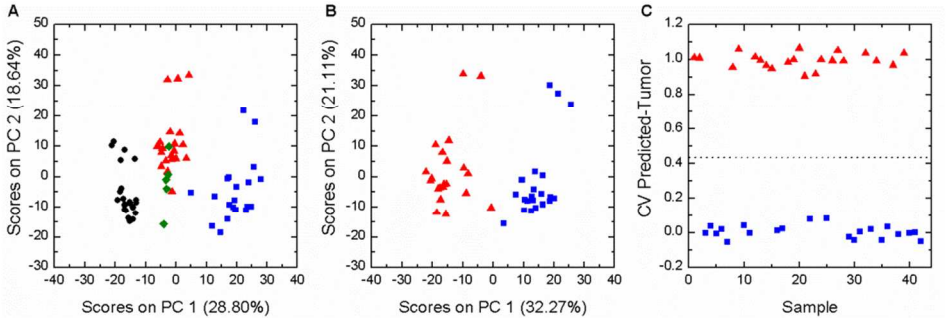

A
Scores on PC 2 (18.64%)
Scores on PC 1 (28.80%)
B
Scores on PC 2 (21.11%)
Scores on PC 1 (32.27%)
C
CV Predicted-Tumor
Sample

Figure 3.

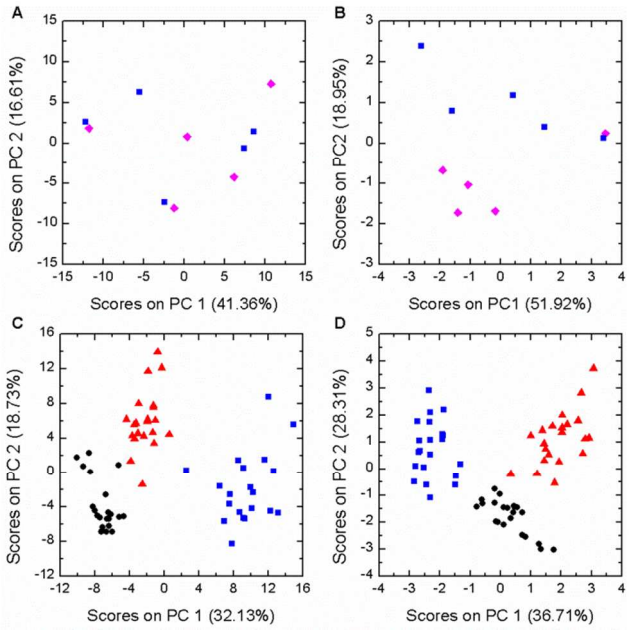

A
Scores on PC 2 (16.61%)
Scores on PC 1 (41.36%)
B
Scores on PC2 (18.95%)
Scores on PC1 (51.92%)
C
Scores on PC 2 (18.73%)
Scores on PC 1 (32.13%)
D
Scores on PC 2 (28.31%)
Scores on PC 1 (36.71%)

Figure 4.

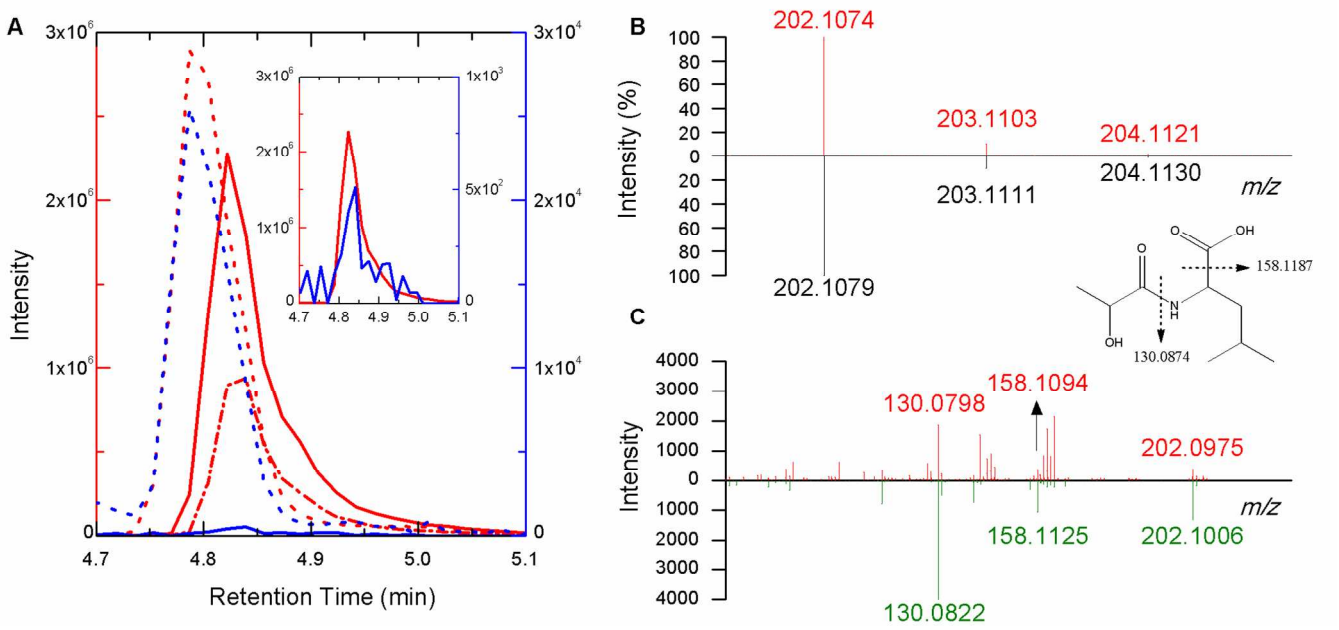

A
Intensity
Retention Time (min)
B
Intensity (%)
202.1074
203.1103
204.1121
204.1130
203.1111
202.1079
m/z
158.1187
130.0874
C
Intensity
130.0798
158.1094
202.0975
158.1125
202.1006
130.0822
m/z

Figure 5.

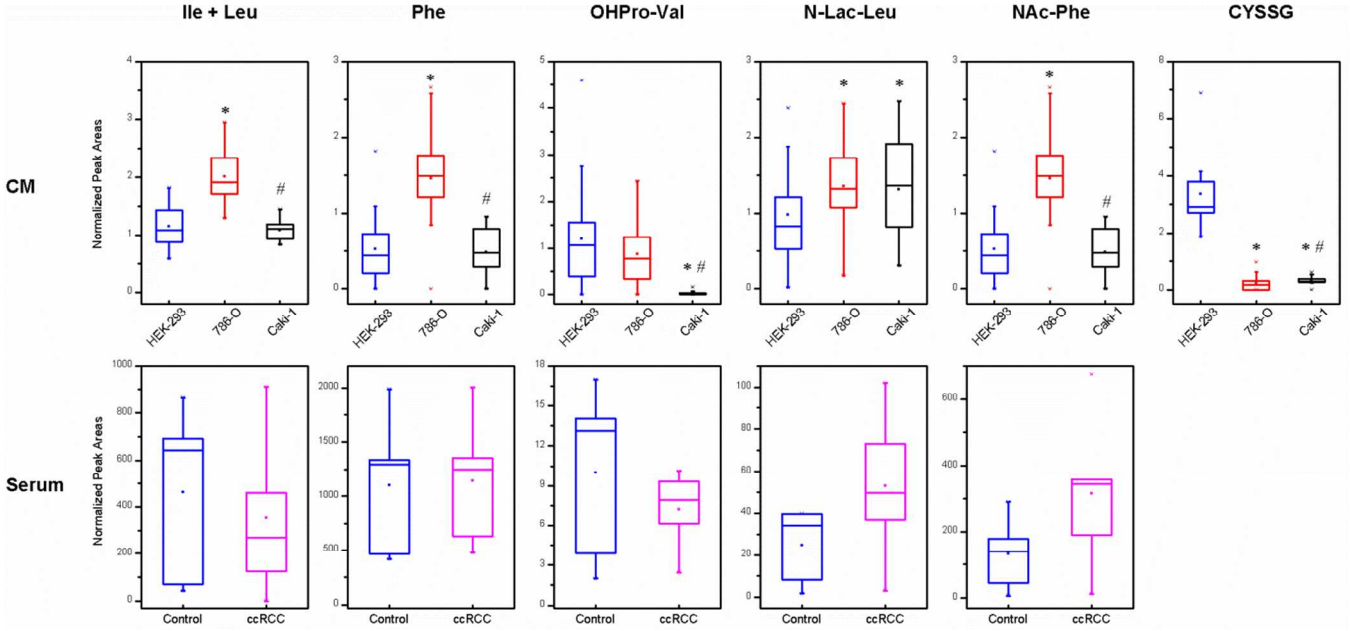

Ile + Leu
Phe
OHPro-Val
N-Lac-Leu
NAc-Phe
CYSSG
CM
Serum
Normalized Peak Areas
HEK-293
786-O
Caki-1
Control
ccRCC

Table 1. Identification of discriminant feature panel, based on accurate mass (a), isotopic pattern (b), MS/MS experiments (c), and retention time match with standards (d). Compounds and features highlighted in bold are detected in serum and metabolites chromatographically confirmed by chemical standards are indicated in italics. Δm is calculated as the mass difference between the *m/z* value obtained from Progenesis and the theoretical mass. Fold changes are calculated as the ratio of median peak areas between CM from 786-O and HEK-293 (CM-7/CM-H), Caki-1 and HEK-293 (CM-C/CM-H), and Caki-1 and 786-O (CM-C/CM-7) samples; except for cases with median value equal to zero. *p* values were calculated using Mann-Whitney U tests. NS: non-significant differences after correction with the Benjamini-Hochberg procedure for multiple comparisons with a FDR of 0.1. Trends in binary comparisons are indicated with arrows, (↑): increased levels, and (↓): decreased levels. For CM samples, fold changes of 1-9.9, 10-99.99 and >100 are indicated with one, two or three arrows, respectively. Non-statistically significant trends are indicated for serum samples in the comparison of SIV ccRCC patients and healthy individuals (HI).

Rt (min)	*m/z*	Ion Type	Elemental Formula	Δm (mDa)	CM-7/CM-H			CM-C/CM-H			CM-C/CM-7			Serum ccRCC/HI	Metabolite Identification	ID Confirmation
					p	Trend	Fold Change	*p*	Trend	Fold Change	*p*	Trend	Fold Change	Trend		
0.92	425.0825	$[M-H]^-$	$C_{13}H_{22}N_4O_8S_2$	0.2	$2.5x10^{-8}$	↓↓	18	$2.1x10^{-8}$	↓	9.9	$6.6x10^{-2}$	↑	1.8	-	*Cysteineglutathione-disulfide*	a, b, c, d
1.09	225.9854	$[M+NaCl-H]^-$	-	-	$2.2x10^{-8}$	-	-	-	-	-	$1.7x10^{-8}$	↓	3.8	-	NO ID	-
1.30	**303.9641**	**$[M+3NaCl-H]^-$**	**$C_6H_{13}NO_2$**	**1.4**	**$1.4x10^{-6}$**	↑	**1.8**	**NS**	**-**	**1.0**	**$3.7x10^{-8}$**	↓	**1.7**	↓	***Isoleucine+Leucine***	**a, b, c, d**
1.67	**569.7810**	**$[M+7NaCl-H]^-$**	**$C_9H_{11}NO_2$**	**-0.5**	**$3.5x10^{-4}$**	↑	**3.4**	**NS**	↑	**1.2**	**$3.4x10^{-5}$**	↓	**2.8**	**-**	***Phenylalanine***	**a, b, c, d**
2.07	**287.0872**	**$[M+NaCl-H]^-$**	**$C_{10}H_{18}N_2O_4$**	**9.7**	**NS**	↓	**1.4**	**$1.3x10^{-6}$**	↓↓↓	**270**	**$5.4x10^{-7}$**	↓↓↓	**200**	↓	**Hydroxyprolyl-valine**	**a, b, c**
2.52	**142.0687**	**$[M-H]^-$**	-	-	**$2.0x10^{-6}$**	↑	**3.1**	**$2.1x10^{-8}$**	↑	**5.7**	**$1.0x10^{-5}$**	↑	**1.8**	↓	**NO ID**	-
3.19	558.9097	$[M+4NaCl-H]^-$	-	-	$2.1x10^{-8}$	-	-	$2.5x10^{-7}$	-	-	$1.2x10^{-3}$	↓	1.9	-	NO ID	-
3.38	230.9961	$[M-H]^-$	-	-	$7.9x10^{-8}$	-	-	$1.3x10^{-5}$	↓	3.6	$1.0x10^{-6}$	-	-	-	NO ID	-
3.97	180.0137	$[M-H]^-$	-	-	NS	↑	1.6	$8.7x10^{-8}$	-	-	$2.6x10^{-6}$	-	-	-	NO ID	-
4.38	**283.0891**	**$[M-H]^-$**	-	-	**$2.0x10^{-8}$**	**-**	**-**	**$1.0x10^{-8}$**	**-**	**-**	**$4.3x10^{-3}$**	↑	**1.3**	↑	**NO ID**	-
4.91	**202.1104**	**$[M-H]^-$**	**$C_9H_{17}NO_4$**	**2.5**	**$3.9x10^{-2}$**	↑	**1.6**	**$6.7x10^{-2}$**	↑	**1.7**	**NS**	**-**	**1.0**	↑	***N-lactoyl-leucine***	**a, b, c, d**
5.03	**264.0427**	**$[M+NaCl-H]^-$**	**$C_{11}H_{13}NO_3$**	**2.4**	**$3.6x10^{-8}$**	↑	**7.3**	**NS**	↓	**1.1**	**$2.8x10^{-8}$**	↓	**7.8**	↑	***N-acetyl-phenylalanine***	**a, b, c, d**
5.40	319.0772	$[M+Cl]^-$	-	-	$3.0x10^{-8}$	↑↑	39	$9.4x10^{-8}$	↑↑	14	$2.8x10^{-5}$	↓	2.8	-	NO ID	-
6.34	206.0815	$[M-H]^-$	-	-	$2.1x10^{-8}$	-	-	$1.9x10^{-8}$	↓↓↓	200	-	-	-	-	NO ID	-
6.58	**138.0199**	**$[M-H]^-$**	-	-	**$3.1x10^{-8}$**	↓	**5.4**	**$2.1x10^{-8}$**	↓	**6.3**	**$7.4x10^{-2}$**	↓	**1.2**	↓	**NO ID**	-
6.81	213.1167	$[M-H]^-$	-	-	$4.1x10^{-5}$	↓	2.2	$1.9x10^{-7}$	↓	5.1	$1.9x10^{-2}$	↓	2.4	-	NO ID	-
7.90	420.0948	$[M-H]^-$	-	-	$9.6x10^{-3}$	↑	1.6	$1.8x10^{-9}$	-	-	$1.8x10^{-9}$	-	-	-	NO ID	-
8.36	333.1066	$[M-H]^-$	-	-	$4.8x10^{-3}$	↑	2.4	NS	↓	1.6	$3.1x10^{-4}$	↓	3.9	-	NO ID	-
8.41	**211.1355**	**$[M-H]^-$**	-	-	**$7.2x10^{-7}$**	↑	**7.6**	**$2.6x10^{-4}$**	↑	**3.0**	**$1.1x10^{-4}$**	↓	**2.5**	↓	**NO ID**	-
9.44	177.0970	$[M+Cl]^-$	-	-	$4.1x10^{-5}$	↓	1.5	$8.7x10^{-7}$	↓	2.1	$7.8x10^{-2}$	↓	1.4	-	NO ID	-
11.29	291.1608	$[M-H]^-$	-	-	$7.2x10^{-7}$	↓	4.3	$2.3x10^{-6}$	↓	3.4	NS	↓	1.3	-	NO ID	-

Supporting Information

The following Supporting Information is available free of charge at ACS website http://pubs.acs.org.

Data Set S1. Metabolic feature matrix for conditioned media samples.

Materials and Methods. Organic Synthesis of N-acetyl-phenylalanine and *N*-lactoyl-leucine.

Figure S1. ^{1}H NMR and ^{13}C NMR spectra (500 MHz, D_2O) of N-acetyl-phenylalanine.

Figure S2. ^{1}H NMR and ^{13}C NMR spectra (500 MHz, D_2O) of N-lactoyl-leucine.

Figure S3. Representative base peak intensity chromatograms obtained for conditioned media and serum samples.

Figure S4. Principal Component Analysis scores plot of conditioned media (CM) samples using the set of 9 metabolic features from the discriminant CM panel that were present in serum samples.

Figure S5. Identification of N-acetyl-phenylalanine.

Figure S6. Identification of phenylalanine.

Figure S7. Identification of isoleucine/leucine.

Figure S8. Identification of hydroxyprolyl-valine.

Figure S9. Identification of cysteineglutathione disulfide.

Table S1. Conditioned media samples collected at different time points based on the cell culture protocol.

References

(1) Linehan, W. M.; Srinivasan, R.; Schmidt, L. S., The genetic basis of kidney cancer: a metabolic disease. *Nat. Rev. Urol.* **2010,** *7*, 277-85.

(2) Hsieh, J. J.; Purdue, M. P.; Signoretti, S.; Swanton, C.; Albiges, L.; Schmidinger, M.; Heng, D. Y.; Larkin, J.; Ficarra, V., Renal cell carcinoma. *Nat. Rev. Dis. Primers* **2017,** *3*, 17009.

(3) Linehan, W. M.; Bratslavsky, G.; Pinto, P. A.; Schmidt, L. S.; Neckers, L.; Bottaro, D. P.; Srinivasan, R., Molecular Diagnosis and Therapy of Kidney Cancer. *Annu. Rev. Med.* **2010,** *61*, 329-343.

(4) IARC, Tumors of the kidney. In *World Health Organization Classification of Tumors. Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs*, Eble, J. N.; Sauter, G.; Epstein, J. I.; Sesterhenn, I. A., Eds. IARC press: Lyon, 2004; pp 9-43.
(5) Gibbons, R. P.; Monte, J. E.; Correa, R. J., Jr.; Mason, J. T., Manifestations of renal cell carcinoma. *Urology* **1976,** *8*, 201-6.
(6) Hu, B.; Lara, P. N., Jr.; Evans, C. P., Defining an Individualized Treatment Strategy for Metastatic Renal Cancer. *Urol. Clin. North Am.* **2012,** *39*, 233-249.
(7) Graves, A.; Hessamodini, H.; Wong, G.; Lim, W. H., Metastatic renal cell carcinoma: update on epidemiology, genetics, and therapeutic modalities. *Immunotargets Ther.* **2013,** *2*, 73-90.
(8) Diamond, E.; Molina, A. M.; Carbonaro, M.; Akhtar, N. H.; Giannakakou, P.; Tagawa, S. T.; Nanus, D. M., Cytotoxic chemotherapy in the treatment of advanced renal cell carcinoma in the era of targeted therapy. *Crit. Rev. Oncol. Hematol.* **2015,** *96*, 518-526.
(9) De Meerleer, G.; Khoo, V.; Escudier, B.; Joniau, S.; Bossi, A.; Ost, P.; Briganti, A.; Fonteyne, V.; Van Vulpen, M.; Lumen, N.; Spahn, M.; Mareel, M., Radiotherapy for renal-cell carcinoma. *Lancet Oncol.* **2014,** *15*, e170-e177.
(10) Rodriguez-Vida, A.; Hutson, T. E.; Bellmunt, J.; Strijbos, M. H., New treatment options for metastatic renal cell carcinoma. *ESMO Open* **2017,** *2*, e000185.
(11) Hakimi, A. A.; Pham, C. G.; Hsieh, J. J., A clear picture of renal cell carcinoma. *Nat. Genet.* **2013,** *45*, 849-850.
(12) Calzada, M. J.; Esteban, M. A.; Feijoo-Cuaresma, M.; Castellanos, M. C.; Naranjo-Suarez, S.; Temes, E.; Mendez, F.; Yanez-Mo, M.; Ohh, M.; Landazuri, M. O., von Hippel-Lindau tumor suppressor protein regulates the assembly of intercellular junctions in renal cancer cells through hypoxia-inducible factor-independent mechanisms. *Cancer Res.* **2006,** *66*, 1553-60.
(13) Banks, R. E.; Tirukonda, P.; Taylor, C.; Hornigold, N.; Astuti, D.; Cohen, D.; Maher, E. R.; Stanley, A. J.; Harnden, P.; Joyce, A.; Knowles, M.; Selby, P. J., Genetic and epigenetic analysis of von Hippel-Lindau (VHL) gene alterations and relationship with clinical variables in sporadic renal cancer. *Cancer Res.* **2006,** *66*, 2000-11.
(14) Godinot, C.; de Laplanche, E.; Hervouet, E.; Simonnet, H., Actuality of Warburg's views in our understanding of renal cancer metabolism. *J. Bioenerg. Biomembr.* **2007,** *39*, 235-41.
(15) Pinthus, J. H.; Whelan, K. F.; Gallino, D.; Lu, J.-P.; Rothschild, N., Metabolic features of clear-cell renal cell carcinoma: mechanisms and clinical implications. *Can. Urol. Assoc. J.* **2011,** *5*, 274-282.
(16) Gebhard, R. L.; Clayman, R. V.; Prigge, W. F.; Figenshau, R.; Staley, N. A.; Reesey, C.; Bear, A., Abnormal cholesterol metabolism in renal clear cell carcinoma. *J. Lipid Res.* **1987,** *28*, 1177-84.
(17) Cuperlovic-Culf, M.; Barnett, D. A.; Culf, A. S.; Chute, I., Cell culture metabolomics: applications and future directions. *Drug Discov. Today* **2010,** *15*, 610-21.
(18) León, Z.; García-Cañaveras, J. C.; Donato, M. T.; Lahoz, A., Mammalian cell metabolomics: Experimental design and sample preparation. *Electrophoresis* **2013,** *34*, 2762-2775.
(19) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L., Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000,** *18*, 1157-61.
(20) Nicholson, J. K.; Lindon, J. C., Systems biology: Metabonomics. *Nature* **2008,** *455*, 1054-6.
(21) Kim, K.; Taylor, S. L.; Ganti, S.; Guo, L. N.; Osier, M. V.; Weiss, R. H., Urine Metabolomic Analysis Identifies Potential Biomarkers and Pathogenic Pathways in Kidney Cancer. *OMICS* **2011,** *15*, 293-303.
(22) Ganti, S.; Taylor, S. L.; Kim, K.; Hoppel, C. L.; Guo, L. N.; Yang, J.; Evans, C.; Weiss, R. H., Urinary acylcarnitines are altered in human kidney cancer. *Int. J. Cancer* **2012,** *130*, 2791-2800.
(23) Ganti, S.; Taylor, S. L.; Abu Aboud, O.; Yang, J.; Evans, C.; Osier, M. V.; Alexander, D. C.; Kim, K.; Weiss, R. H., Kidney Tumor Biomarkers Revealed by Simultaneous Multiple Matrix Metabolomics Analysis. *Cancer Res.* **2012,** *72*, 3471-3479.
(24) Lin, L.; Huang, Z. Z.; Gao, Y.; Yan, X. M.; Xing, J. C.; Hang, W., LC-MS based serum metabonomic analysis for renal cell carcinoma diagnosis, staging, and biomarker discovery. *J. Proteome Res.* **2011,** *10*, 1396-1405.

(25) Hakimi, A. A.; Reznik, E.; Lee, C.-H.; Creighton, Chad J.; Brannon, A. R.; Luna, A.; Aksoy, B. A.; Liu, Eric M.; Shen, R.; Lee, W.; Chen, Y.; Stirdivant, Steve M.; Russo, P.; Chen, Y.-B.; Tickoo, Satish K.; Reuter, Victor E.; Cheng, Emily H.; Sander, C.; Hsieh, James J., An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer Cell* **2016,** *29*, 104-116.

(26) Kell, D. B.; Brown, M.; Davey, H. M.; Dunn, W. B.; Spasic, I.; Oliver, S. G., Metabolic footprinting and systems biology: the medium is the message. *Nat. Rev. Microbiol.* **2005,** *3*, 557-565.

(27) Halama, A., Metabolomics in cell culture--a strategy to study crucial metabolic pathways in cancer development and the response to treatment. *Arch. Biochem. Biophys.* **2014,** *564*, 100-9.

(28) Sinha, R.; Winer, A. G.; Chevinsky, M.; Jakubowski, C.; Chen, Y. B.; Dong, Y.; Tickoo, S. K.; Reuter, V. E.; Russo, P.; Coleman, J. A.; Sander, C.; Hsieh, J. J.; Hakimi, A. A., Analysis of renal cancer cell lines from two major resources enables genomics-guided cell line selection. *Nat Commun.* **2017,** *8*, 15165.

(29) Zheng, X.; Baker, H.; Hancock, W. S.; Fawaz, F.; McCaman, M.; Pungor, E., Jr., Proteomic analysis for the assessment of different lots of fetal bovine serum as a raw material for cell culture. Part IV. Application of proteomics to the manufacture of biological drugs. *Biotechnol. Progr.* **2006,** *22*, 1294-300.

(30) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B., Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **2018,** *14*, 72.

(31) Bateman, K. P.; Castro-Perez, J.; Wrona, M.; Shockcor, J. P.; Yu, K.; Oballa, R.; Nicoll-Griffith, D. A., MSE with mass defect filtering for in vitro and in vivo metabolite identification. *Rapid Commun. Mass Spectrom.* **2007,** *21*, 1485-96.

(32) Kirwan, J. A.; Weber, R. J. M.; Broadhurst, D. I.; Viant, M. R., Direct infusion mass spectrometry metabolomics dataset: a benchmark for data processing and quality control. *Scientific Data* **2014,** *1*, 140012.

(33) Daskalaki, E.; Pillon, N. J.; Krook, A.; Wheelock, C. E.; Checa, A., The influence of culture media upon observed cell secretome metabolite profiles: The balance between cell viability and data interpretability. *Anal. Chim. Acta* **2018,** *DOI: 10.1016/j.aca.2018.04.034*.

(34) Trygg, J.; Wold, S., Orthogonal projections to latent structures (O-PLS). *J. Chemometr.* **2002,** *16*, 119-128.

(35) Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J., OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemometr.* **2006,** *20*, 341-351.

(36) Benjamini, Y.; Hochberg, Y., Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B, Stat. Methodol.* **1995,** *57*, 289-300.

(37) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L., HMDB: the Human Metabolome Database. *Nucleic Acids Res.* **2007,** *35*, D521-6.

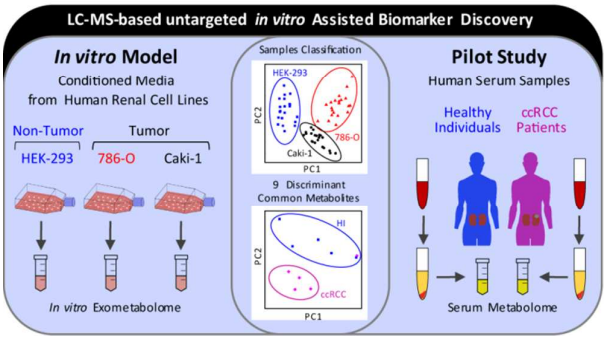
(38) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005,** *27*, 747-51.

(39) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T., MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010,** *45*, 703-714.

(40) Wild, C. P., Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol. Biomarkers Prev.* **2005,** *14*, 1847-1850.
(41) Rappaport, S. M.; Smith, M. T., Environment and Disease Risks. *Science* **2010,** *330*, 460-461.
(42) Nicholson, J. K.; Holmes, E.; Kinross, J.; Burcelin, R.; Gibson, G.; Jia, W.; Pettersson, S., Host-gut microbiota metabolic interactions. *Science* **2012,** *336*, 1262-7.
(43) Kebarle, P.; Verkerk Udo, H., Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spectrom. Rev.* **2009,** *28*, 898-917.
(44) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S., Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013,** *85*, 2-9.
(45) Johnson, C. H.; Ivanisevic, J.; Siuzdak, G., Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **2016,** *17*, 451-459.
(46) Eriksson, B.; Eriksson, S. A., Synthesis and characterization of the L-cysteine-glutathione mixed disulfide. *Acta Chem. Scand.* **1967,** *21*, 1304-1312.
(47) Giustarini, D.; Colombo, G.; Garavaglia, M. L.; Astori, E.; Portinaro, N. M.; Reggiani, F.; Badalamenti, S.; Aloisi, A. M.; Santucci, A.; Rossi, R.; Milzani, A.; Dalle-Donne, I., Assessment of glutathione/glutathione disulphide ratio and S-glutathionylated proteins in human blood, solid tissues, and cultured cells. *Free Radical Biol. Med.* **2017,** *112*, 360-375.
(48) Chakravarthi, S.; Jessop, C. E.; Bulleid, N. J., The role of glutathione in disulphide bond formation and endoplasmic-reticulum-generated oxidative stress. *EMBO Rep.* **2006,** *7*, 271-5.
(49) Meister, A.; Anderson, M. E., Glutathione. *Annu. Rev. Biochem.* **1983,** *52*, 711-60.
(50) Estrela, J. M.; Ortega, A.; Obrador, E., Glutathione in cancer biology and therapy. *Crit. Rev. Clin. Lab. Sci.* **2006,** *43*, 143-81.
(51) Lu, S. C., Glutathione synthesis. *Biochim. Biophys. Acta* **2013,** *1830*, 3143-53.
(52) Ballatori, N.; Krance, S. M.; Marchan, R.; Hammond, C. L., Plasma membrane glutathione transporters and their roles in cell physiology and pathophysiology. *Mol. Aspects Med.* **2009,** *30*, 13-28.
(53) Ganesamoni, R.; Bhattacharyya, S.; Kumar, S.; Chauhan, A.; Mete, U. K.; Agarwal, M. M.; Mavuduru, R.; Kaushik, G.; Mandal, A. K.; Singh, S. K., Status of oxidative stress in patients with renal cell carcinoma. *J. Urol.* **2012,** *187*, 1172-6.
(54) Tew, K. D.; Monks, A.; Barone, L.; Rosser, D.; Akerman, G.; Montali, J. A.; Wheatley, J. B.; Schmidt, D. E., Jr., Glutathione-associated enzymes in the human cell lines of the National Cancer Institute Drug Screening Program. *Mol. Pharmacol.* **1996,** *50*, 149-59.
(55) Li, B.; Qiu, B.; Lee, D. S. M.; Walton, Z. E.; Ochocki, J. D.; Mathew, L. K.; Mancuso, A.; Gade, T. P. F.; Keith, B.; Nissim, I.; Simon, M. C., Fructose-1,6-bisphosphatase opposes renal carcinoma progression. *Nature* **2014,** *513*, 251.
(56) Goldstein, J. L.; Anderson, R. G.; Brown, M. S., Coated pits, coated vesicles, and receptor-mediated endocytosis. *Nature* **1979,** *279*, 679-85.
(57) Mustafa, A.; Gupta, S.; Hudes, G. R.; Egleston, B. L.; Uzzo, R. G.; Kruger, W. D., Serum amino acid levels as a biomarker for renal cell carcinoma. *J. Urol.* **2011,** *186*, 1206-12.
(58) Zhang, F.; Ma, X.; Li, H.; Guo, G.; Li, P.; Gu, L.; Li, X.; Chen, L.; Zhang, X., The predictive and prognostic values of serum amino acid levels for clear cell renal cell carcinoma. *Urol. Oncol.* **2017,** *35*, 392-400.
(59) Gerlo, E.; Van Coster, R.; Lissens, W.; Winckelmans, G.; De Meirleir, L.; Wevers, R., Gas chromatographic-mass spectrometric analysis of N-acetylated amino acids: the first case of aminoacylase I deficiency. *Anal. Chim. Acta* **2006,** *571*, 191-9.
(60) Kelley, M.; Vessey Donald, A., Isolation and characterization of mitochondrial acyl‐coa: Glycine n‐acyltransferases from kidney. *J. Biochem. Mol. Toxicol.* **1993,** *8*, 63-69.
(61) Wettersten, H. I.; Aboud, O. A.; Lara, P. N., Jr.; Weiss, R. H., Metabolic reprogramming in clear cell renal cell carcinoma. *Nat. Rev. Nephrol.* **2017,** *13*, 410-419.

(62) Jansen, R. S.; Addie, R.; Merkx, R.; Fish, A.; Mahakena, S.; Bleijerveld, O. B.; Altelaar, M.; Ijlst, L.; Wanders, R. J.; Borst, P.; van de Wetering, K., N-lactoyl-amino acids are ubiquitous metabolites that originate from CNDP2-mediated reverse proteolysis of lactate and amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112*, 6601.

For TOC only


LC-MS-based untargeted *in vitro* Assisted Biomarker Discovery
In vitro Model
Conditioned Media
from Human Renal Cell Lines
Non-Tumor
Tumor
HEK-293
786-O
Caki-1
In vitro Exometabolome
Samples Classification
HEK-293
786-O
Caki-1
PC1
PC2
9 Discriminant
Common Metabolites
HI
ccRCC
PC1
PC2
Pilot Study
Human Serum Samples
Healthy
Individuals
ccRCC
Patients
Serum Metabolome