sciforum

A Mass Spectrometry-based Lipidomics Study for Early Diagnosis of clear cell Renal Cell Carcinoma

Malena Manzi[1], Martín Palazzo[2,3], Nicolás Zabalegui[1], María Elena Knott[1], Patricio Yankilevich[3], María Isabel Giménez[4], Lyida I. Puricelli[5], María Eugenia Monge*[1]

[1]Centro de Investigaciones en Bionanociencias (CIBION), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2390, C1425FQD, CABA, Argentina. [2]Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA) CONICET, Instituto Partner de la Sociedad Max Planck, Godoy Cruz 2390, C1425FQD, CABA, Argentina. [3] LM2S, Université de Technologié de Troyes 12 rue Marie-Curie, CS42060 Troyes, France. [4]Departamento de Diagnóstico y Tratamiento, Hospital Italiano de Buenos Aires, Tte. Gral. Juan Domingo Perón 4190, C1199ABB, CABA, Argentina.[5]Instituto de Oncología Ángel H. Roffo, Facultad de Medicina, Universidad de Buenos Aires, Av. San Martín 5481, C1417DTB, CABA, Argentina.

***maria.monge@cibion.conicet.gov.ar**

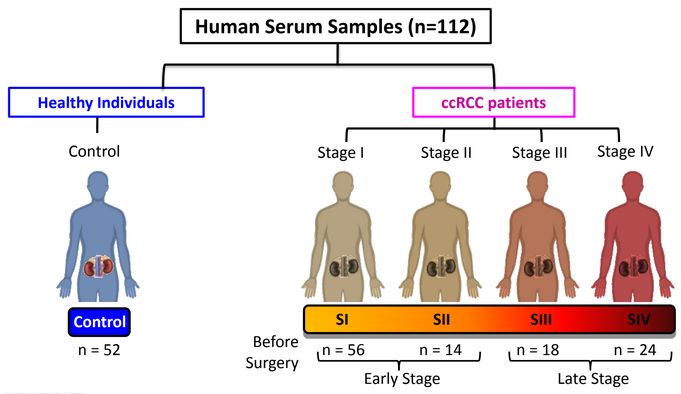
INTRODUCTION

Kidney cancer is fundamentally a metabolic disease[1] and more than 30% of patients, often incidentally diagnosed by imaging procedures, exhibit locally advanced or metastatic renal cell carcinoma (RCC) at the time of diagnosis.[2, 3] The disease is inherently resistant to chemotherapy[4] and radiotherapy.[5] Clear cell RCC (ccRCC) is the most common (75%) lethal subtype, and is considered a glycolytic and lipogenic tumor.[6,7] Current research has shown that several metabolic alterations are associated with RCC and different potential biomarkers have been identified.[7-9] Early diagnosis is needed to reduce the mortality associated to ccRCC, to give more opportunities for early intervention and improved outcome of ccRCC patients. In this context, lipids are candidate molecules to be explored in a non-targeted fashion as potential biomarkers for ccRCC diagnosis by means of lipid profiling experiments.

OBJECTIVES

- Optimize a protocol for extracting and analyzing lipids from human serum samples.
- Profile the lipidome using a discovery-based lipidomics approach via UHPLC-QTOF-MS.
- Compare the serum lipid profiles of ccRCC patients with those from healthy individuals.
- Compare the lipid profiles along disease progression through the analysis of samples from patients with different ccRCC stages (I, II, III, IV).
- Develop a machine learning method applying Support Vector Machines (SVM) with LASSO to find discriminant feature panels for sample classification.
- Analyze the relative level change of discriminant features between the different classes and ccRCC stages.
- Assign identities to the discriminant lipids in order to understand the tumor biology.

Patient Cohort – Retrospective Study



The analysis was performed on age-matched serum samples

Serum samples were provided by the Public Oncologic Serum Bank (BPMSO) from Instituto de Oncología "Ángel H. Roffo" and Hospital Italiano de Buenos Aires

Experimental Design and Analytical Methods

Pre-Analytical Factors → Experiment → Data Pre-Processing → Data Processing, Classification & Prediction → Data Analysis → Biological Interpretation

Experimental Design

Blood Sera Patient Cohort

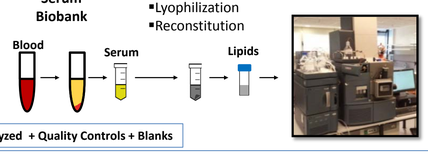
Sample Collection

BPMSO Serum Biobank

Sample Preparation

- Protein Precipitation
- Centrifugation
- Lyophilization
- Reconstitution

Separation and Detection of Metabolites



Samples Randomly analyzed + Quality Controls + Blanks

Reverse Phase Chromatography: Waters Acquity I-class UPLC system. Mobile phases: A= CH_3CN with 0.1% acetic acid (v/v), 10 mM $NH_4CH_3CO_2$ (60:40 v/v); B=isopropanol: CH_3CN (90:10 v/v) with 0.1 % acetic acid (v/v) and 10 mM $NH_4CH_3CO_2$. Waters BEH C18 UPLC column (1.0×100mm;1.7 µm particle size). Column temperature: 55 °C, Autosampler tray temperature: 5 °C.
Mass Spectrometry: Waters Xevo G2S QTOF mass spectrometer operated in negative ESI Mode. Data acquisition in MS^E mode. Probe capillary voltage: 2.3 kV; Sampling cone voltage: 30.0 V; Source temperature: 120 °C; Desolvation gas temperature: 300 °C. Acquisition range: *m/z* 50-1200

Metabolic Feature Extraction Rt_*m/z*; peak area

- Import Data into Progenesis QI
- Retention time alignment
- Peak Picking and Integration
- Adduct Deconvolution
- Solvent Correction
- Prevalence Filter
- Data curation

Feature Matrix: 386 (Rt_ *m/z*) pairs

Sample Classification by Statistical Analysis

- Develop and Train Cross-Validated SVM Models and Find Top Discriminant Features using LASSO (least absolute shrinkage and selection operator)
- Find the Discriminant Feature Panel that Maximizes Classification Accuracy of the Validation Set.
- Evaluate the Model´s Performance with an Independent Test Set.

Identification of Discriminant Features

- Accurate mass
- Isotopic pattern
- Tandem MS experiments
- Database search
- Validation with chemical standards

Pathway Mapping

Hypothesis generation

Biomarker Discovery

RESULTS

Multivariate Statistical Analysis Conducted on a 386-Feature Matrix

Support Vector Machines (SVM) coupled with LASSO Variable Selection Method

Table 1. Sample Distribution in Models: Control vs ccRCC.

Set	Total # of Samples	SI ccRCC	SII ccRCC	SIII ccRCC	SIV ccRCC	Control
Training	80	10	10	10	10	40
Validation	14	2	2	2	2	6
Test	70	44	2	6	12	6
Total	164	56	14	18	24	52

Table 2. Sample Distribution in Models: Early vs Late Stage ccRCC.

Set	Total # of Samples	Early Stage ccRCC SI + SII		Late Stage ccRCC SIII + SIV	
Training	70	24	11	15	20
Validation	14	4	3	3	4
Total	84	42		42	

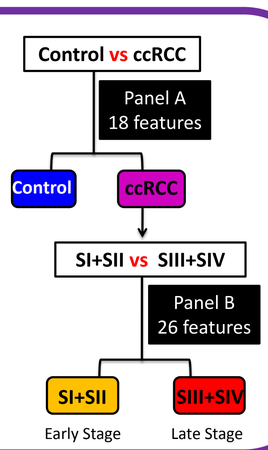
Test Set Stage I ccRCC n=28

Training Set
94% Specificity
100% Sensitivity
97% Accuracy

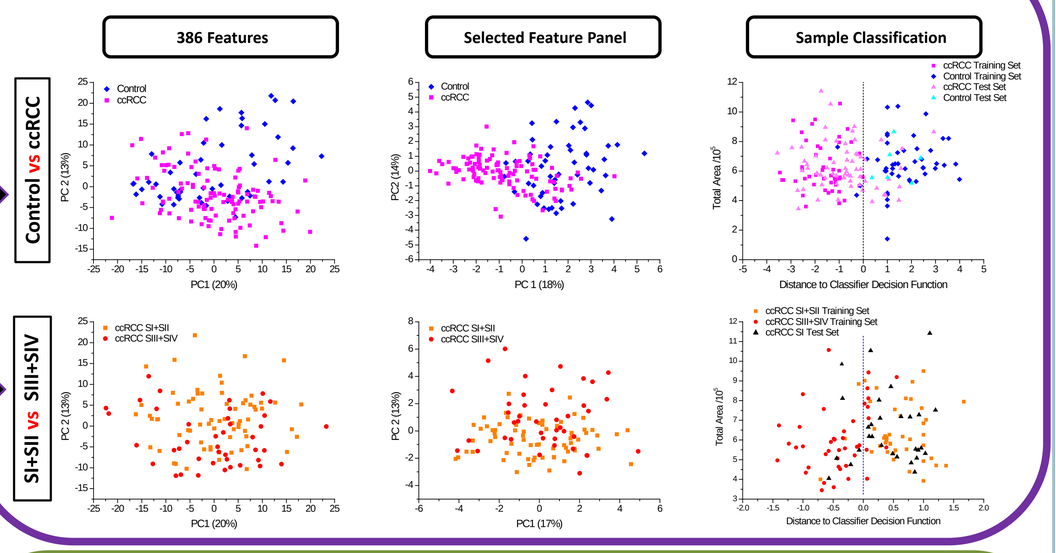
Test Set
100% Specificity
80% Sensitivity
81% Accuracy
AUC 0.89

Training Set
93% Accuracy (SI+SII)
76% Accuracy (SIII+SIV)
85% Accuracy

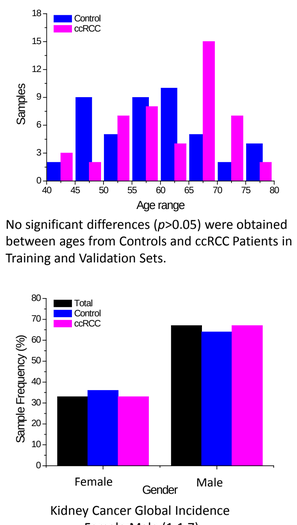
Test Set 82% Accuracy



Principal Component Analysis



SVM Model



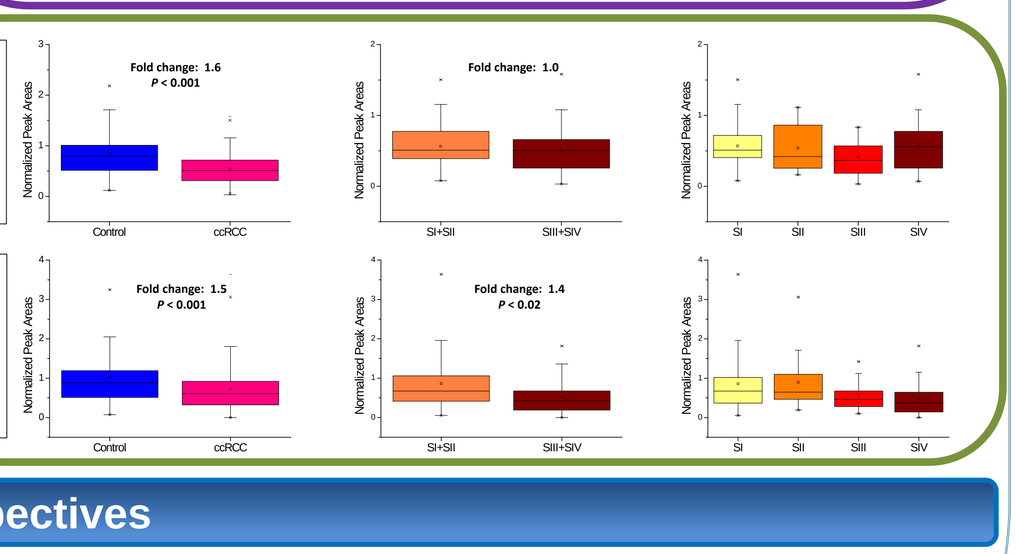
No significant differences (p>0.05) were obtained between ages from Controls and ccRCC Patients in Training and Validation Sets.

Kidney Cancer Global Incidence
Female:Male (1:1.7)

Table 3. Putative Identification of Discriminant Lipids based on Accurate Mass and Isotopic Pattern.

Panel	Features R_t_m/z	Adduct	Fold Change[a] Control/ccRCC or Early/Late	p value	Molecular Formula	Lipid Main Class[b]
B	0.62_369.1736	$[M-H]^-$	1.5	<0.02	$C_{47}H_{86}NO_{10}P$	Glycerophosphoserines
A	0.74_407.2787	$[M-H]^-$	3.1	<0.001	$C_{24}H_{40}O_5$	Bile Acids & Derivates
A & B	0.74_479.2813	$[M-H]^-$	1.6 / 1.0	<0.00002 / NS	$C_{27}H_{44}O_5S$	Secosteroids
A	0.92_243.1955	$[M-H]^-$	1.7	<0.03	$C_{14}H_{28}O_3$	Fatty Acids & Conjugates
A	0.94_183.1385	$[M-H]^-$	1.2	<0.04	$C_{25}H_{50}O_2$	Fatty Acids & Conjugates
B	1.04_197.1542	$[M-H]^-$	1.2	<0.03	$C_{12}H_{22}O_2$	Fatty Acids & Conjugates
A	1.23_199.1698	$[M-H]^-$	1.4	<0.0006	$C_{12}H_{24}O_2$	Fatty Acids & Conjugates
B	5.30_864.5743	$[M-H_2O-H]^-$	1.1	NS	$C_{48}H_{84}NO_{10}P$	Glycerophosphoserines
A	5.57_766.5408	$[M-H]^-$	1.4	<0.001	$C_{43}H_{78}NO_8P$	Glycerophosphocholines
B	7.12_854.5890	$[M-H]^-$	1.0	NS	$C_{47}H_{86}NO_{10}P$	Glycerophosphoserines
A & B	7.24_381.3720	$[M-H]^-$	1.5 / 1.4	<0.001 / <0.02	$C_{25}H_{50}O_2$	Fatty Acids & Conjugates
B	7.54_826.5970	$[M-H]^-$	1.1	NS	$C_{46}H_{88}NO_9P$	Glycerophosphoserines
B	7.71_722.5110	$[M-H]^-$	1.3	NS	$C_{41}H_{74}NO_7P$	Glycerophosphoethanolamines

[a]Fold changes are calculated as the ratio of median peak areas between compared classes. *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.
[b]According to LIPID MAPS Database.



Conclusions & Perspectives

- Lipid profiling coupled with SVM-LASSO multivariate analysis provided 2 discriminant feature panels for serum sample classification and prediction: i) 18 features allowed discriminating controls from ccRCC patients with 81% accuracy in an independent test set, and ii) 26 features allowed classifying stage I from stage III and IV ccRCC patients in an independent test set with 82% accuracy.
- 15 out of 18 discriminant lipids were significantly decreased in ccRCC serum samples compared to controls, in agreement with previous studies.[8]
- Since 2 discriminant lipids are common to both panels, 42 lipids would allow early ccRCC detection.
- Current work involves the identification of the discriminant lipids by tandem MS experiments and comparison with chemical standards.

Acknowledgment

We acknowledge CONICET, ANPCYT, and MINCYT for providing the funding, and the BPMSO Biobank from Instituto de Oncología A. H. Roffo and Hospital Italiano de Buenos Aires for providing the samples.

References

(1) Linehan, WM. *et al Nat. Rev. Urol.* **2010,** *7*, 277.
(2) Hu, B. *et al. Urol. Clin. North Am.* **2012,** *39*, 233.
(3) Graves, A. *et al. Immunotargets Ther.* **2013,** *2*, 73.
(4) Diamond, E. *et al. Crit. Rev. Oncol. Hematol.* **2015,** *96*, 518.
(5) De Meerleer, G. *et al. Lancet Oncol.* **2014,** *15*, e170.
(6) Hsieh, J.J. *et al. Nature Reviews Disease Primers* **2017,** *3*, 17009.
(7) Hakimi, A.A. *et al. Nat. Genet.* **2013,** *45*, 849.
(8) Lin, L. *et al. J. Proteome Res.* **2011**, *10*, 1396.
(9) Knott, M.E.; Manzi, M. *et al. J. Proteome Res.* **2018**, *17*, 3877.