



**HAL**  
open science

# Identification of biomarkers using-omics approach for the early detection of chronic kidney disease and its complications

Valérie Brunchault

► **To cite this version:**

Valérie Brunchault. Identification of biomarkers using-omics approach for the early detection of chronic kidney disease and its complications. Human health and pathology. Université Paul Sabatier - Toulouse III, 2018. English. NNT : 2018TOU30133 . tel-02146040

**HAL Id: tel-02146040**

**<https://theses.hal.science/tel-02146040>**

Submitted on 3 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

---

**Présentée et soutenue par :**

**Valérie BRUNCHAULT**

**le** mercredi 19 septembre 2018

**Titre :**

Identification of biomarkers using -omics approach for the early detection of chronic kidney disease and its complications

---

**École doctorale et discipline ou spécialité :**

ED BSB : Maladie métaboliques et cardiovasculaires

**Unité de recherche :**

INSERM U1048/ Institut des Maladies Métaboliques et Cardiovasculaires (I2MC) - Equipe 12

**Directeur/trice(s) de Thèse :**

Dr. Bénédicte BUFFIN-MEYER

Dr. Julie KLEIN

**Jury :**

Pr. Isabelle CASTAN-LAURELL (Présidente du Jury)

Dr. Fabiola TERZI (Rapporteur)

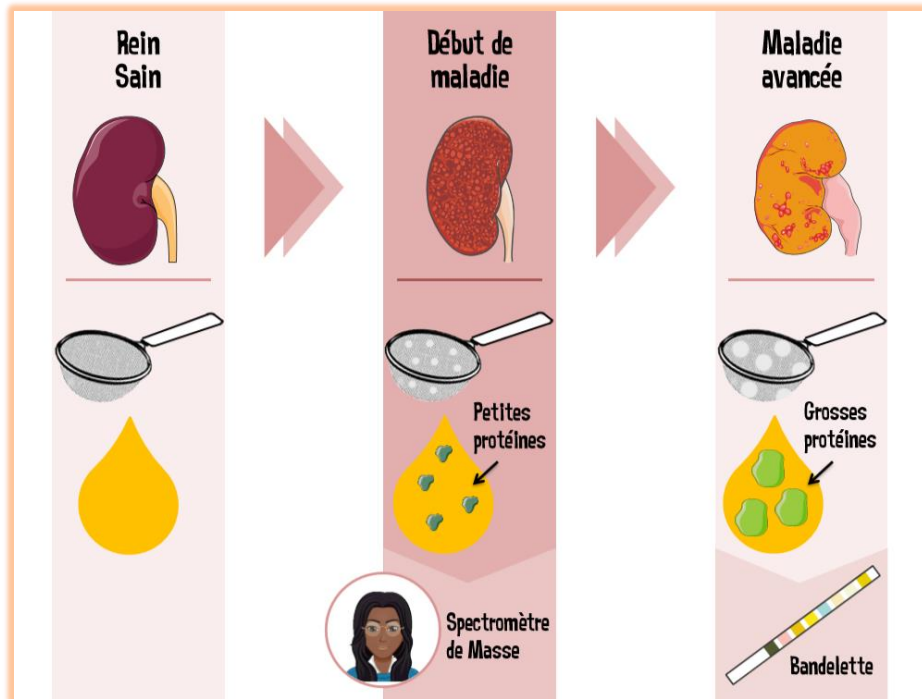
Dr. Christos CHATZANTONIOU (Rapporteur)

Dr. Bénédicte BUFFIN-MEYER

Dr. Julie KLEIN



# Ma thèse en 180 secondes



*\*Faire semblant de tenir un verre contenant un liquide à la main, le tourner légèrement, le sentir, regarder sa couleur, le goûter ....* Nous sommes au 16<sup>ème</sup> siècle, et les médecins observaient la couleur et l'odeur de l'urine pour diagnostiquer des maladies, et mieux encore... ils la dégustaient ! Il paraît que l'urine des patients atteints de diabète est sucrée. Hmm le pipi c'est délicieux ! Mais...d'où vient l'urine ? L'urine est produite par les reins. Ces organes ont la forme d'un haricot et font environ la taille de notre poing. Dans notre corps, ils agissent comme des filtres géants et servent à nettoyer, ♪♪ astiquer le sang toujours pimpant ♪♪ Hmm, oui ! Pour vous donner une idée, un enfant de 10 ans a en moyenne 2,5L de sang et ce sang va passer 80 fois par ses reins, tous les jours, pour être nettoyé. Les déchets filtrés sont éliminés dans l'urine. Cependant, cette urine contient plein d'informations sur notre organisme et plus particulièrement sur l'état de fonctionnement des reins.

En soirée, on me taquine souvent en m'appelant 'Dr Pipi' parce que pour mon projet de thèse, j'étudie la composition de l'urine afin d'aider les médecins à diagnostiquer précocement les maladies rénales chez les enfants. De nos jours, pour diagnostiquer une maladie rénale le médecin trempe une bandelette dans votre urine afin de détecter la présence de grosses protéines. Ces dernières se retrouvent dans l'urine lorsque le filtre est endommagé. En effet, au fur et à mesure que le filtre cumule des dommages, des trous vont s'y former et s'agrandir avec le temps. Une fois ces trous suffisamment grands, ils permettent ainsi aux grosses protéines de passer dans l'urine.

Mais pour pouvoir diagnostiquer plus précocement les dommages aux reins, il faudrait identifier des petites protéines qui passent le filtre lorsque les trous sont encore petits. Pour les identifier, au laboratoire, j'ai analysé des échantillons d'enfants malades et d'enfants en bonne santé avec un appareil qu'on appelle un spectromètre de masse. Il me permet d'identifier<sup>1</sup> des protéines 20 fois plus petites mais aussi de mesurer leur quantité<sup>2</sup> dans l'urine. J'ai ensuite comparé les urines des enfants malades et celles des enfants en bonne santé, afin de voir ce qui pourrait être différent. Cela m'a permis d'identifier une combinaison de petites protéines qui sont présentes très tôt dans l'urine des enfants malades et absentes dans l'urine des enfants sains. Diagnostiquer plus précocement la maladie permettrait ainsi de prendre en charge les patients plus tôt et surtout d'adapter le traitement afin de retarder l'apparition de plus de dommages aux reins. Je vous remercie de votre attention, et j'espère que durant ces 180 secondes je n'aurais pas... « pissé » dans un violon !



# Résumé

---

Diagnostiquer précocement les maladies est un défi à relever pour améliorer la prise en charge des patients concernés et leur offrir une meilleure qualité de vie. Les analyses 'omiques', qui quantifient globalement, simultanément et sans *a priori* l'abondance de milliers de molécules dans les liquides biologiques, s'avèrent très prometteuses pour l'identification de biomarqueurs précoces des maladies complexes. Dans ce contexte, mon travail de thèse avait pour objectif de développer des outils de diagnostics, à partir d'analyses du peptidome et métabolome urinaire, pour détecter précocement la présence d'une maladie rénale chronique (MRC) et la survenue de ses complications cardiovasculaires.

La première étude, insérée dans le projet européen 4C (*Cardiovascular Complications in Children with Chronic kidney disease*), s'est centrée sur les complications cardiovasculaires associées à la MRC en pédiatrie. Ces complications constituent la principale cause de mortalité des enfants en insuffisance rénale, et leur diagnostic précoce est impossible à ce jour. En analysant par électrophorèse capillaire couplée à la spectrométrie de masse (CE-MS) le peptidome urinaire de 86 enfants souffrant, ou non, de complications cardiovasculaires secondaires à la MRC, nous avons identifié des peptides qui permettent de prédire à l'avance les patients à haut risque cardiovasculaire : 190 peptides étaient associés à l'épaississement de la paroi carotidienne (AUC 0.87, sensibilité 80%, spécificité 100%) et 22 peptides prédisaient l'augmentation de la rigidité artérielle (AUC 0.83, sensibilité 83%, spécificité 70%).

Le second projet relevait de la médecine vétérinaire. Dans cette étude menée sur 50 chiens avec et sans MRC, nous avons caractérisé pour la première fois le peptidome urinaire canin *via* la technologie CE-MS et nous avons découvert 133 peptides urinaires associés à la MRC. Ces derniers ont permis de diagnostiquer la présence d'une MRC dans 80% des chiens.

Les métabolites sont mieux corrélés au phénotype que les autres strates moléculaires. Cependant l'apport de la métabolomique en clinique est encore limité, dû au manque de technologies analytiques performantes. Le troisième objectif de ma thèse était donc de mettre au point une procédure de dosage par CE-MS des métabolites urinaires. Grâce à une méthode unique de normalisation interne, basée sur l'utilisation de métabolites endogènes stables, il est maintenant possible d'analyser le contenu en métabolites d'un même échantillon urinaire avec une très haute reproductibilité sur le long terme (4 ans). Comme preuve de concept, nous avons mis en évidence, *via* cette procédure, la présence d'une combinaison de 32 métabolites dans l'urine qui permet de repérer avec une sensibilité de 76% et une spécificité de 86% les nouveau-nés porteurs d'une malformation rénale obstructive.

Enfin, la quatrième problématique s'inscrivait dans une démarche translationnelle. Son but était de développer des aptasenseurs capables de détecter avec de hautes affinités et spécificités les biomarqueurs d'origine omique, pour un diagnostic simple, rapide et à moindre coût. La cible choisie

était un fragment urinaire de l' $\alpha$ -1-antitrypsine, qui est ~1000 fois plus abondant chez les adultes atteints de MRC que les chez les sains. La sélection de l'aptasenseur s'est faite par le Systematic Evolution of Ligands by EXponential enrichment (SELEX). Nous présentons ici les travaux préliminaires de la mise au point du SELEX sur cette cible.

En conclusion, cette thèse démontre le potentiel de l'analyse du contenu urinaire en peptides et métabolites pour le diagnostic précoce des pathologies complexes telles que la MRC et les complications cardiovasculaires associées. De plus l'obtention d'aptasenseurs dirigés contre ces biomarqueurs précoces et utilisables au chevet du patient devrait révolutionner dans le futur les méthodes diagnostiques.

# Abstract

---

Early diagnosis of diseases is a big challenge to improve patients' health and quality of life. 'Omics' analyses, which allow the global and simultaneous quantification of the relative abundance of thousands of molecules in biological fluids are promising for the identification of early biomarkers of complex diseases. In this context, the objective of my thesis was to develop diagnostic tools, based on urinary peptidome and metabolome analyses, for the early detection of chronic kidney disease (CKD) and associated cardiovascular complications.

The first study, as part of the 4C European project (Cardiovascular Complications in Children with Chronic kidney disease), focused on analyzing the cardiovascular complications associated to CKD in children. These complications are the main cause of mortality in children with CKD and their early diagnosis is impossible for now. Analysis of the urinary peptidome of 86 children with or without cardiovascular complications associated to CKD by capillary electrophoresis coupled to mass spectrometry (CE-MS), led to the identification of two sets of peptides for the early prediction of high cardiovascular risk in pediatric patients: 190 peptides were associated to an increase of the carotid intima-media thickness (AUC 0.87, sensitivity 80%, specificity 100%) and 22 peptides were associated to an increase in arterial stiffness (AUC 0.83, sensitivity 83%, specificity 70%).

The second study falls in the field of veterinary medicine. In this study, carried out on 50 dogs with or without CKD, we analyzed for the first time the canine urinary peptidome using the CE-MS technology. We identified 133 urinary peptides associated to CKD allowing an accurate diagnosis of CKD in 80% of the dogs.

Metabolites correlate best to phenotype compared to other molecular traits. However, the use of metabolomics for identification of clinically relevant biomarkers is very limited due to the lack of high-performance analytical technologies. The third part of my thesis was to develop a procedure for the quantification of urinary metabolites by CE-MS. Using a unique method of internal normalization based on endogenous and stable metabolites, we can now analyze the metabolite content of the same urine sample with a high reproducibility over the long-term (4 years). As a proof-of-concept, we demonstrated that this developed procedure led to the identification of a set of 32 urinary metabolites that allow the early identification of newborns with an obstructive kidney anomaly with a sensitivity of 76% and a specificity of 86%.

Finally, the fourth study was dedicated to improving translational research. The aim was to develop aptasensors able to detect 'omics'-identified biomarkers with a high affinity and specificity to obtain a simple, rapid and low-cost diagnostic test. The biomarker chosen as target is a urinary fragment of  $\alpha$ -1-antitrypsin, which is ~1000 more abundant in adults with CKD compared to healthy subjects.



Aptasensors were selected by the Systematic Evolution of Ligands by EXponential enrichment (SELEX). Here we present preliminary work on the development of the SELEX for our target.

In conclusion, this thesis shows the strength of the urinary content, in terms of peptides and metabolites, for the early diagnosis of complex pathologies like CKD and the associated cardiovascular complications. Moreover, the selection of aptasensors targeting these early biomarkers and that can be used at bedside, will revolutionize future diagnostic methods.

# Remerciements

---

Ce travail de thèse clôt un chapitre de ma vie qui est loin d'être le fruit d'un parcours solitaire. Je tiens tout d'abord à remercier mon directeur de laboratoire, **Joost Schanstra**, de m'avoir accueillie au sein de son équipe. Merci à mes deux directrices de thèse, **Bénédicte Buffin-Meyer** et **Julie Klein** pour leur encadrement tout au long de cette thèse. Je vous suis reconnaissante, à tous les trois, pour le temps conséquent que vous m'avez accordé et pour la qualité pédagogique et scientifique enseignée. Merci aussi de m'avoir intégrée à plusieurs projets très stimulants sur le plan scientifique. Sans votre contribution, vos conseils, votre aide, je ne serais pas arrivée là. Merci également à nos proches collaborateurs : **Petra Zürbig**, **Pedro Magalhães**, **William Mullen**, **Lena Pelander**, **Franz Schaefer**, **Kevin Kunzmann**, **Vincent Ecochard**, **François Couderc**, **Jean-José Maoret**, **Frédéric Martins**, **Jason Iacovoni** et **Audrey Ric**.

Je tiens à remercier **Christian Touriol** et **Laurent Debrauwer** qui m'ont suivie tout au long de mon parcours de thèse. Je les remercie pour leurs aides, les encouragements et les conseils qu'ils m'ont prodigués lors de nos réunions afin d'assurer l'avancement de mes travaux de thèse.

Merci également aux membres de mon jury de thèse d'avoir accepté d'évaluer mes travaux : **Pr. Isabelle Castan-Laurell**, **Dr. Fabiola Terzi** et **Dr. Christos Chatziantoniou**. Merci pour cet échange que nous avons pu avoir lors de ma soutenance.

J'adresse aussi toute ma gratitude à un grand nombre de personnes dont la générosité et la bonne humeur m'ont permis d'avancer durant ces trois années.

- Merci à **Ben** de m'avoir formée à l'électrophorèse capillaire couplée à la spectrométrie de masse. La joie n'était pas toujours présente lorsque les appareils étaient quelque peu capricieux, mais la bonne humeur était toujours au rendez-vous. Merci pour ce délicieux Saint-Honoré !

- Merci également à **Marie** et **Hélène** alias grande saucisse, mes deux copines de 'bocal', pour les fous rires et les moments de confidences que nous avons pu avoir.

- Merci à **Eloïse**, notre 'Miss Joie', pour ta gentillesse, ton écoute, ton enthousiasme, ton soutien, ton moral d'acier et ta disponibilité sans faille. Tu es une fille en or, un médecin génial et une superbe danseuse ! Je tiens aussi à remercier **Ophélie**, une autre perle rare du labo. Merci aussi pour ta zenitude (sauf quand tu es au volant), ta patience, ton écoute, ta créativité, ton courage et les fous rires. Merci pour les gâteaux que tu nous as ramenés, tu es vraiment un vrai chef ! (Je n'oublie pas qu'il faut que j'essaie le West Coast !). Enfin merci à **Audrey** pour ton enthousiasme et ta

créativité, sans oublier les goûters et les muffins partagés. Vous êtes exceptionnelles les filles. Bon courage et tout le meilleur pour la suite !

- Merci à **Guyène** pour ta gentillesse et grâce à qui je me sentais moins seule lorsque je ramenais des knackis à midi.

- Merci également à **Adeline** pour les bons moments passés à discuter de danse, plus particulièrement de d@nse en ligne.

- Je tiens aussi à remercier les anciens de l'équipe/de l'étage : **Jean-Loup**, merci pour les encouragements *mèm si ou lé loin*. **Marion** et **Dimitri** merci pour les bons moments et les fous rires que nous avons pu partager dans le bureau. Merci aussi à **Marie-Hélène** de l'équipe 8. Tu es l'une des personnes les plus merveilleuses que j'ai pu rencontrer au labo. Tes encouragements et tes conseils ont vraiment été inestimables. Merci également à **Marion** et **Roberto** de l'équipe 7.

- Merci à mes anciens encadrants de Master, **Eric Lacazette** et **Henrik Laurell**.

-Merci infiniment à **Delphine Milhas** et à **Caroline Conte** pour l'aide durant les enseignements.

- Merci **Eric** pour ta gentillesse et ton aide précieuse sur mon projet des aptamères car il faut le dire, les aptamères c'est la galère !

- Merci à **Chantal Burgos** pour toutes les commandes dont elle s'est occupée durant ma thèse et merci à **Céline Terrade** de s'être occupée des missions pour les déplacements et les congrès.

- A la belle brochette de thésards : **Francky**, **Camille**, **Ana** et plus particulièrement **Alexis** alias 'le schtroumpf grognon'. Je vous souhaite une belle thèse et bon courage pour la suite. N'oubliez pas de continuer les repas sushis/pizza au labo :). Merci infiniment Alexis pour ton aide et tes conseils durant la fin de ma thèse.

- Merci **Colette** d'avoir été notre coach pour les activités de vulgarisation que j'ai découvertes et durant lesquelles je me suis éclatée. Vive l'atelier Dr. Pipi et les haricots magiques !!

- Malgré un agenda plus que rempli, je remercie chaleureusement le **Pr. Stanislas Faguer** d'avoir pris le temps de me donner des conseils avisés sur le projet 4C et pour les encouragements durant ces trois années.

Aux copains qui me soutiennent depuis des années, chacun à leur manière. Tout d'abord **Mathoo**, **Hugo** et **Marjo**, vous m'avez connue dès mes premiers jours en France sur les bancs de la fac. Merci d'avoir toujours été là pour m'encourager. Merci aussi aux amis rencontrés durant ces dernières

années et dont le soutien a été très important : **Jim** - you've always got the right words and were always here at anytime, **Jean-Louïs-Marie** merci pour ta bonne humeur, promis je te ferai des macarons ;) , **Laura** merci d'avoir été là dans les bons et les mauvais moments. Même si géographiquement tu es un peu plus loin, tu restes toujours présente et ton amitié m'est chère. Enfin, merci aussi à **Gwendo** et bon courage en cette fin de thèse et aussi pour le post-doc (au soleil ?!).

Mille mercis aux copains du CBD : Merci à **Cyril**, l'homme aux multiples talents (dessinateurs, cordon-bleu, presque Cupidon) et à **Maleaume**, qui a toujours eu l'oreille très attentive (arrête de raconter aux gens l'histoire de la bino !). Bon courage pour cette fin de thèse. Merci, or should I say thank you, **Jérôme**, pour les bons moments passés. J'espère que tout se passe bien durant ton post-doc à Cambridge. Merci à **Manon** et bon courage pour ce nouveau post-doc à Bruxelles. Merci particulièrement aussi à **Manue**, un vrai bout de femme en or ! merci pour tous tes conseils et tes encouragements. Merci à **Mikhaïl** et **Angie** pour votre soutien et toutes les soirées passées ensemble.

Merci aux copains geek : **Marie**, (petit Cupidon... mais pas seulement ! Tu es géniale :) ) et **JC, Pauline** et **Mathieu** ainsi que **Florie** et **Hubert** pour les bons moments qui en plus de contribuer aux moments de détente durant la thèse, ont aussi été un soutien.

Merci du fond du cœur à **Pascale** et **Raph** auprès de qui j'ai toujours trouvé des conseils, un soutien et du réconfort quand tout n'était pas rose. Merci pour les soirées piscine et les ateliers créatifs'. Merci également à la championne **Aurore**, à **Patrick** le fantastique et à **Julie** la Star ! Merci à P. et P. de m'avoir ouvert les portes de chez vous, merci pour tout.

Enfin, merci à ma famille et à ma belle-famille. **Maman, Papa** et **Tim** merci de m'avoir encouragé et soutenu malgré la distance qui nous sépare. Merci à **Rox** d'avoir pu faire le déplacement pour ma thèse, tu es géniale. Bon courage à toi pour la thèse ! Merci également à la **famille Olivencia** pour tous les moments que nous partageons, qui me font me sentir comme à la maison et qui inconsciemment ont contribué à me détendre surtout durant les moments les plus difficiles de la thèse. Merci beaucoup.

Ma thèse a aussi été une très belle rencontre. Je tiens à remercier particulièrement **Adrien**, celui qui a croisé ma route en milieu de thèse, qui m'a accompagné jusqu'à la fin de celle-ci et qui continue de m'accompagner dans la vie de tous les jours. Merci pour ton soutien sans faille. Je ne souhaite qu'une chose, c'est de continuer ma route avec toi à mes côtés.

A ceux que j'ai oublié, j'espère que vous ne m'en voudrez pas et je vous remercie d'avoir été là pour moi. MERCI à tous !



# Table of Contents

<b>Introduction</b> .....	1-46
<b>I. <u>Cardiovascular complications in children with chronic kidney disease</u>.....</b>	<b>1-15</b>
1. Epidemiology of CKD in children .....	1
2. Etiology of CKD in children .....	1
i. Main etiologies.....	1
ii. Factors influencing CKD development in children.....	4
3. Mortality in children with CKD .....	5
4. Cardiovascular complications associated to CKD in children .....	5
i. Cardiac complications .....	6
a. Left Ventricular Hypertrophy (LVH).....	7
b. Outcome and challenges of cardiac hypertrophy .....	8
ii. Vascular complications .....	9
a. Vascular calcification and arterial stiffness.....	9
b. Atherosclerosis and arterial wall thickening .....	11
a. Outcome and challenges of vascular complications.....	14-15
<b>II. <u>Chronic Kidney Disease in dogs</u> .....</b>	<b>16-26</b>
1. Epidemiology .....	16
2. Etiology of CKD in dogs.....	17
i. Ageing .....	17
ii. Periodontal diseases .....	17
iii. Acute kidney injury (AKI) .....	17
iv. Hypertension .....	18
v. Genetics.....	18
vi. Congenital anomalies .....	19
3. Complications associated to CKD in dogs and their treatments .....	19
i. Nutritional disorders.....	19
ii. Chronic kidney disease mineral bone disorder.....	19
iii. Metabolic acidosis.....	19
iv. Cardiac remodeling .....	20
v. Anemia .....	21
4. Diagnosis of CKD in dogs .....	21
i. Actual clinical biomarkers of CKD in dogs .....	21

a.	Glomerular filtration rate (GFR) .....	21
b.	Urine specific gravity (USG) .....	22
c.	Proteinuria .....	23
d.	Azotemia .....	23
e.	Kidney hyperechogenicity on ultrasound imaging.....	24
ii.	Drawbacks of actual biomarkers for CKD diagnosis in dogs .....	24
a.	Creatinine .....	24
b.	Urine specific gravity .....	25
c.	Proteinuria .....	25
d.	Kidney hyperechogenicity on ultrasound imaging.....	25
iii.	Challenge for diagnosis of early CKD in dogs.....	25-26

### **III. Peptidome and metabolome analysis by mass spectrometry ..... 27-38**

1.	Introduction .....	27
2.	Biomarkers .....	28
i.	Definition .....	28
ii.	Identification of biomarkers .....	28
a.	Clinical issues.....	28
b.	Analytical issues.....	28
c.	Statistical and validation issues.....	29
3.	Mass spectrometry.....	29
i.	Principle .....	30
ii.	Coupling with separation techniques .....	31
a.	Liquid chromatography .....	32
b.	Capillary electrophoresis.....	32
4.	Biological fluids .....	33
5.	Urinary peptidome analysis by CE-MS.....	34
i.	Kidney disease .....	35
ii.	Cardiovascular disease (CVD) .....	36
6.	Urinary metabolome analysis by CE-MS for CKD.....	37
	Challenge for the study of urinary metabolome .....	38

### **IV. Aptamers: 'omics' translation for bed side detection of biomarkers ..... 39-46**

1.	What are aptamers? .....	39
i.	Definition: .....	39

ii. Structure:.....	39
iii. Advantages:.....	39
2. Aptamer selection.....	40
i. Library of oligonucleotides .....	40
ii. SELEX .....	41
a. Principle .....	41
b. SELEX using affinity columns .....	42
c. Capillary electrophoresis SELEX (CE-SELEX).....	43
3. Clinical applications of aptamers .....	44
i. Aptamers as therapeutic agents .....	44
ii. Aptamers as biosensors .....	45
Challenge for the development of biosensors for biomarker detection.....	46
<b>Study 1: Cardiovascular complications (CVD) in children with CKD.....</b>	<b>47-74</b>
<b>Study 2: CKD diagnosis in dogs.....</b>	<b>75-98</b>
<b>Study 3: Identification of metabolite biomarkers by CE-MS technology.....</b>	<b>99-100*</b>
<b>Study 4: Aptamers selection for biosensor development.....</b>	<b>101-113</b>
<b>Conclusion.....</b>	<b>114-117</b>
<b>List of references.....</b>	<b>118-133</b>





# List of Abbreviations

---

2D-PAGE – two-dimensional polyacrylamide gel electrophoresis

ACE – Angiotensin-converting enzyme

ACEi – Angiotensin-converting enzyme inhibitor

ADMA – Asymmetric dimethylarginine

AIC – Arterial intimal calcification

AKI – Acute kidney disease

AMC – Arterial medial calcification

AuNPs – Gold nanoparticles

BUN – Blood urea nitrogen

CE – Capillary electrophoresis

CE-MS – Capillary electrophoresis coupled to mass spectrometry

CE-SELEX – Systematic evolution of ligands by exponential enrichment by capillary electrophoresis

cIMT – carotid intima-media thickness

CKD – Chronic Kidney Disease

CKD MBD – Chronic kidney disease mineral bone disorder

CTAB – hexadecyltrimethylammonium bromide

CVD – Cardiovascular disease

Da - Daltons

dL – deciliter

DNA – Deoxyribonucleic acid

eGFR – estimated glomerular filtration rate

EM – Electrophoretic mobility

EOF – Electroosmotic flux

ESI – Electrospray ionization

ESRD – End-stage renal disease

FGF-23 – Fibroblast growth factor 23

GN – Glomerulonephritis

HA – Hippuric acid

IxS - indoxyl sulphate

IRIS – International Renal Interest Society

kDa - Kilodaltons

KDIGO – Kidney Disease: Improving Global Outcomes guidelines

LC – Liquid chromatography

LDL – low-density lipoproteins

LVH – Left ventricular hypertrophy

LVMI – Left ventricular mass index

MALDI – Matrix-assisted laser desorption ionization

mRNA – messenger RNA

MS – Mass spectrometry

pCS – *p*-cresyl sulphate

pmarp – per million of the age-related population

PWV – Pulse wave velocity

RNA – Ribonucleic acid

RRT – Renal replacement therapy

SELEX – Systematic evolution of ligands by exponential enrichment

SRNS – Steroid-resistant nephrotic syndrome

ssDNA – Single-stranded DNA

TOF – Time-of-flight

UPC – urine protein-creatinine ratio

USG – Urine specific gravity

VSMCs – Vascular smooth muscle cells

# *Introduction*

---

# I. Cardiovascular complications in children with chronic kidney disease

## 1. Epidemiology of CKD in children

Chronic kidney disease (CKD) is characterized by irreversible damage of the kidneys leading, in most of the cases, progressively to end-stage renal disease (ESRD). There are 5 CKD stages, stage 5 being the most severe and is equal to ESRD [1]. The only solution of survival for patients with ESRD is renal replacement therapy (RRT) by dialysis or kidney transplantation. Epidemiological studies report that CKD is the 14<sup>th</sup> leading cause of death worldwide, accounting for 12.2 deaths per 100,000 individuals. According to the projections from the Global Health Observatory, this death rate will continue to rise to reach 14 per 100,000 individuals by 2030 [2].

The pediatric incidence of CKD stages 3-5 in Europe is approximately 11-12 per million of the age-related population (pmarp) and the prevalence is 55-60 pmarp [3]. For the past 30 years, the prevalence and the incidence of RRT in children has increased. In 2008, the worldwide median incidence of RRT in children aged less than 20 years old was about 9 pmarp [4]. A recent study, published in February 2018, examined a large, nationwide cohort of 1,521,000 Israeli adolescents (mean age 17.7 years) before conscription (obligatory military service) and followed them up for 30 years. This study showed that even if there is no evident decline of kidney function during adolescence, after a childhood episode of mild kidney disease or injury, early renal damages and renal scarring contribute to higher risks of ESRD and onset of ESRD at a younger age compared to the normal population [5].

## 2. Etiology of CKD in children

### i. Main etiologies

CKD affects both adults and children, but the causes are different between these two populations. The causes of CKD in children are mostly due to primary kidney disease, as we will see below, while in adults, CKD is mostly secondary to other diseases, mainly diabetes and hypertension [6-9].

Congenital anomalies of the kidney and urinary tract (CAKUT) are the main causes of CKD in children (about 50%), followed by other key causes including steroid resistant nephrotic syndrome (SRNS) and glomerulonephritis both with a frequency of about 10% (Table 1).

**Table 1.** Causes of CKD in children under 25 years old and their relative frequency. From NAPRTCS - North American Pediatric Renal Trials and Collaborative Studies [10].

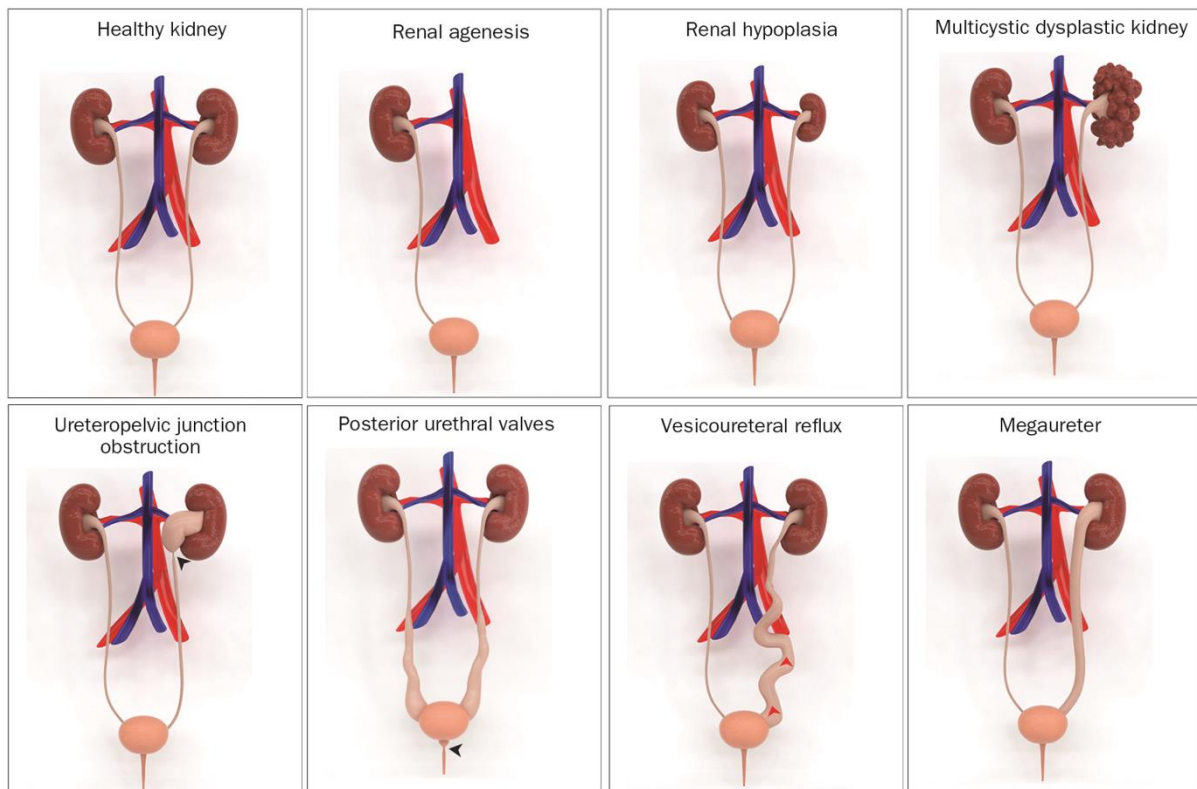
DIAGNOSTIC GROUPS	Total
Congenital anomalies of the kidney and urinary tract (CAKUT)	49.1%
steroid-resistant nephrotic syndrome (SRNS)	10.4%
Chronic glomerulonephritis (GN)	8.1%
Renal cystic ciliopathies	5.3%
Hemolytic uremic syndrome	2.0%
Nephrolithiasis/nephrocalcinosis	1.6%
Other	20.9%
Unknown	2.6%
Total	100% (N=7,037)

CAKUT are due to structural anomalies that occur during kidney development (nephrogenesis) and urinary tract development. They may be further divided into renal agenesis (absence of kidney), hypoplasia (small-sized kidneys with intact nephrons but in reduced number), dysplasia (disrupted and undifferentiated renal parenchyma), ureteropelvic junction obstruction (obstruction of urine outflow at the intersection of the pelvis and the ureter), posterior urethral valves (obstruction of urine outflow at the exit of the bladder), vesicoureteral reflux (lower urinary tract dysfunction due to a variety of reasons) or megaureter (dilated ureter), as shown in Figure 1. CAKUT account for approximately 20-30% of all congenital abnormalities and for each 1000 births, about 3-6 cases are reported [11, 12]. Data collected during the follow-up of patients have shown that CAKUT are the cause of ~50% and 7% of ESRD worldwide in the pediatric and adult population, respectively [12].

Nephrotic syndromes are characterized by excessive proteinuria ( $> 50 \text{ mg.kg}^{-1}.\text{day}^{-1}$  in children). They are classified based on their response to glucocorticoids treatment as either steroid-sensitive (glucocorticoids treatment is efficient and results in complete remission) or steroid-resistant (SRNS) [13]. SRNS may be further subdivided into focal segmental glomerulosclerosis, congenital nephrotic syndrome and membranous nephropathy. Histological analysis of biopsies from SRNS children showed that most of these children had glomerular epithelial damage and podocyte effacement which decrease the stability of the glomerular filtration barrier [14-17] and contribute to kidney lesions leading to CKD. SRNS is responsible of 5-20% of ESRD worldwide [18].

Chronic glomerulonephritis (GN) is the third leading cause of CKD in children. GN is a glomerular inflammation that is detected based on histopathological observations and described by a combination of 3 clinical observations: hematuria, hypertension and proteinuria [19]. They are further subdivided

into lupus nephritis, familial nephritis (Alport Syndrome), chronic glomerulonephritis, membranoproliferative glomerulonephritis, IgA nephritis and idiopathic crescentic GN [20]. On biopsies, GN kidneys display glomerular basement membrane damage, which is also associated to podocyte effacement and consequently leads gradually to nephron loss [21].



**Figure 1.** 3D representations of some distinct types of CAKUT. The classification is based on anatomical changes adapted from [11].

## ii. Factors influencing CKD development in children

CKD has been shown to vary with race. Prospective analyses carried in Europe and the USA have shown that CKD was more frequent in black compared Caucasian children. For example, focal segmental glomerulosclerosis was three times more common in blacks than in Caucasians (19 vs 6%) and particularly among black adolescents (35%) when children aged between 0 – 20 years were compared [2, 22, 23].

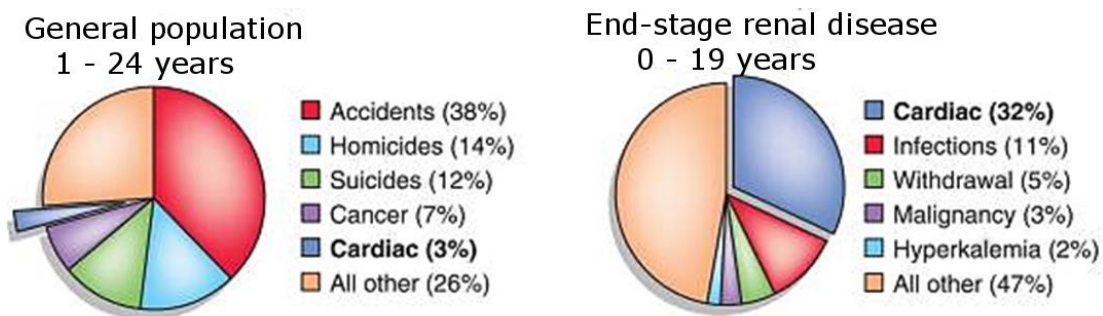
Moreover, certain age groups are more susceptible to develop CKD. A statistical analysis of the different age groups of children revealed that adolescence was the age group with the highest incidence of CKD [4, 23].

Finally, several studies have shown that there was a higher frequency of CKD in boys than girls due to the higher frequency of CAKUT in this gender [3, 11, 24].

This chapter on the different etiologies of CKD in children underlines a critical point to consider which is the occurrence of the disease. CKD occurs very early in children's life and is accompanied by complications associated to CKD, thereby reducing life expectancy.

### 3. Mortality in children with CKD

Data available from 2006-2008 in the US report that the mortality rate in the general pediatric population was 0.31 per 1000 patient-years at risk. However, an analysis of the mortality rate in children on ESRD showed that this rate was much higher: 35.6 for children on dialysis and 3.5 for children who received a kidney transplant per 1000 patient-years at risk. A high number of studies, including observations from international registries, to understand the factors contributing to this increase in death rate showed that cardiovascular diseases (CVD) were the leading causes of death in children with ESRD (Figure 2) [25-28].



**Figure 2.** Main causes of death in the children from the general population and children suffering from ESRD [28].

### 4. Cardiovascular complications associated to CKD in children

As indicated above, the main causes of death in children with ESRD are due to CVD. To improve clinical management of CKD children, several studies have investigated the cardiovascular risk, mechanisms of the disease and markers of CVD complications in the pediatric population with CKD. The first complication that develops in the initial stages of CKD is left ventricular hypertrophy. At more advanced stages vascular calcification and associated wall stiffness, as well as atherosclerosis and associated increased wall thickness are observed.



## **i. Cardiac complications**

### **a. Left Ventricular Hypertrophy (LVH)**

#### **Definition and detection**

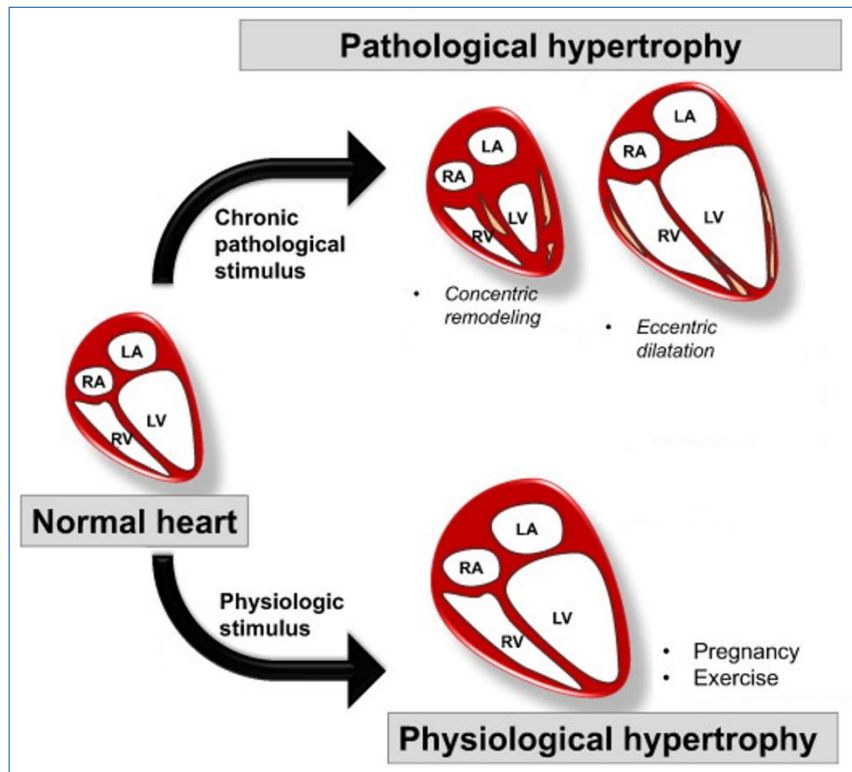
Cardiac hypertrophy is characterized by an increase of the ventricular wall thickness and/or internal chambers size (Figure 3). This enlargement of the heart occurs in physiology consequently to exercise or pregnancy. Apart from these two exceptions, cardiac hypertrophy is a response to pathological conditions causing impaired cardiac function [29].

To assess cardiac hypertrophy, LVH is usually evaluated by two-dimensional echocardiography [30-34]. This allows evaluation of the cardiac geometry together with chamber size in real-time, and quantification of interventricular septal thickness, posterior wall thickness and internal LV diameter both in systole and diastole. Left ventricular mass index (LVMI) is then computed and normalized for height for expression in  $\text{g}/\text{m}^{2.16}$ . LVMI is independent of gender. Moreover, relative wall thickness, i.e. the mean thickness of the septal and posterior wall divided by the LV end-diastolic dimension is also calculated [35]. However, the two-dimensional echocardiography technique has geometrical limitations, depending on the angle at which the measurement is done, and potential interobserver variability, depending upon image quality [36].

LVH is defined as either concentric or eccentric (Figure 3). LVH concentric remodeling is characterized by an elevated wall thickness and a normal left ventricular mass index (LVMI), while eccentric remodeling is characterized by a normal wall thickness and an elevated LVMI due to an increase in chamber size.

#### **LVH in children with CKD**

Cross-sectional studies on the analysis of the cardiac geometry in children suffering from CKD showed that cardiac hypertrophy is one of the most important independent marker of cardiovascular risk associated to CKD, with a prevalence of 17 – 49% in mild to moderate CKD and increasing to 75% in ESRD [37, 38] [39, 40]. Analysis of LVH in children suffering from CKD has shown that eccentric LVH with elevated LVMI (i.e. above the 95<sup>th</sup> percentile of the general healthy population) was more recurrent than concentric LVH [41-45]. Of note, it has been shown that girls have a higher risk of developing LVH than boys during CKD, although the reasons for these sex differences are still unknown [32, 46].



**Figure 3.** The different phenotypes of cardiac hypertrophy in physiological and pathological conditions adapted from [29]. RA: Right atrium; LA: Left atrium; RV; Right ventricle; LV: Left ventricle.

### **Factors and mechanisms contributing to LVH**

LVH occurs in response to hemodynamic, cellular and molecular factors contributing to cardiomyocyte growth. Cardiomyocytes are terminal differentiated cells that cannot re-enter the cell cycle. They respond to external stresses by an increase in size. Eccentric hypertrophy is an increase in cardiomyocyte length, rather than the width, which occurs following assembly of serial contractile-protein units of myosin. In contrast, concentric hypertrophy occurs due to assembly of myosin in parallel such that cardiomyocytes increase in width, rather than in length. During CKD, previous studies have shown that hypertension, and more specifically increase in systolic blood pressure, was the major factor contributing to LVH in children. High blood pressure generates a mechanical stress that induces chronic changes in sarcomere length and collagenous components of the heart [37, 39].

Moreover, decline of kidney function during CKD leads to the retention and accumulation of compounds, which are normally eliminated by the kidneys through urine. A number of these, known as uremic toxins, have shown biological activity that exert negative impacts on body functions [47]. Uremic toxins are broadly classified into 3 groups: small, middle or protein-bound solutes. Examples of uremic toxins recently identified in children with CKD stages 1-5 compared to healthy controls were symmetric dimethyl-arginine (SDMA), fibroblast growth factor 23 (FGF-23),  $\beta$ 2-microglobuline, indoxyl sulphate (IxS), *p*-cresyl sulphate (pCS), indole-acetic acid, 3-carboxy-4-methyl-5-propyl-furanpropionic acid, and hippuric acid (HA) [48]. The association between SMDA, IxS, pCS, HA and LVH has already been

observed in humans and/or animal models but the mechanisms contributing to cardiac hypertrophy are still unknown [49-53]. On the other hand, different studies have explored the link between FGF-23 and LVH. Recent publications showed that circulating FGF-23 was an independent, with respect to eGFR, predictor of LVH in children with mild to moderate CKD [54-56]. In addition, *in-vitro* studies have shown that treatment of rat cardiomyocytes with FGF-23 induced hypertrophy [57]. Moreover, treatment with a pan-FGF receptor (FGFR) inhibitor, PD17307, in a rat model of CKD inhibited initiation of LVH [58]. Further studies on a FGFR4-knockout mice model showed that FGF-23 induces LVH *via* the activation of FGFR4 [59]. This suggests that FGFR4 could be a new pharmacological target to reduce cardiovascular complications associated to CKD. Nonetheless, LVH seems to be a very complex mechanism implying many more molecular actors. More research work is needed to determine the exact triggering mechanisms leading to LVH and, by the same way, to identify new therapeutic targets.

### **Treatment**

In the 'Effect of Strict blood pressure Control and ACE inhibition on Progression of chronic renal failure in pediatric (ESCAPE) patients' trial, the effect of strict blood control was evaluated on baseline LVH in 84 hypertensive CKD children (aged 3-18 years), after either 1-year or 2-years follow-up [60]. The normalization of blood pressure with ACE inhibitors (ACEi) lead to a decrease in LVH prevalence from 38% to 25% during the follow-up. Moreover, a significant improvement in myocardial systolic function was noted over time. However, reducing blood pressure to below normal values did not lead to further decrease in LVMI [60]. These results confirm that hypertension is a major actor in LVH and suggest that LVH could be targeted pharmacologically.

As low hemoglobin level is significantly associated to increased LVH in children [32], it has been hypothesized that increasing hemoglobin levels in these children would increase oxygen transport and cardiac function. However, little evidence is available on the benefits of anemia treatment in children with CKD with respect to LVH and cardiac function. Only one study evaluated short term effect of human recombinant erythropoietin treatment in 11 children on ESRD, aged 2.3-12.3 years, and showed a decrease in LVH after 36 weeks of treatment [61].

#### **b. Outcome and challenges of cardiac hypertrophy**

The main CVD outcome associated to cardiac hypertrophy are cardiomyopathies such as arrhythmia and cardiac arrest/sudden death [62-66]. Another study suggests that the abnormal cardiac function in CKD children with LVH may result in the development of systolic dysfunction and heart failure [40]. For now, the main problem for the management of LVH in CKD children is the lack of identification of high-risk patients. In 2011, Chavers *et al.* carried out an analysis on 656 United States pediatric patients on dialysis, aged 0.7-18 years, to investigate the frequency of the presence of cardiac disease in these

children. This study reported that while LVH was diagnosed in 24% of all the patients, echocardiography was performed in routine in only one third of the ESRD children [67]. The reasons for this lack of echocardiography analysis are unknown. Moreover, children who develop LVH associated to CKD remain asymptomatic for a long-time and are detected at a very late stage. Hence, it is imperative to find diagnostic tools which could allow early identification of LVH in children with CKD. Once hypertension is diagnosed in CKD children, annual testing for cardiac disease such as LVH should be performed [67]. This would allow prescription of the appropriate treatment to these children at the appropriate time, and thus extend their life expectancy.

## **ii. Vascular complications**

### **a. Vascular calcification and arterial stiffness**

#### **Definition and detection**

In physiology, arteries have a relatively elevated elasticity due to the high elastin-to-collagen ratio and the influence of smooth muscle cells. During vascular calcification, hydroxyapatite mineral (calcium and phosphate complex) is deposited in the arterial wall leading to increased stiffness. There are two types of calcification: arterial medial calcification (AMC, also known as Mönckeberg's sclerosis) and intimal calcification (AIC) [68, 69]. In AMC, calcification occurs in a sheet-like manner in the tunica media, not the intima, leading to a concentric thickening of the vascular wall and it is responsible for arterial stiffness. In contrast, AIC occurs in the lipid-rich atheromatous plaque and in a patchy manner, involving pro-inflammatory macrophages and vascular smooth muscle cells (VSMCs). AIC is usually associated with ageing and is rarely seen in the pediatric population. However, it is difficult to clearly distinguish AIC from AMC [69].

To assess rigidity of arteries, three different approaches can be used: 1) assessment of the arterial pressure waveforms, 2) measurement of the change in the diameter (or area) of an artery with respect to the expanding pressure and 3) measurement of the pulse wave velocity (PWV). PWV is considered as the gold standard for evaluation of arterial stiffness in adults [70]. In children, there are no gold standards, but most arterial stiffness evaluations have also been done by measuring the PWV and reference for comparison with the healthy population are available [71, 72]. The basis of PWV principle is that the pressure pulse, generated by the ventricular ejection, is spread along the arterial tree at a speed determined by the geometric and elastic properties of the arterial wall. The PWV is then expressed in meter per second. The arterial pulse wave is measured in proximal and distal arteries, most frequently the carotid and the femoral arteries. The different techniques used to measure PWV are the Doppler ultrasound, applanation tonometry, magnetic resonance imaging and oscillometric pressure cuffs [70]. PWV measurement can also be complemented with methods for evaluation of calcification such as the electron-beam computed tomography or spiral or helical computed tomography imaging. However, they

are less commonly used since the technique is expensive and not readily accessible [70]. One drawback observed with the PWV measure is that it can be difficult to carry out in very young children since their active participation is needed. They need to remain calm for at least 5 minutes in a supine position in a tempered room, and their head and shoulder part elevated at 30° to prevent venous artefacts [73].

### **Vascular calcification in children with CKD**

Vascular calcification occurs in chronic inflammatory conditions such as CKD and dyslipidemia. Calcification is the first vascular damage observed in children with CKD, but very limited data are available on their prevalence in the pediatric population. In 2008, Shroff *et al.* analyzed vascular calcification in 34 children with CKD [74]. The authors showed that accumulation of calcium in the vessel walls starts in early CKD stages without any histological calcification evidence. However, dialysis of children with ESRD accelerated the process to rapidly induce overt calcification (within 2 months of dialysis), as hydroxyapatite deposits were visible in 6 out of the 24 patients on dialysis (25%) together with VSMC apoptosis and osteogenic differentiation.

### **Factors and mechanisms contributing to vascular calcification**

Under physiological conditions, calcification promoters and inhibitors are tightly balanced such that calcium and phosphate, which are supersaturated in body fluids, do not precipitate. However, this balance is disturbed during CKD and induces pro-osteogenic signaling. This is mainly due to another complication associated to CKD which is the chronic kidney disease mineral bone disorder (CKD MBD). The first biochemical abnormality noted during the initial stage of CKD is an increase in circulating FGF-23 [75]. In addition to the pathway involving FGFR4 described above for FGF-23 in the development of LVH, FGF-23 also binds to its co-receptor, Klotho, activating the fibroblast growth factor receptor 1 (FGFR-1). FGFR1 activation reduces renal phosphate reabsorption in the tubules, inhibits renal 1- $\alpha$ -hydroxylase, increases PTH secretion by the parathyroid gland and increases 24-hydroxylase activity. This increase in 24-hydroxylase activity leads to a decrease in calcitriol (vitamin D). Moreover, a decrease in calcium absorption from the intestine occurs during CKD. The net decrease in calcium and phosphate levels, together with the increase in PTH in CKD patients leads to a compensatory mechanism causing calcium and phosphate release from the bones. This release causes renal osteodystrophy and bone fragilization in children. The other consequences usually observed in CKD children is a delay in growth, a short stature and eventually, vascular calcification [76, 77].

Calcium and phosphate are calcification promoters, whose effects are mediated by VSMCs. These cells are responsible of blood vessels contraction and relaxation, but they are also the key cellular effectors of calcification. Indeed, VSMCs are highly plastic cells which are able to adopt several phenotypes, such as osteoblasts or macrophage foam cells, depending upon the molecular signals in the extracellular microenvironment [78]. An increase in phosphate together with even a modest increase in calcium has shown to increase phosphate and calcium uptake in VSMCs *in vitro* and in animal models [79-81].

Calcium initiates the calcification process by inducing the development and release of vesicles laden with hydroxyapatite which will further lead to nucleation of crystalline hydroxyapatite. These calcium-phosphate nanocrystals are then degraded by lysosomes in VSMCs resulting in a peak intracellular calcium concentration and cell death (senescence and apoptosis – Figure 4). Once a nest of calcification forms in the extracellular matrix of VSMCs, its uptake and phagocytosis will induce additional calcification and simultaneously contribute to VSMCs differentiation to osteoblast like cells [69, 79]. To induce this differentiation of VSMCs during CKD, there is an upregulation of osteoblast genes and a synchronized decrease in smooth muscle genes which is mediated by a key transcription factor, Runx2 [82-84].

Finally, calcification may also be promoted in children with CKD due to a decrease in calcification inhibitors such as fetuin-A [85, 86] and osteoprotegerin [87].

### **Treatment**

To reduce progression of arterial stiffness, several treatment strategies are used. These include administration of phosphate binders, vitamin D derivatives supplement and calcimimetics.

To reduce cardiovascular risk, phosphate levels are reduced with the administration of phosphate binders in children with CKD such as calcium carbonate and sevelamer hydrochloride, licensed for use in children [77]. In addition, calcimimetics are given to children to maintain a calcium to phosphate balance during CKD. However, very few data are available on the efficacy of those compounds and their long-term effect in children with CKD [88].

To reduce the secondary hyperthyroidism that occurs following bone mineral disorders in CKD patients, vitamin D derivatives (Ergocalciferol and cholecalciferol) are administered [89]. The active form of vitamin D (1,25-dihydroxyvitamin D<sub>3</sub>) is a hormone involved in the control of mineral and bone homeostasis, and cardiovascular protection. A recent study carried out in 41 children with CKD and 24 healthy controls reports that a high-dose oral cholecalciferol treatment, a vitamin D derivative, reduced endothelial dysfunction and local arterial stiffness after 12 weeks of treatment [90].

### **b. Atherosclerosis and arterial wall thickening**

#### **Definition and detection**

Atherosclerosis is defined as the process of formation of an atherosclerotic plaque, also known as, atheroma [91]. The very first sign of atherosclerosis are fatty streaks developing in the intima. They are formed by the accumulation of T lymphocytes and lipid-rich macrophage (foam cells), which are derived from monocytes. These monocytes have been recruited from the blood circulation, in response to endothelial injury, or from de-differentiation of VSMCs (Figure 4). These foam cells regulate the cholesterol traffic, lipid homeostasis, inflammation and retention of atherogenic factors such as low-density lipoproteins (LDL) in the arterial intima. The persistence of a pro-atherogenic signaling will

then cause a progressive thickening of the intima and fatty streaks evolution to more complex occlusive lesions called fibrous plaques. These fibrous plaques will increase in size over several years and project into the lumen of the artery, hampering blood flow. At advanced stages of atherogenesis, as the lesions become thicker, angiogenesis of the fibrous plaques occurs. At the terminal stages, these capillary and venule-like vessels may modify the structure of the plaques causing cleavage, cracking or ulceration, and finally lead to bleeding from the vascular lumen, or from vessels which developed during the secondary angiogenesis. These finally lead to formation of a terminal thrombus [92].

The gold standard for the analysis of atherosclerotic plaque formation is the evaluation of the carotid intima-media thickness (cIMT). cIMT is measured using ultrasonography. For the measurement, the patient should stay in the supine position for 10 minutes in a controlled environment, to avoid vasoconstriction and/or vasodilation of blood vessels. During the imaging, a probe is placed perpendicular to the carotid artery for optimal imaging of the vessel wall and 2-dimensions images obtained. Using software compatible with the ultrasonogram, cIMT is then calculated [93].

This technique is non-invasive but has two drawbacks for pediatric evaluation of cIMT. The first one, again as in PWV measurements, requires the patients' collaboration and it is often challenging when performed in very young patients. Secondly, in most of the publications concerning the evaluation of cIMT, little attention is given to differences in the timing of measurements during the cardiac cycle and this may lead to variations during atherosclerosis diagnosis. To standardize cIMT measurement, it is suggested to perform the evaluation of thickness at the end diastole, when the artery appears the thickest [70].

### **Atherosclerosis in children with CKD**

In children with CKD, atherosclerosis is the most common vascular damage [91]. Data available on the prevalence of atherosclerosis showed that it affects 10% to 15% of the ESRD population [94]. Analysis of atherosclerotic plaques were done in adult CKD patients only and showed that they were more complex and less stable than in patients with normal kidney function: they have higher lipid content and lower fibrous content [95, 96]. This is certainly due to complications, such as dyslipidemia, inflammation and uremic toxins, driven simultaneously by CKD. Furthermore, calcification is recognized as being an integral part of 80 – 90% of atheromatous plaques [97]. Histomorphological analysis of arteries composition of children with CKD revealed that there were calcium-phosphate deposits, similar to bone, in atheromatous plaques and the media [98]. As discussed above, this vascular calcification leads to an increase in the rigidity of arteries [99-101] and further contribute to morbidity and mortality.

### **Factors and mechanisms contributing to atherosclerosis**

The main traditional atherogenic risk factor known for years now, is dyslipidemia [102, 103] and the pediatric CKD population is a vulnerable population. Analysis of the lipemic profile of 460 children

suffering from stage 2-3 CKD showed that 44% had dyslipidemia [104]. A small histological Turkish study of the internal iliac artery of 12 children with ESRD (aged 11-17 years) obtained during kidney transplantation showed that 7 (58%) of them had atherosclerotic lesions [105].

While dyslipidemia seems to be the main contributor for this clinical observation in children, molecular data showed that atherosclerosis is accelerated due to chronic inflammation. This supports clinical findings which showed ESRD patients had high levels of serum C-reactive protein [106, 107].

Moreover, several studies have shown that atherosclerosis was highest at ESRD, compared to earlier stages of CKD, in children [108]. This implies that uremic toxins could be associated to atherosclerosis in children [109, 110]. The uremic toxin, IxS has shown to induce human and rat smooth muscle cells proliferation *in vitro* [111-113] and IxS administration to rats showed an induction of aortic calcification accompanied by aortic wall thickening [114]. Moreover, IxS and pCS have shown to inhibit endothelial repair *in vitro* and therefore contribute to atherosclerosis [115, 116].

Elevated levels of calcium, phosphate, parathyroid hormone (PTH) and fibroblast growth factor 23 (FGF-23), and the calcium-phosphate product in CKD patients were also found to be associated to vascular calcification in atherosclerosis in children with CKD [117-119].

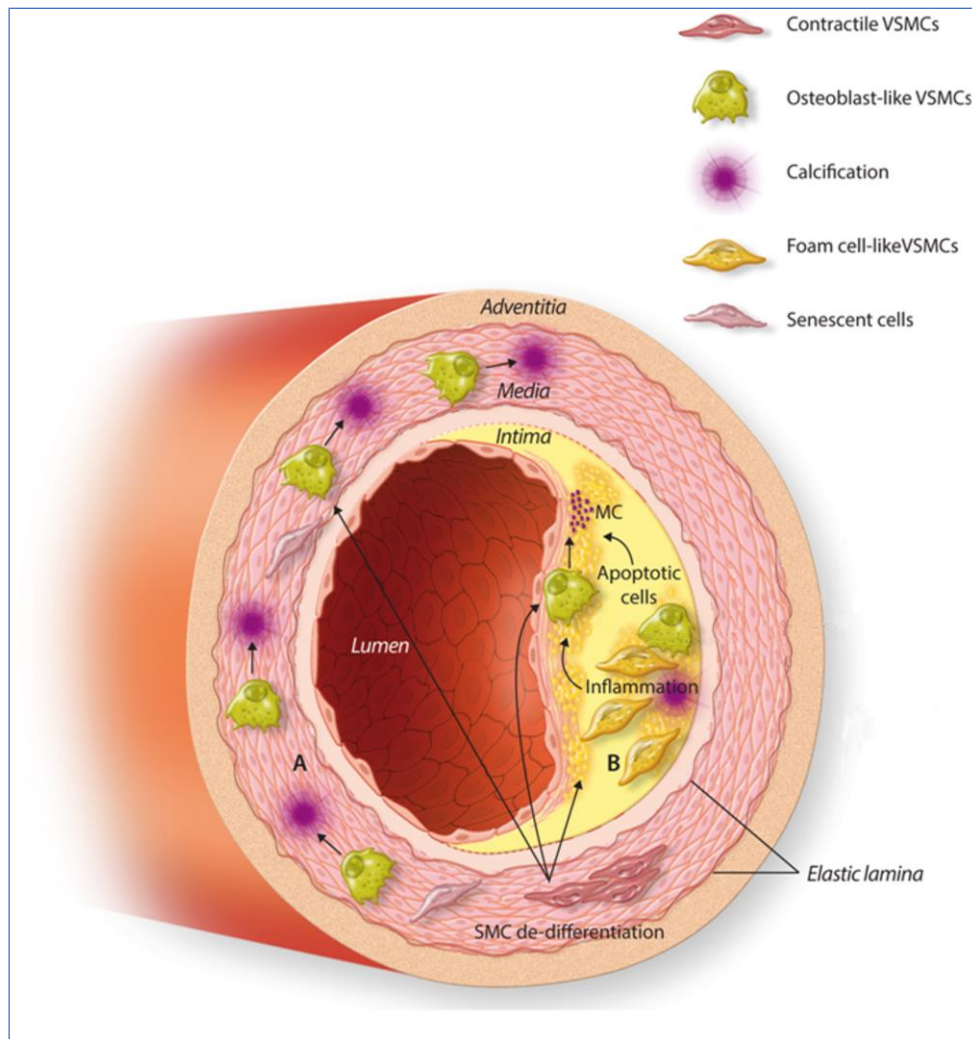
### **Treatment**

Atherosclerosis is already diagnosed during moderate CKD and acceleration in atherosclerosis progression is usually observed when patients are enrolled on dialysis. This is not observed in transplanted patients [91]. Anti-atherosclerotic strategies target dyslipidemia, inflammation and arterial calcification.

According to the current guidelines, dietary modifications and lifestyle changes are the first line treatment for dyslipidemia in children [120]. In the literature, statins are described as an effective treatment to reduce total cholesterol and LDL cholesterol in adults, and consequently reduce the risk of an increase in cIMT and the development of atherosclerosis. Few data are available concerning the pediatric population. Key factors must be considered by the clinician before prescribing statins: Any child/adolescent who begins statin therapy will be prescribed a higher cumulative dose over the course of his/her lifetime compared to most adults. Safety of such long-term statin exposure beginning in adolescence has not yet been validated in the clinic. The 2013 Kidney Disease: Improving Global Outcomes (KDIGO) guidelines do not recommend any statins treatment for children with CKD since no dose escalation has been done to confirm the safety of the lowest starting dose possible and that of higher doses over the short term. Moreover, no data is available concerning the appropriate LDL cholesterol to target in these children. More investigations are required to provide strong evidence that use of statins in CKD children is safe and effective [121].



Treatment of children with CKD with angiotensin-converting enzyme inhibitors (ACEi) seems to contribute to the reduction of atherosclerosis by inhibiting inflammation. Several studies using mice models have shown that ACEi reduced atherosclerosis [122, 123]. Angiotensin II contributes to atherosclerosis by promoting VSMCs migration and proliferation [124], and the release of proinflammatory molecules such as interleukin-6 [125]. ACEi are usually administered during the initial stages of CKD to children in most of the cases. This may contribute to reduction of atherosclerosis early in CKD and explain why atherosclerosis is detected in the late stages of CKD.



**Figure 4.** Schematic representation of an atherosclerotic plaque with medial (A) and intimal (B) calcification, during CKD. Adapted from [78].

#### **a. Outcome and challenges of vascular complications**

Atherosclerosis is less prevalent in children with CKD compared to adults. However, a number of prospective analyses report that adolescents and young adults with increased cIMT are more prone to myocardial infarction [126] and coronary artery disease [102]. Furthermore, according to a study which analyzed the survival and death of 146 children enrolled on peritoneal dialysis, 4 out of the 16 deaths

reported were due hemorrhagic stroke. Autopsy analysis revealed that these children with ESRD had atherosclerosis and vascular calcification [127]. The tools currently available for the diagnosis of vascular calcification and atherosclerosis (cIMT, PWV, computed tomography scan) are not sensitive enough to detect the very early changes of calcification. Moreover, normal/negative tests should be cautiously interpreted [74]. The real challenge is to identify new biomarkers which will allow a sensitive and clear identification of CKD patients with these above-mentioned vascular damages since these patients present high cardiovascular morbidity and mortality risks.

## II. Chronic Kidney Disease in dogs

### 1. Epidemiology

CKD is the most common kidney disease in dogs [128]. Like in humans, CKD is defined as an irreversible kidney damage due to the presence of structural or functional abnormalities of the kidneys for 3 months or more.

However, the incidence and prevalence of CKD is poorly documented in dogs. In a recent study, Pelander *et al.* (2015) report that incidence of kidney diseases varies much with the dogs' breed (table 2) between 16 and 39 cases for 10,000 dog-years at risk [129] and so does CKD. Another study supports this observation and reports that prevalence depends also on the study population and may vary from 0.05 to 3.74% [128].

	Breed	DYAR	Cases	Incidence (95% CI)	Incidence females (95% CI) (cases/10,000 DYAR)	Incidence males (95% CI) (cases/10,000 DYAR)	Mean (median) age at diagnosis
1	Bernese mountain dog	18,641	95	51 (41 to 61)	60 (44 to 75)	41 (28 to 55)	5.0 (4.7)
2	Miniature schnauzer	28,000	108	39 (31 to 46)	40 (30 to 51)	37 (27 to 47)	6.8 (7.2)
3	Boxer	19,190	69	36 (27 to 44)	45 (32 to 59)	26 (16 to 37)	5.3 (5.9)
4	Yorkshire terrier	21,313	73	34 (26 to 42)	33 (22 to 43)	36 (25 to 47)	7.6 (8.1)
5	Dalmatian	17,374	59	34 (25 to 43)	27 (17 to 38)	42 (28 to 57)	5.9 (6.3)
6	Shetland sheepdog	38,330	119	31 (25 to 37)	27 (20 to 35)	35 (27 to 43)	8.0 (8.3)
7	Fox terrier	11,024	34	31 (20 to 41)	28 (14 to 41)	34 (19 to 49)	8.2 (8.5)
8	Shih tzu	19,103	58	30 (23 to 38)	25 (15 to 35)	35 (24 to 47)	7.0 (7.3)
9	Soft-coated wheaten terrier	24,330	73	30 (23 to 37)	30 (21 to 40)	30 (20 to 40)	6.2 (6.4)
10	Cairn terrier	35,931	106	30 (24 to 35)	30 (22 to 38)	29 (21 to 37)	7.6 (8.6)
11	Giant schnauzer	10,655	31	29 (19 to 39)	29 (16 to 44)	29 (14 to 44)	5.0 (5.2)
12	Bearded collie	24,010	69	29 (22 to 36)	25 (16 to 33)	33 (23 to 43)	8.3 (9.6)
13	Standard poodle	24,950	70	28 (21 to 35)	25 (16 to 33)	32 (22 to 42)	7.3 (7.5)
14	Papillon	28,226	78	28 (22 to 34)	26 (17 to 34)	30 (21 to 39)	8.2 (8.5)
15	Flat-coated retriever	33,897	92	27 (22 to 33)	41 (31 to 50)	13 (7 to 18)	6.4 (7.6)
16	Rottweiler	40,409	106	26 (21 to 31)	26 (19 to 33)	27 (19 to 34)	6.7 (7.3)
17	Bichon frisé	26,674	66	25 (19 to 31)	28 (19 to 37)	21 (14 to 29)	7.5 (7.9)
18	Cavalier King Charles spaniel	55,839	134	24 (20 to 28)	25 (19 to 31)	23 (17 to 28)	6.5 (6.8)
19	Collie	41,079	98	24 (19 to 29)	33 (25 to 40)	15 (10 to 20)	6.1 (5.9)
20	English springer spaniel	60,031	138	23 (19 to 27)	21 (16 to 26)	25 (19 to 31)	7.6 (8.5)
21	American cocker spaniel	13,687	31	23 (15 to 31)	19 (9 to 29)	27 (15 to 39)	6.6 (6.4)
22	Whippet	12,899	28	22 (14 to 30)	18 (8 to 28)	26 (13 to 38)	8.6 (9.5)
23	English cocker spaniel	42,548	90	21 (17 to 26)	22 (16 to 28)	21 (14 to 27)	7.7 (8.5)
24	Greyhound	8,535	18	21 (11 to 31)	16 (4 to 27)	27 (11 to 43)	6.3 (7.8)
25	Rhodesian ridgeback	10,619	21	20 (11 to 28)	28 (14 to 43)	11 (2 to 20)	5.3 (4.4)
26	West highland white terrier	24,082	47	20 (14 to 25)	19 (11 to 27)	20 (12 to 28)	8.4 (9.2)
27	Newfoundland	11,754	22	19 (11 to 27)	16 (6 to 26)	22 (10 to 35)	7.1 (7.2)
28	Doberman pinscher	12,536	23	18 (11 to 26)	27 (15 to 40)	8 (1 to 16)	6.2 (7.1)
29	Samoyed	15,479	28	18 (11 to 25)	20 (10 to 30)	16 (7 to 25)	7.3 (6.7)
30	Tibetan spaniel	15,115	25	17 (10 to 23)	12 (4 to 20)	21 (11 to 31)	5.2 (4.9)
31	Golden retriever	156,459	257	16 (14 to 18)	19 (16 to 23)	13 (11 to 16)	7.0 (7.3)

Breeds with an incidence of kidney disease exceeding the mean total incidence for all breeds (15.8 cases/10,000 DYAR). Only breeds with >10,000 DYAR in the database are included in the breed analysis  
DYAR, dog-years at risk

**Table 2.** Incidence of kidney diseases in dogs with respect to their breed. Study carried out on 600,000 insured dogs in Sweden. There was no difference in the incidence of kidney disease between males and females [129].

In dogs, CKD is a typically progressive disease. Therefore, when the dogs are diagnosed with CKD, the disease is expected to be a lifetime condition, even with treatment.

## 2. Etiology of CKD in dogs

Several factors affect the occurrence or progression of CKD, and most of them are analogous to humans.

### i. Ageing

The prevalence of CKD increases considerably with age in dogs. Studies on the remodeling that occurs in the ageing dog kidney identified three major mechanisms which contribute to lifetime wear and tear: autophagy, senescence [130] and telomere shortening [131].

Autophagy is a recycling process during which cellular components are degraded. During this degradation, cytoplasmic components, such as organelles and proteins, are enclosed in a membrane bound structure called an autophagosome which then fuses with lysosomes. The degraded products can then be reused by the cell [132]. In humans and mice, it has been shown that anomalies in autophagy, that occur with age, cause acute tubular epithelial injury, glomerulosclerosis and tubulointerstitial scarring [133] lead to CKD. Senescence, that occurs following arrest in replication, initiates inflammation *via* the production of cytokines, or cell death. This mechanism is often accompanied by telomere shortening [134]. This is in accordance with a study which reports that telomere shortening in dogs is associated to a less effective repair of oxidative damage with age and therefore contributes to CKD [131].

### ii. Periodontal diseases

A retrospective study carried out on mix-breed dogs has shown an association between periodontal diseases and histopathological alterations in renal glomeruli and interstitium leading to CKD [135, 136]. This observation was confirmed on a larger cohort of 164,706 dogs where increasing severity of periodontal disease was correlated to the decrease of glomerular filtration rate [137]. In humans also an association has been found between periodontal disease and CKD [138, 139], however, the mechanisms associating those two are still unknown in both species.

### iii. Acute kidney injury (AKI)

AKI may be defined as a rapidly occurring kidney disease which can be reversed by kidney repair. Many recent studies have shown that AKI and CKD interconnect. Indeed, as in humans, the presence of AKI is a risk factor for initiation and advancement of CKD, and CKD is a risk factor for the active injury leading to AKI in dogs [140]. AKI displays a series of maladaptive cellular repair events which predisposes to CKD. In mice it has been shown that the driver of this maladaptive response is the G2/M arrest in tubular cells. Indeed, in response to an injury, the tubular epithelial cells blocked in G2/M overexpress profibrotic factors. This is accompanied by an induction of senescence in epithelial cells and the release of cytokines such as interleukin 6 and interleukin 8, leading to a persistent parenchymal inflammation [141]. A major feature that results from this maladaptive repair is an increase in the proliferation of myofibroblasts that will further deposit collagen and other extracellular matrix

components leading to fibrosis, a hallmark of CKD [142]. This is in accordance with a histological analysis of the kidneys of dogs suffering from AKI after ingestion of toxin-containing pet foods (melamine and cyanuric acid). Four out of the six dogs with AKI had renal interstitial fibrosis in addition to renal tubular necrosis, intratubular crystals, and interstitial inflammation which are markers of CKD [143].

#### **iv. Hypertension**

In small animals like dogs, hypertension is defined as chronic high blood pressure greater than 150/95mmHg. The cause of hypertension is multifactorial, and the kidneys are one of the first organs to be directly impacted by this high blood pressure. Like in humans, glomerular hypertension and hyperperfusion result in an increase in the surface area and volume of the glomerular tuft. Podocytes maintain the integrity of the glomerular filtration barrier through complex series of interdigitating cell foot processes between adjacent podocytes, connected by structures known as ‘slit diaphragms’. These slit diaphragms limit the trafficking of blood components through the capillary wall. Podocyte number or size does not increase harmoniously as the glomeruli enlarge. A low glomerular podocyte density results in podocyte effacement, focal denudation of the capillary wall and increased capillary permeability, leading to leakage of proteins, such as albumin, in the urine resulting in proteinuria, as well as glomerular sclerosis and atrophy [144]. Moreover, the kidney function is further affected since hypertension also induces an increase in renal loss of sodium and water due to excessive pressure damages to tubules and causes interstitial fibrosis. All these combined contribute to CKD [145, 146].

#### **v. Genetics**

A first example of genetic associated-CKD is X-linked hereditary nephropathy (XLHN), observed in some canine breeds such as English Cocker Spaniels, Navasotas, Samoyed, Dalmatian and Bull Terriers. XLHN is characterized by a defect in type IV collagen in the glomerular basement membrane (GBM). This glomerular disease, which corresponds to the Alport syndrome found in human, is caused by mutations in *COL4A5* gene and leads to a failure in the synthesis of  $\alpha 5$ -collagen chains. Like humans, more males are affected than females. The juvenile-onset of CKD in dogs suffering from XLHN begins with a persistent proteinuria of glomerular origin, as early as 3–6 months old, followed by a reduced glomerular filtration rate GFR and persistent azotemia. These dogs typically progress to ESRD before 1 year of age [147-149].

Another example is polycystic kidney disease (PKD). PKD has been described in several canine breeds: Shih Tzu, border terrier, golden retriever, standard poodle, boxer, Finnish harrier, Rhodesian ridgeback and Dutch kookier. In humans, PKD results from mutations in *PKD1* or *PKD2* genes encoding for polycystin-1 and 2 respectively. In dogs, mechanisms of PKD are unknown and it cannot be ruled out that mutations in PKD genes may be a cause, even if causative mutations have not yet been identified

[150]. Only one study reports that a missense mutation in PKD1 gene of bull terrier is associated to autosomal dominant PKD [151].

#### **vi. Congenital anomalies**

Apart from familial hereditary issues, congenital disorders may also be responsible of CKD in dogs. As in human, these include renal dysplasia, hypoplasia and unilateral renal agenesis [152, 153]. Renal dysplasia and hypoplasia in dogs seem to occur due allelic variation following methylation of DNA in the promoter region of canine cyclooxygenase-2 gene [154, 155]. However, no mechanism has yet been identified for renal agenesis in dogs. To compensate for the missing nephrons, the remaining intact nephrons hypertrophy to increase glomerular filtration. This hypertrophy will then lead to podocyte effacement, focal denudation of the capillary wall and an increase in the permeability of capillaries for blood components, as discussed in the section on hypertension. Moreover, intact remnant nephrons will accumulate injuries consequently leading to inflammation, fibrosis and finally to CKD [156].

### **3. Complications associated to CKD in dogs and their treatments**

#### **i. Nutritional disorders**

The first complication associated to CKD in dogs concerns nutritional disorders: reduced appetite, dehydration, anorexia, nausea, vomiting and body weight decrease due to muscle mass.

Once this is noticed, veterinary nutritionists provide owners with a ‘kidney diet’ suitable for dogs suffering from CKD. This diet consists in reduced protein, phosphorus, and sodium content, increased caloric density, a neutral effect on acid-base balance and the addition of antioxidants. The effect of this new diet is then evaluated regularly by monitoring body weight, general body condition (including coat color and hydration), food intake, serum albumin concentration and quality of life [157].

#### **ii. Chronic kidney disease mineral bone disorder**

As in humans, CKD in dogs is associated to CKD MBD. Mechanisms are like those described in humans since they imply increased FGF-23 levels, reduced renal phosphate reabsorption in the tubules, secondary increased PTH levels as well as decreased plasma calcium levels due to reduced intestinal absorption. This has deleterious effects on bones leading to their release of calcium and phosphate, thereby contributing to bone weakening and increased bone porosity [158]. However, the effect of CKD MBD on soft tissue calcification such as blood vessels has not been reported or thoroughly studied in dogs. Phosphate binders are usually given to dogs to manage CKD MBD [158, 159].

#### **iii. Metabolic acidosis**

Like in humans, metabolic acidosis is also a problem in dogs with CKD [160, 161]. Kidneys ensure a stable serum bicarbonate concentration by reabsorbing bicarbonate ions from the glomerular ultrafiltrate

and by synthesizing bicarbonate ions. During CKD, kidney function is compromised, and bicarbonate ion reabsorption and synthesis are decreased. This leads to a positive H<sup>+</sup> balance in the plasma. As a compensatory mechanism, to bring back a neutral acid-base balance, bones are used as the buffering component mainly by releasing phosphate ions. Consequently, this further exacerbates CKD MBD and increase bone fragilization. Moreover, this metabolic acidosis causes muscle wasting, impaired thyroid metabolism, increased inflammation and contribute to CKD progression [162].

To manage this acidosis, alkalinizing salts such as potassium citrate or sodium bicarbonate are prescribed for dogs to reduce disease progression. To evaluate the efficiency of the alkalinizing treatment, blood gas analysis is performed 10 to 14 days after starting the therapy and any adjustment of the dosage is done until blood gas level is normalized [157].

#### **iv. Cardiac remodeling**

There are increasing evidences of a cardiorenal disorder associated to morbidity and mortality in CKD dogs, like in humans. Cardiac modifications that occur in dogs with CKD are mainly valvular heart disease as opposed to humans which is LVH. Renal damages lead to systemic volume overload which contributes to congestion, valve disease, dilated cardiomyopathy, and diastolic dysfunction (hypertrophic cardiomyopathy and hypertensive heart disease) in dogs [163].

A retrospective study carried out on 124 dogs showed that 50% of the dogs with chronic valvular heart disease also had CKD, and that 70% of CKD dogs were at the most severe stages of the chronic valvular heart disease [164]. However, more research work needs to be done to investigate the direct or indirect mechanisms linking this heart disease and kidney injury, and with respect to reduced GFR [163].

Even if LVH is not the main cardiorenal disorder associated to morbidity and mortality in dogs with CKD, one small scale study investigated the relationship between CKD and LVH in dogs. In this study carried out on 16 dogs, CKD was generated by subtotal nephrectomy and the authors followed left ventricular hypertrophy development. After 6 months of follow-up, CKD dogs displayed increased ventricular fibrillation, higher sympathetic activation, more inflammation and left ventricular hypertrophy compared to controls [165]. These observations have, however, to be validated on a larger cohort of dogs and the type of LVH (eccentric or concentric) needs to be determined. Concerning the possible molecular actors involved, azotemia/uremic toxins adverse effects on cardiomyocytes may be implicated in cardiorenal disorder, but these toxins and their effects need to be investigated in dogs.

Management of cardiac remodeling is challenging in dogs since it relies upon fluid therapy and close monitoring of the amount and type of proteins and phosphorus intake. To manage heart failure and hypertension, ACEi are usually the first drugs to be prescribed to dogs.

## **v. Anemia**

A decline in kidney function, mainly in stage 3 or 4 CKD in dogs, is often accompanied by anemia. Erythropoietin (EPO) is a hormone essential for the synthesis of red blood cells which is produced by the kidneys. As functional mass of the kidney decreases, EPO is less secreted, causing anemia. To manage anemia in dogs, EPO is given to dogs, commercially available as darbopoietin and Aranesp® [166].

## **4. Diagnosis of CKD in dogs**

CKD in dogs is most of the time asymptomatic until the disease has quite advanced. The occurrence of CKD is usually strongly suspected when owners spot signs of age-related diseases such as decreased appetite and weight loss, deteriorating coat, vomiting, bad breath, lethargy, excessive production of urine (polyuria) and excessive thirst (polydipsia), as well as any urination accidents [167].

The veterinary physician begins the examination by gathering all the details concerning the dog: age, gender, breed predisposition, any pertinent history concerning medications use, any exposition to toxins, healing of acute kidney disease, diet and travel, and any exposure to infectious diseases, such as Lyme disease or Leptospirosis, that may favor occurrence of CKD.

The next step is a physical examination of the dog, which consists in palpating any kidney abnormality, weighing the animal, evaluating dehydration, pallor, oral ulcers and hypertensive retinopathy, and more advanced evaluation using more specific diagnostic tools [167]. The main tools available for the assessment of canine CKD diagnosis are discussed below.

### **i. Actual clinical biomarkers of CKD in dogs**

#### **a. Glomerular filtration rate (GFR)**

As in humans, in dogs GFR measurement is the principal test to evaluate kidney function. GFR is determined using specific markers which are eliminated in urine by glomerular filtration without being reabsorbed or secreted by the tubules. Four GFR markers are currently used in dogs: inulin, iohexol, creatinine and cystatin C. The two first markers are exogenous, safe and inert compounds while creatinine is a breakdown product of muscular creatine phosphate and cystatin C is a protease inhibitor produced by all nucleated cells.

Measurement of GFR is achieved by evaluating the plasma clearance of these markers, i.e the volume of plasma theoretically 'free' of the marker per unit time. The most accurate method to measure a marker's renal clearance needs to evaluate the concentration of marker in plasma as well as the quantity of marker excreted in the urine over a 24-hour period. However, collection of urine produced over 24 hours is impractical in the clinic with dogs. The alternative method consists to inject a given amount of exogenous (inulin or iohexol) marker into the dog's blood and then quantify the decrease over time of



the marker plasma concentration. In this case, the marker clearance is calculated as the ratio between the quantity of the administrated marker and the area under the plasma concentration-versus-time curve.

As in human, GFR can also be estimated by measuring endogenous (creatinine and cystatin C) marker concentration in plasma. Indeed, since both creatinine and cystatin C are produced at a constant rate, their accumulation in blood reflects a decrease in GFR. Although less accurate, this approach is less tedious than the previously described method.

The International Renal Interest Society (IRIS, <http://www.iris-kidney.com/guidelines/>) – has classified the progression of CKD in dogs into 4 stages, with stage 4 being the most severe. The staging system is mostly based on creatinine (Table 3) completed with additional observations (i.e. presence of azotemia, extrarenal clinical signs (e.g blood pressure and urine specific gravity). In healthy dogs, blood creatinine is < 1.0 mg/dL [168].

Stage	Blood creatinine	Comments
<b>At risk</b>	<125 µmol/L	Suggests that the animal is at increased risk of developing CKD in the future
	<1.4 mg/dL	
<b>1</b>	<125 µmol/L	Non-azotemia dog but presence of some other renal abnormalities
	<1.4 mg/dL	
<b>2</b>	125 – 180 µmol/L	Mild renal azotemia with almost no clinical signs
	1.4 – 2.0 mg/dL	
<b>3</b>	181 – 440 µmol/L	Moderate renal azotemia with several extrarenal clinical signs.
	2.1 – 5.0 mg/dL	
<b>4</b>	>440 µmol/L	Increasing risk of systemic clinical signs and uremic crises
	>5.0 mg/dL	

**Table 3.** Definition of the 4 distinctive CKD stages in dogs according to the IRIS staging system.

Of note, as in humans GFR is influenced by body weight in very low body weight animals, age, gender, breed and circadian rhythm [169].

#### **b. Urine specific gravity (USG)**

Urine specific gravity (USG) is an estimate of urine osmolality which is used to evaluate the ability of the kidney tubules to concentrate the glomerular filtrate in dogs. USG is a ratio of the density of 1L of urine to the weight 1L of water and it is measured on a drop of urine using a refractometer [170, 171]. The higher the value, the more concentrated is the urine [171]. A USG value >1.030 is obtained when the dog kidneys are able to substantially concentrate urine by water tubular reabsorption and show that the kidneys are healthy [172]. In contrast, a USG between 1.008 and 1.012 indicates abnormalities in water reabsorption and is interpreted as renal dysfunction. In case of intermediate values of USG (1.013 to 1.029), the decrease in urine concentrating ability is not enough to infer about renal tubular dysfunction [173].

However, several factors such as age, diet, sex, fasting status, thirst and the type of refractometer used affect USG [174].

### **c. Proteinuria**

In healthy kidneys, circulating proteins cannot pass in urine due to their higher size and negative charge. However, in most renal diseases, the glomerular barrier is damaged, leading to increased permeability, and plasma proteins flow into urine. Proteinuria is also considered as a prognostic marker of CKD in dogs, like in humans. Available evidence showed that proteinuria increases in nephrectomized dogs [175] and that treatment of dogs with ACEi simultaneously reduces proteinuria and increases lifespan [176, 177].

Proteinuria is evaluated by measuring the urine albumin concentration (microalbuminuria) and/or the urine protein-creatinine ratio (UPC). Albumin is the most abundant protein present in proteinuria [178], and as data on urinary albumin concentration and urinary total protein concentration are similar, both are used as markers of CKD [179]. In dogs, as in humans, the first common screening test is the dipstick test, a colorimetric test sensitive to detect albumin and other proteins in urine. For a quantitative evaluation, commercial kits such as enzyme-linked immunosorbent assays for albumin, or colorimetric tests for protein are used in dogs [179].

Ideally, albumin or proteins must be measured over a 24-hour period but, as previously indicated this is impractical in dogs. The normal albumin concentration in dogs is  $< 0.01\text{g/L}$ . Microalbuminuria is defined as the presence of albumin in urine varies between  $\geq 0.01$  and  $< 0.03\text{g/L}$ , and proteinuria is diagnosed when urinary albumin  $> 0.03\text{g/L}$ . The gold standard test is the determination of UPC performed on a single urine sample, this ratio being closely correlated to the 24-hour urine protein excretion. UPC is  $< 0.2$  in healthy dogs. An UPC between 0.2 and 0.5 is considered as borderline proteinuria, and UPC  $> 0.5$  indicated proteinuria. Of note, the dipstick colorimetric test lower limit of protein detection in urine is about  $0.30\text{g/L}$ . [179, 180].

The main factors that affect proteinuria are hypertension, endogenous or exogenous corticosteroids, dietary protein content, exercise and hyperthermia [179, 181].

### **d. Azotemia**

Azotemia is defined as the presence of high concentration of blood urea nitrogen (BUN) compounds associated to inadequately concentrated urine. These BUN compounds are urea, creatinine and other non-protein nitrogenous compounds, usually cleared by glomerular filtration but retained in the blood in case of reduced GFR.

Azotemia is evaluated by measuring serum creatinine level and urea, and USG. Dogs are considered to have azotemia when the serum creatinine  $> 1.6\text{mg/dL}$  or BUN  $> 30\text{mg/dL}$  or UPC  $\geq 3.5$  over weeks to months together with a USG between 1.008 and 1.029 [182, 183].

The factors that are known to affect azotemia are the amount of protein present in the diet and (de)hydration status [164].

#### **e. Kidney hyperechogenicity on ultrasound imaging**

Ultrasound evaluation helps to distinguish between a normal kidney and a kidney with CKD, giving information about the position, size, shape and internal structure, and hemodynamics of the kidneys.

When increased blood creatinine and urea levels are detected, abdominal ultrasound is performed if the veterinary has access to B-mode ultrasonography techniques (Doppler sonography). Renal cortical echogenicity is compared to the echogenicity of adjacent organs such as the liver and the spleen. Corticomedullary differentiation can also be assessed since the cortex is more hyperechoic, due to its higher cellularity, than the medulla, which has a higher liquid content.

Increased hyperechogenicity has been noticed in 88% of dogs with CKD, decreased corticomedullary differentiation in 54% and absence of differentiation in 35% of dogs with CKD. In 45% of dogs with CKD, low renal volume was observed. In others, as in dogs with PKD, an increase in renal volume was detected, associated with irregular, oval or circular contours, anechoic content (typical of renal cysts), and thin, smooth, and hyperechoic walls sharply defined by distal acoustic shadowing.

An abnormal kidney on imaging requires a renal biopsy to evaluate glomerular lesions, such as glomerulosclerosis, or tubular lesions, such as tubular atrophy, interstitial inflammation and fibrosis indicative of CKD, unless ultrasonography unveils kidney changes described above indicative of ESRD [184, 185].

### **ii. Drawbacks of actual biomarkers for CKD diagnosis in dogs**

The incidence of CKD in dogs is under-estimated. This is due to the lack of standardized protocols and performant diagnostic tools to detect CKD during the initial stages [129]. The drawbacks of the currently used biomarkers are discussed below.

#### **a. Creatinine**

Plasma creatinine is used to estimate GFR and a decreased GFR is considered as a main surrogate marker of CKD. However, the major limitation pointed out in many canine studies is that creatinine is a relevant biomarker for late stage CKD, when up to 75% of the kidney function is lost [186, 187]. Indeed, the kidneys display great compensatory mechanisms which take place during nephron loss, such that surviving nephrons hypertrophy, to ensure normal filtration rate [188]. In addition, since serum creatinine has an inverse but nonlinear relationship with GFR, major modifications in GFR occurring during in initial CKD stages induce minimal detectable changes in serum creatinine while much damage has been done to the kidneys. Second, creatinine provides little information on the disease progression [164]. Third, serum creatinine levels in dogs are affected by gender, breed variation, muscle mass protein intake and circadian rhythm, similarly to humans. Different formulae can be used to correct GFR by

these parameters, but no consensual calculation was found [167]. Additionally, depending on the size of the dogs, blood creatinine values may remain in the normal range and CKD may be detected only at azotemia stages which correspond to stage 3-4. At these stages, the dogs have only a few months to 2 years survival before dying or being euthanized because of the advanced disease [128].

#### **b. Urine specific gravity**

Urine concentration measurement is not very reliable since the value obtained will depend on the refractometer used for measurement. Indeed, the comparison of five refractometers (four optical and one digital) have shown disparities in results and data failed to increase consistently with increasing urine specific gravity, contrary to reference methods (e.g density determination by pycnometer). USG evaluation using refractometer is therefore not a good marker of the kidneys urine concentration efficiency and cannot be used to infer about the presence of CKD [171].

Moreover, according to the IRIS working group, USG does not seem to be a reliable marker. Indeed, a wide range of USGs can be encountered in healthy dogs – 1.015 to 1.045 –, depending on the patient's hydration status. This points out a loophole that USG cannot be used to infer about kidney function [189].

#### **c. Proteinuria**

Proteinuria is only a late marker of CKD, since it is detected when there is significant and irreversible glomerular damage, and cannot predict CKD progression [180, 190]. Moreover, the dipstick colorimetric test can give false positive results in case of highly concentrated or pigmented urine since it has been designed for human urine which is rarely concentrated as dog urine. Finally, caution should be taken when interpreting albuminuria since urinary albumin increases with age in dogs [179].

#### **d. Kidney hyperechogenicity on ultrasound imaging**

As GFR and proteinuria, ultrasound imaging does not allow early diagnosis of CKD [191]. In addition, architectural alterations associated with diffused parenchymal kidney dysfunction are difficult to assess. Moreover, there is no reference standard for the renal size in adult dogs due to the wide variety of dog breeds. Finally, Doppler ultrasound is not routinely used by most veterinary practitioners since costs are high, the examination times are long, and ultrasound data interpretation is highly subjective [184, 185, 192].

### **iii. Challenge for diagnosis of early CKD in dogs**

It is important to point out again that biomarkers used in veterinary clinical practice have been originally found in humans and it is very well possible that a solution for the early diagnosis of CKD in dogs also comes from studies in humans. In human medicine, urinary biomarkers that can diagnose CKD with a high sensitivity and can even predict the progression of the disease have been identified. Moreover, to

diagnose complex diseases such as CKD, a combination of several markers showed to be most efficient [193, 194].

As you will see in the next chapter on '*peptidome and metabolome analyses for the identification of biomarkers for early diagnosis of CKD and CVD*', urinary peptidome analysis has been applied for the identification of novel early biomarkers in human disease. The clinical challenge now is to apply this technology to help in the diagnosis of diseases in dogs by starting with identification of urinary peptide biomarkers for the early diagnosis of canine CKD.

# III. Peptidome and metabolome analysis by mass spectrometry

## 1. Introduction

Complex clinical phenotypes, as exhibited by kidney disease, cannot in general, be described in detail on the level of single molecular features. In contrast, a combination of features - i.e. a panel or profile – can cover a disease complexity, while being less sensitive to inter-individual variations, and appears more suitable to describe or represent a complex clinical presentation. “Omics” studies allow quantitative monitoring of a plethora of such molecular features using high-throughput technologies. The major omics tracks are genetics (genomics), transcripts (transcriptomics), proteins (proteomics) and metabolites (metabolomics) (the “big 4”), but are further complemented by dozens of further “Omes” [195].

Genomics represents the global analyses of variations in genes (coding and noncoding regions) [196]. Genes and mutations therein describe the predisposition for a disease or its progression, but actual disease activity is generally not captured by genomic analysis. Transcriptome levels (i.e. messenger RNA (mRNA) and noncoding(nc) RNA) potentially better describe disease activity, but certainly mRNA expression only serves as an approximation of effective protein concentration and activity [197].

Proteomics consists in the analysis of the global protein content, which are the main cellular components and important functional molecules (catalytic enzymes and signal transduction proteins), in a sample [196]. Moreover, proteomics ties information contained in the genome and the transcriptome together, with post-translational modifications such as phosphorylation and glycosylation. Therefore, the proteome can reflect the dynamics of changes in protein concentrations or post-translational modifications that occur during diseases [198]. Peptidomics has emerged from proteomics and consists in the analysis of low-molecular-weight proteins (peptides) with a molecular weight between 0.5 – 15kDa, in a biological sample [199].

A younger branch of omics research, which is gaining more and more popularity, is metabolomics. This is the study of metabolites (<1200Da) such as carbohydrates, amino acids, dipeptides and organic acids. Metabolites are considered to describe most accurately the physiological state of a biological system since it represents the downstream expression of the genome, transcriptome, proteome together with the influence of other factors such as the environment or lifestyle [200].

In this thesis, we will focus mainly on the body fluid peptidome and metabolome analysis for the identification of biomarkers.

## **2. Biomarkers**

### **i. Definition**

The term *biomarker*, a blendword of '*biological marker*', refers to a characteristic that is measured as an indicator of healthy biological processes, pathogenic processes, or responses to treatment. When the link between any given measurable biomarker and relevant clinical endpoints is established, the biomarkers can be used for diagnosis (identification of patients with a disease or a subset of the disease), monitoring (serial measurements are used to detect a change in the severity/stage of the disease), prognosis (identification of disease progression or a clinical event) or response to a treatment. Biomarkers can be a molecular, histologic, radiographic, or physiologic characteristic [201].

### **ii. Identification of biomarkers**

High-throughput technologies have led to the identification of biomarkers in several biological fluids including saliva, blood, plasma, urine, or amniotic fluid. Such clinical omics have become possible by coupling omics-based approaches to computational, bioinformatics methods and well-phenotyped patient cohorts [202, 203]. However, several issues need to be considered before starting the identification of biomarkers using omics-based approaches.

#### **a. Clinical issues**

Having a clinical sample and being able to carry out omics experiments does not necessarily lead to useful biomarkers. A critical issue is to determine the clinical need and to define whether omics analysis can contribute to solve this clinical need. In addition, appropriate study design is essential for the study and starts with a precisely defined clinical phenotype, inclusion criteria and supply of sufficient demographic and/or clinical data. Appropriate positive and negative controls must be also defined. A healthy control population does not often reflect the clinical situation encountered and biomarkers identified in a defined study will potentially be invalid in follow-up studies (e.g. if we want to identify biomarkers of diabetic nephropathy, the controls should be diabetic patients with absence of nephropathy).

#### **b. Analytical issues**

Appropriate and highly standardized sample procurement is crucial for omics studies. Sufficient information about the sampling methodology including a description of specimen collection, handling and storage (i.e. type of containers, stabilizing solutions) should be available, in addition to the assurance of having all samples collected in the same manner. A bias can be introduced e.g. if urine is collected at home or in the clinic. The number of sample freeze-thaw cycle should also be considered if the omics

trait studied is sensitive to this. Finally, the same data normalization procedure should be applied to all samples to allow for comparison between different samples and avoid biases. Normalization procedures include, for e.g., correction for urine dilution between samples before comparing the samples to identify candidate urinary biomarkers based on relative abundance evaluation [204].

### **c. Statistical and validation issues**

To identify biomarkers, in general a two-step procedure is employed: the first step is called the discovery (or training) phase where biomarkers are identified following appropriate statistical analyses. This step is followed by a validation phase on an independent dataset or cohort, which allows appreciation of the performance of the biomarkers by assessment of their sensitivity and specificity. Good biomarkers have a high sensitivity, which means that it allows detection of the diseased individuals with slight or no overlap between healthy and diseased individuals, and a high specificity, meaning that absence of the disease in individuals identified by the biomarkers as healthy [205].

For quality assurance of clinical omics studies, appropriate and rigorous statistical analyses should be performed. In the discovery phase, normality tests such as Shapiro-Wilk test should be first performed before comparing the groups. Then accordingly, a Student's *t*-test or Wilcoxon-Mann-Whitney test is performed to identify potential biomarkers that are different between two conditions (*e.g.* healthy versus disease or two distinct stages of a disease). Correlation analysis can also be performed to identify potential biomarkers if binary cutoff values are absent. For the identification of robust biomarkers, the next important following step is the multiple testing correction since generally, the number of features is much higher than the number of patients. This includes Benjamini-Hochberg correction which calculates the false discovery rate, to reduce overall type I error rate and false positive results [206]. By the end of this analysis in the discovery phase, individual candidate biomarkers are identified. The next step is the development of multi- biomarker predictive models using a machine-learning algorithm such as support vector machines [207] or random forest algorithms [208].

These multi-marker models need then to be validated. We reach therefore the second phase which is the validation of the model in an independent cohort (*e.g.* a holdout set or a newly recruited set of patients). In this step, we evaluate whether the multi-marker model can be applied to the new group of individuals, and that the model was not highly specific (*i.e.* overfitted) of the discovery cohort [202].

## **3. Mass spectrometry**

Since the completion of genome sequencing, there has been a shift in the characterization of other biological molecules such as proteins and metabolites. With this aim omics studies using mass spectrometry (MS) have been widely carried out.

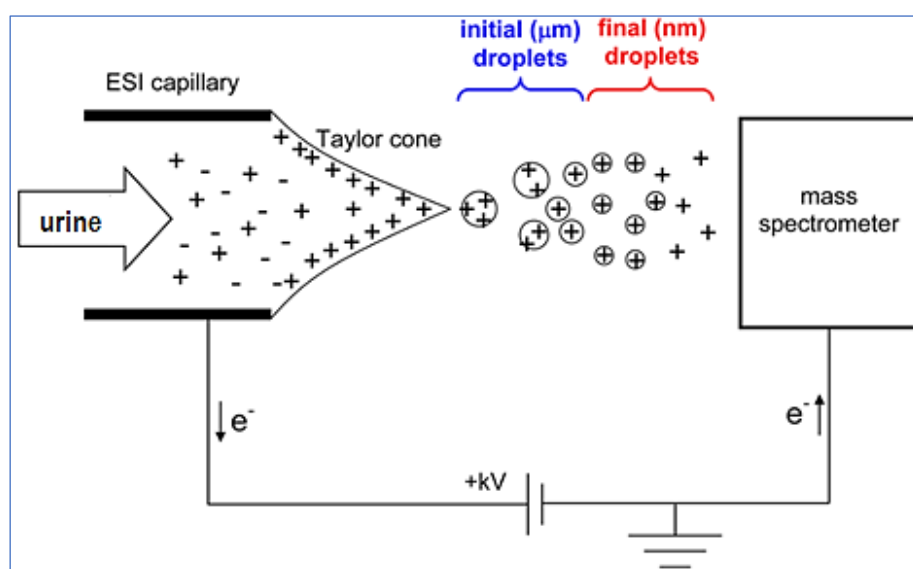


## i. Principle

Basically, MS consists in the analysis of ions in the gaseous phase to measure a mass-to-charge ( $m/z$ ) ratio. For this analysis, mass spectrometers are made-up of 3 main parts: i) an ion source, which converts analyte molecules into the gas phase (ionization), ii) a mass analyzer that separates ionized analytes with respect to their  $m/z$  ratio, and iii) a detector that records the number of ions at each  $m/z$  value [209].

*i) ionization:* The two ionization techniques used in biomedical mass spectrometry are electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI). In this manuscript, we will pay more attention to ESI, since MALDI is a high throughput technology that is mainly used on solid analytes, such as biopsies [210].

ESI is the most compatible method for the analysis of liquids including biological fluids and it has virtually no limit on molecular mass for the compounds under analysis [211]. ESI is usually carried out when the mass spectrometer is coupled to a capillary emitter such as microcapillary chromatography [212], microfluid devices [213], or capillary electrophoresis [214]. The analytes in the biological fluid are ionized within the atmosphere of the laboratory forming a Taylor's cone. The Taylor's cone is a spray of droplets with ions of the analyte, and as the solvent evaporates, the droplet shrinks until completely desolvated, leading to the formation of highly-charged species [215]. The ions are then injected directly into the mass spectrometer with high vacuum at the ion source (Figure 5). This particularity accounts for the fact that ESI may be affected by environmental interference and therefore affect the detection sensitivity, accuracy and reproducibility of compounds of low abundance within the biological sample [214, 216].



**Figure 5.** Schematic depiction of an ESI source, following separation of a biological fluid (in this case desalted urine) components by capillary electrophoresis, operated in positive ion mode as in peptidome and metabolome analyses [217].

ii) *mass separation*: The step following ionization is the determination of  $m/z$  ratio of the ions using the mass analyzers of the mass spectrometer. There are several types of analyzers: quadrupole, ion-mobility, ion traps, and time-of-flight (TOF) analyzers.

In quadrupole analyzers, ions travel across the analyzer in a pulsed beam mode, i.e., the ions are trapped radially by a two-dimensional radio frequency field and axially by stopping potentials applied to end electrodes. Quadrupoles are often used to filter desired masses prior to fragmentation [218].

Ion-mobility analyzers allow separation of ions based on their mass, charge and mobility in an electric field. This analysis gives information about the structures and stabilities of molecules. However, one drawback of this technique is that the time of analysis is quite long [219].

Ion traps confine ions in an electromagnetic field over a prolonged period, within limited volume, in a three-dimensional quadrupole. A voltage is then applied to the electrodes of the trap affecting the trajectories of the ions which become unstable. The ions then leave the trap according to their  $m/z$  ratio and reach the detector. The main drawback of mass spectrometers with these analyzers is the cost of the apparatus, especially compared to TOF-based mass spectrometers [218].

Finally, TOF mass spectrometers are widely used mass analyzers and these mass spectrometers will be of prime interest for this thesis because this is the technology that we use in our laboratory for urine peptidome and metabolome analysis (Figure 6). For the analysis, ionized molecules are accelerated in an electric field and then ejected in a flight tube under vacuum. Smaller ions will fly faster than larger ions in the tube, and this time of flight for each ion will then be measured by the detector. There are two modes of operation of the analyzer: linear or reflectron. In the linear mode, ions fly in the tube and reach the detector. In the reflectron mode, which is more accurate than the linear mode, ions fly towards a reflectron which focuses ions with the same  $m/z$  ratio so that these ions reach the detector at the same time. The advantages of using a TOF analyzer is that the molecular mass range is unlimited, it has a high resolution, a high mass accuracy, and compared to other analyzers such as the quadrupole, most ions will reach the detector (there is fewer ion loss) [218, 220].

iii) *detection*: The last step before obtaining analytes is the detection. In MS, there are several types of detectors. For TOF-MS, in addition to having a high sensitivity, together with a linear and quantitative response, they should be able to have a very rapid readout and response, a very high-count rate ( $>10^6$  counts/second), low or no noise, simultaneous detection, wide mass-range response, short recovery time and a high saturation level [221].

## ii. Coupling with separation techniques

To increase the number of molecular species analyzed in a complex biological sample, mass spectrometers are coupled to separation techniques such as two-dimensional polyacrylamide gel

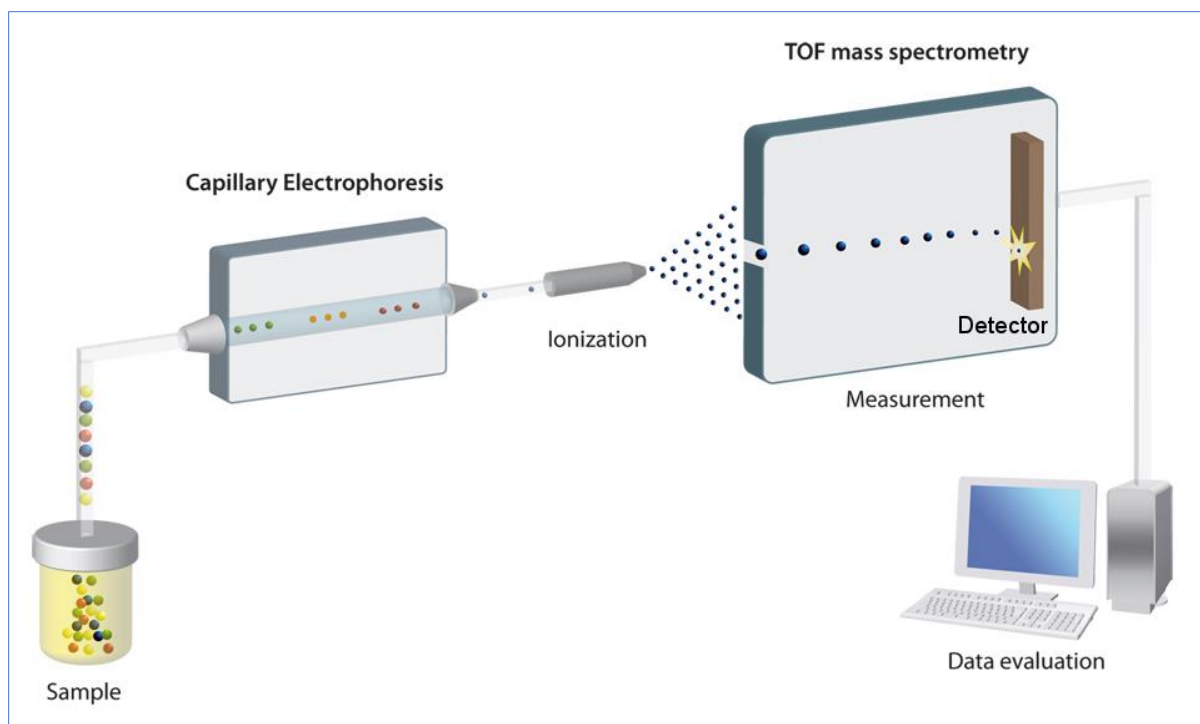
electrophoresis (2D-PAGE), liquid chromatography (LC), or capillary electrophoresis (CE). They allow separation of the samples into small sub fractions, increasing the sensitivity of the analysis. 2D-PAGE is a technique that has been widely used for proteome analysis in the past, but is hardly used anymore because it is time-consuming, cannot be used to analyze small proteins < 10kDa, is not appropriate to analyze highly hydrophobic proteins, displays strong of gel-to-gel variation and finally the quantification of the proteins is affected by the sensitivity of the stain used. Moreover, only up to a maximum of ~400 protein spots can be analyzed per gel [222-224]. 2D-PAGE is therefore not discussed in detail.

#### **a. Liquid chromatography**

LC is a technique routinely used for both proteomic and metabolomic profiling. It consists in separating analytes based upon the rates at which they elute from a stationary phase typically over a mobile phase gradient. Different analytes usually have different affinities for the stationary/mobile phase during separation such that those which will be more attracted to the mobile phase will elute more rapidly (short retention time) than those retained by the stationary phase which will elute more slowly [225]. For omics analysis, most of the time high-performance liquid chromatography is performed. It consists of a column packed with hydrophobic beads for the stationary phase, a pump that creates pressure and moves the mobile phase (which is polar) through the column. Hydrophilic analytes elute more rapidly than hydrophobic ones. Moreover, analytes with the same molecular mass but different polarity will elute at separate times. The advantages of LC coupled to MS are that up to 1500 proteins can be analyzed in a single run, non-polar proteins can be analyzed and that the system can be automated [222]. In addition, the high separation and ionization efficiency of LC separation followed by MS analysis also allows the analysis of metabolites and up to ~1100 metabolites in a single run can be detected [226]. However, sample preparation and instrumentation performance may affect quantification of metabolites, and the whole analysis often lasts for several hours [227].

#### **b. Capillary electrophoresis**

Capillary electrophoresis is a technique which allows separation of analytes with respect to their charge and size within an electric field. It can be used for the analysis of proteins and metabolites. This separation technique is different from chromatography and is more efficient for separation of analytes since there is no transfer between phases [228]. CE-MS is performed with either a sheathless or a sheath liquid interface. In both approaches, the end of the capillary is inserted in a needle at the entry of the ion source of the mass spectrometer (Figure 6). In the sheathless interface, only the contents that elute from the capillary form droplets that are turned into a spray during the electrospray process. In the sheath liquid interface, the capillary content is mixed with a liquid that surrounds the tip of the capillary at the outlet, thereby diluting the content of the capillary. Although some sensitivity is lost due to dilution of the sample, the sheath liquid ensures a more stable interface for electrical connection between the CE and the mass spectrometer to ensure efficient and uninterrupted separation [229].



**Figure 6.** Schematic representation of the CE-MS workflow for urine analysis, adapted from [230].

In contrast to LC, no elution gradients are required for separation, since the migration of the analyte is controlled by the electric field strength [231]. This is a strong advantage as the use of elution gradients could induce interference with subsequent detection by MS, due to a continuous change of composition during separation [232]. On another hand, CE-MS is limited to the detection of small proteins and peptides and only a small sample volume can be loaded onto the capillary, which decreases somewhat the sensitivity of the detection. Overall, the CE-MS technology has been developed into a commonly used proteomic technology, especially in the field of low-molecular weight proteome profiling (i.e. peptidome) and it has also been successfully used to study metabolites [233].

## 4. Biological fluids

Peptide and metabolite biomarkers of kidney disease were most frequently studied in blood and urine.

### Blood

Blood is a relatively easily accessible body fluid, and consequently many groups have analyzed the blood (serum or plasma) -ome. For proteome analysis, serum has the advantage of being depleted from abundant coagulation factors. But it is likely that proteolytic activity persists, leading eventually to protein degradation upon sample treatment and storage, resulting therefore in an alteration of the serum proteome [234]. This contributes to experimental inaccuracy, and hampers comparison of serum proteome or peptidome profiles between individuals. Alternatively, plasma, the most complex body fluid, can be used. Plasma high complexity is mainly due to the dynamic range with which proteins and

peptides are found, spanning 10 orders of magnitude [235, 236]. The 10 most abundant proteins represent >90% of the total plasma content, albumin being first, and may interfere with the identification of less abundant proteins. Highly abundant proteins can be removed by employing depletion techniques, but this removal potentially introduces biases and may induce loss of the low abundance proteins and peptides. It has been shown that depletion of albumin induces an additional loss of ~1000 low abundance proteins and increases variability [237]. Thus, although blood analysis has a large potential to provide us with biomarkers of (kidney) disease, both plasma and serum proteomics have drawbacks that need to be overcome.

For metabolome analysis, both plasma and serum are used [238]. However, some metabolome-targeted studies have revealed that plasma metabolites are unstable and are affected by the anticoagulants used and by storage. The serum metabolome showed to be more stable, resisted freeze-thaw cycles and more than 1 year storage at -70°C [239, 240].

## **Urine**

Urine has been in the center of attention of scientists for many years now. In addition to being collected non-invasively, it is an attractive fluid to study the proteome/peptidome since it contains a very limited amount of proteins. Urine contains approximately 2000 proteins compared to plasma which contains more than 10,000 proteins. The presence of a low protein quantity in a sample decreases the probability of high abundance proteins to bind to and/or cover candidate peptide biomarkers [198, 241]. For laboratory analyses, the stability of urinary peptides is of great advantage compared to urinary proteins since they are the final breakdown products and are unlikely to be further cleaved in the bladder at 37°C. The urinary peptidome is very stable for 6 hours at room temperature, for 3 days at 4°C, for several years at -20°C and/or at -80°C [242, 243].

In addition, urine is also an attractive fluid for metabolites study since it contains few enzymes for metabolite degradation and few high molecular weight compounds which may mask the identification of metabolites. Analysis of human urine samples revealed that it contains ~2600 metabolites [244]. The main source of variability between analyses were mainly due to low intensity signals [245] and to fluctuations associated to hormonal changes and time of exposure to medications or a particular diet for example [246]. Moreover, urinary metabolome has shown to be stable during storage for 26 weeks at -25°C [247] and during freeze-thaw cycles [248].

## **5. Urinary peptidome analysis by CE-MS**

For the study of kidney diseases urine is known as the ‘liquid biopsy’ since it has been estimated that, under physiological conditions, 70% of the urinary proteins originate from the kidney and the urinary tract, while the 30% remaining originate from plasma [243].

## **i. Kidney disease**

One key example of the use of urinary peptidome analysis in adults and two examples in the pediatric population for kidney disease will be given below.

*Chronic kidney disease (CKD):* CKD represents the best studied condition using urinary peptidome analysis. In 2010, Good *et al.* published a combination of 273 urinary peptides, of known sequence, as potential biomarkers of CKD in the adult population. The 273 peptides were identified by CE-MS by comparing the urine samples of 379 healthy subjects and 230 biopsy-proven CKD patients. This analysis showed that fragments of the major circulating proteins were present with increased abundance in urine of CKD patients, suggesting failure of the glomerular barrier filtering characteristics, and different collagen fragments displayed decreased abundance, suggesting an accumulation of intrarenal extracellular matrix. Protein fragments associated with progression of CKD originated mostly from proteins related to inflammation and tissue repair [194, 249]. These 273 peptides were combined in a machine-learning algorithm called the CKD273 model. This model was then validated a first time on an independent blinded cohort of 144 patients, with 34 healthy individuals and 110 CKD patients. The CKD273 model showed a sensitivity of 85.5%, a specificity of 100% and an AUC of 0.955 [193]. This classifier was further validated on a larger cohort of patients and could detect early and late CKD independent of age and gender [230, 250, 251]. Furthermore, an assessment of the performance of the CKD273 model on a large cross-sectional multicenter cohort detected normo- and micro-albuminuric patients and outperformed creatinine and albuminuria predicting a fast CKD progression (eGFR decline >5% per year). Just recently, Currie *et al.* (2018) showed that there was a significant correlation between the CKD273 score and mortality [252]. The most interesting aspects of the CKD273 model are not its diagnostic capacities but the apparent predictive capacities of progressive disease at a very early stage CKD, before apparent clinical signs. It is therefore now used in a multicenter prospective clinical trial, to select at an early stage progressive CKD (ClinicalTrials.gov Identifier: NCT02040441 [253]).

*Ureteropelvic junction obstruction (UPJO):* The identification of urinary biomarkers as diagnostic tool has also been applied for the diagnosis of pediatric kidney anomalies. The first study concerns the detection of hydronephrosis associated to UPJO. UPJO is a frequently encountered clinical situation in which urine flow to the bladder is restricted and if untreated can lead to renal damage. In mild obstruction UPJO resolves naturally and in severe cases surgery is needed. However, for intermediary UPJO, it is not clear whether surgery is required. Neonates with intermediate UPJO need close surveillance with repetitive isotope excretion scans up to the age of 2 years, to identify those in need for surgical removal of the obstruction. Our research team has identified 51 urinary peptide markers which were differentially abundant between children with spontaneous resolution of UPJO and those who needed surgery after birth [254]. A model of those 51 peptides predicted severe UPJO in 34 out of 36 children (94%) up to 15 months in advance of the actual surgery in a blinded cohort [255, 256]. This model of 51 peptides

was further validated on 27 children in a separate study. For 19 children, aged < 1 year, the model predicted surgery with 83% sensitivity and 92% specificity. However, when the model was validated on older children, the sensitivity was of 20% and the specificity of 66%. This suggest that there are changes in urinary proteome occurring with age in children and that urinary peptide models should be used in their context of development, in this case, detection of severe UPJO before 1 years old [257].

*Posterior urethral valves (PUV):* Posterior urethral valves (PUV) is a rare developmental disease which represents the most common cause of lower urinary tract obstruction in males. The majority of the infants with PUV surviving the neonatal period progress to CKD, and approximately 20 to 30% progress to ESRD in the first decade of life [258-260]. An important clinical challenge upon antenatal detection of PUV is to predict post-natal renal function. Current methods to predict these outcomes *in utero* are controversial [261, 262]. A comparative analysis of the fetal urinary peptidome of 15 fetuses with PUV displaying normal post-natal renal function and 18 fetuses with PUV with early ESRD lead to the identification of 26 peptides differentially expressed between the two groups. Out of these 26 peptides, 12 peptides were the most significant and were combined in a machine-learning algorithm called 12PUV model. The 12PUV model was then validated on an independent blinded cohort of 38 PUV patients with an AUC of 0.94, a sensitivity of 88% and a specificity of 95% [263]. This fetal urinary peptide-based model outperformed all conventional clinical parameters such as prenatal ultrasound and fetal urine biochemistry, none of which reached the same level of predictive sensitivity and specificity. The 12PUV model will potentially allow truly informed prenatal counseling, and hopefully will avoid unnecessary termination of pregnancy. Furthermore, prediction of postnatal renal function will be helpful in tailoring clinical follow-up and planning for clinical management of ESRD. Most importantly, reliable prediction of postnatal outcomes would greatly reduce the psychological burden of prognostic uncertainty imposed on the affected families. To study the wider employability of the 12PUV model, it is currently being validated in a multicenter European study called “multicentre vAlidationN of a fetal urine pepTidome-based classifiEr to predict post-natal reNAI function in posterior ureThral vALves” (ANTENATAL; ClinicalTrials.gov Identifier: NCT03116217).

## **ii. Cardiovascular disease (CVD)**

As described in chapter 2, CVD is the leading complication in CKD in children. With the available clinical tools including echocardiography and Doppler ultrasound, CVD complications are often detected at a late stage where lesions are irreversible. Urinary biomarkers of CVD might offer the possibility for the early identification/prediction of CVD of the most at-risk patients and might provide additional information about disease-related pathways. Urinary biomarkers of CVD have been identified, although not yet on a CKD background.

A first study in 2008, on a cohort of 370 adult patients with (n=88) or without (n=282) coronary artery disease (CAD) lead to the identification of 15 urinary peptides that diagnosed CAD with a sensitivity of

98%, a specificity of 83% and an area under curve (ROC) of 0.94 in a blinded validation set. Five out of the 15 urinary peptides were identified by sequencing [264]. All of them were collagen fragments (type I or III fragments) and they were all more abundant in CAD patients' urine compared to controls. In the normal artery, both synthesis and degradation of extracellular matrix proteins are remarkably slow and balanced [265]. During atherosclerosis there is an increased synthesis of many matrix components, including collagen types I and III, elastin, and several proteoglycans. Increased abundance of collagen fragments in the urine of CAD patients may be explained by an alteration in the turnover of these proteins during physiopathology. Collagen types I and III are predominant proteins in the arterial walls and in the thickened intima of atherosclerotic lesions. Furthermore, in accordance with these observations concerning an elevated collagen degradation levels, increased circulating levels of collagenases, such as MMP-9, have been noticed in CAD [264, 266].

Additional urinary peptide-based models were developed for the identification of other CVDs. For the prediction of acute coronary syndrome (ACS) defined as progression of atherosclerotic plaques towards an inflamed, unstable fibroatheromas that are susceptible to cause a thrombotic occlusion of coronary arteries, 75 urinary peptides were identified, on a cohort of 252 patients. They were combined in a model called the ACSP75 model, which successfully diagnosed ACS in asymptomatic patients with a sensitivity of 73.8% and a specificity of 73.8% [267]. In another study comparing the urinary peptidome of 33 controls with 69 patients suffering from stroke 35 urinary peptide biomarkers were identified and combined in a model which had a sensitivity of 56%, a specificity was 93% and an AUC on ROC analysis of 0.86, allowing the diagnosis of acute stroke in patients with mild symptoms [268]. Furthermore, a study on a cohort of 70 patients, with the aim to diagnose early left ventricular diastolic dysfunction in hypertensive patients lead to the identification of 85 urinary peptides with a sensitivity of 69%, a specificity of 94% and an AUC on ROC of 0.84 [269]. Urinary peptide biomarkers have also been identified for diagnosis of heart failure with reduced ejection fraction [270] and heart failure with reduced ejection fraction due to chronic kidney disease [271].

Taken together, urinary peptide analysis showed promising results in predicting CKD and its progression and potential early detection of several CVDs. Of note, studies on the identification of urinary peptide biomarkers of CVD complications in children with CKD are clearly missing in the scientific literature.

## **6. Urinary metabolome analysis by CE-MS for CKD**

The most common metabolite biomarker used in CKD diagnosis is creatinine [272, 273]. However, compared to the identification of urinary peptide-based biomarkers, the literature concerning the identification and validation of urinary metabolite biomarkers for the diagnosis or prognosis of CKD or its complications using MS is poor.



A study on a cohort of 49 adult patients concerning the identification of urinary and plasma metabolites by MS/MS techniques lead to the discovery of 30 metabolites comprising 17 plasma metabolites and 13 urinary metabolites associated to CKD. As observed in previous studies, a significant decrease of asymmetric dimethylarginine (ADMA) in the urine was significantly correlated with a decrease in the eGFR [274]. ADMA is an analogue of L-arginine that occurs naturally in the human circulation. A decrease of urinary ADMA may suggest accumulation in the circulation and studies have shown that high ADMA levels diminish endothelial function and contributes to renal impairment [275]. Another study identified a panel of 13 urine metabolites linked with mitochondrial metabolism. The metabolites were significantly decreased in patients with diabetes and CKD compared with healthy controls [276, 277]. In another study involving 240 patients, adding plasma and urine metabolites to a model of baseline albuminuria and estimated glomerular filtration rate significantly improved risk prediction for the development of macroalbuminuria in diabetic individuals [278].

### **Challenge for the study of urinary metabolome**

Very few studies have investigated the use of the CE-MS technology in the discovery of body fluid metabolite markers. Only one study by Kimura T *et al.* (2016) was carried out on the identification of plasma metabolites to predict progression to ESRD. The study cohort consisted of 112 stage 3-5 CKD patients, with median follow-up period of 4.4 years, and the plasma metabolites were analyzed by LC-MS or CE-MS. On a total of 218 detected plasma metabolites, 16 metabolites could predict progression of the disease to ESRD. However, neither AUC, sensitivity or specificity was reported, and validation was not performed [279].

The use of body fluid metabolome analysis in CKD and its complications is still anecdotal and validation studies are missing.

# IV. Aptamers: 'omics' translation for bedside detection of biomarkers

## 1. What are aptamers?

### i. Definition:

Aptamers are short single-stranded DNA (ssDNA) or RNA which can specifically bind targets with high affinity [280-282]. In the literature, we can find a large diversity of molecules for which aptamers have been selected [283]: toxins [284, 285], dyes [281], ATP [286], metal ions [287, 288], and proteins [289, 290].

### ii. Structure:

Contrary to the information contained in the intrinsic nucleotide chain of genes or mRNA molecules, here it is the molecular spatial arrangement (secondary and tertiary) that gives aptamers their properties. The major difference between the well-known helical DNA structure, and the single-stranded oligonucleotides is that single-stranded molecules fold into three dimensional structures due to intramolecular Watson-Crick interactions, aromatic ring stacking or Van der Waals forces of attraction, that lead to different fascinating shapes depending upon the length and nucleotide arrangement in the sequence (Figure 7). Therefore, aptamers can adopt complex structures such as hairpins, G-quartets or pseudo-knots, that will determine their ability to bind a target [282].

### iii. Advantages:

Aptamers are often compared to antibodies, but they offer much more advantages. First, aptamers have a great discriminatory potential and are very specific for their ligand, as exemplified by the RNA aptamer which targets theophylline and that does not bind to caffeine, knowing that theophylline and caffeine differ only by a methyl group [291]. Second, aptamers are stable, even if they are denatured and renatured several times. Their structure can be modified, by adding base-pairs to the stem structures, to further increase their stability and their capacity of binding can also be improved by modifying the phosphodiester linkage [292]. For use as a detection tool, aptamers can be coupled to reporter molecules such as fluorochromes, without altering their conformation and their binding affinity. For *in vivo* applications, aptamers are small enough to reach tissues easily. Their average molecular weight ranges from 15 to 30 kDa compared to 150 kDa for IgG. They have a low immunogenicity since the human immune system does not recognize oligonucleotides as foreign agents [293]. Finally, while antibody production is laborious, takes months, is very expensive and requires the use of animals [294], aptamers

are chemically synthesized *in vitro* at low cost and in large quantity, with great accuracy, purity and reproducibility. These properties make aptamers relevant alternatives to antibodies as diagnostic, analytical or therapeutic agents.

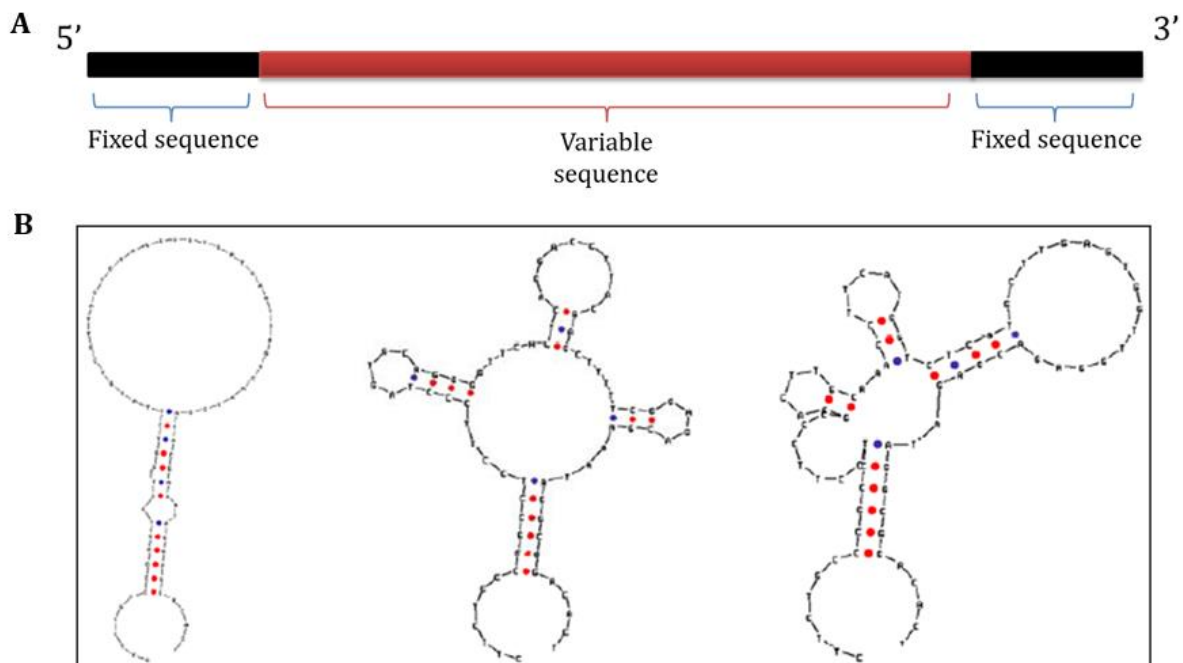
## 2. Aptamer selection

Aptamers are selected *in-vitro* from a library of oligonucleotides by the systematic evolution of ligands by exponential enrichment (SELEX).

### i. Library of oligonucleotides

Aptamers are selected from a library which consists of a pool of  $10^{13}$ - $10^{15}$  RNA or ssDNA molecules. Each molecule has a central random sequence flanked by two fixed sequences. These fixed sequences have a length which varies between 18-22 nucleotides [295] (Figure 7).

The central variable part is responsible for the aptamers' specific secondary/tertiary structures and their size varies from 8 to 50 nucleotides [296].



**Figure 7. A.** Schematic representation of an aptamer showing the random sequence flanked by two fixed sequences, which are complementary to PCR primers. **B.** The random sequence is responsible of the different folding shapes of aptamers.

According to Gold *et al.* (2012), libraries with at least 25 randomized nucleotides are necessary to offer enough length to generate stable secondary and tertiary structures such as hairpins, G-quartets or pseudo-knots [296]. However, correlation analysis between the length of an aptamer and its affinity for a target,

in all reported selection studies between 1990 and 2013, did not reveal any significant association between those two parameters [297]. Intuitively, the longer the randomized sequence, greater are the conformation possibilities and chances of finding an aptamer for the target. Nevertheless, for a successful selection of aptamers recognizing small targets, it has been suggested to start with a pool of short DNA sequences to limit crowding during aptamer-target interaction and increase selection efficiency [296].

## ii. SELEX

### a. Principle

The SELEX consists of rounds of selection to screen the library for sequences having the highest binding affinity for a defined target [280]. The process [295] is divided into four main steps (Figure 8):

**Step 1 – Incubation of the library with the target:** the oligonucleotides are first heated to ~ 95°C for linearization and then cooled down rapidly on ice so that they fold into stable structures. They are then incubated with the target.

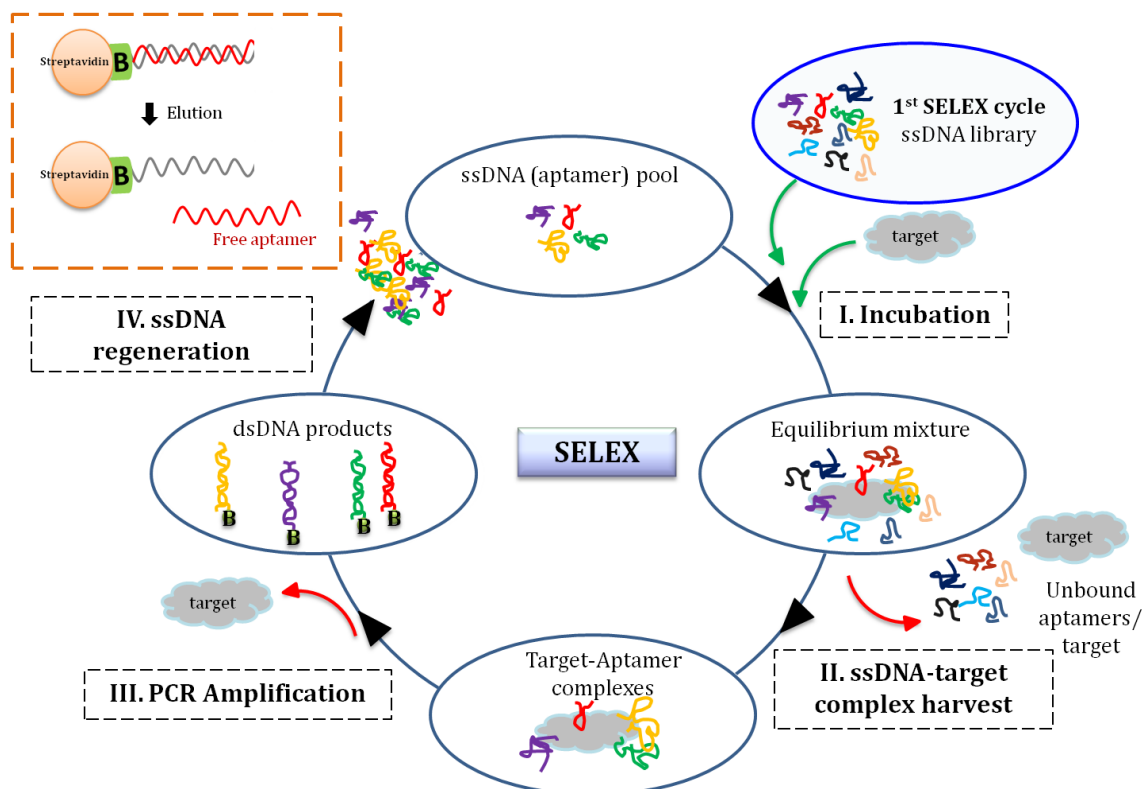
**Step 2 –Oligonucleotide-target complex harvest:** the target-bound sequences are separated from the unbound sequences using for example affinity columns or capillary electrophoresis (CE) and are collected by elution.

**Step 3 – Polymerase Chain Reaction (PCR) amplification of target-bound sequences:** during the denaturation step of the PCR, the bound oligonucleotide sequences are separated from the targets, and the oligonucleotide sequences are amplified by PCR.

**Step 4 –Regeneration of single-stranded oligonucleotides:** for ssDNA aptamers, the complementary strand is trapped on beads using for example the streptavidin-biotin affinity system. In this case, the previous step of PCR involves the use of a biotinylated reverse primer to label the complementary strand. When the PCR products are applied on the streptavidin beads, they get immobilized by the complementary strands and the non-biotinylated strand, i.e aptamer, can be regenerated by elution with a buffer having a high ionic strength [298]. For RNA aptamers, the process is similar except than an additional step for *in vitro* transcription should be added.

This cycle consisting of the four steps is then repeated for several times until the pool of oligonucleotide becomes enriched with strands of similar sequences and which have a high affinity for the target. At each iterative cycle, the stringency of selection is progressively augmented by increasing the aptamers:target relative ratio to increase the competition between the aptamers and the target. In addition, during the different cycles, PCR amplicons are sub-cloned and sequenced to see an enrichment of the pool with similar sequences. According to previous studies, at least 15 SELEX cycles are required to obtain an aptamer for a defined target [299, 300].

Finally, once the sequences are known, aptamers should be synthesized, and their binding affinities and specificities characterized [295, 301].



**Figure 8.** The systematic evolution of ligands by exponential enrichment (SELEX) cycle for selection of aptamers from an ssDNA library. This includes four main steps: incubation, ssDNA-target complex harvest, PCR amplification and ssDNA regeneration. At the first cycle, a ssDNA library is used for incubation with the target but for the next cycles, the ssDNA regenerated in the preceding cycle are used. B: Biotin-label; dsDNA: double-stranded DNA.

The most critical step in this whole SELEX procedure is the partition of the target-bound sequences from unbound sequences and different SELEX strategies can be used, as explained in part below.

### b. SELEX using affinity columns

Conventional SELEX is usually done by immobilizing the target using nitrocellulose filter [299, 302], streptavidin affinity-based methods [300, 303], magnetic beads [304, 305], or *N*-hydroxysuccinimide (NHS)-activated sepharose system [306, 307]. In this lastly mentioned case for example, the target could be any molecule with a free amino group, i.e. mostly peptide or protein. The target is first immobilized on the NHS sepharose column via a very stable amide bond between its amino group and NHS esters [308]. Then the library of oligonucleotides is applied to the column for the incubation step of the SELEX. After elimination of unbound oligonucleotides by a few washes, the bound aptamers are eluted using a strong ionic elution buffer. Before aptamer amplification, a desalting step is required since the high salt concentration of the elution inhibits the PCR reaction [307, 309]. This strategy has been used for the

selection of ssDNA aptamers targeting mucin-1, a glycoprotein, a tumor marker upregulated in several cancers [307].

### c. Capillary electrophoresis SELEX (CE-SELEX)

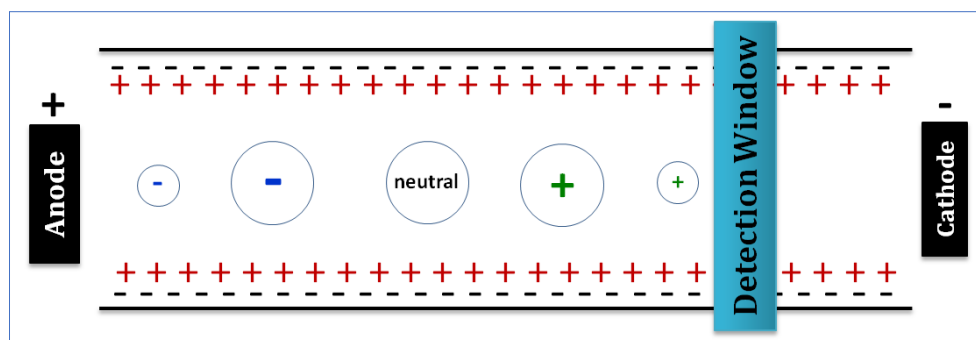
Capillary electrophoresis is an electrophoretic technique in which analytes are separated according to the size/charge [310]. The application of CE for aptamer selection was first published in 2004 by Mendonsa and Bowser. The principle is that oligonucleotides bound to a target undergo a mobility shift compared to free oligonucleotides. The bound sequences can therefore be collected just after the capillary exit separately from unbound sequences either by monitoring UV-absorbance or by laser-induced fluorescence for oligonucleotide libraries labelled fluorescently [311].

For an efficient selection of aptamers using CE-SELEX, care should first be brought to the choice of the capillaries as well as pH and composition of separation buffers.

Capillary walls used for CE-SELEX should not interact with the analyte since any interactions results in low resolution during separation, and peak tailing. Most of the time, bare fused-silica capillaries are used and the pH at which the separation is performed is usually greater than 2. In this case, two processes are involved in separation by CE: electroosmotic flux (EOF) and electrophoretic mobility (EM).

**EOF:** at a pH > 2, the silanol groups present in the capillary wall are deprotonated. This results in a net negative charge on the capillary wall. Therefore, cations present in the solvent are attracted to the negatively charged capillary wall surface creating a positively charged mobile layer. The cation mobile layer is attracted to the cathode, thereby generating the EOF. EOF creates a ‘conveyor belt effect’, carrying any analyte to the capillary outlet at the cathode, independent of their charge [312].

**EM:** EM occurs concurrently to EOF and affects only charged molecules. Positively charged molecules migrate toward the cathode whereas negatively charged compounds are attracted to the anode. The cumulative effect of EOF and EM will lead to separation of compounds. Positively charged molecules will be the first to rush to the cathode with both EOF and EM driving in the same direction, smaller positively charged ions being detected first followed by bigger ones. Then, neutral molecules will exit the capillary since they are carried only with the EOF. Finally, negative molecules will reach the outlet, with EOF and EM driving in opposite directions, with bigger negatively charged compounds being detected before smaller ones [313, 314] (Figure 9).



**Figure 9.** Order (from left to right) in which the different analytes of different mass and charge will reach the detection window in CE. The negative charge of the capillary wall surface is due to the deprotonation of silanol groups at  $\text{pH} > 2$ ; the positive charges come from positive ions in the buffer to ensure electroneutrality.

For an optimal selection, the CE separation buffer should mimic the conditions in which the aptamer will be used. In addition, its ionic composition should not be high to avoid Joule heating. Joule heating is defined as the heat generated inside the capillary when an electric current is applied to an electrolyte. This increase in temperature is removed to a certain extent by the cooling system of the CE apparatus but when the temperature rises too much, this affects the viscosity of the buffer, leading to an increase in the diameter of the capillary, and causes a non-uniform migration of the analytes by affecting the electrophoretic velocity and the molecular diffusion [315].

### 3. Clinical applications of aptamers

#### i. Aptamers as therapeutic agents

Since the discovery of aptamers, several studies have been launched to see whether they could be used as therapeutic agents. The first aptamer to reach into the clinic was NX1838, an RNA aptamer which had a very high affinity for vascular endothelial growth factor 165 ( $\text{VEGF}_{165}$ ) and which prevented it to bind to its receptor [316]. The efficacy of NX1838 has been demonstrated for the treatment of the wet form of age-related macular degeneration (ARMD). Licensed and renamed as Macugen®, this aptamer received the food and drug administration (FDA) approval in 2005. However, Macugen® was not a commercial and a financial success since it was selected to target specifically  $\text{VEGF}_{165}$  while another, smaller VEGF isoform,  $\text{VEGF}_{121}$ , also involved in the development of ARMD, was not targeted by Macugen® [296].

Other aptamers are now in clinical trial. For example, E10300 targets PDFG-B in ARMD and diabetic retinopathy. AS1411 binds to nucleolin and was proposed for acute myeloid leukemia and renal carcinoma treatment. NOX-E36 which targets the pro-inflammatory chemokine C-C motif-ligand 2 (CCL2), was tested in a context of kidney diseases [317]. Safe and well tolerated, this aptamer

(commercialized under the name of Emapticap pegol) reduces glycated hemoglobin (an index of glycemia) as well as albumin:creatinine ratio in patients with type 2 diabetes [318].

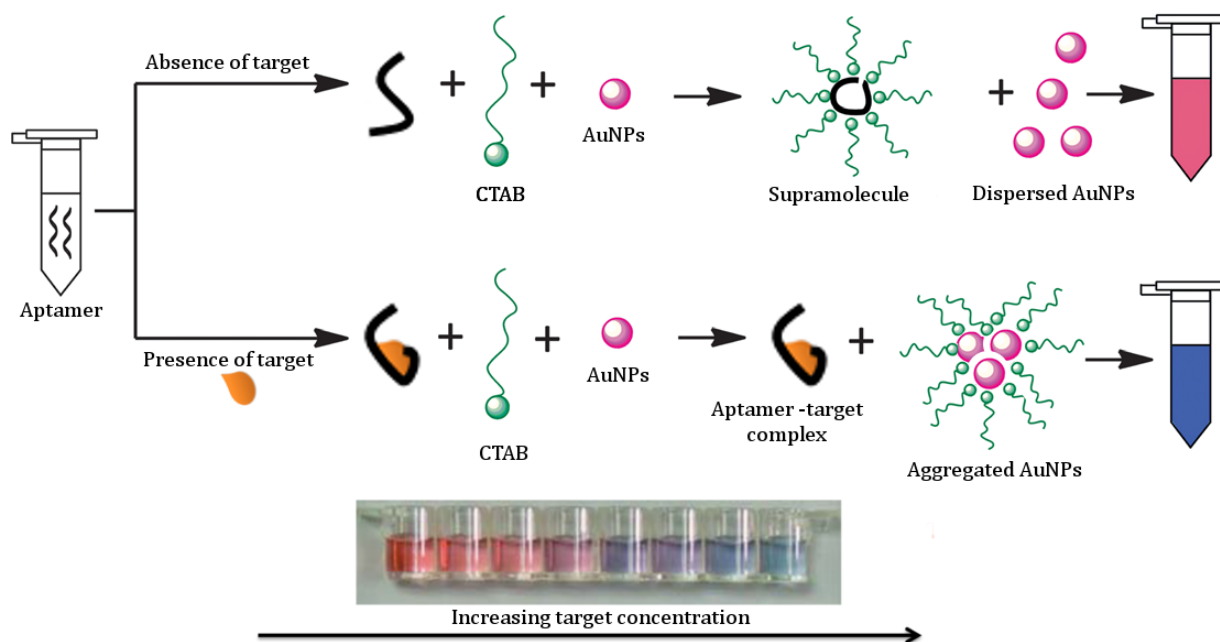
## **ii. Aptamers as biosensors**

Biosensors are tools used for biological recognition that could be used to identify biomarkers at the patient's bedside [319]. When binding to their targets, aptamers undergo a three-dimensional conformation change that can be monitored either with an electrical signal using voltammeter, potentiometry measurement or electric impedance spectroscopy [320, 321], or by fluorescence [322, 323]. Interestingly, aptamers can also be coupled to nanomaterials to transduce the induced conformational change in an easy-to-use colorimetric test [285]. In most cases, the colorimetric sensing method employs gold nanoparticles (AuNPs), since they are biologically inert material [324], in association with a cationic surfactant (e.g. hexadecyltrimethylammonium bromide (CTAB)). The surfactant does not only bind to DNA to form supramolecules but also induce the aggregation AuNPs. The principle of aptamer-target detection relies on the change in the distribution of nanoparticles in the medium due to aptamer-target complexes formation [285]. When AuNPs are dispersed, the inter-particle distance is greater than the diameter of the nanoparticle and the solution appears red. When the nanoparticle aggregates, this inter-particle distance decreases and causes a color change from red to blue [325-327]. In the absence of target, the surfactant and free aptamers form a supramolecule, and thus the subsequent AuNPs cannot aggregate due to the lack of free surfactant (Figure 10). In the presence of a target, the aptamers bind preferentially to their targets and the following surfactant molecules can assemble AuNPs to form aggregates, thereby leading to a color change from red to blue [328, 329] (Figure 10).

This strategy has been used in the context of malaria for the development of an aptasensor-based diagnostic tool targeting *Plasmodium vivax* and *Plasmodium falciparum* lactate dehydrogenase [329].

For now, aptasensors are extensively studied in research and are already available to target thrombin, theophylline, PDGF and cocaine. However, no aptasensor has reached clinical trial up to now [330, 331].





**Figure 10.** Schematic representation of aptasensors based on surfactant (CTAB)-induced aggregation together with the visual color changes that occur with increasing target concentration, adapted from [328].

## Challenge for the development of biosensors for biomarker detection

The use of CE to separate free nucleic acids and collect aptamer-target complexes has been a major improvement. Due to the high resolving power of CE associated to the low probability for nonspecific interaction with the stationary phase (selections are done in free solution), CE-SELEX typically takes less than 15 cycles and is applicable to a large variety of targets, including small molecules or organic target having only no suitable functional groups [332]. Thus, several aptamers have been selected using the CE-SELEX, including aptamers recognizing HIV reverse transcriptase [333], alpha-fetoprotein [334], porphyrin [335] and thrombin [336].

For many years now, our research team has identified several urinary peptide biomarkers. However, this identification requires a technology which cannot be used at the patients' bedside and need some trained personnel. For the detection of biomarkers, previously identified by CE-MS by our research group, we aimed to develop an aptamer-based biosensor that will facilitate the diagnosis system and provide a tool that could be used by anyone at the patient's bedside. For this, our challenge is to use CE-SELEX and select aptamers that would bind to selected biomarkers freely in urine.

## **Study 1:**

# Cardiovascular complications (CVD) in children with CKD

(Manuscript in preparation)

---

## **Aims and Objectives**

CVD the main cause of death in the CKD pediatric population. Biomarkers that could predict progression of cardiovascular complications in children with CKD are unfortunately not available for now. The objective of this work was to identify urinary peptide biomarkers that could predict CVD progression in the largest European cohort of children with CKD. The parameters monitored as surrogate markers of cardiovascular complications, i.e. atherosclerosis, vascular stiffness and cardiac hypertrophy, were carotid intima-media thickness, pulse wave velocity and left ventricular hypertrophy respectively. Such biomarkers could revolutionize identification of high-risk pediatric patients so that their therapeutic treatment could be personalized to delay a cardiovascular event or even death.

# Urinary peptide biomarkers predictive of the progression of cardiovascular complications in children with chronic kidney disease

Brunchault et al. (*in preparation*)

## Introduction

Chronic kidney disease (CKD) affects approximately 10% of the worldwide general population [1] and affects both adults and children. Regardless of age, CKD is associated with several comorbidities and high mortality. In 2017, epidemiological data from the European Society for Paediatric Nephrology/European Renal Association-European Dialysis and Transplant Association (ESPN/ERA-EDTA) registry showed that the overall mortality rate was 32.6 per 1000 children with advanced CKD compared to 3.7 per 1000 children in the European general pediatric population [2]. Using a cohort of 1643 children with CKD, the Australia and New Zealand Dialysis and Transplant Registry also showed that the mortality rates differ according to the type of renal replacement therapy: 4.8 per 100 patient-years among patients receiving hemodialysis, 5.9 per 100 patient-years among those having peritoneal dialysis and 1.1 per 100 patient-years among those who received a renal transplantation [3]. Like in adults, cardiovascular disease (CVD) is the most important comorbidity associated with CKD in the pediatric and young adult population. CVD is responsible for 50% of the deaths, cardiac arrest being the first cause of CKD-related CVD followed by arrhythmia, cardiomyopathy, and stroke [4-7]. As shown in the Chronic Kidney Disease in Children (CKiD) study (United States and Canada) following 586 children, 31% of dialyzed patients experienced a cardiac-related incident within 7 years of follow-up [8]. The most common cardiovascular lesions observed in pediatric patients with CKD are left ventricular hypertrophy [9-11] and vascular remodeling associated with accelerated atherosclerosis and increased arterial stiffness [12, 13]. Non-invasive methods that could help predicting the progression of these complications would strongly improve specific patient clinical management and should result in a decrease in morbidity and mortality due to preventable CV events in patients with CKD.

Urine is increasingly investigated as ‘liquid biopsy’ since it is a potential source of markers of disease. Using approaches based on capillary electrophoresis coupled to mass spectrometry (CE-MS), we and others have shown the usefulness of urinary peptidome analysis for prediction of progression of CKD in both adults [14] and children [15]. Furthermore, in several small-sized studies in non-CKD adult populations, a number of urinary peptide biomarkers have been associated to the occurrence of CVD

such as coronary artery disease [16], hypertensive atherosclerotic CVD [17], heart failure [18], left ventricular diastolic dysfunction [19] and, recently, acute coronary syndromes [20]. However, urinary peptides able to predict development of CVD in the pediatric population have not yet been identified, irrespective of whether those CVD develop on a CKD background.

In the current study we analyzed the urinary peptidome of 86 pediatric patients with CKD (eGFR 10-45 mL/min/1.72m<sup>2</sup>), without overt cardiac dysfunction at inclusion, from the Cardiovascular Comorbidity in Children with CKD (4C) study [21] by CE-MS. We investigated the use of urinary peptides as biomarkers to predict progression of CVD over one year, using left ventricular mass index (LVMI), carotid intima media thickness (cIMT) and pulse wave velocity (PWV) as markers of cardiovascular complications.

## **Materials and Methods**

### ***Study Population***

The 4C study is a prospective observational cohort study in pediatric patients with CKD. It included 700 patients recruited in 55 pediatric nephrology centers in 12 European countries. Inclusion criteria were ages 6–17 years old and GFR (estimated using Schwartz formula [22]) between 10 and 45 ml/min/1.73 m<sup>2</sup>. Non-inclusion criteria were existing transplants, active systemic vasculitis, renal artery stenosis, coexisting primary cardiovascular anomalies, and anomalies of the limbs preventing diagnostic procedures. The study was approved by local ethics committees in all participating centers, and parents or legal guardians provided informed consent for study participation. Eighty-six patients from the 4C cohort whose detailed CVD follow-up information were available were randomly picked for the identification of urinary peptides associated with the progression of CVD.

cIMT, PWV and LVMI were measured as surrogate markers of CVD, namely atherosclerosis, vascular stiffness and cardiac hypertrophy respectively [23], and the rate of changes of these parameters (i.e slopes, measured over one year) were used to determine the CVD progression. For every CVD marker, the 86 patients were randomized in a discovery (2/3 of the population, n= 57) and a validation cohort (1/3 of the population, n = 29) for building and testing the peptidomics-based prognosis tools, respectively. Since no clear clinical cutoffs are available for the definition of patients displaying CVD, we defined the cutoffs by dividing the discovery population in tertiles of progression for the 3 measured parameters (**Figure 1**): T1 was composed of children considered with low-risk for CVD progression; T2 contained patients considered with intermediate risk for CVD progression; T3 was composed of children with high-risk for CVD progression. These threshold values dividing tertiles were then applied to the validation population to define the risk for CVD progression (**Figure 1**).

### ***Cardiovascular Disease Monitoring***

All patients underwent a complete transthoracic echocardiographic examination with commercially available machines. Examinations were stored and analyzed in a central echocardiographic reading center by two independent readers unaware of the clinical data.

Two-dimensional echocardiography images were obtained for the analysis of left ventricular (LV) volumes on three consecutive beats from apical four- and two- chamber views. Wall thickness and chamber dimensions were obtained from the two-dimensional parasternal long axis or M-mode short axis at the midventricular level. The LV mass index (LVMI) was calculated according to the Devereux Equation [24] and indexed to height<sup>2.16</sup> [25]. The sex- and age- specific LVMI partition values of Khoury *et al.* [26] are applied to define left ventricular hypertrophy.

The cIMT is measured according to the Mannheim cIMT consensus [27]. The cIMT is obtained either by five averaged measurements on each side or semi-automatically using a portable ultrasound device (Acuson P50; Siemens Medical Solutions USA, Inc.) with integrated digital image evaluation software (Syngo US Workplace; Siemens Medical Solutions USA, Inc.). Interobserver variation studies showed an intraclass correlation coefficient of 0.42 and an interobserver coefficient of variation (CV) of 7.3%. Because cIMT in children changes with growth, reference values normalized for height and age were established in 1155 healthy children ages 6–18 years old [28].

The central PWV is measured with the Vicorder Oscillometric PWV device using the distance from the suprasternal notch to the femoral recording point via the umbilicus as path length. The method was validated against the gold standard of applanation tomometry [29], and reference values normalized for height and age were established in a large European pediatric population (1003 healthy children ages 6–18 years old) [30]. Intra- and interobserver variability studies showed CVs of 5.6% and 5.8%, respectively, and intraclass correlation coefficients of 0.8 and 1.0, respectively [30].

### ***Sample preparation***

Urine samples were collected at enrollment by the clean-catch method and frozen at -20°C within the hour. Urine samples were shipped on dry ice to Inserm U1048, Toulouse, France and thawed immediately before use. A volume of 0.7 mL was diluted with 0.7 mL 2 M urea, 10 mM NH<sub>4</sub>OH and 0.02 % sodium dodecyl sulfate (SDS). To remove high molecular weight polypeptides, samples were filtered using Centriscart ultracentrifugation filter devices (20 kDa molecular weight cut-off); Sartorius, Goettingen, Germany) at 3000 g until 1.1 mL of filtrate was obtained. The filtrate was desalted with PD-10 column (GE Healthcare, Sweden) equilibrated in 0.01% NH<sub>4</sub>OH in HPLC-grade water. The prepared samples were lyophilized and stored at 4°C. Shortly before capillary electrophoresis coupled to mass

spectrometry (CE-MS) analysis, lyophilisates were resuspended in HPLC-grade water (Merck KGaA, Darmstadt, Germany). The preparation method has previously been described in more detail by Theodorescu *et al.* 2006 [31].

### ***CE-MS analysis and data processing***

CE-MS analysis was performed using a Beckman Coulter Proteome Lab PA800 capillary electrophoresis system (Beckman Coulter, Fullerton, USA) on-line coupled to a micrOTOF II MS (Bruker Daltonic, Bremen, Germany), as previously described [15, 32]. The electro-ionization sprayer (Agilent Technologies, Palo Alto, CA, USA) was grounded, and the ion spray interface potential was set to -4.5 kV. Data acquisition and MS acquisition methods were automatically controlled by the CE via contact-close-relays. Spectra were accumulated every 3 s, over a range of  $m/z$  350 to 3000.

MosaiquesVisu software package was applied to deconvolute mass spectral ion peaks representing identical molecules at different charge states into single masses [33]. Migration time and ion signal intensity (amplitude) were normalized using internal polypeptide standards [34]. Each polypeptide present in the list was defined by its normalized migration time [min], molecular mass [kDa], and signal intensity detected. Using a Microsoft SQL database, all detected polypeptides were deposited, matched, and annotated to allow for further comparison between the groups. The criteria applied to consider a polypeptide identical was that within different samples, the mass deviation was lower than 50 ppm for masses <4kDa, 150 ppm for masses >6kDa, and between 50-150 ppm for masses between 4-6kDa. Acceptable migration time deviation was between 1 and 2.5 minutes.

### ***MS-MS analysis***

CE was connected to an Orbitrap Velos FTMS (Thermo Finnigan, Bremen, Germany). Urine sample was ionized using a modified Proxeon nano spray source fitted with a non-grounded Agilent ESI sprayer operating in positive ion mode. Ionization voltage was 3.4 kV and the capillary temperature was 275°C. The mass spectrometer was operated in MS/MS mode scanning from 350 to 1500 amu. The top five multiply charged ions were selected from each scan for MS/MS analysis using HCD at 40% collision energy. The resolution of ions was 60 000 for MS1 and 7500 for MS2. For LC-MS/MS analysis, aliquots of 5 $\mu$ L were analyzed on a Dionex UltiMate 3000 RSLC nano flow system (Dionex, Camberley, UK) at a flowrate of 5 $\mu$ L/min. The trap and nanoflow column were maintained at 35°C. After loading (5 $\mu$ L) onto a Dionex 0.1 $\times$ 20 mm 5 $\mu$ m C18 nano trap column, elution was performed on an Acclaim PepMap C18 nano column 75 $\mu$ m $\times$ 15 cm, 2 $\mu$ m 100 $\text{\AA}$  at a flowrate of 0.3 $\mu$ L/min. Typically, samples were eluted with a gradient of solvent A: 97.9% H<sub>2</sub>O, 0.1% formic acid, 2% acetonitrile versus solvent B: 80% acetonitrile, 19.9% H<sub>2</sub>O, 0.1% formic acid starting at 1% B for 5 min rising to 20% B after 90 min and finally to 40% B after 120 min. Thereafter, the column was washed and reequilibrated prior to the next injection. The eluant was ionized using a Proxeon nano spray ESI source operating in positive ion mode into an Orbitrap Velos FTMS (Thermo Finnigan, Bremen, Germany). Ionization voltage was 2.6 kV and

the capillary temperature was 250°C. The mass spectrometer was operated in HCD MS/MS mode. Using HCD, the top 20 multiply charged ions were selected from each scan for MS/MS analysis and only charge state one was rejected for MS/MS [36]. The detection limit for the LC- or CE-MS/MS analysis using the Orbitrap Velos mass spectrometer, with 60 000 resolution for MS1 and with 7500 resolution for MS2, was in the range of 0.05–0.2 fmol. To obtain sequence information, CE- and LC-MS/MS were used as complementary approaches. CE-MS/MS has the advantage of direct matching (mass and CE-time) to the peptides quantified by CE-MS. On the other hand, LC-MS/MS exhibits higher sensitivity due to the increased loading capacity of the LC-column, consequently a better coverage of sequence information. However, the retention time in LC does not directly correspond to the CE migration time, which represents a disadvantage of LC-MS/MS in this approach [35].

### ***Protein sequence data analysis***

Data files were searched against the UniProt human nonredundant database using Proteome Discoverer 1.2 (Thermo) and the SEQUEST search engine. Relevant settings were: no fixed modifications, oxidation of methionine and proline as variable modifications. The minimum precursor mass was set to 790 Da, maximum precursor mass to 6000 Da with a minimum peak count of 10. The high-confidence peptides were defined by cross correlation ( $Xcorr \geq 1.9$  and  $rank=1$ ). Precursor mass tolerance was 5 ppm and fragment mass tolerance was 0.05 Da. False discovery rate settings cannot guarantee that the correct sequence is assigned to a fragmentation spectrum generated in an MS/MS experiment [36]. However, a property of CE is that the migration time is dependent on the net positive charge of the peptide. At pH 2 (pH of the running buffer) this (charge) is a function of the number of basic amino acids present [37]. The peptide sequences obtained from the tandem MS analysis were matched to the CE-MS peaks by matching the molecular mass and migration time (in the case of CE-MS/MS) or molecular mass and theoretical migration time based on the number of basic amino acids (in the case of LC-MS/MS). To further allow matching, the absence of cysteine and the absence of hydroxylated proline in non-collagen peptides were required. If a sequence passed all these criteria, it was then assigned to the corresponding CE-MS peak [37].

### ***Statistical Analysis***

*Multidimensional classifiers:* For the identification of peptides associated to markers of CVD progression, we compared the normalized abundance of peptides between T1 and T3 groups from the discovery cohort. Significant peptides were selected by performing a Wilcoxon analysis and then combined using an in-house developed tool in random forest models (for cIMT and LVMI) or support vector machine (SVM) model (for PWV) or to build the prognostic biomarker classifiers. For the SVM model (SVM package of R [38]), the parameters of the radial kernel function (type C) were 1 (cost parameter) and 0.04545455 (kernel width). The score taken as cut-off for prediction was 0.5 for the random forest-based classifiers and 0 for the SVM-based classifier.



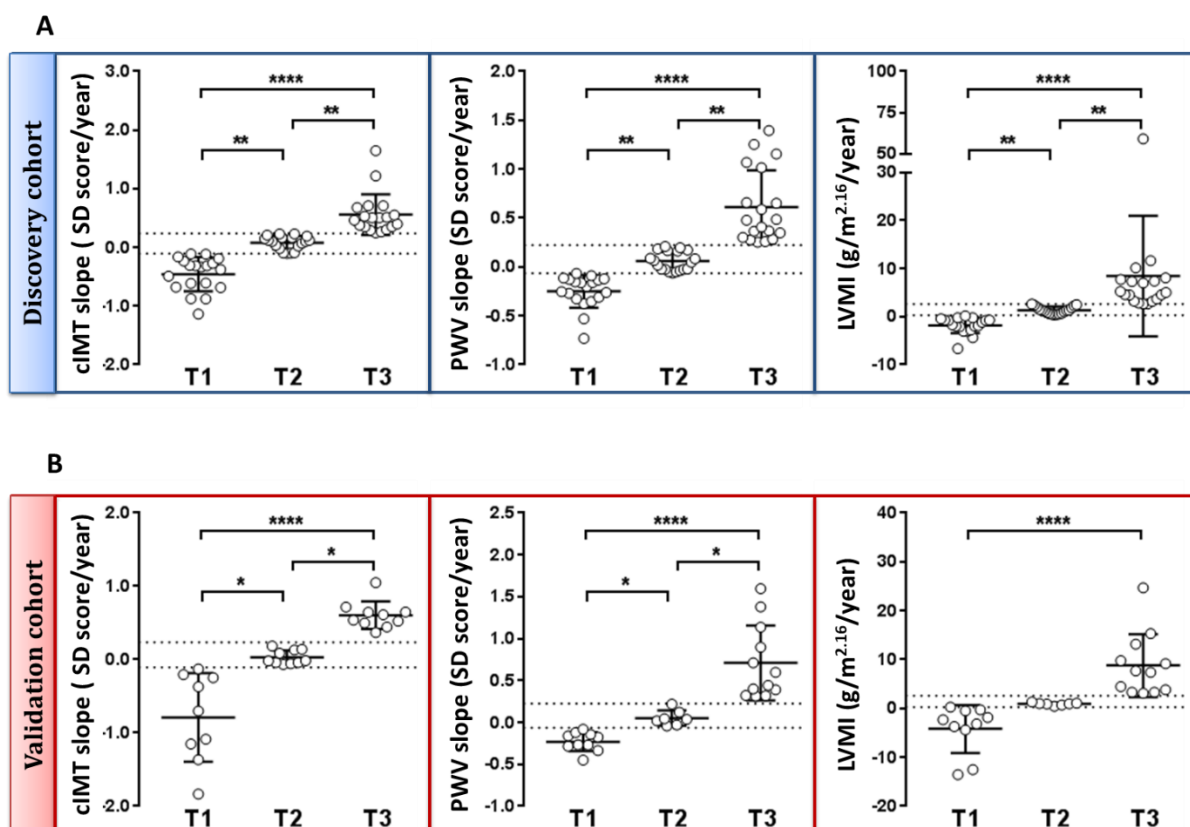
Patients in the validation cohort were then scored using the built classifiers mentioned above. Predictive performances of classifiers were evaluated calculating area under the receiver operating characteristic (ROC) curve (AUC) as well both sensitivity and specificity using GraphPad Prism 5.0 for Windows (GraphPad Software Inc).

*Comparisons:* Characteristics and classifier based-scores of patients were compared using a Mann-Whitney test or student *t*-test or a One-way ANOVA, as indicated. To assess the discriminatory ability of clinical parameters or classifiers, we tested the hypothesis that the AUC is 0.5 [39]. Statistical analyses were performed using GraphPad Software Inc.  $p < 0.05$  was considered as statistically significant.

## Results

### *Description of the study population*

For the identification of urinary peptides associated with the progression of CVD, we studied the urinary peptidome of 86 CKD children from the 4C cohort by CE-MS. The discovery cohort, consisting of 57 patients, was used for the identification of urinary peptide biomarkers and establishment of the prognostic classifiers and the validation cohort, composed of 29 patients, was used for the analysis of the performances of prognostic classifiers. cIMT, PWV and LVMI were considered as surrogate markers of CVD [23] and the rate of changes of these parameters (i.e slopes, measured over one year) were used to determine the CVD progression. For the 3 measured parameters, 3 groups of cardiovascular risks were defined (**Figure 1**): T1 was composed of children considered with low-risk for CVD progression; T2 contained patients considered with intermediate risk for CVD progression; T3 was composed of children with high-risk for CVD progression. Clinical characteristics of individual patients in both discovery and validation cohorts are presented in **Tables 1-3**.



**Figure 1:** Scatter plots showing the distribution of patients with the different CVD risks in the discovery and validation cohorts. **(A)** The discovery cohort was split into tertiles for the 3 CVD parameters, T1 defining the group of patients with low-risk for CVD progression, T2 the group with intermediate risk for CVD progression and T3 the group with high-risk for CVD progression. **(B)** The threshold values defining tertiles in the discovery cohort (dotted lines on the graph) were then applied to the validation. Data are mean  $\pm$  SD; cIMT: carotid intima-media thickness; PWV: pulse wave velocity; LVMI: left ventricular mass index. \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*\*  $P < 0.001$ , One-way ANOVA test for independent samples.

**Table 1:** Characteristics of the cIMT study population.

		Discovery phase (n=38)			Validation phase (n=19)		
		Low risk (T1)	High risk (T3)	p-value <sup>§</sup>	Low risk (T1)	High risk (T3)	p-value <sup>§</sup>
Age (years)	-	12.06 $\pm$ 2.83	11.92 $\pm$ 3.95	0.90	11.40 $\pm$ 2.85	12.35 $\pm$ 3.29	0.66
Gender	f	7	7	1.00	3	1	0.29
	m	12	12		6	9	
eGFR Baseline (ml/min/1.73m <sup>2</sup> )	-	26.97 $\pm$ 5.65	28.20 $\pm$ 6.66	0.54	27.28 $\pm$ 4.36	28.98 $\pm$ 6.60	0.49
eGFR slope (ml/min/1.73m <sup>2</sup> /year)	-	-3.22 $\pm$ 3.37	-2.93 $\pm$ 4.97	0.84	-2.88 $\pm$ 1.25	-4.87 $\pm$ 3.79	0.13
cIMT baseline (SD score)	-	2.61 $\pm$ 1.69	0.88 $\pm$ 0.86	<0.01	2.77 $\pm$ 1.05	0.11 $\pm$ 1.76	<0.01
cIMT slope (SD score/year)*	-	-0.46 $\pm$ 0.29	0.55 $\pm$ 0.35	<0.0001	-0.79 $\pm$ 0.60	0.60 $\pm$ 0.19	<0.0001

PWV baseline (SD score)	-	0.54 ± 1.68	0.01 ± 1.53	0.32	0.44 ± 0.69	0.47 ± 1.99	0.80
PWV slope (SD score/year)	-	0.10 ± 0.51	0.10 ± 0.24	0.98	0.14 ± 0.41	1.18 ± 0.55	0.89
LVMI baseline (g/m <sup>2.16</sup> )	-	41.14 ± 10.03	44.47 ± 13.98	0.40	49.94 ± 16.39	35.05 ± 6.22	<b>0.03</b>
LVMI slope (g/m <sup>2.16</sup> /year)	-	1.66 ± 3.74	4.60 ± 13.51	0.37	0.11 ± 6.10	4.56 ± 5.46	0.10
Systolic BP (SD score)	-	1.15 ± 1.37	1.23 ± 1.29	0.86	1.32 ± 1.87	0.87 ± 0.77	0.59
Diastole BP (SD score)	-	1.02 ± 1.11	1.94 ± 1.22	0.82	1.12 ± 1.55	0.63 ± 0.51	0.42

Values are means ± SD. T1: first tertile; T3: third tertile; GFR: Glomerular filtration rate; cIMT: carotid intima-media thickness; PWV: pulse wave velocity; LVMI: left ventricular mass index; BP: blood pressure. \*See also Fig. 1. §Mann-Whitney test for independent samples.

**Table 2:** Characteristics of the PWV study population.

		Discovery phase (n=38)			Validation phase (n=22)		
		Low risk (T1)	High risk (T3)	p-value <sup>§</sup>	Low risk (T1)	High risk (T3)	p-value <sup>§</sup>
Age (years)	-	-1.94 ± 3.43	-3.01 ± 2.50	0.28	10.40 ± 3.28	13.64 ± 2.34	<b>0.01</b>
Gender	f	9	10	0.75	4	1	0.08
	m	10	9		6	11	
eGFR Baseline (ml/min/1.73m <sup>2</sup> )	-	27.52 ± 6.14	24.43 ± 5.00	0.09	25.98 ± 3.96	30.81 ± 5.27	0.07
eGFR slope (ml/min/1.73m <sup>2</sup> /year)	-	-1.94 ± 3.44	-3.01 ± 2.50	0.28	-2.94 ± 2.55	-6.39 ± 4.10	0.127
cIMT baseline (SD score)	-	1.75 ± 1.91	1.47 ± 1.26	0.60	1.48 ± 1.71	1.64 ± 1.18	0.67
cIMT slope (SD score/year)	-	-0.01 ± 0.55	0.01 ± 0.41	0.9	-0.03 ± 0.64	-0.20 ± 0.78	0.37
PWV baseline (SD score)	-	1.30 ± 1.60	-0.83 ± 1.26	<b>&lt;0.01</b>	1.16 ± 1.58	-0.53 ± 0.90	<b>0.02</b>
PWV slope (SD score/year) *	-	-0.25 ± 1.67	0.61 ± 0.37	<b>&lt;0.0001</b>	-0.23 ± 0.11	0.71 ± 0.45	<b>&lt;0.0001</b>
LVMI baseline (g/m <sup>2.16</sup> )	-	44.76 ± 11.72	40.61 ± 10.03	0.25	42.10 ± 12.63	42.87 ± 13.07	0.86
LVMI slope (g/m <sup>2.16</sup> /year)	-	0.82 ± 3.79	2.52 ± 2.61	0.12	2.56 ± 3.95	3.14 ± 9.92	<b>&lt;0.01</b>
Systolic BP (SD score)	-	1.77 ± 1.46	0.66 ± 1.08	<b>0.01</b>	0.79 ± 1.06	0.67 ± 0.83	0.43
Diastole BP (SD score)	-	1.41 ± 1.12	0.46 ± 0.79	<b>&lt;0.01</b>	0.54 ± 0.76	0.63 ± 0.85	0.93

Values are means ± SD. GFR: T1: first tertile; T3: third tertile; Glomerular filtration rate; cIMT: carotid intima-media thickness; PWV: pulse wave velocity; LVMI: left ventricular mass index; BP: blood pressure. \*See also Fig. 1. §Mann-Whitney test for independent samples.

**Table 3:** Characteristics of the LVMI study population.

		Discovery phase (n=38)			Validation phase (n=22)		
		low risk (T1)	high risk (T3)	p-value <sup>§</sup>	low risk (T1)	high risk (T3)	p-value <sup>§</sup>
Age (years)	-	11.78 ± 2.92	12.30 ± 3.40	0.46	12.46 ± 3.87	12.27 ± 3.76	0.91
Gender	f	10	10	1,00	1	2	0.66
	m	9	9		9	10	
eGFR Baseline (ml/min/1.73m <sup>2</sup> )	-	27.81 ± 7.30	28.60 ± 5.10	0.75	27.45 ± 5.31	27.17 ± 4.66	0.9

<b>eGFR slope (ml/min/1.73m<sup>2</sup>/year)</b>	-	-1.56 ± 3.70	-4.17 ± 5.38	0.09	-4.91 ± 5.68	-4.42 ± 2.49	0.81
<b>cIMT baseline (SD score)</b>	-	1.59 ± 1.47	1.21 ± 1.12	0.39	1.63 ± 1.24	1.34 ± 1.20	0.59
<b>cIMT slope (SD score/year)</b>	-	0.01 ± 0.45	0.19 ± 0.53	0.25	-0.21 ± 0.76	-0.02 ± 0.64	0.55
<b>PWV baseline (SD score)</b>	-	0.48 ± 1.43	0.31 ± 1.81	0.39	-0.35 ± 0.75	0.02 ± 1.14	0.36
<b>PWV slope (SD score/year)</b>	-	0.05 ± 0.34	0.13 ± 0.43	0.54	0.43 ± 0.55	0.16 ± 0.57	0.27
<b>LVMI baseline (g/m<sup>2.16</sup>)</b>	-	45.98 ± 10.97	42.04 ± 15.33	0.37	50.94 ± 13.49	43.60 ± 12.68	0.21
<b>LVMI slope (g/m<sup>2.16</sup>/year) *</b>	-	-1.83 ± 1.66	8.42 ± 12.51	<b>&lt;0.0001</b>	-4.19 ± 4.89	8.77 ± 6.41	<b>&lt;0.0001</b>
<b>Systolic BP (SD score)</b>	-	1.54 ± 1.56	1.04 ± 1.41	0.3	0.93 ± 0.71	0.94 ± 1.97	0.99
<b>Diastole BP (SD score)</b>	-	1.23 ± 1.23	0.85 ± 1.23	0.34	0.64 ± 0.81	0.61 ± 1.45	0.96

Values are means ± SD. T1: first tertile; T3: third tertile; GFR: Glomerular filtration rate; cIMT: carotid intima-media thickness; PWV: pulse wave velocity; LVMI: left ventricular mass index; BP: blood pressure. \*See also Fig. 1. \*Mann-Whitney test for independent samples.

### ***Identification of urinary peptides associated to progression of CVD and development of predictive classifiers***

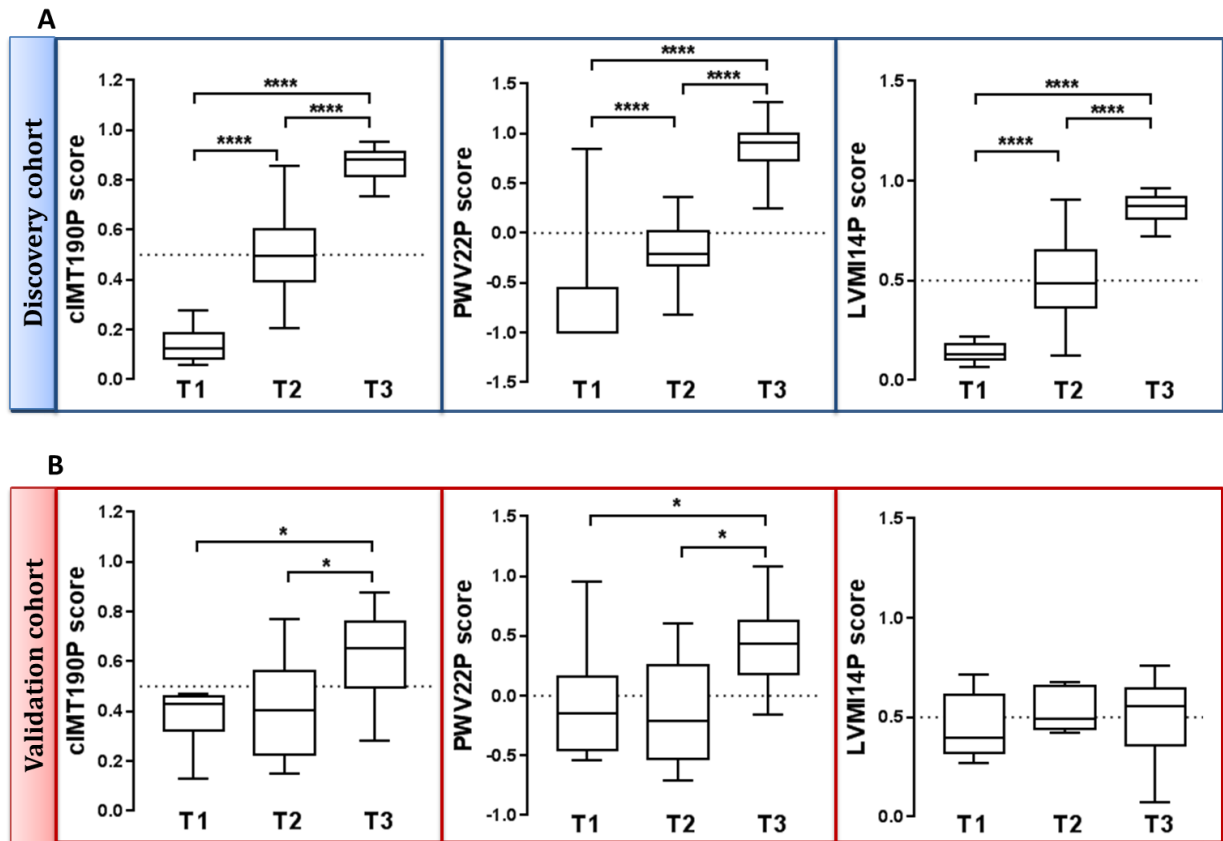
Analysis of the urinary peptidome of the discovery sets by CE-MS led to the identification of a total of 7586 urinary peptides. Comparison of the abundance of the urinary peptides between T1 (low-risk of CVD progression) and T3 (high-risk of CVD progression) groups for the 3 cardiovascular parameters led to the discovery of 190, 22 and 14 urinary peptides associated to cIMT, PWV and LVMI progression, respectively. None of these peptides resisted to correction for multiple testing. We therefore continued the analysis with these non-adjusted peptides.

These identified peptides were then combined in either a random forest or SVM prediction model and were optimized for the classification of patients in this discovery set (**Figure 2A**). This resulted in three models: the cIMT190P model for cIMT slope prediction, the PWV22P model for PWV slope prediction and the LVMI14P model for LVMI slope prediction. Scoring the patients with each of these classifiers clearly and significantly separated the 3 groups according to the severity (low-, intermediate- or high-risk) of CVD damage.

### ***Validation of the predictive classifiers and evaluation of predictive performances in an independent cohort***

The 3 models were subsequently tested on a new set of patients (validation cohort) not previously used for designing the classifiers (**Figure 2B**). After CE-MS analysis of their urinary peptidome, these patients were scored blindly with cIMT190P, PWV22P and LVMI14P classifiers. Interestingly the distribution of both cIMT190P and PWV22P classifiers showed significant separation of the patients at low (T1) and high (T3) risk for developing CVD. The cIMT190P classifier predicted cIMT progression with a sensitivity of 80% [95% confidence interval (CI), 44 to 97], a specificity of 100% (95% CI, 66 to 100) and an AUC of 0.87 (95% CI, 0.68 to 1.00) (**Figure 3A**). Although they failed to distinguish the

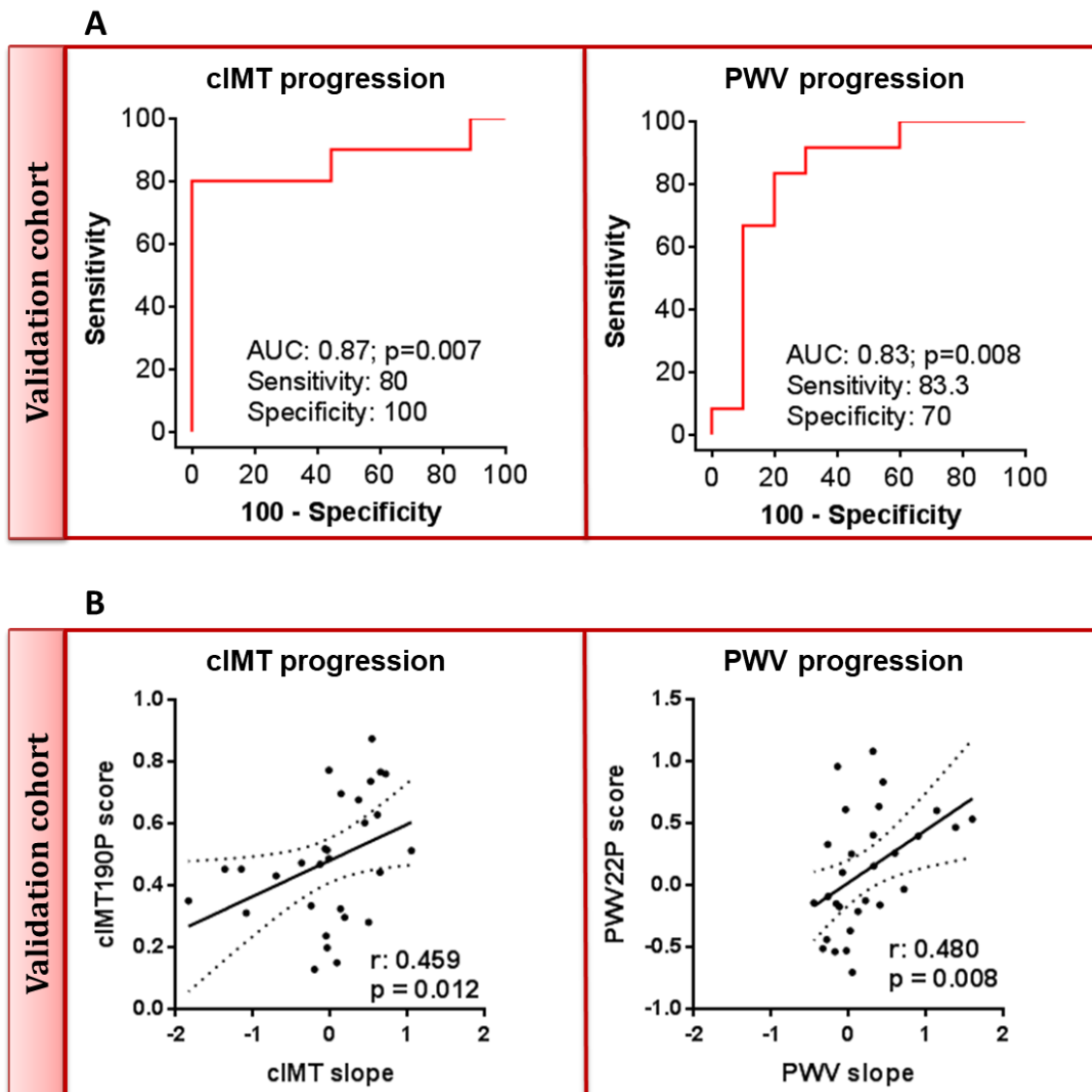
individuals with intermediate risk of CVD progression (T2) from T1, the cIMT190P-based scores allowed to separate T2 from T3 (**Figures 2B**) and correlated with cIMT slope ( $r=0.459$ ;  $p = 0.012$ ) (**Figure 3B**).



**Figure 2:** Urinary peptide-based scores of CVD risk in the discovery and validation cohorts. **(A)** Box-and-whisker plots showing the distribution of scores for cIMT190P model, PWV22P model and LVMI14P model in the discovery cohort for patients with low-risk (T1), intermediate risk (T2) and high-risk for CVD progression (T3). **(B)** Validation of the classifiers in an independent, blinded cohort. The dotted lines represent the cut-off scores for prediction. cIMT: carotid intima-media thickness; PWV: pulse wave velocity; LVMI: left ventricular mass index. \*  $P < 0.05$ , \*\*\*\*  $P < 0.001$ , One-way ANOVA for independent samples.

The PWV22P classifier predicted PWV progression with a sensitivity of 83% [95% CI, 52 to 98], a specificity of 70% (95% CI, 34.8 to 93.3) and an AUC of 0.83 (95% CI, 0.64 to 1.00) (**Figure 3A**). Like the prediction of progression of cIMT slope, the PWV22P model also clearly distinguished T2 from T3 patients, but not T1 from T2 (**Figure 2B**) and the PWV22P-based scores displayed a significant correlation to PWV slope ( $r=0.480$ ;  $p = 0.008$ ) (**Figure 3B**). In contrast, the LVMI model (LVMI14P) was not been validated (AUC: 0.70 (95% CI, 0.42 to 0.89),  $p>0.05$ ) (**Figure 2B** and data not shown). Taken together, these data support the existence of urinary peptides allowing to predict future

development of several CVD such as vascular remodeling (atherosclerosis [related to cIMT] and arterial stiffness [related to PWV]) in a CKD background.



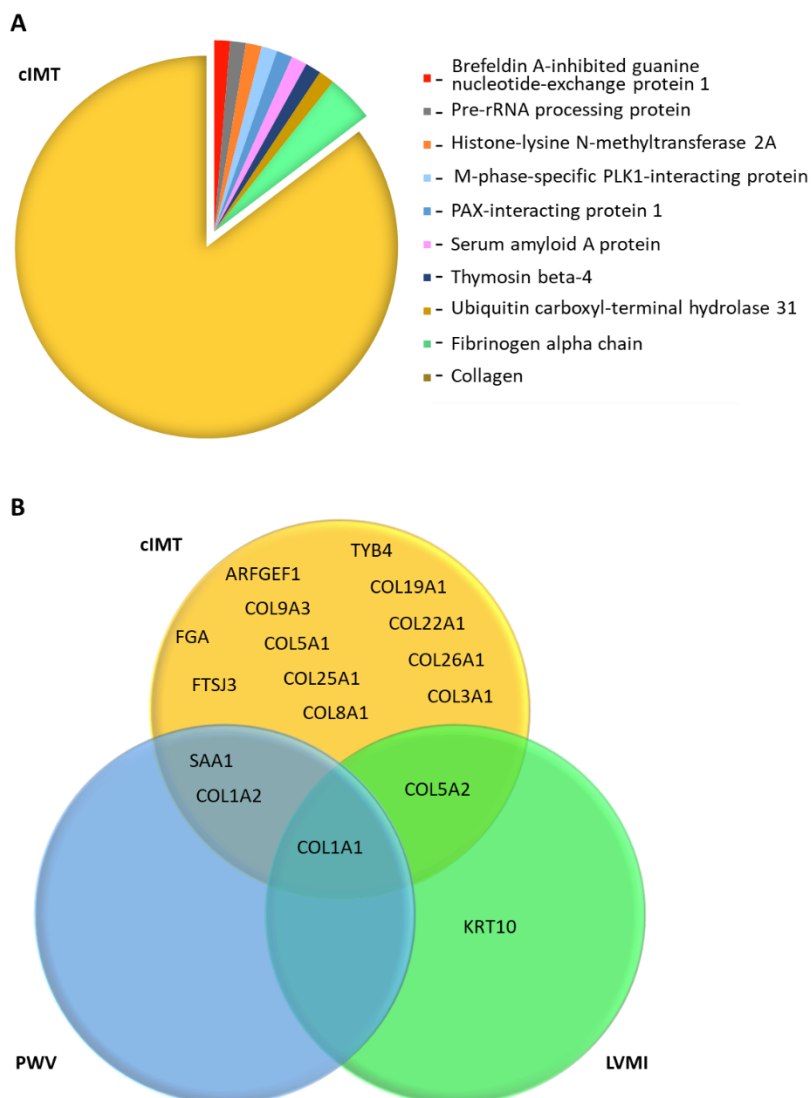
**Figure 3.** Efficiency of urinary peptides in the prediction of cIMT and PWV progression. **(A)** ROC curves for cIMT190P and PWV22P models. **(B)** Correlation between classifier-based scores and the rates of changes of CVD parameters. cIMT: carotid intima-media thickness; PWV: pulse wave velocity; LVMI: left ventricular mass index.

### *Identity of urinary peptides associated to CVD progression*

Tandem mass spectrometry was applied to sequence the peptides predictive for the CVD complications associated to CKD. Among peptides included in the cIMP190P, PWV22P and LVMI14P models, 75 peptides, 9 peptides and 6 peptides were identified by sequencing in each classifier, respectively (**Supplementary tables 2-4**). Regarding the signature of cIMT progression, most of sequenced peptides

are fragments of various collagens (85%), a major constituent of the extracellular matrix; other peptides included fragments of keratin, brefeldin A-inhibited guanine nucleotide-exchange protein 1, pre-rRNA processing protein, histone-lysine N-methyltransferase 2A, M-phase-specific PLK1-interacting protein, PAX-interacting protein 1, serum amyloid A protein, thymosin-β4, ubiquitin carboxyl-terminal hydrolase 31 and fibrinogen alpha chain and (**Figure 4A**). 2/3 of these cIMT-related urinary peptides were less abundant in patients with high-risk to develop CVD.

Next, we compared proteins from which the predictive peptides for cIMT, PWV and LVMI slope originated to find any overlap. Interestingly, although most of the peptide sequences were not the same, we only found collagen alpha-1(I) (COL1A1) as common protein to the 3 analyzed CVD parameters (**Figure 4B**). Among fragments derived from the collagen alpha-1(I), 23/35, 5/7 and 2/4 present in the cIMT190P, PWV22P and LVMI14P classifiers, respectively, were downregulated in patients with high-risk CVD. Peptide sequences which were either more or less abundant, span the entire protein sequence.



**Figure 4.** Protein origin of urinary peptides associated to progression of CVD in CKD children. **(A)** Signature of cIMT progression and **(B)** overlap of protein of origin of the sequenced peptides associated to progression of CVD in CKD children. Brefeldin A-inhibited guanine nucleotide-exchange protein 1 – ARFGEF1; Collagen alpha-1(I) chain – COL1A1; Collagen alpha-2(I) chain – COL1A2; Collagen alpha-1(III) chain – COL3A1; Collagen alpha-1(V) chain – COL5A1; Collagen alpha-2(V) chain – COL5A2; Collagen alpha-1(VIII) chain – COL8A1; Collagen alpha-3(IX) chain – COL9A3; Collagen alpha-1(XIX) chain – COL19A1; Collagen alpha-1(XXII) chain – COL22A1; Collagen alpha-1(XXV) chain – COL25A1; Collagen alpha-1(XXVI) chain – COL26A1; Fibrinogen alpha chain – FGA; pre-rRNA processing protein - FTSJ3; Keratin; type I cytoskeletal 10 – KRT10; Serum amyloid A protein – SAA1; Thymosin beta-4 – TYB4.

Collagen alpha-2(I) (COL1A2) and serum amyloid A protein (SAA1) were the only proteins overlapping between the cIMT190P and PWV22P models: most of fragments originating from collagen alpha-2(I) were downregulated with cIMT progression (12/14), whereas the fragments included in the PWV22P model were upregulated by CVD. In addition, the abundance of the unique sequence derived from serum amyloid A protein was decreased in cIMT190P but increased in the PWV22P model. Collagen alpha-2(V) chain (COL5A2) was the only common protein between cIMT190P and LVMI14P being less abundant in high-risk patients in both models. Moreover, fragments from keratin; type I cytoskeletal 10 (KRT10) was less abundant in high-risk patients for LVMI and was unique to this CVD parameter.

Finally, we evaluated the potential overlap of our peptides with other urinary peptides panels associated to CVD complications/events in adults on a non CKD background: CAD238 as classifier for the diagnosis and prognosis of coronary artery disease [16, 17], HFP which predicts heart failure [18, 40] and ACSP75 for the prediction of acute coronary syndromes [20]. Interestingly, several sequenced peptides present in the cIMP190P model, PWV22P and LVMI14P models were common with those included in the previously published CVD classifiers (**Table 4**).

**Table 4.** Overlap of sequences from cIMT190P, PWV22P and LVMI14P classifiers with peptides previously identified in classifiers predicting cardiovascular events.

Sequence	Classifier	Protein name	Previously published classifiers	Cardiovascular event
SpGPDGKTGPpGP	cIMT190P	Collagen alpha-1(I) chain	CAD238 [16]	coronary artery disease
SpGSPGPDGKTGPpGP	cIMT190P	Collagen alpha-1(I) chain		
TGSpGSpGPDGKTGPPGpAG	cIMT190P	Collagen alpha-1(I) chain		



EGSpGRDGSpGAKGDRGETGP	cIMT190P	Collagen alpha-1(I) chain		
GKNGDDGEAGKpGRpGERGPpGP	cIMT190P	Collagen alpha-1(I) chain		
GKNGDDGEAGKpGRpGERGPpGP	cIMT190P	Collagen alpha-1(I) chain		
LDGAKGDAGPAGPKGEpGSpGENGApG	cIMT190P	Collagen alpha-1(I) chain		
SpGRDGSpGAKGDRGETGP	LVMII14P	Collagen alpha-1(I) chain		
ADGQpGAKGEpGDAGAKGDA GPpGP	PWV22P	Collagen alpha-1(I) chain		
VGPpGpPGPPGPPGPPS	cIMT190P	Collagen alpha -1(I) chain	HFP [40]	Incident heart failure
VGPpGPpGPpGPPGPPS	cIMT190P	Collagen alpha -1(I) chain		
LKGQpGApGVKGEpGApGENGTPGQTGARG	cIMT190P	Collagen alpha -2(I) chain		
SpGERGETGPpGP	cIMT190P	Collagen alpha -1(III) chain		
DEAGSEADHEGTHSTKRGHAKSRPV	cIMT190P	Fibrinogen alpha chain		
KGNSGEPGApGSKGDTGAKGEPGPVG	PWV22P	Collagen alpha -1(I) chain		
VGPpGPpGPpGPPGPPS	cIMT190P	Collagen alpha-1(I) chain	ACSP75 [20]	Acute coronary syndromes
VGPpGPpGPpGpPGPPS	cIMT190P	Collagen alpha-1(I) chain		
PpGPpGPpGpPGPPS	cIMT190P	Collagen alpha-1(I) chain		
SpGERGETGPp	cIMT190P	Collagen alpha-1(III) chain		
SpGERGETGPpGP	cIMT190P	Collagen alpha-1(III) chain		
ADGQpGAKGEpGDAGAKGDA GPpGPA	LVMII14P	Collagen alpha-1(I) chain		
ADGQpGAKGEpGDAGAKGDA GPpGP	PWV22P	Collagen alpha-1(I) chain		

## Discussion

CVD is the leading complication in CKD in children. To improve personalized risk assessment and clinical management of these patients, novel strategies need to be found to detect children the most at risk of cardiovascular lesion progression as earliest as possible. Several panels of urinary peptide markers of cardiovascular complications and events have already been established as diagnostic/prognostic tools in the adult non-CKD population [16, 17, 19, 20, 41]. Here, we evaluated the potential of CE-MS urinary peptidomics analysis to predict the progression of CVD complications associated to CKD in children after 1-year follow-up. In order to monitor the progression of the cardiac lesions, we assessed the slope of three surrogate markers of CVD: cIMT, PWV, and LVMI for atherosclerosis, vascular stiffness and left ventricular hypertrophy respectively. The cohort was split into tertiles of progression for each of the CVD parameter.

We found 190 and 22 urinary peptides associated with atherosclerosis and vascular stiffness, respectively. Combination of those peptides into models (cIMT190P and PWV22P) allowed to predict with high sensitivities and specificities, as well as significant AUCs, rapid *versus* slow progression of both cIMT and PWV. To our knowledge, this study is the first to identify robust urinary biomarkers of CVD in children having a CKD background. The use of both classifiers can provide a significant improvement of the clinical management of CKD children, since early therapeutic treatment to delay the occurrence of a cardiovascular event could be proposed in the patients predicted in the high-risk stratum.

Although they can be used for predicting patients having a high or a low risk to develop CVD, both cIMT190P and PWV22P classifiers were less performant to detect patients with intermediate risk. Nevertheless, we observed a significant positive correlation between the classifier-based scores and the rates of changes of CVD markers (i.e cIMT slopes and PWV slopes), thereby suggesting that maybe the scores should not be interpreted in a dichotomous manner but rather as a continuum without threshold which increases when CVD risks increases.

We also discovered a panel of 14 urinary peptides associated with LWMI. However, these markers did not display efficiency for early screening of pediatric CKD patients with increased risk to develop cardiac hypertrophy. This discrepancy between predictive performances of cIMT190P and PWV22P one side and those of LVMI14P on the other side might be because early molecular modifications preceding vascular structural/functional changes occur earlier or faster compared to those responsible for cardiac modifications in CKD children. A longer follow-up might be necessary to identify urinary peptide biomarker predictive of the progression of left ventricular remodeling.

Among the urinary CVD associated-biomarkers discovered in this study, a total of 90 peptides were sequenced which could relate to key molecular components of pediatric-CKD dependent CVD. Most of the identified peptides originated from collagen. These observations were in accordance with previously described CVD signatures associated to coronary artery disease [16, 17, 41], heart failure [18, 19, 40] and acute coronary syndrome [20]. They strengthen the role of the remodeling of extracellular matrix as a common and important mechanism in CVD process, including also atherosclerosis, vascular stiffness and left ventricular hypertrophy induced by CKD in children.

The alpha 1 chain of type I collagen was the protein represented by the highest number of peptides and is the only common protein of the 3 CVD classifiers. Collagen type I is a major extracellular matrix constituent of the myocardium and the arterial vascular wall, synthesized by fibroblasts in the myocardium and smooth muscle cells in the medial arterial vascular wall [42]. The collagen turnover is controlled by proteinases such as matrix metalloproteinases and the tight control on the enzyme activities is altered during heart failure and atherosclerosis reducing the number or the type of generated fragments [43]. However, the focus here on urinary peptides failed to assess the early underlying molecular changes of connective tissue turnover in CKD-induced CVD. Indeed, both up and down regulations of collagen type I alpha 1 chain, without regionalization on the protein of origin, were observed in patients rapidly progressing toward CVD. In the same order, 16 fragments of collagen present in the cIMT classifier have a C-terminal GxPGP motif, a substrate of MMP2, membrane Pro-x carboxypeptidase and prolyl oligopeptidase [16], but 11/16 were down-regulated while 5/16 were up-regulated in the high risk group.

Interestingly, the cIMT classifier included several peptides not related to collagen (15%) such as thymosin beta 4 or serum amyloid A1 protein (downregulated in patients with fast progressing cIMT) as well as pre-rRNA processing protein (up regulated). Thymosin beta 4 is known for its implication in mechanisms of repair. This protein is able to inhibit oxidative stress and myocardial cell death during cardiac injury [44, 45], and its administration in vivo contributes to vascular regeneration and cardiac repair [46]. Serum amyloid A1 is an apolipoprotein highly expressed in plasma in response to inflammatory stimulation [47], which promotes the development of atherosclerosis [48, 49]. The protein accumulates in HDL particles of patients with CAD and/or CKD [50, 51] where it reverses the anti-inflammatory and vaso-protective properties of HDL [51].

The present study however has some limitations. Peptides associated to the progression of cIMT, PWV and LVMI were identified based on their differential urinary excretion (Wilcoxon test) between children progressing rapidly - versus slowly - toward CVD. However, none of these peptides resisted to correction for multiple testing. Pediatric CKD patients included in this study displayed a broad range of CVD status at enrollment, thereby potentially explaining urinary peptidome heterogeneity. Refinement of the prognosis study in a much larger cohort including children balanced with respect to baseline CVD status must be undertaken to reevaluate the predictive value of urinary peptides for CVD progression.

## Conclusion

In conclusion, the present study showed the potential of CE-MS urinary peptidomics for identification of biomarkers able to predict cardiovascular complications such as atherosclerosis and vascular stiffness in children with CKD. Used in clinic, such biomarkers will allow non-invasive screening of CKD children having a high risk to develop CVD. These will improve management of the disease, without adding traumatic examination for these children already undergoing repetitive medical examinations. Further studies should be carried out for discovering biomarkers predictive of left ventricular hypertrophy, a cardiac disease frequently observed in children with CKD.

## Bibliography – Study 1

1. Eckardt, K.U., et al., *Evolving importance of kidney disease: from subspecialty to global health burden*. Lancet, 2013. **382**(9887): p. 158-69.
2. Chesnaye, N.C., et al., *Mortality risk disparities in children receiving chronic renal replacement therapy for the treatment of end-stage renal disease across Europe: an ESPN-ERA/EDTA registry analysis*. Lancet, 2017. **389**(10084): p. 2128-2137.
3. McDonald, S.P., et al., *Long-term survival of children with end-stage renal disease*. N Engl J Med, 2004. **350**(26): p. 2654-62.
4. Mitsnefes, M.M., *Cardiovascular disease in children with chronic kidney disease*. J Am Soc Nephrol, 2012. **23**(4): p. 578-85.
5. Tonelli, M., et al., *Risk of coronary events in people with chronic kidney disease compared with those with diabetes: a population-level cohort study*. Lancet, 2012. **380**(9844): p. 807-14.
6. Herzog, C.A., et al., *Cardiovascular disease in chronic kidney disease. A clinical update from Kidney Disease: Improving Global Outcomes (KDIGO)*. Kidney Int, 2011. **80**(6): p. 572-86.
7. Chronic Kidney Disease Prognosis, C., et al., *Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis*. Lancet, 2010. **375**(9731): p. 2073-81.
8. Furth, S.L., et al., *Metabolic abnormalities, cardiovascular disease risk factors, and GFR decline in children with chronic kidney disease*. Clin J Am Soc Nephrol, 2011. **6**(9): p. 2132-40.
9. Mitsnefes, M.M., et al., *Changes in left ventricular mass in children and adolescents during chronic dialysis*. Pediatr Nephrol, 2001. **16**(4): p. 318-23.
10. Ros, J.A., et al., *[Acute pericarditis: electrocardiographic review]*. Rev Clin Esp, 1975. **137**(1): p. 53-60.
11. Mitsnefes, M.M., et al., *Severe cardiac hypertrophy and long-term dialysis: the Midwest Pediatric Nephrology Consortium study*. Pediatr Nephrol, 2006. **21**(8): p. 1167-70.
12. Groothoff, J.W., et al., *Increased arterial stiffness in young adults with end-stage renal disease since childhood*. J Am Soc Nephrol, 2002. **13**(12): p. 2953-61.
13. Goodman, W.G., et al., *Coronary-artery calcification in young adults with end-stage renal disease who are undergoing dialysis*. N Engl J Med, 2000. **342**(20): p. 1478-83.
14. Schanstra, J.P., et al., *Diagnosis and Prediction of CKD Progression by Assessment of Urinary Peptides*. J Am Soc Nephrol, 2015. **26**(8): p. 1999-2010.
15. Decramer, S., et al., *Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis*. Nat Med, 2006. **12**(4): p. 398-400.
16. Delles, C., et al., *Urinary proteomic diagnosis of coronary artery disease: identification and clinical validation in 623 individuals*. J Hypertens, 2010. **28**(11): p. 2316-22.

17. Brown, C.E., et al., *Urinary proteomic biomarkers to predict cardiovascular events*. *Proteomics Clin Appl*, 2015. **9**(5-6): p. 610-7.
18. Rossing, K., et al., *Urinary Proteomics Pilot Study for Biomarker Discovery and Diagnosis in Heart Failure with Reduced Ejection Fraction*. *PLoS One*, 2016. **11**(6): p. e0157167.
19. Kuznetsova, T., et al., *Urinary proteome analysis in hypertensive patients with left ventricular diastolic dysfunction*. *Eur Heart J*, 2012. **33**(18): p. 2342-50.
20. Htun, N.M., et al., *Prediction of acute coronary syndromes by urinary proteome analysis*. *PLoS One*, 2017. **12**(3): p. e0172036.
21. Querfeld, U., et al., *The Cardiovascular Comorbidity in Children with Chronic Kidney Disease (4C) study: objectives, design, and methodology*. *Clin J Am Soc Nephrol*, 2010. **5**(9): p. 1642-8.
22. Schwartz, G.J., et al., *New equations to estimate GFR in children with CKD*. *J Am Soc Nephrol*, 2009. **20**(3): p. 629-37.
23. Guerin, A.P., et al., *Assessment and significance of arterial stiffness in patients with chronic kidney disease*. *Curr Opin Nephrol Hypertens*, 2008. **17**(6): p. 635-41.
24. Devereux, R.B., et al., *Echocardiographic assessment of left ventricular hypertrophy: comparison to necropsy findings*. *Am J Cardiol*, 1986. **57**(6): p. 450-8.
25. Chinali, M., et al., *Left Ventricular Mass Indexing in Infants, Children, and Adolescents: A Simplified Approach for the Identification of Left Ventricular Hypertrophy in Clinical Practice*. *J Pediatr*, 2016. **170**: p. 193-8.
26. Khoury, P.R., et al., *Age-specific reference intervals for indexed left ventricular mass in children*. *J Am Soc Echocardiogr*, 2009. **22**(6): p. 709-14.
27. Touboul, P.J., et al., *Mannheim carotid intima-media thickness consensus (2004-2006). An update on behalf of the Advisory Board of the 3rd and 4th Watching the Risk Symposium, 13th and 15th European Stroke Conferences, Mannheim, Germany, 2004, and Brussels, Belgium, 2006*. *Cerebrovasc Dis*, 2007. **23**(1): p. 75-80.
28. Doyon, A., et al., *Carotid artery intima-media thickness and distensibility in children and adolescents: reference values and role of body dimensions*. *Hypertension*, 2013. **62**(3): p. 550-6.
29. Kracht, D., et al., *Validating a new oscillometric device for aortic pulse wave velocity measurements in children and adolescents*. *Am J Hypertens*, 2011. **24**(12): p. 1294-9.
30. Thurn, D., et al., *Aortic Pulse Wave Velocity in Healthy Children and Adolescents: Reference Values for the Vicorder Device and Modifying Factors*. *Am J Hypertens*, 2015. **28**(12): p. 1480-8.
31. Theodorescu, D., et al., *Discovery and validation of urinary biomarkers for prostate cancer*. *Proteomics Clin Appl*, 2008. **2**(4): p. 556-570.
32. Mischak, H., A. Vlahou, and J.P. Ioannidis, *Technical aspects and inter-laboratory variability in native peptide profiling: the CE-MS experience*. *Clin Biochem*, 2013. **46**(6): p. 432-43.
33. Neuhoff, N., et al., *Mass spectrometry for the detection of differentially expressed proteins: a comparison of surface-enhanced laser desorption/ionization and capillary electrophoresis/mass spectrometry*. *Rapid Commun Mass Spectrom*, 2004. **18**(2): p. 149-56.
34. Dissard, R., et al., *Long term metabolic syndrome induced by a high fat high fructose diet leads to minimal renal injury in C57BL/6 mice*. *PLoS One*, 2013. **8**(10): p. e76703.
35. Magalhaes, P., et al., *Comparison of Urine and Plasma Peptidome Indicates Selectivity in Renal Peptide Handling*. *Proteomics Clin Appl*, 2018: p. e1700163.
36. Pejchinovski, M., et al., *Comparison of higher energy collisional dissociation and collision-induced dissociation MS/MS sequencing methods for identification of naturally occurring peptides in human urine*. *Proteomics Clin Appl*, 2015. **9**(5-6): p. 531-42.
37. Zurbig, P., et al., *Biomarker discovery by CE-MS enables sequence analysis via MS/MS with platform-independent separation*. *Electrophoresis*, 2006. **27**(11): p. 2111-25.
38. Meyer, D., et al., *Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*. 2015, <http://CRAN.R-project.org/package=e1071>. p. R package version 16-7.
39. Hanley, J.A. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve*. *Radiology*, 1982. **143**(1): p. 29-36.

40. Zhang, Z.Y., et al., *Novel Urinary Peptidomic Classifier Predicts Incident Heart Failure*. J Am Heart Assoc, 2017. **6**(8).
41. von Zur Muhlen, C., et al., *Evaluation of urine proteome pattern analysis for its potential to reflect coronary artery atherosclerosis in symptomatic patients*. J Proteome Res, 2009. **8**(1): p. 335-45.
42. Myllyharju, J. and K.I. Kivirikko, *Collagens and collagen-related diseases*. Ann Med, 2001. **33**(1): p. 7-21.
43. Heeneman, S., et al., *The dynamic extracellular matrix: intervention strategies during heart failure and atherosclerosis*. J Pathol, 2003. **200**(4): p. 516-25.
44. Kumar, S. and S. Gupta, *Thymosin beta 4 prevents oxidative stress by targeting antioxidant and anti-apoptotic genes in cardiac fibroblasts*. PLoS One, 2011. **6**(10): p. e26912.
45. Wei, C., et al., *Thymosin beta 4 protects cardiomyocytes from oxidative stress by targeting anti-oxidative enzymes and anti-apoptotic genes*. PLoS One, 2012. **7**(8): p. e42586.
46. Shrivastava, S., et al., *Thymosin beta4 and cardiac repair*. Ann N Y Acad Sci, 2010. **1194**: p. 87-96.
47. Urieli-Shoval, S., R.P. Linke, and Y. Matzner, *Expression and function of serum amyloid A, a major acute-phase protein, in normal and disease states*. Curr Opin Hematol, 2000. **7**(1): p. 64-9.
48. Getz, G.S., P.A. Krishack, and C.A. Reardon, *Serum amyloid A and atherosclerosis*. Curr Opin Lipidol, 2016. **27**(5): p. 531-5.
49. Thompson, J.C., et al., *A brief elevation of serum amyloid A is sufficient to increase atherosclerosis*. J Lipid Res, 2015. **56**(2): p. 286-93.
50. Zewinger, S., et al., *Serum amyloid A: high-density lipoproteins interaction and cardiovascular risk*. Eur Heart J, 2015. **36**(43): p. 3007-16.
51. Weichhart, T., et al., *Serum amyloid A in uremic HDL promotes inflammation*. J Am Soc Nephrol, 2012. **23**(5): p. 934-47.

## SUPPLEMENTARY DATA

**Table 1.** Seventy-five sequenced peptides from the list of 190 peptides combined in the 190P model for cIMT slope prediction.

Peptide ID	Sequence	Prot_Symbol	Prot_Name	Prot_Accession	Start AA	Stop AA	Fold Change T3/T1
12495	DGSDSENIQANGIP	ARFGE F1	Brefeldin A-inhibited guanine nucleotide-exchange protein 1	Q9Y6D6	363	376	0.586
61665	GPEGPSGKpGINGKD GIPGAQGIImGKpGD RGpKGERGDQGIP	COL19 A1	Collagen alpha-1(XIX) chain	Q14993	917	958	0.774
12646	GVNVpSYPGpPGPPG	COL19 A1	Collagen alpha-1(XIX) chain	Q14993	834	848	0.816
40610	GPpGPKGNSGEPGAp GSKGDTGAKGEPGP VG	COL1A 1	Collagen alpha-1(I) chain	P02452	425	455	0.176
45333	LDGAKGDAGpAGPK GEPGSpGENGAPGQ mGPRG	COL1A 1	Collagen alpha-1(I) chain	P02452	273	305	0.391
4181	pGPPGpPGppGP	COL1A 1	Collagen alpha-1(I) chain	P02452	141	152	0.458
20736	GAAGEpGKAGERGV pGpPGA	COL1A 1	Collagen alpha-1(I) chain	P02452	587	606	0.481
58619	PGLPGPSGEPGKQGp SGASGERGPPGPMG PPGLAGppGESGR	COL1A 1	Collagen alpha-1(I) chain	P02452	973	1014	0.503
48602	pGPpGPPGLGpNFAP QLSYGYDEKSTGGIS VPG	COL1A 1	Collagen alpha-1(I) chain	P02452	147	179	0.623
4102	GPpGSAGAPGKDG	COL1A 1	Collagen alpha-1(I) chain	P02452	1142	1154	0.669

<b>10513</b>	PpGPPGpPpGPPGPPS	COL1A 1	Collagen alpha-1(I) chain	P02452	1179	1193	0.677
<b>14180</b>	VGpPGPPGpPGPPGP PS	COL1A 1	Collagen alpha-1(I) chain	P02452	1177	1193	0.792
<b>63758</b>	ARGNDGATGAAGpP GPTGPAGPPGFPGA VGAKGEAGpQGpRG SEGpQG	COL1A 1	Collagen alpha-1(I) chain	P02452	321	368	0.802
<b>44522</b>	VMGFpGPKGAAGEp GKAGERGVPGppGA VGPAAG	COL1A 1	Collagen alpha-1(I) chain	P02452	579	611	0.882
<b>44263</b>	NVGApGAKGARGs AGpPGATGFpGAAG RVGPpGP	COL1A 1	Collagen alpha-1(I) chain	P02452	855	888	0.904
<b>21158</b>	GEpGSpGENGApGQ MGPRG	COL1A 1	Collagen alpha-1(I) chain	P02452	287	305	0.930
<b>13681</b>	GpGKNGDDGEAGK PG	COL1A 1	Collagen alpha-1(I) chain	P02452	224	239	0.931
<b>14556</b>	GSpGSpGPDGKTGPP Gp	COL1A 1	Collagen alpha-1(I) chain	P02452	542	558	0.934
<b>45005</b>	FQGLPpGpAGpPGEA GKpGEQGVPGDLGA pGPSG	COL1A 1	Collagen alpha-1(I) chain	P02452	642	674	0.941
<b>40305</b>	ERGSpGPAGPKGSpG EAGRpGEAGLpGAK G	COL1A 1	Collagen alpha-1(I) chain	P02452	510	539	0.946
<b>27973</b>	AGpPGEAGKpGEQG VPGDLGApGP	COL1A 1	Collagen alpha-1(I) chain	P02452	649	672	0.959
<b>28305</b>	AGpPGEAGKpGEQG VpGDLGApGP	COL1A 1	Collagen alpha-1(I) chain	P02452	649	672	0.963
<b>32074</b>	KGNSGEpGAPGSKG DTGAKGEpGpVG	COL1A 1	Collagen alpha-1(I) chain	P02452	430	455	0.967
<b>29755</b>	GKNGDDGEAGKPG RpGERGppGP	COL1A 1	Collagen alpha-1(I) chain	P02452	227	249	0.982
<b>14529</b>	VGPpGPPGpPpGpGP PS	COL1A 1	Collagen alpha-1(I) chain	P02452	1177	1193	0.993



<b>34813</b>	GPpGKNGDDGEAGK pGRpGERGPPGP	COL1A 1	Collagen alpha-1(I) chain	P02452	224	249	0.997
<b>19821</b>	TGSpGSpGPDGKTGP pGPAG	COL1A 1	Collagen alpha-1(I) chain	P02452	541	560	1.001
<b>31732</b>	KGNSGEPGApGSKG DTGAKGEPGpVG	COL1A 1	Collagen alpha-1(I) chain	P02452	430	455	1.019
<b>13342</b>	SpGSpGPDGKTGPPG p	COL1A 1	Collagen alpha-1(I) chain	P02452	543	558	1.021
<b>8050</b>	ApGDRGEpGPPGp	COL1A 1	Collagen alpha-1(I) chain	P02452	798	810	1.025
<b>6660</b>	SpGPDGKTGpPGP	COL1A 1	Collagen alpha-1(I) chain	P02452	546	558	1.046
<b>25386</b>	EGSpGRDGSpGAKG DRGETGP	COL1A 1	Collagen alpha-1(I) chain	P02452	1021	1041	1.047
<b>21723</b>	SpGRDGSpGAKGDR GETGP	COL1A 1	Collagen alpha-1(I) chain	P02452	1023	1041	1.389
<b>6317</b>	pGDRGEpGPpGP	COL1A 1	Collagen alpha-1(I) chain	P02452	799	810	1.434
<b>33377</b>	LDGAKGDAGPAGpK GepGSpGENGApG	COL1A 1	Collagen alpha-1(I) chain	P02452	273	299	1.495
<b>16523</b>	GPIGPpGPAGApGDK GES	COL1A 1	Collagen alpha-1(I) chain	P02452	767	784	1.838
<b>43435</b>	ESGREGAPGAEGSP GRDGSpGAKGDRGE TGp	COL1A 1	Collagen alpha-1(I) chain	P02452	1011	1041	2.040
<b>12101</b>	SpGEAGRpGEAGLpG	COL1A 1	Collagen alpha-1(I) chain	P02452	522	536	2.231
<b>34002</b>	PpGPPGRDGEDGPT GPPGPPGPpGPpG	COL1A 2	Collagen alpha-2(I) chain	P08123	46	72	0.449
<b>30077</b>	pGERGEVGPAGpNG FAGPAGAAGQP	COL1A 2	Collagen alpha-2(I) chain	P08123	711	735	0.519
<b>23575</b>	LpGSpGNIGPAGKEG PVGLpG	COL1A 2	Collagen alpha-2(I) chain	P08123	452	472	0.624
<b>29619</b>	AGPpGKAGEDGHpG KpGRpGERG	COL1A 2	Collagen alpha-2(I) chain	P08123	135	157	0.722

<b>29293</b>	AGPpGKAGEDGHpG KpGRPGERG	COL1A 2	Collagen alpha-2(I) chain	P08123	135	157	0.746
<b>45928</b>	RTGEVGA VGPpGFA GEKGPSGEAGTAGP pGTpGP	COL1A 2	Collagen alpha-2(I) chain	P08123	830	863	0.825
<b>17544</b>	SpGNIGPAGKEGPVG LpG	COL1A 2	Collagen alpha-2(I) chain	P08123	455	472	0.864
<b>58028</b>	DQGPVGR TGEVGA VGPpGFAGEKGPpSG EAGTAGPPGTpGPQ G	COL1A 2	Collagen alpha-2(I) chain	P08123	824	865	0.885
<b>42230</b>	LkGQpGApGVkGEpG ApGENGTpGQTGAR G	COL1A 2	Collagen alpha-2 (I) chain	P08123	188	217	0.890
<b>55841</b>	DQGPVGR TGEVGA VGppGFAGEKGPSG EAGTAGPpGTpGP	COL1A 2	Collagen alpha-2(I) chain	P08123	824	863	0.901
<b>28666</b>	NDGpPGRDQpGHK GERGYpG	COL1A 2	Collagen alpha-2(I) chain	P08123	932	952	0.925
<b>27882</b>	VGEpGpAGSKGESG NKGEPGSAGP	COL1A 2	Collagen alpha-2(I) chain	P08123	345	368	0.943
<b>889</b>	DpGKNGDKG	COL1A 2	Collagen alpha-2(I) chain	P08123	503	511	1.093
<b>5105</b>	ApGEAGRDNpG	COL1A 2	Collagen alpha-2(I) chain	P08123	920	931	1.240
<b>45138</b>	ppGAKGQEGAHGAP GAAGNPGAPGHVG APGPSGpP	COL22 A1	Collagen alpha- 1(XXII) chain	Q8NFW1 -3	899	933	0.679
<b>35934</b>	QGpPGPpGPQGLQGP KGEQGSPIPGM	COL25 A1	Collagen alpha- 1(XXV) chain	Q9BXS0	449	474	2.420
<b>1382</b>	GRpGPpGPpG	COL26 A1	Collagen alpha- 1(XXVI) chain	Q96A83	250	259	0.632
<b>8080</b>	DGVPGKDGPRGPT	COL3A 1	Collagen alpha- 1(III) chain	P02461	752	764	0.516
<b>10644</b>	TGpGGDKGDTGPpG P	COL3A 1	Collagen alpha- 1(III) chain	P02461	623	637	0.772

<b>4746</b>	SpGERGETGPP	COL3A 1	Collagen alpha- 1(III) chain	P02461	796	806	0.804
<b>25922</b>	GSNGNpGpPGPSGSp GKDGPpGP	COL3A 1	Collagen alpha- 1(III) chain	P02461	885	907	1.114
<b>8536</b>	SpGERGETGPPGP	COL3A 1	Collagen alpha- 1(III) chain	P02461	796	808	1.333
<b>52572</b>	NTGApGSPGVSGPK GDAGQpGEKGSpGA QGpPGAPGPLG	COL3A 1	Collagen alpha- 1(III) chain	P02461	910	948	1.432
<b>14888</b>	KGDpGpAGLpGKDG pP	COL5A 1	Collagen alpha-1(V) chain	P20908	1038	1053	1.272
<b>35960</b>	GpTGATGDKGPPGP VGPPGSNGpVGEpGP	COL5A 2	Collagen alpha-2(V) chain	P05997	1020	1048	0.494
<b>5347</b>	GPpGpPGPPGPpA	COL8A 1	Collagen alpha- 1(VIII) chain	P27658	560	572	0.833
<b>3797</b>	DGEkGDPGPpG	COL9A 3	Collagen alpha- 3(IX) chain	Q14050	207	217	1.059
<b>29136</b>	DEAGSEADHEGTHS TKRGHAK	FGA	Fibrinogen alpha chain	P02671	605	625	0.719
<b>38224</b>	DEAGSEADHEGTHS TKRGHAKSRPV	FGA	Fibrinogen alpha chain	P02671	605	629	0.938
<b>50457</b>	DEAGSEADHEGTHS TKRGHAKSRPV	FIBA	Fibrinogen alpha chain	P02671	605	629	0.717
<b>43336</b> <b>95</b>	VEDDGDDTSLDSDL DPE	FTSJ3	pre-rRNA processing protein FTSJ3	Q8IY81	460	476	1.528
<b>25362</b>	GGSGEDEQFLGFGS DEEVR	KMT2A	Histone-lysine N- methyltransferase 2A	Q03164	140	158	0.727
<b>18439</b>	PGPGGGGWGSGSSF RGTPG	MPLKI P	M-phase-specific PLK1-interacting protein	Q8TAP9	14	32	2.276
<b>3152</b>	SPASSQEGSPS	PAXIP1	PAX-interacting protein 1	Q6ZW49	219	229	1.752

<b>55834</b>	FGHGAEDSLADQAA NEWGRSGKDPNHFR PAGLPE	SAA1	Serum amyloid A protein	E9PQD6	87	120	0.900
<b>70785</b>	SDKPDMAEIEKFDK SKLKKKTETQEKNPL PSKETIEQEKQAGES	TYB4	Thymosin beta-4	P62328	2	44	0.776
<b>65617</b>	LPDHPWGTLNPSVS WGGGGPGTGWGTR PMPHPEGIWGINNQ P	USP31	Ubiquitin carboxyl- terminal hydrolase 31	Q70CQ4	1315	1325	3.074

**Table 2.** Nine sequenced peptides from the list of 22 peptides combined in the 22P model for PWV slope prediction.

<b>Peptide ID</b>	<b>Sequence</b>	<b>Prot_ Symbol</b>	<b>Prot_Name</b>	<b>Prot_ Accession</b>	<b>Start AA</b>	<b>Stop AA</b>	<b>Fold Change T3/T1</b>
<b>55834</b>	FGHGAEDSLADQA ANEWGRSGKDPNH FRPAGLPE	SAA1	Serum amyloid A protein	E9PQD6	87	120	1.222
<b>47804</b>	GADGQpGAKGEpG DAGAKGDAGPpGP AGPAGPpGPIG	COL1A 1	Collagen alpha- 1(I) chain	P02452	818	854	0.623
<b>28845</b>	ADGQPGAKGEpGD AGAKGDAGPPGp	COL1A 1	Collagen alpha- 1(I) chain	P02452	819	843	0.969
<b>32074</b>	KGNSGEpGAPGSKG DTGAKGEpGpVG	COL1A 1	Collagen alpha- 1(I) chain	P02452	430	455	1.031
<b>28733</b>	KGNSGEpGApGSKG DTGAKGEpGP	COL1A 1	Collagen alpha- 1(I) chain	P02452	430	453	1.072
<b>22895</b>	GNAGPpGPpGPAGK EGGKGPR	COL1A 1	Collagen alpha- 1(I) chain	P02452	890	910	1.090
<b>40424</b>	KEGKGPRGETGPA GRpGEVGPpGPpGP AG	COL1A 1	Collagen alpha- 1(I) chain	P02452	903	932	1.207

<b>49492</b>	GPpGESGREGAPGA EGSpGRDGSpGAKG DRGETGp	COL1A 1	Collagen alpha- 1(I) chain	P02452	1007	1041	1.217
<b>29293</b>	AGPpGKAGEDGHpG KpGRPGERG	COL1A 2	Collagen alpha- 2(I) chain	P08123	135	157	1.341

**Table 3.** Six sequenced peptides from the list of 14 peptides combined in the 14P model for LVMI slope prediction.

<b>Peptide ID</b>	<b>Sequence</b>	<b>Prot_ Symbol</b>	<b>Prot_Name</b>	<b>Prot_ Accession</b>	<b>Start AA</b>	<b>Stop AA</b>	<b>Fold Change T3/T1</b>
<b>10154</b>	GSpGPDGKTGPpGP A	COL1A1	Collagen alpha- 1(I) chain	P02452	545	559	2,664494
<b>21723</b>	SpGRDGSpGAKGD RGETGP	COL1A1	Collagen alpha- 1(I) chain	P02452	1023	1041	0,514339
<b>30951</b>	ADGQpGAKGepGD AGAKGDAGPpGpA	COL1A1	Collagen alpha- 1(I) chain	P02452	819	844	0,673761
<b>43789</b>	NSGepGApGSKGD TGAKGepGPVGVQ GPpGPAG	COL1A1	Collagen alpha- 1(I) chain	P02452	432	464	3,249639
<b>49422</b>	pTGAVGFAGpQGP DGQPGVKGEpGEp GQKGDAGSP	COL5A2	Collagen alpha- 2(V) chain	P05997	841	875	0,626274
<b>22191</b>	SSSKGSLGGGFSSG GFSGGSF	KRT10	Keratin; type I cytoskeletal 10	P13645	37	57	0,509296

## **Study 2:**

# CKD diagnosis in dogs

(Manuscript submitted to The Veterinary Journal - 2018)

---

## **Aims and Objectives**

Since several years now, research has brought new diagnostic tools based on the identification of urinary peptides as biomarkers for the early identification of CKD in humans. To help veterinary medicine, where the diagnosis of CKD is still struggling, we adapted the CE-MS technology of urinary peptides identification for the canine diagnosis. These biomarkers, which need to be validated on a larger cohort of dogs, will allow an earlier diagnosis of CKD in dogs so that the disease can be treated early to increase life expectancy.

# Urinary peptidome analyses for diagnosis of chronic kidney disease in dogs

Lena Pelander<sup>a</sup> <sup>‡\*</sup>, **Valérie Brunchault<sup>b,c</sup> <sup>‡</sup>**, Bénédicte Buffin-Meyer<sup>b,c</sup>, Julie Klein<sup>b,c</sup>, Benjamin Breuil<sup>b,c</sup>, Petra Züribig<sup>d,e</sup>, Pedro Magalhães<sup>d,e</sup>, William Mullen<sup>f</sup>, Jonathan Elliott<sup>g</sup>, Harriet Syme<sup>h</sup>, Joost P Schanstra<sup>b,c</sup>, Jens Häggström<sup>a</sup>, Ingrid Ljungvall<sup>a</sup>.

<sup>a</sup> Department of Clinical Sciences, University of Agricultural Sciences, Ulls väg 12, 750 07 Uppsala, Sweden. <sup>b</sup> Institut National de la Santé et de la Recherche Médicale (INSERM), U1048, Institut of Cardiovascular and Metabolic Disease, Equipe 12, 1 avenue Jean Poulhès, BP 84225, 31432 Toulouse Cedex 4, France. <sup>c</sup> Université Toulouse III Paul-Sabatier Toulouse, France. <sup>d</sup> Department of Pediatric Nephrology, Hannover Medical School, Hannover, Germany. <sup>e</sup> Mosaiques Diagnostics GmbH, Hannover, Germany. <sup>f</sup> BHF Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, United Kingdom. <sup>g</sup> Comparative Biomedical Sciences, Royal Veterinary College, London, UK. <sup>h</sup> Clinical Science and Services, Royal Veterinary College, North Mymms, UK.

<sup>‡</sup> These authors contributed equally to this work.

Key words: canine, CE-MS, CKD, renal      Total number of words: 5280

## Abbreviations

CKD    chronic kidney disease

mGFR    measured glomerular filtration rate

UPC    urine protein-to-creatinine ratio

IRIS    international renal interest society

SVM    support vector machine

ECM    extra-cellular matrix



## **Abstract**

Chronic kidney disease (CKD) is clinically important in canine medicine. Current diagnostic tools lack sensitivity to identify subclinical canine CKD. The aim of our study was to evaluate if mass spectrometry-based urinary peptidome analysis could be used to diagnose CKD and improve the detection of CKD in dogs. Analysis demonstrated presence of ~5400 peptides in dog urine. Comparison of dogs with and without CKD identified 133 differentially excreted peptides (adjusted p-values <0.05). Sequence information was obtained for 35 peptides out of 133. These 35 and 133 peptides were included in two predictive models of CKD, and validated in an independent cohort of 20 dogs. Both models predicted CKD in this blinded cohort with an identical area under the ROC curve of 0.88 (95% CI: 0.72 to 1.0). Most of the differentially excreted peptides represented fragments of collagen I, indicating possible association with fibrotic processes in CKD, as is the case for the equivalent human urinary peptide CKD model (CKD273). In conclusion, this first study of the urinary peptidome in dogs identified peptides that predicted presence of CKD. Future studies should validate the usefulness of this model for diagnosis and prediction of progression of canine CKD in a clinical setting.

### **Statement of significance of the study**

In this paper CE-MS technology, originally developed for the diagnosis of early chronic kidney disease (CKD) in humans, was applied for the first time in the canine population. Similar to the situation in human medicine, currently available diagnostic tools lack sensitivity to identify early CKD in dogs. The identification and blinded validation of two urinary peptide marker models for diagnosis of canine CKD are described. Most of the peptides in the biomarker models were fragments of extracellular matrix components and it was hypothesized that these fragments represent the early fibrotic process *in situ* in the kidney that constitutes the histological hallmark of CKD. If proved to be a reproducible and specific diagnostic method for dogs, that is, coupled to the intrarenal fibrosing process common to all forms of CKD, it will provide a completely new way of diagnosing CKD. The identification of ongoing renal damage occurring before global renal functional loss is presently only possible with renal biopsy, which is a costly and risky procedure.

To our knowledge, this is the first urinary peptidome-based tool investigated for detection of CKD in dogs. We therefore believe it is suitable for the readership of Proteomics.

## Introduction

Chronic kidney disease (CKD) is a clinically important cause of morbidity and mortality in dogs.<sup>[1]</sup> In veterinary medicine, CKD is defined as “structural or functional abnormalities (present for at least three months) of one or both kidneys”.<sup>[1]</sup> This heterogeneous disease is insidious in onset and often not recognized until late in the course of disease.

Routine methods used for diagnosis of CKD include measurement of serum creatinine concentration and urinalysis, including evaluation of presence of renal proteinuria.<sup>[1]</sup> In addition, diagnostic imaging is often used to detect structural lesions of the kidneys. However, current diagnostic methods are insensitive for detection of early CKD. Also, evidence of persistence of disease, hence, repeated measurements, are needed to confirm presence of CKD.<sup>[1]</sup> Therefore, research with the aim of improving early diagnosis of CKD in dogs has gained attention. Several serum or plasma biomarkers of decreased glomerular filtration rate (GFR) have been suggested as candidate biomarkers for future implementation in the clinic, but because of the immense compensatory capacity of the kidneys a reduction in GFR (even if measured by renal clearance studies or by scintigraphy) does not ensue until this compensatory adaptation fails.<sup>[2]</sup> Therefore, markers of GFR are not likely to ever be able to indicate subclinical progressive renal damage.

Urine (which is non-invasively available in comparably large volumes from most dogs) has evolved as a potential source of biomarkers for diagnostic use for diseases of the kidney and urinary tract.<sup>[3]</sup> Urine as a biological fluid for peptidome analyses is stable, since any proteolytic degradation is considered to be completed in the bladder by the time of urination.<sup>[3, 4]</sup> Urinary peptidomics (i.e. analysis of the low molecular weight proteome), using capillary electrophoresis coupled to mass spectrometry (CE-MS), has shown to be a particularly useful tool to diagnose and predict CKD and its complications in people.<sup>[5]</sup> The sample preparation is robust and several thousands of peptides that compose the human urinary peptidome can be analyzed by employing standard operating protocols and normalization procedures within a short time span of approximately one hour.<sup>[6]</sup> In this context, a model containing 273 urinary peptide biomarkers, called CKD273, was identified by CE-MS and shown to perform significantly better than urinary albumin in early detection and prediction of progression of human CKD.<sup>[4, 7]</sup> This model is

currently used in the PRIORITY trial (NCT02040441) in Europe, in which diabetic patients are screened for presence of early renal lesions.<sup>[8]</sup>

Several studies have investigated the urinary proteome in healthy dogs or in dogs with kidney disease, but only two of these studies have explored proteomics for the diagnosis of CKD.<sup>[9]</sup> No attempt to validate study findings in a separate cohort of dogs was performed in either study.

Therefore, the aim of this study was to evaluate if CE-MS-based urinary peptidome analysis can discriminate healthy dogs from dogs with CKD with high sensitivity and specificity. A secondary aim was to identify the peptides included in the discriminating models.

## **Materials and methods**

### ***Study population***

This cross-sectional observational study was performed at the Swedish University of Agricultural Sciences in Uppsala, after approval by the local ethical committee (Uppsala djurförsöksetiska nämnd, Sweden). All experiments were performed in accordance with relevant guidelines and regulations. Client-owned dogs were included in the study provided that the owner had given informed consent. Dogs with a previous diagnosis (or a strong suspicion of) CKD and healthy dogs, of any breed, body weight (BW) and age, were prospectively recruited. Dogs were considered to have a conclusive diagnosis of CKD if they had multiple renal cysts, persistent azotemia, persistent proteinuria, a persistently decreased measured glomerular filtration rate (mGFR), or a combination thereof. Exclusion criteria were the presence of other systemic or organ related disease. If a dog was receiving an angiotensin converting enzyme inhibitor, the drug was withdrawn a week before inclusion, and reintroduced after study inclusion. Dogs chronically medicated with other drugs (except sodium pentosane polysulfate injections) were excluded, as were dogs with bacteriuria. Oral administration of glucosaminoglycan supplements and feeding a kidney diet was allowed. Dogs referred with a suspicion of kidney disease, for which a diagnosis of CKD could not be confidently confirmed during the clinical investigation, were included as “inconclusive”. Healthy student-, client-, and staff-owned dogs of various breeds and ages were included as controls. Also, six healthy beagles from a Swedish research institution were included as healthy control dogs. These beagles were considered free of kidney disease

based on the absence of clinical signs, a creatinine concentration within the reference range and a normal complete urinalysis, including semi-quantitative biochemical analysis, sediment examination, protein-to-creatinine ratio (UPC), and normal kidney histology.

On the day of enrollment into the study, all dogs (including all control dogs except the six research beagles) underwent repeated blood pressure measurements, a physical examination, collection of venous blood and urine, echocardiographic examination, abdominal ultrasound examination of the entire urinary tract, and a scintigraphic examination for calculation of individual kidney mGFR. Dogs had to be clinically stable and fasted for 12 h on the day of inclusion. Dogs with CKD were staged according to the International Renal Interest Society (IRIS) classification system, based on stable serum creatinine concentration.<sup>[10]</sup>

#### ***Blood pressure measurement***

Indirect blood pressure measurements were performed according to reported guidelines<sup>[11]</sup> by high definition oscillometry (S+B medVET, Germany).

#### ***Blood and urine examinations***

Blood was drawn from the cephalic vein and transferred to the laboratory at the University Animal Hospital for immediate hematological and biochemical analysis. For most dogs, urine was obtained by cystocentesis. When cystocentesis was not possible (n=8), fresh spontaneously voided urine was obtained. In the six beagles, urine was obtained by cystocentesis, snap frozen and stored at -70°C. Remaining urine from the beagles was cooled and analyzed within 24 hours (dipstick and sediment examinations, specific gravity, urine protein-to-creatinine ratio (UPC) and aerobic culture).

#### ***Abdominal ultrasound examination***

Complete upper and lower urinary tract ultrasound examinations were conducted by radiologists at the university animal hospital diagnostic imaging clinic in Uppsala according to a pre-defined protocol.

### ***Glomerular filtration rate measurement by scintigraphy***

Estimation of individual kidney GFR was performed by a board-certified radiologist after renal scintigraphy, using the plasma volume method as previously described.<sup>[12]</sup>

### ***Sample preparation***

Urine samples were shipped on dry ice to Inserm U1048, Toulouse, France, and thawed immediately before use. A volume of 0.7 mL was diluted with 0.7 mL 2 M urea, 10 mM NH<sub>4</sub>OH and 0.02 % sodium dodecyl sulfate (SDS). In order to remove high molecular weight polypeptides, samples were filtered using Centriscart ultracentrifugation filter devices (20 kDa molecular weight cut-off); Sartorius, Goettingen, Germany) at 3000 g until 1.1 mL of filtrate was obtained. The filtrate was desalted with PD-10 column (GE Healthcare, Sweden) equilibrated in 0.01% NH<sub>4</sub>OH in HPLC-grade water. The prepared samples were lyophilized and stored at 4°C. Shortly before CE-MS analysis, lyophilised samples were resuspended in HPLC-grade water (Merck KGaA, Darmstadt, Germany). The preparation method has previously been described in more detail by Theodorescu *et al.* 2006.<sup>[13]</sup>

### ***CE-MS analysis and data processing***

CE-MS analysis was performed as previously described.<sup>[6]</sup> Briefly, CE-MS analyses were performed using a Beckman Coulter Proteome Lab PA800 capillary electrophoresis system (Beckman Coulter, Fullerton, USA) on-line coupled to a micrOTOF II MS (Bruker Daltonic, Bremen, Germany). The electro-ionization sprayer (Agilent Technologies, Palo Alto, CA, USA) was grounded, and the ion spray interface potential was set to -4.5 kV. Data acquisition and MS acquisition methods were automatically controlled by the CE via contact-close-relays. Spectra were accumulated every 3 s, over a range of m/z 350 to 3000. In the next step the MosaiquesVisu software package was applied to deconvolute mass spectral ion peaks, because ionization produced ions at different charged states from the original urinary peptides. This deconvolution step groups these differently charged ions into single peptides with unique real mass. Only signals observed in a minimum of three consecutive spectra with a signal-to-noise ratio of at least 4 were considered. Signals with a calculated charge of 1+ were automatically excluded to minimize interference with matrix compounds or drugs. Capillary electrophoresis migration time and

MS-detected mass were normalized by the definition of 950 clusters of peptides covering a range of 17.23 to 47.74 minutes in CE migration time and 807 to 16399 kDa in molecular mass. Peptide abundance (intensity) calibration was based on 141 endogenous internal urinary polypeptide standards with > 81% frequency and < 61% amplitude deviation to compensate for differences in dehydration and urine dilution between dogs.<sup>[14]</sup> Each polypeptide present in the list was defined by its normalized migration time [min], molecular mass [kDa], and signal intensity detected. Using a Microsoft Structured Query Language database, all detected polypeptides were deposited, matched, and annotated in order to allow for further comparison between the groups. The criteria applied to consider a polypeptide identical was that within different samples, the mass deviation was lower than 50 ppm for masses < 4 kDa, 150 ppm for masses > 6 kDa, and between 50-150 ppm for masses between 4 and 6 kDa. Acceptable migration time deviation was between 1 and 2.5 minutes.

### *Sequencing of peptides*

Candidate biomarkers and other native peptides from dog urine were sequenced using LC-MS/MS and CE-MS/MS analysis.<sup>[15]</sup> LC-MS/MS analysis experiments were performed on a Dionex Ultimate 3000 RSLC nano flow system (Dionex, Camberly UK). For CE-MS/MS, the samples were injected under constant flow and pressure conditions at a pH of 2.2 to ensure that all peptides are positively charged. Both CE and LC were directly interfaced with an LTQ-Orbitrap XL (Thermo Finnigan, Bremen, Germany), using data-dependent high-energy collisional dissociation (HCD) MS/MS sequencing of a maximum of the top 20 ions. All resultant MS/MS data were analyzed using Proteome Discoverer 1.3 (activation type: HCD; min-max precursor mass: 790-6000; precursor mass tolerance: 10 ppm; fragment mass tolerance: 0.05 Da; S/N threshold: 1) and were searched against the Uniprot canine non-redundant database without enzyme specificity. No fixed modifications were selected, oxidation of methionine and proline were selected as variable modifications. The peptide data were extracted using high confidence peptides, defined by an Xcorr  $\geq$  1.9, a delta mass between experimental and theoretical mass  $\pm$  5 ppm, absence of cysteine in the sequence as without reduction and alkylation it forms disulphide bonds, absence of oxidized proline in protein precursors other than collagens or elastin, and top one peptide rank filters.

For further validation of peptide identification, the strict correlation between peptide charge at pH 2 and CE-migration time was utilized to minimize false-positive identification rates.<sup>[16]</sup> Calculated CE-migration time of the sequence candidate based on its peptide sequence (number of basic amino acids) was compared to the charges of the peptides and the experimental CE-migration time. Peptides were accepted only if they had a mass deviation below  $\pm 90$  ppm and a CE-migration time deviation below  $\pm 2$  min.

### ***Biomarker selection and modelling***

For the identification of candidate urinary biomarkers, the reported P-values were calculated using the Wilcoxon Rank-Sum test (R software package, version 3.1.3) followed by adjustment for multiple testing using the method described by Benjamini and Hochberg.<sup>[17]</sup> Peptides that were detectable in at least 75% of dogs in one of the two groups (healthy *versus* CKD) and reached an adjusted p-value of  $<0.05$  were further considered as relevant. An R-based (version 3.1.3) support vector machine (SVM)-package and leave-one-out feature selection approach were used to generate biomarker models. Sensitivity and specificity were calculated based on the number of properly classified samples. The overall yield of the polypeptide pattern was evaluated by receiver operating characteristic (ROC) and area under curve (AUC) plots using the Prism 7.00 GraphPad software.

## **Results**

### ***Study setup and patient data***

In total, 53 dogs (25 dogs with CKD, 25 healthy dogs and 3 inconclusive dogs) were included in the study. Clinical data for the 25 dogs with CKD are summarized in **Table 1**.



**Table 1: Clinical characteristics of dogs used for diagnosis of CKD.**

	<b>Discovery (CKD dogs, n=15)</b>	<b>Validation (CKD dogs, n=10)</b>	<b>P</b>
<b>Persistent azotemia</b>	10 (68%)	6 (60%)	0.73
<b>Persistent proteinuria</b>	7 (47%)	7 (70%)	0.25
<b>Persistently decreased mGFR</b>	12 (80%)	7 (70%)	0.57
<b>IRIS stage 1/2/3/4</b>	5/4/4/2	4/3/2/1	0.97
<b>Multiple cysts in both kidneys</b>	1 (7%)	1 (10%)	

CKD: chronic kidney disease, mGFR: glomerular filtration rate, IRIS stage: international renal interest society staging of CKD ([www.IRIS-kidney.com](http://www.IRIS-kidney.com)).

Dogs were considered to have a conclusive diagnosis of CKD if they had multiple renal cysts, persistent azotemia, persistent proteinuria, a persistently decreased mGFR, or a combination thereof. Healthy student-, client-, and staff-owned dogs of various breeds and ages, and six healthy beagles from a Swedish research institution, were included as healthy control dogs. According to guidelines for human clinical proteomics,<sup>[18]</sup> the population of 50 dogs with known clinical status was divided into two cohorts: a discovery cohort, in which the clinical status of each dog was known to those performing the analyses and an independent validation cohort, in which the clinical status of the dogs was blinded to those performing the analyses. The minimum number of individuals to include in a discovery cohort in a proteomic study has been suggested to be 24 (12+12).<sup>[19]</sup> For this reason, 30 dogs (15 healthy and 15 CKD dogs) were included in the discovery cohort. Thus, 20 dogs comprised the independent validation cohort (10 healthy dogs and 10 CKD dogs). An attempt was made to accomplish an even distribution of dogs in different IRIS CKD stages in the discovery- and validation groups, respectively. Inconclusive dogs were not used in the statistical modelling. Clinical data for dogs in the different groups are presented in **Table 2**.

**Table 2: Descriptive statistics of clinical variables by group** (median, interquartile range (IQR)).

	Discovery (D)			Validation (V)			D vs V
	Control (n=15)	CKD (n=15)	p	Control (n=10)	CKD (n=10)	p	p
<b>Creatinine (µmol/L)</b>	78 (72-88)	171(100-236)		77 (67-96)	132(86-212)		0.62
<b>UPC</b>	0,06 (0.03-0.12)	0,44 (0.15-1.67)		0,06 (0.03-0.13)	1,5 (0.17-3.42)		0.37
<b>mGFR* (ml/min/L)</b>	52 (48-71)	15 (13-30)		51 (45-73)	29 (25-47)		0.02
<b>Age (years)</b>	3.9 (1.9–7.7)	7.1 (2.5–9.8)	0.33	2.7 (1.6–6.5)	6.9 (3.9–9.5)	0.08	0.96
<b>Sex (F vs M)</b>	11 vs 4	10 vs 5	0.69	8 vs 2	2 vs 8	0.007	0.02
<b>SBP** (mmHg)</b>	147 (127-164)	158 (129-169)	0.47	123 (103-127)	144 (133-165)	0.001	0.94
<b>USG***</b>	1.036 (1.029-1.050)	1.024 (1.016-1.029)	0.007	1.028 (1.021-1.040)	1.022 (1.016-1.037)	0.25	0.90
<b>Urinary creatinine (µmol/L)</b>	19303 (10371-23826)	9923 (4791-19315)	0.032	13242 (8496-15183)	7701 (4290-14730)	0.19	0.95

UPC: urinary protein to creatinine ratio; mGFR: measured glomerular filtration rate; SBP: mean systolic blood pressure; USG: urine specific gravity

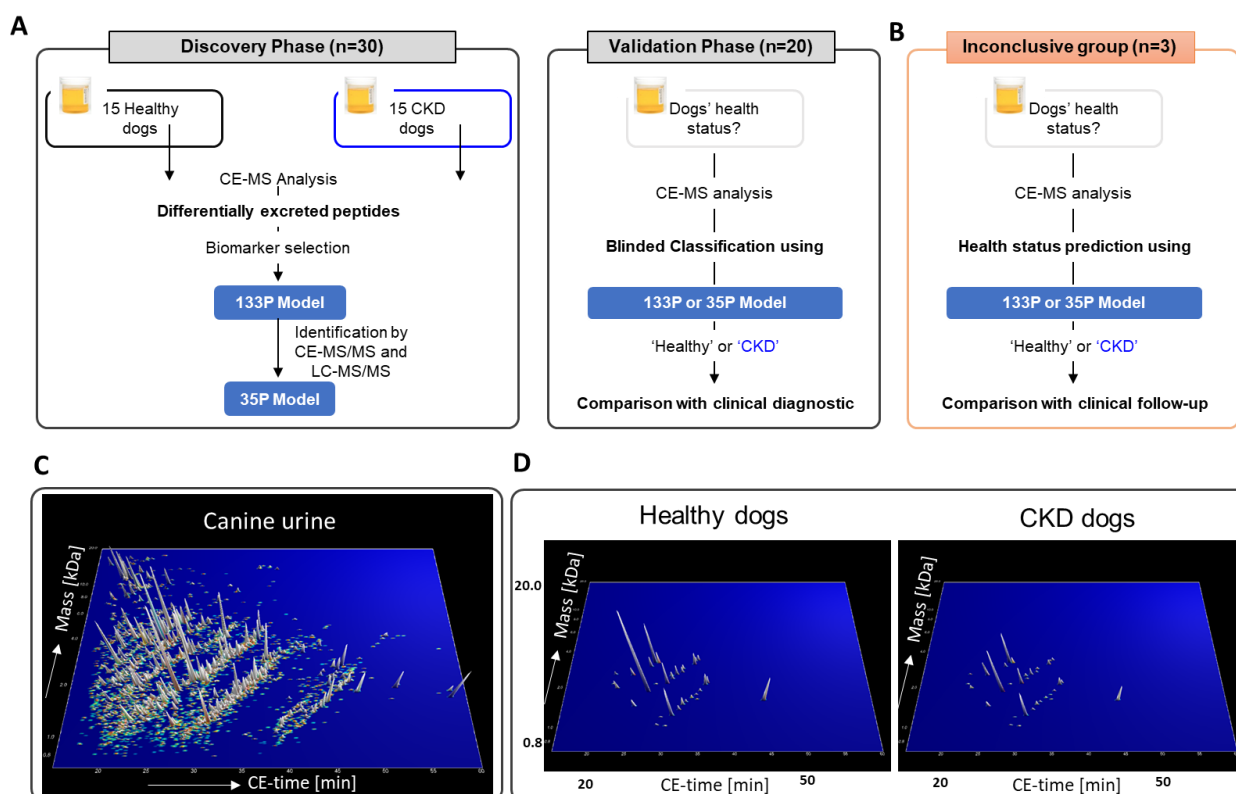
\*Measurement of mGFR was not available for all dogs. Discovery group: n=12 (control) and n=13 (CKD), validation group: n=7 (control) and n=8 (CKD).

\*\*Blood pressure measurement was not available for all dogs. Discovery group: n=7 (control) and n=12 (CKD), validation group n=7 (control) and n=8 (CKD).

\*\*\*USG and urinary creatinine were not available for two dogs. One from the discovery group and one from the validation group.

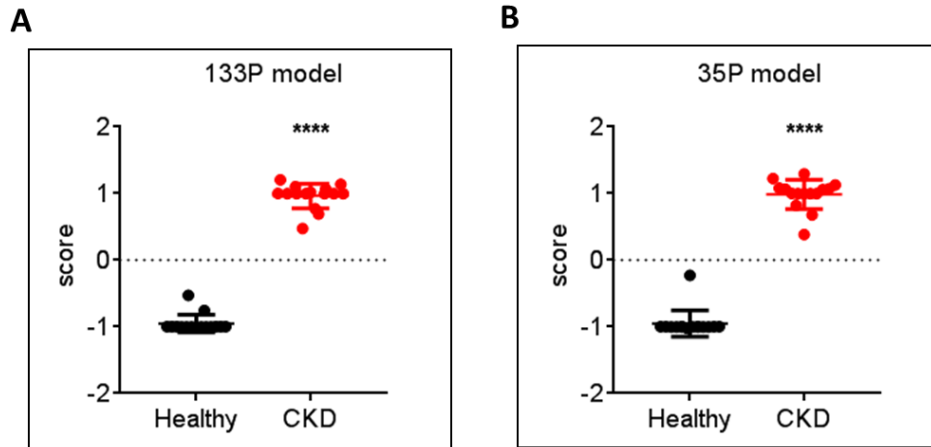
### ***Identification of urinary peptides associated with CKD***

Fifteen healthy and 15 dogs with CKD were used in the discovery phase of the study (**Fig. 1a**). CE-MS analysis of canine urine led to the identification of in total 5398 different peptides in all samples (**Fig 1C**).



**Fig. 1 Study set-up and CE-MS analysis of canine urine.** (a) The analysis was performed in two separate phases: a discovery phase and a validation phase. Urine from 30 dogs (15 healthy and 15 with CKD) was analyzed, leading to the identification of 133 differentially secreted peptides (133P model). Of these peptides, 35 were identified by CE-MS/MS and LC-MS/MS sequencing and included in the 35P model. In the validation phase, the two different models were tested on an independent, blinded cohort of dogs (n=20) to evaluate their predictive value. (b) Prediction of health status of dogs in the inconclusive group (n=3) was performed using models developed in “a” and validated in “b”. (c) Peptide pattern showing the compiled datasets of 30 canine urine samples. Each peptide (n=5398) was identified based on CE migration time and specific mass (kD), with relative abundance represented by the peak height. (d) Sequenced peptide (n=35) pattern distinguishing dogs with CKD from healthy dogs.

For inter sample comparability, all samples were normalized by 141 endogenous urinary peptides displaying the highest frequency and stability in all analyzed samples as described previously for people<sup>[4]</sup> and in detail in the M&M section. Comparison of canine urine from healthy and CKD dogs in this discovery phase resulted in the identification of 133 differentially excreted peptides after correction for multiple testing. These 133 peptides were combined in an SVM model, resulting in a model called 133P, which clearly separated healthy and CKD dogs in the discovery cohort (**Fig 2a**). In the next step we tried to obtain sequence information of the 133 urinary peptides by CE-MS/MS and LC-MS/MS (**Fig. 1a**).

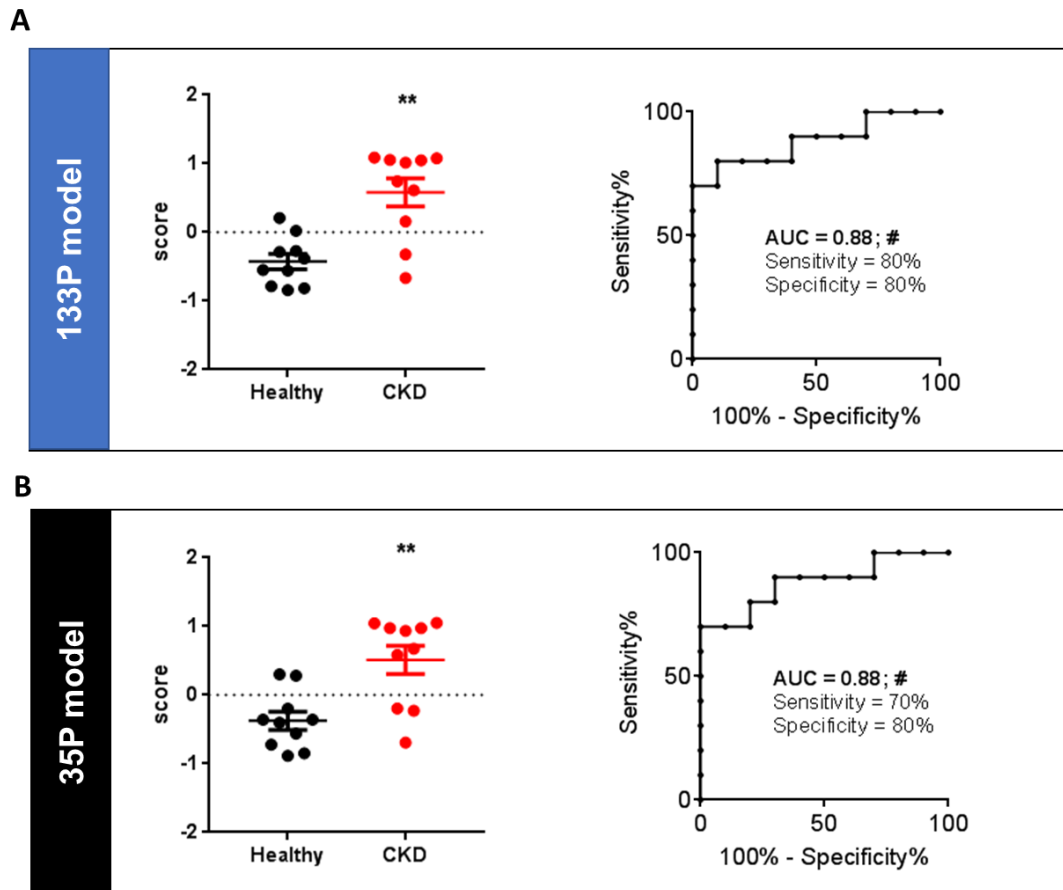


**Fig. 2** Distribution scores of urinary peptide models for CKD vs. healthy dogs in the discovery cohort. (a) 133P model, (b) 35P model. \*\*\*\* $p < 0.0001$  versus healthy dogs, Mann-Whitney test for independent samples.

This led to identification of 35 peptides: 33 collagen (I and IV) and 2 uromodulin fragments (**Supplementary table**). All 35 peptides were down-regulated in dogs suffering from CKD (**Fig. 1d**). These 35 peptides were combined in an SVM model, resulting in a model called 35P, also clearly separating healthy and CKD dogs in the discovery cohort (**Fig. 2b**).

#### *Validation of 133P and 35P models*

The two previously established models were validated in an independent, blinded cohort of 20 dogs (10 healthy dogs and 10 dogs with CKD) not used in the discovery phase (**Table 2**). The samples were scored using the 133P and 35P models and then given a prediction as to either an individual dog is healthy or suffering from CKD. A positive score ( $>0$ ) predicts a dog with CKD. These predictions were then compared to the clinical status of the dog.



**Fig 3 Validation of urinary peptide models in an independent, blinded cohort.** Classification of healthy and CKD dogs in the validation cohort according to (a) 133P model scores and ROC curve for the 133 P model (AUC 0.88 [0.721 to 1.04]); (b) 35P model scores and ROC curve for the 35P model (AUC 0.88 [0.723 to 1.04]). \*\* P < 0.01 versus healthy dogs, Mann-Whitney test for independent samples; AUC: Area under curve; [95% CI]; # : p<0.005.

The distribution of 133P and 35P scores (**Fig 3a-b**) showed significant separation of healthy and CKD dogs. The 133P model predicted CKD with a sensitivity of 80% [95%confidence interval (CI): 44 to 97%], a specificity of 80% (95% CI: 44 to 97%) and an area under the curve (AUC) of 0.88 (95% CI: 0.72 to 1.0) (**Fig 3a**). The 35P model predicted CKD with a sensitivity of 70% (95% CI: 35 to 93%), a specificity of 80% (95% CI: 44 to 97%) and an AUC of 0.88 (95% CI: 0.72 to 1.0) (**Fig 3b**).

Using the 133P and 35P models, two out of the three inconclusive dogs (**Fig. 1b**) were classified as healthy, and one dog as CKD.

## Discussion

Using urinary peptidome analysis, we developed and blindly validated two peptide models for discrimination between dogs with CKD and healthy dogs. Such a tool, based on a non-invasive, one point in time, urine analysis might, if validated and considered useful in further studies, allow early detection of CKD in dogs. Diagnosis of canine CKD currently mainly relies on repeated measurements (renal persistent proteinuria or azotemia) or on reliable parenchymal abnormalities detected by ultrasonography, all of which are insensitive for diagnosis of subclinical disease.<sup>[20]</sup>

Thirty-five of the peptides differentially secreted between CKD dogs and healthy dogs were sequenced in this study. Of the 35 sequenced peptides, 33 were collagen fragments (for the major part fragments of interstitial collagen I), which were less abundant in the urine of dogs with CKD than in healthy dogs. This is in agreement with studies in people with CKD.<sup>[4, 21]</sup> Most urinary peptides are results of proteolytic activity, probably reflecting differences in activity of different collagenases.<sup>[21]</sup> It was hypothesized that decreased urinary collagen fragments are associated with decreased collagenase activity and increased intrarenal extra-cellular matrix (ECM) deposition.<sup>[21]</sup> Intrarenal ECM deposition leading to fibrosis is a hallmark of CKD across species and therefore this reasoning most likely also holds true in dogs.

In the present cross-sectional study, we identified peptide models associated to the presence of CKD, which might contribute to early detection of CKD in dogs. An additional important issue is whether CKD is progressive. Regarding the human counterpart peptide model, CKD273, which was identified under similar conditions as those in this study, there is initial evidence that it can predict CKD progression.<sup>[7]</sup> This could be linked to the presence of uromodulin fragments in the model as described above. In addition, recently, a significant association between CKD273 score and degree of fibrosis in renal biopsies was shown.<sup>[22]</sup> These findings all further strengthen the hypothesis that these peptide-based models probably identify the fibrotic process that constitutes the histological hallmark of progressive CKD. If our models can identify progressive CKD, they detect something different than our current golden standard (the clinical diagnosis of canine CKD, based mainly on interpretation or measurement of GFR) does and, consequently, represent a new, single step option for diagnosis of progressive CKD.

In this study, we also investigated whether we could use only sequenced peptides to build another model, 35P, without losing performance. This model performed similarly to the 133P-model. However, biomarker models with higher numbers of included peptides are more robust and therefore more likely to perform better in independent cohorts.<sup>[23]</sup> In addition, the heterogenous nature of CKD is probably reflected in the urinary peptidome of patients with CKD, and therefore over-fitting is a risk when reducing numbers of peptides in a model that is intended for use as a diagnostic tool for canine CKD of all causes.

In order to investigate early diagnosis, ideal individuals with CKD to include in test- and validation sets are those that have progressive disease but not (yet) azotemia or clinical signs. Three dogs with a suspicion of CKD, for which no conclusive diagnosis of CKD could be made even after extensive diagnostic workup, were included as “inconclusive”. Two of these dogs were classified as “healthy” by both models, and none had evidence of reduced kidney function on follow up examinations at least two years after study inclusion. The third dog was included six months after treatment of, and recovery from, pyelonephritis and was non-azotemic, non-proteinuric with normal mGFR, and therefore could not be included in the “CKD”-group. However, mGFR in one kidney was half of that of the other kidney (which would not have been appreciated using only routine diagnostics), and both models categorized this dog as CKD. This dog was considered healthy by the owners two and a half years after inclusion but on follow up, mGFR was further decreased. Our interpretation is that this dog had progressive CKD at inclusion, as reflected by both models, but not detectable with routine diagnostic methods.

An important future task in the validation of the urinary peptide models is to investigate specificity of the model, most importantly in dogs with other diseases mimicking early CKD clinically, i.e. hyperadrenocorticism, diabetes insipidus and psychogenic polydipsia. An additional potential future application of this technique could be for specific diagnosis of different etiologies of CKD (potentially lessening the need for kidney biopsy), as this has been recently shown for urinary peptides in people using the same technological platform as in this study.<sup>[8]</sup>

In conclusion, urinary peptide-based models were able to discriminate healthy dogs from dogs with CKD, with an AUC of 0.88, in an independent cohort of dogs. Thirty-five of the 133 peptides

differentially expressed between healthy and CKD dogs were identified and most of them were collagen fragments. Peptidome analysis by CE-MS is a promising future tool for early diagnosis of canine CKD. Although further investigation of these models is necessary in order to validate their usefulness for early CKD diagnosis and prediction of progression in a clinical setting, it is the first urinary peptidome-based tool investigated for early detection of CKD in dogs.

### **Data availability**

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

### **Acknowledgements**

The authors thank AGRIA/SKK Research Foundation, Michael Forsgren Foundation, Thure F and Karin Forsberg Foundation and the French Ministry of Higher Education, Research and Innovation for funding this study. They also thank AstraZeneca.

### **Conflicts of interest**

PZ and PM are employees of Mosaiques Diagnostics. P.M. received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 642937 (RENALTRACT; MSCA-ITN-2014-642937).

### **References**

- [1] D. Polzin, in *Textbook of Veterinary Internal Medicine*, Vol. 2 (Ed: S. J. Ettinger, Feldman EC, Coté E), Elsevier, St. Louis, Missouri 2017, 1938.
- [2] S. A. Brown, D. R. Finco, W. A. Crowell, D. C. Choat, L. G. Navar, *The American journal of physiology* 1990, 258, F495.
- [3] S. Decramer, A. Gonzalez de Peredo, B. Breuil, H. Mischak, B. Monsarrat, J. L. Bascands, J. P. Schanstra, *Molecular & cellular proteomics : MCP* 2008, 7, 1850.
- [4] D. M. Good, P. Zurbig, A. Argiles, H. W. Bauer, G. Behrens, J. J. Coon, M. Dakna, S. Decramer, C. Delles, A. F. Dominiczak, J. H. Ehrich, F. Eitner, D. Fliser, M. Frommberger, A. Ganser, M. A. Girolami, I. Golovko, W. Gwinner, M. Haubitz, S. Herget-Rosenthal, J. Jankowski, H. Jahn, G. Jerums, B. A. Julian, M. Kellmann, V. Kliem, W. Kolch, A. S. Krolewski, M. Luppi, Z. Massy, M. Melter, C. Neususs, J. Novak, K. Peter, K. Rossing, H. Rupperecht, J. P. Schanstra, E. Schiffer, J. U. Stolzenburg,



- L. Tarnow, D. Theodorescu, V. Thongboonkerd, R. Vanholder, E. M. Weissinger, H. Mischak, P. Schmitt-Kopplin, *Molecular & cellular proteomics : MCP* 2010, 9, 2424.
- [5] J. P. Schanstra, H. Mischak, *Pediatric nephrology (Berlin, Germany)* 2015, 30, 713.
- [6] H. Mischak, A. Vlahou, J. P. Ioannidis, *Clinical biochemistry* 2013, 46, 432.
- [7] J. P. Schanstra, P. Zurbig, A. Alkhalaf, A. Argiles, S. J. Bakker, J. Beige, H. J. Bilo, C. Chatzikyrkou, M. Dakna, J. Dawson, C. Delles, H. Haller, M. Haubitz, H. Husi, J. Jankowski, G. Jerums, N. Kleefstra, T. Kuznetsova, D. M. Maahs, J. Menne, W. Mullen, A. Ortiz, F. Persson, P. Rossing, P. Ruggenenti, I. Rychlik, A. L. Serra, J. Siwy, J. Snell-Bergeon, G. Spasovski, J. A. Staessen, A. Vlahou, H. Mischak, R. Vanholder, *Journal of the American Society of Nephrology : JASN* 2015, 26, 1999.
- [8] J. Siwy, P. Zurbig, A. Argiles, J. Beige, M. Haubitz, J. Jankowski, B. A. Julian, P. G. Linde, D. Marx, H. Mischak, W. Mullen, J. Novak, A. Ortiz, F. Persson, C. Pontillo, P. Rossing, H. Rupperecht, J. P. Schanstra, A. Vlahou, R. Vanholder, *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2016.
- [9] S. Forterre, J. Raila, F. J. Schweigert, *Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc* 2004, 16, 271; M. B. Nabity, G. E. Lees, L. J. Dangott, R. Cianciolo, J. S. Suchodolski, J. M. Steiner, *Veterinary clinical pathology / American Society for Veterinary Clinical Pathology* 2011, 40, 222.
- [10] D. J. Polzin, *The Veterinary clinics of North America. Small animal practice* 2011, 41, 15.
- [11] S. Brown, C. Atkins, R. Bagley, A. Carr, L. Cowgill, M. Davidson, B. Egner, J. Elliott, R. Henik, M. Labato, M. Littman, D. Polzin, L. Ross, P. Snyder, R. Stepien, *Journal of veterinary internal medicine / American College of Veterinary Internal Medicine* 2007, 21, 542.
- [12] F. Westgren, C. J. Ley, N. Kampa, P. Lord, *Veterinary radiology & ultrasound : the official journal of the American College of Veterinary Radiology and the International Veterinary Radiology Association* 2014, 55, 632.
- [13] D. Theodorescu, E. Schiffer, H. W. Bauer, F. Douwes, F. Eichhorn, R. Polley, T. Schmidt, W. Schofer, P. Zurbig, D. M. Good, J. J. Coon, H. Mischak, *Proteomics. Clinical applications* 2008, 2, 556.
- [14] R. Dissard, J. Klein, C. Caubet, B. Breuil, J. Siwy, J. Hoffman, L. Sicard, L. Ducasse, S. Rascalou, B. Payre, M. Buleon, W. Mullen, H. Mischak, I. Tack, J. L. Bascands, B. Buffin-Meyer, J. P. Schanstra, *PloS one* 2013, 8, e76703.
- [15] J. Klein, T. Papadopoulos, H. Mischak, W. Mullen, *Electrophoresis* 2014, 35, 1060.
- [16] P. Zurbig, M. B. Renfrow, E. Schiffer, J. Novak, M. Walden, S. Wittke, I. Just, M. Pelzing, C. Neususs, D. Theodorescu, K. E. Root, M. M. Ross, H. Mischak, *Electrophoresis* 2006, 27, 2111.
- [17] Y. Benjamini, Y. Hochberg, *Journal of the Royal Statistical Society. Series B (Methodological)* 1995, 57, 289.
- [18] H. Mischak, R. Apweiler, R. E. Banks, M. Conaway, J. Coon, A. Dominiczak, J. H. Ehrich, D. Fliser, M. Girolami, H. Hermjakob, D. Hochstrasser, J. Jankowski, B. A. Julian, W. Kolch, Z. A. Massy, C. Neusuess, J. Novak, K. Peter, K. Rossing, J. Schanstra, O. J. Semmes, D. Theodorescu, V. Thongboonkerd, E. M. Weissinger, J. E. Van Eyk, T. Yamamoto, *Proteomics. Clinical applications* 2007, 1, 148; H. Mischak, G. Allmaier, R. Apweiler, T. Attwood, M. Baumann, A. Benigni, S. E. Bennett, R. Bischoff, E. Bongcam-Rudloff, G. Capasso, J. J. Coon, P. D'Haese, A. F. Dominiczak, M. Dakna, H. Dihazi, J. H. Ehrich, P. Fernandez-Llama, D. Fliser, J. Frokiaer, J. Garin, M. Girolami, W. S. Hancock, M. Haubitz, D. Hochstrasser, R. R. Holman, J. P. Ioannidis, J. Jankowski, B. A. Julian, J. B. Klein, W. Kolch, T. Luider, Z. Massy, W. B. Mattes, F. Molina, B. Monsarrat, J. Novak, K. Peter, P. Rossing, M. Sanchez-Carbayo, J. P. Schanstra, O. J. Semmes, G. Spasovski, D. Theodorescu, V. Thongboonkerd, R. Vanholder, T. D. Veenstra, E. Weissinger, T. Yamamoto, A. Vlahou, *Science translational medicine* 2010, 2, 46ps42.
- [19] M. Dakna, K. Harris, A. Kalousis, S. Carpentier, W. Kolch, J. P. Schanstra, M. Haubitz, A. Vlahou, H. Mischak, M. Girolami, *BMC bioinformatics* 2010, 11, 594.
- [20] J. P. Braun, H. P. Lefebvre, A. D. Watson, *Veterinary clinical pathology / American Society for Veterinary Clinical Pathology* 2003, 32, 162.
- [21] K. Rossing, H. Mischak, P. Rossing, J. P. Schanstra, A. Wiseman, D. M. Maahs, *Proteomics. Clinical applications* 2008, 2, 997.
- [22] P. Magalhaes, M. Pejchinovski, K. Markoska, M. Banasik, M. Klinger, D. Svec-Billa, I. Rychlik, M. Rroji, A. Restivo, G. Capasso, F. Bob, A. Schiller, A. Ortiz, M. V. Perez-Gomez, P. Cannata, M. D. Sanchez-Nino, R. Naumovic, V. Brkovic, M. Polenakovic, W. Mullen, A. Vlahou, P.

Zurbig, L. Pape, F. Ferrario, C. Denis, G. Spasovski, H. Mischak, J. P. Schanstra, *Scientific reports* 2017, 7, 16915.  
[23] D. Fliser, J. Novak, V. Thongboonkerd, A. Argiles, V. Jankowski, M. A. Girolami, J. Jankowski, H. Mischak, *Journal of the American Society of Nephrology : JASN* 2007, 18, 1057.

**SUPPLEMENTARY Table:**

**Peptide sequences of thirty-five urinary peptides, differentially excreted between healthy and CKD dogs in the discovery cohort.**

Peptide ID	Peptide sequence	Protein Name	UniProt Name	p-value (Wilkinson)	Adjusted p-value (BH)*
1118	DGRpGpPpGpG	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00105003	0.0112353
1287	GDRGEpGpPpGP	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00618982	0.03679507
2544	GpPpGESGREGSpG	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00897202	0.04465147
2873	ApGDRGEpGpPpGP	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00794034	0.04212972
4264	ApGDRGEpGpPpGPAG	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00030786	0.00732031
5134	DGQPGAKGEpGDAGAK	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00369059	0.02755065
5497	GSpGSpGPDGKTGPPGp	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00049369	0.00880411
6130	SpGSpGPDGKTGpPpGPAG	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00057598	0.00897333
6513	VGpPGpPpGpPpGPPGPPSGG	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00161971	0.015292
6662	VGpPGpPpGpPpGPPGPPSGG	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00479476	0.03173438
6802	DQGPVGRGTGETGASGpPG	Collagen alpha-2(I) chain	CO1A2_CANLF	0.00042247	0.00874917
7219	NGApGNDGAKGDAGApGApG	Collagen alpha-1(I) chain	CO1A1_CANLF	0.0024626	0.02107989
8115	GEKGPSGEpGTAGPpGTpGP	Collagen alpha-2(I) chain	CO1A2_CANLF	0.00022146	0.00681417
8907	SGGIIDQSRVNLGPITR	Uromodulin	UROM_CANLF	6.8368E-05	0.00337633
9040	TGEKGpSGEpGTAGPpGTpGP	Collagen alpha-2(I) chain	CO1A2_CANLF	0.00078021	0.00963253
9966	SGGIIDQSRVNLGPITRK	Uromodulin	UROM_CANLF	0.00049369	0.00880411
11110	AGPpGEAGKpGEQGVPGDLGApGP	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00105003	0.0112353
11265	AGPpGEAGKpGEQGVpGDLGApGP	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00545214	0.03431641

11289	GPpGpPGGMKGEKGEQGEpGKR	Collagen alpha-5(IV) chain	CO4A5_CANLF	0.00322672	0.02468753
11294	PGpDGKTGPPGPAGQDGRPGPPGP	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00618982	0.03679507
11526	RGAPGDRGEpGPpGPAGFAGppGA	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00545214	0.03431641
11728	FTGEKGPSGEpGTAGPpGTPGpQG	Collagen alpha-2(I) chain	CO1A2_CANLF	6.8368E-05	0.00337633
13263	LDGAKGDAGPAGPKGEpGSpGENGApG	Collagen alpha-1(I) chain	CO1A1_CANLF	2.7983E-05	0.00337633
14251	VNGApGEAGRDNpGNDGPpGRDQGAG	Collagen alpha-2(I) chain	CO1A2_CANLF	0.00369059	0.02755065
14554	pGDKGEAGPSGpAGpTGARGApGDRGEP	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00420995	0.02970094
14824	AGPpGApGApGAPGPVGPAGKNGDRGETGP	Collagen alpha-1(I) chain	CO1A1_CANLF	0.0101218	0.04885862
14848	KEGGKARGETGPAGRpGEVGPpGPpGP	Collagen alpha-1(I) chain	CO1A1_CANLF	5.6936E-05	0.00337633
15775	GSRGDGpGpGATGFPGAAGRTGPpGpSGITG	Collagen alpha-2(I) chain	CO1A2_CANLF	1.9352E-05	0.00337633
17053	NGPpGPAGSRGDGpPGATGFpGAAAGRTGpPGP	Collagen alpha-2(I) chain	CO1A2_CANLF	0.00018795	0.00639273
18031	LDGAKGDAGPAGPKGEpGSpGENGApGQMGPRG	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00102315	0.0112353
19095	GADGQPGAKGEpGDAGAKGDAGPpGPAGPTGPpGPIG	Collagen alpha-1(I) chain	CO1A1_CANLF	0.00078021	0.00963253
19687	AAGEpGKAGERGVpGpGAVGPAGKDGEAGAQQPPGP	Collagen alpha-1(I) chain	CO1A1_CANLF	8.134E-05	0.00373003
19943	GpAGVRGPNNGDSGRPGEPGLmGpRGFPGAPGNVGP	Collagen alpha-2(I) chain	CO1A2_CANLF	0.00161971	0.015292
20535	ARGNDGATGAAGPpGPTGPAGPpGpGAVGAKGEAGpQG	Collagen alpha-1(I) chain	CO1A1_CANLF	6.786E-05	0.00337633
20952	GPpGADGQPGAKGEpGDAGAKGDAGpPGPAGPTGPpGPIG	Collagen alpha-1(I) chain	CO1A1_CANLF	0.004748	0.03173438

Abbreviations in sequence: p, hydroxyproline; k, hydroxylysine; m, hydroxymethionine.

\* p-value adjusted by the method of Benjamini and Hochberg<sup>[17]</sup>



## Study 3:

# Identification of metabolite biomarkers using CE-MS technology

Boizard F\*, Brunchault V\* *et al.*, 2016, Scientific Reports


\*Equal contribution

---

## **Aims and Objectives**

Metabolites are believed to be the closest molecules associated to a phenotype since they integrate information from the genome, transcriptome, proteome and respond to subtle variations such as diet, medication or changes during disease development/progression. The objective of this project was to optimize the CE-MS set-up and pipeline for the analysis of urine metabolome in clinical research. As a proof of concept of this novel approach, we focused on its use for the identification of metabolite biomarkers of a kidney anomaly, the ureteropelvic junction obstruction, in newborns.

# SCIENTIFIC REPORTS



OPEN

## A capillary electrophoresis coupled to mass spectrometry pipeline for long term comparable assessment of the urinary metabolome

Received: 20 July 2016  
Accepted: 14 September 2016  
Published: 03 October 2016

Franck Boizard<sup>1,2,\*</sup>, Valérie Brunchault<sup>1,2,\*</sup>, Panagiotis Moulos<sup>3</sup>, Benjamin Breuil<sup>1,2</sup>, Julie Klein<sup>1,2</sup>, Nadia Lounis<sup>4</sup>, Cécile Caubet<sup>1,2</sup>, Stéphanie Tellier<sup>5</sup>, Jean-Loup Bascands<sup>1,2</sup>, Stéphane Decramer<sup>1,2,5</sup>, Joost P. Schanstra<sup>1,2</sup> & Bénédicte Buffin-Meyer<sup>1,2</sup>

Although capillary electrophoresis coupled to mass spectrometry (CE-MS) has potential application in the field of metabolite profiling, very few studies actually used CE-MS to identify clinically useful body fluid metabolites. Here we present an optimized CE-MS setup and analysis pipeline to reproducibly explore the metabolite content of urine. We show that the use of a beveled tip capillary improves the sensitivity of detection over a flat tip. We also present a novel normalization procedure based on the use of endogenous stable urinary metabolites identified in the combined metabolome of 75 different urine samples from healthy and diseased individuals. This method allows a highly reproducible comparison of the same sample analyzed nearly 130 times over a range of 4 years. To demonstrate the use of this pipeline in clinical research we compared the urinary metabolome of 34 newborns with ureteropelvic junction (UPJ) obstruction and 15 healthy newborns. We identified 32 features with differential urinary abundance. Combination of the 32 compounds in a SVM classifier predicted with 76% sensitivity and 86% specificity UPJ obstruction in a separate validation cohort of 24 individuals. Thus, this study demonstrates the feasibility to use CE-MS as a tool for the identification of clinically relevant urinary metabolites.

'Omics'-based strategies appear to be promising tools for the identification of diagnostic and prognostic biomarkers of disease. They can lead to the design of multimarker models which are potentially better suited than single biomarkers to describe complex pathophysiological mechanisms<sup>1-3</sup>. Metabolomics, defined as the analysis of the low-molecular-weight compound (<1500 Da) content of a sample, offers advantages compared to the other omics traits. Indeed, being the downstream products of cellular function, metabolites represent a sensitive measure of the actions of upstream molecular species such as genes, transcripts, and enzymes, including the effects of disease, drugs, toxicity, and the environment<sup>4,5</sup>. However sensitivity to these many perturbants also contributes to potential issues about the high variability in metabolome exploration<sup>6</sup>.

Analysis of urine plays a central role in clinical diagnostics as it can be collected non-invasively, often in large quantities, and requires minimal sample pre-treatment due to its low complexity and protein content. In addition, we and others have already shown that urine is an excellent reservoir of biomarkers (peptides, proteins and metabolites) of many diseases<sup>7-17</sup>.

Metabolomics studies mostly use NMR spectroscopy and liquid chromatography coupled to mass spectrometry (LC-MS) that provide complementary readouts<sup>4</sup>. NMR spectroscopy allows both identification and quantification of metabolites. It is a highly reproducible and non-destructive method which requires minimal sample preparation thereby minimizing contamination and maintenance issues and enabling the routine and

<sup>1</sup>Institut National de la Santé et de la Recherche Médicale (INSERM), U1048, Institut of Cardiovascular and Metabolic Disease, Equipe 12, 1 avenue Jean Poulhès, BP 84225, 31432 Toulouse Cedex 4, France. <sup>2</sup>Université Toulouse III Paul-Sabatier Toulouse, France. <sup>3</sup>HybridStat Predictive Analytics, Athens, Greece. <sup>4</sup>Unité de Recherche Clinique Pédiatrique, Module Plurithématique Pédiatrique, Centre d'Investigation Clinique - Hôpital des Enfants, Toulouse, France. <sup>5</sup>CHU Toulouse, Hôpital des Enfants, Service de Néphrologie – Médecine Interne – Hypertension Pédiatrique, Toulouse, France. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.P.S. (email: joost-peter.schanstra@inserm.fr) or B.B.-M. (email: benedicte.buffin-meyer@inserm.fr)



high-throughput analysis of hundreds to thousands of samples<sup>4,18</sup>. The inherent low sensitivity of NMR, however, restricts the detection limit to about  $1\ \mu\text{M}$ <sup>18,19</sup>. Moreover, the interpretation of NMR data is challenging<sup>18,20</sup>. In contrast, LC-MS allows the detection, quantification and structure elucidation of metabolites in the picomolar to nanomolar range of several thousand metabolites in a single measurement<sup>5</sup>. Unfortunately, the coupling of chromatographic separations with MS platforms requires an elevated level of maintenance, as the samples come in direct contact with many components of these platforms, contaminate surfaces and cause drift in the measured response and retention time over relatively short analysis periods<sup>4,5</sup>, thereby preventing the comparison of large numbers of samples. Relevant progress in the field of LC-MS was made with the introduction of ultra high performance liquid chromatography (UPLC) leading to improvement of analysis speed as well as sensitivity and resolution<sup>19,21,22</sup>. In particular, the potential of miniaturized UPLC-MS, based on the optimized use of microbore columns, was recently demonstrated for large-scale metabolomic studies<sup>23,24</sup>.

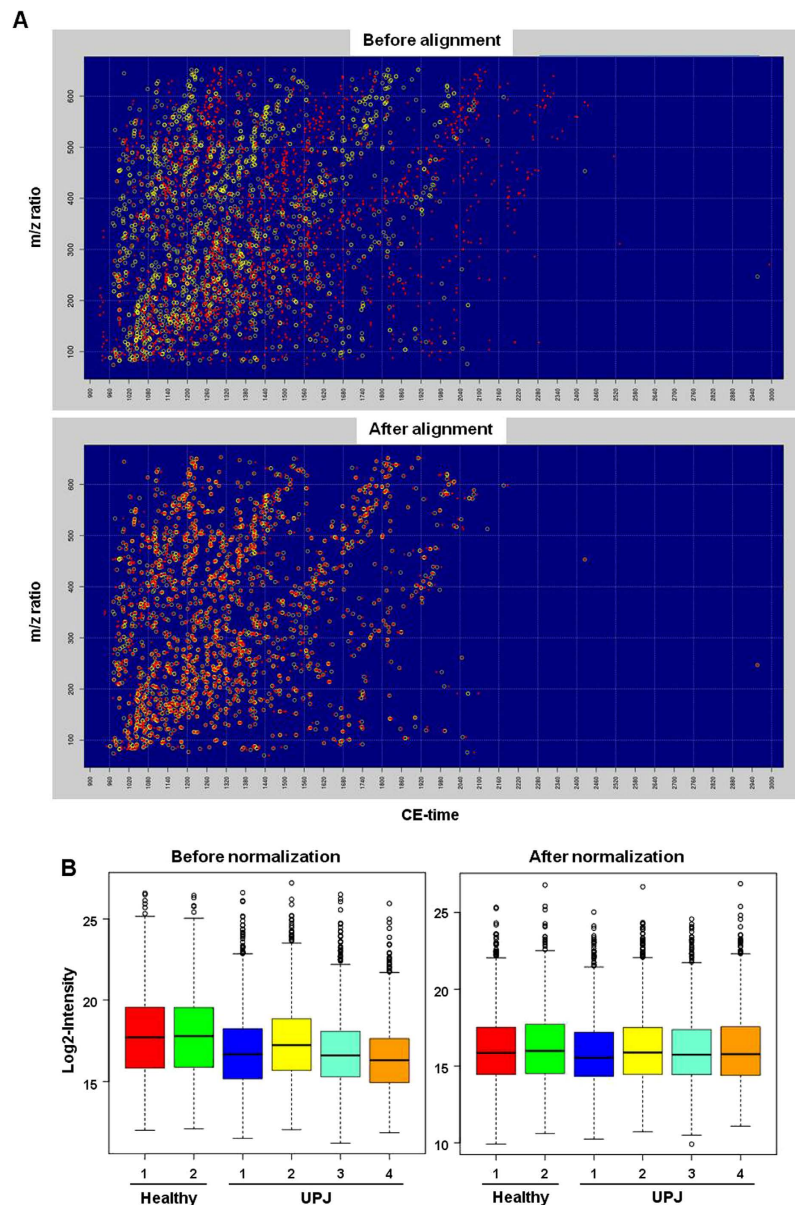
Until approximately ten years ago, capillary electrophoresis coupled to mass spectrometry (CE-MS) has only been rarely used for metabolome analysis. This was potentially due to issues related to stable coupling of CE to the MS instrument and the limited loading capacity of CE capillaries. However the significantly increased sensitivity of modern mass spectrometers and optimized methods for coupling of CE to MS have transformed CE-MS into a potential appropriate tool for profiling of disease associated metabolites in clinical relevant body fluid samples<sup>20,25–29</sup>. A number of recent studies now report the use of CE-MS for metabolome analysis of clinically relevant samples, with in particular those recently conducted by Soga and coworkers<sup>30–32</sup>, the first group to develop CE-MS for the comprehensive profiling of metabolites in biological samples<sup>33</sup>. However, the use of CE-MS for the discovery and validation of clinically relevant metabolic markers of human disease requires evaluation of its performance in terms of long term reproducibility and comparability.

Here, we present an optimized CE-MS setup and data analysis pipeline. Using a normalization procedure based on a set of “housekeeping” metabolites, this method allows to compare the metabolite content in urine samples analyzed over a period of several years. As proof of concept, we demonstrate the clinical relevance of this pipeline for the urinary metabolome based-detection of obstructive nephropathy in infants.

## Results

**Identification of metabolite internal standards for CE-MS normalization.** As a first measure towards improved comparison of large numbers of clinical samples over time, we developed a new method that allows to normalize the metabolite content of a biofluid sample. This method is based on the use of a set of persistent and stable metabolites across disease and healthy urine samples. In order to identify these so-called stable endogenous metabolites, 54 CE-MS runs of urine obtained from various kidney and urinary tract pathologies together with 21 control CE-MS runs of urine from healthy patients (Supplementary Table S1) were processed using the Bioconductor package *xcms*<sup>34</sup>. Each metabolite feature was identified by a unique identifier (ID) on the basis of the specific mass-to-charge ratio and migration time with a peak height representing the relative abundance. After preprocessing of the mass spectra (including mass calibration and migration time window restriction), the *xcms* pipeline (see Materials and Methods) identified 9642 distinct molecule features in terms of *m/z* and migration time pairs across all 75 samples. From this initial list, only features present (no-null abundance) in at least 50% of the total samples were considered for further analysis. The 6044 remaining metabolite features spanned a CE migration time from 16 to 50 min and a *m/z* range from 30–650. This reference dataset of 6044 metabolite features was then interrogated for the presence of stable molecule features, in terms of intensity, that would comprise the basis for a set of CE-MS internal normalization standards. For this, several established algorithms from the ‘rank invariant’ family of normalization methods present in the DNA microarray literature were deployed. Specifically, the Rank Invariant normalization method implemented in the *dChip* algorithm<sup>35</sup>, the Rank Invariant normalization algorithms for Illumina BeadArrays implemented in the *lumi* Bioconductor package<sup>36</sup> and the GRSN algorithm<sup>37</sup> were tested. However, each one of these suffered from several drawbacks, including among others unstable housekeeping sets because of their selection algorithm (*dChip*), selection preference in higher (*dChip*), lower (*lumi*) or medium (GRSN) intensities instead of spanning the whole metabolite abundance range, very high number of metabolites to achieve proper normalization (*lumi*) or poor normalization efficiency (*dChip*). The failure of present methodologies (partially due to the different nature of CE-MS data as compared to microarrays) to detect a stable set of metabolites led to the development of two new different internal standard selection strategies. Specifically, the first approach used the residuals of Robust Linear Regression models<sup>38,39</sup> to identify sets of metabolites presenting low variability across samples and the second, more geometrical than statistical, approach was based on the Euclidean distance of each metabolite abundance vector from the identity ‘hyperline’ in the sample space. The final set of stable metabolites for each method was derived using a Forward Selection procedure with the purpose of finding the smallest possible subset of metabolites with the greater normalization power (detailed description of the methods in the ‘Materials and Methods’ section). The method that was finally followed was the geometrical approach as it was found to yield more robust results in terms of metabolite intensity coverage, normalization power, smaller number of stable metabolites and its application did not require any assumptions for a baseline as compared to the RLM approach which requires a baseline. This led to the identification of 267 endogenous housekeeping metabolic features among the 6044 features detected (Supplementary Table S2) which spanned a CE migration time from 17 to 36 min and a *m/z* range from 82 to 650. These stable endogenous metabolite features were implemented in the CE-MS normalization pipeline. Hence, the CE migration time is normalized in a first step (Fig. 1A) followed by normalization of the metabolite abundance using the endogenous housekeeping metabolic features, as exemplified on a random selection of six samples (Fig. 1B).

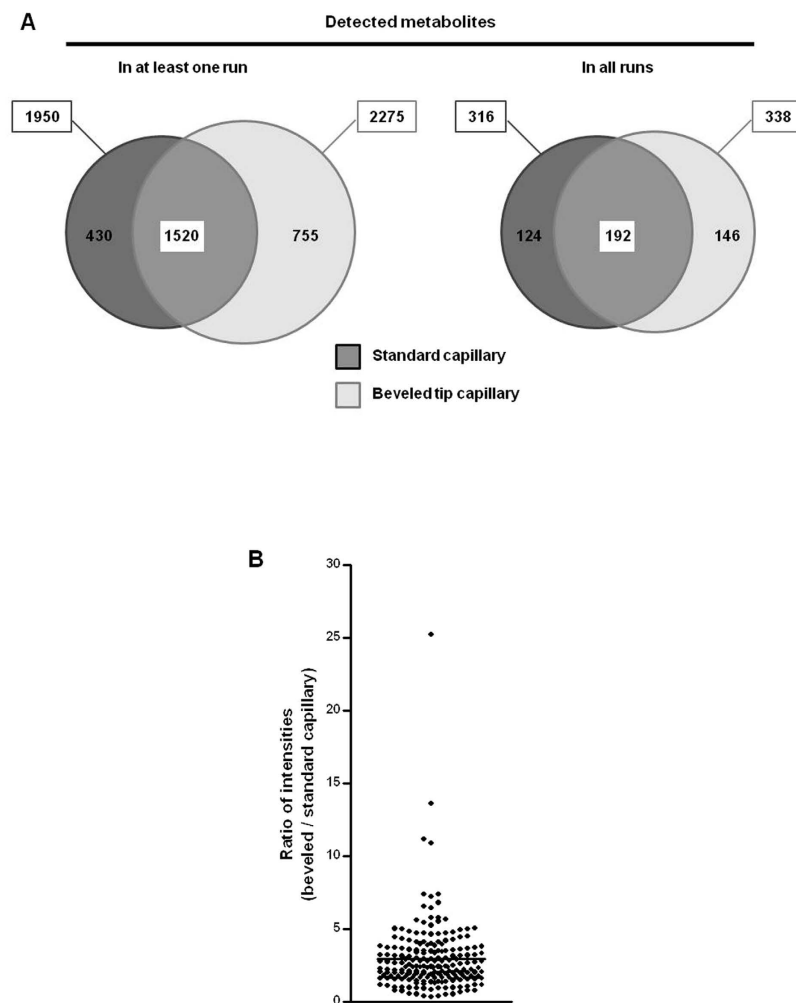
**Use of a beveled capillary improves the sensitivity of metabolite detection.** CE coupling to MS via electrospray ionization (ESI) can be performed using either a sheathless or a sheath flow interface<sup>40</sup>. The use



**Figure 1. Processing and normalization of samples.** Urinary samples were analyzed in CE-MS, processed and then normalized using the stable endogenous metabolites-based procedure described in the Materials and Methods section. **(A)** Representative distribution profile of urinary metabolite features before and after migration time alignment against reference dataset. Each circle is a unique peak processed with xcms. Red: metabolite features detected in a random urine sample and matching the reference; yellow: equivalent features in the reference dataset. **(B)** Box-whisker plot for metabolite abundance of exemplary healthy (2) and UPJ obstruction (4) patients before and after intensity normalization.

of sheathless systems is promising. In particular, the potential usefulness of a sheathless porous tip interface for CE-MS has been recently demonstrated for the analysis of the urinary metabolome<sup>28,29</sup>. Nevertheless this porous tip has not yet been adopted as a routine method for CE-MS coupling. So far, the sheath flow interface has been most widely used for CE-MS in metabolomics<sup>26,40,41</sup>. This type of coupling is stable and provides good sensitivity, its implementation is relatively easy and allows using a wide range of buffers. However, the CE-effluent is diluted in this configuration, thereby reducing the achievable sensitivity of the method<sup>28,29,40,41</sup>. As part of a continuous effort to improve the interface between CE and MS, Tseng *et al.*<sup>41</sup> have developed a beveled tapered tip emitter in order to reduce the sheath flow leading to decreased sample dilution. By analyzing synthetic drugs and triazine mixtures, they demonstrated that the use of beveled tip provides better sensitivity for detection than conventional sheath liquid interface which uses flat capillary tips<sup>41</sup>.

Therefore in an attempt to optimize the sensitivity of the detection of urinary metabolites, we compared the performance of a standard flat tip and a beveled tip sheath-liquid ESI interface. A QC urine sample was analyzed by CE-MS using either a standard (ten consecutive runs) or beveled capillary (ten consecutive runs) for CE. Of

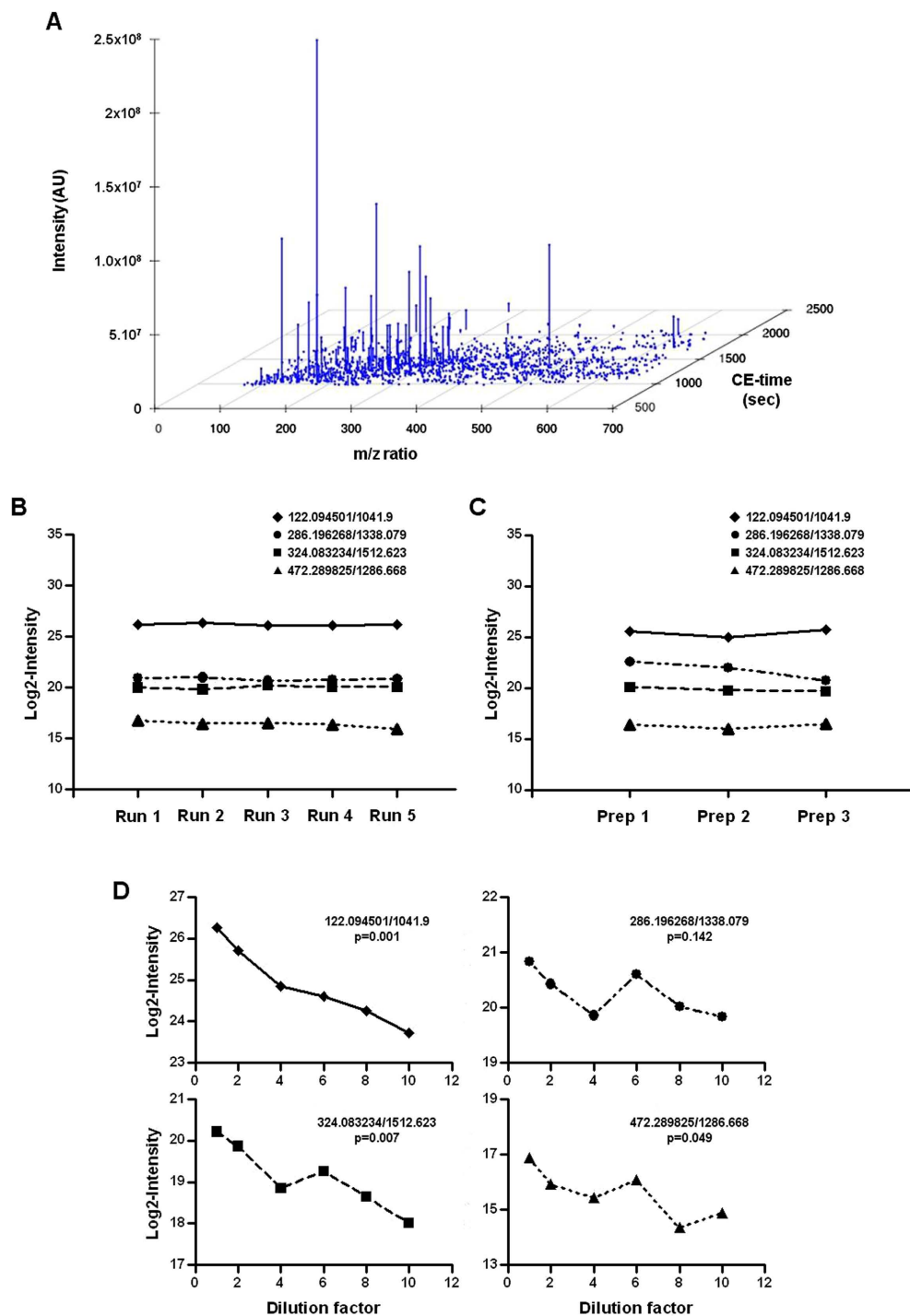


**Figure 2. Effect of the capillary on sensitivity of metabolite detection.** The same sample was analyzed in CE-MS using either standard (10 times) or beveled tip capillary (10 times) for CE. **(A)** Euler diagrams showing for each capillary the number of metabolite features detected at least once (left) or every time (right). Dark gray: standard (flat tip) capillary; light gray: beveled tip capillary. **(B)** For each metabolite detected in every run and with both types of capillaries ( $n = 192$ ), the mean intensity was calculated and then the ratio between intensity measured with beveled tip capillary and intensity measured with classical capillary was calculated. Graph shows the mean ratio  $\pm$  SEM, indicating that metabolite detection was more sensitive with beveled tip than with standard capillary.

note, the previously described 267 stable endogenous metabolites required for normalization procedure were identified using a beveled tip. After normalization, 2275 and 1950 distinct molecule features were detected in at least one run using the beveled tip and standard flat tip, respectively (Fig. 2A). Moreover, 338 and 316 metabolite features were detected consistently in all ten runs using the beveled tip and conventional tip, respectively (Fig. 2A). Although the absolute number of features detected is only slightly higher using the beveled tip, comparison of the intensities of 192 features detected in all runs with both types of capillary revealed a significant 3 fold gain in sensitivity using the modified capillary (Fig. 2B). Of note, robustness of the beveled tip was not decreased compared to flat tip (resisting to 40–50 runs [data not shown]). Therefore, the use of beveled tip as sheath-flow interface for CE-MS displays increased sensitivity towards the detection of urinary metabolites. We used the beveled tip for the remainder of the experiments.

**QC-based validation of CE-MS pipeline for urine metabolome profiling.** In order to estimate the analytical variability of the CE-MS pipeline, a set of experiments for validation was performed: repeatability (intra-assay precision), postpreparation stability, postdilution stability, and long-term (intermediate) precision were evaluated.

Repeatability expresses the precision under the same operating conditions over a short interval of time. Repeatability of the CE-MS pipeline was examined by analyzing the QC urine sample in five consecutive runs, covering a total run time of  $\approx 8$  h. Among 6044 potential metabolites, 1342 (22%) features were detected on average in each run. Figure 3A shows a typical plot of a CE-MS analysis of a QC sample, giving an indication of the distribution of mass-to-charge ratio and CE migration times encountered for this typical sample. To obtain



**Figure 3. Short term performance characteristics of metabolomic CE-MS platform.** The data from QC analyses were investigated to assess intra-assay precision, postpreparation stability and postdilution stability for molecule intensities. (A) Typical plot from the CE-MS analysis of the QC sample: Each metabolite was identified by a unique identifier (ID) on the basis of the specific mass-to-charge ratio and migration time. Graph shows the distribution of metabolite mass-to-charge ratio ( $m/z$ ) with CE-migration time for a representative QC injection. (B) Short term precision: The QC was analyzed in five consecutive runs and the intensity in each run was shown for four exemplary randomly selected metabolite features. The coefficient of variance (CV) for amplitude was between 0.7 and 1.9% for these individual features, thereby demonstrating the repeatability in peak height. (C) Variability according to preparation: QC sample was prepared on three different dates using different lots of buffer, and then analyzed in consecutive runs. The intensity in each run was shown for four exemplary randomly selected metabolite features. The obtained CV for abundance was between 1.1 and 4.3%, showing a stability depending of the preparation. (D) Stability according to dilution: QC sample was prepared at different concentrations and then analyzed in consecutive runs. The intensity of four exemplary randomly selected metabolite features was plotted against the dilution factor.

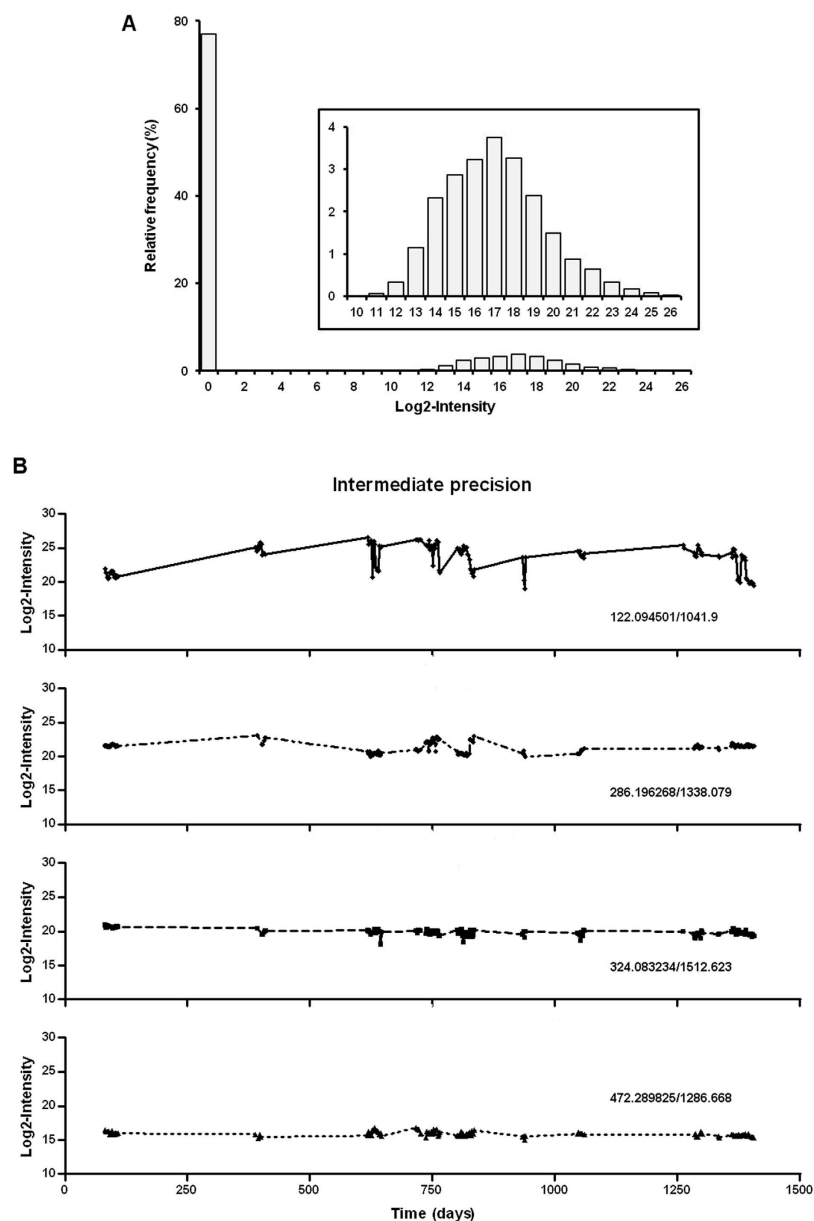
information on the run-to-run precision, four metabolite features were randomly selected for evaluation of intensity variation. The abundance variation of these four metabolite features was found to be negligible (Fig. 3B), with coefficient of variation (CV) values less than 2%, thereby indicating high performance of CE-MS platform in terms of repeatability. Next, the effect of different sample preparations was studied (post-preparation stability). We prepared QC sample according to the same procedure but using three different lots of buffer before CE-MS analysis in 3 consecutive runs. As shown in Fig. 3C, the intensity of the four exemplary selected metabolite features was constant in preparations, with a low CV, below 4.3%. Third, in order to test linearity of detection, the QC sample was prepared at six different concentrations and then analyzed by CE-MS in consecutive runs. Figure 3D depicts the abundance of the four randomly selected molecule features as a function of the dilution factor of a urine sample. For three of them, a significant negative correlation was observed between dilution and abundance whereas only a trend was observed for the fourth (Fig. 3D), thereby suggesting the relative stability of CE-MS platform when urine samples are diluted.

Finally, we evaluated intermediate precision of CE-MS platform which expresses the precision within laboratory variations. This assay involved analysis of QC urine metabolites at different days by different operators over a long period of time. It included different lot numbers of buffers, solvents and chemicals and also implies annual maintenance service of both CE and MS devices. This evaluation is important in the field of clinically useful metabolite biomarkers where durable use of CE-MS is necessary. For the long-term stability assay, the QC sample was analyzed repeatedly 128 times over a range of 4 years (from 2011 to 2014). Among 6044 potential metabolite features, 1389 (23%) were detected on average in each run, this result being similar to the previously reported value. A mean of 67.7% of all metabolite features and 30.5% of the stable endogenous metabolites in the QC samples from these 128 runs matched against the reference dataset. The analysis of our data set revealed that the distribution of intensities is bimodal, with a strong proportion of values at a point-mass at zero (*point-of-mass values* [PMVs] corresponding to missing values [NaN], zero intensity data being treated as missing data) and a continuous component (Fig. 4A). The occurrence of zero component in the data matrix is a recurrent issue encountered in MS data<sup>42</sup>. The origin of PMVs may either be biological, eg absence of a specific metabolite in biological sample, or technical, eg the inability of the mass spectrometer to detect the specific metabolite or of the algorithm to identify the peak. Next, as it is recommended that the coefficient of variation should not exceed 15%<sup>4,43</sup>, we examined CE-MS results using similar acceptance criteria as a means of determining the quality of the data. For this, the abundance of four exemplary randomly chosen molecule features was plotted over time (Fig. 4B). The statistical spread for these metabolite features was between 2.2 and 8.6%, indicating that CE-MS platform exhibits long-term stability. In addition, CV of intensities was calculated for all metabolite features across the QC samples. A data subset was considered including features which were detected in at least one of the 128 QC injections (4879 entities) and different filters of selected metabolites were considered to evaluate improvement of the proportion of peaks being acceptable. Using this subset, we observed that 4487 (92%) of the 4879 molecule features displayed a variation of  $\leq 10\%$ , whilst 2892 (59%) exhibited a variation of  $\leq 5\%$  level. Altogether, these results demonstrated the long-term stability of CE-MS platform and thus suggest that the optimized CE-MS setup and analysis pipeline allows to compare the metabolite content in urine samples regardless of the time of analysis.

**CE-MS for clinical metabolomics: application to diagnosis of UPJ obstruction.** Next we analyzed the capacity of the aforementioned pipeline in clinical research for the identification of diagnostic/prognostic biomarkers of disease. Newborns with UPJ obstruction were chosen for our proof of principle study. Two different cohorts of infants were employed: one discovery cohort ( $n = 49$ ) for the identification of urinary metabolite biomarkers of UPJ obstruction (15 healthy newborns and 34 patients with UPJ obstruction; Table 1 and Supplementary Table S3) and one cohort ( $n = 24$ ) for the blinded validation of urinary biomarkers (7 healthy newborns and 17 patients with UPJ obstruction; Table 2 and Supplementary Table S4). All urine samples were analyzed by CE-MS for their metabolite content and normalized using the above developed stable endogenous metabolites-based normalization procedure.

*Metabolic profiling of urine samples from patients with UPJ obstruction and healthy children.* The urinary metabolome of the discovery cohort, composed of 15 healthy children and 34 patients with severe UPJ obstruction (Table 1 and Supplementary Table S3) was studied by CE-MS. A mean of 42.0% of the stable endogenous metabolites in urine samples matched against the reference dataset. Among 6044 potential metabolite features, 1889 (31%) were detected on average in each sample. Only the features detected in at least 75% of the urine samples in each group (healthy and UPJ) were further investigated. This noise-filtering process reduced the number of features to 388 entities (Fig. 5A). The distribution of the metabolite intensities for all the 388 selected metabolite features showed, as for QC sample data, a bimodal distribution characterized by a proportion of PMVs (Fig. 5B) and a continuous component. In order to explore the origin of PMVs, metabolite features with consonant or dissonant differences were quantified. In the former case, the group with the higher proportion of PMVs has the smaller mean in the continuous part, while in the latter case the group with the higher PMV proportion also has the higher mean. An example of each type is shown in Fig. 5C. Although this definition does not distinguish between technical and biological PMVs, technical PMVs naturally correspond to consonant compounds whereas biological PMVs generally allow for both types<sup>44,45</sup>. The data employed here contains 357 (92%) consonant compounds, 15 (4%) dissonant and 16 (4%) without point-mass component. The high proportion of consonant markers associated with the low number of dissonant markers suggests that PMVs in present metabolomics data originated from technical considerations rather than biological (Fig. 5D).

*Identification of urinary metabolites associated to UPJ obstruction.* Comparing urinary metabolites from UPJ and healthy patients led to the identification of 32 adjusted (Benjamini and Hochberg<sup>46</sup>) differentially excreted



**Figure 4. Long term performance characteristics of metabolomic CE-MS platform.** The data from QC analyses were investigated to evaluate intermediate precision for molecule intensities. **(A)** Histograms of the distribution of abundance: The mean frequency of all features in QC sample was plotted against the logarithm (2) of the intensity. Profiles show a point-mass at zero and a continuous component. The zero component arises because the molecule features are either absent or their concentration is below the detection limit. Insert: magnification of the continuous distribution. **(B)** Long term variability: The QC sample was analyzed 128 times between 2011 and 2014. The intensity of four exemplary randomly selected metabolite features was plotted against the time.

metabolite features (Fig. 6A,B and Supplementary Table S5). Matching 32 features against databases (HMDB, ChEBI and KEGG) led to determination of real mass for 9 metabolite features; 5 of 9 were annotated for chemical formulas (Table 3). Of note, abundances of two compounds (227.111791/989.758 and 228.114334/990.108) corresponding to the same annotation were highly correlated ( $R^2 = 0.94$ ,  $p < 0.0001$ , data not shown). The 32 metabolite features of interest were then used to develop a support vector machine (SVM) discrimination model that we called “UPJMetab32”. Scoring the patients from the discovery cohort with the UPJMetab32 classifier clearly separated UPJ from healthy patients (Fig. 6C).

*Validation of UPJMeta32 in a separate, blinded cohort.* In the next step, following the recommendations for biomarker identification<sup>47</sup>, the UPJMetab32 model was validated in a separate, blinded study using urine from 7 healthy and 17 UPJ patients not used in the discovery cohort (Table 2 and Supplementary Table S4). These urine samples were analyzed by CE-MS and scored using the UPJMetab32 model (Supplementary Table S4).

	All patients	Healthy	UPJ obstruction
n	49	15	34
Gender			
M	46 (93.9%)	15 (100%)	31 (91.2%)
F	3 (6.1%)		3 (8.8%)
Age			
Mean (months)	2.25 +/- 0.27	2.29 +/- 0.62	2.22 +/- 0.29
Median (months)	1.45 (range 0 to 7.0)	1.58 (range 0 to 6.1)	1.45 (range 0.7 to 7.0)

**Table 1. Discovery cohort.**

	All patients	Healthy	UPJ obstruction
n	24	7	17
Gender			
M	21 (87.5%)	7 (100%)	14 (82.4%)
F	3 (12.5%)	0	3 (17.6%)
Age			
Mean (months)	2.51 +/- 0.58	1.35 +/- 1.21	2.99 +/- 0.65
Median (months)	1.28 (range 0 to 8.6)	0.03 (range 0 to 8.6)	1.61 (range 0.8 to 8.6)

**Table 2. Validation cohort.**

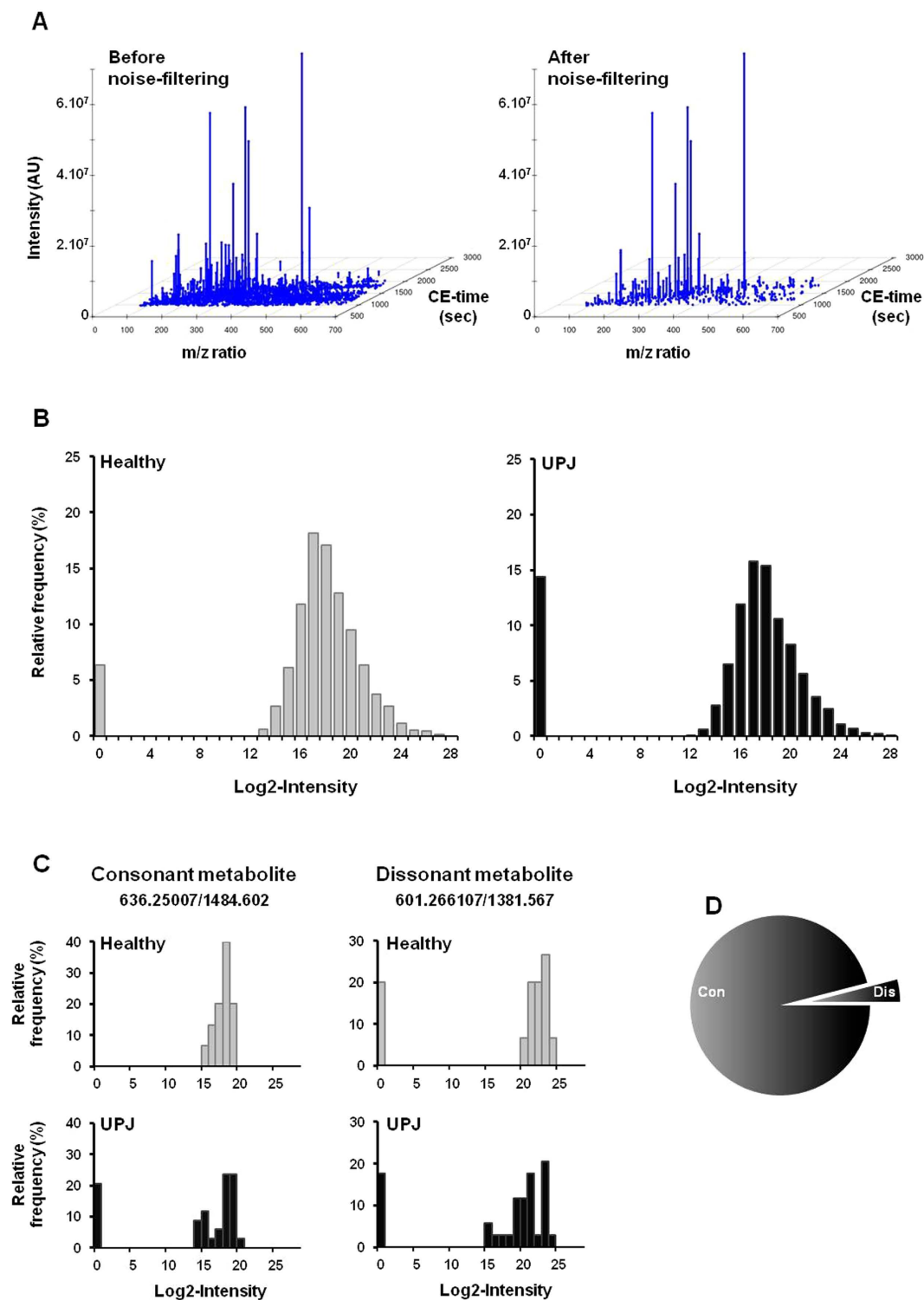
A UPJMetab32 score  $>0$  predicts patients with UPJ obstruction. These predictions were compared to the clinical criteria based status. The UPJMetab32 classifier diagnosed clinical status (healthy versus UPJ) with a sensitivity of 76.5%, a specificity of 85.7%, and an area under the curve (AUC) of 0.90 [95% CI: 0.707 to 0.984] (Fig. 7A). The UPJMetab32 model predicted 13 out of 17 UPJ cases correctly, showing the efficacy of the model to detect patients with severe UPJ. In addition, it predicted 6 out of 7 control cases correctly. The distribution of the UPJMetab32 scores for the validation cohort showed significant separation of the two patient populations (Fig. 7B).

## Discussion

We have explored the use of CE-MS and endogenous stable urinary metabolites for long-term, reproducible and comparable analysis of the urinary metabolome. The developed pipeline allowed comparison of urinary metabolite content analyzed over a 4 year timespan. As proof-of-concept we have used this pipeline to discover and validate urinary metabolites associated to a frequently encountered renal pathology in newborns.

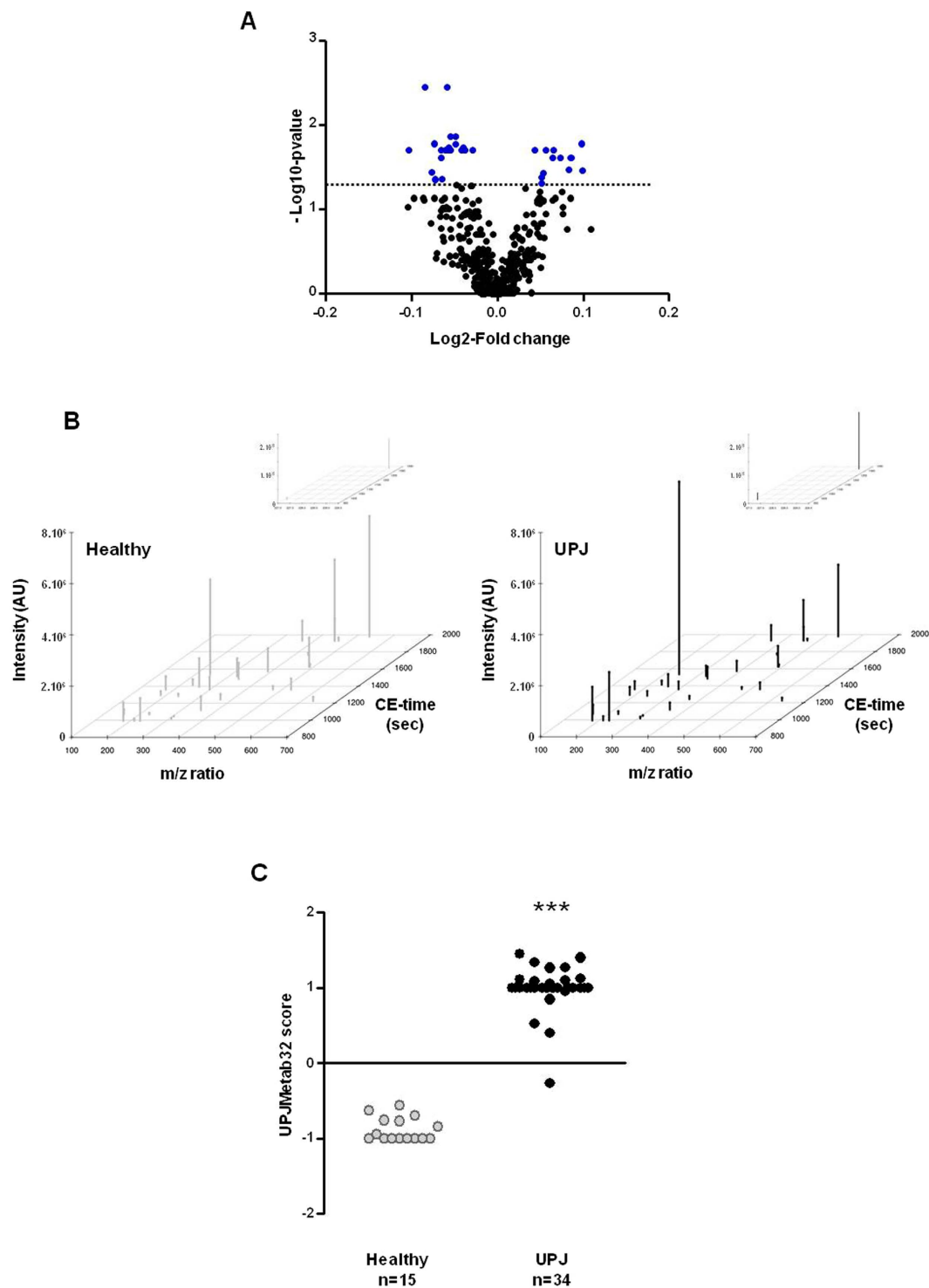
Clinical metabolomics aims at the detection of clinically useful metabolites that can be extracted from a diverse range of sample types. Amongst those samples, easily accessible bodyfluids like urine and blood are most suited for clinical use. Although the field of metabolomics has advanced significantly in the past 10 years<sup>4</sup>, there has been little progress in the identification of clinically useful urinary metabolite biomarkers. To enable the discovery and the validation of diagnostic/predictive biomarkers, medium-to-large-scale epidemiological studies are required in order to take into account the substantial diversity observed in physiology/physiopathology, metabolic status and lifestyle in the general human population. This involves the use of analytical methods able to analyze large numbers of samples over periods of many months or years with both high reproducibility and high sensitivity<sup>4</sup>. We explored the potential of CE which offers multiple advantages: (i) as CE separates compounds on the basis of their charge and size<sup>25</sup>, it demonstrates high-resolution power for separation of small ionogenic metabolites which are important constituents of the urinary metabolome; (ii) CE separations require a low sample volume and consume very little solvent<sup>25</sup>, thereby reducing the matrix effect that can cause ion suppression and then insufficient ionization and lower peak intensity in MS; (iii) CE displays high reproducibility when analyzing large numbers of samples since no gradients are applied. Indeed, we observed high stability of urinary metabolite abundance when analyzing the same sample nearly 130 times over a range of 4 years. A few studies report stability evaluation of pipelines, such as for example over 535 runs covering a timespan of 5 months (GC-TOF-MS<sup>48</sup>) or over 120 runs covering a timespan of 3 years (UPLC-TOF-MS<sup>49</sup>). However, such a long term assessment of reproducibility and comparability is only rarely performed. Hence our 4 years proof of stability of the developed pipeline, associated with its use in the UPJ obstruction, validates its potential use in the clinic field.

Establishing long term stability has therefore been a major objective of the study. Although CE-MS is a reproducible analytical tool, some variations induced by sample concentration (especially for urine where individual urine outputs are dependent of water uptake, diet, ...), interfering compounds and injection volume differences might still be observed. Several normalization strategies, such as normalization to creatinine, osmolarity and total area normalization are frequently employed in urine metabolomics studies. However, these commonly used normalization methods are not well adapted. For example, the creatinine level can be impacted by factors such as kidney function impairment, gender difference, and lean body mass<sup>50,51</sup>. The osmolarity normalization procedure is often affected by insoluble components, such as urine particles<sup>50,52</sup>. Adjusting the total peak area might yield biased results since the background noise and ion suppression due to the matrix may greatly interfere with the



**Figure 5. Metabolomic CE-MS analysis in urine of patients with UPJ.** The urine metabolome of 15 healthy and 34 UPJ patients of the discovery cohort was analyzed. (A) Representative figure showing abundance of the CE-MS detected-urinary metabolite features: on the left, before application of a filter; on the right: after selection of features present in at least 75% of the samples in each group. (B) Histograms of distribution: The frequency of all metabolite features in healthy and UPJ samples was plotted against the logarithm (2) of the intensity. As for QC sample data, profiles show a point-mass at zero and a continuous component. (C) Histograms of distribution of two selected metabolite features from example dataset. Metabolite feature ID: 636.25007/1484.602 (left): consonant; metabolite feature ID: 601.266107/1381.567 (right): dissonant. (D) Repartition of compounds with consonant and dissonant differences between healthy and UPJ obstruction groups.





**Figure 6. Identification of a classifier: UPJMeta32.** The urine metabolome of 15 healthy and 34 UPJ patients (discovery cohort) was analyzed. **(A)** Volcano plot showing fold-changes (Log<sub>2</sub>) between UPJ obstruction and healthy groups as well as statistical significance (-Log<sub>10</sub> of p-value) for 388 considered metabolite features. The dashed line shows where  $p = 0.05$ . Points above the line had  $p < 0.05$  and corresponding metabolite features (32) have been considered as significantly differentially excreted by UPJ patients. **(B)** Compared abundance of the 32 urine metabolite features which were identified as differentially excreted between UPJ patients and healthy subjects in the discovery cohort. Insert: two strongly abundant metabolite features **(C)** Cross-validation score of an SVM metabolite model, called UPJMeta32, consisting of 32 differentially excreted metabolite features. \*\*\* $p < 0.0001$  versus healthy subjects. Mann-Whitney test for independent samples.

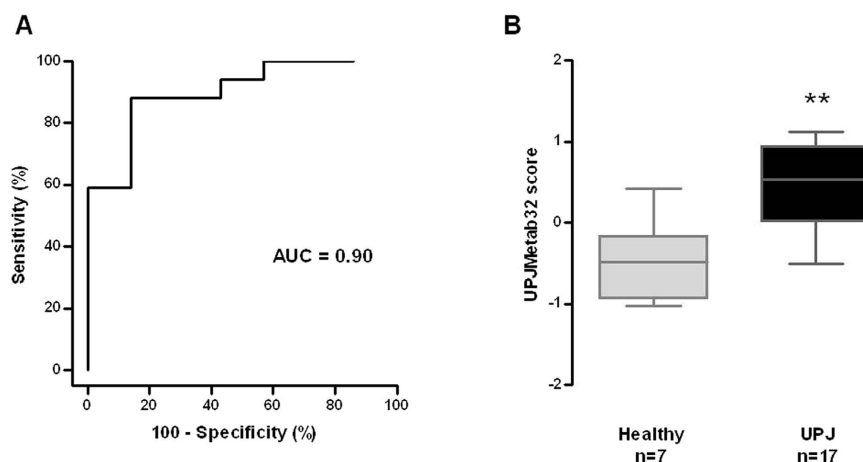
total signal. Furthermore, the total signal for samples with different metabolite distributions does not reflect the total concentration differences as ionization efficiency is compound dependent<sup>50</sup>. Variations can also be corrected

ID	Isotope	Adduct	Real Mass	Database	Database code	Proposed Formula	Proposed Name
227.111791/989.758	[56][M]+	[M+H]+	226.104515	HMDB	HMDB00033	C9H14N4O3	Carnosine
228.114334/990.108	[56][M+1]+			KEGG	cpd:C00386	C9H14N4O3	Carnosine
				HMDB	HMDB12482	C9H14N4O3	Hydroxypterin
				HMDB	HMDB00245	C10H14N2O4	Porphobilinogen
				KEGG	cpd:C00931	C10H14N2O4	Porphobilinogen
				KEGG	cpd:C02345	C15H14O2	(2S)-Flavan-4-ol
				KEGG	cpd:C15598	C15H14O2	Favan-3-ol
				KEGG	cpd:C09757	C15H14O2	7-Hydroxyflavan
				KEGG	cpd:C10276	C15H14O2	Pinosylvinmethylether
				KEGG	cpd:C10325	C15H14O2	Deoxylapachol
				KEGG	cpd:C13632	C15H14O2	4,4'-Dihydroxy-alpha-methylstilbene
				KEGG	cpd:C07205	C14H14N2O	Metyrapone
				ChEBI	55316	C7H16BrNO2	Acetylcholine bromide
				ChEBI	50426	H4O6P2S2	Disulfanediybis(phosphonic acid)
229.117309/1322.695	[8][M]+	[M+H]+	228.110033	HMDB	HMDB06695	C10H16N2O4	Prolylhydroxyproline
				KEGG	cpd:C13733	C10H16N2O4	(S)-ATPA
				KEGG	cpd:C10371	C15H16O2	MansononeC
				KEGG	cpd:C13624	C15H16O2	BisphenolA
				KEGG	cpd:C15210	C15H16O2	1,1-Bis(4-hydroxyphenyl)propane
				KEGG	cpd:C17424	C15H16O2	Lindenone
				ChEBI	58089	C5H11NO7P	5-phosphonato-D-ribosylaminium(1-)
				ChEBI	58681	C5H11NO7P	5-phospho-β-D-ribosylaminium(1-)
				KEGG	cpd:C18436	C9H16N4OS	Tebuthiuron
				ChEBI	53648	C7H4N2O7	2-hydroxy-3,5-dinitrobenzoic acid
355.071351/1117.064	[306][M]+		354.064075	KEGG	cpd:C01268	C9H15N4O9P	5-Amino-6-(5'-phosphoribosylamino)uracil
				KEGG	cpd:C02927	C15H14O10	2-Caffeoylsucitate
				KEGG	cpd:C07952	C17H19ClN2S.HCl	Chloropromazinemonohydrochloride
				KEGG	cpd:C12600	C19H14O5S	Phenolsulfonphthalein
488.133087/1622.375	[453][M]+		487.125811	KEGG	cpd:C02555	C26H21N3O5S	Luciferyl sulfate
				KEGG	cpd:C18429	C18H22FN5O8S	Flucetosulfuron
366.599792/1929.853	[461][M]+		365.592516				
438.677677/1763.369	[443][M]+		437.670401				
474.701959/1359.882	[359][M]+		473.694683				
526.161131/1357.627	[160][M]+		525.153855				

**Table 3. Annotation for potential chemical formulas and names.** HMDB: Human Metabolome Data Base; KEGG: Kyoto Encyclopedia of Genes and Genomes, ChEBI: Chemical Entities of Biological Interest.

by addition of exogenous standards but this method assumes that those are representative of the thousands of injected metabolites<sup>53</sup>. In the present study, we have opted for the selection of a set of most stable endogenous metabolites observed in a range of samples. This method offers several advantages. Firstly, for the selection of these stable endogenous compounds, we have chosen 75 urine samples potentially representing the diversity of (pediatric) diseases to be encountered in future studies. Therefore, we anticipate that the 267 derived stable endogenous metabolites can be used for the discovery of metabolite-based biomarkers in a number of pediatric diseases of the kidney and the urinary tract. Secondly, such a high number of stable endogenous metabolites for normalization spanning a CE migration time from 17 to 36 min and a m/z range from 82–650 allows that signal normalization can be performed 'locally', using metabolites with comparable ionization efficiency since close in terms of CE migration time and m/z ratio. In addition, this inclusion ensures that for every new sample, there will be a sufficient number of endogenous internal standards so as to span the whole intensity range of the new sample. Thirdly, as a result of this high number of stable endogenous metabolites, we observed that significant numbers of metabolite features are available for robust normalization in nearly all cases (we identified a mean of 42.0% of stable endogenous metabolites in the UPJ experiments). A potential drawback of the use of these endogenous metabolites for normalization could be that those are stable in the specific case of kidney disease and are excluded for the selection of biomarkers of disease. Selection of novel endogenous stable metabolites might thus be required in order to discover biomarkers for disease affecting other organs than the kidney/urinary tract.

Analysis of urinary metabolome is extremely attractive since changes reflect modifications of the entire organism in its equilibrium with the environment including particularly contributions from nutritive substances, drugs and gut microbial activities<sup>19</sup>. However, the variability induced by these factors can introduce a day-to-day intrapersonal variability as well as interpersonal differences, being a major drawback in studies aiming at disease diagnosis/prognosis. In order to address the sources of urinary metabolome variation throughout the day, Kim



**Figure 7. Validation of urinary metabolite classifier UPJMetab32 in a separate population.** The diagnostic value of the UPJMetab32 model was tested in an independent cohort (7 healthy subjects and 17 UPJ patients) by a blinded analysis. **(A)** ROC curve for the UPJMetab32 classifier. **(B)** Box-whisker plot for classification of healthy and UPJ patients in the validation set according to the UPJMetab32 score. \*\* $p < 0.005$  versus healthy subjects. Mann-Whitney test for independent samples.

*et al.*<sup>6</sup> have performed LC-MS metabolomics analysis of urine in subjects receiving a standardized and weight-based diet. The largest source of instability was attributable to technical issues such as sample preparation and analysis; to a lesser extent, an inconstancy subject-to-subjects as well as intrapersonal variability due to meals and time of day were observed; day-to-day fluctuation was minimal<sup>6</sup>. Despite that, several studies suggest the existence of a stable part (time scale: months to years) of the urine metabolomic profile which seems to be specific to each individual<sup>54,55</sup>. Under unrestricted lifestyle conditions, multiple collections of urine samples can be used to reduce the metabolic noise and retrieve the individual phenotype<sup>56</sup>. In the current study, differences in alimentation are most likely not a confounding factor since alimentation of newborns/infants is significantly less variable than in adults.

We have show-cased the use of the pipeline in a frequently encountered renal pathology in newborns<sup>11,57</sup>. We were able to identify 32 metabolic features associated to UPJ obstruction. Combination of the 32 metabolite features in a SVM classifier predicted with 76% sensitivity and 86% specificity UPJ obstruction in a separate validation cohort, thereby demonstrating the efficacy of the model to detect patients with UPJ obstruction. Increased carnosine excretion in UPJ was attributed to two highly correlating isotopes of a same metabolite. Carnosine is a dipeptide synthesized from alanine and histidine by the carnosine synthase in muscle, brain and other tissues such as kidney. It is degraded by the carnosinase predominantly in the liver but also in kidneys. Carnosine from animal food can also be absorbed in the small intestine, and at least part of it enters the blood intactly upon oral ingestion. Finally, kidneys filter plasma carnosine, reabsorb a part of carnosine *via* specific transporters and excrete the remaining in urine<sup>58,59</sup>. In order to understand the origin of the elevated urinary level of carnosine from UPJ obstruction patients, further experiments measuring expression of carnosine related-enzymes and transporter proteins in both obstructed and contralateral kidneys should be performed. The dipeptide possesses also strong antioxidant and free radical scavenging activities<sup>58</sup>. Interestingly, protective effects of carnosine have been demonstrated in rodent models of kidney disease<sup>60–62</sup> and in patients with diabetic nephropathy<sup>63</sup> or children with glomerulopathies<sup>64</sup>. Thus, increased urinary excretion of carnosine in UPJ obstruction could be an adaptive rather than a deteriorating mechanism.

In conclusion, we have developed a robust setup and analysis pipeline for the exploration by CE-MS of the metabolite content of urine and found that the long-term reproducibility of the metabolite data generated was excellent. As proof of concept, we demonstrated the feasibility to use CE-MS as a tool for the identification of clinically relevant urinary metabolites.

## Materials and Methods

**Patients and urine collection.** *Samples used for optimization of the CE-MS normalization procedure.* Fifty-four urinary samples from various kidney and urinary tract pathologies together with 21 control CE-MS samples from healthy patients (Supplementary Table S1) were used. We considered that these samples represent the potential diversity to be encountered in clinical samples and hence used those samples for the development of CE-MS normalization procedure.

**Quality control (QC).** The QC sample was a mixture of urine samples of 9 healthy individuals (3 females and 6 males, mean age  $34.1 \pm 2.8$  years).

**Ureteropelvic junction (UPJ) obstruction and healthy patients.** UPJ obstruction patients ( $n = 51$ ) and healthy individuals ( $n = 22$ ) of less than one year old were recruited in Toulouse Hospital and included in our study. The UPJ obstruction group was composed of patients scheduled for pyeloplasty with a pelvic dilatation of at least 16 mm and grade 3 and 4 hydronephrosis. Renographies were performed as soon as possible after birth,

generally between week 3 and 6 to establish baseline differential renal function (DMSAscan) and washout pattern (MAG3-scan). Healthy and UPJ obstruction patients were randomly divided into two cohorts: a discovery cohort ( $n = 49$ ; Table 1 and Supplementary Table S3) and a blinded cohort for validation ( $n = 24$ ; Table 2 and Supplementary Table S4). Mann Whitney analysis revealed no significant difference in the age of healthy and UPJ obstruction newborns included in both discovery and validation cohorts ( $p = 0.26$ ). In addition, the use of Chi Squared test also revealed no gender bias ( $p = 0.45$ ). Urine from newborns was collected in the morning during 30 min using a sterile pediatric urine collection pouch (B. Braun, Boulogne, France) during hospital consultation. Urine from healthy controls was collected from newborns in the maternity hospital and at home using the same sterile collection bags and a pair of gloves. Care was taken to not take the first morning urine. After collection, all urines were frozen within the hour at  $-20^{\circ}\text{C}$  both in the hospital (dedicated  $-20^{\circ}\text{C}$  freezer in the clinic) and at home. Transport was done using ice blocks in both cases and the samples were finally stored at  $-80^{\circ}\text{C}$  in the laboratory. The UPJ study was performed in accordance with the ethical principles in the Declaration of Helsinki and Good Clinical Practice. The study and its experimental protocols were approved by the ethics committee of the French Ministry of National Education, Higher Education and Research (number DC-2008-452). Written informed consent was obtained from all participants (parents of the newborns).

**Sample Preparation.** A  $170\ \mu\text{l}$  aliquot of urine was diluted with the same volume of a denaturing solution composed of 2 M urea, 0.0125%  $\text{NH}_4\text{OH}$ , 100 mM NaCl and 0.01% SDS. To remove higher molecular mass proteins, the sample was ultrafiltered using a Centriscan 20 kDa cut-off centrifugal filter device (Satorius, Göttingen, Germany) at  $2000 \times g$  for 45 min at  $4^{\circ}\text{C}$ . In order to remove urea, electrolytes and SDS,  $200\ \mu\text{l}$  of filtrate was applied onto a NAP5 gel filtration column (GE Healthcare Bio Sciences, Uppsala, Sweden), washed and then eluted with  $700\ \mu\text{l}$  of 0.01%  $\text{NH}_4\text{OH}$ . Finally, all samples were lyophilized in a Savant speedvac SVC100H connected to a Virtis 3L Sentry freeze dryer (Fischer Scientific, Illkirch, France). At this step, samples can be stored at  $4^{\circ}\text{C}$  until use and re-suspended in HPLC grade water shortly before CE-MS analysis. The resuspension volume was adjusted to yield  $1\ \mu\text{g}/\mu\text{l}$  protein as measured by BCA assay (Pierce Biotechnology, Rockford, USA).

**CE-MS analysis.** CE-MS analyses were performed as previously described<sup>11,12,65</sup> using a Beckman Coulter Proteome Lab PA800 capillary electrophoresis system (Beckman Coulter, Fullerton, USA) on-line coupled to a microTOF II MS (Bruker Daltonic, Bremen, Germany). The electro-ionization sprayer (ESI, Agilent Technologies, Palo Alto, CA, USA) was grounded, and the ion spray interface potential was set between  $-4$  and  $-4.5$  kV. The CE separation buffer contained 20% (v/v) acetonitrile and 250 mM formic acid (Sigma-Aldrich) in HPLC-grade water. The CE-system was equipped with a 95 cm (internal diameter:  $50\ \mu\text{m}$ ) bare fused silica capillary. Two types of CE-ESI-MS interfaces were tested (see results section); either a flatted or a tapered and beveled needle surrounding the capillary terminus. Data and MS acquisition methods were automatically controlled by the CE via contact-close-relays. Spectra were accumulated every 2 s, over a range of  $m/z$  30 to 650.

**CE-MS sample preprocessing for stable endogenous metabolites identification.** After mass calibration using the measurement of sodium formate salts at the start of each run, the raw MS-data were converted into NetCDF format (<http://www.unidata.ucar.edu/software/netcdf/>) through the Bruker software (DataAnalysis version 4.0). The NetCDF files were filtered by excluding spectra corresponding to a migration time less than 520 or greater than 3650 seconds prior to preprocessing using the Bioconductor package *xcms*<sup>34</sup> as previously described<sup>25</sup>. All the standard *xcms* pipeline parameters were kept to their defaults apart from *steps* which was set to 3 and *bw* which was set to 20. In addition, the total number of migration time alignment iterations was set to 5, using the LOESS approach of *xcms*. The resulting molecule features derived from the execution of the *xcms* pipeline (in terms of  $m/z$  and migration time pairs) were further filtered for their presence across samples by including only those molecule features present in at least 50% of the total samples. The latter ensured the robustness of the initial set of molecule intensities which would be later interrogated for the presence of stable (in terms of intensity) molecule features that would serve as a set of CE-MS internal normalization standards.

**Stable endogenous metabolites identification.** The final filtered set of *xcms* preprocessed and identified  $m/z$  – migration time pairs was further interrogated for the potential presence of a set of ‘housekeeping’ metabolites with stable intensity across pathologies and spanning the whole intensity range. To this end, a subset of ‘rank invariant’ family of normalization algorithms from the DNA microarray literature was applied with the purpose of identifying stable molecule features that would represent the ‘invariant set’ as referenced in the microarray bibliography<sup>66</sup>. Specifically, the algorithms described in<sup>35</sup> (*dChip* algorithm)<sup>36</sup>, (*lumi* Bioconductor package) and<sup>37</sup> (*GRSN* algorithm) were applied and sets of rank invariant metabolite abundances were retrieved. However, graphical assessment of the performance of these algorithms (see main text) revealed that the nature of CE-MS data prohibited the usage of these algorithms for the identification of a set of internal standards. Therefore, the following two strategies were applied:

1. The first strategy is based on the assumption that the majority of identified metabolites do not present differential abundance across samples (a similar assumption made for the normalization methods in the DNA microarray literature) and as a result, the relationship among different sample abundances is close to linear, after *xcms* preprocessing. Specifically, this approach includes the fitting of a set of Robust Linear Regression models<sup>38,39</sup>, either among all possible sample pairs, or for all samples against a baseline (e.g. the median metabolite abundances across samples) and the calculation of each model residuals. The set of stable

- metabolites is iteratively constructed by aggregating those ones whose abundance presented very low residuals in each model, implying low divergence from the model and subsequently, among samples.
- The second strategy does not make any assumptions about the differential abundance distribution of the metabolites but requires noise preprocessing, as performed by the xcms pipeline and is based on the geometrical distance of each metabolite abundance vector in the sample space from the identity 'hyperline'. Specifically, this approach includes the construction of the identity 'hyperline'  $\vec{Y} = \vec{X}$ , in the  $n$ -dimensional sample space, where  $\vec{Y} = (y_1, y_2, \dots, y_n)$ ,  $\vec{X} = (x_1, x_2, \dots, x_n)$ , and  $n$  is the number of samples. Then for each metabolite abundance vector  $A_i = (a_{i1}, a_{i2}, \dots, a_{in})$ , the Euclidean distance  $d_i$  from an equally spaced grid distributed along  $\vec{Y} = \vec{X}$  is calculated. The set of stable metabolites is constructed by aggregating metabolites with small  $d_i$  which imply both very high correlation as well as low inter-sample variability.

In both strategies (i) and (ii), the optimal number of metabolites with stable abundances is selected according to the normalizing potential of forward selected subsets of metabolites. The Forward Selection approach was selected as the number of stable metabolites should be kept to the minimum possible also for later purposes (exploration of prognostic or diagnostic values). Specifically, the initial candidate list is constructed by retrieving the first 1000 metabolites with the smallest Euclidean distance from the identity hyperline (or with the smallest residual value from an RLM) and sorting it in ascending order (of distance or residual value). Then, starting from a minimum set  $S$  of 10 metabolites, the whole dataset is normalized by fitting a LOESS curve  $L$  in this set and using it as the normalization reference. In each iteration one member of the stable metabolite candidate list is added to  $S$ ,  $L$  is recalculated, the whole dataset is normalized and the following dataset variability metric is calculated:

$$M = MAD(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n), \text{ where } = (x_{1j}, x_{2j}, \dots, x_{mj})$$

and  $x_{ij}$  the normalized abundance of metabolite  $i$  in sample  $j$ ,  $i = 1, \dots, m$  ( $m$  the number of metabolites in the total dataset),  $j = 1, \dots, n$  ( $n$  the number of CE-MS samples).  $M$  reflects the total variability of the normalized intensity matrix by firstly summarizing each column (sample) by taking its median value and then calculating the variability of this summarization, by taking the Median Absolute Deviation (MAD) of the column medians distribution. The final number of the stable metabolites is the size of  $S$  that minimizes  $M$  and has thus the best normalizing potential while at the same time being as small as possible.

**Processing and normalization of new samples.** New CE-MS urine samples are preprocessed up to filtering (exclusion of spectra corresponding to a CE-time less than on average 840 [sodium salts] or greater than 3000 seconds) and peak-picking (no migration time alignment) with xcms as described above. Then, the masses of the new samples are matched against the reference dataset (consisting of 75 disease and control runs as described above) with a tolerance of 0.01 mass units, and the molecule features that do not match the reference are excluded from further analysis. The migration time alignment of the new samples is performed with an iterative procedure, similarly to the one followed by the xcms package but using the urine specific internal standards instead of the ones that are identified for independent datasets by xcms. Specifically, the migration times of the internal standards subset which is specific to the new sample (identified as described above) and span the whole range of the new sample's migration times, are used as seeds for the creation of migration time clusters using *k-means* clustering with the  $k$  parameter equal to the number of matching internal standards. Then, a LOESS curve is fitted to each cluster and used as a reference for the alignment of migration times in each cluster. The intensity normalization of the new samples is performed as described above ('stable endogenous metabolites identification' section), using the proper subset of the internal standards set according to the aforementioned mass match procedure.

**Metabolite features annotation.** The final set was matched against HMDB<sup>67</sup>, ChEBI<sup>68</sup> and KEGG<sup>69</sup> for known molecules and annotated for potential chemical formulas using the CAMERA Bioconductor package<sup>70</sup>. From the two aforementioned methods, RLM and identity hyperline, the latter was selected as it was found to yield more robust results in terms of metabolite intensity coverage ( $S$  contained features spanning a sufficient range of intensities), normalization power, cardinality of  $S$  and its application did not require any assumptions for a baseline.

**Statistical analysis. Biomarker identification and modelling.** For the identification of potential metabolite biomarkers, the normalized levels of urinary metabolite features were compared between the healthy and UPJ obstruction patient groups. Only molecule features that were detected with a minimal frequency of 75% in every of the discovery groups were investigated for statistical analysis. Missing values (recorded as "Not a Number" [NaN]) from the discovery cohort were replaced by the average of the metabolite intensities found in the corresponding group (UPJ obstruction patients or healthy newborns). However, in the validation cohort where the belonging of the sample is unknown, we used the mean abundance of all patients from discovery set as imputation methods for missing values. Of note, zero values were considered as missing values. P-values were calculated for the comparison between healthy and UPJ obstruction patient groups using the Wilcoxon test followed by adjustment for multiple testing using the method described by Benjamini and Hochberg<sup>46</sup>. Only metabolite features with a corrected  $p < 0.05$  were considered significant. Using an in-house developed tool, we next used a support vector machine (SVM)-based approach (SVM package e1071 of R)<sup>71</sup> to generate a prognostic biomarker classifier based on 32 biomarkers associated with UPJ obstruction. The parameters of the radial kernel function

(type C) for the multi-dimensional hyperplane were: cost parameter (C) of 1 and kernel width ( $\gamma$ ) of 0.03125. Sensitivity and specificity were calculated using receiver operating characteristic (ROC) plots *via* the software R.

**Comparison of svm scores.** Statistical analyses were performed using GraphPad Prism 5.0 for Windows (GraphPad Software Inc) and comparisons between two groups were assessed using a Mann-Whitney test for independent samples.  $p < 0.05$  was considered as statistically significant.

## References

- Decramer, S. *et al.* Urine in clinical proteomics. *Mol Cell Proteomics*. **7**, 1850–62 (2008).
- Frantzi, M. *et al.* Developing proteomic biomarkers for bladder cancer: towards clinical application. *Nat Rev Urol*. **12**, 317–30 (2015).
- Mischak, H. *et al.* Clinical proteomics: A need to define the field and to begin to set adequate standards. *Proteomics Clin Appl*. **1**, 148–56 (2007).
- Dunn, W. B. *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc*. **6**, 1060–83 (2011).
- Gowda, G. A. & Djukovic, D. Overview of mass spectrometry-based metabolomics: opportunities and challenges. *Methods Mol Biol*. **1198**, 3–12 (2014).
- Kim, K. *et al.* Mealtime, temporal, and daily variability of the human urinary and plasma metabolomes in a tightly controlled environment. *PLoS One*. **9**, e86223 (2014).
- Brown, C. E. *et al.* Urinary proteomic biomarkers to predict cardiovascular events. *Proteomics Clin Appl*. **9**, 610–7 (2015).
- Metzger, J. *et al.* Diagnosis of subclinical and clinical acute T-cell-mediated rejection in renal transplant patients by urinary proteome analysis. *Proteomics Clin Appl*. **5**, 322–33 (2011).
- Metzger, J. *et al.* Urine proteomic analysis differentiates cholangiocarcinoma from primary sclerosing cholangitis and other benign biliary disorders. *Gut*. **62**, 122–30 (2013).
- Zimmerli, L. U. *et al.* Urinary proteomic biomarkers in coronary artery disease. *Mol Cell Proteomics*. **7**, 290–8 (2008).
- Decramer, S. *et al.* Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis. *Nat Med*. **12**, 398–400 (2006).
- Klein, J. *et al.* Fetal urinary peptides to predict postnatal outcome of renal disease in fetuses with posterior urethral valves (PUV). *Sci Transl Med*. **5**, 198ra106 (2013).
- Schanstra, J. P. *et al.* Diagnosis and Prediction of CKD Progression by Assessment of Urinary Peptides. *J Am Soc Nephrol*. **26**, 1999–2010 (2015).
- Schonemeier, B. *et al.* Urinary Peptide Analysis Differentiates Pancreatic Cancer From Chronic Pancreatitis. *Pancreas* (2016).
- Posada-Ayala, M. *et al.* Identification of a urine metabolomic signature in patients with advanced-stage chronic kidney disease. *Kidney Int*. **85**, 103–11 (2013).
- Zhao, X. *et al.* Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. *Metabolomics*. **6**, 362–374 (2010).
- Wang, X. *et al.* Urine metabolomics analysis for biomarker discovery and detection of jaundice syndrome in patients with liver disease. *Mol Cell Proteomics*. **11**, 370–80 (2012).
- Emwas, A. H. *et al.* Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. *Metabolomics*. **11**, 872–894 (2015).
- Nicholson, J. K. *et al.* Metabolic phenotyping in clinical and surgical environments. *Nature*. **491**, 384–92 (2012).
- Ramautar, R. Capillary Electrophoresis-Mass Spectrometry for Clinical Metabolomics. *Adv Clin Chem*. **74**, 1–34 (2016).
- Gika, H. G., Macpherson, E., Theodoridis, G. A. & Wilson, I. D. Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples. *J Chromatogr B Analyt Technol Biomed Life Sci*. **871**, 299–305 (2008).
- Novakova, L., Matysova, L. & Solich, P. Advantages of application of UPLC in pharmaceutical analysis. *Talanta*. **68**, 908–18 (2006).
- Gray, N. *et al.* Development of a Rapid Microbore Metabolic Profiling Ultra-performance Liquid Chromatography-Mass Spectrometry Approach for High-Throughput Phenotyping Studies. *Anal Chem*. **88**, 5742–51 (2016).
- Gray, N., Lewis, M. R., Plumb, R. S., Wilson, I. D. & Nicholson, J. K. High-Throughput Microbore UPLC-MS Metabolic Phenotyping of Urine for Large-Scale Epidemiology Studies. *J Proteome Res*. **14**, 2714–21 (2015).
- Ramautar, R., Somsen, G. W. & de Jong, G. J. CE-MS for metabolomics: developments and applications in the period 2010–2012. *Electrophoresis*. **34**, 86–98 (2012).
- Ramautar, R., Somsen, G. W. & de Jong, G. J. CE-MS for metabolomics: developments and applications in the period 2012–2014. *Electrophoresis*. **36**, 212–24 (2015).
- Kuehnbaum, N. L., Kormendi, A. & Britz-McKibbin, P. Multisegment injection-capillary electrophoresis-mass spectrometry: a high-throughput platform for metabolomics with high data fidelity. *Anal Chem*. **85**, 10664–9 (2013).
- Ramautar, R., Busnel, J. M., Deelder, A. M. & Mayboroda, O. A. Enhancing the coverage of the urinary metabolome by sheathless capillary electrophoresis-mass spectrometry. *Anal Chem*. **84**, 885–92 (2012).
- Zhang, W., Hankemeier, T. & Ramautar, R. Next-generation capillary electrophoresis-mass spectrometry approaches in metabolomics. *Curr Opin Biotechnol*. **43**, 1–7 (2016).
- Harada, S. *et al.* Metabolomic profiling reveals novel biomarkers of alcohol intake and alcohol-induced liver injury in community-dwelling men. *Environ Health Prev Med*. **21**, 18–26 (2016).
- Kami, K. *et al.* Metabolomic profiling of lung and prostate tumor tissues by capillary electrophoresis time-of-flight mass spectrometry. *Metabolomics*. **9**, 444–453 (2013).
- Kimura, T. *et al.* Identification of biomarkers for development of end-stage kidney disease in chronic kidney disease by metabolomic profiling. *Sci Rep*. **6**, 26138 (2016).
- Soga, T. *et al.* Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal Chem*. **74**, 2233–9 (2002).
- Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*. **78**, 779–87 (2006).
- Li, J. & Wong, L. Emerging patterns and gene expression data. *Genome Inform*. **12**, 3–13 (2001).
- Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. **24**, 1547–8 (2008).
- Pelz, C. R., Kulesz-Martin, M., Bagby, G. & Sears, R. C. Global rank-invariant set normalization (GRSN) to reduce systematic distortions in microarray data. *BMC Bioinformatics*. **9**, 520 (2008).
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. *Robust Statistics: The Approach Based on Influence Functions*. (New York: Wiley) (1986).
- Huber, P. J. *Robust Statistics*. (New York: Wiley) (1981).
- Maxwell, E. J. & Chen, D. D. Twenty years of interface development for capillary electrophoresis-electrospray ionization-mass spectrometry. *Anal Chim Acta*. **627**, 25–33 (2008).

41. Tseng, M. C., Chen, Y. R. & Her, G. R. A beveled tip sheath liquid interface for capillary electrophoresis-electrospray ionization-mass spectrometry. *Electrophoresis* **25**, 2084–9 (2004).
42. Gleiss, A., Dakna, M., Mischak, H. & Heinze, G. Two-group comparisons of zero-inflated intensity values: the choice of test statistic matters. *Bioinformatics* **31**, 2310–7 (2015).
43. FDA & Industry, G. f. Bioanalytical Method Validation, Food and Drug Administration: A Guidance. *Centre for Drug Valuation and Research (CDER)* (2001).
44. Taylor, S. & Pollard, K. Hypothesis tests for point-mass mixture data with application to 'omics data with many zero values. *Stat Appl Genet Mol Biol.* **8**, Article 8 (2009).
45. Dakna, M. *et al.* Addressing the challenge of defining valid proteomic biomarkers and classifiers. *BMC Bioinformatics* **11**, 594 (2010).
46. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.* **57**, 289–300 (1995).
47. Mischak, H. *et al.* Recommendations for biomarker identification and qualification in clinical proteomics. *Sci Transl Med.* **2**, 46ps42 (2010).
48. Begley, P. *et al.* Development and performance of a gas chromatography-time-of-flight mass spectrometry analysis for large-scale nontargeted metabolomic studies of human serum. *Anal Chem.* **81**, 7038–46 (2009).
49. Zelena, E. *et al.* Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal Chem.* **81**, 1357–64 (2009).
50. Chen, G. Y., Liao, H. W., Tseng, Y. J., Tsai, I. L. & Kuo, C. H. A matrix-induced ion suppression method to normalize concentration in urinary metabolomics studies using flow injection analysis electrospray ionization mass spectrometry. *Anal Chim Acta.* **864**, 21–9 (2015).
51. Waikar, S. S., Sabbiseti, V. S. & Bonventre, J. V. Normalization of urinary biomarkers to creatinine during changes in glomerular filtration rate. *Kidney Int.* **78**, 486–94 (2010).
52. Chadha, V., Garg, U. & Alon, U. S. Measurement of urinary concentration: a critical appraisal of methodologies. *Pediatr Nephrol.* **16**, 374–82 (2001).
53. Wu, Y. & Li, L. Determination of total concentration of chemically labeled metabolites as a means of metabolome sample normalization and sample loading optimization in mass spectrometry-based metabolomics. *Anal Chem.* **84**, 10723–31 (2012).
54. Assfalg, M. *et al.* Evidence of different metabolic phenotypes in humans. *Proc Natl Acad Sci USA* **105**, 1420–4 (2008).
55. Bernini, P. *et al.* Individual human phenotypes in metabolic space and time. *J Proteome Res.* **8**, 4264–71 (2009).
56. Wallner-Liebmann, S. *et al.* The impact of free or standardized lifestyle and urine sampling protocol on metabolome recognition accuracy. *Genes Nutr.* **10**, 441 (2015).
57. Decramer, S., Bascands, J. L. & Schanstra, J. P. Non-invasive markers of ureteropelvic junction obstruction. *World J Urol.* **25**, 457–65 (2007).
58. Boldyrev, A. A., Aldini, G. & Derave, W. Physiology and pathophysiology of carnosine. *Physiol Rev* **93**, 1803–45 (2013).
59. Peters, V. *et al.* Intrinsic carnosine metabolism in the human kidney. *Amino Acids.* **47**, 2541–50 (2015).
60. Kurata, H. *et al.* Renoprotective effects of l-carnosine on ischemia/reperfusion-induced renal injury in rats. *J Pharmacol Exp Ther.* **319**, 640–7 (2006).
61. Riedl, E. *et al.* Carnosine prevents apoptosis of glomerular cells and podocyte loss in STZ diabetic rats. *Cell Physiol Biochem.* **28**, 279–88 (2011).
62. Yay, A. *et al.* Antioxidant effect of carnosine treatment on renal oxidative stress in streptozotocin-induced diabetic rats. *Biotech Histochem.* **89**, 552–7 (2014).
63. Janssen, B. *et al.* Carnosine as a protective factor in diabetic nephropathy: association with a leucine repeat of the carnosinase gene CNDP1. *Diabetes.* **54**, 2320–7 (2005).
64. Peters, V. *et al.* CNDP1 genotype and renal survival in pediatric nephropathies. *J Pediatr Endocrinol Metab* (2016).
65. Desveaux, C. *et al.* Identification of Symptomatic Fetuses Infected with Cytomegalovirus Using Amniotic Fluid Peptide Biomarkers. *PLoS Pathog.* **12**, e1005395 (2016).
66. Tseng, G. C., Oh, M. K., Rohlin, L., Liao, J. C. & Wong, W. H. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res.* **29**, 2549–57 (2001).
67. Wishart, D. S. *et al.* HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res.* **41**, D801–7 (2013).
68. Hastings, J. *et al.* The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.* **41**, D456–63 (2013).
69. Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42**, D199–205 (2014).
70. Kuhl, C., Tautenhahn, R., Bottcher, C., Larson, T. R. & Neumann, S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem.* **84**, 283–9 (2012).
71. Meyer, D., Dimitriadou, E., Hornik, K. A. W. & F., L. E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-7. <http://CRAN.R-project.org/package=e1071> (2015).

## Acknowledgements

This work was supported in part by a research grant from the French Ministry of Education, Research and Technology for VP, by the “Programme Hospitalier de Recherche Clinique” (PHRC) [number No. 06 223 01 - No. RCB 2007-A00854-49] project for NL, ST, SD, and by the “European Consortium for High-Throughput Research in Rare Kidney Diseases” [EUrenOmics, GA2012-305608] for JK, SD, JK, BBM.

## Author Contributions

J.-L.B., S.D., J.P.S. and B.B.-M. designed the experiments; F.B., V.B., P.M. and B.B. performed the experiments; F.B., V.B., P.M., J.K., J.P.S. and B.B.-M. carried out analysis and interpretation of data; B.B., N.L., C.C., S.T. and S.D. collected clinical data and banked human urines; P.M., J.P.S. and B.B.-M. wrote the manuscript. All authors have reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

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

**How to cite this article:** Boizard, F. *et al.* A capillary electrophoresis coupled to mass spectrometry pipeline for long term comparable assessment of the urinary metabolome. *Sci. Rep.* **6**, 34453; doi: 10.1038/srep34453 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016



## Supplementary Information

### **A capillary electrophoresis coupled to mass spectrometry pipeline for long term comparable assessment of the urinary metabolome**

Franck Boizard<sup>1,2,#</sup>, Valérie Brunchault<sup>1,2,#</sup>, Panagiotis Moulos<sup>3</sup>, Benjamin Breuil<sup>1,2</sup>, Julie Klein<sup>1,2</sup>, Nadia Lounis<sup>4</sup>, Cécile Caubet<sup>1,2</sup>, Stéphanie Tellier<sup>5</sup>, Jean-Loup Bascands<sup>1,2</sup>, Stéphane Decramer<sup>1,2,5</sup>, Joost P Schanstra<sup>1,2\*</sup>, Bénédicte Buffin-Meyer<sup>1,2\*</sup>

<sup>1</sup> Institut National de la Santé et de la Recherche Médicale (INSERM), U1048, Institut of Cardiovascular and Metabolic Disease.

<sup>2</sup> Université Toulouse III Paul-Sabatier Toulouse, France.

<sup>3</sup> HybridStat Predictive Analytics, Athens, Greece.

<sup>4</sup> Unité de Recherche Clinique Pédiatrique, Module Plurithématique Pédiatrique, Centre d'Investigation Clinique - Hôpital des Enfants, Toulouse, France

<sup>5</sup> CHU Toulouse, Hôpital des Enfants, Service de Néphrologie – Médecine Interne – Hypertension Pédiatrique, Toulouse, France.

# Franck Boizard and Valérie Brunchault contributed equally to this work.

\* Co-correspondence to:

[joost-peter.schanstra@inserm.fr](mailto:joost-peter.schanstra@inserm.fr) and [benedicte.buffin-meyer@inserm.fr](mailto:benedicte.buffin-meyer@inserm.fr)

**Supplementary Table S1: Characteristics of the cohort used for identification of the endogenous housekeeping metabolic features.**

---

<b>Gender</b>		
M	46	61.3%
F	29	38.7%

---

<b>Clinic symptoms of renal pathology</b>		
no renal pathology	21	28.0%
ureteropelvic junction obstruction	25	33.3%
pelvicalyceal dilatation	4	5.3%
posterior urethral valves	4	5.3%
megaureter	3	4.0%
pyelonephritis	3	4.0%
hyperechogen kidney	3	4.0%
ureteral hydronephrosis	3	4.0%
bilateral renal hypoplasia	2	2.7%
multicystic kidney	2	2.7%
unique kidney	2	2.7%
renal cystic dysplasia	1	1.3%
glomerulonephritis	1	1.3%
maternity	1	1.3%

---

**Supplementary Table S2: List of endogenous housekeeping metabolic features**

<b>ID</b>	<b>mass-to-charge ratio</b>	<b>CE-time (min)</b>	<b>Log2-intensity</b>
82.064873/1368.589	82,06487274	22,8098192	14,68
100.074386/1290.237	100,0743857	21,50395422	21,50
126.090263/1290.666	126,0902629	21,51110489	19,34
131.04454/1408.987	131,0445404	23,48310929	19,25
134.043624/1361.785	134,0436241	22,69641784	20,62
152.579489/1215.841	152,5794887	20,26401059	13,50
164.107226/1230.268	164,1072264	20,50447002	18,05
164.770715/1144.141	164,7707148	19,06901749	12,66
165.537382/1144.661	165,5373824	19,07768342	12,69
166.597455/1194.84	166,5974554	19,914004	13,16
179.098782/1060.965	179,0987821	17,68274446	19,37
191.064945/1361.393	191,064945	22,68987572	20,19
198.601243/1201.964	198,6012432	20,03273491	13,73
202.500848/1344.512	202,5008476	22,4085376	13,52
203.136101/1300.548	203,1361008	21,67580248	23,07
204.14024/1296.284	204,1402399	21,60474124	20,45
212.615842/1208.194	212,6158423	20,13656831	13,98
222.60553/1222.569	222,6055304	20,37615281	13,36
228.589951/1160.622	228,5899513	19,34369431	13,66
231.607345/1221.787	231,6073454	20,36311795	14,50
235.604147/1214.61	235,6041466	20,24350635	13,78
236.611114/1286.885	236,6111144	21,44807627	13,69
246.611275/1469.58	246,6112747	24,49299902	13,87
250.61946/1288.443	250,6194598	21,47405246	13,33
254.612435/1213.68	254,612435	20,22800526	13,78
255.622652/1222.513	255,6226521	20,37522417	13,07
256.614786/1262.341	256,6147862	21,03901713	14,60
260.189836/1017.354	260,1898359	16,95589886	21,14
260.740627/1465.128	260,7406272	24,41880129	14,03
261.25445/1017.45	261,2544502	16,95750329	15,26
262.112181/1421.234	262,1121809	23,68723898	23,07
265.633773/1179.398	265,6337732	19,65664051	14,62
268.045356/1743.858	268,0453562	29,0642968	16,60
268.622288/1258.926	268,6222881	20,9820975	13,79
269.635473/1230.377	269,6354726	20,50627711	14,72
270.6254/1212.424	270,6253996	20,20706497	13,98
270.63942/1359.559	270,63942	22,65931457	13,73
271.636821/1202.838	271,6368213	20,04730355	16,10
273.628936/1194.162	273,6289358	19,90270664	14,30
274.629864/1191.02	274,6298635	19,85034113	14,13
274.630437/1262.888	274,6304374	21,0481343	14,41
281.143806/1204.434	281,1438063	20,07390411	19,58
282.665381/1153.988	282,665381	19,23313141	13,84
285.575663/1815.207	285,5756634	30,25345542	13,36
285.612072/1263.691	285,6120721	21,0615105	14,13

289.616079/1364.636	289,6160794	22,74393441	14,04
289.633751/1214.821	289,6337508	20,24700843	15,15
290.636694/1206.118	290,6366936	20,10197092	14,74
292.621738/1205.03	292,621738	20,08383204	16,21
294.630044/1199.84	294,6300443	19,9973258	15,79
294.64226/1362.038	294,6422602	22,70063152	13,41
296.654961/1271.452	296,654961	21,19086441	14,91
298.631463/1220.712	298,6314635	20,34519762	14,37
298.635515/1293.916	298,6355149	21,56527263	14,62
299.646088/1197.092	299,6460876	19,95153484	15,99
300.590944/1408.706	300,5909444	23,47843393	13,71
300.650874/1217.639	300,6508739	20,29397918	17,35
302.133273/1411.849	302,1332732	23,53082121	23,15
302.135969/1287.865	302,135969	21,46440995	20,12
303.13451/1413.158	303,1345101	23,55262564	20,95
304.111015/1593.291	304,1110154	26,55484677	21,92
304.648005/1268.013	304,6480047	21,13354378	16,32
305.63435/1302.773	305,6343504	21,71287654	15,12
307.146323/1369.066	307,1463235	22,817772	18,62
307.148705/1230.369	307,1487049	20,50614996	19,07
308.665297/1169.656	308,6652966	19,49426678	14,04
310.638959/1255.701	310,6389587	20,92835776	13,79
314.101344/1743.897	314,1013436	29,0649441	18,79
314.634715/1251.179	314,6347146	20,85298135	15,97
318.643809/1223.993	318,6438086	20,39988076	14,52
319.159081/1102.727	319,1590809	18,37878525	19,39
320.13298/1485.611	320,1329803	24,7601903	20,37
320.635477/1213.982	320,6354768	20,2330331	15,35
321.641887/1305.62	321,6418868	21,76034121	15,11
321.650428/1208.409	321,6504277	20,14014278	15,39
323.630609/1801.983	323,6306086	30,03305525	13,83
323.640607/1249.625	323,6406065	20,82708473	16,90
324.150607/1248.284	324,150607	20,8047335	16,67
324.647138/1249.004	324,6471381	20,81674097	15,53
325.649757/1288.357	325,6497575	21,47261704	14,11
327.161675/1369.801	327,1616752	22,83001681	17,43
327.162473/1230.925	327,1624728	20,51541556	19,53
328.124035/1141.391	328,1240353	19,0231778	19,09
328.671614/1184.054	328,6716143	19,73424161	14,87
331.644158/1207.185	331,6441578	20,11974746	15,90
333.646652/1259.013	333,6466523	20,98354474	13,84
335.15515/1273.104	335,1551495	21,21840159	16,63
337.653498/1390.386	337,653498	23,17309764	13,95
342.135439/1743.897	342,1354392	29,0649441	19,19
343.661957/1224.444	343,6619573	20,40740632	15,98
348.147758/1301.265	348,1477582	21,68775823	21,03
348.64621/1326.937	348,6462097	22,11562184	14,59
349.150317/1301.565	349,1503175	21,69275782	18,52
351.687587/1345.166	351,687587	22,41942605	14,58
352.060866/1538.349	352,060866	25,63914717	17,44
352.696232/1254.963	352,6962316	20,9160515	16,68
353.636588/1256.701	353,6365877	20,9450182	14,62

353.655923/1336.987	353,6559232	22,28311962	14,00
356.043978/1484.516	356,0439784	24,74192537	16,78
357.240469/1045.095	357,2404687	17,41825461	19,67
357.668794/1243.614	357,6687942	20,72690523	15,21
358.218428/1045.336	358,2184278	17,42227492	18,22
360.652745/1399.776	360,6527447	23,3296071	13,83
360.661529/1303.782	360,661529	21,72970736	14,15
361.718911/1215.219	361,7189114	20,25364527	14,91
365.654915/1256.415	365,6549154	20,94025609	16,08
366.599792/1929.853	366,5997925	32,16421764	18,61
368.654864/1319.063	368,6548638	21,9843821	14,65
369.66691/1355.09	369,6669101	22,58483945	13,91
370.676917/1353.474	370,676917	22,55789869	14,67
373.168936/1550.871	373,168936	25,84785716	19,76
375.22096/1068.053	375,2209604	17,80087744	21,27
375.645865/1258.961	375,6458654	20,98268539	13,77
376.218952/1070.119	376,2189515	17,83531475	19,05
377.16708/1163.562	377,1670795	19,39269307	17,61
377.655953/1359.433	377,6559532	22,65721418	15,56
378.110994/1784.113	378,1109936	29,73521871	18,01
378.659033/1291.424	378,6590332	21,52372702	17,01
379.671845/1355.508	379,6718446	22,59179673	14,88
381.688107/1297.209	381,6881069	21,62015418	15,24
385.166963/1664.281	385,1669625	27,73802389	18,91
385.173767/1266.758	385,1737671	21,11262804	20,18
385.648308/1368.425	385,6483075	22,80709165	14,80
387.174605/1288.03	387,1746047	21,46716792	19,04
389.197597/1336.085	389,1975971	22,26808683	17,09
391.178399/1158.617	391,1783986	19,31029045	17,20
392.18842/1099.733	392,1884198	18,32888225	17,71
392.637684/1364.239	392,6376836	22,73731927	14,12
395.127839/1864.154	395,1278386	31,06923162	16,68
395.629514/1865.679	395,6295141	31,09464439	14,84
397.678509/1356.868	397,6785088	22,61446805	14,82
398.676458/1217.373	398,676458	20,28954232	14,60
401.165412/1630.778	401,1654115	27,17962894	17,16
401.172738/1694.078	401,1727382	28,23463127	17,85
405.160292/1750.745	405,1602925	29,17908622	17,69
407.681908/1288.138	407,6819084	21,46896308	15,95
408.196932/1586.633	408,1969318	26,44388156	18,85
410.173096/1595.747	410,1730957	26,59578908	19,37
413.200371/1192.737	413,2003708	19,87894394	18,29
413.683915/1159.411	413,6839146	19,32352332	15,16
416.138318/1457.816	416,1383182	24,29692872	21,18
417.138051/1457.816	417,1380508	24,29692872	19,16
422.208861/1585.041	422,208861	26,41734307	19,27
423.21461/1582.481	423,2146104	26,37469114	17,03
428.171946/2161.302	428,1719461	36,02170451	19,85
430.200632/1216.775	430,2006318	20,27957778	18,74
434.672277/1449.443	434,6722771	24,15738558	13,56
435.181292/1392.421	435,1812918	23,20701655	20,18
436.224365/1604.09	436,2243654	26,73483786	19,68

437.22758/1603.401	437,22758	26,72335754	17,54
438.677677/1763.369	438,6776767	29,38947692	15,76
442.195188/1213.07	442,1951882	20,21783167	17,66
444.678335/1389.577	444,6783355	23,15962007	14,81
445.234116/1131.271	445,234116	18,85450941	18,79
445.685726/1317.599	445,6857257	21,95997874	13,37
447.13451/1743.8	447,13451	29,06333393	18,09
448.120871/1977.551	448,1208713	32,95918744	17,53
451.199547/1640.605	451,1995466	27,3434211	16,94
451.698063/1344.384	451,6980627	22,40640803	15,51
453.752029/1209.719	453,7520288	20,1619817	14,68
456.630578/1929.795	456,6305782	32,16324359	19,55
457.133081/1929.812	457,1330812	32,16354081	17,98
457.635126/1929.771	457,6351265	32,16285129	16,27
461.71289/1220.865	461,7128898	20,34775286	13,90
462.190898/1220.712	462,1908982	20,34519762	17,51
462.706609/1339.049	462,7066092	22,31747984	14,64
464.684514/1212.616	464,684514	20,21026703	13,28
468.703106/1451.689	468,703106	24,19482171	14,32
470.225944/1653.39	470,225944	27,55650688	17,43
470.703741/1183.775	470,7037412	19,72958001	14,72
481.238133/1219.328	481,2381329	20,32213214	15,33
484.698435/1368.309	484,6984353	22,80514295	15,24
486.247369/1708.019	486,2473691	28,4669869	20,13
488.133087/1622.375	488,1330872	27,03958472	19,29
488.214118/1369.828	488,2141182	22,83047402	22,28
489.215511/1369.069	489,2155112	22,81781094	20,28
489.229227/1205.593	489,2292272	20,09322211	19,03
489.665135/1342.397	489,6651353	22,37328039	14,50
490.134078/1623.722	490,134078	27,06203208	16,47
490.21951/1368.282	490,21951	22,80470357	17,63
492.205293/1471.231	492,2052929	24,52051938	16,79
493.253588/1677.895	493,2535878	27,96491312	15,97
495.748052/1194.549	495,7480519	19,90914215	14,95
497.661528/1356.495	497,6615285	22,60824506	13,63
498.255692/1806.746	498,2556915	30,11242689	15,39
499.721026/1363.67	499,7210256	22,72782957	14,53
501.226919/1696.798	501,2269192	28,279971	16,28
502.224384/1707.009	502,2243842	28,45015414	16,28
506.200924/1752.808	506,2009238	29,2134718	16,15
515.24634/1400.671	515,2463404	23,34451668	15,51
515.248321/1729.411	515,2483211	28,82351853	16,06
518.231079/1746.072	518,2310791	29,10120139	15,29
518.249367/1644.683	518,2493668	27,41138073	14,93
518.730279/1191.19	518,7302786	19,8531684	14,93
519.202828/1716.55	519,2028283	28,60916124	17,18
519.261834/1679.677	519,261834	27,99461568	16,29
524.18742/1537.513	524,1874198	25,62521491	15,28
532.237327/1715.671	532,2373266	28,59451929	15,34
537.252966/1745.497	537,2529662	29,09161056	16,26
537.259601/1206.691	537,2596009	20,1115228	15,45
537.65526/1977.926	537,6552596	32,96542568	19,10

538.15717/1977.551	538,1571699	32,95918744	17,90
538.256933/1745.852	538,2569334	29,09753333	15,13
541.262762/1216.547	541,2627622	20,27579099	15,15
542.266486/1216.08	542,2664862	20,26799739	14,90
543.733933/1491.624	543,7339327	24,86039627	13,59
545.181617/1534.794	545,1816167	25,57989247	17,51
548.236652/1770.107	548,2366522	29,50178164	16,56
551.230866/1687.373	551,2308658	28,12287597	15,00
556.233638/1715.476	556,2336378	28,59126324	16,25
557.22316/1733.993	557,2231598	28,89988669	15,04
557.728983/1522.141	557,7289831	25,36900963	14,74
558.246165/1439.659	558,2461647	23,99431211	18,32
559.235568/1778.845	559,2355682	29,64742487	15,95
559.261777/1211.451	559,2617769	20,19084567	15,80
564.228867/1194.728	564,2288667	19,9121261	15,60
564.755231/1194.859	564,7552315	19,91431623	14,18
565.264247/1755.841	565,2642474	29,26401596	16,42
566.26682/1756.102	566,2668204	29,2683656	14,90
567.306418/1742.115	567,3064183	29,03524965	18,11
568.275617/1738.142	568,2756168	28,9690395	17,85
570.254059/1734.858	570,2540592	28,91429373	15,80
570.736694/1175.967	570,7366943	19,5994569	15,63
571.12954/1927.947	571,1295405	32,13245331	13,78
571.233572/1187.381	571,2335721	19,78968591	17,25
571.75515/1187.613	571,7551504	19,79354428	15,92
572.247057/1190.971	572,2470572	19,84951362	16,04
574.245882/1803.627	574,245882	30,06045483	15,62
575.230134/1758.306	575,2301345	29,30509964	14,74
581.170349/2003.087	581,1703491	33,38478874	15,85
581.261495/1791.803	581,2614954	29,86337724	16,17
582.245846/1711.533	582,2458461	28,52555629	15,55
582.261043/1789.684	582,2610432	29,82807281	15,32
583.284852/1764.808	583,2848516	29,41346918	15,38
584.253179/1203.079	584,2531791	20,05132495	15,30
593.28699/1271.06	593,28699	21,18432928	13,91
594.258391/1285.899	594,2583907	21,43164897	16,89
594.67484/2007.316	594,6748398	33,45526924	14,17
595.274204/1800.156	595,2742037	30,00259454	15,46
597.298388/1782.142	597,2983879	29,70237409	14,01
598.243112/1817.154	598,2431123	30,28589325	14,82
599.157187/2007.515	599,1571868	33,45858828	18,10
599.254308/1775.366	599,254308	29,58942879	14,82
599.294007/1215.768	599,294007	20,26279527	14,74
599.745744/1190.383	599,745744	19,83972129	15,56
599.761164/2005.377	599,7611635	33,42294345	15,66
600.686619/2008.125	600,6866186	33,46875825	13,92
600.737245/1190.894	600,7372445	19,84823332	13,56
602.240525/1285.603	602,2405248	21,42672213	16,22
603.25773/1872.357	603,2577304	31,20595449	16,31
607.287674/1266.867	607,2876741	21,11445337	17,24
613.281398/1205.717	613,2813976	20,09527859	14,73
614.292271/1810.528	614,2922713	30,17547232	15,77

614.2963/1730.712	614,2963001	28,84519658	14,01
620.247793/1820.414	620,2477933	30,34023657	14,89
627.195496/1743.897	627,1954956	29,0649441	21,08
628.198697/1743.893	628,1986965	29,06487812	19,08
629.20105/1743.897	629,2010498	29,0649441	17,17
630.204347/1743.893	630,2043469	29,06487812	14,68
630.329301/1777.574	630,3293012	29,62623844	14,54
633.245859/1808.377	633,2458589	30,1396222	14,01
638.284268/1843.128	638,2842685	30,71880022	13,96
639.191648/2019.722	639,1916478	33,66203958	14,90
645.302376/1782.53	645,3023764	29,70884146	13,61
647.769309/1214.697	647,7693085	20,24494297	12,80
649.261459/1503.246	649,2614586	25,0540934	15,08
650.18319/1749.625	650,1831897	29,1604185	14,96

---



**Supplementary Table S3: Clinical characteristics of UPJ discovery cohort**

Sample code	Clinical characteristics			Gender	Age (days)	Cohort
	Status	HN grade	Pelvis dilatation			
C63-1-14	Healthy control			M	1	discovery
C63-1-7	Healthy control			M	1	discovery
C63-1-3	Healthy control			M	2	discovery
C63-1-8	Healthy control			M	2	discovery
C63-1-4	Healthy control			M	3	discovery
C63-1-13	Healthy control			M	4	discovery
C63-1-5	Healthy control			M	5	discovery
C30-4-3	Healthy control			M	48	discovery
C47-1-2	Healthy control			M	94	discovery
C30-3-16	Healthy control			M	104	discovery
C59-1-5	Healthy control			M	123	discovery
C63-1-16	Healthy control			M	143	discovery
C63-2-7	Healthy control			M	152	discovery
C63-2-5	Healthy control			M	176	discovery
C47-1-3	Healthy control			M	187	discovery
C22-2-2	UPJ obstruction	3	16	M	38	discovery
C11-4-14	UPJ obstruction	3	21	M	44	discovery
C42-4-11	UPJ obstruction	3	25	M	44	discovery
C42-4-4	UPJ obstruction	3	26	F	37	discovery
C49-4-12	UPJ obstruction	3	43	M	34	discovery
C42-4-14	UPJ obstruction	4	17	M	29	discovery
C3-1-1	UPJ obstruction	4	22	M	53	discovery
C7-1-20	UPJ obstruction	4	22	M	53	discovery
C5-3-11	UPJ obstruction	4	23	M	33	discovery
C2-3-13	UPJ obstruction	4	24	M	37	discovery
C49-3-4	UPJ obstruction	4	24	M	106	discovery
C63-2-2	UPJ obstruction	4	25	M	50	discovery
C16-4-15	UPJ obstruction	4	25	M	58	discovery
C58-2-6	UPJ obstruction	4	26	M	30	discovery
C57-1-21	UPJ obstruction	4	28	F	56	discovery
C63-2-3	UPJ obstruction	4	30	M	20	discovery
C11-3-17	UPJ obstruction	4	30	M	26	discovery
C58-4-9	UPJ obstruction	4	30	M	39	discovery
C55-1-5	UPJ obstruction	4	31	M	40	discovery
C31-1-17	UPJ obstruction	4	33	M	39	discovery
C17-3-14	UPJ obstruction	4	34	M	58	discovery
C16-3-3	UPJ obstruction	4	35	M	40	discovery
C50-1-6	UPJ obstruction	4	50	M	20	discovery
C50-2-1	UPJ obstruction	4	50	M	36	discovery
C5-3-6	UPJ obstruction	ND	25	M	35	discovery
C6-3-16	UPJ obstruction	ND	16	M	167	discovery
C63-2-4	UPJ obstruction	ND	20	M	133	discovery
C6-3-19	UPJ obstruction	ND	22	M	112	discovery
C5-3-10	UPJ obstruction	ND	30	M	213	discovery

C6-3-17	UPJ obstruction	ND	34	M	105	discovery
C6-4-11	UPJ obstruction	ND	40	M	69	discovery
C5-3-1	UPJ obstruction	ND	ND	F	80	discovery
C5-4-15	UPJ obstruction	ND	ND	M	174	discovery
C6-4-21	UPJ obstruction	ND	ND	M	193	discovery

---

HN: Hydronephrosis. ND: undetermined.

**Supplementary Table S4: Clinical characteristics of UPJ validation cohort**

Sample code	Score UPJMetab32	Clinical characteristics			Gender	Age (days)	Cohort
		Status	HN grade	Pelvis dilatation			
C30-4-5	-0,1674	Healthy control			M	18	validation
C63-1-10	-1,0277	Healthy control			M	1	validation
C63-1-2	-0,9342	Healthy control			M	1	validation
C63-1-1	-0,4827	Healthy control			M	1	validation
C63-1-9	-0,7862	Healthy control			M	4	validation
C30-4-2	0,4222	Healthy control			M	261	validation
C63-1-6	-0,2766	Healthy control			M	1	validation
C2-3-1	-0,3163	UPJ obstruction	ND	15	M	193	validation
C5-4-5	0,0726	UPJ obstruction	ND	15	M	232	validation
C5-3-17	0,9125	UPJ obstruction	ND	16	M	153	validation
C28-4-12	1,0583	UPJ obstruction	3	17	M	49	validation
C47-3-10	0,9652	UPJ obstruction	4	19	M	47	validation
C47-1-10	0,5623	UPJ obstruction	4	20	M	64	validation
C5-3-20	-0,0424	UPJ obstruction	4	23	M	196	validation
C55-2-11	0,4952	UPJ obstruction	4	23	F	42	validation
C44-4-18	0,2127	UPJ obstruction	4	25	M	69	validation
C9-2-7	-0,0509	UPJ obstruction	3	27	F	27	validation
C7-1-18	1,1222	UPJ obstruction	4	31	M	29	validation
C31-4-10	0,8644	UPJ obstruction	4	33	M	33	validation
C3-1-15	0,3647	UPJ obstruction	4	37	M	24	validation
C50-1-20	1,0617	UPJ obstruction	4	46	F	29	validation
C55-1-1	0,6135	UPJ obstruction	4	50	M	36	validation
C56-1-2	0,5371	UPJ obstruction	4	67	M	60	validation
C5-3-19	-0,510	UPJ obstruction	ND	ND	M	262	validation

**Supplementary Table S5: List of 32 urinary metabolites differentially excreted between healthy and UPJ patients in the discovery cohort.**

ID	Raw-intensity			Log2-intensity		p-value	
	Healthy	UPJ obstruction	FC	Healthy	UPJ obstruction	Wilcoxon	BH
324.083234/1512.623	708506	406112	0,57	19,34	18,56	1,8E-05	0,0035
456.630578/1929.795	3173450	1588974	0,50	21,50	20,26	1,8E-05	0,0035
177.031281/1357.129	526987	318408	0,60	18,92	18,20	1,3E-04	0,0135
366.599792/1929.853	788580	605801	0,77	19,54	18,89	1,4E-04	0,0135
227.111791/989.758	10196251	24720725	2,42	22,67	24,26	2,5E-04	0,0164
467.617961/1929.795	129782	73106	0,56	16,83	15,99	2,5E-04	0,0164
457.133081/1929.812	864131	540265	0,63	19,58	18,91	3,0E-04	0,0166
438.677677/1763.369	83802	53536	0,64	16,26	15,63	6,7E-04	0,0182
490.134078/1623.722	114760	85880	0,75	16,76	16,30	8,0E-04	0,0182
157.077555/1067.986	198756	369487	1,86	17,56	18,25	7,9E-04	0,0193
242.074921/1533.731	3719793	7549206	2,03	21,63	22,61	5,9E-04	0,0193
299.056259/1357.852	500388	304343	0,61	18,82	18,10	7,4E-04	0,0193
338.098419/1482.724	633012	468298	0,74	19,20	18,81	5,8E-04	0,0193
390.245275/1571.028	905682	401355	0,44	19,42	18,07	3,3E-03	0,0193
474.701959/1359.882	132854	80873	0,61	16,85	16,22	1,4E-03	0,0193
488.133087/1622.375	1150184	811380	0,71	20,07	19,55	5,8E-04	0,0193
229.117309/1322.695	108917815	204227951	1,88	26,49	27,30	9,3E-04	0,0194
234.129064/1411.802	230142	158724	0,69	17,72	17,20	1,7E-03	0,0194
368.181291/1241.422	219228	127700	0,58	17,58	16,85	2,4E-03	0,0194
632.28172/1223.924	166033	115919	0,70	17,26	16,49	3,7E-03	0,0194
211.087719/991.444	88269	164219	1,86	16,27	17,10	1,4E-03	0,0240
228.167817/1066.426	64195	110790	1,73	15,75	16,46	1,4E-03	0,0240
237.193077/1281.495	96751	187287	1,94	16,28	17,26	1,5E-03	0,0240
526.161131/1357.627	432409	272417	0,63	18,56	17,72	1,2E-03	0,0240
181.105924/990.207	732059	1324472	1,81	19,03	20,14	2,3E-03	0,0333
228.114334/990.108	894846	1902390	2,13	19,20	20,55	2,1E-03	0,0344
259.090975/1390.694	1117825	485984	0,43	19,48	18,46	3,8E-03	0,0356
306.619622/1015.236	45530	67407	1,48	15,35	15,92	6,0E-03	0,0366
303.615001/1045.311	16068	24344	1,52	13,91	14,41	1,4E-03	0,0407
355.071351/1117.064	547815	283969	0,52	18,80	17,97	4,0E-03	0,0424
537.65526/1977.926	4714555	2813304	0,60	21,96	20,87	1,2E-02	0,0435
185.125274/1291.16	165754	317070	1,91	17,18	17,79	2,3E-03	0,0481

FC: fold change between UPJ obstruction and healthy patients. BH: Benjamini and Hochberg.



## **Study 4:**

# Aptamer selection for biosensor development

---

## **Aims and Objectives**

Identification of urinary biomarkers by CE-MS gave promising results but this technology is quite heavy and requires trained personnel to analyze urine samples. The aim of this last part of my thesis was to develop biosensor using aptamers, that could be used as an easy-to-use diagnostic tool and that would give, for e.g., a color change in the presence of a urinary peptide biomarker.

## Introduction

Identification of urinary peptide biomarkers by CE-MS has emerged as a powerful tool allowing CKD diagnosis and prognosis as discussed in chapter III in the introduction. The CKD273 classifier, a panel of 273 urinary peptide biomarkers, has reached the highest step in clinical validation and has even received the FDA support for further validation (<https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/UCM508790.pdf>) [1].

However, this technology has some limitations. Trained personnel are needed to run the sample analysis, the results are not obtained on the spot, and it is quite expensive since regular servicing needs to be done to ensure that the apparatuses are still correctly calibrated for the analyses. To go one step ahead, tools with the same predictive potential but that are easier, more rapid and cheaper to use are constantly looked forward. As discussed in chapter IV in the introduction, aptamers seem to be optimal biosensors that could help developing such clinical tools. However, aptamer selection takes place on one target at a time.

In the laboratory, we thus aimed to develop a biosensor based on the aptamer technology that would detect, directly in urine, one peptide biomarker previously identified by CE-MS. For this, we first selected a peptide that displayed appropriate physico-chemical properties for optimal aptamer selection and good predictive performance in the context of CKD, from the CKD273 panel. Then, we focused on the selection of aptamers directed against this peptide using CE-SELEX.

## Materials and methods

### *Alpha-1-antitrypsin fragment, ssDNA library sequences and conjugated ssDNA-peptide synthesis.*

The  $\alpha$ -1-antitrypsin fragment (MIEQNTKSPLFMGKVVNPTQK, MW=2391.0 g/mole) was purchased at Eurogentec (France) with a purity >85% and resuspended in ultrapure water at 25mM. A ssDNA library was designed and purchased at Integrated DNA Technologies (Belgium). The synthesized library was delivered at a concentration of 100 $\mu$ M in IDTE buffer (pH 8.0). Each ssDNA sequence of the library contained 66 nucleotides (MW~20000 g/mole): 30 randomized nucleotides (N30) flanked by two 18-nucleotide primer sequences (5' - CGCCTCGGATAACGGATT-N30-TTCGCAGGTTACAGGACA - 3').

A peptide conjugated with a fixed ssDNA sequence of 66 nucleotides was also purchased at Eurogentec (France). The  $\alpha$ -1-antitrypsin fragment amino acid sequence was slightly modified to add a cysteine (C) so that the oligonucleotide could be added by covalent bonding. To note, this modification negligibly affected the isoelectric point, size and hydrophobicity of the peptide. The 21-amino acid was MIEQNCKSPLFMGKVVNPTQK and the ssDNA sequence added was 5'-CGCCTCGGATAACGGATTTAGACTGATGAGACTCTACTAATGTTGCGCTTCGCAGGTTACAGGACA-3' (MW=23159.8 g/mole).



### ***Incubation step of CE-SELEX***

For our CE-CELEX process, 25 $\mu$ M ssDNA sequences were resuspended in the incubation buffer containing 100mM potassium phosphate (Sigma, USA), 1mM magnesium chloride (Sigma, USA) and 10mM NaCl (Merck, Germany) adjusted to pH 8.2 to a final volume of 10 $\mu$ L and then heated at 94°C for 10mins before being rapidly cooled on ice. Ten microliters of 25 $\mu$ M  $\alpha$ -1-antitrypsin fragment was also resuspended in the incubation buffer and then added to the ssDNA solution for incubation for 30 minutes at room temperature.

### ***CE migration buffer***

For an efficient separation of our molecular species, we tested 4 different buffers: i) **10mM potassium phosphate buffer** (10mM potassium phosphate (Sigma, USA), 1mM magnesium chloride (Sigma, USA) and 10mM NaCl (Merck, Germany) adjusted to pH 8.2), ii) **20mM potassium phosphate buffer** (20mM potassium phosphate (Sigma, USA), 1mM magnesium chloride (Sigma, USA) and 10mM NaCl (Merck, Germany) adjusted to pH 8.2), iii) **100mM potassium phosphate buffer** (100mM potassium phosphate (Sigma, USA), 1mM magnesium chloride (Sigma, USA) and 10mM NaCl (Merck, Germany) adjusted to pH 8.2) and iv) **10mM sodium acetate buffer** (10mM sodium acetate (Sigma, USA), 1mM magnesium chloride (Sigma, USA) and 10mM NaCl (Merck, Germany) adjusted to pH 8.2).

### ***CE instrumentation***

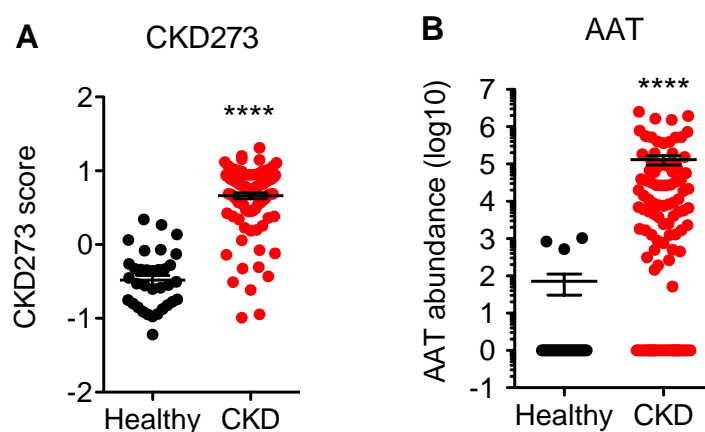
Analyzes were done using a Beckman Coulter Proteome Lab PA800 capillary electrophoresis system (Beckman Coulter, Fullerton, USA) equipped with a 100 $\mu$ m internal diameter fused silica capillary (Polymicro Technologies, Phoenix, Arizona) of a total length of 50 cm on which a detection window was done at 39.5 cm. Using the on-instrument ultraviolet (UV) photodiode array detection system and standard filters, the migration of molecular species was monitored at 214nm ( $\alpha$ -1-antitrypsin fragment and conjugated peptide) and 254nm (ssDNA). For the experiments, the electrodes were cleaned using milli-Q water and in between different CE separation runs, the capillary was washed with 0.1M NaOH. The CE migration buffers volumes of the vials at anode and cathode was 1.5 mL. Each time a new capillary was used, it was conditioned by washing with 0.1M HCl (30.0 psi for 4mins), 0.1M NaOH (30.0 psi for 4mins), milli-Q water (30.0 psi for 4mins), and finally the CE separation buffer 0.1M HCl (30.0 psi for 4mins). The voltage applied was of +22 kV (current intensity: 90  $\mu$ A).

## **Results**

### ***Choice of urinary peptide biomarker as target for aptamer selection***

To start our selection by CE-SELEX, we decided to focus on a peptide from the CKD273 panel. The choice of the peptide was based on two main selection criteria. First, for affinity studies using CE, an important requirement is that there should be an electrophoretic mobility difference between the

molecular species under analysis. This means that there should be a difference in the size or charge between free or bound species that will cause a change in their migration patterns [2]. In the CE, at pH~8, the ssDNA sequences are negatively charged [3, 4]. Therefore, to optimize the separation, the candidate peptide should have a net positive charge at this pH, i.e., an isoelectric point > 8. In addition, the candidate peptide should be able to significantly discriminate between controls and patient with CKD. Based on these criteria, we selected a fragment of  $\alpha$ -1-antitrypsin (AAT) from the CKD273 (peptide ID 90840 [5]). This peptide has a 21-amino acid sequence MIEQNTKSPLFMGKVVNPTQK and an isoelectric point of 9.7. At pH~8, the peptide has a net positive charge due to the ionization of the three lysine (K) and one glutamate (E). Moreover, this AAT peptide is ~1000 more abundant in urine of patients with CKD compared to controls, and has a good discriminatory power compared to the CKD273 score (Figure 1).



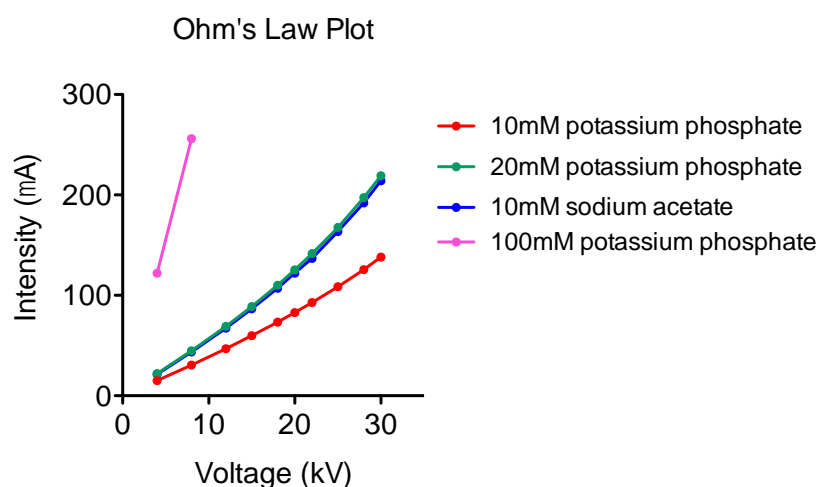
**Figure 1.** Scatterplots showing the classification of CKD patients and patients with no kidney disease (controls) using the scores established by the CKD273 classifier (A) and by the relative urinary abundance of AAT fragment (B). Adapted from [5] supplementary data.

### *Selection of optimal CE separation buffer*

For affinity analysis by CE, the first step was to find the correct buffer composition for separation of the different molecules. Optimal composition of a CE separation buffer is defined by the absence of heat generated (Joule heating) when an electric current is applied to the solution since this will affect the migration of our electrolytes. We thus compared 4 different buffers: i) 10mM potassium phosphate buffer, ii) 20mM potassium phosphate buffer, iii) 100mM potassium phosphate buffer and iv) 10mM sodium acetate buffer. For all of them the pH was kept to 8.2. To find the optimal buffer composition and the maximum voltage that can be applied to limit Joule heating, an Ohm's law plot was plotted by recording the intensity ( $\mu$ A) generated in response to the application of increasing voltages to a capillary filled with the 4 different buffers. The Ohm law states that ratio of voltage to current (intensity) is a

constant and gives a linear relationship when voltage and current are plotted. If the relationship is non-linear, this means that the material does not obey the Ohm law and that the temperature will change [6].

As shown in Figure 2, the 20mM potassium phosphate (green line) and 10mM sodium acetate (blue line) buffers showed a non-linear relationship between voltage and intensity. For the 100mM potassium phosphate buffer (pink line), the intensity rose too rapidly to 122 and 265  $\mu\text{A}$  upon application of low voltages of 4 and 8 kV respectively. By comparison, the 10mM potassium phosphate buffer (red line) was the only one showing a linear relationship between voltage and intensity, hence obeying the Ohm's law. This buffer was then selected for separation of molecular species during CE-SELEX.

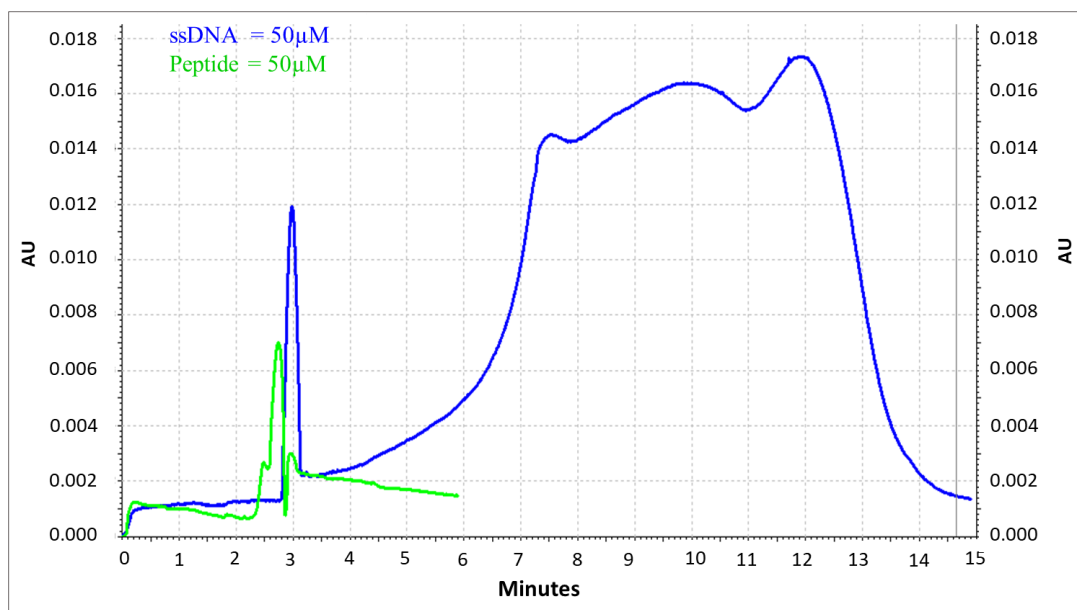


**Figure 2.** Ohm's law plot showing the intensity recorded with increasing voltage application to the different CE separation buffers: 10mM potassium phosphate buffer, 20mM potassium phosphate buffer, 10mM sodium acetate buffer and 100mM potassium phosphate buffer.

### *Migration time assessment*

The next step was to define clearly the migration time of the unbound AAT fragment, unbound ssDNA and the conjugated peptide-ssDNA complex to define the optimal time windows for complex collection.

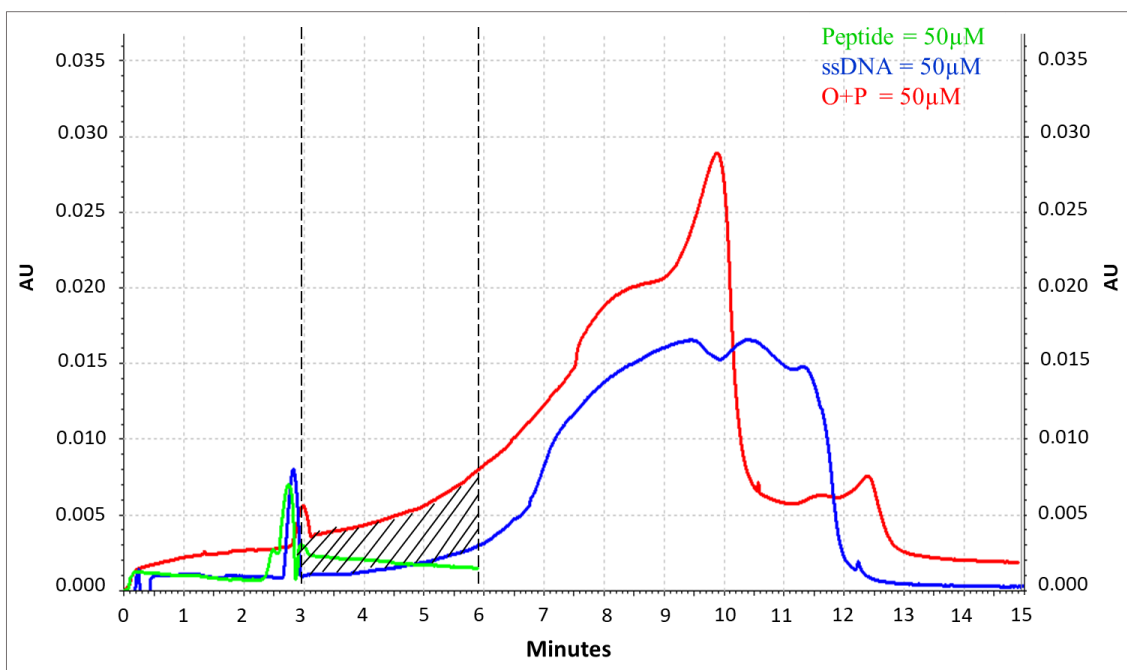
For the analysis we first ran the AAT fragment alone (50 $\mu\text{M}$ ), then the ssDNA library alone (50 $\mu\text{M}$ ). With the peptide, a sharp peak was detected between 2.38 ( $\pm$  0.20) and 2.91 ( $\pm$  0.02) mins as shown in Figure 3 (green line). With the ssDNA, two peaks were detected: a first one between 2.77 ( $\pm$  0.07) and 3.06 ( $\pm$  0.07) mins, slightly overlapping the time of detection of AAT fragment, and the other detected between 5.71 ( $\pm$  0.45) and 13.4 ( $\pm$  1.20) mins (Figure 3, blue line). These results confirmed the mobility shift between the AAT peptide and the ssDNA library that migrate in two different time windows.



**Figure 3.** Overlay of the migration pattern of 50µM AAT peptide (green curve) and 50µM ssDNA library (blue curve). Detection is monitored at 214nm for AAT fragment and 254nm for ssDNA. AU: absorbance unit.

We next monitored the electrophoretic mobility of the complex. The peptide being positively charged, and the ssDNA sequences being negatively charged at pH 8.2, the formation of a complex between the two molecules would imply a charge shift, towards neutrality, and a detection window during the time interval, between the free AAT and free ssDNA sequences.

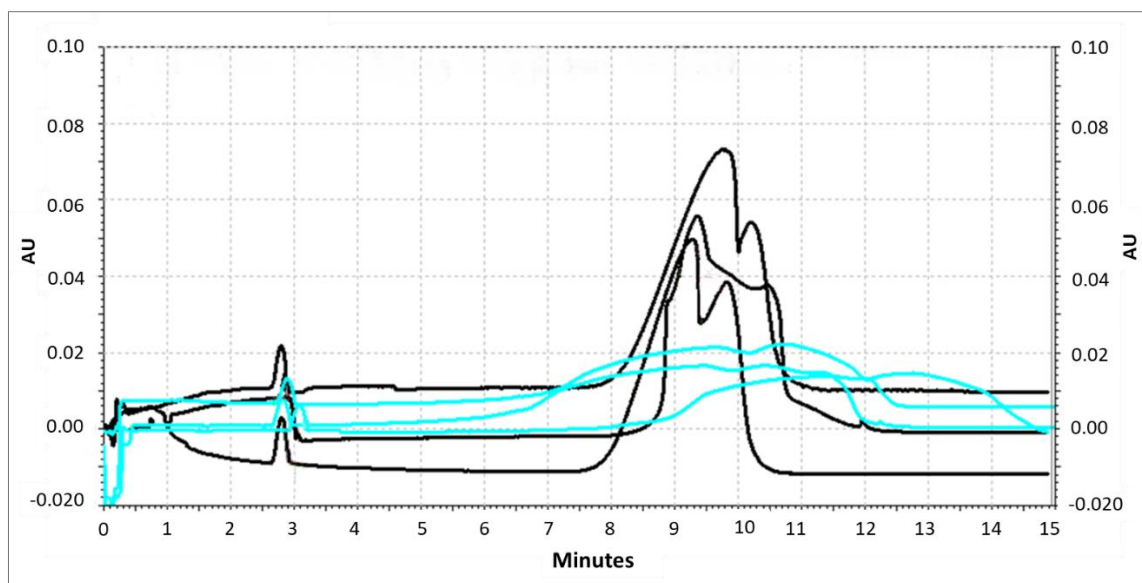
To evaluate AAT-ssDNA complex migration and the time window during which the complex should be collected we purchased the same AAT fragment on which an oligonucleotide of 66 nucleotides has been conjugated. We ran 50µM of this conjugated peptide and obtained the migration pattern shown in Figure 4 (red line). First, a small peak, overlapping with the detection time of free AAT fragment and free ssDNA is obtained. Then, a second peak is detected, starting immediately after the first one, and slowly increasing to attain maximum intensity at about 10 minutes, followed by a sharp collapse. Optimal time window for complex collection should be placed between the two peaks observed for the free ssDNA library so that there is no contamination of complexes with unbound DNA during the CE-SELEX analysis. This window, represented by the shaded area in Figure 4, corresponds to a time-lapse of approximately three minutes.



**Figure 4.** Overlay of the migration pattern of AAT peptide (green curve), ssDNA library (blue curve) and the peptide conjugated to the oligonucleotide (O+P) (red curve). Detection is measured at 214nm for AAT fragment and the conjugated peptide and at 254nm for ssDNA. The shaded region represents the time zone during which AAT peptide fragment – ssDNA complexes can be collected without contamination from unbound DNA. AU: absorbance unit.

#### *CE migration reproducibility of ssDNA library*

We next assessed the reproducibility of the migration of the ssDNA during CE to validate the time window for the complex collection. While the first peak at ~2.8 mins was still present and stable over 17 runs, the second peak showed a very strong instability in the resolution of separation over time (Figure 5). These results show that our parameters for CE-SELEX were not reproducible and that the CE analysis could not be optimized for aptamers selection.



**Figure 5.** Overlap of different CE migration runs during AAT peptide-ssDNA complexes collection, all carried out using the same capillary. The black curves represent the early runs (5<sup>th</sup> – 7<sup>th</sup>) and the blue curves represent later runs (15<sup>th</sup>-17<sup>th</sup>). AU: Absorbance unit.

## Discussion

Selection of aptamers using CE-SELEX has several advantages over other methods, including that low sample and reagent volumes are required, a large variety of molecules may be analyzed, separation is automated, and the number of SELEX cycles required for the selection of aptamers is reduced. Moreover, its use during separation of a ligand-protein complexes have already been described [2, 7].

The composition of the buffers used for incubation and separation of the complexes is of utmost importance in the CE-SELEX process, as this will influence the formation and the stabilization of the aptamer structure, the binding of the aptamer to its target, and the migration in the CE. In the literature, Tris-based buffers for CE analysis reigned as the ‘optimal’ buffer for DNA analysis and is the most widely used buffer nowadays [8, 9]. However, several years ago, Ray, T. *et al.* described for the first time that during electrophoresis Tris induced ssDNA degradation due to electrolysis of water altering the pH of the buffers in the vials [4, 10, 11]. For this reason, we tested the use of other buffers without Tris, to avoid DNA degradation, and favored the presence of K<sup>+</sup>, Na<sup>+</sup> and Mg<sup>2+</sup>. These cations play key roles in nucleic acid folding, such as the formation of G-quadruplex, which seems to be an important DNA structure that has been described in a variety of aptamers targeting for e.g. thrombin, ATP, VEGF or insulin [12]. G-quadruplex is a cyclic arrangement of four hydrogen-bonded guanines called a G-quartet. This structure is induced and stabilized by cations which are able to fit in the cavity in the middle of the G-quartet (tetrad plane) [13]. Na<sup>+</sup> allows formation of G-quadruplex structures by fitting only in a tetrad plane but does not allow stabilization of the structure. K<sup>+</sup> is best since it is big enough to fit in the middle of the cavity, formed in between two-tetrad planes, and allows strong stabilization of the G-

quadruplex [12]. According to NMR magnetic studies of G-quadruplex, 100mM of cation is an optimal concentration for G-quadruplex stabilization [14, 15], although for  $K^+$  the ion concentration may be as low as 10mM [16]. Furthermore,  $Mg^{2+}$  is needed to stabilize DNA duplexes and at least 1mM are required [17, 18]. A combination of all these cations increases DNA folding, increasing the chances that structured DNA could bind to our peptide.

To test the efficiency of a separation buffer for CE, it should be certified that there is no heat generated (Joule heating) when the electric current is applied since this will affect the migration of our electrolytes. Heat generated when a voltage is applied will vary according to parameters such as length and internal diameter of the capillary. After testing 4 different buffer compositions, we selected a 10mM phosphate buffer for the CE-SELEX as it did not generate Joule heating.

In addition, since we want to use aptamers for the detection of biomarkers directly in urine, the pH and ionic composition of the buffers should closely mimic that of urine. Analysis of urine revealed that the pH of individuals varies from 5.4 to 8.5 [19] and that pH of urine is usually low in patients with kidney diseases [20, 21]. For this reason, the pH of the buffer was set at 8.2.

Our target for aptamer selection was chosen from a panel of 273 peptides used in the diagnosis and prognosis of CKD [1]. We selected a fragment of AAT, a protein closely associated to inflammation and progression of CKD [1, 22]. The selection of the AAT peptide was mainly driven by its great discriminating power between CKD and non-CKD patients and its physicochemical properties that seemed optimal for the CE-SELEX process. For the selection of aptamers, we used a library containing  $10^{17}$  ssDNA molecules with each sequence having 66 nucleotides, with a random 30-nucleotide central region (N30). The size of the random central region is important. Too short, the complexity may be too low such that the sequences could not form relevant secondary and tertiary structures for target binding. Too long, sequences would be more complex but not all could be represented in the library during chemical synthesis resulting in a loss of useful motifs [23]. In the literature, at least a ~25-nucleotide central region is favored since the whole molecular diversity may be represented in a purchased ssDNA library ( $4^{25} \sim 10^{15}$ ). Therefore, we selected a N30-ssDNA library as it provided a good compromise between assortment and complexity.

Due to its positive charge, the AAT peptide migrated rapidly towards the cathode as shown by the single, resolute peak on CE analysis. Conversely, ssDNA being negatively charged, it was detected much later than AAT peptide. The small peak obtained during ssDNA migration at  $2.77 (\pm 0.07)$  to  $3.06 (\pm 0.07)$  mins is due to positively charged impurities left in the ssDNA library following chemical synthesis. In our analysis, we noted a large standard deviation between the CE runs of free ssDNA ( $\pm 1.11$  mins). It may be hypothesized that the impurities contained in the ssDNA library may accumulate on the capillary wall with successive runs. This accumulation may affect efficiency of migration and could account for the delay in detection of unbound ssDNA and loss in separation resolution.

After incubation, the bound peptide-ssDNA complexes should be separated from the unbound and collected. Such collection at a specific time and over a certain time lapse after CE separation is possible by automation. Due to the lack of migration reproducibility, precise determination of time window for the complex collection and automation of complex collection will not be possible here.

## Conclusion

In this section, we presented preliminary development of CE-SELEX and showed that the use of CE for SELEX did not display sufficient separation resolution and reproducibility for optimal development. However, other options are available such as immobilization of the target on an affinity column. After solving this issue, the next steps, PCR and ssDNA regeneration need to be optimized to proceed the selection of aptamers. The development of a biosensor to screen CKD at the initial stages, using a very sensitive method and at bedside will certainly help to identify most at-risk patient so that they can be treated at the very beginning of the disease and reduce disease progression.

## Bibliography – Study 4

1. Pontillo, C. and H. Mischak, *Urinary peptide-based classifier CKD273: towards clinical application in chronic kidney disease*. Clin Kidney J, 2017. **10**(2): p. 192-201.
2. Ostergaard, J. and N.H. Heegaard, *Capillary electrophoresis frontal analysis: principles and applications for the study of drug-plasma protein binding*. Electrophoresis, 2003. **24**(17): p. 2903-13.
3. Rosenblum, B.B., et al., *Improved single-strand DNA sizing accuracy in capillary electrophoresis*. Nucleic Acids Res, 1997. **25**(19): p. 3925-9.
4. Ric, A., et al., *ssDNA degradation along capillary electrophoresis process using a Tris buffer*. Electrophoresis, 2017. **38**(12): p. 1624-1631.
5. Good, D.M., et al., *Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease*. Mol Cell Proteomics, 2010. **9**(11): p. 2424-37.
6. Waygood, A., *An introduction to electrical science*. 1st ed. 2013, New York: Routledge.
7. Mendonsa, S.D. and M.T. Bowser, *In vitro evolution of functional DNA using capillary electrophoresis*. J Am Chem Soc, 2004. **126**(1): p. 20-1.
8. Brody, J.R. and S.E. Kern, *History and principles of conductive media for standard DNA electrophoresis*. Anal Biochem, 2004. **333**(1): p. 1-13.
9. Kanoatov, M., et al., *Using nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM) for simultaneous determination of concentration and equilibrium constant*. Anal Chem, 2015. **87**(5): p. 3099-106.
10. Ray, T., A. Mills, and P. Dyson, *Tris-dependent oxidative DNA strand scission during electrophoresis*. Electrophoresis, 1995. **16**(6): p. 888-94.
11. Ray, T., J. Weaden, and P. Dyson, *Tris-dependent site-specific cleavage of Streptomyces lividans DNA*. FEMS Microbiol Lett, 1992. **75**(2-3): p. 247-52.
12. Tucker, W.O., K.T. Shum, and J.A. Tanner, *G-quadruplex DNA aptamers and their ligands: structure, function and application*. Curr Pharm Des, 2012. **18**(14): p. 2014-26.
13. Williamson, J.R., M.K. Raghuraman, and T.R. Cech, *Monovalent cation-induced structure of telomeric DNA: the G-quartet model*. Cell, 1989. **59**(5): p. 871-80.
14. Saintome, C., et al., *The exception that confirms the rule: a higher-order telomeric G-quadruplex structure more stable in sodium than in potassium*. Nucleic Acids Res, 2016. **44**(6): p. 2926-35.



15. Marchand, A. and V. Gabelica, *Folding and misfolding pathways of G-quadruplex DNA*. Nucleic Acids Res, 2016. **44**(22): p. 10999-11012.
16. Ambrus, A., et al., *Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution*. Nucleic Acids Res, 2006. **34**(9): p. 2723-35.
17. Miduturu, C.V. and S.K. Silverman, *Modulation of DNA constraints that control macromolecular folding*. Angew Chem Int Ed Engl, 2006. **45**(12): p. 1918-21.
18. Owczarzy, R., et al., *Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations*. Biochemistry, 2008. **47**(19): p. 5336-53.
19. Eggleton, M.G., *Urine acidity in alcohol diuresis in man*. J Physiol, 1946. **104**(3): p. 312-20.
20. Nakanishi, N., et al., *Low urine pH Is a predictor of chronic kidney disease*. Kidney Blood Press Res, 2012. **35**(2): p. 77-81.
21. Ogawa, S., et al., *Lower urinary pH is useful for predicting renovascular disorder onset in patients with diabetes*. BMJ Open Diabetes Res Care, 2015. **3**(1): p. e000097.
22. Cho, J.H., et al., *Alpha1-Antitrypsin Attenuates Renal Fibrosis by Inhibiting TGF-beta1-Induced Epithelial Mesenchymal Transition*. PLoS One, 2016. **11**(9): p. e0162186.
23. Eaton, R.M., et al., *Selection of DNA aptamers for ovarian cancer biomarker HE4 using CE-SELEX and high-throughput sequencing*. Anal Bioanal Chem, 2015. **407**(23): p. 6965-73.

# *Conclusion*

---

In conclusion, my thesis focused on the development and the use of biomarkers in the context of kidney disease and its associated CVD complications.

I have presented here, and for the first time, the identification of urinary peptide biomarkers which predict with high accuracy the development of CVD in children on a CKD background. However, the peptides did not resist to multiple testing correction. This might be improved by increasing the size of the cohorts which would increase the statistical power of the identified peptides or by moving from urine to blood for peptidome profiling. Since blood is in direct contact with the heart and blood vessels, blood-derived peptides may even better reflect early changes in cardiac and vascular remodeling and will not suffer from changes with respect to glomerular filtration and tubular uptake in children with CKD. Moreover, these blood-derived peptides may be informative about the mechanisms involved in the pathophysiology of CVD.

Urinary metabolites could also be examined in this context since metabolites are theoretically closer to the phenotype and might reflect more closely the ongoing biological processes tying up information from the genome, the transcriptome, and the proteome. To fast-forward the identification of such metabolite biomarkers, we have established the proof of concept of the use of a CE-MS pipeline for the analysis of the urinary metabolome. The development of this new promising technology is certainly another step towards improvement of tools for early and more accurate diagnosis and prognosis of diseases. The actual CE-MS pipeline used for analysis of the peptidome allows identification of biomarkers in several biological fluids. In future works, it would be interesting to use the CE-MS pipeline developed in this thesis for the identification of metabolite biomarkers in other biological fluids, such as serum and plasma. Moreover, the identification of biomarkers of different molecular species, for e.g. metabolites and peptides, and/or from different biological fluid, for e.g. urine and plasma, could be combined in classifiers to improve the power of a classifier.

Furthermore, for the first time, we identified urinary peptide biomarkers for the diagnosis of CKD in dogs. The results are very promising since these urinary peptides would ensure an early diagnosis and a more standardized protocol in clinical practice. Interestingly, modifications of the canine urinary peptidome during CKD closely resemble those of the human urinary peptidome in the same context (not shown in the publication). Comparison between the two species showed several ortholog fragments (**CKD273 classifier peptide IDs:** 92410, 43543, 46649, 78332, 93417, 152967, 15593; **35P classifier peptide IDs:** 13263, 5497, 6130, 11110, 5134, 11265, 1287, respectively) originating from collagen alpha-1 (I) chain that were all less abundant in the urine in the context of CKD. This implies that both in humans and dogs, mechanisms involved in the alteration of collagen turnover are similar, have been conserved during evolution and may explain why renal fibrosis is observed as a hallmark of CKD in both species. In the past, dogs have extensively been used in nephrology in the very first studies of transplantation and dialysis [337, 338]. Our results suggest that the man's best friend could also be an

animal model that could boost our knowledge on the development and the progression of CKD in humans. For further comprehension, it is planned to continue the identification of the 98 remaining unknown peptides of the 133P classifier so that we can identify more orthologs and gain more knowledge on the mechanisms involved in the pathophysiology of CKD. Moreover, the 133P and 35P classifiers need to be validated on a larger cohort of dogs. In such larger study, it would be interesting to investigate whether these classifiers could be used to discriminate between CKD stages in dogs and more importantly be used for the prognosis of canine CKD as it has been done for the human counterpart CKD273.

Finally, while a plethora of urinary biomarkers have been identified by CE-MS, the development of biosensors for the detection of biomarkers, directly in urine using aptamers is expected to significantly reduce the cost of detection and facilitate identification at bedside, like the dipstick test. When looking at the flotilla of proof-of-principle biosensor studies available, aptamers targeting distinct types of biomarker molecules could be selected (peptides, proteins or metabolites) since each of aptamer will be very specific to its target depending upon its nucleotide sequence. This is particularly interesting since different aptamers targeting different, but most discriminating, biomarkers could be used in multiplex test in urine for a more accurate and robust diagnosis in the future [339].



# *List of references*

---

1. Webster, A.C., et al., *Chronic Kidney Disease*. Lancet, 2017. **389**(10075): p. 1238-1252.
2. Mong Hiep, T.T., et al., *Clinical characteristics and outcomes of children with stage 3-5 chronic kidney disease*. Pediatr Nephrol, 2010. **25**(5): p. 935-40.
3. Becherucci, F., et al., *Chronic kidney disease in children*. Clin Kidney J, 2016. **9**(4): p. 583-91.
4. Kaspar, C.D., R. Bholah, and T.E. Bunchman, *A Review of Pediatric Chronic Kidney Disease*. Blood Purif, 2016. **41**(1-3): p. 211-7.
5. Calderon-Margalit, R., et al., *History of Childhood Kidney Disease and Risk of Adult End-Stage Renal Disease*. N Engl J Med, 2018. **378**(5): p. 428-438.
6. Bailey, R.A., et al., *Chronic kidney disease in US adults with type 2 diabetes: an updated national estimate of prevalence based on Kidney Disease: Improving Global Outcomes (KDIGO) staging*. BMC Res Notes, 2014. **7**: p. 415.
7. Gatwood, J., et al., *Evidence of chronic kidney disease in veterans with incident diabetes mellitus*. PLoS One, 2018. **13**(2): p. e0192712.
8. Bakris, G.L., et al., *Preserving renal function in adults with hypertension and diabetes: a consensus approach. National Kidney Foundation Hypertension and Diabetes Executive Committees Working Group*. Am J Kidney Dis, 2000. **36**(3): p. 646-61.
9. Cai, G., et al., *Prevalence, awareness, treatment, and control of hypertension in elderly adults with chronic kidney disease: results from the survey of Prevalence, Awareness, and Treatment Rates in Chronic Kidney Disease Patients with Hypertension in China*. J Am Geriatr Soc, 2013. **61**(12): p. 2160-7.
10. Smith, J.M., et al., *Contributions of the Transplant Registry: The 2006 Annual Report of the North American Pediatric Renal Trials and Collaborative Studies (NAPRTCS)*. Pediatr Transplant, 2007. **11**(4): p. 366-73.
11. Nicolaou, N., et al., *Genetic, environmental, and epigenetic factors involved in CAKUT*. Nat Rev Nephrol, 2015. **11**(12): p. 720-31.
12. Capone, V.P., et al., *Genetics of Congenital Anomalies of the Kidney and Urinary Tract: The Current State of Play*. Int J Mol Sci, 2017. **18**(4).
13. Ginzburg, K., et al., *Acute Stress Disorder Symptoms Predict All-Cause Mortality Among Myocardial Infarction Patients: a 15-Year Longitudinal Study*. Ann Behav Med, 2016. **50**(2): p. 177-86.
14. Ibrahim Seif, E., et al., *Histological patterns of idiopathic steroid resistant nephrotic syndrome in Egyptian children: A single centre study*. J Nephropathol, 2013. **2**(1): p. 53-60.
15. D'Agati, V.D., *Pathobiology of focal segmental glomerulosclerosis: new developments*. Curr Opin Nephrol Hypertens, 2012. **21**(3): p. 243-50.
16. Kari, J.A., et al., *Clinico-pathological correlations of congenital and infantile nephrotic syndrome over twenty years*. Pediatr Nephrol, 2014. **29**(11): p. 2173-80.
17. Huttunen, N.P., et al., *Renal pathology in congenital nephrotic syndrome of Finnish type: a quantitative light microscopic study on 50 patients*. Int J Pediatr Nephrol, 1980. **1**(1): p. 10-6.
18. Kitiyakara, C., P. Eggers, and J.B. Kopp, *Twenty-one-year trend in ESRD due to focal segmental glomerulosclerosis in the United States*. Am J Kidney Dis, 2004. **44**(5): p. 815-25.
19. Liu, Y.C. and J. Chun, *Prospects for Precision Medicine in Glomerulonephritis Treatment*. Can J Kidney Health Dis, 2018. **5**: p. 2054358117753617.
20. Szymanik-Grzelak, H., et al., *Lupus nephritis in children - 10 years' experience*. Cent Eur J Immunol, 2016. **41**(3): p. 248-254.
21. Wenderfer, S.E. and J.P. Gaut, *Glomerular Diseases in Children*. Adv Chronic Kidney Dis, 2017. **24**(6): p. 364-371.
22. Ardissino, G., et al., *Epidemiology of chronic renal failure in children: data from the Italkid project*. Pediatrics, 2003. **111**(4 Pt 1): p. e382-7.
23. *The EMMES Corporation, in Annual report*. 2008: Rockville, MD.
24. Rodriguez, M.M., *Congenital Anomalies of the Kidney and the Urinary Tract (CAKUT)*. Fetal Pediatr Pathol, 2014. **33**(5-6): p. 293-320.

25. McDonald, S.P., et al., *Long-term survival of children with end-stage renal disease*. N Engl J Med, 2004. **350**(26): p. 2654-62.
26. Groothoff, J.W., et al., *Mortality and causes of death of end-stage renal disease in children: a Dutch cohort study*. Kidney Int, 2002. **61**(2): p. 621-9.
27. Oh, J., et al., *Advanced coronary and carotid arteriopathy in young adults with childhood-onset chronic renal failure*. Circulation, 2002. **106**(1): p. 100-5.
28. Mitsnefes, M.M., *Cardiovascular disease in children with chronic kidney disease*. J Am Soc Nephrol, 2012. **23**(4): p. 578-85.
29. Johnson, E.J., B.P. Dieter, and S.A. Marsh, *Evidence for distinct effects of exercise in different cardiac hypertrophic disorders*. Life Sci, 2015. **123**: p. 100-6.
30. Drukker, A., J. Urbach, and J. Glaser, *Hypertrophic cardiomyopathy in children with end-stage renal disease and hypertension*. Proc Eur Dial Transplant Assoc, 1981. **18**: p. 542-7.
31. Goren, A., J. Glaser, and A. Drukker, *Diastolic function in children and adolescents on dialysis and after kidney transplantation: an echocardiographic assessment*. Pediatr Nephrol, 1993. **7**(6): p. 725-8.
32. Mitsnefes, M.M., et al., *Severe left ventricular hypertrophy in pediatric dialysis: prevalence and predictors*. Pediatr Nephrol, 2000. **14**(10-11): p. 898-902.
33. Ulinski, T., et al., *Reduction of left ventricular hypertrophy in children undergoing hemodialysis*. Pediatr Nephrol, 2006. **21**(8): p. 1171-8.
34. Chinali, M., et al., *Reduced systolic myocardial function in children with chronic renal insufficiency*. J Am Soc Nephrol, 2007. **18**(2): p. 593-8.
35. Chinali, M., et al., *Left Ventricular Mass Indexing in Infants, Children, and Adolescents: A Simplified Approach for the Identification of Left Ventricular Hypertrophy in Clinical Practice*. J Pediatr, 2016. **170**: p. 193-8.
36. Boyd, A.C., N.B. Schiller, and L. Thomas, *Principles of transthoracic echocardiographic evaluation*. Nat Rev Cardiol, 2015. **12**(7): p. 426-40.
37. Kupferman, J.C., et al., *BP control and left ventricular hypertrophy regression in children with CKD*. J Am Soc Nephrol, 2014. **25**(1): p. 167-74.
38. Sinha, M.D., et al., *Blood pressure control and left ventricular mass in children with chronic kidney disease*. Clin J Am Soc Nephrol, 2011. **6**(3): p. 543-51.
39. Matteucci, M.C., et al., *Left ventricular geometry in children with mild to moderate chronic renal insufficiency*. J Am Soc Nephrol, 2006. **17**(1): p. 218-26.
40. Mitsnefes, M.M., et al., *Left ventricular mass and systolic performance in pediatric patients with chronic renal failure*. Circulation, 2003. **107**(6): p. 864-8.
41. Bullington, N., et al., *Left ventricular hypertrophy in pediatric kidney transplant recipients: long-term follow-up study*. Pediatr Transplant, 2006. **10**(7): p. 811-5.
42. Weaver, D.J., Jr., et al., *Cardiac output and associated left ventricular hypertrophy in pediatric chronic kidney disease*. Pediatr Nephrol, 2009. **24**(3): p. 565-70.
43. Wilson, A.C., et al., *High prevalence of the metabolic syndrome and associated left ventricular hypertrophy in pediatric renal transplant recipients*. Pediatr Transplant, 2010. **14**(1): p. 52-60.
44. Mitsnefes, M., et al., *Masked hypertension associates with left ventricular hypertrophy in children with CKD*. J Am Soc Nephrol, 2010. **21**(1): p. 137-44.
45. Cho, H., et al., *Influence of the Method of Definition on the Prevalence of Left-Ventricular Hypertrophy in Children with Chronic Kidney Disease: Data from the Know-Ped CKD Study*. Kidney Blood Press Res, 2017. **42**(3): p. 406-415.
46. Ruebner, R.L., et al., *Cardiovascular Disease Risk Factors and Left Ventricular Hypertrophy in Girls and Boys With CKD*. Clin J Am Soc Nephrol, 2016. **11**(11): p. 1962-1968.
47. Vanholder, R., S. Van Laecke, and G. Glorieux, *What is new in uremic toxicity?* Pediatr Nephrol, 2008. **23**(8): p. 1211-21.
48. Snauwaert, E., et al., *Concentrations of representative uraemic toxins in a healthy versus non-dialysis chronic kidney disease paediatric population*. Nephrol Dial Transplant, 2017.



49. Zoccali, C., et al., *Left ventricular hypertrophy, cardiac remodeling and asymmetric dimethylarginine (ADMA) in hemodialysis patients*. *Kidney Int*, 2002. **62**(1): p. 339-45.
50. Mervaala, E., et al., *Metabolomics in angiotensin II-induced cardiac hypertrophy*. *Hypertension*, 2010. **55**(2): p. 508-15.
51. Kuwahara, M., et al., *Cardiac remodeling associated with protein increase and lipid accumulation in early-stage chronic kidney disease in rats*. *Biochim Biophys Acta*, 2014. **1842**(9): p. 1433-43.
52. Hung, S.C., et al., *Indoxyl Sulfate: A Novel Cardiovascular Risk Factor in Chronic Kidney Disease*. *J Am Heart Assoc*, 2017. **6**(2).
53. Cao, X.S., et al., *Association of indoxyl sulfate with heart failure among patients on hemodialysis*. *Clin J Am Soc Nephrol*, 2015. **10**(1): p. 111-9.
54. Mitsnefes, M.M., et al., *FGF23 and Left Ventricular Hypertrophy in Children with CKD*. *Clin J Am Soc Nephrol*, 2018. **13**(1): p. 45-52.
55. Seeherunvong, W., et al., *Fibroblast growth factor 23 and left ventricular hypertrophy in children on dialysis*. *Pediatr Nephrol*, 2012. **27**(11): p. 2129-2136.
56. Smith, K., et al., *Fibroblast growth factor 23, high-sensitivity cardiac troponin, and left ventricular hypertrophy in CKD*. *Am J Kidney Dis*, 2013. **61**(1): p. 67-73.
57. Grabner, A., et al., *FGF23/FGFR4-mediated left ventricular hypertrophy is reversible*. *Sci Rep*, 2017. **7**(1): p. 1993.
58. Di Marco, G.S., et al., *Treatment of established left ventricular hypertrophy with fibroblast growth factor receptor blockade in an animal model of CKD*. *Nephrol Dial Transplant*, 2014. **29**(11): p. 2028-35.
59. Grabner, A., et al., *Activation of Cardiac Fibroblast Growth Factor Receptor 4 Causes Left Ventricular Hypertrophy*. *Cell Metab*, 2015. **22**(6): p. 1020-32.
60. Matteucci, M.C., et al., *Change in cardiac geometry and function in CKD children during strict BP control: a randomized study*. *Clin J Am Soc Nephrol*, 2013. **8**(2): p. 203-10.
61. Morris, K.P., et al., *Non-cardiac benefits of human recombinant erythropoietin in end stage renal failure and anaemia*. *Arch Dis Child*, 1993. **69**(5): p. 580-6.
62. Chesnaye, N.C., et al., *Mortality risk in European children with end-stage renal disease on dialysis*. *Kidney Int*, 2016. **89**(6): p. 1355-62.
63. Maron, B.J., et al., *Relationship of race to sudden cardiac death in competitive athletes with hypertrophic cardiomyopathy*. *J Am Coll Cardiol*, 2003. **41**(6): p. 974-80.
64. Spirito, P., et al., *Magnitude of left ventricular hypertrophy and risk of sudden death in hypertrophic cardiomyopathy*. *N Engl J Med*, 2000. **342**(24): p. 1778-85.
65. Griffin, M.L., et al., *Dilated cardiomyopathy in infants and children*. *J Am Coll Cardiol*, 1988. **11**(1): p. 139-44.
66. Wiles, H.B., et al., *Prognostic features of children with idiopathic dilated cardiomyopathy*. *Am J Cardiol*, 1991. **68**(13): p. 1372-6.
67. Chavers, B.M., et al., *Diagnosis of cardiac disease in pediatric end-stage renal disease*. *Nephrol Dial Transplant*, 2011. **26**(5): p. 1640-5.
68. Schlieper, G., et al., *Vascular calcification in chronic kidney disease: an update*. *Nephrol Dial Transplant*, 2016. **31**(1): p. 31-9.
69. Shroff, R., D.A. Long, and C. Shanahan, *Mechanistic insights into vascular calcification in CKD*. *J Am Soc Nephrol*, 2013. **24**(2): p. 179-89.
70. Urbina, E.M., et al., *Noninvasive assessment of subclinical atherosclerosis in children and adolescents: recommendations for standard assessment for clinical research: a scientific statement from the American Heart Association*. *Hypertension*, 2009. **54**(5): p. 919-50.
71. Hidvegi, E.V., et al., *Reference values of aortic pulse wave velocity in a large healthy population aged between 3 and 18 years*. *J Hypertens*, 2012. **30**(12): p. 2314-21.
72. Reusz, G.S., et al., *Reference values of pulse wave velocity in healthy children and teenagers*. *Hypertension*, 2010. **56**(2): p. 217-24.

73. Thurn, D., et al., *Aortic Pulse Wave Velocity in Healthy Children and Adolescents: Reference Values for the Vicorder Device and Modifying Factors*. Am J Hypertens, 2015. **28**(12): p. 1480-8.
74. Shroff, R.C., et al., *Dialysis accelerates medial vascular calcification in part by triggering smooth muscle cell apoptosis*. Circulation, 2008. **118**(17): p. 1748-57.
75. Wesseling-Perry, K., et al., *Early skeletal and biochemical alterations in pediatric chronic kidney disease*. Clin J Am Soc Nephrol, 2012. **7**(1): p. 146-52.
76. Bacchetta, J., et al., *The consequences of chronic kidney disease on bone metabolism and growth in children*. Nephrol Dial Transplant, 2012. **27**(8): p. 3063-71.
77. Rees, L. and R. Shroff, *The demise of calcium-based phosphate binders-is this appropriate for children?* Pediatr Nephrol, 2015. **30**(12): p. 2061-71.
78. Durham, A.L., et al., *Role of smooth muscle cells in vascular calcification: implications in atherosclerosis and arterial stiffness*. Cardiovasc Res, 2018. **114**(4): p. 590-600.
79. Reynolds, J.L., et al., *Human vascular smooth muscle cells undergo vesicle-mediated calcification in response to changes in extracellular calcium and phosphate concentrations: a potential mechanism for accelerated vascular calcification in ESRD*. J Am Soc Nephrol, 2004. **15**(11): p. 2857-67.
80. Tani, T., et al., *Development of a novel chronic kidney disease mouse model to evaluate the progression of hyperphosphatemia and associated mineral bone disease*. Sci Rep, 2017. **7**(1): p. 2233.
81. Chavkin, N.W., et al., *Phosphate uptake-independent signaling functions of the type III sodium-dependent phosphate transporter, PiT-1, in vascular smooth muscle cells*. Exp Cell Res, 2015. **333**(1): p. 39-48.
82. Speer, M.Y., et al., *Runx2/Cbfa1, but not loss of myocardin, is required for smooth muscle cell lineage reprogramming toward osteochondrogenesis*. J Cell Biochem, 2010. **110**(4): p. 935-47.
83. Lin, M.E., et al., *Runx2 Expression in Smooth Muscle Cells Is Required for Arterial Medial Calcification in Mice*. Am J Pathol, 2015. **185**(7): p. 1958-69.
84. Graciolli, F.G., et al., *Phosphorus overload and PTH induce aortic expression of Runx2 in experimental uraemia*. Nephrol Dial Transplant, 2009. **24**(5): p. 1416-21.
85. Jahnen-Dechent, W., et al., *Fetuin-A regulation of calcified matrix metabolism*. Circ Res, 2011. **108**(12): p. 1494-509.
86. Schaible, J., et al., *Serum fetuin-A and vitamin D in children with mild-to-severe chronic kidney disease: a cross-sectional study*. Nephrol Dial Transplant, 2012. **27**(3): p. 1107-13.
87. Shroff, R.C., et al., *The circulating calcification inhibitors, fetuin-A and osteoprotegerin, but not matrix Gla protein, are associated with vascular stiffness and calcification in children on dialysis*. Nephrol Dial Transplant, 2008. **23**(10): p. 3263-71.
88. Wesseling-Perry, K. and I.B. Salusky, *Phosphate binders, vitamin D and calcimimetics in the management of chronic kidney disease-mineral bone disorders (CKD-MBD) in children*. Pediatr Nephrol, 2013. **28**(4): p. 617-25.
89. Shroff, R., et al., *Ergocalciferol supplementation in children with CKD delays the onset of secondary hyperparathyroidism: a randomized trial*. Clin J Am Soc Nephrol, 2012. **7**(2): p. 216-23.
90. Aytac, M.B., et al., *Effect of cholecalciferol on local arterial stiffness and endothelial dysfunction in children with chronic kidney disease*. Pediatr Nephrol, 2016. **31**(2): p. 267-77.
91. Mitsnes, M.M., *Cardiovascular complications of pediatric chronic kidney disease*. Pediatr Nephrol, 2008. **23**(1): p. 27-39.
92. Ross, R., *The pathogenesis of atherosclerosis: a perspective for the 1990s*. Nature, 1993. **362**(6423): p. 801-9.
93. Brady, T.M., et al., *Carotid intima-media thickness in children with CKD: results from the CKiD study*. Clin J Am Soc Nephrol, 2012. **7**(12): p. 1930-7.
94. Chavers, B.M. and C.A. Herzog, *The spectrum of cardiovascular disease in children with predialysis chronic kidney disease*. Adv Chronic Kidney Dis, 2004. **11**(3): p. 319-27.

95. Miyagi, M., et al., *Impact of renal function on coronary plaque composition*. Nephrol Dial Transplant, 2010. **25**(1): p. 175-81.
96. Hayano, S., et al., *Relation between estimated glomerular filtration rate and composition of coronary arterial atherosclerotic plaques*. Am J Cardiol, 2012. **109**(8): p. 1131-6.
97. Goodman, W.G., et al., *Vascular calcification in chronic kidney disease*. Am J Kidney Dis, 2004. **43**(3): p. 572-9.
98. Doherty, T.M., et al., *Calcification in atherosclerosis: bone biology and chronic inflammation at the arterial crossroads*. Proc Natl Acad Sci U S A, 2003. **100**(20): p. 11201-6.
99. Alexopoulos, N. and P. Raggi, *Calcification in atherosclerosis*. Nat Rev Cardiol, 2009. **6**(11): p. 681-8.
100. Odink, A.E., et al., *The association of arterial stiffness and arterial calcification: the Rotterdam study*. J Hum Hypertens, 2008. **22**(3): p. 205-7.
101. van Popele, N.M., et al., *Association between arterial stiffness and atherosclerosis: the Rotterdam Study*. Stroke, 2001. **32**(2): p. 454-60.
102. Berenson, G.S., et al., *Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. The Bogalusa Heart Study*. N Engl J Med, 1998. **338**(23): p. 1650-6.
103. Lauer, R.M., et al., *Coronary heart disease risk factors in school children: the Muscatine study*. J Pediatr, 1975. **86**(5): p. 697-706.
104. Wilson, A.C., et al., *Prevalence and correlates of multiple cardiovascular risk factors in children with chronic kidney disease*. Clin J Am Soc Nephrol, 2011. **6**(12): p. 2759-65.
105. Nayir, A., et al., *Arterial changes in paediatric haemodialysis patients undergoing renal transplantation*. Nephrol Dial Transplant, 2001. **16**(10): p. 2041-7.
106. Libby, P., *Inflammation in atherosclerosis*. Nature, 2002. **420**(6917): p. 868-74.
107. Goldstein, S.L., et al., *Acute and chronic inflammation in pediatric patients receiving hemodialysis*. J Pediatr, 2003. **143**(5): p. 653-7.
108. Lindner, A., et al., *Accelerated atherosclerosis in prolonged maintenance hemodialysis*. N Engl J Med, 1974. **290**(13): p. 697-701.
109. Milliner, D.S., et al., *Soft tissue calcification in pediatric patients with end-stage renal disease*. Kidney Int, 1990. **38**(5): p. 931-6.
110. Ibels, L.S., et al., *Arterial calcification and pathology in uremic patients undergoing dialysis*. Am J Med, 1979. **66**(5): p. 790-6.
111. Muteliefu, G., A. Enomoto, and T. Niwa, *Indoxyl sulfate promotes proliferation of human aortic smooth muscle cells by inducing oxidative stress*. J Ren Nutr, 2009. **19**(1): p. 29-32.
112. Yamamoto, H., et al., *Indoxyl sulfate stimulates proliferation of rat vascular smooth muscle cells*. Kidney Int, 2006. **69**(10): p. 1780-5.
113. Shimizu, H., et al., *ROS and PDGF-beta [corrected] receptors are critically involved in indoxyl sulfate actions that promote vascular smooth muscle cell proliferation and migration*. Am J Physiol Cell Physiol, 2009. **297**(2): p. C389-96.
114. Adijiang, A., et al., *Indoxyl sulphate promotes aortic calcification with expression of osteoblast-specific proteins in hypertensive rats*. Nephrol Dial Transplant, 2008. **23**(6): p. 1892-901.
115. Dou, L., et al., *The uremic solutes p-cresol and indoxyl sulfate inhibit endothelial proliferation and wound repair*. Kidney Int, 2004. **65**(2): p. 442-51.
116. Brunet, P., et al., *Does uremia cause vascular dysfunction?* Kidney Blood Press Res, 2011. **34**(4): p. 284-90.
117. Siomou, E., et al., *Serum osteoprotegerin, RANKL and fibroblast growth factor-23 in children with chronic kidney disease*. Pediatr Nephrol, 2011. **26**(7): p. 1105-14.
118. Portale, A.A., et al., *Disordered FGF23 and mineral metabolism in children with CKD*. Clin J Am Soc Nephrol, 2014. **9**(2): p. 344-53.
119. Yasin, A., et al., *Fibroblast growth factor-23 and calcium phosphate product in young chronic kidney disease patients: a cross-sectional study*. BMC Nephrol, 2013. **14**: p. 39.

120. Yoon, J.M., *Dyslipidemia in children and adolescents: when and how to diagnose and treat?* *Pediatr Gastroenterol Hepatol Nutr*, 2014. **17**(2): p. 85-92.
121. *Clinical practice guidelines for nutrition in chronic renal failure. K/DOQI, National Kidney Foundation.* *Am J Kidney Dis*, 2000. **35**(6 Suppl 2): p. S1-140.
122. Lu, H., et al., *Renin inhibition reduces hypercholesterolemia-induced atherosclerosis in mice.* *J Clin Invest*, 2008. **118**(3): p. 984-93.
123. Lu, H., et al., *Comparative effects of different modes of renin angiotensin system inhibition on hypercholesterolaemia-induced atherosclerosis.* *Br J Pharmacol*, 2012. **165**(6): p. 2000-2008.
124. Stenvinkel, P., R. Pecoits-Filho, and B. Lindholm, *Coronary artery disease in end-stage renal disease: no longer a simple plumbing problem.* *J Am Soc Nephrol*, 2003. **14**(7): p. 1927-39.
125. Schieffer, B., et al., *Role of NAD(P)H oxidase in angiotensin II-induced JAK/STAT signaling and cytokine induction.* *Circ Res*, 2000. **87**(12): p. 1195-201.
126. Barra, S., et al., *Early increase of carotid intima-media thickness in children with parental history of premature myocardial infarction.* *Heart*, 2009. **95**(8): p. 642-5.
127. Litwin, M., et al., *Patient survival and causes of death on hemodialysis and peritoneal dialysis-single-center study.* *Pediatr Nephrol*, 2001. **16**(12): p. 996-1001.
128. O'Neill, D.G., et al., *Chronic kidney disease in dogs in UK veterinary practices: prevalence, risk factors, and survival.* *J Vet Intern Med*, 2013. **27**(4): p. 814-21.
129. Pelander, L., et al., *Incidence of and mortality from kidney disease in over 600,000 insured Swedish dogs.* *Vet Rec*, 2015. **176**(25): p. 656.
130. Cianciolo, R.E., S.L. Benali, and L. Aresu, *Aging in the Canine Kidney.* *Vet Pathol*, 2016. **53**(2): p. 299-308.
131. Fick, L.J., et al., *Telomere length correlates with life span of dog breeds.* *Cell Rep*, 2012. **2**(6): p. 1530-6.
132. Mizushima, N., *Autophagy: process and function.* *Genes Dev*, 2007. **21**(22): p. 2861-73.
133. Fougeray, S. and N. Pallet, *Mechanisms and biological functions of autophagy in diseased and ageing kidneys.* *Nat Rev Nephrol*, 2015. **11**(1): p. 34-45.
134. Valentijn, F.A., et al., *Cellular senescence in the aging and diseased kidney.* *J Cell Commun Signal*, 2018. **12**(1): p. 69-82.
135. DeBowes, L.J., et al., *Association of periodontal disease and histologic lesions in multiple organs from 45 dogs.* *J Vet Dent*, 1996. **13**(2): p. 57-60.
136. Pavlica, Z., et al., *Periodontal disease burden and pathological changes in organs of dogs.* *J Vet Dent*, 2008. **25**(2): p. 97-105.
137. Glickman, L.T., et al., *Association between chronic azotemic kidney disease and the severity of periodontal disease in dogs.* *Prev Vet Med*, 2011. **99**(2-4): p. 193-200.
138. Fisher, M.A. and G.W. Taylor, *A prediction model for chronic kidney disease includes periodontal disease.* *J Periodontol*, 2009. **80**(1): p. 16-23.
139. Fisher, M.A. and Z.Q. Ma, *The association between periodontitis and high C-reactive protein in adults with chronic kidney disease is not clearly established.* *J Evid Based Dent Pract*, 2012. **12**(3): p. 171-3.
140. De Loor, J., et al., *Urinary biomarkers for acute kidney injury in dogs.* *J Vet Intern Med*, 2013. **27**(5): p. 998-1010.
141. Bonventre, J.V. and L. Yang, *Cellular pathophysiology of ischemic acute kidney injury.* *J Clin Invest*, 2011. **121**(11): p. 4210-21.
142. Ferenbach, D.A. and J.V. Bonventre, *Mechanisms of maladaptive repair after AKI leading to accelerated kidney ageing and CKD.* *Nat Rev Nephrol*, 2015. **11**(5): p. 264-76.
143. Brown, C.A., et al., *Outbreaks of renal failure associated with melamine and cyanuric acid in dogs and cats in 2004 and 2007.* *J Vet Diagn Invest*, 2007. **19**(5): p. 525-31.
144. Bartges J. and Polzin D., *Nephrology and Urology of Small Animals.* 2011: Wiley-Blackwell.
145. Acierno, M.J. and M.A. Labato, *Hypertension in renal disease: diagnosis and treatment.* *Clin Tech Small Anim Pract*, 2005. **20**(1): p. 23-30.

146. Syme, H., *Hypertension in small animal kidney disease*. Vet Clin North Am Small Anim Pract, 2011. **41**(1): p. 63-89.
147. Benali, S.L., et al., *X-Linked Hereditary Nephropathy in Navasota Dogs: Clinical Pathology, Morphology, and Gene Expression During Disease Progression*. Vet Pathol, 2016. **53**(4): p. 803-12.
148. Lees, G.E., et al., *A model of autosomal recessive Alport syndrome in English cocker spaniel dogs*. Kidney Int, 1998. **54**(3): p. 706-19.
149. Chu, C.P., et al., *RNA-seq of serial kidney biopsies obtained during progression of chronic kidney disease from dogs with X-linked hereditary nephropathy*. Sci Rep, 2017. **7**(1): p. 16776.
150. Kim, J., et al., *Multicystic dysplastic kidney disease in a dog*. Can Vet J, 2011. **52**(6): p. 645-9.
151. Gharahkhani, P., et al., *A non-synonymous mutation in the canine Pkd1 gene is associated with autosomal dominant polycystic kidney disease in Bull Terriers*. PLoS One, 2011. **6**(7): p. e22455.
152. Morita, T., et al., *Renal dysplasia with unilateral renal agenesis in a dog*. J Comp Pathol, 2005. **133**(1): p. 64-7.
153. Bruder, M.C., et al., *Renal dysplasia in Beagle dogs: four cases*. Toxicol Pathol, 2010. **38**(7): p. 1051-7.
154. Whiteley, M.H., J.S. Bell, and D.A. Rothman, *Novel allelic variants in the canine cyclooxygenase-2 (Cox-2) promoter are associated with renal dysplasia in dogs*. PLoS One, 2011. **6**(2): p. e16684.
155. Whiteley, M.H., *Allelic variation in the canine Cox-2 promoter causes hypermethylation of the canine Cox-2 promoter in clinical cases of renal dysplasia*. Clin Epigenetics, 2014. **6**(1): p. 7.
156. Schnaper, H.W., *Remnant nephron physiology and the progression of chronic kidney disease*. Pediatr Nephrol, 2014. **29**(2): p. 193-202.
157. Polzin, D.J., *Evidence-based step-wise approach to managing chronic kidney disease in dogs and cats*. J Vet Emerg Crit Care (San Antonio), 2013. **23**(2): p. 205-15.
158. Harjes, L.M., et al., *Fibroblast Growth Factor-23 Concentration in Dogs with Chronic Kidney Disease*. J Vet Intern Med, 2017. **31**(3): p. 784-790.
159. Shipov, A., et al., *The Influence of Chronic Kidney Disease on the Structural and Mechanical Properties of Canine Bone*. J Vet Intern Med, 2018. **32**(1): p. 280-287.
160. Marone, C.C., et al., *Effects of metabolic alkalosis on calcium excretion in the conscious dog*. J Lab Clin Med, 1983. **101**(2): p. 264-73.
161. Sutton, R.A., N.L. Wong, and J.H. Dirks, *Effects of metabolic acidosis and alkalosis on sodium and calcium transport in the dog kidney*. Kidney Int, 1979. **15**(5): p. 520-33.
162. Kraut, J.A. and N.E. Madias, *Consequences and therapy of the metabolic acidosis of chronic kidney disease*. Pediatr Nephrol, 2011. **26**(1): p. 19-28.
163. Pouchelon, J.L., et al., *Cardiovascular-renal axis disorders in the domestic dog and cat: a veterinary consensus statement*. J Small Anim Pract, 2015. **56**(9): p. 537-52.
164. Nicolle, A.P., et al., *Azotemia and glomerular filtration rate in dogs with chronic valvular disease*. J Vet Intern Med, 2007. **21**(5): p. 943-9.
165. Tang, X., et al., *Renal denervation decreases susceptibility of the heart to ventricular fibrillation in a canine model of chronic kidney disease*. Exp Physiol, 2017. **102**(11): p. 1414-1423.
166. Fiocchi, E.H., et al., *The Use of Darbepoetin to Stimulate Erythropoiesis in the Treatment of Anemia of Chronic Kidney Disease in Dogs*. J Vet Intern Med, 2017. **31**(2): p. 476-485.
167. Relford, R., J. Robertson, and C. Clements, *Symmetric Dimethylarginine: Improving the Diagnosis and Staging of Chronic Kidney Disease in Small Animals*. Vet Clin North Am Small Anim Pract, 2016. **46**(6): p. 941-60.
168. Babyak, J.M., et al., *Prevalence of Elevated Serum Creatinine Concentration in Dogs Presenting to a Veterinary Academic Medical Center (2010-2014)*. J Vet Intern Med, 2017. **31**(6): p. 1757-1764.
169. Von Hendy-Willson, V.E. and B.M. Pressler, *An overview of glomerular filtration rate testing in dogs and cats*. Vet J, 2011. **188**(2): p. 156-65.

170. Tvedten, H.W. and A. Noren, *Comparison of a Schmidt and Haensch refractometer and an Atago PAL-USG Cat refractometer for determination of urine specific gravity in dogs and cats.* Vet Clin Pathol, 2014. **43**(1): p. 63-6.
171. Tvedten, H.W., H. Ouchterlony, and I.E. Lilliehook, *Comparison of specific gravity analysis of feline and canine urine, using five refractometers, to pycnometric analysis and total solids by drying.* N Z Vet J, 2015. **63**(5): p. 254-9.
172. Miyagawa, Y., et al., *Development of correction formulas for canine and feline urine specific gravity measured using a Japanese refractometer.* J Vet Med Sci, 2011. **73**(5): p. 679-81.
173. Willard, M.D. and H. Tvedten, *Small animal clinical diagnosis by laboratory methods.* 5th ed. 2012: Elsevier.
174. Rishniw, M. and R. Bicalho, *Factors affecting urine specific gravity in apparently healthy cats presenting to first opinion practice for routine evaluation.* J Feline Med Surg, 2015. **17**(4): p. 329-37.
175. Brown, S.A., et al., *Single-nephron adaptations to partial renal ablation in the dog.* Am J Physiol, 1990. **258**(3 Pt 2): p. F495-503.
176. Grauer, G.F., et al., *Effects of enalapril versus placebo as a treatment for canine idiopathic glomerulonephritis.* J Vet Intern Med, 2000. **14**(5): p. 526-33.
177. King, J.N., et al., *Effects of Benazepril on Survival of Dogs with Chronic Kidney Disease: A Multicenter, Randomized, Blinded, Placebo-Controlled Clinical Trial.* J Vet Intern Med, 2017. **31**(4): p. 1113-1122.
178. Yalcin, A. and M. Cetin, *Electrophoretic separation of urine proteins of healthy dogs and dogs with nephropathy and detection of some urine proteins in dogs using immunoblotting.* . Revue de Médecine Vétérinaire, 2004. **155**: p. 104-112.
179. Harley, L. and C. Langston, *Proteinuria in dogs and cats.* Can Vet J, 2012. **53**(6): p. 631-8.
180. Lees, G.E., et al., *Assessment and management of proteinuria in dogs and cats: 2004 ACVIM Forum Consensus Statement (small animal).* J Vet Intern Med, 2005. **19**(3): p. 377-85.
181. Duffy, M.E., A. Specht, and R.C. Hill, *Comparison between Urine Protein: Creatinine Ratios of Samples Obtained from Dogs in Home and Hospital Settings.* J Vet Intern Med, 2015. **29**(4): p. 1029-35.
182. Harison, E., et al., *Acute azotemia as a predictor of mortality in dogs and cats.* J Vet Intern Med, 2012. **26**(5): p. 1093-8.
183. Buranakarl, C., et al., *Relationships between degree of azotaemia and blood pressure, urinary protein:creatinine ratio and fractional excretion of electrolytes in dogs with renal azotaemia.* Vet Res Commun, 2007. **31**(3): p. 245-57.
184. Zotti, A., et al., *Correlation of renal histopathology with renal echogenicity in dogs and cats: an ex-vivo quantitative study.* BMC Vet Res, 2015. **11**: p. 99.
185. Bragato, N., N.C. Borges, and M.C.S. Fioravanti, *B-mode and Doppler ultrasound of chronic kidney disease in dogs and cats.* Vet Res Commun, 2017. **41**(4): p. 307-315.
186. Pressler, B.M., *Clinical approach to advanced renal function testing in dogs and cats.* Vet Clin North Am Small Anim Pract, 2013. **43**(6): p. 1193-208, v.
187. Braun, J.P., H.P. Lefebvre, and A.D. Watson, *Creatinine in the dog: a review.* Vet Clin Pathol, 2003. **32**(4): p. 162-79.
188. Yerramilli, M., et al., *Kidney Disease and the Nexus of Chronic Kidney Disease and Acute Kidney Injury: The Role of Novel Biomarkers as Early and Accurate Diagnostics.* Vet Clin North Am Small Anim Pract, 2016. **46**(6): p. 961-93.
189. Lulich, J.P., *Urine specific gravity - The most underutilized test in veterinary medicine, in European Veterinary Conference VOORJAARSDAGEN.* 2017.
190. Grauer, G.F., *Canine glomerulonephritis: new thoughts on proteinuria and treatment.* J Small Anim Pract, 2005. **46**(10): p. 469-78.
191. Herget-Rosenthal, S., *Imaging techniques in the management of chronic kidney disease: current developments and future perspectives.* Semin Nephrol, 2011. **31**(3): p. 283-90.

192. Stock, E., et al., *Contrast-Enhanced Ultrasound Examination for the Assessment of Renal Perfusion in Cats with Chronic Kidney Disease*. J Vet Intern Med, 2018. **32**(1): p. 260-266.
193. Good, D.M., et al., *Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease*. Mol Cell Proteomics, 2010. **9**(11): p. 2424-37.
194. Schanstra, J.P., et al., *Diagnosis and Prediction of CKD Progression by Assessment of Urinary Peptides*. J Am Soc Nephrol, 2015. **26**(8): p. 1999-2010.
195. Prohaska, S.J. and P.F. Stadler, *The use and abuse of -omes*. Methods Mol Biol, 2011. **719**: p. 173-96.
196. Zhang, W., F. Li, and L. Nie, *Integrating multiple 'omics' analysis for microbial biology: application and methodologies*. Microbiology, 2010. **156**(Pt 2): p. 287-301.
197. Maier, T., M. Guell, and L. Serrano, *Correlation of mRNA and protein in complex biological samples*. FEBS Lett, 2009. **583**(24): p. 3966-73.
198. Bauca, J.M., E. Martinez-Morillo, and E.P. Diamandis, *Peptidomics of urine and other biofluids for cancer diagnostics*. Clin Chem, 2014. **60**(8): p. 1052-61.
199. Schulz-Knappe, P., et al., *Peptidomics: the comprehensive analysis of peptides in complex biological mixtures*. Comb Chem High Throughput Screen, 2001. **4**(2): p. 207-17.
200. Hocher, B. and J. Adamski, *Metabolomics for clinical use and research in chronic kidney disease*. Nat Rev Nephrol, 2017. **13**(5): p. 269-284.
201. Califf, R.M., *Biomarker definitions and their applications*. Exp Biol Med (Maywood), 2018. **243**(3): p. 213-221.
202. Mischak, H., et al., *Recommendations for biomarker identification and qualification in clinical proteomics*. Sci Transl Med, 2010. **2**(46): p. 46ps42.
203. Mischak, H., et al., *Clinical proteomics: A need to define the field and to begin to set adequate standards*. Proteomics Clin Appl, 2007. **1**(2): p. 148-56.
204. Harpole, M., J. Davis, and V. Espina, *Current state of the art for enhancing urine biomarker discovery*. Expert Rev Proteomics, 2016. **13**(6): p. 609-26.
205. Orton, D.J. and A.A. Doucette, *Proteomic Workflows for Biomarker Identification Using Mass Spectrometry - Technical and Statistical Considerations during Initial Discovery*. Proteomes, 2013. **1**(2): p. 109-127.
206. Li, Z., et al., *Use of the local false discovery rate for identification of metabolic biomarkers in rat urine following Genkwa Flos-induced hepatotoxicity*. PLoS One, 2013. **8**(7): p. e67451.
207. Yang, Z.R., *Biological applications of support vector machines*. Brief Bioinform, 2004. **5**(4): p. 328-38.
208. Wolf, B.J., et al., *Development of Biomarker Models to Predict Outcomes in Lupus Nephritis*. Arthritis Rheumatol, 2016. **68**(8): p. 1955-63.
209. Han, X., A. Aslanian, and J.R. Yates, 3rd, *Mass spectrometry for proteomics*. Curr Opin Chem Biol, 2008. **12**(5): p. 483-90.
210. Yalcin, E.B. and S.M. de la Monte, *Review of matrix-assisted laser desorption ionization-imaging mass spectrometry for lipid biochemical histopathology*. J Histochem Cytochem, 2015. **63**(10): p. 762-71.
211. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules*. Science, 1989. **246**(4926): p. 64-71.
212. Yi, E.C., et al., *A microcapillary trap cartridge-microcapillary high-performance liquid chromatography electrospray ionization emitter device capable of peptide tandem mass spectrometry at the attomole level on an ion trap mass spectrometer with automated routine operation*. Rapid Commun Mass Spectrom, 2003. **17**(18): p. 2093-8.
213. Hoffmann, P., et al., *Microfluidic glass chips with an integrated nanospray emitter for coupling to a mass spectrometer*. Angew Chem Int Ed Engl, 2007. **46**(26): p. 4913-6.
214. Sun, L., et al., *Ultrasensitive and fast bottom-up analysis of femtogram amounts of complex proteome digests*. Angew Chem Int Ed Engl, 2013. **52**(51): p. 13661-4.
215. Page, J.S., et al., *Ionization and transmission efficiency in an electrospray ionization-mass spectrometry interface*. J Am Soc Mass Spectrom, 2007. **18**(9): p. 1582-90.

216. Schlosser, A. and R. Volkmer-Engert, *Volatile polydimethylcyclsiloxanes in the ambient laboratory air identified as source of extreme background signals in nanoelectrospray mass spectrometry*. J Mass Spectrom, 2003. **38**(5): p. 523-5.
217. Konermann, L., et al., *Unraveling the mechanism of electrospray ionization*. Anal Chem, 2013. **85**(1): p. 2-9.
218. Demartini, D.R., *A short overview of the components in mass spectrometry instrumentation for proteomics analyses*. Tandem Mass spectrometry. 2013: IntechOpen.
219. Bohrer, B.C., et al., *Biomolecule analysis by ion mobility spectrometry*. Annu Rev Anal Chem (Palo Alto Calif), 2008. **1**: p. 293-327.
220. Walther, T.C. and M. Mann, *Mass spectrometry-based proteomics in cell biology*. J Cell Biol, 2010. **190**(4): p. 491-500.
221. Koppenaal, D.W., et al., *MS detectors*. Anal Chem, 2005. **77**(21): p. 418A-427A.
222. Thongboonkerd, V., *Practical points in urinary proteomics*. J Proteome Res, 2007. **6**(10): p. 3881-90.
223. Rogowska-Wrzesinska, A., et al., *2D gels still have a niche in proteomics*. J Proteomics, 2013. **88**: p. 4-13.
224. Jerebtsova, M. and S. Nekhai, *Quantitative mass spectrometry of urinary biomarkers*. J Integr OMICS, 2014. **4**(2): p. 69-78.
225. Allwood, J.W. and R. Goodacre, *An introduction to liquid chromatography-mass spectrometry instrumentation applied in plant metabolomic analyses*. Phytochem Anal, 2010. **21**(1): p. 33-47.
226. Guo, K., F. Bamforth, and L. Li, *Qualitative metabolome analysis of human cerebrospinal fluid by <sup>13</sup>C/<sup>12</sup>C-isotope dansylation labeling combined with liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry*. J Am Soc Mass Spectrom, 2011. **22**(2): p. 339-47.
227. Metz, T.O., et al., *The future of liquid chromatography-mass spectrometry (LC-MS) in metabolic profiling and metabolomic studies for biomarker discovery*. Biomark Med, 2007. **1**(1): p. 159-185.
228. Zhang, W., T. Hankemeier, and R. Ramautar, *Next-generation capillary electrophoresis-mass spectrometry approaches in metabolomics*. Curr Opin Biotechnol, 2017. **43**: p. 1-7.
229. Haselberg, R., G.J. de Jong, and G.W. Somsen, *Capillary electrophoresis-mass spectrometry for the analysis of intact proteins 2007-2010*. Electrophoresis, 2011. **32**(1): p. 66-82.
230. Pontillo, C. and H. Mischak, *Urinary peptide-based classifier CKD273: towards clinical application in chronic kidney disease*. Clin Kidney J, 2017. **10**(2): p. 192-201.
231. Gaspar, A., et al., *Trends in CE-MS 2005-2006*. Electrophoresis, 2008. **29**(1): p. 66-79.
232. Neususs, C., M. Pelzing, and M. Macht, *A robust approach for the analysis of peptides in the low femtomole range by capillary electrophoresis-tandem mass spectrometry*. Electrophoresis, 2002. **23**(18): p. 3149-59.
233. Ramautar, R., G.W. Somsen, and G.J. de Jong, *CE-MS for metabolomics: developments and applications in the period 2012-2014*. Electrophoresis, 2015. **36**(1): p. 212-24.
234. Kolch, W., et al., *Capillary electrophoresis-mass spectrometry as a powerful tool in clinical diagnosis and biomarker discovery*. Mass Spectrom Rev, 2005. **24**(6): p. 959-77.
235. Avent, N.D., et al., *Post-genomics studies and their application to non-invasive prenatal diagnosis*. Semin Fetal Neonatal Med, 2008. **13**(2): p. 91-8.
236. Metzger, J., et al., *Adapting mass spectrometry-based platforms for clinical proteomics applications: The capillary electrophoresis coupled mass spectrometry paradigm*. Crit Rev Clin Lab Sci, 2009. **46**(3): p. 129-52.
237. Shen, Y., et al., *Characterization of the human blood plasma proteome*. Proteomics, 2005. **5**(15): p. 4034-45.
238. Yu, Z., et al., *Differences between human plasma and serum metabolite profiles*. PLoS One, 2011. **6**(7): p. e21230.



239. Du, L., D.G. Musson, and A.Q. Wang, *Stability studies of vorinostat and its two metabolites in human plasma, serum and urine*. J Pharm Biomed Anal, 2006. **42**(5): p. 556-64.
240. Kamlage, B., et al., *Impact of Prolonged Blood Incubation and Extended Serum Storage at Room Temperature on the Human Serum Metabolome*. Metabolites, 2018. **8**(1).
241. Mischak, H., et al., *Epidemiologic design and analysis for proteomic studies: a primer on -omic technologies*. Am J Epidemiol, 2015. **181**(9): p. 635-47.
242. Klein, J., et al., *The role of urinary peptidomics in kidney disease research*. Kidney Int, 2016. **89**(3): p. 539-45.
243. Good, D.M., et al., *Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future*. J Proteome Res, 2007. **6**(12): p. 4549-55.
244. Bouatra, S., et al., *The human urine metabolome*. PLoS One, 2013. **8**(9): p. e73076.
245. Gika, H.G., et al., *Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples*. J Chromatogr B Analyt Technol Biomed Life Sci, 2008. **871**(2): p. 299-305.
246. Xiao, Q., et al., *Sources of variability in metabolite measurements from urinary samples*. PLoS One, 2014. **9**(5): p. e95749.
247. Lauridsen, M., et al., *Human urine as test material in 1H NMR-based metabonomics: recommendations for sample preparation and storage*. Anal Chem, 2007. **79**(3): p. 1181-6.
248. Rotter, M., et al., *Stability of targeted metabolite profiles of urine samples under different storage conditions*. Metabolomics, 2017. **13**(1): p. 4.
249. Critselis, E. and H. Lambers Heerspink, *Utility of the CKD273 peptide classifier in predicting chronic kidney disease progression*. Nephrol Dial Transplant, 2016. **31**(2): p. 249-54.
250. Argiles, A., et al., *CKD273, a new proteomics classifier assessing CKD and its prognosis*. PLoS One, 2013. **8**(5): p. e62837.
251. Siwy, J., et al., *Multicentre prospective validation of a urinary peptidome-based classifier for the diagnosis of type 2 diabetic nephropathy*. Nephrol Dial Transplant, 2014. **29**(8): p. 1563-70.
252. Currie, G.E., et al., *Urinary proteomics for prediction of mortality in patients with type 2 diabetes and microalbuminuria*. Cardiovasc Diabetol, 2018. **17**(1): p. 50.
253. Lindhardt, M., et al., *Proteomic prediction and Renin angiotensin aldosterone system Inhibition prevention Of early diabetic nephropathy in Type 2 diabetic patients with normoalbuminuria (PRIORITY): essential study design and rationale of a randomised clinical multicentre trial*. BMJ Open, 2016. **6**(3): p. e010310.
254. Decramer, S., et al., *Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis*. Nat Med, 2006. **12**(4): p. 398-400.
255. Decramer, S., et al., *Identification of urinary biomarkers by proteomics in newborns: use in obstructive nephropathy*. Contrib Nephrol, 2008. **160**: p. 127-41.
256. Decramer, S., et al., *Urine in clinical proteomics*. Mol Cell Proteomics, 2008. **7**(10): p. 1850-62.
257. Drube, J., et al., *Urinary proteome analysis identifies infants but not older children requiring pyeloplasty*. Pediatr Nephrol, 2010. **25**(9): p. 1673-8.
258. Drozd, D., et al., *Progression to end-stage renal disease in children with posterior urethral valves*. Pediatr Nephrol, 1998. **12**(8): p. 630-6.
259. Parkhouse, H.F., et al., *Long-term outcome of boys with posterior urethral valves*. Br J Urol, 1988. **62**(1): p. 59-62.
260. Lopez Pereira, P., et al., *Posterior urethral valves: prognostic factors*. BJU Int, 2003. **91**(7): p. 687-90.
261. Morris, R.K., et al., *Systematic review of accuracy of fetal urine analysis to predict poor postnatal renal function in cases of congenital urinary tract obstruction*. Prenat Diagn, 2007. **27**(10): p. 900-11.
262. Morris, R.K., et al., *Antenatal ultrasound to predict postnatal renal function in congenital lower urinary tract obstruction: systematic review of test accuracy*. BJOG, 2009. **116**(10): p. 1290-9.

263. Klein, J., et al., *Fetal urinary peptides to predict postnatal outcome of renal disease in fetuses with posterior urethral valves (PUV)*. *Sci Transl Med*, 2013. **5**(198): p. 198ra106.
264. Zimmerli, L.U., et al., *Urinary proteomic biomarkers in coronary artery disease*. *Mol Cell Proteomics*, 2008. **7**(2): p. 290-8.
265. Ponticos, M. and B.D. Smith, *Extracellular matrix synthesis in vascular disease: hypertension, and atherosclerosis*. *J Biomed Res*, 2014. **28**(1): p. 25-39.
266. von Zur Muhlen, C., et al., *Evaluation of urine proteome pattern analysis for its potential to reflect coronary artery atherosclerosis in symptomatic patients*. *J Proteome Res*, 2009. **8**(1): p. 335-45.
267. Htun, N.M., et al., *Prediction of acute coronary syndromes by urinary proteome analysis*. *PLoS One*, 2017. **12**(3): p. e0172036.
268. Dawson, J., et al., *Urinary proteomics to support diagnosis of stroke*. *PLoS One*, 2012. **7**(5): p. e35879.
269. Kuznetsova, T., et al., *Urinary proteome analysis in hypertensive patients with left ventricular diastolic dysfunction*. *Eur Heart J*, 2012. **33**(18): p. 2342-50.
270. Rossing, K., et al., *Urinary Proteomics Pilot Study for Biomarker Discovery and Diagnosis in Heart Failure with Reduced Ejection Fraction*. *PLoS One*, 2016. **11**(6): p. e0157167.
271. Farmakis, D., et al., *Urine proteome analysis in heart failure with reduced ejection fraction complicated by chronic kidney disease: feasibility, and clinical and pathogenetic correlates*. *Eur J Heart Fail*, 2016. **18**(7): p. 822-9.
272. Madero, M. and M.J. Sarnak, *Creatinine-based formulae for estimating glomerular filtration rate: is it time to change to chronic kidney disease epidemiology collaboration equation?* *Curr Opin Nephrol Hypertens*, 2011. **20**(6): p. 622-30.
273. Levey, A.S., L.A. Inker, and J. Coresh, *GFR estimation: from physiology to public health*. *Am J Kidney Dis*, 2014. **63**(5): p. 820-34.
274. Nkuipou-Kenfack, E., et al., *Assessment of metabolomic and proteomic biomarkers in detection and prognosis of progression of renal function in chronic kidney disease*. *PLoS One*, 2014. **9**(5): p. e96955.
275. Rysz, J., et al., *Novel Biomarkers in the Diagnosis of Chronic Kidney Disease and the Prediction of Its Outcome*. *Int J Mol Sci*, 2017. **18**(8).
276. Sharma, K., et al., *Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease*. *J Am Soc Nephrol*, 2013. **24**(11): p. 1901-12.
277. Pena, M.J., et al., *Prognostic clinical and molecular biomarkers of renal disease in type 2 diabetes*. *Nephrol Dial Transplant*, 2015. **30 Suppl 4**: p. iv86-95.
278. Pena, M.J., et al., *Urine and plasma metabolites predict the development of diabetic nephropathy in individuals with Type 2 diabetes mellitus*. *Diabet Med*, 2014. **31**(9): p. 1138-47.
279. Kimura, T., et al., *Identification of biomarkers for development of end-stage kidney disease in chronic kidney disease by metabolomic profiling*. *Sci Rep*, 2016. **6**: p. 26138.
280. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase*. *Science*, 1990. **249**(4968): p. 505-10.
281. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands*. *Nature*, 1990. **346**(6287): p. 818-22.
282. Gold, L., *SELEX: How It Happened and Where It will Go*. *J Mol Evol*, 2015. **81**(5-6): p. 140-3.
283. Ruscito, A. and M.C. DeRosa, *Small-Molecule Binding Aptamers: Selection Strategies, Characterization, and Applications*. *Front Chem*, 2016. **4**: p. 14.
284. Handy, S.M., et al., *First report of the use of a saxitoxin-protein conjugate to develop a DNA aptamer to a small molecule toxin*. *Toxicon*, 2013. **61**: p. 30-7.
285. Citartan, M., et al., *Assays for aptamer-based platforms*. *Biosens Bioelectron*, 2012. **34**(1): p. 1-11.
286. Lato, S.M., et al., *Boron-containing aptamers to ATP*. *Nucleic Acids Res*, 2002. **30**(6): p. 1401-7.

287. Long, F., et al., *Rapid on-site/in-situ detection of heavy metal ions in environmental water using a structure-switching DNA optical biosensor*. Sci Rep, 2013. **3**: p. 2308.
288. Chung, C.H., et al., *Nuclease-resistant DNA aptamer on gold nanoparticles for the simultaneous detection of Pb<sup>2+</sup> and Hg<sup>2+</sup> in human serum*. Biosens Bioelectron, 2013. **41**: p. 827-32.
289. Bock, L.C., et al., *Selection of single-stranded DNA molecules that bind and inhibit human thrombin*. Nature, 1992. **355**(6360): p. 564-6.
290. Dua, P., S. Kim, and D.K. Lee, *Nucleic acid aptamers targeting cell-surface proteins*. Methods, 2011. **54**(2): p. 215-25.
291. Jenison, R.D., et al., *High-resolution molecular discrimination by RNA*. Science, 1994. **263**(5152): p. 1425-9.
292. Volk, D.E. and G.L.R. Lokesh, *Development of Phosphorothioate DNA and DNA Thioaptamers*. Biomedicines, 2017. **5**(3).
293. Song, K.M., S. Lee, and C. Ban, *Aptamers and their biological applications*. Sensors (Basel), 2012. **12**(1): p. 612-31.
294. Birch, J.R. and A.J. Racher, *Antibody production*. Adv Drug Deliv Rev, 2006. **58**(5-6): p. 671-85.
295. Gopinath, S.C., *Antiviral aptamers*. Arch Virol, 2007. **152**(12): p. 2137-57.
296. Gold, L., et al., *Aptamers and the RNA world, past and present*. Cold Spring Harb Perspect Biol, 2012. **4**(3).
297. McKeague, M., et al., *Analysis of In Vitro Aptamer Selection Parameters*. J Mol Evol, 2015. **81**(5-6): p. 150-61.
298. Paul, A., et al., *Streptavidin-coated magnetic beads for DNA strand separation implicate a multitude of problems during cell-SELEX*. Oligonucleotides, 2009. **19**(3): p. 243-54.
299. Wiegand, T.W., et al., *High-affinity oligonucleotide ligands to human IgE inhibit binding to Fc epsilon receptor I*. J Immunol, 1996. **157**(1): p. 221-30.
300. Cox, J.C., et al., *Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer*. Nucleic Acids Res, 2002. **30**(20): p. e108.
301. Thiviyanathan, V. and D.G. Gorenstein, *Aptamers and the next generation of diagnostic reagents*. Proteomics Clin Appl, 2012. **6**(11-12): p. 563-73.
302. Mondal, B., et al., *A combinatorial systematic evolution of ligands by exponential enrichment method for selection of aptamer against protein targets*. Appl Microbiol Biotechnol, 2015. **99**(22): p. 9791-803.
303. Murphy, M.B., et al., *An improved method for the in vitro evolution of aptamers and applications in protein detection and purification*. Nucleic Acids Res, 2003. **31**(18): p. e110.
304. Duan, N., et al., *An ssDNA library immobilized SELEX technique for selection of an aptamer against ractopamine*. Anal Chim Acta, 2017. **961**: p. 100-105.
305. Hunniger, T., et al., *Just in time-selection: A rapid semiautomated SELEX of DNA aptamers using magnetic separation and BEAMing*. Anal Chem, 2014. **86**(21): p. 10940-7.
306. Forier, C., et al., *DNA aptamer affinity ligands for highly selective purification of human plasma-related proteins from multiple sources*. J Chromatogr A, 2017. **1489**: p. 39-50.
307. Ferreira, C.S., C.S. Matthews, and S. Missailidis, *DNA aptamers that bind to MUC1 tumour marker: design and characterization of MUC1-binding single-stranded DNA aptamers*. Tumour Biol, 2006. **27**(6): p. 289-301.
308. Patil, U.S., et al., *Labeling primary amine groups in peptides and proteins with N-hydroxysuccinimidyl ester modified Fe<sub>3</sub>O<sub>4</sub>@SiO<sub>2</sub> nanoparticles containing cleavable disulfide-bond linkers*. Bioconjug Chem, 2013. **24**(9): p. 1562-9.
309. Ferreira, C.S., et al., *DNA aptamers against the MUC1 tumour marker: design of aptamer-antibody sandwich ELISA for the early diagnosis of epithelial tumours*. Anal Bioanal Chem, 2008. **390**(4): p. 1039-50.
310. Skoog, D.A., F.J. Holler, and T.A. Nieman, *In Principles of Instrumental Analysis*. 1998: Saunders College Publishing. 778-795.
311. Mendonsa, S.D. and M.T. Bowser, *In vitro evolution of functional DNA using capillary electrophoresis*. J Am Chem Soc, 2004. **126**(1): p. 20-1.

312. Lim, C.Y., A.E. Lim, and Y.C. Lam, *Ionic Origin of Electro-osmotic Flow Hysteresis*. Sci Rep, 2016. **6**: p. 22329.
313. Nowak, P.M., et al., *Improving repeatability of capillary electrophoresis—a critical comparison of ten different capillary inner surfaces and three criteria of peak identification*. Anal Bioanal Chem, 2017. **409**(18): p. 4383-4393.
314. Li, S.F.Y., *Capillary electrophoresis, principles, practice and applications*. Journal of Chromatography library. Vol. 52. 1992: Elsevier science publishers B.V.
315. Xuan, X., et al., *Electroosmotic flow with Joule heating effects*. Lab Chip, 2004. **4**(3): p. 230-6.
316. Ruckman, J., et al., *2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain*. J Biol Chem, 1998. **273**(32): p. 20556-67.
317. Boels, M.G.S., et al., *Systemic Monocyte Chemotactic Protein-1 Inhibition Modifies Renal Macrophages and Restores Glomerular Endothelial Glycocalyx and Barrier Function in Diabetic Nephropathy*. Am J Pathol, 2017. **187**(11): p. 2430-2440.
318. Menne, J., et al., *C-C motif-ligand 2 inhibition with emapticap pegol (NOX-E36) in type 2 diabetic patients with albuminuria*. Nephrol Dial Transplant, 2017. **32**(2): p. 307-315.
319. Ilgu, M. and M. Nilsen-Hamilton, *Aptamers in analytics*. Analyst, 2016. **141**(5): p. 1551-68.
320. Liu, Y., et al., *Aptamer-based electrochemical biosensor for interferon gamma detection*. Anal Chem, 2010. **82**(19): p. 8131-6.
321. Li, X., et al., *Electrochemical impedance spectroscopy for study of aptamer-thrombin interfacial interactions*. Biosens Bioelectron, 2008. **23**(11): p. 1624-30.
322. Holeman, L.A., et al., *Isolation and characterization of fluorophore-binding RNA aptamers*. Fold Des, 1998. **3**(6): p. 423-31.
323. Gomes de Castro, M.A., C. Hobartner, and F. Opazo, *Aptamers provide superior stainings of cellular receptors studied under super-resolution microscopy*. PLoS One, 2017. **12**(2): p. e0173050.
324. Lim, Z.Z., et al., *Gold nanoparticles in cancer therapy*. Acta Pharmacol Sin, 2011. **32**(8): p. 983-90.
325. Mirkin, C.A., et al., *A DNA-based method for rationally assembling nanoparticles into macroscopic materials*. Nature, 1996. **382**(6592): p. 607-9.
326. Cao, Y.C., R. Jin, and C.A. Mirkin, *Nanoparticles with Raman spectroscopic fingerprints for DNA and RNA detection*. Science, 2002. **297**(5586): p. 1536-40.
327. Xia, F., et al., *Colorimetric detection of DNA, small molecules, proteins, and ions using unmodified gold nanoparticles and conjugated polyelectrolytes*. Proc Natl Acad Sci U S A, 2010. **107**(24): p. 10837-41.
328. Wu, Y., et al., *Ultrasensitive aptamer biosensor for arsenic(III) detection in aqueous solution based on surfactant-induced aggregation of gold nanoparticles*. Analyst, 2012. **137**(18): p. 4171-8.
329. Lee, S., et al., *Cationic surfactant-based colorimetric detection of Plasmodium lactate dehydrogenase, a biomarker for malaria, using the specific DNA aptamer*. PLoS One, 2014. **9**(7): p. e100847.
330. Han, K., Z. Liang, and N. Zhou, *Design strategies for aptamer-based biosensors*. Sensors (Basel), 2010. **10**(5): p. 4541-57.
331. Seo, H.B. and M.B. Gu, *Aptamer-based sandwich-type biosensors*. J Biol Eng, 2017. **11**: p. 11.
332. Mendonsa, S.D. and M.T. Bowser, *In vitro selection of aptamers with affinity for neuropeptide Y using capillary electrophoresis*. J Am Chem Soc, 2005. **127**(26): p. 9382-3.
333. Mosing, R.K., S.D. Mendonsa, and M.T. Bowser, *Capillary electrophoresis-SELEX selection of aptamers with affinity for HIV-1 reverse transcriptase*. Anal Chem, 2005. **77**(19): p. 6107-12.
334. Dong, L., et al., *Screening and Identifying a Novel ssDNA Aptamer against Alpha-fetoprotein Using CE-SELEX*. Sci Rep, 2015. **5**: p. 15552.

335. Yang, J. and M.T. Bowser, *Capillary electrophoresis-SELEX selection of catalytic DNA aptamers for a small-molecule porphyrin target*. *Anal Chem*, 2013. **85**(3): p. 1525-30.
336. Hao, L., et al., *Affinity capillary electrophoresis with laser induced fluorescence detection for thrombin analysis using nuclease-resistant RNA aptamers*. *J Chromatogr A*, 2016. **1476**: p. 124-129.
337. Carrel, A. and C.C. Guthrie, *Functions of a Transplanted Kidney*. *Science*, 1905. **22**(563): p. 473.
338. Abel, J.J., L.G. Rowntree, and B.B. Turner, *On the removal of diffusible substances from the circulating blood by means of dialysis*. *Transactions of the Association of American Physicians, 1913*. *Transfus Sci*, 1990. **11**(2): p. 164-5.
339. Gold, L., et al., *Aptamer-based multiplexed proteomic technology for biomarker discovery*. *PLoS One*, 2010. **5**(12): p. e15004.