



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à l'École Supérieure de Physique et de Chimie  
Industrielles de la ville de Paris

**Caractérisation protéomique du matrisome des maladies  
des petits vaisseaux cérébraux par spectrométrie de  
masse**

Soutenue par

**Iman HADDAD**

Le 3 décembre 2020

Ecole doctorale n° 388

**Chimie physique et chimie  
analytique de Paris Centre**

Spécialité

**Chimie Analytique**

Composition du jury :

Anne, VARENNE Professeur, Chimie ParisTech	<i>Présidente</i>
Guillaume, VAN DER REST Professeur, Université Paris Sud	<i>Rapporteur</i>
Willy, BIENVENUT Chargé de Recherche, INRAE	<i>Rapporteur</i>
Hugues, CHABRIAT Professeur, Université Paris VII	<i>Examineur</i>
Laurent, MULLER Chargé de Recherche, Collège de France	<i>Invité</i>
Joëlle, VINH Directrice de Recherche, CNRS ESPCI	<i>Directrice de thèse</i>
Anne, JOUTEL Directrice de Recherche, INSERM	<i>Co-directrice</i>

Confidentielle jusqu'au 03/03/2022



A mes amis,  
A ma famille,

"I Can Only Show You The Door. You're The  
One That Has To Walk Through It."

Morpheus



# Remerciements (ou pas)

---

La réalisation de ce projet de thèse a été possible grâce au concours de nombreuses personnes à qui je souhaiterais témoigner toute ma reconnaissance :

Je tiens tout d'abord à remercier très chaleureusement ma directrice de thèse Joëlle Vinh, qui m'a donné la chance de pouvoir réaliser ce projet et qui m'a accordé sa confiance et son aide tout au long de cette thèse. Sans elle, je n'en serai pas là aujourd'hui, elle a été ma première chef d'équipe, ensuite directrice de laboratoire, et tout ce que je sais aujourd'hui est en grande partie grâce à elle.

Je remercie également le Dr Anne Joutel. Cette thèse est le fruit d'une collaboration de plus de dix années avec elle. C'est à ses côtés que j'ai compris ce que rigueur et précision voulaient dire et qui m'a fait partager ses brillantes intuitions.

Je souhaiterais remercier tout particulièrement Guillaume van der Rest et Willy Bienvenu pour avoir accepté de lire mon travail et d'être mes rapporteurs. Vos remarques m'ont été très précieuses pour améliorer mon manuscrit.

J'exprime ma gratitude à Anne Varenne et Hugues Chabriat qui m'ont fait l'honneur d'être membre de mon jury et d'avoir pris du temps pour lire mon travail.

Je tiens également à remercier Laurent Muller, grâce à qui les collagènes n'ont plus aucun secret pour moi (enfin surtout COL1A1 et COL1A2).

\*\*\*

Avant de remercier mes collègues, amis et famille, je souhaiterais, une fois n'est pas coutume, pour ce projet de thèse<sup>1</sup> ne pas remercier :

- Mon pc, qui malgré mes soins constants, a rendu non pas une mais deux fois l'âme pendant ce projet de thèse
- Les nombreuses pannes de LC ou de spectromètre de masse qui ont étayé mes campagnes d'analyses (un vendredi soir de préférence, je sais que je parle au nom de tous les doctorants)
- Le ou la COVID19 qui a rendu cette dernière année bien compliquée

\*\*\*

J'aimerai dire un grand merci à Valérie Domenga, Virginie Dubosclard, Julien Ratelade, Lamia Ghezali et Emilie Puard pour leur aide et leur bonne compagnie lorsque je travaillais sur le site Villemin, j'en ai d'excellents souvenirs.

Un énorme merci à mes collègues du SMBP qui m'ont supporté tout ce temps (de thèse, normalement je suis sympa). Merci Manu pour ton aide, merci de m'avoir fait sortir lors des alertes incendies et de m'avoir obligé à écrire mes protocoles et les stocker (maintenant je les trouve). Merci Gio pour ses longues discussions scientifiques, voir philosophiques par moment. Merci Yann pour ton humour, ta patience, ton calme et d'avoir été chercher à manger à chaque fois que je ne pouvais pas. Merci aux membres de mon bureau du soleil, Alex pour tes chants et ton eau chaude, Martha pour ta bonne humeur permanente, Jean pour tes adorables imitations de moi et Massamba pour m'avoir aidé pendant ces derniers mois et m'avoir encouragé. Merci au bureau de doctorants de l'autre bout du couloir, Wafa, grâce à toi, je suis toujours au fait des dernières tendances sur le net, Ranjith, thanks for your help in english and your delicious meal, Zeyuan merci pour nos discussions du matin et tous ces gâteaux que tu ramènes. Je n'oublie pas les anciens, Shakir et Nicolas en particulier qui ont suivi pas à pas mes péripéties et qui m'ont encouragé à continuer. Et évidemment, je remercie à nouveau Joëlle pour sa patience et sa capacité à résoudre, ou en tout cas à relativiser dans les pires situations.

Un grand merci également à Jean-Jacques, Emmanuel et Frédéric pour leur aide et leur patience, notamment lors de mes crashes de pc.

Merci à tous mes amis qui m'ont apporté leur soutien moral pendant ces années, qui ont accepté de m'écouter expliquer mon sujet sans comprendre et sans se plaindre, particulièrement toi Isa et toi Carole.

Enfin, un très grand MERCI à toute ma famille qui m'a gratifié de son amour, de sa patience, de son soutien pendant 6 ans et a contribué à l'aboutissement de ce travail.

# Résumé

---

Les maladies des petits vaisseaux cérébraux sont responsables de lésions de la substance blanche du cerveau et d'infarctus cérébraux profonds multiples. Elles sont à l'origine de plus de 25% des accidents vasculaires cérébraux et constituent la deuxième cause de démence après la démence d'Alzheimer. C'est un ensemble de processus pathologiques, qui affectent les petites artères, les artérioles, veinules ou capillaires cérébraux de moins de 400µm. Le matrisome cérébrovasculaire semble être une voie pathologique convergente entre les différentes maladies des petits vaisseaux de type génétique et de type sporadique. Le matrisome est l'ensemble des protéines constituant la matrice extracellulaire (MEC) ainsi que les protéines associées, leurs rôles consistent non seulement dans le soutien et l'ancrage des cellules mais également dans divers processus fondamentaux tels que la différenciation, la prolifération, la survie ou la migration des cellules. La diversité structurale et physico-chimique de ces protéines rend cependant leur analyse particulièrement délicate. En effet, certaines protéines du *core* matrisome, protéines de haut poids moléculaire, sont particulièrement insolubles et d'autres protéines du matrisome associé sont généralement plus petites et moins abondantes.

Dans le cadre de cette thèse nous nous proposons de caractériser de manière quantitative et qualitative le matrisome microvasculaire dans le cadre des maladies des petits de vaisseaux, ainsi que d'identifier des anomalies communes ou spécifiques à chaque maladie. Pour cela nous avons, dans un premier temps, développé une approche protéomique quantitative en spectrométrie de masse de type *label-free* sur des vaisseaux cérébraux et périphériques isolés, et nous l'avons validée dans une première étude sur un modèle murin préclinique de CADASIL.

Ensuite nous avons appliqué cette approche sur deux autres modèles murins pour les maladies génétiques CARASIL et la maladie du collagène de type IV, et deux modèles murins pour l'hypertension et l'âge, modèle pour leur caractère sporadique.

Nous avons développé et validé une nouvelle méthode robuste et sensible pour l'analyse quantitative sans marquage des changements du matrisome des artères cérébrales de souris. L'application de notre méthode sur les artères des différents modèles murins étudiés nous a permis d'identifier quelques pistes intéressantes pour chaque maladie de manière indépendante mais a également mis en évidence quelques signatures communes entre les différentes études. En effet, nous avons mis en évidence une protéine commune à toutes ces formes de cSVDs, PRSS23, une serine protéase dans les 5 modèles étudiée et HTRA1, une autre serine protéase, dans CADASIL, la maladie du collagène de type IV et l'âge.

# Table des matières

<b>REMERCIEMENTS (OU PAS)</b> .....	<b>5</b>
<b>RESUME</b> .....	<b>7</b>
<b>LISTE DES TABLEAUX</b> .....	<b>10</b>
<b>LISTE DES FIGURES</b> .....	<b>11</b>
<b>LISTE DES ABREVIATIONS</b> .....	<b>14</b>
<b>INTRODUCTION GENERALE</b> .....	<b>17</b>
<b>I. I. CONTEXTE SCIENTIFIQUE ET APPROCHES TECHNIQUES</b> .....	<b>21</b>
I.1 LES MALADIES DES PETITS VAISSEAUX.....	23
I.1.1 <i>L'unité neurovasculaire</i> .....	23
I.1.2 <i>Les maladies des petits vaisseaux cérébraux et leur diagnostic</i> .....	26
I.1.1 <i>Différents types de maladie des petits vaisseaux cérébraux</i> .....	26
I.2 LE MATRISOME .....	36
I.2.1 <i>Un peu d'histoire</i> .....	36
I.2.2 <i>Les différentes matrices extracellulaires</i> .....	38
I.2.3 <i>Définition</i> .....	39
I.2.4 <i>Maladie des petites vaisseaux cérébraux et matrisome</i> .....	56
I.2.5 <i>Protéomique de la matrice extracellulaire</i> .....	58
I.3 PROTEOMIQUE, SPECTROMETRIE DE MASSE (MS) ET QUANTIFICATION .....	64
I.3.1 <i>Pourquoi analyser le Protéome ?</i> .....	65
I.3.2 <i>Traitement de l'échantillon de type tissulaire, de l'extraction à la digestion</i> .....	65
I.3.3 <i>La spectrométrie de masse et l'analyse protéomique</i> .....	70
I.3.4 <i>La quantification relative en protéomique par spectrométrie de masse</i> .....	82
<b>II. OBJECTIFS</b> .....	<b>89</b>
<b>III. RESULTATS ET DISCUSSIONS</b> .....	<b>91</b>
III.1 DEVELOPPEMENT D'UNE NOUVELLE METHODE POUR L'ETUDE DU MATRISOME DES VAISSEAUX CEREBRAUX CHEZ LA SOURIS (ARTICLE2).....	92
III.1.1 <i>Introduction</i> .....	92
III.1.2 <i>Objectif</i> .....	92
III.1.3 <i>Méthode</i> .....	93
III.1.4 <i>Résultats et Discussions</i> .....	95
III.2 ÉTUDE DES CHANGEMENTS DU MATRISOME MICROVASCULAIRE DANS LES MALADIES DES PETITS VAISSEAUX.....	131
III.2.1 <i>Introduction</i> .....	131
III.2.2 <i>Objectifs</i> .....	132
III.2.3 <i>Matériel et méthode</i> .....	132
III.2.4 <i>Résultats</i> .....	134

<i>III.2.5 Discussion</i> .....	148
<b>SYNTHESE ET CONCLUSION GENERALE</b> .....	<b>151</b>
<b>ANNEXES</b> .....	<b>155</b>
<b>I. ANNEXE I : ARTICLE 1</b> .....	<b>157</b>
<b>II. ANNEXE 2 : INFORMATIONS COMPLEMENTAIRES</b> .....	<b>175</b>
II.1 VALIDATION DES ANALYSES ET IDENTIFICATION DES PROTEINES DANS LES PETITS VAISSEAUX CEREBRAUX DES SOURIS CADASIL.....	175
II.2 VALIDATION DES ANALYSES ET IDENTIFICATION DES PROTEINES DANS LES PETITS VAISSEAUX CEREBRAUX DES SOURIS CARASIL .....	179
II.3 VALIDATION DES ANALYSES ET IDENTIFICATION DES PROTEINES DANS LES PETITS VAISSEAUX CEREBRAUX DES SOURIS MODELE POUR LA MALADIE DU COLLAGENE DE TYPE IV.....	183
II.4 VALIDATION DES ANALYSES ET IDENTIFICATION DES PROTEINES DANS LES PETITS VAISSEAUX CEREBRAUX DES SOURIS MODELE POUR L'EFFET DE L'AGE .....	187
II.5 VALIDATION DES ANALYSES ET IDENTIFICATION DES PROTEINES DANS LES PETITS VAISSEAUX CEREBRAUX DES SOURIS MODELE POUR L'EFFET DE L'HYPERTENSION .....	191
<b>III. ANNEXE 3 : PROTOCOLES EXPERIMENTAUX</b> .....	<b>195</b>
III.1 PREPARATION DES ARTERES PIALES POUR LA LC MS/MS.....	195
III.2 ANALYSES LC MS/MS .....	195
III.3 TRAITEMENT ET VALIDATION DES DONNEES .....	196
III.4 QUANTIFICATION, VALIDATION ET STATISTIQUES .....	196
<b>IV. ANNEXE 4 : SCRIPT R POUR LE TRAITEMENT DES DONNEES</b> .....	<b>199</b>
<b>BIBLIOGRAPHIE</b> .....	<b>211</b>

# Liste des tableaux

---

Table 1 : Classifications des SVD en fonction des modifications pathologique cérébrovasculaires (inspiré de Pantoni,2010 <sup>5</sup> ).....	27
Table 2 : Les différents tissus et matrices extracellulaires( inspiré de Dupont, 2018 <sup>59</sup> ).....	38
Table 3 : Tableau des différents types de collagène avec leur structure, leur distribution dans les tissus et quelques caractéristiques (Adapté de Ricard-Blum, 2011 <sup>68</sup> ).....	42
Table 4 : Quelques groupes de glycoprotéines de la MEC, leurs fonctions connues et les articles qui en réfèrent.....	49
Table 5 : Quelques famille de protéines régulatrice de la MEC, leur(s) fonction(s), leur(s) substrat(s) et les articles qui en réfèrent.....	51
Table 6 : Tableau récapitulatif et quantitatif pour les 3 maladies génétiques.....	139
Table 7 : Tableau récapitulatif et quantitatif pour les 2 maladies sporadiques.....	143
Table 8 : Table récapitulative des analyses MSMS pour l'étude CADASIL.....	175
Table 9 : Table récapitulative des analyses MSMS pour l'étude CARASIL.....	179
Table 10 : Table récapitulative des analyses MSMS pour l'étude de la maladie du collagène de type IV.....	183
Table 11 : Table récapitulative des analyses MSMS pour l'étude de la maladie de l'effet de l'âge ...	187
Table 12 : Table récapitulative des analyses MSMS pour l'étude de la maladie de l'effet de l'hypertension.....	191

# Liste des figures

---

Figure 1 : Représentation de l'arbre cérébrovasculaire de l'artère carotide interne (ACI) aux artérioles. ....	24
Figure 2 : Représentation de l'unité vasculaire (Adapté de Iadecola, 2004 <sup>2</sup> ) .....	24
Figure 3 : La matrice extracellulaire de Bordeu à aujourd'hui .....	37
Figure 4 : Distribution des différentes catégories de protéines du matrisome <i>in silico</i> .....	40
Figure 5 : Description et formation de la structure en réseau du collagène de type IV (adapté de Kalluri 2003 <sup>94</sup> ) .....	53
Figure 6 : Description de l'interaction entre les différentes protéines majoritaires de la MB (adapté de Mak 2017 <sup>93</sup> ) .....	55
Figure 7 : A : Schéma illustrant la transformation progressive des artères cérébrales en capillaires (d'après Girouard, 2006) <sup>101</sup> . B : Coupe transversale des différents types de vaisseaux ainsi que leur composition (d'après Haffner, 2016) <sup>102</sup> . ....	56
Figure 8 : Stratégie du <i>bottom-up</i> en protéomique .....	64
Figure 9 : Géométrie de l'Orbitrap où R1 = rayon de l'électrode centrale et R2 = rayon de l'électrode externe (Makarov, 2000 <sup>175</sup> ) .....	71
Figure 10 : Détails de l'appareil QExactive (adapté de Michalski, 2011 <sup>179</sup> ) .....	73
Figure 11 : Schéma de fragmentation original de Roepstorff et Fohlmann 1984 <sup>187</sup> .....	77
Figure 12 : De la spectrométrie de masse à l'identification protéique .....	79
Figure 13 : Distribution des scores des identifications correctes et incorrectes (Käll, 2008 <sup>199</sup> ) .....	81
Figure 14 : Les différents flux de travail selon la stratégie de quantification utilisée, de l'échantillon à l'interprétation (adapté de Bantscheff 2012 <sup>200</sup> ) .....	82
Figure 15 : <i>Workflow</i> des données de spectrométrie de masse de la normalisation, du traitement des données manquantes à l'imputation. ....	87
Figure 16 : Distribution des intensités des peptides de Notch3 en fonction de leur position sur la séquence .....	135
Figure 17 : Mesure des intensités normalisées pour tous les échantillons pour les protéines contrôles des protéines NOTCH3, TIMP3, VTN et HTRA1 dans les artères cérébrales des souris TgNotch3R169C et TgNotch3WT. ....	136
Figure 18 : Répartition des intensités peptidiques sur la séquence de la protéine HTRA1 .....	137
Figure 19 : Mesures des intensités protéiques normalisées des protéines COL4A1 et COL4A2 dans l'étude de la maladie du collagène IV dans les artères cérébrales des souris Col4a1G1064D/+ et Col4a1+/+ .....	138
Figure 20 : représentation graphique de type Volcano plot pour l'étude CADASIL des rapports d'intensités des protéines $\log_2(\text{ratio})$ entre les deux conditions avec leur significativité $-\log_{10}(q\text{value})$ . ....	140
Figure 21 : représentation graphique de type Volcano plot pour l'étude CARASIL, des rapports d'intensités des protéines $\log_2(\text{ratio})$ entre les deux conditions avec leur significativité $-\log_{10}(q\text{value})$ . ....	140
Figure 22 : représentation graphique de type Volcano plot pour l'étude de la maladie du collagène de type IV, des rapports d'intensités des protéines $\log_2(\text{ratio})$ entre les deux conditions avec leur significativité $-\log_{10}(q\text{value})$ . ....	141

Figure 23 : Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les 3 maladies des cSVD génétiques .....	143
Figure 24 : Heatmap des analyses protéomiques pour l'étude de l'effet de l'âge .....	145
Figure 25 : Diagramme de Venn du recouvrement protéique entre l'effet de l'âge et l'hypertension .....	146
Figure 26 : Diagramme de Venn Upset <sup>258</sup> pour le recouvrement protéique des différentes maladies étudiées .....	147
Figure 27: Nombre de protéines identifiées par MS/MS et par transfert d'identification pour l'étude de la maladie CADASIL.....	175
Figure 28 : Multiscatterplot et distribution de log2 (intensités) des souris TgNotch3 <sup>wt</sup> et TgNotch3 <sup>R169c</sup> .....	176
Figure 29 : Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris TgNotch3 <sup>WT</sup> .....	177
Figure 30 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris TgNotch3 <sup>R169c</sup> .....	177
Figure 31 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris TgNotch3 <sup>wt</sup> .....	178
Figure 32 : Nombre de protéines identifiées par MS/MS et par transfert d'identification pour l'étude de la maladie CARASIL .....	179
Figure 33 : Multiscatterplot et distribution de log2 (intensités) des souris CAR et WT.....	180
Figure 34 : Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris CARASIL.....	180
Figure 35: Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris WT .....	181
Figure 36 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris CARASIL.....	182
Figure 37 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris WT .....	182
Figure 38 : Nombre de protéines identifiées par MS/MS et par transfert d'identification pour l'étude de la maladie collagène IV.....	183
Figure 39 : Multiscatterplot et distribution de log2 (intensités) des souris COL4A1 et WT .....	184
Figure 40 : Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris COL4.....	184
Figure 41 : Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris WT.....	185
Figure 42 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris COL4.....	186
Figure 43 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris WT .....	186
Figure 44 : Nombre de protéines identifiées par MS/MS et par transfert d'identification pour l'étude de l'effet de l'âge.....	187
Figure 45 : Multiscatterplot et distribution de log2 (intensités) des souris de 2 et 20 mois .....	188
Figure 46 : Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris de 20 mois.....	189

Figure 47 : Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris de 2 mois.....	189
Figure 48 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris de 20 mois.....	190
Figure 49 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris de 2 mois.....	190
Figure 50: Nombre de protéines identifiées par MS/MS et par transfert d'identification pour l'étude de l'effet de l'hypertension .....	192
Figure 51 : Multiscatterplot et distribution de log2 (intensités) des souris CTL, BPH et BPH traitées	192
Figure 52 : Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris hypertendues BPH .....	193
Figure 53 : Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris normaux tendues CTL .....	193
Figure 54 : Diagramme de Venn de type Upset <sup>258</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris hypertendues traitées BPHT.....	193
Figure 55 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris CTL.....	194
Figure 56 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris BPHT .....	194
Figure 57 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris BPH.....	194

# Liste des abréviations

---

ACI : Artère Carotide Interne

ACP : Analyse en composantes principales

ADAM: *A Disintegrin And Metalloprotease*

AVC : accidents vasculaires cérébraux (AVC)

BBB: *Blood Brain Barrier*

CADASIL: *Cerebral Autosomal Dominant Arteriopathy With Subcortical Ischaemic strokes and Leukoencephalopathy*

CARASAL: *Cathepsin-A-Related arteriopathy with Strokes And Leukoencephalopathy*

CARASIL: *Cerebral Autosomal Recessive Arteriopathy With Subcortical Ischaemic strokes and Leukoencephalopathy*

CID : *Collision-Induced Dissociation*

CML : Cellule Musculaire Lisse

cSVD: *cerebral Small Vessel Disease*

C-trap: *Curved rf-only quadrupole*

DDA: *Data Dependant Acquisition*

DDM: n-dodecyl  $\beta$ -D-maltoside

DTT: Dithiothréitol

EC : *Enzyme Commission*

EGFR: *Epidermal Growth Factor Repeat*

ELISA: *Enzyme Linked ImmunoSorbent Assay*

ESI: *ElectroSpray Ionization*

FACIT: *Fibril Associated Collagen with Interrupted Triple helixes*

FDR: *False Discovery Rate*

FT-ICR: *Fourier-Transform Ion Cyclotron Resonance*

GAG: GlycosAminoGlycane

GO: *Gene Ontology*

GOM: *Granular Osmiophilic Materials*

GPI: glycosylphosphatidylinositol

HCD: *High Collision Dissociation*

HF: *High Field*

HPLC: *High Performance Liquid Chromatography*

HSPG: *Heparane Sulfate ProteoGlycan*

HUPO: *HUman Proteomic Organization*

ICAT : *Isotope-Coded Affinity Tags*

IRM : *Imagerie par Résonance Magnétique*

IT: *Ion Trap*

iTRAQ: *isobaric Tags for Relative and Absolute Quantitation*

IUBMB: *The International Union of Biochemistry and Molecular Biology*

LOX: *Lysyl OXydase*

MACIT: *Membrane-Associated Collagens with Interrupted Triple-helices*

MALDI: *Matrix Associated Laser Dissociation Ionization*

MAR: *Missing At Random*

MB: *Membrane Basale*

MCAR: *Missing Completely At Random*

MEC : *Matrice ExtraCellulaire*

MMP : *Metalloprotease Matricielle Protein*

MNAR: *Missing Not At Random*

MS: *Mass Spectrometry*

MS/MS : *Spectrométrie de masse tandem*

m/z : *rapport masse sur charge*

NaN : *Not a Number*

NOTCH3: *Neurogenic locus notch homolog protein 3*

NOTCH3<sup>ECD</sup>: *Notch3 ExtraCellular Domain*

PEP: *Posterior Error Probability*

PG: ProtéoGlycanes

PRM: *Parallel Reaction Monitoring*

PSM : *Peptide-Spectrum Match*

Q: Quadripôle

RE : Réticulum endoplasmique

SC : *Spectral Count*

SDC : Sodium DéoxyCholate

SDS : Dodécyl Sulfate de Sodium

S-lens: *Stacked ring ion guide*

SLRP: *Small Leucine-Rich Proteoglycans*

SPARC: *Secreted Protein Acidic and Cysteine Rich*

STRIVE: *STandards for ReportIng Vascular changes on nEuroimaging*

TCEP: Tri-Carboxy-Ethyl-Phosphine

TIMP: *Tissue Inhibitor of Metalloproteinases*

TMIC : *TransMembrane IntraCellular*

TMT: *Tandem Mass Tags*

TOF: *Time Of Flight*

XIC : *eXtracted Ion Chromatogram*

# Introduction générale

---

Les tissus ne sont pas exclusivement constitués de cellules, l'espace extracellulaire contribue également à une partie, parfois très importante, de leur volume. Cet espace est ce qu'on appelle la matrice extracellulaire (MEC). Le matrisome est l'ensemble des protéines de la matrice extracellulaire (MEC) ainsi que les protéines associées à cette MEC.

Les protéines de la MEC se décomposent en deux catégories : le cœur du matrisome et les protéines associées au matrisome. Les premières sont des protéines de type structural, qui vont conférer à la MEC sa structure, sa résistance ou encore son élasticité. Elles contiennent majoritairement les collagènes, les protéoglycanes et glycoprotéines. Les protéines associées vont regrouper plutôt des protéines régulatrices telles que des enzymes, des protéines sécrétées tels que des facteurs de croissance, et les protéines qui ont des similarités structurales avec d'autres protéines de la matrice, ou qui sont connues pour être associées à une protéine de la matrice.

La diversité structurale et physico-chimique de ces protéines rend cependant leur analyse particulièrement délicate. En effet, le cœur du matrisome, représenté par des protéines de haut poids moléculaire, est particulièrement insoluble tandis que les protéines associées au matrisome sont généralement, mais pas obligatoirement, plus petites et moins abondantes.

Dans le cadre de diverses maladies, des altérations de l'organisation ou de la composition de la MEC ont été observées. C'est pourquoi une meilleure caractérisation de la composition, du métabolisme et de la biologie de la MEC peut conduire à l'identification de nouveaux marqueurs pronostiques et diagnostics.

Dans notre étude, nous nous sommes intéressés plus particulièrement aux maladies des petits vaisseaux cérébraux. Les maladies des petits vaisseaux cérébraux sont responsables de lésions de la substance blanche du cerveau et d'infarctus cérébraux profonds multiples. Elles sont à l'origine de plus de 25% des accidents vasculaires cérébraux et constituent la deuxième cause de démence après la démence d'Alzheimer. C'est un ensemble de processus pathologiques, qui affectent les petites artères, les artérioles, veinules ou capillaires cérébraux de moins de 400 $\mu$ m.

Les modifications du matrisome cérébrovasculaire semblent être une voie pathologique convergente entre les différentes maladies des petits vaisseaux de type génétique, et très probablement aussi pour les maladies de type sporadique, les symptômes cliniques étant très similaires. En effet pour une multitude de maladies génétiques, la caractéristique commune est la présence de remaniements importants de la MEC des microvaisseaux cérébraux.

Mon projet doctoral à terme vise à mieux comprendre les mécanismes pathogéniques des maladies des petits vaisseaux.

Dans le cadre de cette thèse nous nous proposons de (i) caractériser de manière quantitative et (ii) qualitative le matrisome microvasculaire dans les maladies des petits de vaisseaux, ainsi que (iii) d'identifier des anomalies communes ou spécifique à chaque maladie. Pour atteindre cet objectif, a été développée une approche protéomique quantitative sans marquage (*label free*) sur les vaisseaux cérébraux de plusieurs modèles murins précliniques des différentes maladies étudiées.

Dans la première partie, sont introduites les maladies des petits vaisseaux cérébraux de type non amyloïdes (I.1) ainsi que la notion de matrisome (I.2), ses différentes divisions et les méthodes existantes aujourd'hui pour son étude. Nous abordons également la spectrométrie de masse en général (I.3) et plus particulièrement les techniques utilisées pour ce projet de thèse.

Ensuite sont présentés les divers résultats obtenus durant cette étude (III) : tout d'abord l'optimisation du traitement de l'échantillon et des données qui sont présentés notamment sous forme d'article (en soumission) (III.1), ensuite les résultats de l'analyse protéomique différentielle des différentes maladies des petits vaisseaux cérébraux étudiées (III.2).







# I. I. Contexte scientifique et approches techniques

---

I.1	LES MALADIES DES PETITS VAISSEAUX.....	23
I.1.1	<i>L'unité neurovasculaire.....</i>	23
I.1.1.1	L'anatomie de l'arbre cérébrovasculaire .....	23
I.1.1.2	Les différents acteurs de l'unité neurovasculaire et leur fonction .....	25
I.1.2	<i>Les maladies des petits vaisseaux cérébraux et leur diagnostic.....</i>	26
I.1.1	<i>Différents types de maladie des petits vaisseaux cérébraux.....</i>	26
I.1.1.1	Les maladies des petits vaisseaux cérébraux de type sporadique.....	28
I.1.1.2	Maladies héréditaires ou génétiques de petits vaisseaux cérébraux .....	28
I.1.1.2.1	CADASIL (cerebral autosomal dominant arteriopathy with subcortical ischaemic strokes and leukoencephalopathy) .....	29
I.1.1.2.2	CARASIL (cerebral autosomal recessive arteriopathy with subcortical ischaemic strokes and leukoencephalopathy) .....	30
I.1.1.2.3	Maladie du Collagène de type IV .....	31



## **I.1 Les maladies des petits vaisseaux**

Ce premier chapitre sera consacré à une brève introduction sur les maladies des petits vaisseaux cérébraux de type non amyloïde, en commençant par décrire les principaux organes atteints et leur fonctionnement, ensuite sera présenté l'état de l'art des différentes pathologies étudiées durant cette thèse.

### **I.1.1 L'unité neurovasculaire**

Le cerveau, l'organe le plus complexe et le plus sophistiqué du corps humain, présente cependant une faiblesse non négligeable, son alimentation. Il consomme une grande quantité d'énergie mais ne possède pas de réservoir pour stocker son carburant. Il reçoit ses substrats énergétiques en fonction de ces besoins dynamiques et régionaux imposés par son activité. Le flux sanguin doit arriver au bon endroit, au bon moment et tout arrêt de plus de quelques minutes peut être irrémédiable voir fatal. L'intérêt porté à la manière dont le cerveau régule son propre apport énergétique a fait l'objet de beaucoup d'études depuis le début de la neuroscience, et est motivé non seulement par le désir de mieux comprendre les effets néfastes d'une vascularisation cérébrale déficiente mais également des changements régionaux du débit sanguin cérébral, le sujet d'étude de l'imagerie fonctionnelle. Le concept d'unité neurovasculaire met en évidence ce lien existant entre les cellules du cerveau et le système vasculaire cérébral, et maintenir la santé neurovasculaire va de pair avec le maintien de la santé cérébrale. Voici une description brève de l'anatomie de l'unité neurovasculaire pour mieux comprendre ses mécanismes <sup>2</sup>.

#### ***I.1.1.1 L'anatomie de l'arbre cérébrovasculaire***

Le cerveau est irrigué par 4 artères extra crâniennes : 2 artères carotides internes qui forment la circulation antérieure et 2 artères vertébrales qui fusionnent pour former le tronc basilaire et la circulation postérieure. Les circulations antérieures et postérieures irriguent des régions différentes du cerveau, elles sont néanmoins reliées à la base du crâne par un système d'anastomose artérielle, appelé le Polygone de Willis, qui relie une partie des branches terminales des artères carotides internes et du tronc basilaire. Du polygone de Willis partent des artères, qui se ramifient en piales et en artérioles, qui courent à la surface du cerveau formant un réseau fortement interconnecté ; il s'agit des artères piales appelées aussi leptoméningées. De ces artères naissent des branches qui pénètrent dans le cerveau (artères pénétrantes ou perforantes, voir figure 1). Dans leur partie initiale, elles sont entourées par l'espace périvasculaire, un espace virtuel délimité par la membrane basale vasculaire et la glia limitans. La paroi des artères et artérioles est constituée de multiples couches de cellules musculaires lisses (CML), dont le nombre varie en fonction du diamètre du vaisseau mais aussi de l'espèce, bien que les artères piales ne soient pas en contact direct avec les cellules neuronales, elles possèdent de nombreuses innervations.

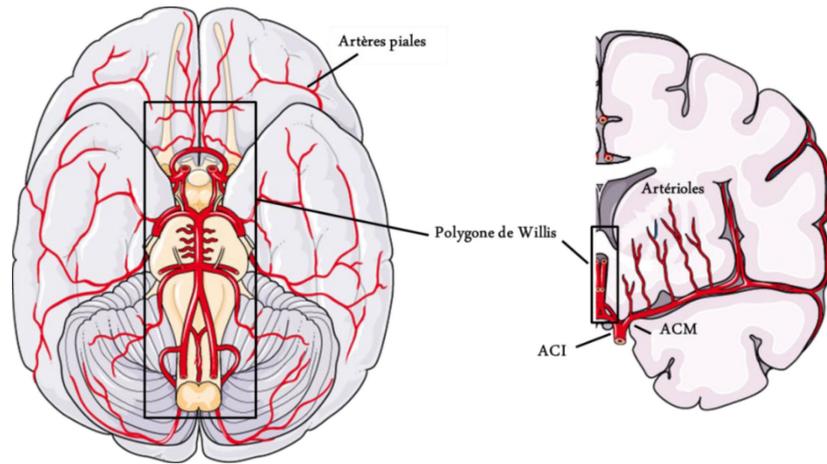


Figure 1 : Représentation de l'arbre cérébrovasculaire de l'artère carotide interne (ACI) aux artéριοles.

A mesure que les artéριοles pénètrent plus profondément dans le cerveau, la membrane gliale et la membrane basale fusionnent effaçant ainsi l'espace périvasculaire, appelé également espace de Virchow-Robin, et les vaisseaux sont enfermés dans les pieds d'extrémité astrocytaire (artéριοles intra parenchymateuses). Dotées d'une seule couche de CML, les artéριοles intra parenchymateuses manquent de nerfs périvasculaires. Les capillaires sont dépourvus de CML mais sont dotés de péricytes, qui sont entièrement enfermés par la membrane basale de l'endothélium (Figure2).

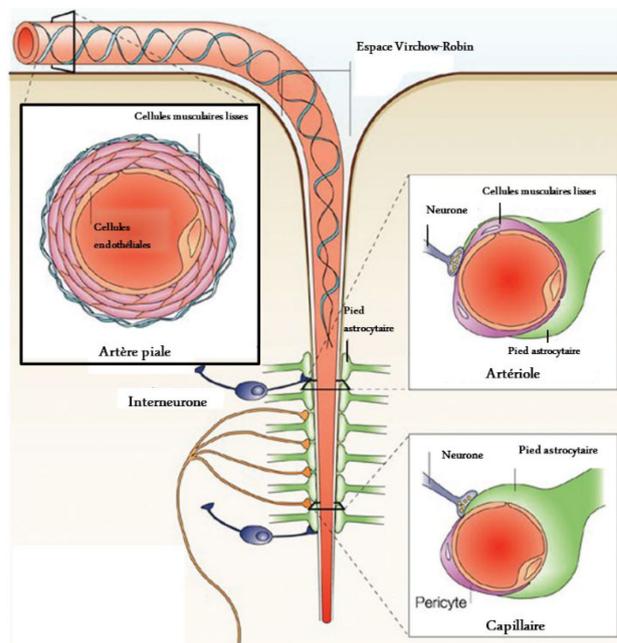


Figure 2 : Représentation de l'unité vasculaire (Adapté de Iadecola, 2004<sup>3</sup>)

### *1.1.1.2 Les différents acteurs de l'unité neurovasculaire et leur fonction*

Neurones, astrocytes, cellules endothéliales, cellules musculaires lisses et macrophages périvasculaires sont les composants cellulaires de l'unité neurovasculaire. Ces cellules participent à la transmission de signaux moléculaires dont l'une des finalités est l'adaptation du débit sanguin cérébral en fonction de l'activité neuronale ou des variations de pression artérielle.

Les neurones et les astrocytes vont principalement générer des signaux qui vont de façon directe ou indirecte par l'intermédiaire d'autres cellules, agir sur les vaisseaux pour initier la réponse vasculaire. En effet, l'augmentation de divers substrats va entraîner la vasodilatation du vaisseau, ce qui a pour conséquence une augmentation du débit sanguin.

Les cellules endothéliales, dans lesquelles se trouve des quantités non négligeables de nombreux agents vasodilatateurs (ou vasoconstricteurs) tels que les prostanoïdes ou l'endothéline, sont suspectées d'avoir un rôle dans la propagation rétrograde des signaux neurovasculaires. Elles sont « activées » en réponse à des composés chimiques ou ioniques (comme le potassium  $K^+$ ), déclenchant la libération de substances vasodilatatrices comme le monoxyde d'azote ou la propagation d'une hyperpolarisation.

Les cellules musculaires lisses sont les principales cellules responsables des réponses vasomotrices et de la régulation du débit. Elles se contractent et se relâchent en réponse à une grande variété d'agent vasoactifs et de stimuli. Elles sont également capables de répondre aux variations de pression artérielle, propriété représentée par le tonus myogénique. Cette propriété joue un rôle très important la stabilité du débit sanguin cérébral, lors de modifications de la pression artérielle, du moins jusqu'à un certain seuil.

Au niveau des capillaires, les cellules murales sont appelées péricytes, cellules connues pour leur rôle important dans le maintien de la barrière hématoencéphalique. Cependant il reste des controverses sur la définition exacte des péricytes et sur leur implication dans la régulation du débit sanguin.

Si l'unité neurovasculaire a un rôle essentiel dans la perfusion sanguine cérébrale et dans l'intégrité de la barrière hématoencéphalique, elle joue également un rôle dans l'élimination des toxines via son compartiment périvasculaire via le système lymphatique<sup>4</sup>. Toute altération de l'unité neurovasculaire peut entraîner de graves conséquences au niveau cérébral or, le cerveau est le régulateur central de tous les organes du corps.

## **I.1.2 Les maladies des petits vaisseaux cérébraux et leur diagnostic**

Les maladies des petits vaisseaux cérébraux, autrement appelées *Cerebral Small Vessel Diseases* en anglais (cSVD), désignent un ensemble de processus pathologiques qui touchent les petites artères, les artérioles, les veinules ou les capillaires du cerveau. Les manifestations cliniques de ces maladies sont principalement des accidents vasculaires cérébraux ischémiques ou hémorragiques, des troubles cognitifs pouvant aller jusqu'à la démence, des manifestations psychiatriques et des troubles de la marche. Les maladies des petits vaisseaux cérébraux sont responsables de lésions diffuses de la substance blanche du cerveau, d'infarctus cérébraux de petite taille profonds multiples, de microsaignements, de dilatation des espaces périvasculaires et d'une atrophie cérébrale. Contrairement aux gros vaisseaux, les petits vaisseaux cérébraux sont très difficilement visualisés actuellement du vivant du patient. Par conséquent, les lésions du parenchyme, visibles essentiellement par imagerie par résonance magnétique (IRM), supposées être causées par ces modifications dans les vaisseaux, ont été adoptées comme marqueur de ces affections<sup>5</sup>.

### **I.1.1 Différents types de maladie des petits vaisseaux cérébraux**

S'il existe plusieurs types de maladies des petits vaisseaux cérébraux, aiguës ou chroniques, inflammatoires ou non, amyloïdes ou non, ne seront abordés dans cette étude que les maladies des petits vaisseaux chroniques non amyloïdes (cSVD) (Table1). Celles-ci sont le plus souvent sporadiques et pourraient être provoquées par un mélange relativement complexe de facteurs génétiques et environnementaux, notamment l'âge et l'hypertension, qui sont les deux principaux facteurs de risque. Mais il a été découvert récemment des formes génétiques mendéliennes, dont les manifestations cliniques et radiologiques sont très semblables aux formes sporadiques<sup>6</sup>. Les études en IRM de populations ont montré que les cSVD évoluent à bas bruit sans manifestation clinique pendant de nombreuses années suggérant qu'il existe une fenêtre thérapeutique<sup>7</sup>. Aujourd'hui, il n'y a aucune mesure curative connue pour traiter les personnes qui en sont atteintes. Néanmoins, il existe un certain nombre de mesures préventives, tels que la normalisation de la pression artérielle, la prise d'anti-agrégants plaquettaires en cas d'infarctus et l'arrêt du tabac. La terminologie et les définitions des lésions IRM dans les cSVD varient considérablement d'une étude à l'autre, rendant difficile les comparaisons entre les différentes études. Dans ce contexte, un groupe de travail international réunissant des experts des maladies des petits vaisseaux a, en 2013, établi une nouvelle terminologie pour désigner ces lésions et publié un certain nombre de recommandations pour l'acquisition et l'analyse des images IRM dans les études cliniques (STandards for ReportIng Vascular changes on nEuroimaging, STRIVE)<sup>5</sup>.

Table 1 : Classifications des SVD en fonction des modifications pathologique cérébrovasculaires (inspiré de Pantoni,2010 <sup>6</sup>).

Type	Classification	Exemples, principales caractéristiques
1	Artériosclérose ou SVD liées à l'âge et aux facteurs de risques vasculaires	Nécrose fibrinoïde : Dépôt de fibrine dans les vaisseaux ; Hyalinose : dépôt de matériel acellulaire amorphe dans la paroi des vaisseaux pouvant être associé à de la fibrose (fibrohyalinose) ou la présence de macrophages chargés de cholestérol (lipohyalinose) Microathérome : accumulation de diverses molécules (calcium, cholestérol) dans les vaisseaux. Microanévrisme, désorganisation artérielle segmentaire
2	Angiopathie amyloïde cérébrale sporadique et héréditaire	Accumulation progressive de protéine $\beta$ amyloïde dans les parois d'artères de petite à moyenne taille et d'artéoles situées principalement dans l'espace leptoméningé, le cortex.
3	Maladies héréditaires ou génétiques de petits vaisseaux distinctes de l'angiopathie cérébrale amyloïde	Gènes mutés responsables et leur maladie : Notch3 -> CADASIL (cerebral autosomal dominant arteriopathy with subcortical ischaemic strokes and leukoencephalopathy) HTRA1 -> CARASIL (cerebral autosomal recessive arteriopathy with subcortical ischaemic strokes and leukoencephalopathy) COL4A1 et COL4A2
4	Maladies inflammatoires et à médiation immunologique	Groupe hétérogène de maladies rares caractérisées par la présence de cellules inflammatoires dans la paroi des vaisseaux (vascularite) et font généralement partie d'une maladie systémique
5	Collagénose veineuse	Dépôts de collagène dans la paroi des veines et des veinules près des ventricules latéraux pouvant entraîner un rétrécissement de la lumière et, parfois, une occlusion.
6	Autres SVD	

### ***1.1.1.1 Les maladies des petits vaisseaux cérébraux de type sporadique***

Elles sont classiquement associées à des facteurs de risque cardiovasculaire tel l'hypertension, le diabète ou l'âge.

Un examen approfondi des données histopathologiques chez les patients révèle que les modifications des vaisseaux consistent principalement en des modifications majeures de la MEC associée à une dégénérescence plus ou moins sévère des cellules musculaires lisses. Les modifications de la MEC peuvent aller de l'épaississement de la membrane basale (MB), au dépôt de protéines de types collagènes normalement présents et éventuellement d'autres composants de la MEC dans l'*intima*, la *média* et l'*adventice*. Il est cependant difficile de tirer des conclusions sur le mécanisme de ces diverses modifications, les observations étant réalisées post mortem, ce qui correspond généralement à un stade très avancé. Une hypothèse serait que ce dépôt anormal de MEC serait présent comme pour un mécanisme de réparation, en réponse à la dégénérescence des cellules musculaires lisses<sup>8</sup>.

De nombreuses études chez l'homme ou l'animal ont montré que l'hypertension, pouvait entraîner un rétrécissement de la lumière artérielle, sans modification de l'épaisseur de la paroi artérielle, modification appelée aussi remodelage interne<sup>9</sup>. Les données actuelles, tendraient à dire que ce remodelage vers l'intérieur résulterait à la fois de variations dans la contractilité des cellules musculaires lisses et dans la composition de la MEC, avec une contribution potentielle de l'activation des métalloprotéase matricielle (MMP). Il existe quelques études portant sur le rôle potentiel de la MEC, avec notamment les dépôts de collagènes et de fibrogènes ou encore sur les modifications de sa composition dans l'étude de l'hypertension mais elles restent relativement encore rares. Cependant une tendance se dégage sur le processus du remodelage artériel au cours de l'hypertension chronique, qui constituerait dans un premier temps un mécanisme d'adaptation/réparation visant à réduire le stress sur la paroi vasculaire et à protéger le lit vasculaire en aval d'une augmentation de la pression artérielle. Mais dans un second temps, ce processus pourrait s'avérer néfaste, entraînant une altération de la vasodilatation, une réduction du débit sanguin, et pour finir, des lésions ischémiques<sup>8</sup>.

### ***1.1.1.2 Maladies héréditaires ou génétiques de petits vaisseaux cérébraux***

Il existe plusieurs maladies génétiques des petits vaisseaux mais seules seront décrites celles étudiées dans ce projet de thèse. Cependant différentes revues récentes auxquelles nous vous renvoyons, listent les maladies identifiées jusqu'à aujourd'hui<sup>6,10,11</sup>.

A l'exception d'un âge de début plus précoce, les manifestations cliniques et IRM de ces formes monogéniques de SVD sont assez semblables à celles des formes sporadiques. Par conséquent elles sont très utiles pour la compréhension de la pathogenèse des formes sporadiques. En particulier, l'identification des gènes responsables de ces formes

monogéniques a permis le développement de modèles animaux, permettant d'étudier la maladie à différents stades, d'identifier les protéines impactées et de comprendre mieux les mécanismes de ces pathologies<sup>12,13</sup>.

#### I.1.1.2.1 CADASIL (cerebral autosomal dominant arteriopathy with subcortical ischaemic strokes and leukoencephalopathy)

CADASIL, l'acronyme en anglais pour « artériopathie cérébrale autosomique dominante avec infarctus sous-corticaux et leucoencéphalopathie », est la cause héréditaire la plus fréquente d'accidents vasculaires cérébraux (AVC) et de démence vasculaire chez l'adulte. Si ces symptômes cliniques sont très similaires à ceux des patients atteints de maladies des petits vaisseaux de type sporadique, ils apparaissent cependant de manière beaucoup plus précoce<sup>14</sup>.

La première description a probablement été réalisée par le neurologue van Bogaert en 1955<sup>15</sup> et, bien que le tableau clinique de CADASIL puisse varier entre les familles et au sein d'une même famille, la maladie est principalement caractérisée par quatre symptômes : migraine avec aura, accidents ischémiques sous-corticaux, troubles de l'humeur, apathie et troubles cognitifs. Néanmoins, la fréquence de ces symptômes varie avec l'âge et la durée de la maladie. Par exemple, les épisodes de migraine avec aura sont observés chez environ 30% des patients avant l'âge de 30 ans. En revanche, l'apathie et les troubles cognitifs sont très fréquents chez les patients âgés de plus de 50 ans. La maladie a également une signature histologique spécifique : la présence de dépôts granulaires osmiophiles, appelée GOMs (*Granular Osmiophilic Materials* en anglais) dans lesquels s'accumule, entre autres, le domaine extracellulaire de Notch3 (*Neurogenic locus notch homolog protein 3*), la protéine responsable de la pathologie<sup>16,17</sup>.

Notch3, le gène dont les mutations sont responsables de la pathologie, code pour une protéine transmembranaire et appartient à la famille des récepteurs Notch, protéines impliquées dans de nombreuses fonctions<sup>18,19</sup>. La voie de signalisation Notch a joué, en particulier, un rôle très important dans le développement vasculaire et la physiologie des vaisseaux<sup>20</sup>.

Notch3 est principalement exprimé dans les cellules musculaires lisses des vaisseaux et dans les péricytes des capillaires. Des études réalisées chez la souris invalidée pour le gène Notch3 démontrent que la protéine Notch3 n'est pas essentielle au développement embryonnaire ou à la fertilité chez la souris<sup>21</sup>. En revanche, ces mêmes études ont démontré que Notch3 jouait un rôle clé dans la différenciation artérielle et la maturation de cellules musculaires lisses des petites artères. Sur le plan fonctionnel, Notch3 joue également un rôle très important dans le contrôle du tonus myogénique des artères cérébrales et dans l'autorégulation du débit sanguin cérébral<sup>22</sup>.

Notch3 est une protéine de 2321 acides aminés, elle est constituée d'un domaine extracellulaire, trois domaines LIN12 et d'un seul domaine transmembranaire. Son domaine extracellulaire comporte 34 domaines EGFR (*Epidermal Growth Factor Repeat*) d'environ 50 acides aminés, chacun contenant un nombre invariant de 6 résidus cystéine. Les mutations du gène Notch3 dans la maladie CADASIL ont pour impact la modification du nombre de résidus cystéine porté par l'un des domaines EGFR. La majorité des mutations observées sont de type faux-sens, elles entraînent la disparition ou l'ajout d'un résidu cystéine. Quelques cas de délétions ou insertions, ne modifiant pas le cadre de lecture ont également été décrites. Quel que soit le type de mutation, elles conduisent toutes au même résultat : un nombre de cystéines impair<sup>23,24</sup>. Le mécanisme pathogénique des mutations CADASIL reste controversé, et si l'idée de perte de fonction est encore étudiée<sup>25</sup>, la plupart des études tendent vers un mécanisme pathogénique de type néomorphique, où Notch3 gagnerait une nouvelle fonction. Ces études ont révélé que seules un petit nombre de mutations, siégeant dans le domaine de liaison au ligand, abolissaient l'activité du récepteur, les mutations les plus fréquentes étaient associées a contrario à une augmentation de l'activité du récepteur<sup>26-29</sup>. En revanche, d'autres études ont clairement démontré que les mutations CADASIL favorisaient l'interaction de la protéine Notch3 mutée avec des protéines de la matrice extracellulaire, telles que TIMP3 et VTN, et leur accumulation anormale dans le vaisseau est à l'origine de la dysfonction des vaisseaux cérébraux<sup>30,31</sup>

#### I.1.1.2.2 CARASIL (cerebral autosomal recessive arteriopathy with subcortical ischaemic strokes and leukoencephalopathy)

CARASIL est l'acronyme en anglais pour « artériopathie cérébrale autosomique récessive avec infarctus sous-corticaux et leucoencéphalopathie ». Sa première description date de 1976<sup>32</sup> et la découverte du gène responsable HTRA1 (High-Temperature Requirement A Serine Peptidase 1) en 2009<sup>33</sup>. Si ces symptômes cliniques sont très similaires à ceux des patients CADASIL, comme son nom l'indique, sa transmission est de type récessif et sa fréquence est donc beaucoup plus rare. Cependant, des cas de mutations du gène HTRA1 sur un seul allèle ont également été associées à une maladie des petits vaisseaux, avec un mode de transmission autosomique dominant<sup>34,35</sup>.

Les symptômes cliniques sont très similaires à CADASIL avec le déficit cognitif, la dysfonction motrice et les accidents vasculaires cérébraux sous-corticaux conduisant à la démence et à une invalidité importante vers 30-40 ans. Il existe toutefois des symptômes cliniques propres à CARASIL, à savoir une alopecie diffuse et des attaques de lombalgies sévères dues à la spondylose déformante et à la dégénérescence discale. Le dysfonctionnement de la mémoire chez CARASIL semble être plus grave que chez CADASIL. Les changements microvasculaires dans CARASIL consistent en une perte importante de cellules musculaires lisses médiales artérielles et de la matrice extracellulaire, avec en particulier des anomalies importantes de la

limitante élastique interne. Les mutations rapportées dans la littérature comprennent des mutations homozygotes non-sens ainsi que des mutations homozygotes faux-sens dans le domaine de la sérine protéase, entraînant la perte de l'expression de la protéine HTRA1 ou de l'activité de la protéase<sup>36</sup>.

HTRA1 appartient à la famille des sérines protéases de type HTRA, qui se distinguent par leur homologie de séquence, leur domaine PDZ et leur architecture oligomérique. Les principales fonctions connues des membres de la famille HTRA sont d'être des contrôleurs-qualité des protéines et d'éliminer les protéines endommagées. Ils participent donc à divers processus physiologiques, tels que la signalisation cellulaire et la dégradation des protéines. HTRA1 est de ce fait, omniprésente dans les tissus humains mais son expression varie en fonction du tissu<sup>37</sup>. HTRA1 chez l'homme est composé de neuf exons codant pour une protéine composée de 480 résidus et comprenant quatre domaines fonctionnels : le domaine protéique de liaison au facteur de croissance (analogue à l'insuline), un domaine d'inhibiteur de sérine protéase de type Kazal, un domaine de la sérine protéase de type trypsine et un domaine PDZ.

HTRA1, comme sa structure l'indique, possède une activité sérine protéase et différents substrats potentiels ont déjà été identifiés<sup>38</sup>, notamment les membres de la voie de signalisation du facteur de croissance  $\beta$  (TGF-  $\beta$ )<sup>39</sup>. Plusieurs études ont été réalisées pour comprendre les mécanismes mis en œuvre, dont certaines se contredisent en particulier sur le fait que HTRA1 est un régulateur positif ou négatif de la voie de signalisation de TGF  $\beta$ <sup>40-43</sup>. Une publication récente suggère notamment une régulation de Notch3 (gène responsable dans la maladie CADASIL) par HTRA1 dans une étude réalisée sur des souris invalidées pour le gène HTRA1<sup>44</sup>.

#### I.1.1.2.3 Maladie du Collagène de type IV

Les maladies des petits vaisseaux cérébraux dues à des modifications du collagène de type IV, sont associées à des manifestations cliniques diverses impliquant le plus souvent le cerveau et les yeux, mais également les reins et les muscles squelettiques. La gravité de la maladie est liée à la survenue d'hémorragies intracérébrales qui peuvent survenir à tout âge de la vie. Les premières familles identifiées datent des années 80<sup>45</sup> et l'identification du premier gène responsable, COL4A1, a pu être résolue par l'utilisation d'un modèle de souris généré par mutagenèse aléatoire vingt ans plus tard<sup>46</sup>. Ce nouveau mutant de souris présentait une hémorragie cérébrale périnatale sévère semblable à celles décrites précédemment dans une grande famille humaine avec une porencéphalie autosomique dominante de type 1<sup>45,47</sup>. Quelques années plus tard, des familles porteuses d'une mutation dans le gène COL4A2 ont été également identifiées<sup>48,49</sup>.

Les collagènes de type IV sont des protéines de la membrane basale qui sont exprimées dans tous les tissus, y compris le système vasculaire. Le COL4A1 (chaîne  $\alpha 1$ ) et le COL4A2 (chaîne  $\alpha 2$ ) sont les collagènes de type IV les plus abondants et forment des hétérotrimères avec une stœchiométrie 2: 1 ( $\alpha 1\alpha 1\alpha 2$ ). Chez les mammifères, la famille du collagène IV comprend six chaînes hautement homologues, mais génétiquement distinctes. Chaque chaîne contient trois domaines structurellement distincts : un domaine amino-terminal riche en résidus de cystéine et de lysine (domaine 7s), un domaine collagène comportant des séquences de répétition triple d'acide aminé Gly-Xaa-Yaa, où Xaa et Yaa peuvent être n'importe quel résidu, mais Yaa est souvent une proline, et un domaine non-collagénique (NC1) en C-terminal. La présence de séquences riches en résidus cystéine et en lysine est essentielle pour la réticulation interchaîne de molécules quadruple hélicoïdales via des liaisons disulfures et des réticulations de lysine-hydroxylysine<sup>50</sup>

La plupart des mutations COL4A1 ou COL4A2 sont des mutations faux-sens entraînant la substitution d'un des résidus glycine dans les répétitions Gly-Xaa-Yaa. L'expression des mutations de COL4A1 est extrêmement variable, même entre les membres d'une même famille, une partie de cette variabilité étant apparemment liée au contexte génétique et aux facteurs environnementaux. Chez les souris, les mutations du domaine triple-hélice près de l'extrémité C-terminal sont associées à des hémorragies intracérébrales plus sévères par exemple que les mutations plus près de l'extrémité N-terminale<sup>51</sup>. Le mécanisme par lequel les mutations dans les gènes COL4A1 ou COL42 exercent leur effet pathogène est encore mal connu. Quelques études ont montré que des mutations du gène COL4A1 pouvaient entraîner une rétention intracellulaire du collagène IV et une diminution de sa sécrétion, à l'origine d'un stress du réticulum endoplasmique. L'observation dans un modèle souris dans lequel l'exon 41 du gène Col4a1 est anormalement épissé, la gravité des hémorragies intracérébrales était associée au niveau d'accumulation intracellulaire mais pas à une réduction de son expression à la MB. Cela suggère que l'accumulation intracellulaire pourrait jouer un rôle dans la survenue des lésions vasculaires à l'origine des hémorragies cérébrales<sup>51</sup>. En revanche, une étude plus récente a démontré dans un autre modèle souris porteur d'une mutation identifiée chez des patients (mutation Gly498Val) l'absence de rétention intracellulaire du collagène de type IV dans les vaisseaux cérébraux mais une diminution importante de l'expression du collagène IV au niveau de la MB<sup>52</sup>.







## I. Contexte scientifique et approches techniques

I.2	LE MATRISOME .....	36
1.2.1	<i>Un peu d'histoire</i> .....	36
1.2.2	<i>Les différentes matrices extracellulaires</i> .....	38
1.2.3	<i>Définition</i> .....	39
1.2.3.1	Le cœur du matrisome (ou core matrisome).....	40
1.2.3.1.1	Les collagènes.....	40
1.2.3.1.1.1	Type et structure .....	41
1.2.3.1.1.2	Synthèse et sécrétion du collagène.....	43
1.2.3.1.1.3	Régulation et dégradation.....	43
1.2.3.1.1.4	Rôles et fonctions du collagène.....	43
1.2.3.1.2	Les protéoglycanes .....	44
1.2.3.1.2.1	Structure.....	44
1.2.3.1.2.2	Synthèse des protéoglycanes .....	45
1.2.3.1.2.3	Rôles et fonctions des protéoglycanes.....	45
1.2.3.1.3	Les glycoprotéines .....	46
1.2.3.1.3.1	Structure.....	47
1.2.3.1.3.2	Synthèse .....	47
1.2.3.1.3.3	Rôles et fonctions.....	48
1.2.3.2	Les protéines associées au matrisome .....	50
1.2.3.2.1	Les protéines régulatrices de la MEC.....	50
1.2.3.2.2	Les protéines sécrétées .....	50
1.2.3.2.3	Les protéines affiliées à la MEC .....	52
1.2.3.3	Forme particulière de la MEC : la membrane basale (MB).....	52
1.2.4	<i>Maladie des petites vaisseaux cérébraux et matrisome</i> .....	56
1.2.5	<i>Protéomique de la matrice extracellulaire</i> .....	58
1.2.5.1	La décellularisation.....	59
1.2.5.2	L'extraction séquentielle des protéines de la MEC.....	60
I.3	PROTEOMIQUE, SPECTROMETRIE DE MASSE (MS) ET QUANTIFICATION .....	64

## I.2 Le Matrisome

Ce second chapitre sera consacré au matrisome, ses compositions et ses différentes fonctions. Je commencerai par un peu d'histoire pour expliquer son concept, un peu plus abstrait, flou que les autres composants cellulaires. Ensuite seront décrites les différentes catégories de protéines de la matrice extracellulaire, et je décrirais principalement un type bien particulier de matrisome, la membrane basale, matrisome principale dans les petits vaisseaux cérébraux. Pour finir, les différentes maladies vues précédemment seront décrites dans le contexte du matrisome.

### I.2.1 Un peu d'histoire

Le matrisome est l'ensemble des protéines de la matrice extracellulaire et des protéines interagissant avec la matrice extracellulaire. Mais qu'est-ce que la matrice extracellulaire ?

Le concept de tissu, ce qu'on sait être aujourd'hui l'ensemble « cellules + matrice extracellulaire », est apparu dans la littérature médicale au cours du XVIIIème siècle notamment dans les textes de Bordeu<sup>53</sup> et de Félix Vicq d'Azyr<sup>54</sup>, le dernier médecin de Marie-Antoinette d'Autriche, dernière reine de France. Mais c'est Xavier Bichat, physiologiste au du XIXème siècle, qui élabore ce concept en tant que doctrine physiopathologique<sup>55</sup>, en ne considérant pas les organes en particulier mais bien les éléments entrant dans leur structure. Très rapidement, l'histologie (Mayer, Heusinger 1819) a permis avec ses méthodes de colorations spécifiques, d'identifier différents constituants macromoléculaires sur les tissus. Cela est suivi de près par le second axiome de la théorie cellulaire : tout organisme vivant est composé de cellule (Schleiden et Schwann 1838). Mais c'est Warren H. Lewis, embryologiste américain, qui va poser la première notion de la matrice extracellulaire en 1922<sup>56</sup> telle que nous la connaissons biochimiquement aujourd'hui :

*« Pourtant, la plupart de nos tissus et organes sont constitués de cellules simplement collées les unes aux autres. Si ces différents types de cellules perdaient leur adhérence mutuelle et **les fibres blanches extracellulaires, réticules**, etc., nos corps se désintégraient immédiatement et s'écouleraient sur le sol dans un flux mixte d'ectoderme, de muscle, de mésenchyme, cellules hépatiques, pancréatiques et de nombreux autres types de cellules. »*

Jusqu'au milieu du XXème, la MEC était considérée comme inerte, servant à remplir juste l'espace extracellulaire mais apportant cependant un support architectural et une résistance mécanique aux tissus. Nous savons aujourd'hui que la MEC est également un composant fondamental des organismes multicellulaires qui orchestrent l'organisation et les fonctions des cellules et des tissus. En effet, c'est un maillage complexe de protéines réticulées fournissant des signaux biophysiques et biochimiques qui sont d'importants régulateurs de la prolifération cellulaire, de la survie, de la différenciation et de la migration<sup>57</sup>. Sont repris en Figure 3 les différents événements importants dans l'histoire de l'étude de la matrice extracellulaire.

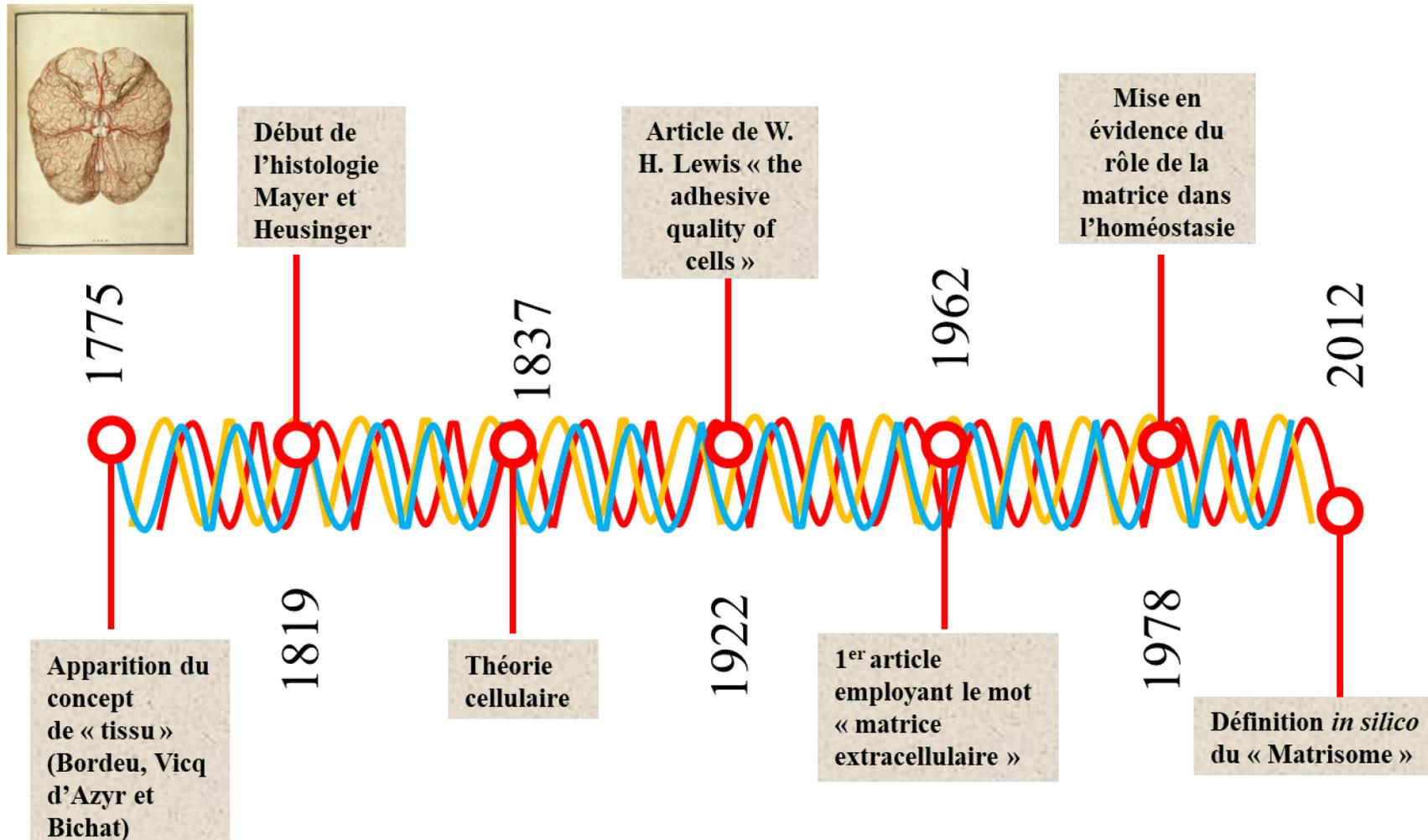


Figure 3 : La matrice extracellulaire de Bordeu à aujourd'hui

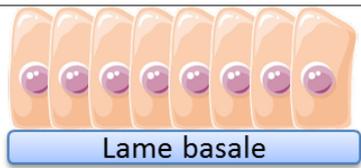
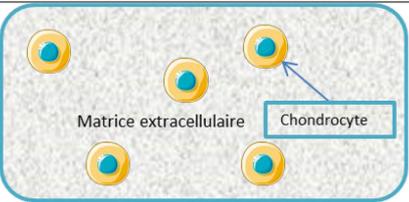
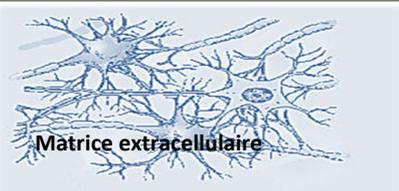
## I.2.2 Les différentes matrices extracellulaires

Chaque organe a une MEC avec une composition unique, la fonction de la MEC va au-delà du soutien physique à l'intégrité et à l'élasticité des tissus : il s'agit d'une structure dynamique constamment remodelée pour contrôler l'homéostasie des tissus. Elle est constituée de diverses protéines, sucres et autres composants<sup>58</sup>.

Chaque type de tissu a une composition et une quantité particulière de MEC, mais on peut les catégoriser en fonction de leur organisation vis-à-vis des cellules du tissu (Table 2).

Quand la MEC est très abondante et les cellules y sont distribuées de manière éparses, c'est une matrice interstitielle, qui est retrouvée principalement dans les tissus conjonctifs. Dans ce cas, c'est la matrice qui supporte la plus grande partie des tensions auxquelles le tissu est soumis, plutôt que les cellules.

Table 2 : Les différents tissus et matrices extracellulaires( inspiré de Dupont, 2018<sup>59</sup>)

Type	Schéma	Exemple
<b>Epithéliale</b>		Epiderme, glandes, muqueuses
<b>Conjonctif</b>		Cartilage, os, ligament, sang
<b>Musculaire</b>		Muscles squelettiques, cœur, parois vasculaires
<b>Nerveux</b>		Cerveau, moelle spinale, nerfs périphériques

Dans les tissus épithéliaux, à l'inverse, les cellules sont étroitement liées entre elles, et associées en feuillets. C'est la matrice péricellulaire. La matrice extracellulaire y est peu abondante et consiste pour l'essentiel en une fine couche : la lame basale. Ce sont les cellules qui subissent la plus grande partie des tensions. Les feuillets de cellules épithéliales tapissent toutes les cavités et les surfaces libres du corps et grâce aux jonctions spécialisées

qui existent entre les cellules, ils forment une protection à la circulation de l'eau, de solutés et des cellules d'un compartiment à un autre. Il existe d'autres types de tissus, dont l'organisation matrice-cellule se situera entre celle de l'organisation du tissu conjonctif et celle du tissu épithélial<sup>60,61</sup>.

### I.2.3 Définition

La matrice extracellulaire (MEC) est un composant essentiel des organismes métazoaires fournissant un support architectural et un ancrage aux cellules, elle est constituée d'eau, de polysaccharides et de protéines. La composition de la MEC détermine les caractéristiques biophysiques du tissu, telle que sa rigidité ou sa déformabilité mais a également un impact sur la régulation cellulaire et ses fonctions. De plus, l'altération de la MEC est associée à diverses maladies telles que le cancer, les fibroses ou encore des maladies du squelette. Le Matrisome est l'ensemble des protéines de la MEC mais également l'ensemble des protéines qui interagissent avec elle. Il est donc nécessaire de mieux caractériser le matrisome afin d'identifier les potentiels biomarqueurs de maladie, et peut-être à terme de soigner ces maladies en agissant sur les différents acteurs.

Si aujourd'hui, plusieurs études ont été réalisées sur différents matrisomes et différents organes, et ce, particulièrement dans le cadre de pathologie tel que le cancer<sup>62</sup> ou encore des maladies neuromusculaires<sup>63</sup>, d'autres études portent plutôt sur la caractérisation du matrisome de manière *in silico*, études qui essaient de prédire la fonction matrisome des protéines codantes ou non, en se basant principalement sur les domaines conservés de ces protéines<sup>57,64</sup>. En effet, les domaines sont des unités structurellement capables de se replier indépendamment de la protéine, ils peuvent être aussi porteurs de fonctions spécifiques (site catalytique, liaison de ligand,..) et ont une séquence homologue que l'on peut rencontrer dans diverses taxonomies<sup>65</sup>.

Très récemment, l'équipe du laboratoire de Richard Hynes a mis au point une définition « bio-informatique » du matrisome en se basant sur la caractéristique des domaines de protéines de la MEC connues. Une liste de 55 domaines a été construite à partir des protéines MEC à l'aide d'Interpro (<https://www.ebi.ac.uk/interpro/>), la base de données de domaines et de signaux protéiques, et celle-ci a été ensuite utilisée pour identifier d'autres protéines dans la banque de données publique Uniprot (<https://www.uniprot.org/>), qui serait encore inconnue en tant que protéine de la MEC. Certains domaines sont présents dans certaines protéines transmembranaires ou impliqués dans l'adhésion cellulaire, et qui n'appartiennent pas à la MEC. Ces protéines possèdent, d'autres domaines particuliers incompatibles avec la définition de protéines MEC. Ce qui a permis une sélection « négative » sur 20 domaines et une prédiction d'un domaine transmembranaire. Le reste de la liste a été compilé sur la base de connaissances antérieures et de données de la littérature<sup>57,66</sup>.

Tout le travail de cette thèse sera réalisé à partir de cette liste de protéines que l'on peut retrouver, (et qui est d'ailleurs régulièrement mis à jour avec de nouveaux organismes, tissus...), sur le site du « Matrisome Project » : <http://matrisomeproject.mit.edu/>. Il est à noter que les critères d'attribution de certaines protéines peuvent être débattus. Par exemple, la désignation des divers facteurs sécrétés a été délibérément inclusive car, bien que beaucoup d'entre eux ne sont pas (encore ?) connus pour être associé à la MEC.

Le matrisome est catégorisé en 2 groupes, le cœur du matrisome ou encore le matrisome central (« *core matrisome* ») et le matrisome associé (« *Matrisome-associated* »). Les premières sont des protéines de types structurales, qui forment la structure, la résistance ou encore l'élasticité de la matrice. Elles sont représentées par les collagènes, les protéoglycanes et les glycoprotéines. Les protéines associées regroupent plutôt des régulatrices telles que des enzymes, des protéines sécrétées comme les facteurs de croissance et les protéines qui ont des similarités structurales avec d'autres protéines de la matrice, ou qui sont connues pour être associées à une protéine de la matrice <sup>57</sup>. La distribution des différentes catégories des protéines du matrisome *in silico*, générée à partir des données humaines publiées sur le site <http://matrisomeproject.mit.edu/>, est présentée sur la Figure 4.

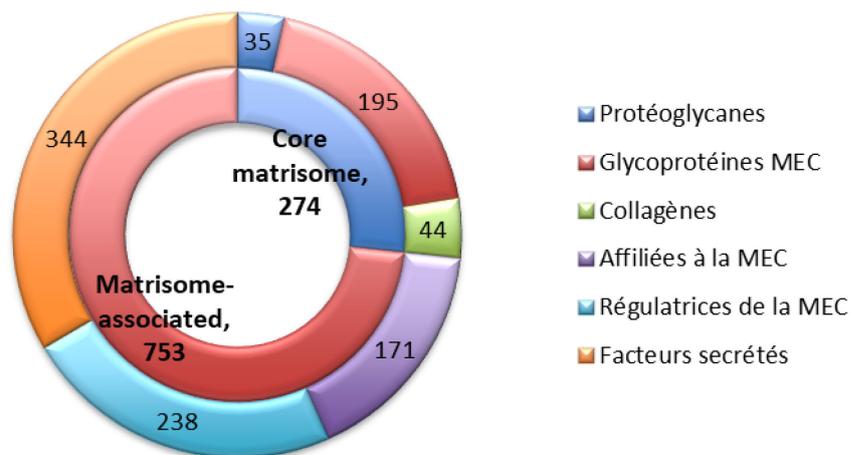


Figure 4 : Distribution des différentes catégories de protéines du matrisome *in silico*

### I.2.3.1 Le cœur du matrisome (ou core matrisome)

#### I.2.3.1.1 Les collagènes

Les collagènes sont les protéines les plus abondantes chez les mammifères, ils constituent environ 30 % de la masse protéique totale. Cette superfamille comporte 28 types différents connus à ce jour, et sont les composants majeurs des os ou encore de la peau. Les collagènes

ont non seulement un rôle architectural dans l'organisation ou la forme des tissus, mais ils contribuent également à diverses régulations cellulaires<sup>67</sup>.

#### *1.2.3.1.1.1 Type et structure*

La structure commune des collagènes est la présence d'une triple hélice, constituée de 3 chaînes peptidiques, nommée chaîne  $\alpha$ , et numérotée en chiffre arabe alors que les collagènes eux sont identifiés par des chiffres romains. La taille des chaînes  $\alpha$  peuvent varier de 662 à 3152 acides aminés, formant une séquence répétitive Gly-X-Y, où X est très fréquemment la proline et Y, l'hydroxyproline. Les trois chaînes peuvent être identiques (homotrimère) ou différentes (hétérotrimères). La triple hélice est stabilisée par la présence d'une glycine tous les 3 résidus, une composition élevée en proline et hydroxyproline, des interactions électrostatiques impliquant la lysine et l'aspartate ainsi que des liaisons hydrogènes. Le degré d'hydroxylation des lysines varie en fonction du type de collagène et de son assemblage dans l'espace extracellulaire. L'association des trois chaînes  $\alpha$  pour former la triple hélice diffère aussi en fonction du type de collagène, mais les mécanismes sont encore peu décrits aujourd'hui.

Les collagènes peuvent également être catégorisés en sous-famille en fonction de leur assemblage supramoléculaire, et qui sont représenté dans la table 3 avec leur structure, leur distribution dans les tissus et quelques caractéristiques.

Les collagènes sont notamment des protéines multidomaines, en effet, si les collagènes fibrillaires contiennent un domaine triple hélicoïdal majeur, les FACIT (*Fibril Associated Collagen with Interrupted Triple helixes*) ou les multiplexines contiennent plusieurs domaines à triple hélice. La découverte du collagène IX, le premier FACIT, a montré que des domaines collagéniques à triple hélice pouvaient être intercalés entre des domaines non collagéniques. Les domaines non collagéniques participent non seulement à l'assemblage mais également à l'activité biologique de la protéine.

Les critères pour nommer une protéine « collagène » ne sont cependant pas totalement définis, et d'autres protéines contenant un domaine en triple hélice ne font pas partie de la famille des collagènes stricto sensu.

**Table 3 : Tableau des différents types de collagène avec leur structure, leur distribution dans les tissus et quelques caractéristiques (Adapté de Ricard-Blum, 2011 <sup>68</sup>)**

Catégorie	Type	Gènes codant pour des protéines typiques de la composition de collagène	Description
<b>Fibrillaire</b>	I	<i>COL1A1</i> et <i>COL1A2</i>	Assemblages des plusieurs molécules de collagène décalées les unes par rapport aux autres d'environ ¼ de leur longueur, qu'on retrouve principalement dans les matrices fibreuses du stroma, telles que la peau, les os, les tendons et les ligaments mais aussi dans la paroi des vaisseaux cérébraux
	II	<i>COL2A1</i>	
	III	<i>COL3A1</i>	
	V	<i>COL5A1</i> , <i>COL5A2</i> et <i>COL5A3</i>	
	XI	<i>COL11A1</i> , <i>COL11A2</i> et <i>COL11A3</i>	
	XXIV	<i>COL24A1</i>	
	XXVII	<i>COL27A1</i>	
<b>FACIT</b>	IX	<i>COL9A1</i> , <i>COL9A2</i> et <i>COL9A3</i>	Ponts moléculaires associés aux fibrilles associés aux collagènes fibrillaire de type I (XII, XVI, XIX, XXI) et de type II (IX, XVI, XIX)
	XII	<i>COL12A1</i>	
	XIV	<i>COL14A1</i>	
	XVI	<i>COL16A1</i>	
	XIX	<i>COL19A1</i>	
	XX	<i>COL20A1</i>	
	XXI	<i>COL21A1</i>	
	XXII	<i>COL22A1</i>	
<b>Membrane basale/Réseau</b>	IV	<i>COL4A1</i> , <i>COL4A2</i> , <i>COL4A3</i> , <i>COL4A4</i> , <i>COL4A5</i> et <i>COL4A6</i>	Structure de réseau composée de laminines et de protéines de la membrane basale
<b>Longue chaîne</b>	VII	<i>COL7A13</i>	Ancrage des fibrilles associées à la membrane basale
<b>Filamenteux</b>	VI	<i>COL6A1</i> , <i>COL6A2</i> , <i>COL6A3</i> , <i>COL6A5</i> et <i>COL6A6</i>	Microfibrilles perlées
<b>Chaîne courte</b>	VIII	<i>COL8A1</i>	Structure en treillis hexagonal
	X	<i>COL10A1</i>	Ces collagènes régulent et sont régulés par l'hypertrophie du cartilage.
<b>Les multiplexines</b>	XV	<i>COL15A1</i>	Plusieurs domaines de triple hélice avec des interruptions contenant du glycosaminoglycane de sulfate de chondroïtine et d'héparine sulfate
	XVIII	<i>COL18A1</i>	
<b>MACIT (Membrane-Associated Collagens with Interrupted Triple-helices)</b>	XIII	<i>COL13A1</i>	Molécules de surface cellulaire avec des domaines extracellulaires, transmembranaires et intracellulaires
	XVII	<i>COL17A1</i>	
	XXIII	<i>COL23A1</i>	

#### *1.2.3.1.1.2 Synthèse et sécrétion du collagène*

Les collagènes sont d'abord synthétisés par les ribosomes pour donner les tropocollagènes, des précurseurs des chaînes  $\alpha$  avec des extensions peptidiques non hélicoïdales aux extrémités terminales. Les chaînes  $\alpha$  sont ensuite soumises aux différentes modifications telles que l'hydroxylation des lysines ou proline ou encore des glycosylations au niveau des hydroxylsines dans le réticulum endoplasmique avant la formation des trois hélices qui débute par une association des 3 chaînes à leur extrémité C-terminale. Cette structure globulaire est stabilisée par la formation de ponts disulfure intra et inter-chaînes, et avec l'aide de plusieurs protéines chaperonnes HSP47, ces trois chaînes vont s'enrouler entre elles afin de former la triple hélice jusqu'à l'extrémité N-terminale. Les trois chaînes associées, nommées à ce stade procollagène, transitent par l'appareil de Golgi avant d'être relarguées dans l'espace extracellulaire, ce qui entraîne l'action des enzymes N et C-peptidases, qui vont cliver les deux extrémités propeptidique et former la protéine de collagène.

Pour la protéine de collagène I, la perte de ses extrémités N et C-terminale lors de sa sécrétion va diminuer d'un facteur 1000 sa solubilité et entraîner l'agrégation avec d'autres protéines de collagène : les fibrilles de collagène<sup>69</sup>. Mais pour d'autres collagènes (type IV, VII, IX et XII), non fibrillaires, ces extrémités terminales ne seront pas clivées.

#### *1.2.3.1.1.3 Régulation et dégradation*

Explorer toutes les informations concernant la régulation, dépasserait le cadre de cette introduction, mais nous savons que la régulation des collagènes dépend du type de collagène, du type de cellules, de l'âge, ... Elle se fait au niveau transcriptomique évidemment mais également au niveau de son assemblage et de sa dégradation. Nous savons également qu'interviennent notamment de nombreux facteurs de croissances et des cytokines.

La dégradation des collagènes est principalement due à l'action des métalloprotéinases de type MMP<sup>70</sup>, des enzymes connues pour leur rôle dans les phénomènes physiologiques tels que le développement ou la réparation des tissus mais également pathologiques telle la tumorigenèse. Ces protéines seront décrites un peu plus loin dans ce texte, car ce sont également des protéines du matrisome.

#### *1.2.3.1.1.4 Rôles et fonctions du collagène*

Les collagènes fibrillaires, les plus abondants chez les vertébrés, contribuent à l'organisation architecturale moléculaire, à la forme et la résistance des tissus. Le collagène de type IV est un constituant majeur de la MB, qui sera décrite également plus loin. Mais d'autres collagènes, dit « mineurs » sont tout aussi essentiels à l'intégrité des tissus pour exemple le cartilage est

principalement constitué, en terme de collagène, du type II, mais on trouve également la présence de collagène de type IX (1% à l'âge adulte) et de type XI (3 % à l'âge adulte)<sup>71</sup>.

Quant à leur fonction de régulation cellulaire, on peut citer le rôle du collagène de type VII dans la cicatrisation, intervenant également dans l'ancrage des fibrilles. Ou encore l'intervention du collagène de type XXVII qui jouerait un rôle lors de la calcification du cartilage et de la transition du cartilage vers l'os<sup>72</sup>.

#### I.2.3.1.2 Les protéoglycanes

Les protéoglycanes (PG) sont des protéines que l'on trouve dispersées parmi les fibrilles de collagène dans différentes MEC. Au lieu de contribuer à la résistance structurelle, elles confèrent des propriétés supplémentaires telles que la lubrification ou encore des fonctions de remplissage. En effet, elles sont composées de protéine et de chaînes polysaccharidiques appelées glycosaminoglycanes (GAGs). L'ajout de GAG fournit aux protéoglycanes une charge négative élevée, et cette charge négative, rend la molécule très hydrophile et est capable d'attirer des quantités considérables de molécules d'eau ainsi que des cations tels que le calcium ou le sodium. Par exemple dans les cartilages, cette propriété permet aux PG de résister aux forces de pression, en déplaçant l'eau retenue. D'autre part, certains GAGs, en particulier les sulfates d'héparane ou HSPG (*Heparane Sulfate ProteoGlycan*), sont un des constituants principaux de la membrane basale mais lient également de nombreux facteurs de sécrétion et de croissance à la MEC<sup>73</sup>.

##### I.2.3.1.2.1 Structure

La multitude de contributions apportée par les protéoglycanes à des processus physiologiques tels que la prolifération cellulaire, la migration, l'immunité et l'assemblage de la matrice est en grande partie due à la diversité structurelle de cette superfamille moléculaire.

Les protéoglycanes sont constitués en un noyau protéique auquel des chaînes polysaccharidiques linéaires chargées négativement sont liées par covalence. Les GAG sont composés de répétitions de disaccharides et varient en ce qui concerne leurs schémas de sulfatation. On distingue plusieurs types de GAG, classés en fonction de la structure de leurs séquences de chaînes disaccharidiques, le nombre et la position du groupement sulfate : les héparanesulfates ou héparine, les chondroïtines sulfates, les kératines sulfates et les dermatanes sulfates. Ces chaînes structurellement distinctes donnent des propriétés de liaison et des capacités de signalisation différentes aux protéines, dont le cœur protéique peut également varier<sup>74</sup>. En effet, certains cœurs protéiques ont la particularité de pouvoir être exprimés sans l'addition de GAG, on dit qu'ils sont PG à temps partiel, contrairement à d'autres qu'on ne retrouve pas autrement, comme les syndecanes. D'autres sont encore

limités à type bien précis de GAG, alors que d'autres peuvent subir de multiples modifications<sup>74</sup>.

#### *1.2.3.1.2.2 Synthèse des protéoglycanes*

Les principales étapes de la synthèse des protéoglycanes (excepté le sulfate de kératine KS) sont les suivantes : la formation de la protéine de cœur, la formation du tétrasaccharide particulier (GlcA-Gal-Gal)-Xyl) qui se liera à certaines serines de la protéine de cœur et ensuite les chaînes de GAG sont polymérisées par ajout successif de motifs disaccharidiques, avant d'être potentiellement modifiées<sup>75</sup>. Ces étapes sont coordonnées par de nombreuses enzymes telles que les glycosyltransférases ou sulfotransférases, présentes à la membrane du réticulum endoplasmique (RE) et dans l'appareil de Golgi. Après synthèse, les PG sont transportés de l'appareil de Golgi vers leur destination : la MEC, la surface cellulaire ou les organites intracellulaires. La variabilité observée dans la structure de la chaîne GAG d'un type cellulaire à l'autre indique que la localisation cellulaire des PG dépend en grande partie de la composition des GAG<sup>76</sup>.

#### *1.2.3.1.2.3 Rôles et fonctions des protéoglycanes*

La diversité de la fonction des protéoglycanes rapportée dans la littérature est quantitativement égale à la diversité de la structure des protéoglycanes, cependant, comme les collagènes, les protéoglycanes sont des protéines clés de la MEC, apportant non seulement un soutien structurel mais également en influençant le comportement cellulaire dans les processus physiologiques et pathologiques<sup>77</sup>.

Néanmoins, une classification des PG basée sur la localisation cellulaire, l'homologie globale protéine/gène et l'association protéine/GAG permet d'avoir un aperçu des différentes fonctions existantes pour ce type de protéines. Voici les quatre grandes classes de PG et quelques PG pour exemple :

- Le protéoglycane intracellulaire :  
La serglycine, présente dans les granules de mastocytes, et qui sert notamment de colle biologique aux protéases présentes dans les granules. Elle est seule dans sa catégorie.
- Les protéoglycanes à la surface cellulaire :  
Les syndécans, des protéines transmembranaires à héparane sulfate (HSPG), dont les différentes fonctions sont l'adhésion à la MEC, en se liant notamment avec les collagènes de type I, II et V, ou encore de nombreuses liaisons avec divers facteurs de croissance.

Les glypicanes, qui sont également des HSPGs, sont des protéines dont l'extrémité C-terminale est liée à la membrane plasmique via une ancre glycosylphosphatidylinositol (GPI). Elles sont impliquées dans le contrôle de la croissance tumorale et de l'angiogenèse.

- Les protéoglycanes péri-cellulaires et de la membrane basale

Les perlécane, autre HSPG, sont exprimés dans divers tissus (vasculaire, avasculaire, membrane basale), ils interviennent entre autres dans la modulation de l'adhésion cellulaire, dans le métabolisme des lipides.

Les agrines, HSPG à haut poids moléculaire, possèdent 9 domaines homologues aux inhibiteurs de protéase. Leur forme transmembranaire est très exprimée dans les axones et les dendrites, et pourrait être en s'associant avec les neurites, un récepteur ou un corécepteur. Comme le perlécane, leurs fonctions sont diverses et variées grâce leur structure particulière et modulable.

- Les protéoglycanes extracellulaires :

Les hyalectanes sont une famille de PG comprenant l'aggrécane, le versicane, le neurocane et le brévicane, des protéines ayant une structure tridomaine. Un domaine N-terminal qui se lie à l'hyaluronane, un domaine central contenant les chaînes latérales GAG et un domaine C-terminal qui lie les lectines. Elles agissent souvent comme des ponts moléculaires entre la surface des cellules et les matrices extracellulaires.

- Les SLRPs (*Small Leucine-Rich Proteoglycans* = petites protéoglycanes riches en leucine), sont des protéines de masse moléculaire relativement bas comparativement aux autres PG, comme leur nom l'indique, constituée de répétition riche en leucine. Elles sont présentes dans la plupart des MEC et sont fortement exprimées au cours du développement dans les membranes minces enveloppant les organes essentiels tels que les méninges, le péricarde ou encore l'endomètre. Cette omniprésence suggère qu'elles seraient impliquées dans la régulation de la forme et de la taille des organes pendant le développement embryonnaire et l'homéostasie. Elles seraient également impliquées dans l'assemblage du collagène fibrillaire <sup>78</sup>.

#### 1.2.3.1.3 Les glycoprotéines

Les glycoprotéines<sup>79</sup> sont des protéines porteuses de manière covalente d'un ou plusieurs groupes oligosaccharidiques. Ce sont, comme les protéoglycanes, des protéines relativement volumineuses et complexes. Leur rôle est fondamental dans de nombreux processus biologiques, notamment la défense immunitaire, la croissance et l'adhésion cellulaire, la fécondation ou encore la réplication virale. Il existe plusieurs types de glycosylation, mais les

plus courantes sont la N-glycosylation et la O-glycosylation. Cette modification peut influencer de manière significative la conformation de la protéine, et par conséquent sa fonction ou son interaction avec d'autres protéines. La glycosylation est la modification post-traductionnelle la plus abondante mais aussi la plus complexe, plus de la moitié des protéines naturelles sont glycosylées<sup>80</sup>.

#### *1.2.3.1.3.1 Structure*

La structure des glycoprotéines peut être extrêmement variable, en effet, elles peuvent posséder un ou plusieurs sites de N ou O-glycosylation, et différents sites peuvent être glycosylés, ou pas, avec des structures de glycosylation différentes selon les conditions (âge, pathologie, ...). Une même protéine peut avoir différentes activités en fonction de ses glycoformes. Sept monosaccharides composent les différents glycanes liés aux protéines : le mannose, le glucose, le fucose, le galactose, la N-acétylgalactosamine, la N-acétylglucosamine et les acides sialiques. Ils peuvent être reliés entre eux par chaînes linéaires ou ramifiées de longueur et de complexité variable, et parfois aussi modifiés (sulfatation, phosphorylation...), ce qui augmente considérablement la diversité des glycoprotéines.

Lors d'une N-glycosylation, la liaison entre la protéine et la chaîne saccharidique se fait entre l'azote (N) du groupement amide de l'asparagine et une N-acétylglucosamine. Les arbres glucidiques de ce type de liaison peuvent être très étendus avec une diversité de glycanes souvent grande et volumineuse.

La O-glycosylation, a un encombrement stérique moindre que la N-glycosylation. Les sucres sont généralement plus petits et moins ramifiés, cependant ils ne partagent pas de noyau commun bien défini et leur structure peut-être très diverse.

#### *1.2.3.1.3.2 Synthèse*

La synthèse des glycoprotéines va naturellement varier en fonction du type de glycosylation. Par exemple, la N-glycosylation est une modification essentiellement co-traductionnelle, le glycanne est rattaché à la protéine lors de son transport dans le RE. Ensuite, la biosynthèse continue avec un assemblage de glycanes précurseurs, qui lors de la phase de maturation dans l'appareil de Golgi, seront combinés par différents clivage et addition d'oses, réalisé par des enzymes stockés dans les citernes de l'appareil de Golgi. Ce qui donnera cette grande variété de glycoprotéines.

La O-glycosylation est une modification catalysée généralement par des glycosyltransferases, qui ajouteront les résidus de polysaccharide directement sur les acides aminés cibles les uns après les autres.

#### *1.2.3.1.3.3 Rôles et fonctions*

Les glycoprotéines sont fondamentales dans de nombreux processus biologiques importants, notamment la fécondation, la défense immunitaire, la réplication virale, les infections parasitaires, la croissance cellulaire, l'adhésion cellulaire, la dégradation des caillots sanguins et l'inflammation. Les glycoprotéines sont des composants majeurs de la surface externe des cellules. Les structures des oligosaccharides changent radicalement au cours du développement et il a été démontré que des ensembles spécifiques d'oligosaccharides sont exprimés à des stades distincts de la différenciation. En outre, les altérations des oligosaccharides à la surface des cellules sont associées à diverses conditions pathologiques.

La majorité des protéines de surface cellulaire ou secrétées sont glycosylées. Le fragment glucidique d'une glycoprotéine peut participer directement à divers événements de reconnaissance, mais également modifier les propriétés de la protéine. En effet, la grande taille des oligosaccharides leur permet de couvrir des zones importantes du point de vue fonctionnel et de moduler les interactions des protéines avec d'autres molécules<sup>81</sup>.

Voici dans la table 4 une liste non exhaustive de différentes familles de glycoprotéine de la MEC ainsi que leur activité/ fonction biologique.

Table 4 : Quelques groupes de glycoprotéines de la MEC, leurs fonctions connues et les articles qui en réfèrent

Glycoprotéines	Description et fonctions	Localisation	Références
<b>Fibulines</b>	<p>Groupe de 7 glycoprotéines liées au Ca<sup>2+</sup></p> <ul style="list-style-type: none"> <li>• Rôle dans l'assemblage des fibres élastique dans la MEC</li> <li>• Rôle dans plusieurs processus cellulaires tels que la croissance, la différenciation ou l'angiogenèse.</li> </ul>	Présentes dans la membrane basale, sang et les artères	82
<b>Fibronectines</b>	<p>Grande glycoprotéine, c'est un dimère formé de 2 sous-unités liées par 2 ponts disulfures près de leur extrémité C-terminale</p> <ul style="list-style-type: none"> <li>• Régule l'adhésion en se liant avec l'intégrine (protéine transmembranaire)</li> <li>• Acteur important dans la fibrillogénèse</li> </ul>	Présentes dans toutes les matrices	82, 83
<b>Fibrinogènes</b>	<p>Grandes glycoprotéines fibreuses à trois paires de chaînes polypeptidiques</p> <ul style="list-style-type: none"> <li>• Joue un rôle important dans la coagulation du sang, la fibrinolyse, les interactions entre cellules et matrice, l'inflammation, la cicatrisation des plaies et la néoplasie</li> </ul>	Principalement dans le sang	82, 84
<b>Laminines</b>	<p>Protéines hétérotrimériques de grandes tailles</p> <ul style="list-style-type: none"> <li>• Contribue à la structure globale en interagissant avec les intégrines, dystroglycane, COL4, etc...</li> <li>• Intervient dans le fonctionnement normal de divers organes tels que les nerfs, les vaisseaux sanguins ou les reins.</li> </ul>	Présentes dans la membrane basale	82
<b>Elastine</b>	<p>Protéine de 68 kDa, insoluble, et dont l'hydrophobicité lui confère des propriétés élastiques.</p> <ul style="list-style-type: none"> <li>• Un des constituants principaux des fibres élastiques présentes dans les tissus.</li> </ul>	Artères, tissu conjonctif, poumons, peau et vessie	85

### 1.2.3.2 Les protéines associées au matrisome

Dans cette catégorie, il y a principalement les protéines connues pour s'associer à la MEC, et les protéines possédant des similitudes avec celles-ci. Ce groupe est également divisée en trois catégories, catégorisation basée sur leur activité, fonction biologique, et non sur leur structure contrairement aux protéines du core matrisome <sup>66</sup>.

#### 1.2.3.2.1 Les protéines régulatrices de la MEC

Les fonctions de ces protéines sont de se lier (transferase), de couper (hydrolases) ou de modifier (les « modification » -ases) les protéines de structure de la MEC.

Pour les classifier, il existe une nomenclature dite EC, pour *Enzyme Commission*, développée par le comité des nomenclatures de l'Union Internationale de Biochimie et de Biologie Moléculaire (NC –IUBMB), en se basant sur le type de réactions catalysées <sup>86</sup>, ce qui nous ramène au 3 premières classes sur les 6 existantes : les oxydoréductases, les transférases et les hydrolases qui représentent plus de la moitié des protéines régulatrices .

Cependant, il existe également une bases de données un peu plus spécifique, MEROPS (<https://www.ebi.ac.uk/merops/index.shtml>), dans laquelle sont représentées huit familles de protéases et leur(s) inhibiteur(s) dont la classification est basée cette fois-ci sur le fragment chimique qui participe à l'hydrolyse <sup>87</sup> et non plus à la réaction. C'est en partie à partir de cette base de classification que sera représentée une liste non exhaustive des différentes familles de ces protéines régulatrices de la MEC (Table 5)

#### 1.2.3.2.2 Les protéines secrétées

La MEC possède également un « réservoir » de nombreux facteurs. De cette manière, la MEC peut conserver certains facteurs de croissance qui peuvent être libérés par protéolyse et peuvent induire localement une prolifération, croissance ou différenciation. Nous y trouvons des facteurs de croissances, notés généralement XGF, X pour le tissu concerné et GF pour *Growth Factor* (facteur de croissance), tels que le FGF (*Fibroblast Grow Factor*) ou encore l'EGF (*Epidermal Grow Factor*), des cytokines ou encore d'autres facteurs secrétés qui vont se lier avec différentes protéines de la MEC, telle que la fibronectine, la vitronectine ou encore la syndécane.

Table 5 : Quelques famille de protéines régulatrice de la MEC, leur(s) fonction(s), leur(s) substrat(s) et les articles qui en réfèrent

Famille de protéines	Classification	Quelques protéines MEC	Références
<b>Hydroxylases, oxydases, dioxygénases</b>	Type EC 1.14.11.-, 1.4.3.13.— Réactions d'oxydo-réduction	<b>Les lysyl oxydase (LOX)</b> : forment des aldéhydes à partir de lysine →réticulation du collagène, de l'élastine	88
<b>Transférases ou phosphorylase</b>	Type EC 2.7.1.-, Réactions de phosphorylation	<b>Les glycosaminoglycan xylosylkinase</b> sont responsable de la phosphorylation du xylose dans la région de liaison glycosaminoglycane-protéine des protéoglycanes, régulant ainsi la quantité de chaînes de GAGs matures	Uniprot www.uniprot.org
<b>Sérines protéases</b>	Type EC 3.4.21.- ; identifiant MEROPS : S01 Réaction d'hydrolyse d'un lien peptidique avec un résidu sérine sur le site actif de l'enzyme	<b>HTRAs, Plasminogènes, Trypsines, chymotrypsines...</b>	89
<b>Cathepsines</b>	Type EC 3.4.22.- ; identifiant MEROPS : C01 Réaction d'hydrolyse d'un lien peptidique avec un résidu cystéine sur le site actif de l'enzyme	<b>Les cathepsines</b> à cystéine sont des protéases de plusieurs protéines de la MEC (laminines, collagène...)	90
<b>Les metalloprotéases</b>	Type EC 3.4.24.- ; identifiant MEROPS : M10 (matrix metallopeptidase) et M12 (adamylsines) Réaction d'hydrolyse d'un lien peptidique avec un un ion Zn <sup>2+</sup> lié à 3 résidus histidine sur le site actif de l'enzyme	<b>Les protéines ADAMs (A Disintegrin And Metalloprotease)</b> sont capables d'adhérer aux intégrines grâce à leur séquence riche en cystéine, et comme <b>les MMPs (Matrix MetalloProtease)</b> , peuvent dégrader des composants de la MEC. Les inhibiteurs des <b>MMPs</b> sont les Tissue Inhibitor of Metalloproteoinases , les <b>TIMPs</b>	91 , 70

### I.2.3.2.3 Les protéines affiliées à la MEC

Dans ce groupe, nous trouvons des familles de protéines qui partagent des similitudes architecturales ou biochimiques avec d'autres protéines de la MEC ou qui sont connues pour être associées avec des protéines de la MEC<sup>57</sup>.

**La MEC est donc un échafaudage dynamique constitué de divers matériaux (les collagènes, les glycoprotéines et les protéoglycanes) mêlés entre eux, modifiés, ajustés, découpés par des ouvriers spécialisés chacun dans un domaine particulier (les protéines régulatrices) mais est également un réservoir de matières premières, d'outillage (les protéines secrétée) nécessaire pour le bon fonctionnement de cette charpente auto-adaptable.**

***Chaque groupe contribue à sa manière afin de former des assemblages tout en maintenant une homéostasie structurale et fonctionnelle des tissus.***

### *I.2.3.3 Forme particulière de la MEC : la membrane basale (MB)*

La membrane basale (MB) est une matrice de type péricellulaire, elle se situe entre les cellules épithéliales et le tissu conjonctif. Le sujet de cette étude portant sur l'analyse des petits vaisseaux, nous nous intéressons particulièrement à cette forme notable de la MEC. Si la composition des molécules qui la constitue varie en fonction du tissu, la membrane basale est généralement composée de collagène de type IV, de nidogène (ou entactine), de laminines, de perlecane et parfois un peu de collagène de type XV et XVIII. Sa fonction est de soutenir les cellules endothéliales, musculaires et adipeuses mais également de jouer le rôle de filtre entre les cellules épithéliales et le compartiment sanguin, ou entre deux types de tissus conjonctif différents.

La microscopie électronique en transmission révèle une structure en forme de feuille de 50 à 100 nm d'épaisseur, d'aspect amorphe. Ces structures morphologiques se retrouvent généralement sous-jacentes à des feuilles de cellules endothéliales et entourant la plupart des cellules musculaires, cellules adipeuses et axones nerveux périphériques<sup>92</sup>

Bien que la «lame basale» et la «membrane basale» soient souvent utilisées de manière interchangeable dans certains tissus, tels que le muscle squelettique, la membrane basale est composée de trois structures distinctes appelées lamina<sup>93</sup> :

- Au plus près de la surface épithéliale, se trouve la lamina lucida, une couche transparente aux électrons, d'environ 50 nm d'épaisseur. Elle est riche en laminines, qui vont se lier à la portion extracellulaire des intégrines.

- La lamina densa est la couche intermédiaire, elle est constituée de collagène de type IV mais également de laminines, de nidogène et de perlécane. Elle a une largeur variable de 20 à 300 nm, selon l'emplacement.
- La lamina fibroreticularis, fibrillaire comme son nom l'indique, contient des composant des fibrilles d'épaisseur variable qui se fondent imperceptiblement dans la matrice du tissu conjonctif. Elle est constituée de collagène de type I, III et VII mais également de fibronectine qui la relie aux autres laminae.

Les composants de la membrane basale sont de grandes protéines modulaires (figure 6). Plus de 20 gènes dans la membrane basale ont été identifiés et séquencés.

- **Le collagène de type IV** est le composant de la membrane basale le plus répandu et est considéré comme responsable de la stabilité mécanique de la membrane basale. C'est un collagène non fibrillaire, il fait partie des collagènes en réseau (Table 3). Chez les mammifères, il existe six chaînes génétiquement distinctes de Col IV,  $\alpha 1$  (IV) à  $\alpha 6$  (IV), tous ont des structures de domaine similaires. Il est composé d'un domaine court non collagénique, d'un domaine collagène à triple hélice moyen long et d'un domaine globulaire non collagénique. La partie hélicoïdale centrale de la chaîne  $\alpha$  contient 1400 Gly-X-Y, cette séquence est interrompue par environ 20 courtes régions non collagènes, ce qui explique la souplesse des membranes basales. Trois chaînes  $\alpha$  (2  $\alpha 1$  et 1  $\alpha 2$ ) s'associent pour former un protomère (ou monomère), ensuite 2 protomères vont générer un dimère, jusqu'à l'élaboration d'un échafaudage supramoléculaire comprenant des tétramères de collagène de type IV sous forme de réseau polygonal (Figure 5).

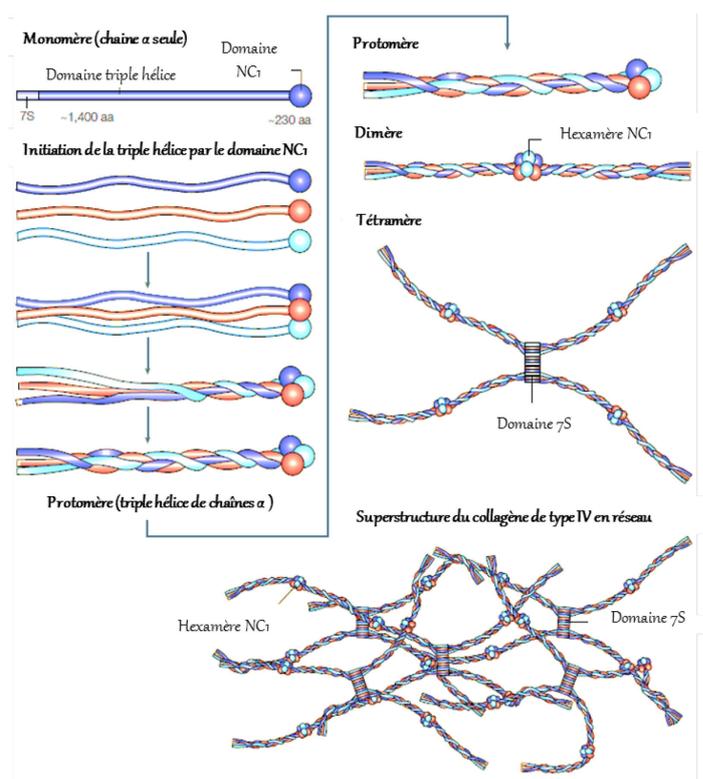


Figure 5 : Description et formation de la structure en réseau du collagène de type IV (adapté de Kalluri 2003<sup>94</sup>)

- **Les laminines**, une famille de glycoprotéines présente uniquement dans les membranes basales, sont les glycoprotéines non collagéniques les plus abondantes dans cette matrice. Les laminines sont des trimères composées de chaîne  $\alpha$ ,  $\beta$  et  $\gamma$ , classées selon une nomenclature bien spécifique <sup>95</sup>. La laminine-111, une protéine avec une composition de chaînes  $\alpha 1$ ,  $\beta 1$  et  $\gamma 1$ , dont l'autoassemblage est dépendant du calcium, est le type de laminine le plus abondant dans les membranes basales, d'autres types de laminine sont exprimées dans la membrane basale mais d'une manière spécifique au tissu. La chaîne  $\alpha 1$  longue d'environ 200 kDa, possède trois domaines globulaires au niveau de son extrémité N-terminale, séparés par des séquences EGF, une séquence hélicoïdale et un domaine G, comportant 5 domaines globulaires. Les chaînes  $\beta$  et  $\gamma$ , sont composés de 2 domaines globulaires en N-terminal séparé par 2 EGF répétés et d'une séquence hélicoïdale. L'assemblage de ces trois chaînes forme une croix, le bras long de la croix étant les séquences en triple des chaînes  $\alpha 1$ ,  $\beta 1$  et  $\gamma 1$ , avec le domaine G de la chaîne  $\alpha 1$ . Les bras courts sont respectivement les N-terminaux de chaque chaîne. Le domaine G est le site principal d'adhésion cellulaire de la laminine et le site de liaison avec le nidogène (ou entactine) est localisé dans les répétitions EGF du domaine  $\gamma 1$ . Les LM-111, LM-211 et LM-221 s'auto-assemblent par gélification thermique dépendante du calcium.
- Le **nidogène** <sup>96</sup> ou **entactine** <sup>97</sup>, est une glycoprotéine de 150 kDa, a une structure sous forme d'haltère comportant trois domaines globulaires séparés par deux segments en forme de tige. Elle se lie à la laminine via la chaîne  $\gamma 1$  et son domaine G2 interagit avec la région triple hélice du collagène de type IV.
- **Le perlécane**, un protéoglycane comportant des héparanes sulfates (HSPG) d'environ 500 kDa, est également abondamment exprimé dans les membranes basales. Il doit son nom à sa forme ressemblant à une série de perles lorsqu'il est visualisé au microscope électronique. La protéine comporte 5 domaines distincts et N-terminal est lié à trois héparane sulfate qui alourdit sa masse de 800 kDa. Chaque domaine possède une fonction distincte mais cette introduction se limitera aux modules en lien avec les protéines de la membrane basale. Le domaine I qui représente la région N-terminal et où sont fixés les chaînes héparane sulfate, ne contient pas de résidu cystéine et son rôle principale est la séquestration (et la libération en temps voulu) des facteurs de croissance. Fait relativement rare pour être relevé, aucune homologie n'existe entre ce module et d'autres protéines connues, présentes dans les bases de données protéiques <sup>98</sup>.

Le domaine III et une partie du domaine V, homologues du bras court de la laminine, comprend trois domaines globulaires de laminine parmi 11 répétitions EGF<sup>99</sup>. Outre ces quatre protéines majeures, il existe d'autres composants de la membrane basale, qui vont notamment varier d'un tissu à l'autre.

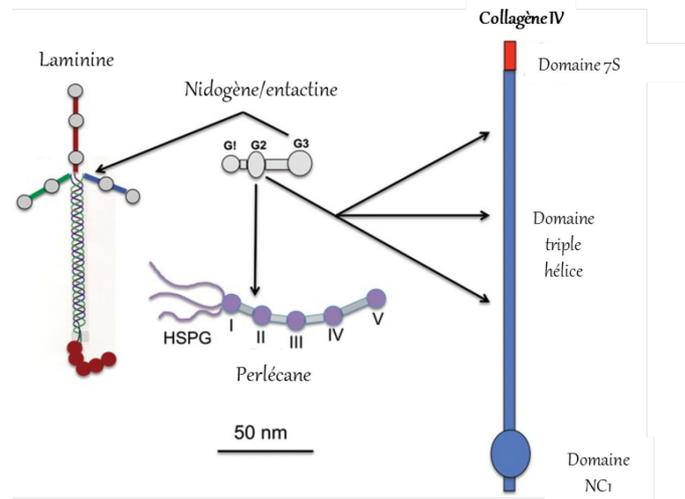


Figure 6 : Description de l'interaction entre les différentes protéines majoritaires de la MB (adapté de Mak 2017<sup>93</sup>)

## I.2.4 Maladie des petites vaisseaux cérébraux et matrisome

Pour rappel (voir paragraphe 1.1), le cercle de Willis se divise en 3 artères principales et à leur tour, elles vont donner plusieurs artères qui vont progressivement devenir plus petites et courir à la surface du cerveau <sup>100</sup>.

Selon le type, l'emplacement ou encore le diamètre des vaisseaux sanguins, le matrisome sera différent. Dans les grandes artères et artérioles, la MEC est une structure spacieuse dans laquelle les cellules musculaires lisses sont présentes tandis que les capillaires comportent uniquement une très fine membrane basale, qui sépare l'endothélium des péricytes, des astrocytes et des neurones (Figure7).

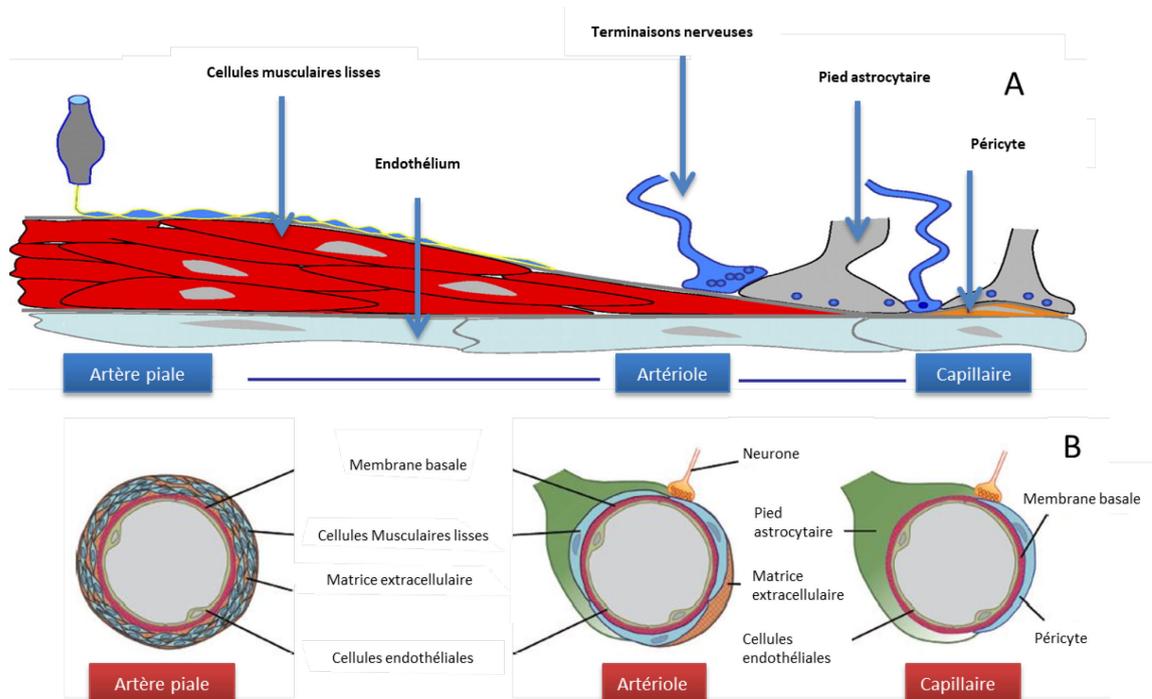


Figure 7 : A : Schéma illustrant la transformation progressive des artères cérébrales en capillaires (d'après Girouard,2006)<sup>101</sup>. B : Coupe transversale des différents types de vaisseaux ainsi que leur composition (d'après Haffner, 2016)<sup>102</sup>.

Le matrisome, notamment la membrane basale est donc un lieu très stratégique entre la microcirculation cérébrale et les astrocytes. La membrane basale des vaisseaux contribue à l'intégrité de la barrière hémato-encéphalique (BBB, *Blood Brain Barrier*), qui est formée par les cellules endothéliales et les capillaires cérébraux. Ces cellules reçoivent le soutien de péricytes et des pieds astrocytaires. La membrane basale des petits vaisseaux cérébraux

forme un réseau protéique tridimensionnel composé principalement de laminine, de collagène IV, de nidogène et de protéoglycanes sulfate d'héparane (perlécane), et plus particulièrement la fibronectine insoluble, la fibuline 1 et 2, le collagène de type XVIII, les thrombospondines 1 et la protéine SPARC (*secreted protein acidic and cysteine rich*). L'assemblage de la membrane basale vasculaire dépend d'un autoassemblage initial de laminine en une feuille suivie d'une liaison au nidogène et aux HSPG. Pour compléter la formation de la membrane basale vasculaire, le nidogène et les HSPG sont liés au collagène IV, permettant ainsi à ce dernier de former un réseau polymère. Selon la taille et l'emplacement du vaisseau, c'est principalement la composition et les isoformes de laminine qui seront modifiés. Cette membrane soutient mutuellement les interactions entre les cellules endothéliales, les péricytes et les astrocytes. Des changements majeurs dans la composition moléculaire de la membrane basale sont observés dans des contextes pathologiques. Lors d'un accident vasculaire cérébral, la perte d'intégrité de la BBB s'accompagne d'une régulation positive des enzymes protéolytiques et d'une dégradation des protéines de la membrane basale<sup>103</sup>.

Les données histopathologiques dans le cadre des maladies des petits vaisseaux indiquent que les modifications sont principalement des modifications majeures de la MEC associée à une dégénérescence plus ou moins sévère des cellules musculaires lisses. En effet, les maladies des petits vaisseaux sont décrites physiologiquement comme une expansion massive de matrice extracellulaire, appelée également fibrose<sup>104</sup> qui va entraîner l'épaississement de la MB ou encore l'accumulation excessive de protéines, les agrégats de protéines telles que les protéines du collagène<sup>105</sup> dans les *intima*, les *media* et les *adventices*.

Très récemment, deux revues ont mis évidence le lien entre les maladies des petits vaisseaux et la MEC qui entoure les petits vaisseaux<sup>8,106</sup>. La plupart des formes monogéniques de maladies des petits vaisseaux sont causées par des modifications des protéines du matrisome.

Le cas le plus évident étant la maladie du **collagène de type IV**, constituant fondamental de la membrane basale, la modification génétique entraîne de fortes altérations fonctionnelles de la membrane basale des vaisseaux cérébraux. Le collagène de type IV interagit avec des composés important pour la formation et la stabilité de la MB tels que la laminine, le nidogène et le perlécane<sup>107,108</sup> mais il a été démontré également que le réseau COL4A1/COL4A2 interagit directement avec plusieurs voies de signalisation impliquées dans l'angiogenèse telles que le facteur de croissance-TGFβ<sup>109</sup>, voie de production de la MEC, incontournable pour la stabilité et l'intégrité des vaisseaux<sup>39,110</sup>.

Dans le cadre de **CARASIL**, la protéine HTRA1 est non seulement une protéine de la MEC mais également une enzyme importante dans la régulation de divers processus physiologiques dans la MEC. Elle a aussi probablement rôle dans la voie de signalisation du TGFβ, bien que certaines études se contredisent<sup>40-43</sup>. Et en tant que protéase, elle a de nombreux substrats,

notamment dans la MEC, telles que la fibronectine<sup>111</sup>, l'aggrécane<sup>112</sup>, ou encore la fibromoduline, la vitronectine et la protéine ADAM9<sup>38</sup>.

Dans la pathologie **CADASIL**, nous avons montré que si Notch3 n'est pas une protéine de la MEC, son action et son aggrégation sont directement liées à plusieurs protéines de la MEC telle que TIMP3 ou encore la vitronectine<sup>113</sup>. En effet, une étude réalisée sur la dérégulation positive de ces deux protéines, anormalement recrutées par NOTCH3<sup>ECD</sup> (Notch3 ExtraCellular Domain), d'un modèle préclinique de souris CADASIL avait pour conséquence la réduction du tonus myogénique (la capacité des artères de résistance à se contracter suite à une élévation de la pression) et l'apparition de lésions de la substance blanche. Cependant, malgré la réduction de ces protéines par la manipulation du niveau d'expression suggère que d'autres protéines pourraient être en jeu<sup>31</sup>. LTBP-1, la protéine de liaison latente TGF- $\beta$  1, protéine de régulation important dans la voie de signalisation du TGF $\beta$  est également recrutée dans les GOMs, ce qui confirmerait à nouveau une dérégulation de cette voie de signalisation essentielle<sup>114</sup>. HTRA1 a également une augmentation de son expression dans CADASIL, bien qu'elle ne soit pas recrutée dans les GOMs, et une étude récente suggère que sa perte de fonction, serine protéase, serait un facteur important dans la pathologie<sup>115</sup>.

En comparant les maladies monogéniques et les maladies sporadiques des petits vaisseaux cérébraux, il a pu être constaté que les symptômes cliniques et les données histopathologiques sont très similaires, notamment CADASIL et CARASIL, ce qui signifie que des perturbations spécifiques du matrisome ont la capacité de produire de lésions vasculaires équivalentes. Quelques études protéomiques ont été réalisées dans le cadre des maladies monogéniques, surtout CADASIL, que cela soit sur de l'humain (en stade terminal généralement) ou sur des animaux<sup>25,113,115</sup>, il y en a beaucoup moins pour les maladies sporadiques des cSVD, et peu se concentrent particulièrement sur la matrice extracellulaire<sup>113,115</sup>, qui est le principal sujet de cette thèse.

### **1.2.5 Protéomique de la matrice extracellulaire**

Bien que la spectrométrie de masse en protéomique soit une méthodologie extrêmement puissante, permettant d'analyser et même de quantifier plusieurs milliers de protéines en quelques heures, l'étape de traitement de l'échantillon, qui va consister principalement à extraire les protéines d'intérêts de cellules ou de tissu et à les rendre compatibles avec une analyse par le spectromètre de masse, est fondamentale. L'isolement ou la purification des protéines est la première étape nécessaire, elle sera ensuite suivie d'une étape de clivage spécifique des liaisons peptidiques avec une ou plusieurs enzymes, étape qui est communément appelée digestion des protéines. Le chapitre « 1.3 Protéomique, spectrométrie de masse et quantification » décrira de manière plus générale les méthodologies pour l'extraction et la solubilisation des protéines des tissus.

Il existe différentes méthodes d'extraction en fonction du matériel initial, que cela soit des cellules<sup>116-118</sup>, des fluides<sup>119</sup> ou encore un tissu<sup>120,121</sup> mais il faut tenir compte de la question à laquelle il faut répondre. En effet, si la recherche porte sur les différents substrats d'une protéine, d'une enzyme particulière<sup>122,123</sup>, d'une seule protéine et de ses modifications post-traductionnelles<sup>124</sup>, d'une particularité physicochimique comme les protéines membranaires<sup>125,126</sup>, ou encore de protéines issues d'une localisation particulière en utilisant en particulier le fractionnement subcellulaire<sup>127,128</sup>, le traitement apporté sera très différent. Plusieurs approches ont été développées pour traiter et enrichir les composants du matrisome et éliminer les protéines intracellulaires, cependant elles consistent principalement en deux méthodes distinctes, la décellularisation et l'extraction protéique. Elles sont très souvent combinées<sup>129</sup>.

### *1.2.5.1 La décellularisation*

Comme son nom l'indique, cette méthode va consister à extraire les cellules pour ne laisser que la matrice extracellulaire environnante. Elle implique l'utilisation de détergents faibles, et/ou d'enzymes telle que la trypsine ou de tampons à faible force ionique pour perturber l'organisation des membranes cellulaires ce qui mènera à la libération de protéines intracellulaires en solution tout en préservant la construction MEC insoluble.

L'utilisation de tissus décellularisés est de plus en plus fréquente dans les stratégies de médecine régénérative et d'ingénierie tissulaire. Ce processus mis au point par Stephen F. Badylak<sup>130</sup>, permet aujourd'hui aux patients d'avoir une greffe de tissu en minimisant le risque de rejets.

La procédure est relativement simple : suite à l'immersion du tissu dans un tampon dénaturant tel que le SDS, le Triton X-100 ou l'hydroxyde d'ammonium<sup>131-135</sup> pour désorganiser les composants cellulaires et les éliminer, les maillages de MEC décellularisés peuvent ensuite être homogénéisés, solubilisés et digérés par des protéases pour l'analyse protéomique des composants de la MEC. Ces techniques de décellularisation, développées initialement à des fins d'ingénierie tissulaire où l'élimination complète des protéines cellulaires est essentielle pour éviter d'éventuelles réponses immunitaires indésirables, demandent plusieurs heures à plusieurs jours de traitement physique et chimique ce qui peut conduire à la dégradation des protéines.

Selon le type de tissus ou de cellules, le rendement est différent, cependant les différentes études ayant utilisé cette technique<sup>131,133,134</sup> arrivent à la même conclusion : cette méthode enrichit principalement les protéines constitutives du cœur matrisome, notamment les collagènes et les glycoprotéines, au détriment des protéines associées au matrisome, qui sont pour beaucoup d'entre elles solubles et seront donc éliminées lors des différents lavages. D'après une étude comparative sur les différentes méthodes d'enrichissement du matrisome<sup>134</sup>, un traitement sans enrichissement des tissus de types pulmonaire et mammaire peut

conduire au même taux d'identification des protéines du matrisome voire à une meilleure couverture du matrisome qu'une étude avec un enrichissement au préalable. Il est à noter cependant que ce type de méthode a été développé initialement pour des préparations de matrice pour l'ingénierie médicale tissulaire, et le but tend vers l'élimination des cellules, de l'ADN tout en préservant la structure de la MEC, et non du matrisome.

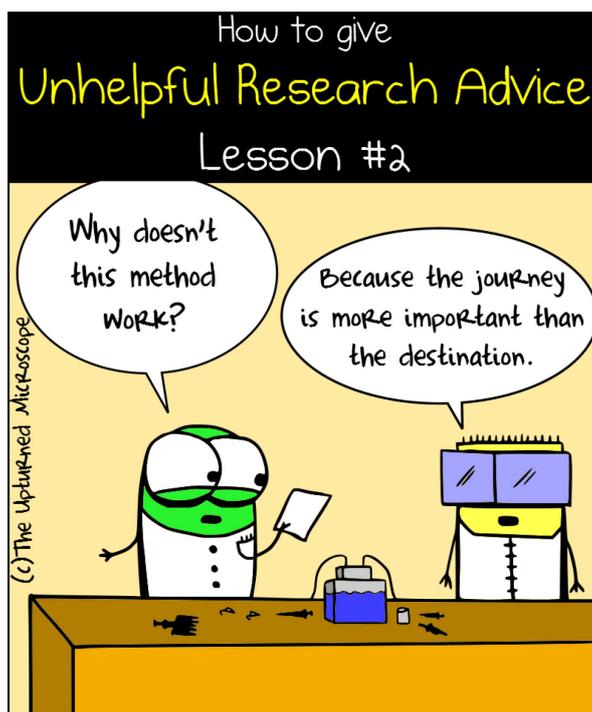
#### *1.2.5.2 L'extraction séquentielle des protéines de la MEC*

Une autre méthode couramment utilisée pour enrichir et extraire les protéines de divers tissus est l'utilisation de divers solvants et détergents pour extraire séquentiellement les protéines des différents compartiments cellulaires et récupérer la fraction la plus insoluble, les protéines de la MEC. Selon les tissus et les protocoles, il peut y avoir quelques différences mais dans l'ensemble, le principe reste le même, utiliser des solvants de plus en plus stringents jusqu'à récupérer le culot insoluble contenant les protéines de la MEC<sup>136-138</sup>. Ensuite, une autre étape importante est de le solubiliser en utilisant d'autres détergents ou encore des agents chaotropes comme le chlorure de guanidium<sup>137</sup>, du diméthylsulfoxyde<sup>131</sup> mais aussi des protéases telles que la déglycosylase, la pepsine ou la collagénase<sup>131,136,137,139</sup>.

Avec un rendement légèrement plus élevé que la décellularisation, l'identification des protéines associées au matrisome reste relativement bas par rapport à celui des protéines constitutives du cœur matrisome. Les différentes étapes d'extraction sont très probablement responsables de la perte de cette catégorie de protéines<sup>136</sup>.







## I. Contexte scientifique et approches techniques

<b>I.3.1</b>	<b><i>Pourquoi analyser le Protéome ?</i></b> .....	<b>65</b>
<b>I.3.2</b>	<b><i>Traitement de l'échantillon de type tissulaire, de l'extraction à la digestion</i></b> .....	<b>65</b>
I.3.2.1	Homogénéisation .....	66
I.3.2.2	Méthode d'extraction et de solubilisation .....	67
I.3.2.3	Digestion enzymatique des protéines .....	68
I.3.2.4	Séparation par chromatographie liquide des mélanges peptidiques complexes .....	69
<b>I.3.3</b>	<b><i>La spectrométrie de masse et l'analyse protéomique</i></b> .....	<b>70</b>
I.3.3.1	Un nouveau type de détecteur : le piège à ions orbital (Orbitrap).....	70
I.3.3.2	Le QqOrbitrap ou Q Exactive .....	72
I.3.3.3	Les différents modes d'acquisition en spectrométrie de masse .....	74
I.3.3.4	Identification des protéines et moteurs de recherche .....	76
I.3.3.4.1	Comment se fragmentent les peptides et comment interpréter ces spectres MS/MS de fragmentation ? .....	76
I.3.3.4.2	Les moteurs de recherches dans les banques de données protéiques.....	78
<b>I.3.4</b>	<b><i>La quantification relative en protéomique par spectrométrie de masse</i></b> .....	<b>82</b>
I.3.4.1	Quantification avec marquage .....	82
I.3.4.2	Quantification sans marquage (« Label Free ») .....	83
I.3.4.2.1	Normalisation des données .....	85
I.3.4.2.2	Traitement des données manquantes.....	86

### I.3 Protéomique, spectrométrie de masse (MS) et quantification

Les sujets abordés dans ce chapitre seront plus méthodologiques et analytiques, il sera question des traitements de l'échantillon de type tissu en vue d'une analyse en spectrométrie de masse. Les principes de la spectrométrie de masse pour l'analyse des protéines, ainsi que son couplage avec la chromatographie liquide sont également décrits, avant de terminer avec les différentes méthodes de quantification relative utilisées en spectrométrie de masse, le traitement des données ainsi que leur interprétation.

Dans ce manuscrit ne sera pas abordée l'étude des protéines entières ni leur mode d'analyse en spectrométrie de masse, le *top-down* (approche descendante) ou la spectrométrie de masse native pour les complexes. Seule l'approche descendante, *bottom-up* en anglais, utilisée pendant ce travail de thèse, sera abordée. Pour cette approche, les protéines (*up*) sont lysées enzymatiquement ou chimiquement et ce sont leurs peptides qui sont, éventuellement séparés par chromatographie liquide (HPLC), détectés en MS (*bottom*) et séquencés en MS/MS (figure 8). Ces analyses sont suivies d'un traitement bio-informatique permettant de mettre en évidence les protéines candidates. Les masses ou les éléments de séquence des peptides obtenus expérimentalement vont être comparés aux masses et séquences obtenues en simulant l'hydrolyse de toutes les protéines répertoriées dans les bases de données. Des algorithmes permettent alors de déterminer avec une grande robustesse les protéines analysées par comparaison.

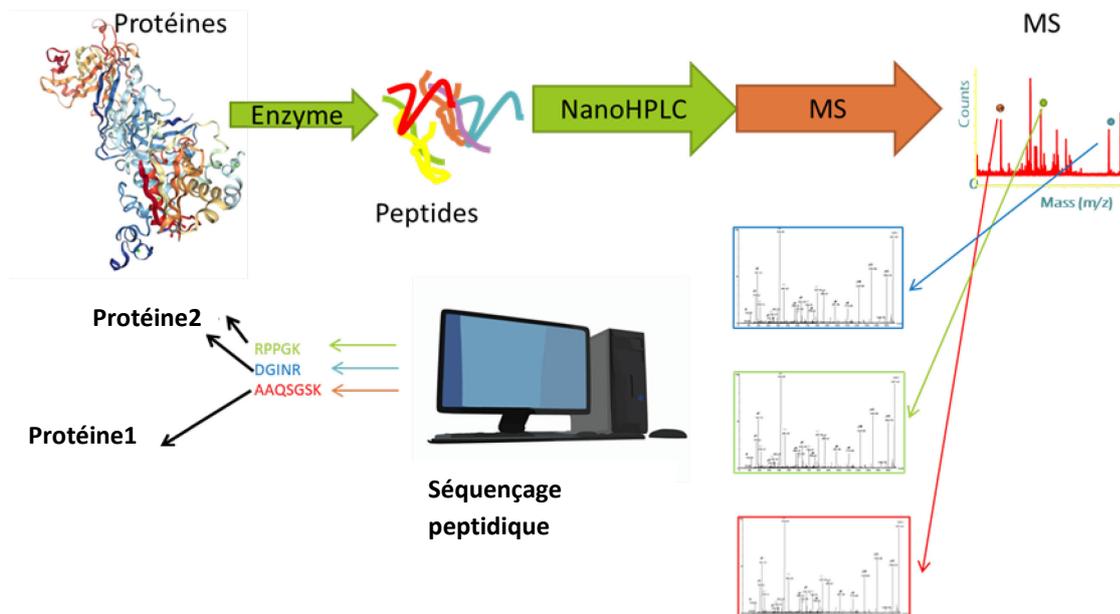


Figure 8 : Stratégie du *bottom-up* en protéomique

### **I.3.1 Pourquoi analyser le Protéome ?**

Le protéome, terme qui désigne l'ensemble des produits du génome d'un être vivant, est l'ensemble des protéines exprimées dans un organe, tissu ou cellule donné à temps donné dans des conditions données. La protéomique est une approche avancée qui vise à caractériser une ou des protéines, ou un protéome, en une ou plusieurs expériences. A l'inverse de l'analyse d'une protéine unique utilisant des techniques basées sur des anticorps, telles que le Western blot, le dosage immuno-enzymatique (ELISA) et la microscopie fluorescente, et qui ne peut être que ciblé sur une protéine et ses isoformes, l'utilisation de la spectrométrie de masse offre une opportunité précieuse d'identifier la plupart des protéines de manière complètement exploratoire dans une condition définie. La spectrométrie de masse peut également, après une expérimentation exploratoire, analyser des échantillons de manière plus ciblée.

Si dans un premier temps, l'étude protéomique en spectrométrie de masse avait pour but de déterminer, dans la mesure du possible, la composition en protéines la plus exhaustive possible d'un type cellulaire ou tissulaire, très logiquement (et rapidement), les chercheurs ont souhaité l'utiliser pour comparer les extraits protéiques provenant de cellules ou de tissus dans deux ou plusieurs conditions différentes. Décrire et comprendre le protéome complet et de façon quantitative ainsi que corréliser sa structure, sa fonction et sa dynamique est un défi central et fondamental de la biologie <sup>140</sup>.

### **I.3.2 Traitement de l'échantillon de type tissulaire, de l'extraction à la digestion**

De nombreuses stratégies de préparation d'échantillons ont été développées afin de permettre une production complète et reproductible de peptides protéolytiques à partir de protéomes extraits d'une grande variété de types d'échantillons. Une quantification précise du protéome dans les cellules et les tissus est nécessaire pour une interprétation adéquate des conséquences fonctionnelles des changements dans les protéoformes pour l'étude de diverses pathologies. Bien que la culture cellulaire soit un modèle de choix important pour étudier les processus se produisant dans les cellules, sa pertinence et sa valeur sont limitées pour une étude au niveau d'un organisme <sup>141</sup>. De plus, les tissus préservent l'organisation et la communication cellulaires natives en trois dimensions. L'analyse directe des protéomes tissulaires est donc devenue indispensable, particulièrement dans la protéomique clinique. Les méthodes pour en extraire les protéines sont sensiblement les mêmes, un tissu étant composé de cellules, les différentes techniques présentées ci-dessous peuvent être appliquées sur les deux types d'échantillons. L'extraction et la solubilisation simultanées de toutes les protéines d'un tissu reste malheureusement une utopie, compte tenu de

l'hétérogénéité des différents molécules présentes, protéines et autre. Il n'y a pas de méthode universelle mais un traitement adapté à chaque type d'échantillon. Il est indispensable de le déterminer expérimentalement.

### *1.3.2.1 Homogénéisation*

Cette première étape va consister à homogénéiser l'échantillon et l'action à mener est dépendante de l'échantillon à traiter. Nous supposons que l'échantillon va modifier ses propriétés physiques sans aucun changement dans la chimie des composants.

Pour les tissus, la technique la plus rapide et la plus efficace est le broyage ou le cisaillement, qu'il soit manuel ou mécanique à l'aide d'une machine type rotor stator ou broyeur à billes. Ces manipulations nécessitent la présence de tampons de force dénaturante adaptée au type de protéines ciblées. Elle doivent être réalisées à une température de 4°C pour limiter la dégradation de l'échantillon protéique et éventuellement, cela dépend du type de tissu, en présence d'inhibiteur de protéases <sup>142</sup>.

L'ultrasonication, qui consiste à appliquer une énergie par ondes ultra-sonores pour perturber (par cavitation) et déstabiliser jusqu'à la rupture de certaines liaisons des molécules diverses d'un échantillon, est aussi une méthode utilisée pour homogénéiser des tissus, qui peut être réalisée dans un bain de sonication ou par une sonde en contact direct avec le tissu <sup>143,144</sup>. L'ultrasonication est utilisée sur des tissus dit mou, tels que le foie ou le cerveau. Cependant elle est déconseillée dans les études de complexes non covalents. En effet c'est également un outil utilisé pour la modification des structures protéiques, notamment dans l'industrie alimentaire <sup>145</sup>.

Si la lyse cellulaire, de type osmotique ou à l'aide de détergent n'est pas une méthode proprement dite d'homogénéisation des tissus, elle accompagne généralement les méthodes « physiques » citées précédemment. Quand les cellules sont plongées dans un environnement hypotonique ou hypertonique, on parlera de lyse osmotique, le déséquilibre osmotique va entraîner un excès d'eau dans ou hors de la cellule, ce qui permettra de récupérer, ou d'éliminer, les protéines plasmatiques. L'utilisation de détergents pour lyser les cellules est expliquée dans le paragraphe suivant, puisqu'ils permettent également de solubiliser les protéines hydrophobes.

### 1.3.2.2 Méthode d'extraction et de solubilisation

L'extraction et la solubilisation optimales des protéines à partir de tissus est indispensable pour accéder à l'ensemble des protéines disponibles pour l'analyse protéomique. Cela permet d'une part d'optimiser la reproductibilité mais également de réduire les contaminants.

La structure secondaire d'une protéine, représentée notamment par le repliement local de sa chaîne principale, est essentiellement constituée par un réseau d'interactions entre les acides aminés constitutifs. Les interactions entre les protéines sont basées sur des interactions équivalentes. Pour accéder à l'entièreté de la protéine, il faut donc briser ces interactions tout en préservant la structure primaire. La rupture des interactions peut être réalisée à l'aide de plusieurs méthodes dont les modes d'action diffèrent mais dont les conséquences sont similaires : déployer la protéine et la solubiliser.

- Les détergents :

Les détergents utilisés ici sont des composés amphiphiles, ils sont composés d'un groupe hydrophile et un groupe hydrophobe séparés d'une chaîne hydrocarbonée, et c'est cette propriété qui va affaiblir et rompre les différentes interactions existantes entre les protéines, les membranes cellulaires et permettre la solubilisation des protéines jusqu'aux plus hydrophobes.

La partie polaire peut être non ionique, anionique, cationique ou zwitterionique. L'étendue de la solubilisation et la stabilité de la protéine solubilisée dépendront du type et de la concentration du détergent<sup>146</sup> mais également de l'échantillon, de la température et du temps d'incubation.

Les détergents ioniques sont souvent plus efficaces, leur mode d'action est de se lier à la protéine, puis modifier sa charge apparente et sa structure. Le détergent le plus couramment utilisé est le dodécyl sulfate de sodium (SDS), mais il est peu compatible avec la trypsine et totalement incompatible avec la chromatographie à polarité de phase inversée, et avec les sources d'ionisation en spectrométrie de masse. De plus, il est difficile de s'en débarrasser pour la suite des expérimentations<sup>147-149</sup>. Bien que plusieurs méthodes d'élimination existent<sup>150-152</sup>, la plus ancienne étant le gel d'électrophorèse<sup>153</sup>, elles peuvent être relativement laborieuses et entraîner une perte partielle de l'échantillon<sup>154</sup> ce qui peut être préjudiciable.

Le sodium deoxycholate (SDC) est également un détergent ionique, naturel, présent dans la bile, et bien qu'il soit un peu moins efficace que le SDS en termes de solubilisation et de dénaturation, il a l'avantage d'être compatible avec la trypsine, et son élimination est très simple et efficace. Il est de plus en plus utilisé en protéomique<sup>126,155-157</sup>.

Les détergents non ioniques, tels que n-dodécyl  $\beta$ -D-maltoside(DDM) ou le Triton X-100 sont plus doux et non dénaturants et leur action est beaucoup moins agressive que les détergents ioniques, ce qui est intéressant lors d'études d'interactions entre

deux protéines, mais ont un rendement de solubilisation généralement plus faible<sup>158-160</sup>.

La force d'interaction des détergents zwitterioniques, qui ont un groupe à la fois positif et négatif, se situe en termes de force entre les détergents ioniques et les non-ioniques, ils sont également non dénaturants mais suffisamment puissants pour rompre les liaisons inter-protéine tout en conservant l'état natif et la charge des protéines isolées<sup>161,162</sup>.

- Agents chaotropiques

A l'instar des détergents, les agents chaotropiques sont utilisés pour affaiblir l'effet des propriétés hydrophobes de certaines protéines en détruisant les liaisons hydrogènes entre les molécules d'eau. Bien que les liaisons ioniques ne soient pas directement affectées par les chaotropes non ioniques tels que l'urée et la thiourée, l'influence de ces chaotropes sur la constante diélectrique de l'eau modifie également la force des liaisons ioniques. Cependant, les chaotropes seuls ne sont pas efficaces pour les échantillons de membranes complexes. Par exemple, ils sont incapables de solubiliser la bicouche lipidique des membranes, qui est nécessaire pour la solubilisation des protéines membranaires. De plus, en dépliant des protéines, les chaotropes augmentent le nombre de résidus hydrophobes exposés au solvant, et donc les possibilités d'agrégation des protéines<sup>163</sup>.

- Les agents réducteurs

Le rôle des agents réducteurs est de rompre les liaisons disulfures entre les résidus des cystéines, favorisant ainsi le dépliement des protéines et permettre ainsi une meilleure accessibilité à l'enzyme de digestion. Classiquement le dithiothréitol (DTT) est utilisé en protéomique, mais sont utilisés également le tri-carboxy-ethyl-phosphine (TCEP), qui a l'avantage de ne pas réagir avec les alkylants nécessaires pour bloquer les thiols afin que les ponts ne se reforment pas, le tributyl phosphine ou le beta-mercaptoéthanol.

### ***1.3.2.3 Digestion enzymatique des protéines***

La protéolyse des protéines, par voie enzymatique ou chimique (ce dernier aspect ne sera pas abordé dans ce paragraphe), est l'ultime étape importante et indispensable, avant l'analyse en spectrométrie de masse, pour l'identification, la caractérisation et la quantification des protéines par des stratégies protéomiques *bottom-up*. Toutefois une étape de dessalage, d'enrichissement ou de fractionnement, ou une combinaison de ces actions, peut être réalisée avant l'analyse.

Après extraction et solubilisation, les protéines sont mises en présence d'une endoprotéase, qui coupe à l'intérieur de la chaîne peptidique générant des fragments de longueur variable selon la nature du(des) site(s) de coupure(s) préférentielle(s). La trypsine est l'enzyme généralement utilisée lors d'analyses de type exploratoire, en effet, elle a la spécificité de couper en C-terminal des résidus lysine (K) et arginine(R) (sauf s'ils sont suivis d'une proline (P)), deux acides aminés basiques, facilitant ensuite des analyses en fragmentation par spectrométrie de masse tandem. Ainsi, ceux-ci sont chargés à leurs deux extrémités et se fragmentent de façon très régulière. L'arginine et la lysine sont des acides aminés généralement bien répartis, 5% en moyenne sur une protéine, ce qui permet d'obtenir des peptides d'environ 10 à 15 résidus d'acides aminés, taille favorable pour les analyses en spectrométrie de masse tandem <sup>164</sup>.

Cependant, selon l'étude réalisée, il peut être nécessaire d'utiliser d'autres protéases ou même de les combiner pour augmenter le taux d'identification et de couverture protéique<sup>165</sup>. Il existe des enzymes très spécifique comme l'Arg-C, qui coupera en C-terminal des arginines mais aussi des enzymes dont la spécificité est très « large » comme la pronase ou la chymotrypsine.

#### ***1.3.2.4 Séparation par chromatographie liquide des mélanges peptidiques complexes***

Tout échantillon biologique donné contient typiquement un large mélange de biomolécules avec une gamme diversifiée d'affinités d'adsorption et par conséquent, la seule méthode pratique pour identifier son contenu est une séparation par gradient. La chromatographie en phase liquide est la technique de séparation privilégiée pour l'analyse de mélanges peptidiques en séparations mono- ou multidimensionnelles. La chromatographie à polarité de phases inversée (*Reversed Phase Liquid Chromatography*) est la méthode plus répandue pour la séparation de peptides, lorsqu'elle est suivie par la spectrométrie de masse. Les composés d'intérêt sont séparés en fonction de leur hydrophobicité, par interaction avec une phase stationnaire apolaire hydrophobe. Les protéines sont identifiées par la mesure des rapports m/z de leurs peptides et la fragmentation de leurs peptides protéolytiques, une séparation efficace, représentée par la résolution, est nécessaire pour une identification sans ambiguïté<sup>166,167</sup>. Il existe toutefois de nombreuses autres stratégies de séparation peptidique<sup>168,169</sup> qui ne seront pas abordées dans ce document.

### I.3.3 La spectrométrie de masse et l'analyse protéomique

Si les premiers développements de la spectrométrie de masse en tandem (MS/MS) visaient à l'origine l'étude des petites molécules très vite son application majeure devient la détermination de structure de biomolécules. Les techniques d'ionisation douce des biomolécules produisent des ions moléculaires de faible énergie interne qui se fragmentent peu, ce qui permet la mesure de masse de grosses biomolécules. Elles valurent d'ailleurs à leurs inventeurs, John B. Fenn pour l'ESI (*ElectroSpray Ionization*)<sup>170</sup> et Koichi Tanaka pour le MALDI (*Matrix Associated Laser Dissociation Ionization*)<sup>171</sup> le prix Nobel de chimie en 2002.

Un spectromètre de masse est nécessairement constitué de 3 éléments : la source d'ions, un analyseur pour séparer les ions selon leur rapport masse sur charge ( $m/z$ ) et un détecteur pour enregistrer le signal des ions séparés par l'analyseur. Le principe est simple : un mélange de peptides est injecté dans le spectromètre de masse et est ionisé par la source d'ions, l'analyseur mesure ensuite la masse de ces différents ions avant de les fragmenter pour obtenir des informations de séquences qui vont permettre de remonter à la protéine initiale. Bien que facultative, l'utilisation d'un préfractionnement de l'échantillon peptidique est souvent nécessaire lors d'analyses de mélange de protéines.

La protéomique nécessite aujourd'hui l'analyse de mélanges extrêmement complexes avec pour but la détection d'analytes d'une grande variété sur une large gamme de concentrations. Une mesure de masse précise, associée à une résolution suffisante, permet de réduire le nombre de formules possibles qui pourraient être représentées par une masse moléculaire particulière. Ce challenge demande des instruments plus performants, plus résolutifs et plus précis. L'évolution technologique dans ce domaine n'a jamais cessé depuis trente ans et c'est cette demande qui va pousser Alexander Makarov à mettre au point un analyseur révolutionnaire en spectrométrie de masse, un piège orbital, ou encore son nom commercial « Orbitrap », qui utilise le principe de la transformé de Fourier<sup>172</sup>.

Il existe de nombreux appareils composés de différentes sources, analyseurs et détecteur, mais seul sera décrit le spectromètre de masse utilisé durant cette étude : **le quadrupole Orbitrap HF ou Q Exactive HF.**

#### I.3.3.1 Un nouveau type de détecteur : le piège à ions orbital (Orbitrap)

Les analyseurs dans un spectromètre de masse ont pour rôle de guider les ions pour les séparer en fonction des rapports masse/charge. Chaque analyseur de masse possède des propriétés uniques, telles que la gamme de masse accessible, la vitesse d'analyse, la résolution, la sensibilité, la transmission ionique et la gamme dynamique. Il existe diverses catégories d'analyseurs, les analyseurs à faisceau ionique, tels que les quadripôles Q, à piège

à ions, tels que les trappes (IT, Ion Trap), et les analyseurs à transformée de Fourier (FT) tels que les FT-ICR (Fourier-transform ion cyclotron resonance) et les Orbitrap. Le TOF (*Time Of Flight*, principal analyseur derrière une source MALDI) sépare les ions en fonction de leur temps de vol qui est caractéristique de leur  $m/z$ , alors que les FT-ICR et Orbitrap basent leur distinction sur la fréquence de résonance des ions dans un champ spécifique de leur rapport  $m/z$ . Les hybrides Q-TOF ou Q-FT sont relativement performants pour certaines gammes d'applications, mais souffrent néanmoins d'une plus faible transmission ionique (entraînant une faible sensibilité MS/MS et des limites de détection) et une plage d'intensité limitée sur laquelle des données de masse précises peuvent être acquises. Ces faiblesses sont balayées par l'introduction d'une étape de piégeage des ions intermédiaire, pour accumuler les ions avant analyse.

Pour rappel, le principe de mesure de masse sur un FT-ICR, dont s'est inspiré le créateur de l'Orbitrap, est basé sur le confinement des ions dans un piège de Penning<sup>173,174</sup>. Les ions sont piégés et entrent en résonance cyclotron dans un champ magnétique uniforme constant relativement intense (7 à 15 T majoritairement) combiné à un champ électrique faible, le courant ionique généré est détecté, enregistré en fonction du temps et converti par transformée de Fourier pour en extraire l'information de fréquence et ainsi des rapport  $m/z$  des ions constituant les espèces ioniques présentes. Bien que le FT-ICR possède une résolution et une précision de masse particulièrement élevées, la technologie nécessaire (aimant supra-conducteur à large entrefer) est plus complexe d'utilisation et demande un grand coût de maintenance, il s'est donc peu démocratisé. Tout en conservant la précision obtenue grâce à la transformée de Fourier, de nouvelles approches d'analyse ont été développées, et c'est le piégeage de type orbital, le piège de Kingdon<sup>175</sup>, qui a donné naissance à l'Orbitrap. La géométrie de ce piège, modifiée par rapport au modèle initial, est composée de deux électrodes externes en forme de barillet et d'une électrode centrale de forme fuselée le long de l'axe (Figure 9).

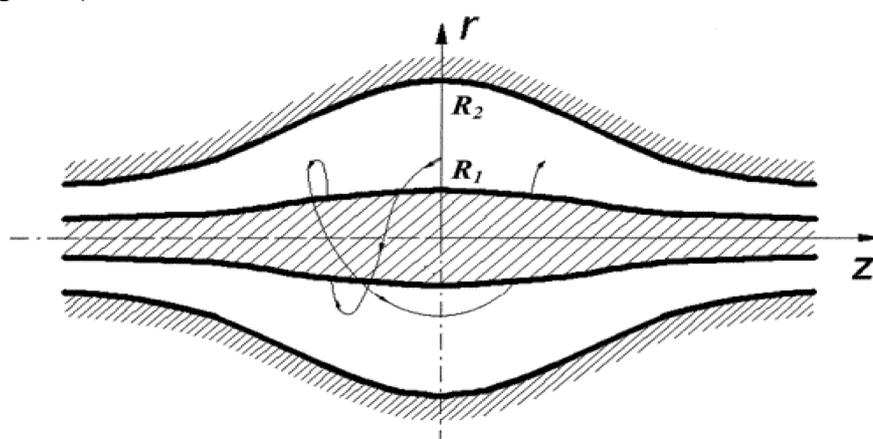


Figure 9 : Géométrie de l'Orbitrap où  $R_1$  = rayon de l'électrode centrale et  $R_2$ =rayon de l'électrode externe (Makarov,2000<sup>176</sup>)

Les électrodes sont connectées de façon indépendante, ce qui permet de faire varier leur potentiel électrique pour créer un champ électrique. Ce champ électrique est nécessaire pour mettre en orbite les ions entrants, il doit rester stable tout le long de l'analyse. Une fois piégés,

les ions oscillent autour de l'électrode centrale et le courant induit est mesuré. La fréquence d'oscillation des ions le long de l'axe de l'électrode centrale est inversement proportionnelle à la racine carrée du rapport  $m/z$  des ions. D'autres oscillations de la trajectoire sont observées mais sont indépendantes de ce rapport. Le courant induit sur les électrodes externes peut donc être enregistré et le rapport  $m/z$  sera déterminé par transformée de Fourier du signal détecté <sup>172</sup>. L'Orbitrap présente ainsi des caractéristiques analogues quoique légèrement inférieures à un spectromètre FT-ICR en termes de résolution et d'exactitude de masse, mais sans la charge d'un aimant supraconducteur coûteux. Afin de garantir une bonne précision l'Orbitrap nécessite en contrepartie un étalonnage plus fréquent. Les Orbitraps seront très rapidement commercialisés par la société Thermo Electron (aujourd'hui Thermo Fisher Scientific) dès 2005 <sup>177</sup> et cette commercialisation a été suivie par de nombreuses extensions importantes de la même famille technologique. En raison des vitesses ioniques élevées et de l'absence de refroidissement collisionnel à l'intérieur de l'Orbitrap, un stockage à ions supplémentaire a dû être ajouté : la C-trap (*Curved rf-only quadrupole*) qui permet d'obtenir des paquets d'ions cohérents pour leur détection dans l'Orbitrap. Ensuite l'ajout d'une cellule de collision après la C-trap dans LTQ Orbitrap XL (2007) a ouvert la voie à l'utilisation de collisions à plus haute énergie, l'HCD pour *High Collision Dissociation*<sup>178</sup>. L'Orbitrap Velos (2009) a été équipé d'une nouvelle lentille S-lens (*Stacked Ring Ion Guide*), constituée d'anneaux juxtaposés, ce qui a pour conséquence d'augmenter l'efficacité de transfert d'un facteur 10 en mode MS / MS et de 3 à 5 pour un spectre de masse à balayage complet. Un double piège à ions permet d'augmenter la fréquence d'acquisition en mode MS / MS<sup>179</sup>. Les instruments LTQ Orbitrap représentent une combinaison de pièges à plusieurs étages. En mode MS, le piège linéaire remplit la fonction de collecter la population d'ions, de les transmettre à la C-trap intermédiaire pour injection et analyse dans l'analyseur Orbitrap à haute résolution. En mode MS / MS, le piège à ions linéaire ne conserve qu'une fenêtre de masse choisie, qui est activée par un champ radiofréquence menant à la fragmentation des ions précurseurs piégés, et enregistre le signal d'un balayage dépendant de la masse à basse résolution et très haute fréquence d'acquisition.

### ***1.3.3.2 Le QqOrbitrap ou Q Exactive***

En 2008, un spectromètre de masse avec un analyseur Orbitrap autonome, l'Exactive, a été construit. La source d'ions était directement reliée à la C-trap, via des guides d'ions donc sans possibilité de sélection de masse, un spectromètre de masse dont l'application serait principalement pour les études des petites molécules. En 2011, une combinaison de cet appareil avec un quadripôle a été lancée sous le nom de Q Exactive. Cet appareil, qui a la capacité de sélectionner et de fragmenter les ions à très haut débit en analysant à haute résolution pendant que les ions pour l'analyse suivante sont générés dans la cellule de fragmentation en parallèle, est devenu très populaire pour la protéomique et le criblage à haut débit. En comparaison au piège à ions linéaire, les filtres de masse quadripolaires ont l'avantage d'être capables de sélectionner simultanément une gamme de masse relativement

étroite sans balayage en modulant le champ radiofréquence <sup>180</sup>, de façon très rapide et très efficace.

Le Q Exactive utilisé dans cette étude est composé d'une source nanoESI (*nanoElectroSpray Ionization*), suivi d'une succession de lentilles et guides d'ions avant d'arriver au quadripôle. La gamme de masse classique du quadripôle est de  $m/z$  50 - 4000. Les ions sont transmis via une optique spécifique (C-Trap et Z-Lens) avant d'être détectés dans l'Orbitrap. Les ions peuvent également subir des fragmentations dans la cellule de collision HCD et les ions fragments sont également détectés dans l'Orbitrap (Figure 10).

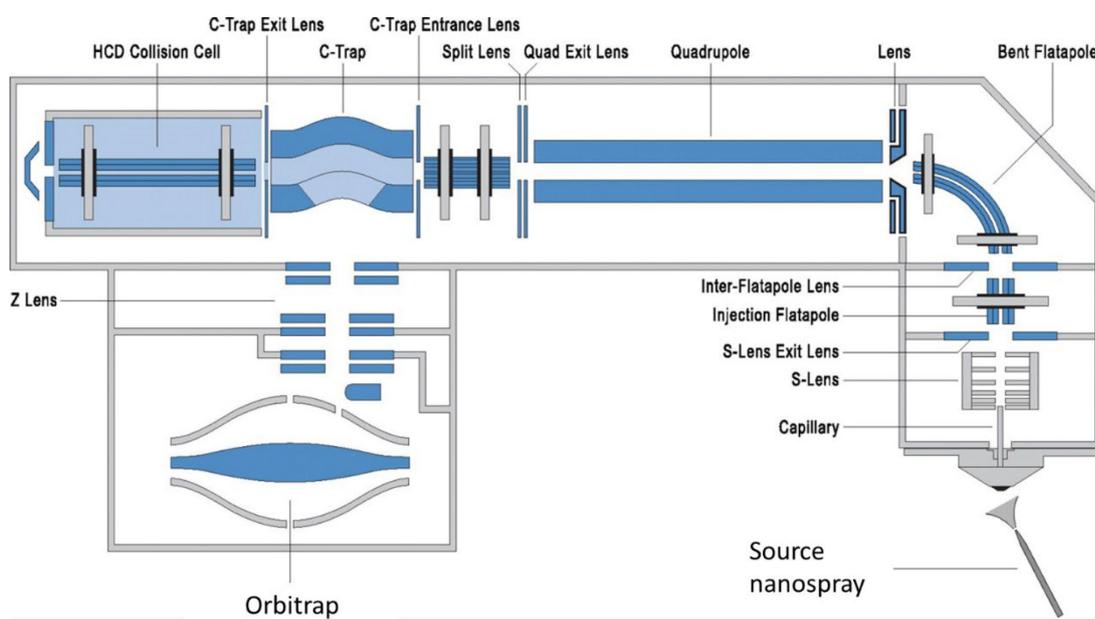


Figure 10 : Détails de l'appareil QExactive (adapté de Michalski, 2011<sup>180</sup>)

La combinaison QqOrbitrap combine deux avantages : le premier quadripôle permet de sélectionner très rapidement les ions d'intérêt car les délais d'application de tension sont très faibles et il n'y a pas de balayage, et le second quadripôle en mode HCD permet de fragmenter les peptides sur une échelle de temps similaire. De plus la séparation dans l'espace et la possibilité de travailler à haute résolution en mode MS et MS/MS ouvre le champ aux analyses en multiplexe. Il est possible de travailler selon différents modes d'acquisition grâce à cette configuration : en DDA (*Data Dependant Acquisition*), mode d'acquisition de routine en exploratoire, PRM (*Parallel Reaction Monitoring*), mode d'acquisition utilisé en quantification ciblée, qui dérive du MRM (*Multiple Reaction Monitoring*) et les multiplexages. Seul le mode d'acquisition DDA a été utilisé durant cette étude et sera développé plus loin dans ce document.

Le spectromètre de masse utilisé pendant cette étude est un QExactive amélioré, le QExactive HF (*High Field*). Cette nouvelle configuration a été introduite en 2014. Les améliorations apportées sont à trois niveaux :

- La taille de l'Orbitrap est d'abord réduite et la tension appliquée passera de 3,5 kV à 5 kV, ce qui entrainera une fréquence d'acquisition plus grande, pour une même résolution, et un taux d'identification des peptides plus élevé aussi.
- Le quadripôle est également modifié afin d'améliorer la focalisation des ions et leur transmission, même pour les m/z élevés.
- Plusieurs lentilles et guides d'ions sont ajoutés pour diminuer l'encrassement de la source.

Toutes ces améliorations vont permettre un meilleur taux d'identification des peptides et donc de leurs protéines, grâce à une grande fréquence d'acquisition, une meilleure transmission des ions mais également par un filtrage des ions indésirables<sup>181</sup>.

### *1.3.3.3 Les différents modes d'acquisition en spectrométrie de masse*

Dans ce paragraphe, chaque mode d'acquisition sera décrit selon ses applications et en énumérant ses avantages et ses inconvénients. Les spectres d'ions précurseurs (les peptides) sont nommés MS1 et les spectres de fragmentations des peptides, isolé ou non, sont nommés MS2 dans ce texte.

- Le mode DDA, pour *Data Dependant Acquisition*, données dépendantes de l'acquisition, est le mode classiquement utilisé en spectrométrie de masse protéomique. Le principe est le suivant, un sous ensemble d'ions précurseurs les plus abondants sélectionnés pendant une MS1, sont ensuite isolés individuellement et fragmentés dans une MS2. C'est cette information qui sera utilisée pour identifier le peptide fragmenté (voir paragraphe 1.3.3.4) dans les moteurs de recherches. Ce cycle est répété pendant toute la durée du gradient de la chromatographie liquide en amont. La plupart des instruments aujourd'hui sont capables d'effectuer un cycle DDA complet, comprenant une MS1 et 20 MS2 en 10 s. Cette méthode permet ainsi d'identifier des dizaines de milliers de peptides et de milliers de protéines. Les inconvénients de cette méthode sont principalement sa faible reproductibilité et dans le cas de co-élution de nombreux ions, sa conception empêche la mesure des peptides de faible abondance. En effet, la reproductibilité entre deux analyses d'un même échantillon, injecté l'un à la suite de l'autre dans exactement les mêmes conditions (instruments, solvant, température) n'est que de 35-60 % au niveau peptidique. Ces variabilités, qui impactent plus les peptides que les protéines au final, sont dues notamment la vitesse d'acquisition du spectromètre de masse qui n'a pas le temps de sélectionner et fragmenter tous les peptides présents dans une MS1<sup>182</sup>. Cependant, cela n'empêche plus la quantification sans marquage depuis l'implantation des

méthodes de type *Match Between Run* (MBR), qui permet un transfert d'information de séquence d'une analyse à l'autre (voir paragraphe I3.4.2.2). Son principal avantage, est sa simplicité de mise en œuvre, elle peut être réalisée sur n'importe quel spectromètre de masse dédié à la protéomique et elle ne demande aucune donnée au préalable en comparaison avec les autres méthodes. Elle est d'ailleurs la base nécessaire à la construction de ces méthodes.

- Les modes de type MRM/PRM (*Multiple Reaction Monitoring/Parallel Reaction Monitoring*) sont des méthodes dites ciblées. Elles se focalisent sur des peptides dont la fragmentation est connue. Elles sélectionnent uniquement les précurseurs et leurs fragments à un temps de rétention déterminé. Les informations de masse  $m/z$  du peptide, de ses fragments et du temps rétention sont obtenues lors d'une acquisition en mode DDA au préalable. Le mode MRM est une technique sans balayage, réalisée sur des instruments à triple quadripôle (QQQ) dans lesquels la fragmentation est utilisée comme un moyen d'augmenter la sélectivité. En PRM, le troisième quadripôle est remplacé par un analyseur de type Orbitrap, ce qui permet d'avoir une haute résolution et une grande précision de mesure des fragments.

Dans le mode MRM<sup>183</sup>, pour chaque spectre MS1, seul un sous-ensemble d'ions fragments est mesuré dans le spectre MS2 suivant, alors qu'en mode PRM<sup>184</sup>, tous les ions fragments sont monitorés en parallèle. Ces méthodes d'acquisition sont utilisées pour les quantifications et sont extrêmement sensibles et reproductibles. Dans les deux cas, les mêmes précurseurs sont sélectionnés et fragmentés plusieurs fois pour acquérir une quantification plus précise.

Le principal inconvénient est la mise en place de la méthode. En effet, il faut dans un premier temps réaliser une ou plusieurs analyses en mode DDA afin d'enregistrer les informations protéiques, peptidiques et le temps de rétention. Ensuite il faut choisir quelques peptides pour représenter la protéine d'intérêt en tenant compte de différents facteurs : le peptide doit être unique, non modifié (et non modifiable artificiellement) et doit avoir une observabilité certaine c'est-à-dire qu'il doit systématiquement être identifiée lors de la présence de la protéine d'intérêt. Après un choix judicieux des peptides, en mode MRM, il faut également sélectionner les transitions pour chaque peptide, ce qui consiste à choisir les fragments les plus intenses, en PRM comme tous les ions peuvent être suivie en parallèle, cette étape n'est plus nécessaire.

- Le mode DIA, pour *Data Independant Acquisition*, données indépendantes de l'acquisition, est un mode de plus en plus utilisé en protéomique quantitative, bien que les premiers essais datent de 2004<sup>185,186</sup>. Le principe est simple, il n'y a pas de sélection de précurseur ni d'analyses séquentielle d'ions individuels, le DIA fragmente tous les ions détectables dans une gamme de  $m/z$  limitée, quelle que soit leur intensité. Ensuite ce spectre MS2, dans lequel tous les fragments de tous les peptides présents dans la

gamme de masse choisie sont mélangés, est confronté à une banque de spectre de référence, généré au préalable à partir d'acquisition en mode DDA. La principale difficulté réside dans le traitement de données qui doit lier les fragments au bon précurseur.

### *1.3.3.4 Identification des protéines et moteurs de recherche*

Une analyse en spectrométrie de masse exploratoire va générer des milliers, voire des dizaines de milliers, de spectres de fragmentation des ions précurseurs en mode DDA. C'est l'information contenue dans ces spectres qui permettra de remonter à l'information protéique. La stratégie d'identification consiste à confronter les données de l'expérience aux données théorique à l'aide de moteurs de recherches auxquels seront transmises les informations de l'expérience, afin de restreindre le champ des possibilités. Il existe cependant d'autres méthodes qui seront décrites succinctement dans les paragraphes suivants.

#### *1.3.3.4.1 Comment se fragmentent les peptides et comment interpréter ces spectres MS/MS de fragmentation ?*

Un peptide est un polymère d'acides aminés, reliés entre eux par des liaisons amides ( $\text{—CO—NH—}$ ). Il comporte une extrémité N-terminale (groupe  $\text{—NH}_2$ ) et une extrémité C-terminale (groupe  $\text{—COOH}$ ). Un acide aminé est constitué d'une fonction amine, d'un groupe carboxyle et d'un groupe caractéristique variable R formant ainsi un répertoire de 20 acides aminés. Le modèle le plus complet actuellement disponible pour décrire comment les peptides protonés se dissocient, est nommé le modèle du « proton mobile ». Pendant l'ionisation, des espèces chargées sont formées, dans lesquelles les protons sont initialement localisés sur les fonctions basiques (amines) dans le cas d'une digestion tryptique, comme c'est le cas en phase liquide. En phase gazeuse, les molécules de gaz vont activer les peptides et les exciter. Les protons alors vont se déplacer vers les sites de plus grande affinité protonique le long des liaisons peptidiques, c'est à dire vers les liaisons amides et amines. Or la liaison amide est une liaison stabilisée par la délocalisation de charge entre les doublets de l'azote et ceux de la fonction carbonyle. La protonation de cette liaison va entraîner l'abolition de cet équilibre et fragiliser la liaison amide jusqu'à sa rupture. La fragmentation entraîne la formation d'ions fragment de type y (charge sur la chaîne latérale de l'acide aminé en C-terminal) et d'ions fragment de type b (charge sur la chaîne latérale de l'acide aminé en N-terminal)<sup>187</sup>

Même si la prédictibilité de la fragmentation des peptides n'est jamais totale, elle est très bien étudiée et a pu être même nomenclaturée. Le séquençage et la nomenclature sont le résultat des travaux de Roepstorff & Fohlmann<sup>188</sup> et Biemann<sup>189</sup> dans les années 80.

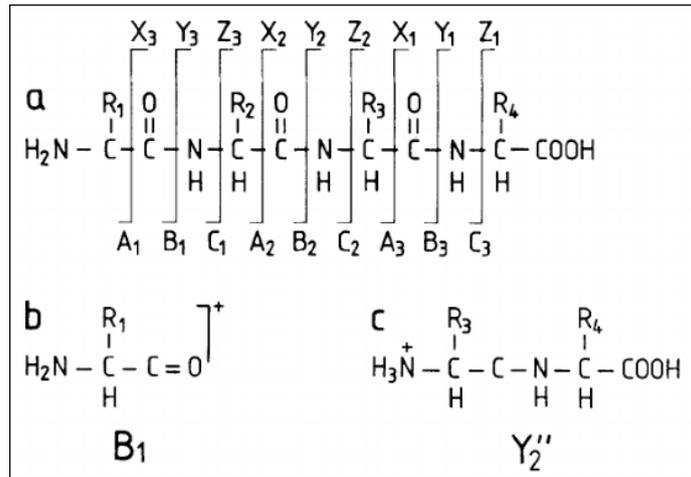


Figure 11 : Schéma de fragmentation original de Roepstorff et Fohlmann 1984<sup>188</sup>

Dans la Figure 11, sont représentés les différents fragments résultant de ruptures le long du squelette peptidique. L'indice accolé en bas à droite de chaque lettre indique le nombre de résidus contenus dans le fragment. Les ions fragments les plus fréquemment observés dans les peptides tryptiques par fragmentation de basse énergie activée par collision avec un gaz neutre sont de type y et b : il s'agit de fragments générés par la rupture des liaisons amides du peptide, et qui contiennent respectivement l'extrémité C- ou N-terminale du précurseur. Les annotations des fragments ont été modifiées par Biemann en 1988<sup>189</sup>, notamment le passage des majuscules aux minuscules, pour éviter des conflits avec d'autres nomenclatures. D'autres fragments sont possibles, comme les ruptures avant la liaison amide (fragments a et x) ou après la liaison amide (fragments c et z), ou les pertes de chaîne latérale ou des fragments internes, mais ils ne sont présent que lors de combinaisons d'acides aminés particuliers ou à l'utilisation d'un mode de fragmentation autre que le classique CID (Collision-Induced Dissociation) ou HCD (High Collision Dissociation)<sup>190</sup>, notamment les modes ETD (Electron Transfer Dissociation) et ECD (Electron-Capture Dissociation)<sup>191</sup>.

#### I.3.3.4.2 Les moteurs de recherches dans les banques de données protéiques

L'essor technologique de la spectrométrie de masse a nécessité très rapidement le développement de stratégies en bio-informatique dédiées pour le traitement des données et notamment en protéomique. En effet, la quantité et la complexité des données générées peuvent rapidement devenir un facteur limitant de l'analyse. Un grand nombre d'algorithmes et de programmes informatiques ont été développés afin d'identifier des peptides et des protéines à partir de données MS/MS (Figure 12). Il existe cependant plusieurs approches pour identifier des peptides et leur protéine à partir des spectres de fragmentation.

- Recherche en banque de données protéiques : après digestion *in silico* de la banque protéique, le logiciel liste toutes les séquences peptidiques dont la masse est proche, dépendant de l'erreur de mesure de l'appareil, et comme les règles de fragmentation peptidique sont relativement bien décrites, il est capable de construire pour chaque peptide potentiel un modèle représentant sa fragmentation théorique qui sera comparée à la fragmentation expérimentale. C'est l'approche la plus classique utilisée aujourd'hui <sup>192</sup>. Il existe plusieurs dizaines de logiciels et donc d'algorithmes qui effectuent ce type d'identification, leur fonctionnement est un peu près similaire, leur différence se situe surtout au niveau de la validation de l'identification du peptide, corrélée avec celle de la protéine. Cette validation est représentée par le score attribué au(x) spectre(s). Les algorithmes d'identification possèdent une fonction permettant de calculer les scores d'identification des peptides, et par extension de protéines, qui peut être séparées en deux catégories :
  - a) Les algorithmes heuristiques qui travaillent uniquement sur la corrélation entre le spectre expérimental et le spectre théorique. Le score est alors calculé à partir de la similarité entre les deux spectres, comme par exemple dans les logiciels Sequest<sup>193</sup> ou XTandem<sup>194</sup>.
  - b) Les algorithmes probabilistes où la fonction déterminant le score prend en considération, en plus de la similarité, la probabilité que la séquence peptidique identifiée puisse être obtenue par hasard. En effet, chaque peptide est pondéré d'une manière statistique. Mascot<sup>195</sup> et Andromeda<sup>196</sup> (moteur de recherche de la suite logiciel MaxQuant) utilisent ce type d'algorithme.

L'approche définie par les concepteurs pour calculer le score étant différente, chaque programme aura ses forces et ses faiblesses lors de la détermination d'une séquence d'acides aminés par rapport à un spectre de masse MS/MS. Par conséquent, il n'est pas rare d'observer des variations dans l'identification des peptides.

- Recherche dans les banques spectrales : une comparaison est directement effectuée entre les spectres MS/MS bruts, expérimentaux et des spectres MS/MS stockés dans une bibliothèque spectrale. Cette méthode est bien plus performante sur le plan de la vitesse, de la sensibilité et du taux d'erreur que la recherche en banque de données mais nécessite un stockage au préalable de l'analyse spectrale de chaque peptide.
- Sequence tag : un séquençage partiel du spectre MS/MS avec l'information de masse du précurseur va permettre d'identifier le peptide dans la banque protéique. Méthode très peu utilisée aujourd'hui et qui nécessite toujours une banque de données protéiques.
- Séquençage de Novo : les logiciels de séquençage de Novo identifient la séquence complète du peptide à partir du spectre MS/MS, ce sont des outils particulièrement intéressants quand il n'existe pas de banques protéiques pour les protéines analysées. Il suffit ensuite de comparer par alignement en utilisant des algorithmes de type BLAST (*Basic Local Alignment Search Tool*)<sup>197</sup> les séquences proposées de Novo avec les séquences des banques d'organismes proches et séquencés. Mais cela nécessite des spectres MS/MS d'excellente qualité et un temps de calcul important.

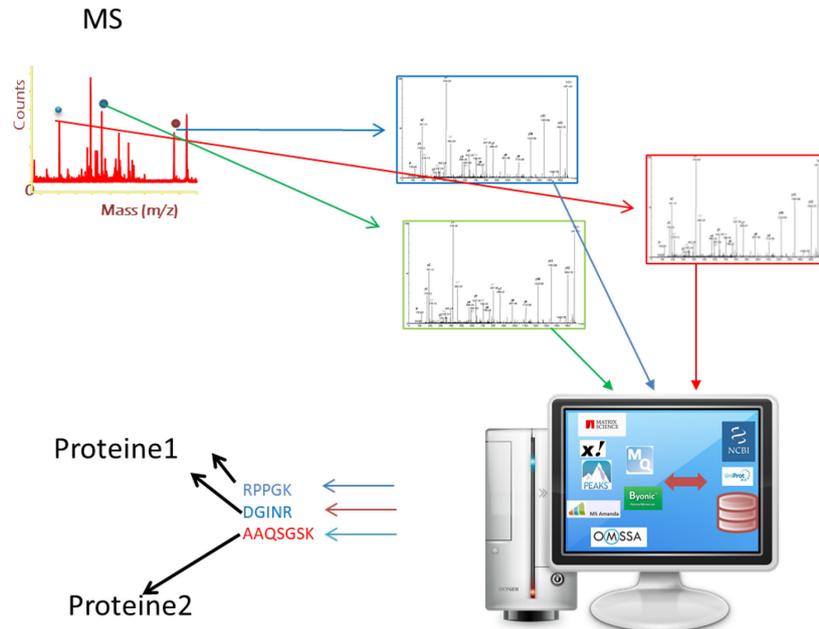


Figure 12 : De la spectrométrie de masse à l'identification protéique

#### I.3.3.4.2.1 Méthodes statistiques pour la validation des résultats d'identification

L'automatisation et l'analyse *in silico* des spectres MS/MS présente de nombreuses sources d'erreurs. La qualité des spectres MS/MS, la pertinence des paramètres de recherche, l'exhaustivité de la banque de séquences protéiques sont des facteurs qui peuvent fausser l'identification des protéines de l'échantillon. La validation des résultats d'analyse protéomique est donc une étape importante sur laquelle de nombreux chercheurs se sont penchés afin de minimiser les erreurs, que l'on appelle les faux positifs <sup>198</sup> (ou les faux négatifs).

Des directives pour la publication de données sur l'identification de peptides et de protéines sont désormais définies avec notamment les « Proteomics Standard Initiative », mise en place depuis 2008 avec l'organisation HUPO (HUman Proteomic Organization), plus de 100 publications sont à disposition pour cela, aussi bien au niveau analyse en spectrométrie de masse, que pour la validation ou pour le retraitement des données. D'une manière générale, les chercheurs doivent maintenant justifier leurs résultats en apportant des informations sur les méthodes, les seuils et les moteurs de recherche utilisés. Les principaux problèmes sont la présence d'un nombre plus ou moins élevé de faux positifs dans les résultats fournis, et le fait que chaque logiciel possède son propre algorithme de score. Etant donné l'absence de standards, il était important de mettre en place des normes pour assurer la qualité des informations publiées <sup>199</sup>.

Si l'on considère le nombre très élevé de spectres MS/MS pouvant être acquis lors d'une analyse protéomique de type « *shotgun* », il devient alors impossible d'imaginer une validation manuelle des assignements peptidiques obtenus. Les méthodes statistiques, qui ont été développées de manière intensive ces dernières années, sont aujourd'hui capables d'estimer des taux de faux positifs présents dans les jeux de données fournis par les moteurs de recherche.

Deux informations peuvent être estimées, le taux de faux positifs FDR (*False Discovery Rate* en anglais) et la probabilité postérieure d'erreur PEP (*Posterior Error Probability*). Le FDR est une estimation globale du taux de faux positifs de la recherche alors que le PEP, attribué à chaque peptide, donne la probabilité que celui-ci soit effectivement présent, ce sont deux valeurs liées et complémentaires<sup>200</sup>. Dans la figure 13, si C représente la distribution des attributions incorrectes et A les attributions correctes, alors le FDR est le rapport de B, le nombre d'identifications incorrectes avec au seuil x choisi, à ( A + B ), le nombre total d'identifications totales au seuil x. Le PEP, en revanche, est un rapport des hauteurs correspondantes de la distribution c'est-à-dire le nombre b d'identifications incorrects avec un score de x divisé par le nombre total ( a + b ) d'identification avec un score de x.

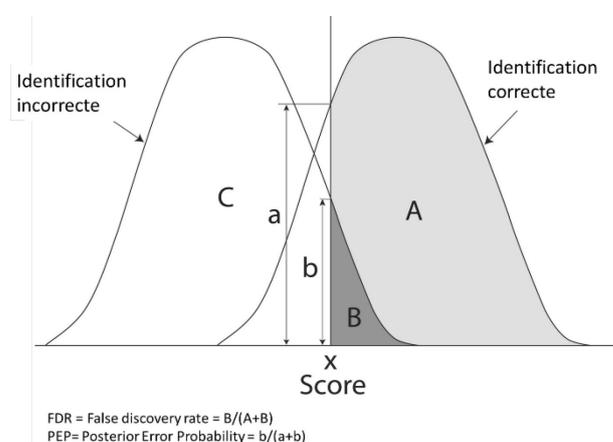


Figure 13 : Distribution des scores des identifications correctes et incorrectes (Käll, 2008<sup>200</sup>)

### I.3.4 La quantification relative en protéomique par spectrométrie de masse

La spectrométrie de masse n'est pas, de par son mode de fonctionnement, intrinsèquement quantitative. En effet, l'efficacité d'ionisation et la réponse de détection dépendent des propriétés physico-chimiques des peptides (charge, hydrophobicité, ...). Ainsi la réponse en spectrométrie de masse diffère d'un peptide à un autre, rendant la technique non quantitative entre molécules différentes. Pour une quantification précise, il est donc généralement nécessaire de comparer chaque peptide individuel entre les expériences. Il existe aujourd'hui différentes méthodes de quantification des protéines avec ou sans marquage, possédant chacune des avantages et des inconvénients. Ce sous-chapitre n'abordera pas la quantification absolue.

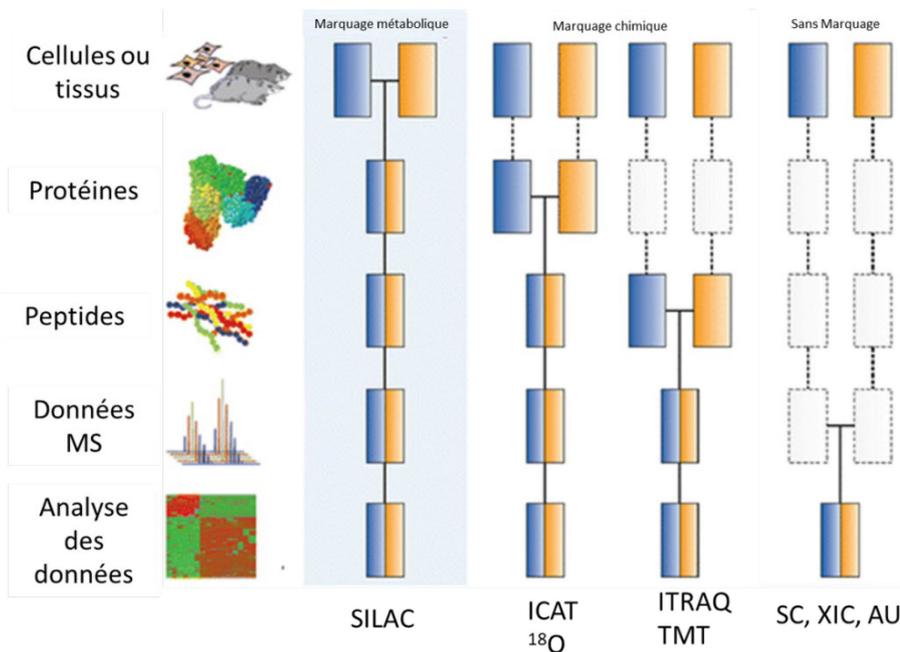


Figure 14 : Les différents flux de travail selon la stratégie de quantification utilisée, de l'échantillon à l'interprétation (adapté de Bantscheff 2012<sup>201</sup>)

Sur la figure 14, les différentes stratégies de quantifications sont représentées schématiquement, les rectangles jaunes et bleus symbolisent 2 conditions expérimentales différentes, les lignes horizontales lorsque les échantillons sont combinés et les lignes en pointillé représentent les étapes sur lesquelles des variations expérimentales peuvent apparaître.

#### I.3.4.1 Quantification avec marquage

Les premiers marquages isotopiques stables en protéomique arrivent en 1999<sup>202-204</sup>, ils peuvent être introduits comme étalon interne dans les acides aminés métaboliquement<sup>205,206</sup> ou chimiquement<sup>203,207,208</sup> et même de manière enzymatique<sup>209</sup>. La plupart de ces techniques de marquage, la quantification relative est obtenue par l'intégration du signal MS des peptides marqués « lourds » versus les peptides « légers » dans les spectres.

Le **marquage métabolique** a l'indéniable avantage d'être présent au début de l'expérience, puisque les marqueurs sont introduits pendant la culture cellulaire. Le mélange des deux échantillons étant réalisé très tôt dans la procédure, toutes les erreurs dues à la manipulation, les conditions opératoires de la chromatographie ou la mesure de masse, se répercutent de la même manière sur les deux conditions. Cependant, ce marquage ne peut être réalisé que sur certaines cultures cellulaires<sup>210</sup>, bien que plusieurs tentatives, coûteuses à mettre en place et à maintenir, aient été réalisées sur des animaux<sup>211,212</sup> ou des plantes<sup>213</sup>. L'incorporation des acides aminés par les cellules nécessite également un certain temps.

Le **marquage chimique** peut être réalisé au niveau protéique pour les méthodes ICAT (*Isotope-Coded Affinity Tags*)<sup>203</sup> ou le marquage à l'<sup>18</sup>O, initié par l'enzyme de digestion<sup>209</sup>, ou au niveau peptidique, pour les méthodes iTRAQ (*isobaric Tags for Relative and Absolute Quantitation*)<sup>207</sup> et TMT (*Tandem Mass Tags*)<sup>208</sup>. Ces dernières méthodes, iTRAQ et TMT sont basées sur un concept différent pour la quantification qui sera réalisée non pas en MS mais en MS/MS. En effet, les peptides sont marqués de manière isobare et co-éluent lors des séparations chromatographiques. Ils sont indiscernables en MS, mais lors de la fragmentation peptidique, apparaîtront des ions rapporteurs spécifiques pour chaque condition de marquage. Certaines méthodes ont la capacité d'analyser jusqu'à 10 conditions simultanément. Néanmoins, si le marquage chimique au niveau protéique a l'avantage d'être moins biaisé dans le flux de l'expérience, puisque les protéines sont mélangées avant la digestion, il est primordial de savoir que la trypsine ne clive pas les lysines modifiées, ce qui a pour conséquence des peptides plus longs, plus difficiles à identifier<sup>214,215</sup>. Le marquage à l'<sup>18</sup>O, réalisé à l'aide de la trypsine<sup>216</sup> ou suite à la digestion avec une différence de m/z de 2 par atome d'oxygène, a malheureusement un marquage incomplet, les taux d'incorporation diffèrent en fonction des peptides, ce qui complique considérablement la quantification des protéines<sup>217</sup>. Un autre inconvénient récurrent du marquage est qu'il peut souvent mener à des produits secondaires préjudiciables et dont l'abondance est non négligeable pour une quantification<sup>201,218,219</sup>.

#### ***1.3.4.2 Quantification sans marquage (« Label Free »)***

Les méthodes de quantification sans marquage, ou *label free*, sont une alternative simple et économique aux méthodes de quantification avec marquage. En effet, elles ne nécessitent pas d'étapes laborieuses, généralement coûteuses en réactifs. De plus, le nombre de conditions à comparer est illimité, la plage dynamique de détection des peptides est plus grande et la complexité du mélange n'est pas augmentée, ce qui facilite amplement le traitement des données. La contrepartie de ces bénéfices est la simplification maximale, dans la mesure du possible, afin d'avoir reproductibilité optimale de chaque étape. De plus, les échantillons issus des différentes conditions étant analysés séparément, cela augmente le temps d'analyse. L'évaluation de la reproductibilité analytique, de la reproductibilité technique globale, ainsi que de la reproductibilité biologique est nécessaire avant que les analyses réalisées et leur

quantification puissent être interprétées de manière fiable. Deux méthodes de quantification label free sont actuellement utilisées en protéomique, une méthode basée sur les données d'identification à partir des spectres MS/MS, le « *spectral count* » (**SC**) et une méthode basée sur l'intensité du signal des peptides, le **XIC** (*eXtracted Ion Chromatogram*), à partir des spectres MS.

Le **SC** utilise les données de MS/MS pour quantifier l'abondance d'une protéine et son principe est basé sur l'observation que les protéines dont l'abondance est élevée, auront un nombre de peptides protéolytiques détecté plus grand, et par conséquent un nombre de spectres MS/MS plus élevé que les protéines de faible abondance<sup>220-222</sup>. Son mode de fonctionnement est sa principale limitation, en effet, ne seront pris en compte que les peptides fragmentés lors de l'analyse, avec le risque non négligeable d'exclure les protéines de faible abondance ou celle dont la masse ne dépasse pas 20kDa, puisqu'elles génèrent moins de peptides. En outre, cette stratégie exige un grand nombre de spectres MS/MS pour calculer et valider une variation d'abondance entre deux conditions. Et l'utilisation usuelle et nécessaire de l'exclusion dynamique<sup>223</sup>, option activable lors d'une acquisition qui va permettre « suspendre » la fragmentation des peptides abondants au profit des peptides coélus moins abondants, a pour conséquence d'augmenter les défauts de cette méthode, notamment sur sa précision de quantification pour les protéines peu abondantes et/ou de petite masse<sup>224</sup>. Idéalement les peptides devraient arriver à la source d'ions à une vitesse compatible à celle d'acquisition des données MS / MS afin que tous les précurseurs puissent être fragmentés, mais cela est rarement le cas dans le cadre de mélanges complexes.

L'autre approche classique de quantification label free repose sur le **XIC**, stratégie rendue possible par les progrès de la bio-informatique et les améliorations en termes de résolution et de précision de l'instrumentation de la spectrométrie de masse. Chaque peptide est caractérisé par sa masse exacte, son temps de rétention et son profil de fragmentation. L'intensité de son profil isotopique donne une évaluation de son abondance relative. Les acquisitions à comparer sont alignées à l'aide de la mesure de masse exacte et de l'identification des peptides comme signature spécifique pour corriger la dérive du temps de rétention. Elles sont également normalisées pour réduire le biais induit par la variation de l'échantillon injecté. Enfin, les intensités relatives entre analyses, après alignement et normalisation, révèlent s'il existe un changement relatif dans l'abondance d'un peptide, entre les échantillons<sup>225</sup>. Les variations d'abondance des protéines entre les échantillons peuvent ensuite être déduites par la somme, la moyenne ou la médiane des différences mesurées pour tous les peptides correspondant à la protéine d'intérêt<sup>226</sup>. La difficulté de cette méthode est que les XIC individuels puissent être attribués sans équivoque aux peptides respectifs. À cette fin, l'utilisation de spectromètres de masse avec un pouvoir de résolution élevé, ce qui sous-entend également l'utilisation d'une chromatographie robuste, et une précision de masse élevée est avantageuse car le premier minimise l'influence des signaux interférents provenant

de peptides co-élus de masse similaire ou quasi isobares et le dernier assure une cartographie fiable des XIC aux peptides identifiés. Le sujet du traitement des données obtenues après les analyses, est également un point critique de l'expérimentation. S'il existe aujourd'hui plusieurs algorithmes disponibles pour effectuer l'extraction des XIC, pour aligner les temps de rétention ou pour normaliser l'intensité du signal, il est important cependant de sélectionner soigneusement les *workflows* les plus adaptés à chaque conception expérimentale<sup>227,228</sup>. Les paragraphes suivants aborderont les différents traitements à appliquer sur les résultats obtenus lors de quantification, quel que soit le domaine d'études (génomique, métabolomique), mais appliqué particulièrement ici à la quantification de type *label free* basée sur le XIC en protéomique.

#### I.3.4.2.1 Normalisation des données

Pour effectuer la quantification et parce que les biais systématiques ne peuvent être évités à chaque étape du traitement des échantillons, il est essentiel de normaliser les données expérimentales. Ne pas tenir compte de ces biais pourrait conduire à des conclusions erronées concernant l'analyse semi quantitative. De multiples méthodes de normalisation, avec des stratégies différentes, ont été proposées. L'instabilité de l'acquisition peut être diminuée par exemple en ajoutant des standards peptidiques internes comme références avec une concentration connue dans chaque échantillon et en normalisant les intensités de signal sur ces références internes connues. Il existe également différentes méthodes sans ajout de standard interne. Beaucoup de méthodes de normalisation utilisées aujourd'hui en protéomique proviennent de la génomique, notamment pour l'étude des puces à ADN<sup>229,230</sup>. Classiquement il y a les méthodes globales ou à tendances centrales, telles que le centrage autour d'une référence commune comme la moyenne, la médiane ou les étalons internes enrichis, ou les « gènes domestiques », méthodes suffisantes pour éliminer les biais linéaires techniques, telle que la variation dans le volume d'injection de l'échantillon avant l'analyse, cependant elles ne prennent pas en compte les biais non linéaires, tels que les effets de suppression d'ion ou la saturation d'un signal peptidique trop intense. Ces techniques de normalisation centrent les ratios d'abondance des peptides ou protéines autour donc d'une constante telle que la médiane, la moyenne ou l'intensité totale. Il existe également diverses méthodes de régressions linéaires, basé sur l'hypothèse que le biais systématique dépend linéairement de la grandeur des pics peptidiques, à mesure que l'intensité protéique mesurée augmente, le biais augmente également. A contrario, il y a également des méthodes qui supposent que justement le biais est non linéairement dépendant de cette grandeur, en tenant compte des effets de suppression d'ions ou de la saturation du signal. Une autre méthode de normalisation très répandue est la technique du quantile, dont l'hypothèse est que la distribution statistique des échantillons devrait être la même au sein d'une même condition, et par un ajustement global va transformer la distribution statistique des intensités pour qu'elles soient identiques entre deux analyses. Il existe néanmoins d'autres nombreuses

méthodes, certaines découlent des techniques citées précédemment, qui ne seront pas décrites dans ce texte mais dont voici quelques références <sup>231-235</sup>.

#### I.3.4.2.2 Traitement des données manquantes

Les valeurs manquantes sont courantes dans les grands ensembles de données de spectrométrie de masse et peuvent provenir de plusieurs causes, y compris des raisons techniques et biologiques. Actuellement, on en sait peu sur leurs origines et leurs distributions dans les ensembles de données pour les empêcher. Cependant, le flux de traitement des données pour obtenir des résultats précis et les plus complets possible devrait en tenir compte et inclure un moyen optimal de leur attribuer des valeurs avant l'interprétation des résultats. Des valeurs manquantes surviennent lorsque, par exemple, un peptide est identifié dans certains échantillons mais pas dans d'autres. Pour les échantillons dans lesquels le peptide n'a pas été identifié, les abondances ne sont pas attribuées ou se voient attribuer une valeur NAN (*Not A Number* = pas un nombre). Un peptide peut manquer parce qu'il n'est pas présent dans l'échantillon, ou parce qu'il est présent mais à une concentration inférieure à la limite de détection de l'instrument, ou parce qu'il est présent et n'est pas identifié correctement ou parce qu'il n'est pas détecté en raison d'un effet de suppression spectrale <sup>231</sup>.

L'un des avantages à l'utilisation de la quantification en *label free* basée sur le XIC permet de réduire ce nombre de valeurs manquantes en transférant l'identification d'une analyse à une autre. L'algorithme « **Match Between Run** » MBR de la suite logicielle MaxQuant <sup>236</sup>, l'un des premiers dans le domaine de la protéomique quantitative, évalue chaque pic identifié dans un spectre MS à partir d'une analyse LC – MS / MS et compare son temps de rétention à des pics non identifiés dans une autre. Une identification est transférée si un pic non identifié, a le même m/z, le même état de charge et le même temps de rétention RT. Après l'utilisation du MBR, il reste néanmoins des données manquantes qui peuvent être catégorisé en trois classes: la «*Missing Not At Random*» (MNAR), manquante mais non aléatoire, est causée par la faible abondance et la limite de détection du spectromètre de masse, la «*Missing Completely At Random*» (MCAR), manquante complètement au hasard, est causée par le hasard les erreurs d'expérience ou les fluctuations lors du processus d'acquisition des données et la «*Missing At Random*» (MAR), manquante par hasard, est causée par une erreur dans le traitement des données, considérée comme corrigible, telle qu'une détection de pic imprécise <sup>237</sup>. Les données manquantes peuvent être soit imputées, en fonction de leur catégorie, soit être invalidées, mais ce dernier cas pourrait rejeter des protéines potentiellement intéressantes. Pour une imputation, selon le type de donnée manquante, le traitement doit être adapté, il existe notamment une « règle » dite des 80% <sup>238</sup>. Trois classes d'identification sont possibles avec cette règle : la protéine est identifiée dans au moins 80% des essais (MAR ou MCAR), elle est considérée comme présente, la protéine est supprimée car elle a trop de données manquantes (MNAR) dans toutes les conditions (identifiée dans <80% dans toutes

les conditions), et le cas où la protéine est considérée comme absente dans l'une des conditions ( $<20\%$  dans l'une des conditions,  $\geq 80\%$  dans l'autre) (Figure15). Si l'imputation n'est pas obligatoire pour une comparaison entre plusieurs conditions, elle peut être indiquée pour certaines analyses en aval telles que les analyses en composantes principales (ACP). Comme pour la normalisation, il existe un large choix dans les méthodes du calcul d'imputation, comme la moyenne ou la valeur observé la plus basse, qui sont à éviter impérativement, car imputer la même valeur plusieurs fois diminue la variabilité biologique, des approches de similarités locale en se basant sur le profil d'expression de plusieurs autres peptides ayant des profils semblable et les approches basées sur la structure globale qui appliquent des techniques de réduction de dimension pour décomposer la matrice de données, puis reconstruire de manière itérative les valeurs manquantes<sup>239,240</sup>.

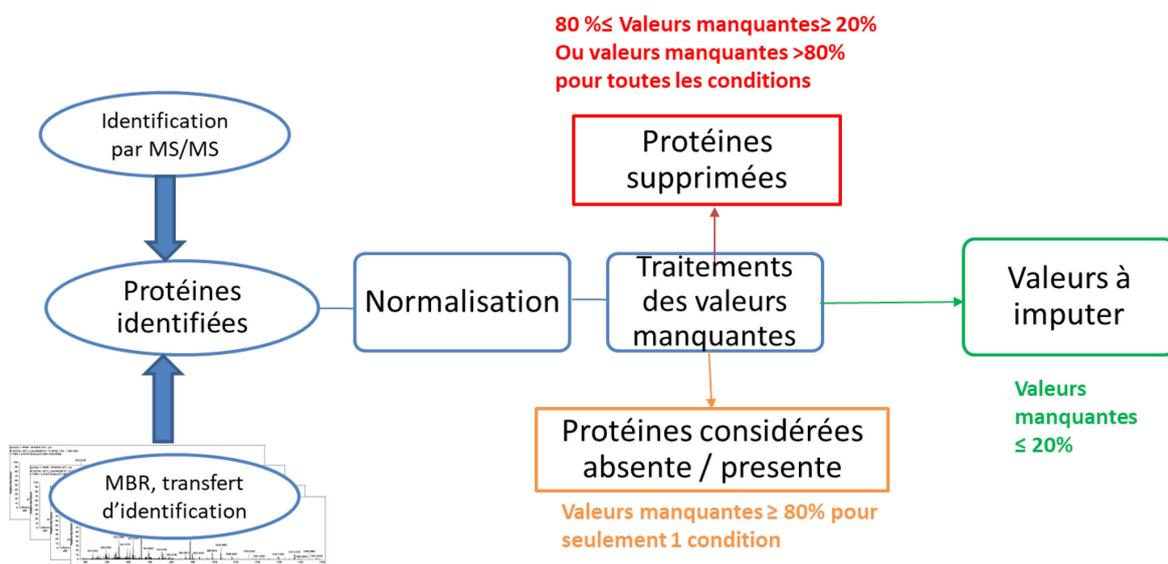


Figure 15 : Workflow des données de spectrométrie de masse de la normalisation, du traitement des données manquantes à l'imputation.



## II. Objectifs

---

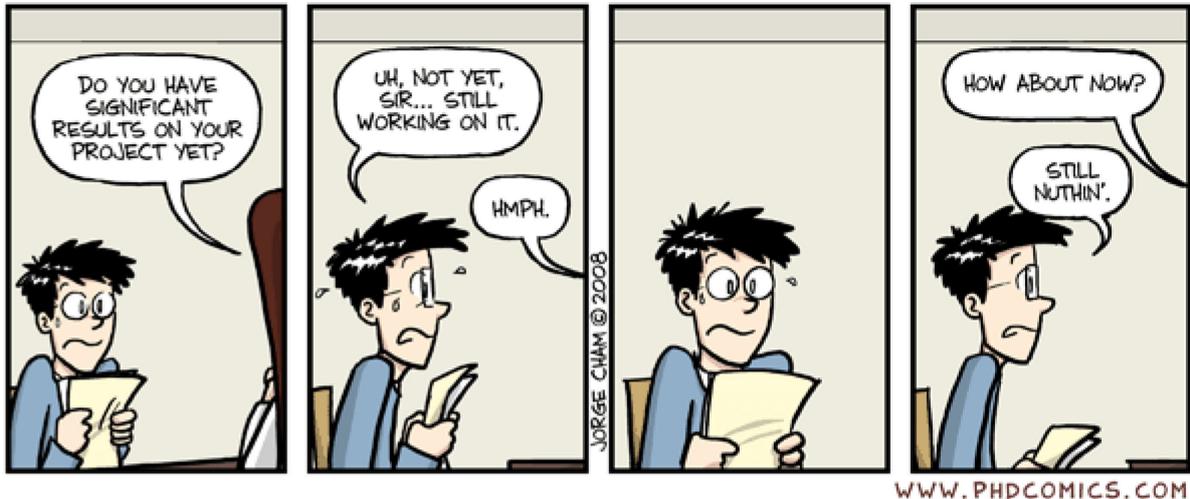
Les modifications de la MEC, associées à une dégénérescence plus ou moins sévère des cellules musculaires lisses, constituent une lésion histologique constante des vaisseaux cérébraux dans les données histopathologiques dans le cadre des cSVD. De plus, des travaux récents ont montré que les gènes mutés impliqués dans les formes génétiques de cSVDs codent pour des protéines du matrisome, comme COL4A1, COL4A2, HTRA1 ou Cathepsin A dans les maladies du collagène de type IV, CARASIL ou CARASAL respectivement, ou bien alors sont responsables de modifications pathogènes du matrisome comme NOTCH3 dans CADASIL. Enfin, la MEC occupe une position stratégique à l'interface entre les vaisseaux cérébraux et le parenchyme cérébral. L'ensemble de ces observations nous a fait émettre l'hypothèse que les perturbations du matrisome des vaisseaux cérébraux jouent un rôle clé dans les mécanismes physiopathogéniques des cSVDs, ce qui a permis une première publication durant ce projet de thèse<sup>8</sup> (Article 1 en Annexe 1).

L'objectif de cette thèse a été d'étudier les modifications du matrisome des vaisseaux cérébraux dans le contexte des cSVDs par une approche de type protéomique.

Tout d'abord, un premier travail de développement technique a été mis en place afin de traiter des échantillons de type petits vaisseaux cérébraux murin afin de caractériser et de quantifier leur matrisome de manière robuste et répétable.

Ensuite, une étude protéomique quantitative de type *label free* a été réalisée sur trois modèles génétiques de cSVDs et deux autres modèles souris mimant les facteurs risque majeurs des cSVDs, à savoir, l'âge et l'hypertension artérielle.





### III. Résultats et discussions

III.1 DEVELOPPEMENT D'UNE NOUVELLE METHODE POUR L'ETUDE DU MATRISOME DES VAISSEUX CEREBRAUX CHEZ LA SOURIS (ARTICLE2).....	92
<i>III.1.1 Introduction</i> .....	92
<i>III.1.2 Objectif</i> .....	92
<i>III.1.3 Méthode</i> .....	93
<i>III.1.4 Résultats et Discussions</i> .....	95
III.2 ÉTUDE DES CHANGEMENTS DU MATRISOME MICROVASCULAIRE DANS LES MALADIES DES PETITS VAISSEUX.....	131
<i>III.2.1 Introduction</i> .....	131
<i>III.2.2 Objectifs</i> .....	132
<i>III.2.3 Matériel et méthode</i> .....	132
III.2.3.1 Modèles souris des formes génétiques de cSVDs.....	132
III.2.3.2 Modèles souris des formes sporadiques de cSVDs.....	133
<i>III.2.4 Résultats</i> .....	134
III.2.4.1 Contrôles « qualité » dans les modèles génétiques de SVDs.....	134
III.2.4.1.1 CADASIL.....	134
III.2.4.1.2 CARASIL.....	136
III.2.4.1.3 Maladie du collagène IV.....	138
III.2.4.2 Analyse des protéines du matrisome différemment exprimés dans les modèles génétiques de maladies des petits vaisseaux cérébraux.....	139
III.2.4.3 Analyse des protéines du matrisome différemment exprimés dans les modèles sporadiques de maladies des petits vaisseaux cérébraux.....	143
III.2.4.4 Comparaison des différentes maladies et conclusions.....	147
<i>III.2.5 Discussion</i> .....	148

## **III.1 Développement d'une nouvelle méthode pour l'étude du matrisome des vaisseaux cérébraux chez la souris (Article2)**

Cette première partie de résultats, décrite dans un article qui sera prochainement soumis, résume le développement et la validation d'une nouvelle approche pour l'étude du matrisome des vaisseaux cérébraux chez la souris.

### **III.1.1 Introduction**

L'étude protéomique, du matrisome en particulier, des petits vaisseaux cérébraux nous paraît essentielle pour la compréhension des mécanismes moléculaires impliqués dans les cSVD. Il existe quelques publications dans lesquelles les vaisseaux cérébraux sont étudiés en protéomique<sup>115,241-243</sup>, mais aucune jusqu'à présent ne s'est concentrée spécifiquement sur l'étude du matrisome de ces vaisseaux.

Le matrisome de nombreux organes, tels que le cœur, le poumon ou de gros vaisseaux tels que l'aorte, les artères coronaires ou les veines ont été étudiés, chez l'homme, la souris ou le porc, sain ou malade<sup>57,132,244-246</sup>. Toutes ces études utilisent des protocoles combinant une décellularisation ou des fractionnements subcellulaires. De plus la quantité de tissu utilisé par échantillon dépasse habituellement les 100 mg.

La quantité de tissu représentée par des vaisseaux cérébraux disséqués ou isolés à partir d'une souris n'excède pas 10 mg. Cette très petite quantité de tissu n'est pas adaptée à des protocoles de décellularisation ou de fractionnements subcellulaires extensifs, tels qu'ils sont couramment utilisés pour l'analyse du matrisome.

### **III.1.2 Objectif**

L'objectif de cette première étude est de développer et de valider une nouvelle méthode de préparation et d'analyse des vaisseaux cérébraux de souris en protéomique, afin d'identifier et de quantifier de manière robuste et la plus complète possible le matrisome de ces vaisseaux.

### III.1.3 Méthode

Le matériel de départ est constitué d'artères piales d'une souris, disséquées sous loupe. Ce type de matériel est pertinent dans le cadre de l'étude des cSVDs puisque les artères piales sont affectées dans ces maladies. Un des avantages présentés par ce type de matériel est que la dissection est très reproductible d'un animal à l'autre.

#### **Les différentes étapes du protocole**

La première étape a consisté à optimiser la préparation des échantillons avec des artères piales de souris sauvages. L'élément clé de ce protocole est la réalisation de l'extraction et de la digestion des protéines en peptides en solution dans un seul tube.

Les réactifs sont ajoutés les uns après les autres dans le tube contenant l'échantillon jusqu'à la récupération des peptides protéolytiques. Le SDC a été choisi comme détergent car il peut être efficacement précipité en fin de digestion par acidification, nécessaire également pour l'inhibition de la trypsine, contrairement au SDS, dont l'élimination nécessite une manipulation fastidieuse (FASP, gel). Pour le dessalage, afin de minimiser la perte de peptides très polaires, des dispositifs de dessalage de grande capacité (30 µg) sont utilisés. L'étape de dessalage a toujours été réalisée de manière extemporanée juste avant les analyses nanoLC MS / MS pour chaque échantillon et son contrôle. Des résultats reproductibles nécessitent des quantités injectées d'échantillon égalisées. Nous avons donc utilisé 20% de chaque échantillon pour doser la quantité de peptides, par mesure d'absorbance à 215 nm, avant les analyses nanoLC MS / MS. De plus, pour une reproductibilité optimale, tous les échantillons d'une série ont été préparés simultanément, avec les mêmes réactifs et solvants.

Les analyses LC-MS étant également des sources de variabilité (volume d'injection, instabilité du *spray*), nous avons réduit cet effet en utilisant des doublons techniques pour chaque échantillon. Les données de spectrométrie de masse sont traitées dans un premier temps à l'aide du logiciel Proteome Discoverer (Thermo Scientific, v2.2) et ensuite les recherches sont réalisées à l'aide du logiciel MASCOT v2.6.0 (Matrix Science, Londres, Royaume-Uni) avec à la base de données Mouse SwissProt (publiée en avril 2018) avec les paramètres suivants: 10 ppm pour la tolérance de masse sur les précurseurs en MS, 0.02 Da pour la tolérance de masse en MS / MS, la trypsine comme enzyme de digestion, la carbamidométhylation de la cystéine (C), la déamidation de l'asparagine (N) et de la glutamine (Q), oxydation de la méthionine (M) comme modifications des variables, un maximum de 2 erreurs de clivage et avec un taux de 1% de faux positif. Les annotations sur les protéines sont fournies par la base de données Uniprot et les annotations du matrisome sont obtenues par la base de données Matrisome (version 2.0) construite par le Hyne Lab (<http://matrisomeproject.mit.edu/>).

Chaque réplicat technique a été considéré pour la quantification uniquement après validation individuelle en fonction du nombre de MS / MS, du nombre de PSM (*Peptide Spectrum Match*)

ou du nombre de protéines avant la quantification d'un minimum de 2000 protéines dont 150 protéines du matrisome, par échantillon, identifiées par au moins 2 peptides avec le moteur de recherche Mascot. Toutes les analyses doivent générer des résultats similaires pour une condition donnée, et les valeurs aberrantes ont été écartées pour la quantification finale.

La 2<sup>e</sup> étape a consisté à optimiser chacune des étapes de l'analyse des données de MS/MS, de l'identification des protéines à leur quantification.

La quantification des protéines est réalisée à l'aide de MaxQuant (version 1.5.3.8) avec à la base de données Mouse Swissprot (publiée en avril 2018). L'oxydation (M), la carbamidométhylation (C) et la déamidation de l'asparagine (N) et de la glutamine (Q) ont été définies comme des modifications variables, avec jusqu'à 2 sites de clivage tryptique non protéolysés, l'utilisation de l'option MBR qui permet le transfert d'une identification MS/MS d'une analyse à l'autre et avec un taux de 1% de faux positifs pour les protéines et les peptides.

Les intensités des protéines (somme des intensités des peptides) sont normalisées par une méthode de type RLE (*Relative Expression Log*), méthode issue de la génomique, en partant du postulat que la quantité totale de protéine est équivalente entre les échantillons<sup>247</sup>. Nous avons ensuite appliqué une stratégie d'imputation des valeurs manquantes avec une méthode de type LSA (*Least square adaptive*), principe basé sur la méthode des moindres carrés. Ce principe est basé sur la construction d'une fonction qui permette de réduire au minimum les erreurs, un peu comme le ferait une régression linéaire, et à partir de cette fonction, pouvoir remplir les valeurs manquantes de certaines protéines. Cependant, cette fonction n'est pas construite sur les valeurs d'une seule protéine mais bien sur l'ensemble. Evidemment, cette imputation n'est appliquée que suite au tri des données manquantes inspirée de la règle des 80% (voir paragraphe I.3.4.2.2).

Nous avons filtré les protéines du matrisome en utilisant la base de données développée dans le laboratoire de Richard Hynes, avant d'appliquer un test statistique afin d'identifier les protéines différentiellement exprimées de manière significative entre les deux conditions.

Celles-ci sont identifiées par le test SAM pour *Significance Analysis of Microarrays*<sup>248</sup>, un test non paramétrique mais basé sur le test de Student (ou test t) à laquelle est ajouté une correction par permutations multiples afin de minimiser le nombre de faux positifs. Pour rappel, le test t de Student est un test permettant de comparer les moyennes de deux conditions, et en tenant compte des variations entre les différents échantillons de la condition, la variance, de déterminer si les moyennes des deux groupes sont significativement différentes. Comme le test est réalisé sur des milliers de protéines, il est indispensable de réaliser une correction. Le principe est le suivant : un test t des données réelles est comparé au test t des données permutées au hasard, et si la distribution des données réelles est identique à celles permutées, la différence annoncée est due au hasard, l'expression de la protéine n'est donc pas significativement différente entre les deux conditions. SAM, une

méthode issue de la génomique, utilise ce principe combiné avec un facteur de correction  $s_0$  arbitraire, afin de ne pas pénaliser les variances plus faibles.

Toutes les étapes de traitement des données sont réalisées par des scripts R fait maison à l'aide des packages R suivants : DESeq pour la normalisation, imp4p avec la méthode SLSA (*Structured Least Square Algorithm*)<sup>249</sup> pour les imputations et Siggenes pour le test statistique SAM.

### **La validation du protocole**

Pour valider ce protocole, nous avons utilisé un modèle de souris, bien caractérisé, de la maladie CADASIL, car dans cette forme génétique de cSVDs, des anomalies du matrisome ont déjà été identifiées. Pour cela nous avons utilisé des artères piales de 4 souris CADASIL (TgNotch3<sup>R169C</sup>) et 4 souris sauvages (TgNotch3<sup>wt</sup>). Les artères ont été disséquées chez des souris âgées de 12 mois car à cet âge les souris développent les manifestations caractéristiques pré-symptomatiques de la maladie, à savoir l'accumulation de Notch3, la dysfonction cérébrovasculaire et les lésions de la substance blanche. En revanche, il n'existe pas de remodelage visible en histologie de la MEC dans les vaisseaux cérébraux des souris mutantes.

Enfin, nous avons évalué l'utilité de ce nouveau protocole en étudiant les changements du matrisome des vaisseaux cérébraux associés au vieillissement, qui est le facteur de risque majeur des cSVDs.

## **III.1.4 Résultats et Discussions**

### **Robustesse de la méthode**

Ce protocole permet d'identifier et de quantifier de façon reproductible dans les artères cérébrales de souris, de différents âges et génotype, en moyenne 250 et 200 protéines du matrisome, avec une proportion comparable de protéines du core matrisome et du matrisome associé. Dans la littérature, les protocoles combinant décellularisation ou fractionnements extensifs identifient en moyenne entre 150 et 250 protéines du matrisome, avec un biais pour les protéines du core matrisome<sup>134</sup>. Le pourcentage de couverture de séquence des protéines du matrisome est comparable à celui des autres protéines. Les nuages de points (multiscatterplot) valident l'homogénéité des mesures des signaux protéiques entre les différents échantillons d'une même condition ou génotype, avec des coefficients de corrélation supérieurs à 0.91. Les Upset, équivalents aux diagrammes de Venn mettent en évidence la proportion de protéines communes entre les échantillons d'un même groupe, mise en évidence également par la distribution des annotations de Gene Ontology (GO)<sup>250</sup> pour les différentes localisations subcellulaires et celle du matrisome. Toutes les représentations graphiques utilisées lors de cette étude sont en Annexe II.

## **Validation de la méthode**

Cette méthode appliquée aux souris CADASIL et contrôles a permis d'identifier les quatre protéines du matrisome (TIMP3, Vitronectine, HTRA1 et COL18A1), connues comme étant dérégulées dans les artères cérébrales des souris et patients CADASIL<sup>25,113,115</sup>. Elle a également identifié six autres protéines du matrisome, pour lesquelles des études protéomiques antérieures réalisées sur du tissu cérébral post-mortem de patients CADASIL, et donc à un stade plus tardif de la maladie, avaient montré une anomalie de leur abondance. Enfin, elle a permis d'identifier de nouvelles protéines du matrisome différenciellement exprimées dans les artères cérébrales des souris CADASIL, qui feront l'objet d'études ultérieures.

## **Utilité de cette nouvelle méthode**

L'application de notre méthode sur des souris jeunes et âgées a révélé que l'âge entraînait des modifications majeures du matrisome des artères cérébrales chez la souris, qu'il reste à confirmer sur des artères cérébrales d'humains. De façon intéressante, certaines de ces modifications sont communes avec celles identifiées chez les souris CADASIL.

En conclusion, nous avons développé et validé une nouvelle méthode robuste et sensible pour l'analyse quantitative sans marquage des changements du matrisome des artères cérébrales de souris en tenant compte des différentes contraintes spécifiques à cette étude.

En effet, pour minimiser la perte d'échantillon, conséquence inévitable dans toute expérimentation mais dont l'impact est notable pour cette quantité de tissu, nous avons décidé d'extraire et de digérer les protéines dans le même tube. Dans cette même optique, aucun fractionnement, ni protéique, ni peptidique n'a été réalisé. Pour minimiser les biais de quantités de peptides analysés en MS, nous avons également normalisé les concentrations de peptides après dessalage tout en assurant une extraction qualitative la plus complète possible de l'échantillon.

Une surveillance de différents indicateurs tels que le nombre de MS/MS ou de peptides identifiés a été réalisé pour chaque analyse, afin de s'assurer du bon fonctionnement des différentes étapes de l'expérience.

Nous avons ainsi démontré par diverses analyses statistiques la robustesse de la méthode, validé notre méthode en identifiant des protéines largement connues par diverses autres études publiées et nous avons identifié environ 200 protéines du matrisome avec une répartition égale entre les protéines du *core* matrisome et celle du matrisome associé. L'étude de l'effet de l'âge nous a permis à nouveau de valider la robustesse de la méthode et de nos points de contrôle.





# A novel workflow for quantitative profiling of the murine cerebral vascular matrisome in Small Vessel Diseases

## Authors

Iman Haddad <sup>1</sup>, Valérie Domenga-Denier <sup>2</sup>, Virginie Dubosclard <sup>2</sup>, Anne Joutel <sup>2,3,4,#</sup>, Joelle Vinh<sup>1,#</sup>

# equal contribution

## Affiliations

1- Biological Mass Spectrometry and Proteomics (SMBP), ESPCI Paris, Université PSL, CNRS, 75005 Paris, France.

2- Institut de Psychiatrie et Neurosciences de Paris (IPNP), UMR\_S1266, INSERM, Université de Paris, 102-108 rue de la Santé, F-75014, Paris, France.

3- GHU Paris Psychiatrie et Neurosciences, Paris, France.

4- Department of Pharmacology, College of Medicine, University of Vermont, Burlington, VT, USA

## Corresponding authors:

Joelle Vinh, Biological Mass Spectrometry and Proteomics (SMBP), ESPCI Paris, Université PSL, CNRS, 75005 Paris, France.

Email: [joelle.vinh@espci.fr](mailto:joelle.vinh@espci.fr). Tel : +33 1 40 79 51 78

Anne Joutel, Inserm UMR 1266, Institute of Psychiatry and Neurosciences of Paris, 102-108 rue de la Santé, 75014 Paris, France

Email: [anne.joutel@inserm.fr](mailto:anne.joutel@inserm.fr). Tel: +33 1 40 78 92 96

Running title: Brain vascular proteome profiling in SVD

Text word counts (does not include title, abstract, ref, table, legends):

6 figures and 1 table

Keywords: Label free, quantitative proteomics, matrisome, extracellular matrix, small vessel disease

## Abstract

Alterations in the extracellular matrix (ECM) are commonly observed in small vessel diseases of the brain (SVDs), but yet poorly understood. Proteomic analysis of the ECM and ECM-associated proteins, collectively known as the matrisome, is a challenging task that usually requires extensive fractionation/extractions due to the complexity and insolubility of these proteins. Analysis of the matrisome from murine brain vessels is even more challenging due to the limited amount of tissue. Here, we developed a novel highly sensitive workflow dedicated to the quantitative profiling of the matrisome from murine brain arteries without prior ECM enrichment and fractionation: brain arteries from a single mouse are lysed and digested in-solution into peptides in a single-step, peptide identity and signal intensity are analyzed with the MaxQuant algorithm with the “match between runs” capability, the R packages Deseq2 and imp4p are used to normalize signal intensities and impute missing values respectively, and matrisome proteins are extracted in silico using the matrisome database (v2.0). This approach applied to brain arteries from a well-characterized mouse model of the CADASIL hereditary SVD, allowed to quantify 203 matrisome proteins per sample, including 50 % of core matrisome and 50% of matrisome-associated proteins. This identified 19 matrisome proteins differentially expressed in CADASIL versus control brain arteries after correcting for multiple testing, including notably all 4 matrisome proteins, ie TIMP3, vitronectin, HTRA1, COL18A1 known to accumulate in CADASIL. To explore the utility of this novel workflow, we assessed matrisome changes that occur upon aging, a major risk factor of SVD. We quantified 214 matrisome proteins and identified 129 proteins significantly up (110) or downregulated (19) in 20- versus 2-month-old brain arteries. Remarkably, samples could be clustered into the two overarching young/old groups based on the expression profile of matrisome proteins, indicating that young and old arteries have a unique matrisome profile. Our proof-of-principle study demonstrates the robustness of our workflow for quantitative profiling of the brain microvascular matrisome. This may facilitate better understanding of the potential roles that matrisome proteins play in SVDs.

## Introduction

The extracellular matrix (ECM) is a complex 3D network of highly cross-linked proteins that interact bidirectionally with surrounding cells to regulate the organization and homeostasis of tissues. ECM proteins and ECM-associated proteins, collectively known as matrisome proteins, act not only as physical scaffolds but also as master regulators of fundamental cellular processes including proliferation, differentiation, senescence and death (1–3). The matrisome is commonly divided in two groups, the core matrisome and the matrisome-associated. The core matrisome comprises around 300 proteins that include collagens, proteoglycans and ECM glycoproteins. The matrisome-associated protein category includes a list of around 800 proteins among which are secreted factors known or suspected to bind core ECM proteins, enzymes or other proteins that modify the ECM (ECM regulators) and proteins known to be associated or share some similarities with ECM proteins (ECM-affiliated) (4).

The matrisome of a wide range of organs, including but not limited to the heart, large arteries, intestine, lung, colon has been characterized in healthy or diseased humans or mice using shotgun proteomic approaches(5). Matrisome proteins are notoriously highly insoluble and far less abundant than most intracellular proteins. Hence, proteomic analysis of the ECM typically employs decellularization or subcellular fractionation protocols to enrich for highly insoluble proteins that are then solubilized, deglycosylated and analyzed by label free LC-MS/MS. Such enrichment based approach usually leads to the identification of 150 to 250 matrisome proteins for a given tissue, with a high level of tissue specificity. Since soluble matrisome proteins may be lost during the enrichment procedure(6), others have explored the possibility of characterizing the ECM without prior enrichment. Schiller et al. showed that the lung ECM from a mouse model of bleomycine-induced lung injury could be quantitatively profiled by extracting four distinct fractions from the total lung, using detergents of increasing strength, and analyzing the four fractions separately(7).

Cerebral small vessel diseases (SVDs) encompass an heterogeneous group of disorders which affect the structure or function of small intracranial vessels. These diseases account for 25 to 30% of ischemic stroke, the vast majority of spontaneous intracerebral hemorrhages and are a leading cause of age-related cognitive deficits. SVDs are deemed to result from a complex mix of genetic and cardiovascular risk factors, among which age and hypertension are the most important (8–10). The salient brain vessel changes include major alterations of mural cells (smooth muscle cells/ pericytes) and remodeling of the ECM, which may range from thickening of the basement membranes, splitting, duplication or fragmentation of the internal elastical lamina and fibrosis of the vessel wall (11). In recent years, several monogenic forms of SVDs, largely indistinguishable from sporadic SVD, have

been characterized. Remarkably, mutated genes, like COL4A1, COL4A2, HTRA1, cathepsin A (CTSA), encode matrisome proteins, or like NOTCH3, in the CADASIL genetic SVD, leads to the pathogenic accumulation of matrisome proteins (12). These observations combined to the fact that the vascular ECM occupies a strategic position in the communication pathway between the cerebral circulation and brain parenchymal cells have fueled the hypothesis that perturbations of the brain vascular matrisome are a common denominator in SVDs (12). Yet, we have a poor knowledge of matrisome changes that occur in SVDs.

Mouse models of sporadic or genetic SVDs offer the opportunity to explore the brain microvascular matrisome, especially at an early stage of the disease process. However, one challenge is that brain arteries dissected from a single mouse represents minute amount of tissue, whereas established protocols typically starts with at least 100 mg of tissue. Moreover, decellularization, subcellular or sequential fractionation are likely to further increase the risk of losing material. Therefore, in this study, we aimed to develop a novel workflow dedicated to the quantitative profiling of the matrisome of murine brain vessels. We validated this workflow by analyzing brain artery samples from a well-established genetic model of SVD (CADASIL) in which matrisome changes have been characterized. We further studied the utility of this workflow by assessing matrisome changes that occur upon aging, a major risk factor of SVD.

## Material and Methods

### Reagents

Phosphate buffered saline tablet (PBS), ammonium bicarbonate (ABC, ≥99.0%), dithiothreitol (DTT, ≥98.0%), sodium deoxycholate (SDC, ≥97.0%), iodoacetamide (IAM) were purchased from Sigma-Aldrich/Merck (France). Trifluoroacetic acid (TFA, LCMS Grade) and formic acid (FA, LCMS Grade) were purchased from Pierce. The mix trypsin-LysC was obtained by Promega and acetonitrile (ACN, LC-MS optima Grade) by Fisher.

### Mouse strains and brain artery samples

Animal experiments were conducted in full accordance with the guidelines of our local institutional Animal Care and Use Committees (CEEA9 and CEEA34). *TgNotch3<sup>R169C</sup>* mice overexpress the rat NOTCH3 protein containing the mutation R169C found in CADASIL patients and were maintained on an FVB/N background. As controls, we used *TgNotch3<sup>WT</sup>* mice, overexpressing wild type rat NOTCH3. To study the effect of aging we used 2 month-old and 20 month-old wildtype mice, also maintained on an FVB/N background.

Mice were sacrificed by a lethal injection of Dolethal. Brain arteries (pial arteries) were dissected under the scope, frozen in liquid nitrogen and stored at -80°C until use.

### Brain artery sample processing

Brain arteries from a single mouse were processed in a single tube. Brain arteries were first extensively washed in PBS to remove plasma proteins, then lysed in 50 µl of a solution containing 2% SDC, 10mM DTT in 50 mM ABC at 95°C for 5 min. After cooling to room temperature (RT), 5.5 µl of 500mM IAM (50mM final concentration) was added to alkylate free cysteine residues for 20 min at RT. Samples were further diluted with 144.5 µl of 50 mM ABC to reach a final concentration of 0.5% SDC before adding the mix trypsin-LysC (2 µg) and the digestion was carried out at 37°C overnight at 37°C with constant shaking.

Tryptic digests were centrifuged at 21,000 g for 15 min to recover the supernatant. 10.7 µl of TFA 10% (0.5% final concentration) was added to inactivate trypsin and LysC and precipitate SDC as previously described (13). Peptides were desalted using Peptide Cleanup C18 Spin Tubes (Agilent). Purified peptides were eluted with 70% ACN and 0.1% TFA. After removal of acetonitrile with a SpeedVac (Savant ISS110), desalted peptides were resuspended in solvent A' (H<sub>2</sub>O/ACN 98:2 (v/v), 0.1% TFA). Peptides concentration was determined by measuring the absorbance at 215nm (DeNovix, model DS-11, and Thermo Fisher Scientific, model NanoDrop One/One).

## Nano LC-MS/MS analyses

Between 2 and 4 µg of peptides was injected in duplicate on a reversed-phase trap column (C18 Acclaim PepMap100, 5 µm, 300 µm i.d. × 5 mm) and eluted to a reversed-phase column (C18 Acclaim PepMap100, 3 µm, 75 µm i.d. × 50 cm) at a constant flow rate of 300 nL/min, with a gradient of 2–50% of solvent B (H<sub>2</sub>O/ACN 10:90 (v/v), 0.1% FA) in solvent A (H<sub>2</sub>O/ACN 98:2 (v/v), 0.1% FA) over 150 min. The LC was coupled to a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific) equipped with a nano-electrospray ion source. Top 20 data-dependent-acquisition MS experiments consisted of one MS survey scan in the Orbitrap cell (resolution of 60000, scan range 375–1500 m/z, AGC target 3e6, maximum injection fill time 60 ms) followed by the MS/MS scans of the twenty most intense precursors in HCD mode (dynamic exclusion 30 s, no singly charged precursors, resolution 15000, NCE 28, AGC target 1e4, maximum injection time 60 ms, isolation window 2 m/z).

## Computational MS analysis

Number of MS/MS, peptide spectrum-match (PSM), peptides and proteins. Raw data were first processed using Proteome Discoverer software (Thermo Scientific, v2.2) and Mascot version v2.6.0 (Matrix Science, London, U.K.) against the Mouse SwissProt database (version April 2018) with the following parameters: 10 ppm MS mass tolerance for precursor ions, 0.02 Da MS/MS mass tolerance, enzyme trypsin, carbamidomethylation of cysteine (C), deamidation of asparagine(N) and glutamine(Q), oxidation of methionine (M) as variables modifications and up to two miscleavages of trypsin and 1% false discovery rate (FDR). Protein annotations were given by the Uniprot database and the matrisome database (v2.0) assembled by the Hyne's Lab (<http://matrisomeproject.mit.edu/>)(14).

Protein quantification. Data files were searched using MaxQuant (version 1.5.3.8) against the mouse Swissprot database (released April 2018) (15). The search parameters used the match-between-runs (MBR) capability, which assesses each peak identified and detected in an MS from any LC-MS/MS run and compares its retention time to other peaks not identified in any other run (15). Up to two miscleavages were allowed. Oxidation (M), carbamidomethylation (C) and deamidation of asparagine (N) and glutamine (Q) were set as variable modifications. A 1% FDR was used to filter candidate peptide and protein identification with 20 ppm MS mass tolerance for precursor ions and 20 ppm MS/MS mass tolerance. Raw intensities were extracted from the MaxQuant files and normalized using the Relative Log Expression (RLE) method implemented in the R package DeSeq2 (16,17). Briefly, this method first computes a pseudo-reference sample as the geometric mean across all samples. Next, it centers every sample compared to the pseudo-reference sample by calculating for every protein in a given sample, the ratio sample/pseudo-reference sample. Then, it applies a normalization factor to every sample calculated as the median value of all ratios. Finally, normalized intensity values are Log<sub>2</sub>-

transformed. Missing values were handled using the following strategy: i) a protein was classified as present/absent if missing values were detected in  $\leq 25\%$  of samples from one group and  $\geq 75\%$  of samples from the other group, ii) a protein was discarded and considered as not-quantified if missing values were detected in 50% of samples from at least one group, iii) in all other situations, missing values were imputed using the least-squares adaptive (LSA) method with the “imp4p” R package(18)(Gianetto et al, 2020, Bioarchive). Matrisome proteins were filtered using the matrisome database (v2.0) assembled by the Hyne Lab (<http://matrisomeproject.mit.edu/>)(14). Differentially expressed proteins were identified using the Significance Analysis of Microarrays (SAM) method (19) with the “SigGene”s R package (<http://bioconductor.org/packages/release/bioc/html/siggenes.html>), with a number of permutations set at 250. An adjusted p-value cutoff of 0.05 was applied to determine significantly deregulated proteins for each pairwise comparison.

### **Software packages for statistics and data visualization**

All scripts performing statistical analysis were implemented in R. Hierarchical clustering was performed using Heatmaper (<http://www.heatmapper.ca/>)(20). The dendrogram was generated with the average linkage method and we used the Euclidean method for calculating the distance matrix. Bar graphs and Whisker plots were generated using the GraphPad Prism 7.0 software.

### **Data access and proteomics raw data repository**

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier XXXXX.

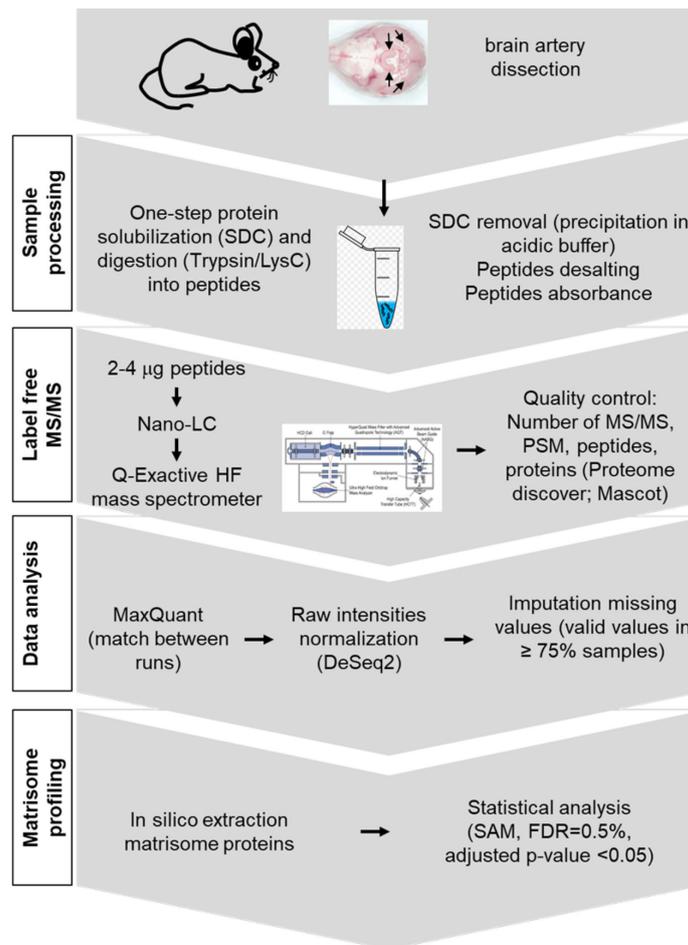
## Results and discussion

### Description of the workflow

Because brain arteries dissected from a single mouse represent less than 5 mg of tissue, we devised the following workflow summarized on Figure 1. For sample processing, we confine every step to a single tube until peptide purification to minimize sample loss. Protein solubilization and digestion into peptides are performed in-solution using a combination of detergent and enzymatic lysis. As a detergent, we selected the strong ionic SDC detergent because of its compatibility with trypsin (21). Also, due to its acid-insolubility, SDC can be efficiently removed from tryptic digests via precipitation in acidic buffers (21). In addition, the desalting step is performed extemporaneously just before nanoLC MS/MS analyses, to minimize peptide degradation or sticking to the vial. In a pilot study using brain arteries from wildtype mice (n=3), this procedure identifies per sample  $192,322 \pm 3,912$  MS/MS, with 33 % of peptide spectrum matches (PSM) generated from LC-MS/MS,  $24,692 \pm 206$  peptides and  $2541 \pm 38$  proteins including  $150 \pm 6$  matrisome proteins (52% core matrisome and 48% matrisome-associated) (Supplementary Table 1).

To identify and quantify protein abundance, we devised the following bioinformatic pipeline, using part of the MaxQuant software environment (15). Raw Intensities (summed peptide intensities) are extracted from MAXQuant. To increase proteome coverage, we employ the MaxQuant's match-between-runs (MBR) algorithm, which transfers identification through PSM from any one run to peaks having no MS/MS information in another run (15). Indeed, Murgia et al. showed that this approach allows to almost double the number of quantified proteins in human single skeletal muscle fibers (22). To correct for inter-sample variation in the amount of injected peptides, raw intensities extracted from the MaxQuant files are normalized. Assuming that most proteins are not differentially expressed and thus have an equal abundance across different samples, we apply a correction factor to every sample, calculated as the median of the ratio, for each protein, of its intensity value over its geometric mean across all samples (23). Missing values are imputed, provided that valid values are detected in  $\geq 75\%$  of samples from all groups, using a local similarity approach, which is based on the profiles of other proteins with similar intensity profiles across samples from the same data set (18).

Then, from the list of quantified proteins, we perform an "in silico screen" of matrisome proteins using the matrisome database (v2.0) assembled by the Hyne's Lab (<http://matrisomeproject.mit.edu/>)(14). Finally, to identify differentially expressed matrisome proteins, taking into account multiple testing, we implement the Significance Analysis of Microarray method, a nonparametric, permutation-based approach, initially developed for large genomic data sets, with a FDR set at 5%(24).

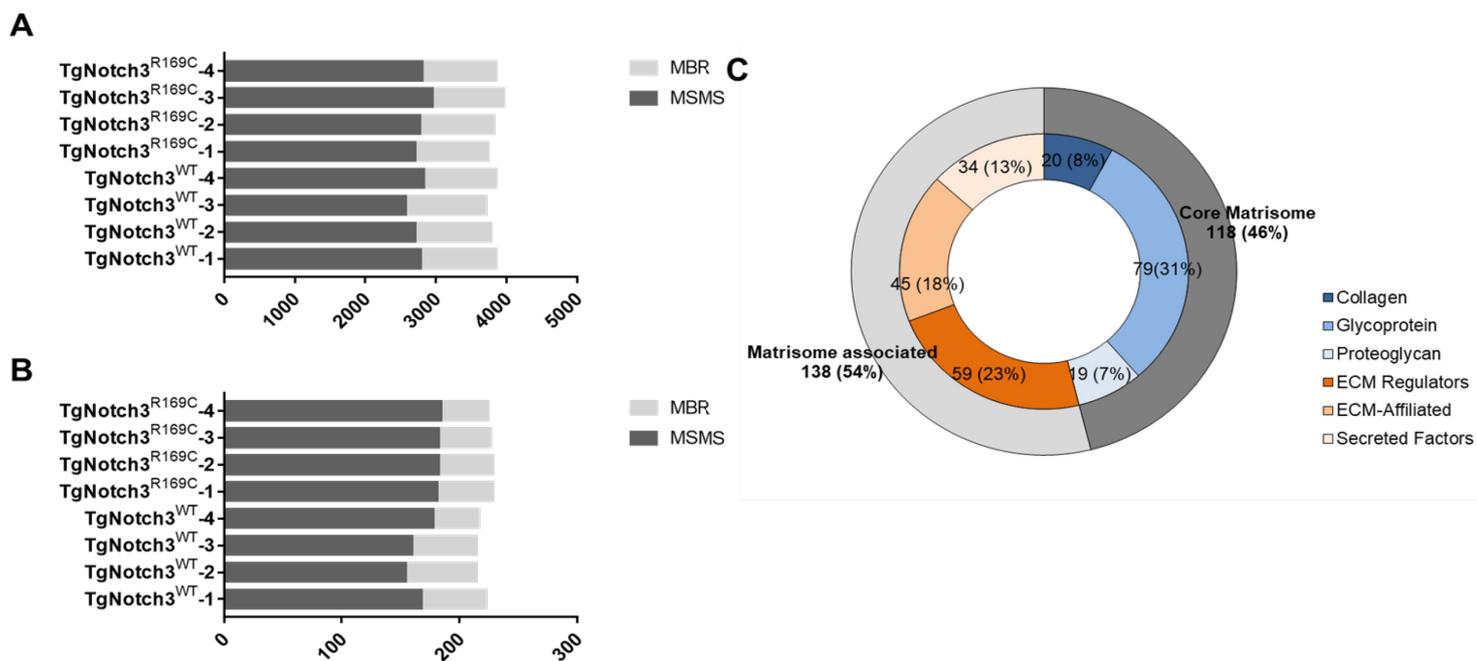


**Figure 1: Schematic of the experimental workflow devised for quantitative profiling of the matrisome from murine brain arteries**

### **Quantitative profiling of the brain vascular matrisome in the CADASIL genetic SVD**

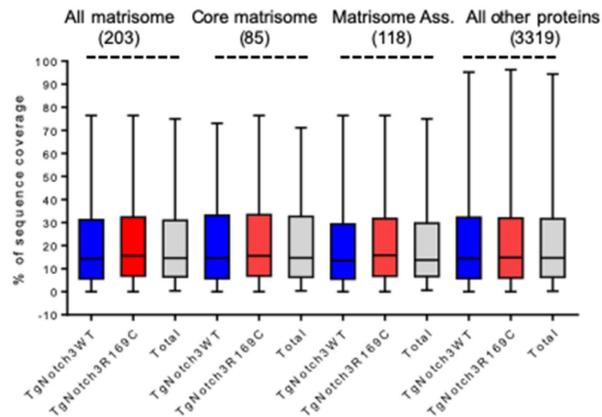
To test the validity of our workflow, we applied it to a mouse model of CADASIL, a genetic SVD in which changes in the brain microvascular matrisome have been characterized (25–27). We used the TgNotch3R169C model, a well-established mouse model of CADASIL which overexpresses a mutant NOTCH3 receptor with an archetypal CADASIL mutation. Transgenic mice which overexpress similar amount of the wildtype NOTCH3 (TgNotch3WT) were used as controls. Brain arteries were dissected at 12 months of age when mutant mice recapitulate the presymptomatic stage of the disease with abnormal accumulation of NOTCH3 in brain vessels, cerebrovascular dysfunction and white matter lesions (28–30). We identified per sample (n=4 TgNotch3R169C and n=4 TgNotch3WT), in the mascot software environment,  $184,754 \pm 6,700$  MS/MS, with 31% of PSM,  $21,322 \pm 1,277$  peptides and  $2417 \pm 115$  proteins including  $158 \pm 11$  matrisome proteins (53% core matrisome and 47% matrisome-associated) (Supplementary Table 2).

Using the bioinformatic pipeline described above (Figure 1), we identified in total 4444 proteins, by two or more unique peptides, of which 74% were identified in all 4 mutant or all 4 control samples. On average, we obtained 3818 proteins per sample including 2759 (72%) proteins identified by tandem MS and 1059 proteins (28%) identified by MBR using a library made of both the 8 CADASIL and control samples and all additional brain artery samples analyzed as part of this study (Figure 2A). Among these, there were 256 proteins identified as matrisome proteins (118 core matrisome and 138 matrisome-associated). On average 221 matrisome proteins were identified per sample (Figure 2B,C). Only proteins with valid values in all wildtype and mutant samples were quantified. In total, we quantified 3522 proteins), including 203 matrisome proteins (103 core matrisome and 100 matrisome associated). Notably, median sequence coverage of matrisome proteins and all other proteins was comparable; likewise median sequence coverages of core matrisome and matrisome associated proteins did not differ. Moreover, sequence coverage of matrisome proteins, core matrisome, matrisome associated and all other proteins were comparable between genotypes (Figure 3). Therefore, our protocol allows the identification of at least as many matrisome proteins than prior studies using ECM enrichment procedures (5). In addition our results indicate that this protocol does not bias against matrisome proteins in terms of sequence coverage.



**Figure 2: Characterization of the proteome of murine brain arteries from CADASIL and control mice**

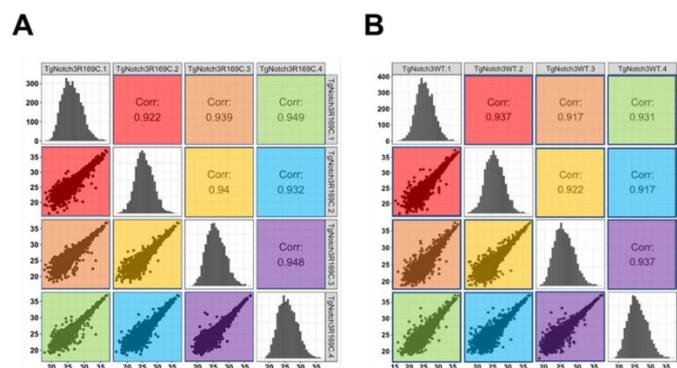
**A.** Number of total proteins identified with at least 2 unique peptides per sample by MSMS (dark grey bars) or match between run (light grey bars) **B.** Number of matrisome proteins identified per sample with at least 2 unique peptides **C.** Protein numbers from different groups of core matrisome and matrisome associated



**Figure 3: Protein sequence coverage of quantified proteins in CADASIL and control brain artery samples**

The box and whisker plots (Interquartile range 25-75%) depict the distribution of protein sequence coverage (coverage of possible tryptic peptides per protein in %) of quantified proteins in the matrisome (203 proteins), core matrisome (85 proteins) or matrisome associated (118 proteins) and all other proteins (3319) from CADASIL (TgNotch3<sup>R169C</sup>), control (TgNotch3<sup>WT</sup>) and both CADASIL and control samples (total).

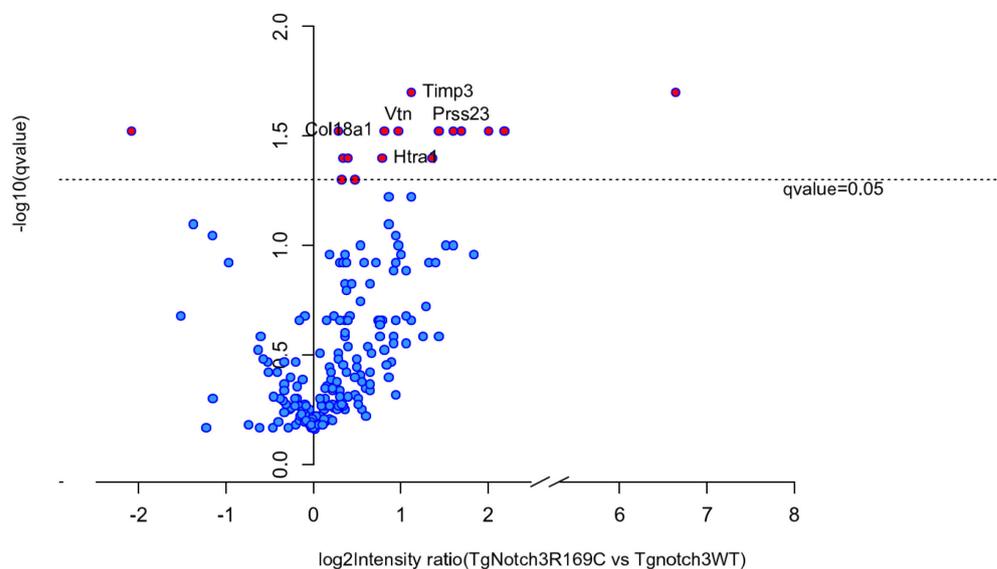
We then assessed the variability of brain artery proteome among mice of the same genotype. Normalized intensities were used to generate multiscatterplot of intensities and compute the Pearson's correlation coefficients between samples of the same genotype. We found that the mean Pearson's correlation coefficients were 0.93 and 0.94 in mutant and control samples respectively, ranging from 0.92 to 0.95, indicating a very low level of variability within each genotype (Figure 4). As an additional quality control, we compared NOTCH3 abundance in CADASIL versus control samples, and found a 30-fold increase in the abundance of NOTCH3 in mutant arteries compared to wildtype arteries (Supplementary Figure 1).



**Figure 4: Multiscatterplot and matrix of correlations between label-free quantification intensities of CADASIL and control samples.**

Multiscatterplot of intensities and Pearson correlation coefficients are reported for each comparison of CADASIL (TgNotch3<sup>R169C</sup>) (A) and control (TgNotch3<sup>WT</sup>) samples (B). The color code correlates the Pearson coefficient with the distribution of intensities.

Next, we compared the abundance of matrisome proteins between wildtype and mutant mice. Among the 203 quantified matrisome proteins, we identified 19 matrisome proteins (8 matrisome and 11 matrisome-associated) that were significantly differentially expressed (adjusted p-value <0.05) between wildtype and mutant samples. Eighteen proteins had increased abundance whereas 1 had lower abundance in mutant arteries (Figure 5, Table 1). Remarkably, this includes the 4 matrisome proteins, namely TIMP3, vitronectin, HTRA1 and COL18A1 with well-established increased abundance in brain arteries from both CADASIL mice and CADASIL patients (25–27). Moreover, 6 other differentially expressed matrisome proteins (CXCL12, EGFL8, HSPG2, MFGE8, LAMA5, PRSS23) have been shown to have a higher abundance in CADASIL brain vessels in prior MS/MS analysis, yet performed at a later stage of the disease process on human autopsy tissue specimens (25,27) (Table 1). In addition, 8 matrisome proteins were detected as present in murine CADASIL brain arteries but absent in the controls (Supplementary Table 3). Among these, Procollagen C-endopeptidase enhancer 1 is of peculiar interest because this matrisome glycoprotein is involved in the maturation of fibrillar collagen types I-III which abnormally accumulates in human CADASIL brain arteries and account for the fibrosis of vessel walls (31). Altogether, these data establish the validity of our workflow to quantitatively profile the brain vascular proteome in CADASIL mice.



**Figure 5: Volcano plot of  $\log_2$  ratios (CADASIL vs control) and  $-\log_{10}$  adjusted p-value (q-value) of all matrisome proteins.** Red circles indicate proteins with a significant change in abundance (FDR= 0.05, adjusted p-value < 0.05) and those labeled with gene names indicate the 4 matrisome proteins with established dysregulation from prior studies.

**Table 1 : List of matrisome proteins differentially expressed (adjusted p-value < 0.05) in brain arteries from CADASIL mice**

Gene names	Protein names	FC*	adjusted p-value		Category	Identified as differentially expressed in prior proteomic study (ref)
<b>Lgals8</b>	Galectin-8	111.20	0.0196	Matrisome-associated	ECM-affiliated Proteins	
<b>Adamts1</b>	ADAMTS-like protein 1	4.48	0.0269	Matrisome-associated	ECM Regulators	
<b>Cxcl12</b>	Stromal cell-derived factor 1	4.44	0.0269	Matrisome-associated	Secreted Factors	CADASIL patients (27)
<b>Thbs1</b>	Thrombospondin-1	3.97	0.0269	Core matrisome	ECM Glycoproteins	
<b>Hapln3</b>	Hyaluronan and proteoglycan link protein 3	3.20	0.0314	Core matrisome	Proteoglycans	
<b>Prss23</b>	Serine protease 23	3.19	0.0261	Matrisome-associated	ECM Regulators	CADASIL patients (27)
<b>Mfge8</b>	Lactadherin	3.02	0.0269	Core matrisome	ECM Glycoproteins	CADASIL patients (25)
<b>Pf4</b>	Platelet factor 4	2.71	0.0336	Matrisome-associated	Secreted Factors	
<b>Serpina3k</b>	Serine protease inhibitor A3K	2.69	0.0327	Matrisome-associated	ECM Regulators	
<b>Igfbp7</b>	Insulin-like growth factor-binding protein 7	2.55	0.0446	Core matrisome	ECM Glycoproteins	
<b>Timp3</b>	Metalloproteinase inhibitor 3	2.15	0.0196	Matrisome-associated	ECM Regulators	CADASIL patients and CADASIL mice (26,27)
<b>Vtn</b>	Vitronectin	1.97	0.0305	Core matrisome	ECM Glycoproteins	CADASIL patients and CADASIL mice (26,27)
<b>Col18a1</b>	Collagen alpha-1(XVIII) chain;Endostatin	1.75	0.0321	Core matrisome	Collagens	CADASIL patients and CADASIL mice (25)
<b>Htra1</b>	Serine protease HTRA1	1.71	0.0444	Matrisome-associated	ECM Regulators	CADASIL patients and CADASIL mice (26,27)
<b>Egfl8</b>	Epidermal growth factor-like protein 8	1.39	0.0454	Matrisome-associated	Secreted Factors	CADASIL patients (27)

<b>Itih5</b>	Inter-alpha-trypsin inhibitor heavy chain H5	1.32	0.0446	Matrisome-associated	ECM Regulators	
<b>Lama5</b>	Laminin subunit alpha-5	1.27	0.0446	Core matrisome	ECM Glycoproteins	CADASIL patients (25)
<b>Hspg2</b>	Basement membrane-specific heparan sulfate proteoglycan core protein	1.22	0.0336	Core matrisome	Proteoglycans	CADASIL patients (25)
<b>Kn1</b>	Kininogen-1	0.23	0.0269	Matrisome-associated	ECM Regulators	

FC, Fold change CADASIL versus Control

## Young and old brain arteries have a unique matrisome profile

We then tested the utility of our workflow to define matrisome changes that occur upon aging, a major risk factor in the most common SVDs. Brain arteries were dissected from 2-month-old (n=6) and 20-month-old wildtype mice (n=6) and processed as described above. The number of MS/MS, PSM, peptides, total proteins and matrisome proteins identified using the mascot software was roughly comparable to those identified in wildtype samples from the pilot study as well as from CADASIL and control samples (Supplementary Table 4). Using our pipeline, we identified in total 4857 proteins, of which 3893 were quantified in all 12 samples. Among these we identified 266 matrisome proteins (120 core matrisome and 146 matrisome associated proteins), of which 214 matrisome proteins (104 core matrisome and 110 matrisome associated proteins) were quantified in all samples. The mean Pearson's correlation coefficients computed from the normalized intensities were 0.93 (0.91-0.95) and 0.94 (0.91-0.96) in samples from 2-month-old and 20-month-old mice respectively, indicating a low inter-samples variability within each group (Supplementary figure 2). Also, the median sequence coverage of matrisome proteins compared to all other proteins was roughly similar in both 2-month-old and 20-month-old (Supplementary Table 5) .

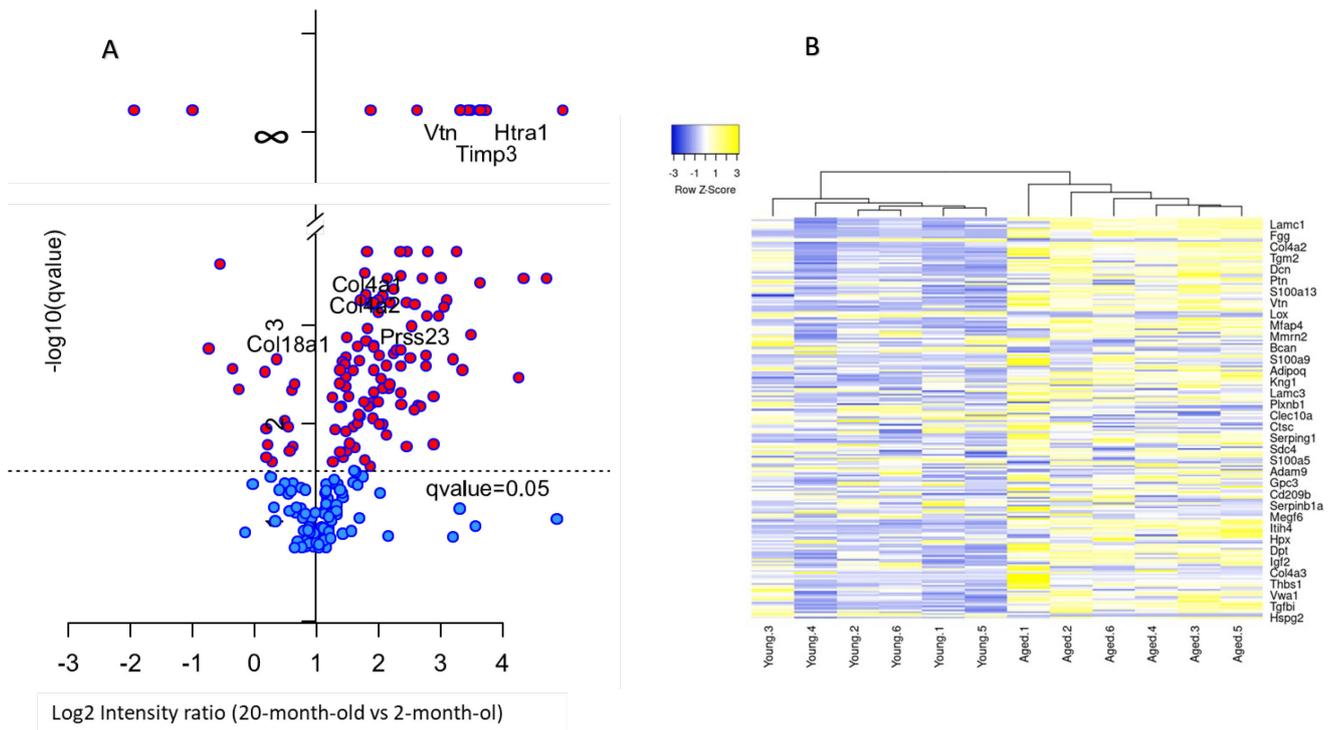
Comparison of the abundance of matrisome proteins between 2-month-old and 20-month-old samples revealed that 129 proteins (67 core matrisome and 62 matrisome associated) out the 214 quantified proteins (60%) were differentially expressed (adjusted p-value < 0.05) between young and aged samples, of which 110 were upregulated and 19 downregulated in 20-month-old arteries compared to 2-month-old arteries (Figure 6A). Remarkably, using protein expression of all matrisome proteins, unsupervised hierarchical clustering robustly divided the samples into the two overarching young/old groups, indicating and young and old brain arteries have a unique matrisome profile (Figure 6B). Interestingly, the 20 most strongly enriched matrisome proteins in 20-month-old arteries include HTRA, TIMP3, Vitronectin, MFGE8 and THBS1, five matrisome proteins with increased abundance in murine CADASIL brain arteries, suggesting some overlap with the CADASIL and the aged brain arterial proteome (Supplementary Table 6).

### Figure 6: Aging induces major changes in the matrisome of murine brain arteries

**A.** Volcano plot of  $\log_2$  ratios (20-month-old vs 2-month-old) and  $-\log_{10}$  adjusted p-values (q-value) of all matrisome proteins. Red circles indicate proteins with a significant change in abundance (FDR= 0.05, adjusted p-value < 0.05)

### Figure 6: Aging induces major changes in the matrisome of murine brain arteries

**A.** Volcano plot of  $\log_2$  ratios (20-month-old vs 2-month-old) and  $-\log_{10}$  adjusted p-values (q-value) of all matrisome proteins. Red circles indicate proteins with a significant change in abundance (FDR= 0.05, adjusted p-value < 0.05)



**Figure 6: Aging induces major changes in the matrisome of murine brain arteries**

**A.** Volcano plot of  $\log_2$  ratios (20-month-old vs 2-month-old) and  $-\log_{10}$  adjusted p-values (q-value) of all matrisome proteins. Red circles indicate proteins with a significant change in abundance (FDR=0.05, adjusted p-value < 0.05)

**B.** Matrisome proteins (214) were grouped using unsupervised hierarchical clustering of MS intensities across all 12 samples, showing a clear separation between the two age groups. Shown are only a subset of proteins names because of space constraint.

## Discussion

We provide here a detailed description of a novel highly sensitive workflow dedicated to the quantitative profiling of the matrisome from murine brain arteries, using shotgun label free mass spectrometry. This protocol, which starts with less than 5mg of tissue and does not require any ECM enrichment or fractionation, allows the identification and quantification of, on average, 250 and 200 matrisome proteins respectively with low inter-sample variability and no bias against matrisome proteins in terms of sequence coverage. We have established the validity of our workflow by applying it to a mouse model of CADASIL, a genetic SVD in which changes in the brain microvascular matrisome have been characterized. Quantitative profiling of brain arteries from CADASIL and control mice uncovered all 4 matrisome proteins, with established dysregulation in brain arteries from CADASIL mice and CADASIL patients, as well as novel additional matrisome proteins that are differentially expressed in CADASIL brain arteries. We demonstrated the utility of our workflow in the assessment

of the impact of aging on matrisome changes of brain arteries. We found that aging induces profound changes in the composition of the matrisome, consistent with the fact that aging is the most important risk factor of common SVDs.

This optimized workflow and the described methods should improve the understanding of the molecular mechanisms governing SVDs, which may help in the development of novel therapies.

## **ACKNOWLEDGEMENTS**

### **SOURCES OF FUNDING**

This work was supported by the Conseil regional d'Île-de-France (SESAME 2010 N°10022268) to JV, the European Union (Horizon 2020 Research and Innovation Programme SVDs@target under the grant agreement n° 666881) and the National Research Agency, France (ANR-16-RHUS-0004) to AJ.

## References

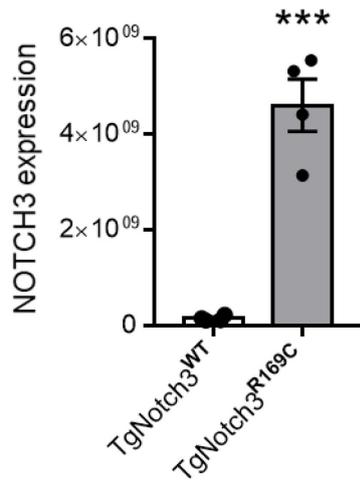
1. Hynes RO. Integrins: bidirectional, allosteric signaling machines. *Cell*. 20 sept 2002;110(6):673-87.
2. Hynes RO. The extracellular matrix: not just pretty fibrils. *Science*. 27 nov 2009;326(5957):1216-9.
3. Hynes RO, Naba A. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harb Perspect Biol*. janv 2012;4(1):a004903.
4. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol Cell Proteomics MCP*. avr 2012;11(4):M111.014647.
5. Taha IN, Naba A. Exploring the extracellular matrix in health and disease using proteomics. *Essays Biochem*. 13 2019;63(3):417-32.
6. Krasny L, Paul A, Wai P, Howard BA, Natrajan RC, Huang PH. Comparative proteomic assessment of matrisome enrichment methodologies. *Biochem J*. 1 nov 2016;473(21):3979-95.
7. Schiller HB, Fernandez IE, Burgstaller G, Schaab C, Scheltema RA, Schwarzmayr T, et al. Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. *Mol Syst Biol*. 14 juill 2015;11(7):819.
8. Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol*. juill 2010;9(7):689-701.
9. Wardlaw JM, Smith C, Dichgans M. Small vessel disease: mechanisms and clinical implications. *Lancet Neurol*. 2019;18(7):684-96.
10. Joutel A. Prospects for Diminishing the Impact of Nonamyloid Small-Vessel Diseases of the Brain. *Annu Rev Pharmacol Toxicol*. 06 2020;60:437-56.
11. Charidimou A, Pantoni L, Love S. The concept of sporadic cerebral small vessel disease: A road map on key definitions and current concepts. *Int J Stroke Off J Int Stroke Soc*. janv 2016;11(1):6-18.
12. Joutel A, Haddad I, Ratelade J, Nelson MT. Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain? *J Cereb Blood Flow Metab Off J Int Soc Cereb Blood Flow Metab*. janv 2016;36(1):143-57.

13. Lin Y, Zhou J, Bi D, Chen P, Wang X, Liang S. Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Anal Biochem.* 15 juin 2008;377(2):259-66.
14. Naba A, Clauser KR, Ding H, Whittaker CA, Carr SA, Hynes RO. The extracellular matrix: Tools and insights for the « omics » era. *Matrix Biol J Int Soc Matrix Biol.* janv 2016;49:10-24.
15. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* déc 2008;26(12):1367-72.
16. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11(10):R106.
17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
18. Webb-Robertson B-JM, Wiberg HK, Matzke MM, Brown JN, Wang J, McDermott JE, et al. Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics. *J Proteome Res.* 1 mai 2015;14(5):1993-2001.
19. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A.* 24 avr 2001;98(9):5116-21.
20. Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, et al. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* 08 2016;44(W1):W147-153.
21. Lin Y, Zhou J, Bi D, Chen P, Wang X, Liang S. Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Anal Biochem.* 15 juin 2008;377(2):259-66.
22. Murgia M, Toniolo L, Nagaraj N, Ciciliot S, Vindigni V, Schiaffino S, et al. Single Muscle Fiber Proteomics Reveals Fiber-Type-Specific Features of Human Muscle Aging. *Cell Rep.* 13 2017;19(11):2396-409.
23. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11(10):R106.
24. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A.* 24 avr 2001;98(9):5116-21.
25. Arboleda-Velasquez JF, Manent J, Lee JH, Tikka S, Ospina C, Vanderburg CR, et al. Hypomorphic Notch 3 alleles link Notch signaling to ischemic cerebral small-vessel disease. *Proc Natl Acad Sci U S A.* 2011;108(21):E128-35.

26. Monet-Leprêtre M, Haddad I, Baron-Menguy C, Fouillot-Panchal M, Riani M, Domenga-Denier V, et al. Abnormal recruitment of extracellular matrix proteins by excess Notch3 ECD: a new pathomechanism in CADASIL. *Brain J Neurol*. juin 2013;136(Pt 6):1830-45.
27. Zellner A, Scharrer E, Arzberger T, Oka C, Domenga-Denier V, Joutel A, et al. CADASIL brain vessels show a HTRA1 loss-of-function profile. *Acta Neuropathol (Berl)*. 2018;136(1):111-25.
28. Joutel A, Monet-Lepretre M, Gosele C, Baron-Menguy C, Hammes A, Schmidt S, et al. Cerebrovascular dysfunction and microcirculation rarefaction precede white matter lesions in a mouse genetic model of cerebral ischemic small vessel disease. *J Clin Invest*. 2010;120(2):433-45.
29. Capone C, Cognat E, Ghezali L, Baron-Menguy C, Aubin D, Mesnard L, et al. Reducing Timp3 or vitronectin ameliorates disease manifestations in CADASIL mice. *Ann Neurol*. mars 2016;79(3):387-403.
30. Rajani RM, Ratelade J, Domenga-Denier V, Hase Y, Kalimo H, Kalaria RN, et al. Blood brain barrier leakage is not a consistent feature of white matter lesions in CADASIL. *Acta Neuropathol Commun*. 21 2019;7(1):187.

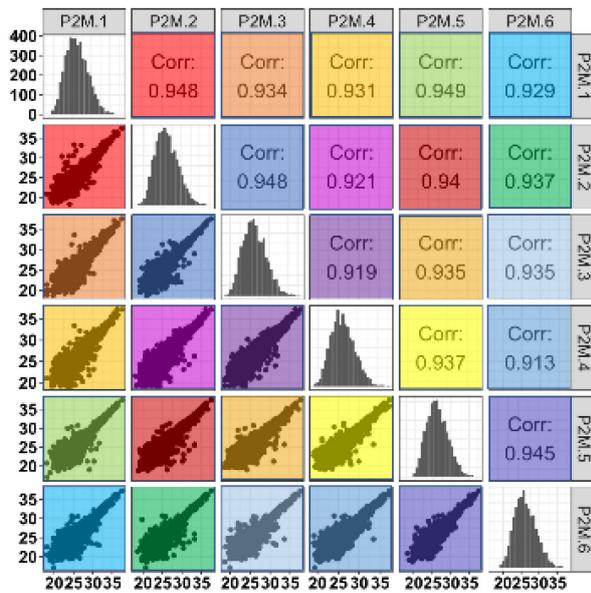
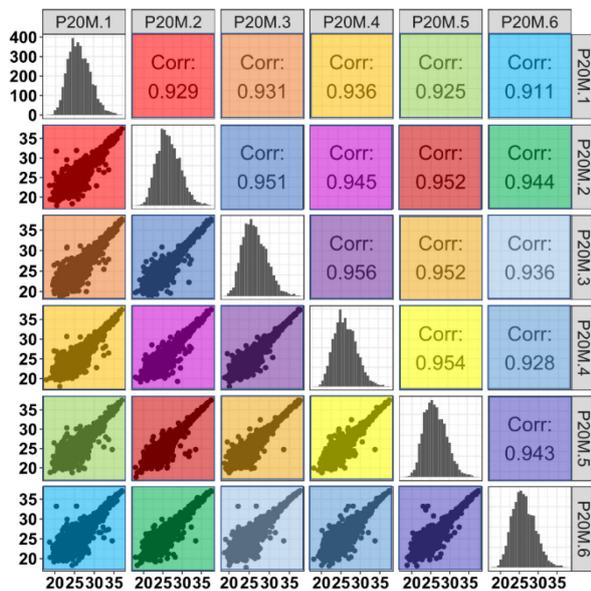
## **Supplementary Material**

**A novel workflow for quantitative profiling of the murine cerebral vascular matrisome in Small Vessel Diseases**



**Supplementary figure 1: NOTCH3 accumulation in CADASIL brain arteries**

Shown is the abundance of NOTCH3 protein (summed peptide intensity) in brain artery samples from CADASIL (TgNotch3<sup>R169C</sup>) and Control (TgNotch3<sup>WT</sup>) mice, with a 30-fold increase expression in CADASIL samples (Student's t-test, sem, \*\*\* P<0.001),

**A****B**

**Supplementary Figure 2: Multiscatterplot and matrix of correlations between label-free quantification intensities of young and aged brain artery samples.**

Multiscatterplot of intensities and Pearson correlation coefficients are reported for each comparison of brain artery samples from 2-month-old (A) and 20-month-old samples (B). The color code correlates the Pearson coefficient with the distribution of intensities.

## Supplementary Tables

**Supplementary Table 1:** Number of MS/MS, PSM, peptides and proteins identified in wildtype brain arteries (pilot study) using the Mascot Environment

	MS/MS (nbre)	PSM (% of MS/MS)	Peptides (nbre)	Proteins id $\geq$ 2 peptides (nbre)	Matrisome id $\geq$ 2 peptides (nbre)	Core matrisome (%)	Matrisome associated (%)
PV1	188.706	33	24.818	2584	155	51	49
PV2	196.474	32	24.454	2519	143	52	48
PV3	191.786	34	24.804	2519	151	54	46

**Supplementary Table 2:** MS/MS. PSM. peptides and protein number identified in brain arteries from wild type and CADASIL mice using the Mascot Environment

	MS/MS (nbre)	PSM (% of MS/MS)	Peptides (nbre)	Proteins id $\geq$ 2 peptides (nbre)	Matrisome id $\geq$ 2 peptides (nbre)	Core matrisome (%)	Matrisome associated (%)
TgNotch3 <sup>WT</sup> -1	176.586	33	20.867	2412	145	53	47
TgNotch3 <sup>WT</sup> -2	172.842	32	20.264	2337	142	54	46
TgNotch3 <sup>WT</sup> -3	185.000	28	19.770	2262	149	53	47
TgNotch3 <sup>WT</sup> -4	190.337	32	22.784	2480	157	54	47
TgNotch3 <sup>R169C</sup> -1	188.692	31	21.211	2360	164	51	49
TgNotch3 <sup>R169C</sup> -2	184.685	30	20.825	2399	166	53	47
TgNotch3 <sup>R169C</sup> -3	188.376	33	23.620	2651	173	52	48
TgNotch3 <sup>R169C</sup> -4	191.516	30	21.240	2442	166	53	47

**Supplementary Table 3:** List of matrisome proteins identified as present in CADASIL arteries and absent in Control arteries

Gene names	Protein names	Present/absent	Category	
<b>Pcolce</b>	Procollagen C-endopeptidase enhancer 1	Detected in 4 TgNotch3 <sup>R169C</sup> and 1 TgNotch3 <sup>WT</sup>	Core matrisome	ECM Glycoproteins
<b>Srpx</b>	Sushi-repeat-containing protein SRPX	Detected in 4 TgNotch3 <sup>R169C</sup> and 1 TgNotch3 <sup>WT</sup>	Core matrisome	ECM Glycoproteins
<b>Angptl2</b>	Angiopoietin-related protein 2	Detected in 3TgNotch3 <sup>R169C</sup> and absent in TgNotch3 <sup>WT</sup>	Matrisome-associated	Secreted Factors
<b>Clec1b</b>	C-type lectin domain family 1 member B	Detected in 3TgNotch3 <sup>R169C</sup> and absent in TgNotch3 <sup>WT</sup>	Matrisome-associated	ECM-affiliated Proteins

<b>Ctss</b>	Cathepsin S	Detected in 3TgNotch3 <sup>R169C</sup> and absent in TgNotch3 <sup>WT</sup>	Matrisome-associated	ECM Regulators
<b>Itih1</b>	Inter-alpha-trypsin inhibitor heavy chain H1	Detected in 3TgNotch3 <sup>R169C</sup> and in 1 TgNotch3 <sup>WT</sup>	Matrisome-associated	ECM Regulators
<b>Serpina1d</b>	Alpha-1-antitrypsin 1-4	Detected in 3TgNotch3 <sup>R169C</sup> and absent in TgNotch3 <sup>WT</sup>	Matrisome-associated	ECM Regulators
<b>Wnt5b</b>	Protein Wnt-5b	Detected in 4 TgNotch3 <sup>R169C</sup> and in 1 TgNotch3 <sup>WT</sup>	Matrisome-associated	Secreted Factors

**Supplementary Table 4:** MS/MS. PSM. peptides and protein number identified in brain arteries from 2-month-old and 20-month-old mice using the Mascot Environment

	MS/MS (nbre)	PSM (% of MS/MS)	Peptides (nbre)	Proteins id ≥ 2 peptides (nbre)	Matrisome id ≥ 2 peptides (nbre)	Core matrisome (%)	Matrisome associated (%)
2 month-old-1	195.742	26	20.784	2360	140	50	50
2 month-old-2	195.574	29	23.287	2824	138	54	46
2 month-old-3	195.650	27	21.372	2568	168	55	45
2 month-old-4	194.329	35	27.067	2955	164	50	50
2 month-old-5	194.382	31	24.596	2739	140	53	47
2 month-old-6	196.914	31	24.249	2749	156	51	49
20-month-old-1	194.121	31	23.426	2622	169	58	42
20-month-old-2	194.167	26	20.415	2403	151	57	43
20-month-old-3	195.900	38	29.229	3100	166	57	43
20-month-old-4	196.060	32	23.719	2617	169	51	49
20-month-old-5	193.286	27	21.118	2426	159	56	44
20-month-old-6	195.823	32	24.653	2738	173	57	43

**Supplementary Table 5:** Protein sequence coverage of quantified proteins in 2-month-old and 20-month-old samples, (per protein in %)

		Median	IQR (25-75%)
Matrisome (214 proteins)	2-month-old	13.3	4,8-28,1
	20-month-old	13.3	5,5-28,2
	Total	13,6	5,7-26,7
	20-month-old		
	Total		
All other proteins (3679)	2-month-old	14,3	5,8-30,1
	20-month-old	12,8	5,1-28,3
	Total	13,9	5,9-28,6

**Supplementary Table 6:** List of matrisome proteins differentially expressed (adjusted p-value < 0.05) in brain arteries from 20-mo versus 2-month old mice

Gene names	Protein names	Significant (FDR =0.05)	adjusted p-value	FC	Division	Category
Htra1	Serine protease HTRA1	+	0	15.68	Matrisome-associated	ECM Regulators
Itih4	Inter alpha-trypsin inhibitor. heavy chain 4	+	0	13.04	Matrisome-associated	ECM Regulators
Plg	Plasminogen;Plasmin heavy chain A;Activation peptide;Angiostatin;Plasmin heavy chain A. short form;Plasmin light chain B	+	0	10.11	Matrisome-associated	ECM Regulators
S100a8	Protein S100-A8	+	0.01	9.59	Matrisome-associated	Secreted Factors
Timp3	Metalloproteinase inhibitor 3	+	0	6.68	Matrisome-associated	ECM Regulators
Mfge8	Lactadherin	+	0	6.50	Core matrisome	ECM Glycoproteins
Serpinf2	Alpha-2-antiplasmin	+	0	6.24	Matrisome-associated	ECM Regulators
Mbl2	Mannose-binding protein C	+	0	6.22	Matrisome-associated	ECM-affiliated Proteins
F13b	Coagulation factor XIII B chain	+	0	6.18	Matrisome-associated	ECM Regulators
Ntn1	Netrin-1	+	0	5.63	Core matrisome	ECM Glycoproteins
Vtn	Vitronectin	+	0	5.63	Core matrisome	ECM Glycoproteins
Fgb	Fibrinogen beta chain;Fibrinopeptide B;Fibrinogen beta chain	+	0	5.43	Core matrisome	ECM Glycoproteins
C1qa	Complement C1q subcomponent subunit A	+	0	5.11	Matrisome-associated	ECM-affiliated Proteins
Fga	Fibrinogen alpha chain;Fibrinopeptide A;Fibrinogen alpha chain	+	0	5.03	Core matrisome	ECM Glycoproteins
Fgg	Fibrinogen gamma chain	+	0	4.99	Core matrisome	ECM Glycoproteins
Wnt6	Protein Wnt-6	+	0	4.97	Matrisome-associated	Secreted Factors
Ecm1	Extracellular matrix protein 1	+	0	4.77	Core matrisome	ECM Glycoproteins
Aspn	Asporin	+	0	4.59	Core matrisome	Proteoglycans
Thbs1	Thrombospondin-1	+	0	4.27	Core matrisome	ECM Glycoproteins
Hrg	Histidine-rich glycoprotein	+	0	4.14	Matrisome-associated	ECM Regulators
Kng1	Kininogen-1;Kininogen-1 heavy chain;Bradykinin;Kininogen-1 light chain	+	0	4.01	Matrisome-associated	ECM Regulators
Angptl2	Angiopoietin-related protein 2	+	0	3.91	Matrisome-associated	Secreted Factors
Hpx	Hemopexin	+	0.01	3.70	Matrisome-associated	ECM-affiliated Proteins
S100a9	Protein S100-A9	+	0.03	3.70	Matrisome-associated	Secreted Factors
Serpine2	Glia-derived nexin	+	0	3.48	Matrisome-associated	ECM Regulators

Emilin2	EMILIN-2	+	0	3.44	Core matrisome	ECM Glycoproteins
F2	Prothrombin;Activation peptide fragment 1;Activation peptide fragment 2;Thrombin light chain;Thrombin heavy chain	+	0	3.40	Matrisome-associated	ECM Regulators
C1qc	Complement C1q subcomponent subunit C	+	0	3.39	Matrisome-associated	ECM-affiliated Proteins
Wnt4	Protein Wnt-4	+	0	3.26	Matrisome-associated	Secreted Factors
Adamts1	ADAMTS-like protein 1	+	0.01	3.21	Matrisome-associated	ECM Regulators
Acan	Aggrecan core protein	+	0.01	3.11	Core matrisome	Proteoglycans
Hapln1	Hyaluronan and proteoglycan link protein 1	+	0	3.08	Core matrisome	Proteoglycans
Prss23	Serine protease 23	+	0	3.00	Matrisome-associated	ECM Regulators
Colec12	Collectin-12	+	0.01	2.98	Matrisome-associated	ECM-affiliated Proteins
Dpt	Dermatopontin	+	0	2.90	Core matrisome	ECM Glycoproteins
Serping1	Plasma protease C1 inhibitor	+	0	2.85	Matrisome-associated	ECM Regulators
Bgn	Biglycan	+	0	2.76	Core matrisome	Proteoglycans
S100a4	Protein S100-A4	+	0	2.74	Matrisome-associated	Secreted Factors
Lgals3	Galectin-3	+	0.03	2.74	Matrisome-associated	ECM-affiliated Proteins
Chadl	Chondroadherin-like protein	+	0.01	2.58	Core matrisome	Proteoglycans
Ogn	Mimiccan	+	0.01	2.57	Core matrisome	Proteoglycans
Fmod	Fibromodulin	+	0	2.57	Core matrisome	Proteoglycans
Cntf	Ciliary neurotrophic factor	+	0	2.56	Matrisome-associated	Secreted Factors
Pf4	Platelet factor 4	+	0	2.56	Matrisome-associated	Secreted Factors
Col12a1	Collagen alpha-1(XII) chain	+	0	2.55	Core matrisome	Collagens
Cd109	CD109 antigen	+	0	2.45	Matrisome-associated	ECM Regulators
F13a1	Coagulation factor XIII A chain	+	0	2.37	Matrisome-associated	ECM Regulators
Itih2	Inter-alpha-trypsin inhibitor heavy chain H2	+	0	2.36	Matrisome-associated	ECM Regulators
Col1a1	Collagen alpha-1(I) chain	+	0	2.27	Core matrisome	Collagens
Serpinc1	Antithrombin-III	+	0.01	2.26	Matrisome-associated	ECM Regulators
Tgfb1	Transforming growth factor beta-1;Latency-associated peptide	+	0.01	2.24	Matrisome-associated	Secreted Factors
Prg4	Proteoglycan 4;Proteoglycan 4 C-terminal part	+	0	2.20	Core matrisome	Proteoglycans

Plat	Tissue-type plasminogen activator;Tissue-type plasminogen activator chain A;Tissue-type plasminogen activator chain B	+	0.02	2.18	Matrisome-associated	ECM Regulators
A2m	Alpha-2-macroglobulin;Alpha-2-macroglobulin 165 kDa subunit;Alpha-2-macroglobulin 35 kDa subunit	+	0	2.18	Matrisome-associated	ECM Regulators
Fgf2	Fibroblast growth factor 2	+	0.01	2.10	Matrisome-associated	Secreted Factors
Lum	Lumican	+	0.02	2.09	Core matrisome	Proteoglycans
Col6a2	Collagen alpha-2(VI) chain	+	0	2.09	Core matrisome	Collagens
Col16a1	Collagen alpha-1(XVI) chain	+	0.01	2.05	Core matrisome	Collagens
Adamtsl4	ADAMTS-like protein 4	+	0.02	2.01	Matrisome-associated	ECM Regulators
Lgi3	Leucine-rich repeat LGI family member 3	+	0	2.01	Core matrisome	ECM Glycoproteins
Lgals9	Galectin-9	+	0.01	2.00	Matrisome-associated	ECM-affiliated Proteins
Prelp	Prolargin	+	0	2.00	Core matrisome	Proteoglycans
Sbspon	Somatomedin-B and thrombospondin type-1 domain-containing protein	+	0	1.99	Core matrisome	ECM Glycoproteins
Col6a1	Collagen alpha-1(VI) chain	+	0	1.91	Core matrisome	Collagens
Col1a2	Collagen alpha-2(I) chain	+	0	1.90	Core matrisome	Collagens
Mfap4	Microfibril-associated glycoprotein 4	+	0	1.89	Core matrisome	ECM Glycoproteins
Dcn	Decorin	+	0	1.89	Core matrisome	Proteoglycans
Itih5	Inter-alpha-trypsin inhibitor heavy chain H5	+	0	1.89	Matrisome-associated	ECM Regulators
Serpina1b	Alpha-1-antitrypsin 1-2	+	0.01	1.89	Matrisome-associated	ECM Regulators
Lama2	Laminin subunit alpha-2	+	0.02	1.88	Core matrisome	ECM Glycoproteins
Tgfbi	Transforming growth factor-beta-induced protein ig-h3	+	0	1.86	Core matrisome	ECM Glycoproteins
Mmrn1	Multimerin-1	+	0.01	1.85	Core matrisome	ECM Glycoproteins
Fn1	Fibronectin;Anastellin	+	0	1.83	Core matrisome	ECM Glycoproteins
Col3a1	Collagen alpha-1(III) chain	+	0.05	1.82	Core matrisome	Collagens
C1qb	Complement C1q subcomponent subunit B	+	0.01	1.79	Matrisome-associated	ECM-affiliated Proteins
Hapln4	Hyaluronan and proteoglycan link protein 4	+	0	1.77	Core matrisome	Proteoglycans
Col4a1	Collagen alpha-1(IV) chain;Arresten	+	0	1.76	Core matrisome	Collagens

Lama1	Laminin subunit alpha-1	+	0	1.74	Core matrisome	ECM Glycoproteins
Vwa1	von Willebrand factor A domain-containing protein 1	+	0	1.73	Core matrisome	ECM Glycoproteins
Col5a2	Collagen alpha-2(V) chain	+	0.04	1.72	Core matrisome	Collagens
Col4a2	Collagen alpha-2(IV) chain;Canstatin	+	0	1.72	Core matrisome	Collagens
S100b	Protein S100-B	+	0.01	1.70	Matrisome-associated	Secreted Factors
Tinagl1	Tubulointerstitial nephritis antigen-like	+	0	1.64	Core matrisome	ECM Glycoproteins
Mfap2	Microfibrillar-associated protein 2	+	0.02	1.63	Core matrisome	ECM Glycoproteins
Igf2	Insulin-like growth factor II;Insulin-like growth factor II;Preptin	+	0	1.62	Matrisome-associated	Secreted Factors
Anxa4	Annexin A4	+	0.01	1.60	Matrisome-associated	ECM-affiliated Proteins
Ctgf	Connective tissue growth factor	+	0.02	1.58	Core matrisome	ECM Glycoproteins
Col18a1	Collagen alpha-1(XVIII) chain;Endostatin	+	0	1.58	Core matrisome	Collagens
S100a11	Protein S100-A11	+	0.03	1.53	Matrisome-associated	Secreted Factors
Anxa11	Annexin A11	+	0	1.51	Matrisome-associated	ECM-affiliated Proteins
Adam10	Disintegrin and metalloproteinase domain-containing protein 10	+	0.02	1.51	Matrisome-associated	ECM Regulators
Lamc3	Laminin subunit gamma-3	+	0.03	1.45	Core matrisome	ECM Glycoproteins
Npnt	Nephronectin	+	0.01	1.43	Core matrisome	ECM Glycoproteins
S100a10	Protein S100-A10	+	0	1.40	Matrisome-associated	Secreted Factors
Fbn1	Fibrillin-1	+	0.04	1.40	Core matrisome	ECM Glycoproteins
Anxa2	Annexin A2	+	0.01	1.39	Matrisome-associated	ECM-affiliated Proteins
Lamc1	Laminin subunit gamma-1	+	0	1.39	Core matrisome	ECM Glycoproteins
Anxa1	Annexin A1	+	0.02	1.39	Matrisome-associated	ECM-affiliated Proteins
Nid1	Nidogen-1	+	0.01	1.38	Core matrisome	ECM Glycoproteins
Lamb2	Laminin subunit beta-2	+	0	1.37	Core matrisome	ECM Glycoproteins
Anxa6	Annexin A6	+	0.04	1.35	Matrisome-associated	ECM-affiliated Proteins
Hspg2	Basement membrane-specific heparan sulfate proteoglycan core protein;Endorepellin;LG3 peptide	+	0	1.33	Core matrisome	Proteoglycans
Sema3g	Semaphorin-3G	+	0.01	1.32	Matrisome-associated	ECM-affiliated Proteins

Vwf	von Willebrand factor;von Willebrand antigen 2	+	0	1.30	Core matrisome	ECM Glycoproteins
Cspg4	Chondroitin sulfate proteoglycan 4	+	0.04	1.30	Matrisome-associated	ECM-affiliated Proteins
Lama5	Laminin subunit alpha-5	+	0.01	1.29	Core matrisome	ECM Glycoproteins
Ctsb	Cathepsin B;Cathepsin B light chain;Cathepsin B heavy chain	+	0.01	1.29	Matrisome-associated	ECM Regulators
Plxn2	Plexin-B2	+	0.02	1.23	Matrisome-associated	ECM-affiliated Proteins
Agrn	Agrin;Agrin N-terminal 110 kDa subunit;Agrin C-terminal 110 kDa subunit;Agrin C-terminal 90 kDa fragment;Agrin C-terminal 22 kDa fragment	+	0.05	1.20	Core matrisome	ECM Glycoproteins
S100a13	Protein S100-A13	+	0.01	1.19	Matrisome-associated	Secreted Factors
Fbln2	Fibulin-2	+	0.01	0.78	Core matrisome	ECM Glycoproteins
Sparc	SPARC	+	0.03	0.77	Core matrisome	ECM Glycoproteins
Serpinh1	Serpin H1	+	0.01	0.76	Matrisome-associated	ECM Regulators
Sparcl1	SPARC-like protein 1	+	0.04	0.74	Core matrisome	ECM Glycoproteins
Plod3	Procollagen-lysine.2-oxoglutarate 5-dioxygenase 3	+	0.02	0.73	Matrisome-associated	ECM Regulators
Creld2	Cysteine-rich with EGF-like domain protein 2	+	0.02	0.70	Core matrisome	ECM Glycoproteins
Eln	Elastin	+	0	0.64	Core matrisome	ECM Glycoproteins
Itih3	Inter-alpha-trypsin inhibitor heavy chain H3	+	0.05	0.61	Matrisome-associated	ECM Regulators
Plxna4	Plexin-A4	+	0.03	0.58	Matrisome-associated	ECM-affiliated Proteins
Bcan	Brevican core protein	+	0.02	0.57	Core matrisome	Proteoglycans
Ltbp1	Latent-transforming growth factor beta-binding protein 1	+	0.04	0.57	Core matrisome	ECM Glycoproteins
S100a5	Protein S100-A5	+	0.01	0.56	Matrisome-associated	Secreted Factors
Tnr	Tenascin-R	+	0.01	0.42	Core matrisome	ECM Glycoproteins
Lgi1	Leucine-rich glioma-inactivated protein 1	+	0	0.39	Core matrisome	ECM Glycoproteins
Lox	Protein-lysine 6-oxidase	+	0	0.34	Matrisome-associated	ECM Regulators
Adam22	Disintegrin and metalloproteinase domain-containing protein 22	+	0	0.30	Matrisome-associated	ECM Regulators
Fbn2	Fibrillin-2	+	0	0.25	Core matrisome	ECM Glycoproteins
Fbln5	Fibulin-5	+	0	0.25	Core matrisome	ECM Glycoproteins
Ltbp4	Latent-transforming growth factor beta-binding protein 4	+	0	0.13	Core matrisome	ECM Glycoproteins

FC, Fold change 20-month-old versus 2-month-old. Highlighted in yellow are the 5 matrisome proteins in common with the “CADASIL” list.

## III.2 Étude des changements du matrisome microvasculaire dans les maladies des petits vaisseaux

L'objectif de cette seconde partie de résultats est l'analyse quantitative différentielle du matrisome dans les formes génétiques et sporadiques de maladies des petits vaisseaux cérébraux. Les différents modèles souris utilisés pour ces études sont présentés ainsi qu'une brève description des protocoles utilisés, validés précédemment. Pour chaque modèle génétique étudié, nous avons réalisé une 1ère étape de contrôle qualité afin de s'assurer de la validité de notre méthode d'analyse protéomique. Ensuite une première comparaison a été réalisée entre les résultats obtenus pour les différentes maladies génétiques, afin d'identifier les signatures communes potentielles ainsi que les divergences. Une seconde comparaison a été effectuée entre les différentes maladies sporadiques, pour déterminer également s'il existe des protéines impactées communes à cette catégorie de maladie. Et pour finir, une comparaison globale a été réalisée sur toutes les maladies.

### III.2.1 Introduction

Les maladies des petits vaisseaux sont un ensemble de processus pathologiques qui touchent les petites artères, les artérioles, les veinules ou les capillaires du cerveau. Les maladies des petits vaisseaux cérébraux sont responsables de lésions diffuses de la substance blanche du cerveau, d'infarctus cérébraux de petite taille profonds et multiples, de microsaignements, de dilatation des espaces périvasculaires et d'une atrophie cérébrale.

Les cSVD sont des maladies essentiellement sporadiques pour lesquelles l'âge et l'hypertension s'avèreraient être des facteurs importants. Cependant, il existe également des formes monogéniques, indifférenciables cliniquement des formes sporadiques, mais dont certains gènes responsables ont été identifiés aujourd'hui. De par leur similarité sur le plan clinique, neuroradiologique et histopathologique, la question de l'existence de mécanismes communs entre les formes sporadiques et les formes génétiques se pose et notamment de l'existence d'anomalies du matrisome communes entre ces différentes formes de cSVD. Dans une revue de la littérature publiée en 2016<sup>7</sup>, nous avons apporté un certain nombre d'arguments suggérant que les perturbations du matrisome pourraient constituer un dénominateur commun dans les mécanismes physiopathologiques à l'origine des cSVDs. Cette hypothèse de travail a depuis été reprise par d'autres auteurs dans d'autres revues <sup>106,251</sup>.

### III.2.2 Objectifs

L'objectif de cette étude est l'analyse protéomique qualitative et quantitative de plusieurs cSVD de type sporadique et génétique. Elle aura pour but, d'une part, d'identifier les protéines du matrisome dont l'expression diffère pour chaque maladie mais également d'identifier des protéines communes entre les différentes maladies étudiées, afin de potentiellement déterminer une voie ou une signature commune.

### III.2.3 Matériel et méthode

#### III.2.3.1 Modèles souris des formes génétiques de cSVDs

##### **CADASIL**

Cette maladie est causée par des mutations très stéréotypées du récepteur Notch3, conduisant à la présence d'un nombre impair de cystéine dans l'un des 34 domaines EGFR du domaine extracellulaire. Le modèle de souris de la maladie CADASIL a été décrit dans le précédent chapitre, il s'agit des souris TgNotch3<sup>R169C</sup>, qui surexpriment une protéine Notch3 de rat avec la mutation R169C, environ 4 fois plus que le Notch3 endogène et les souris contrôle, TgNotch3<sup>WT</sup>, qui surexpriment la protéine Notch3 de rat avec la séquence sauvage, environ 4 fois plus que le NOTCH3 endogène<sup>252</sup>. Les artères cérébrales ont été disséquées à l'âge de 12 mois (n=4/ génotype), pour les raisons décrites dans le précédent chapitre, c'est-à-dire que les souris développent les manifestations caractéristiques pré-symptomatiques de la maladie, à savoir l'accumulation de Notch3, la dysfonction cérébrovasculaire et les lésions de la substance blanche.

##### **CARASIL**

Celle-ci est causée par des mutations perte de fonction dans le gène HTRA1. Nous avons donc utilisé des souris dans lesquelles le gène Htra1 a été invalidé par une approche de type « *gene trap* ». Il s'agit des souris Htra1<sup>Gt(OST394864)Lex</sup>, obtenues grâce à une collaboration avec le laboratoire du Pr. Martin Dichgans à Munich (modèle non publié). Les souris contrôles sont les souris sauvages Htra1<sup>+/+</sup> issues des mêmes portées. Les souris mutées, malgré l'absence de HTRA1, ne développent pas de phénotype évident. Les artères cérébrales ont été disséquées à l'âge de 12 mois (n=6/ génotype).

## Maladie du collagène de type IV

Cette maladie est causée par des mutations très stéréotypées de la chaîne  $\alpha 1$  (COL4A1) ou  $\alpha 2$  (COL4A2) du Collagène IV, qui affectent le plus souvent la glycine du triplet GXY dans le domaine collagénique. Nous avons utilisé les souris Col4a1<sup>G1064D/+</sup>, qui expriment la mutation G1064D à l'état hétérozygote. Ces souris récapitulent les manifestations cliniques développées par les patients<sup>253</sup>. Les souris contrôles sont les souris sauvages des mêmes portées. Les artères cérébrales ont été disséquées à l'âge de 4 mois (n=6/ génotype), quand les souris mutées sont symptomatiques.

### *III.2.3.2 Modèles souris des formes sporadiques de cSVDs*

Les 2 principaux facteurs de risque des cSVDs communes sont l'âge et l'hypertension artérielle.

Pour étudier **l'effet de l'âge**, nous avons utilisé des souris sauvages âgées de 20 mois versus des souris sauvages âgées de 2 mois. Ces souris (n=6/ âge) ont été décrites dans le chapitre précédent.

Pour étudier **l'effet de l'hypertension**, nous avons utilisé le modèle bien établi des souris BPH (*Blood Pressure High*, Hypertendues) qui développent spontanément une hypertension artérielle vers l'âge de 5 semaines<sup>254</sup> et les souris contrôles BPN (*Blood Pressure Normal*, normotendues). Les souris BPH sont reconnues comme un modèle de souris d'hypertension neurogénique, résultant d'une suractivité du système nerveux sympathique. Comme ces deux lignées ne sont pas exactement sur le même fond génétique, nous avons également utilisé des souris BPH traitées avec l'amlodipine, un antihypertenseur de type anti-calcique qui a normalisé la pression artérielle. Les souris BPH développent une dysfonction cérébrovasculaire. Les artères cérébrales ont été disséquées à l'âge de 8 mois (n=6/ groupe).

La méthode de préparation des échantillons pour l'analyse MS/MS ainsi que la méthode d'analyse des résultats est résumée dans le précédent chapitre et décrite en détail dans l'article. Les données brutes concernant les souris CADASIL ainsi que les souris de 2 mois et 20 mois sont celles de la 1ère étude décrite dans le chapitre précédent.

## III.2.4 Résultats

Chaque résultat décrit est suivi d'une discussion, indépendamment des autres.

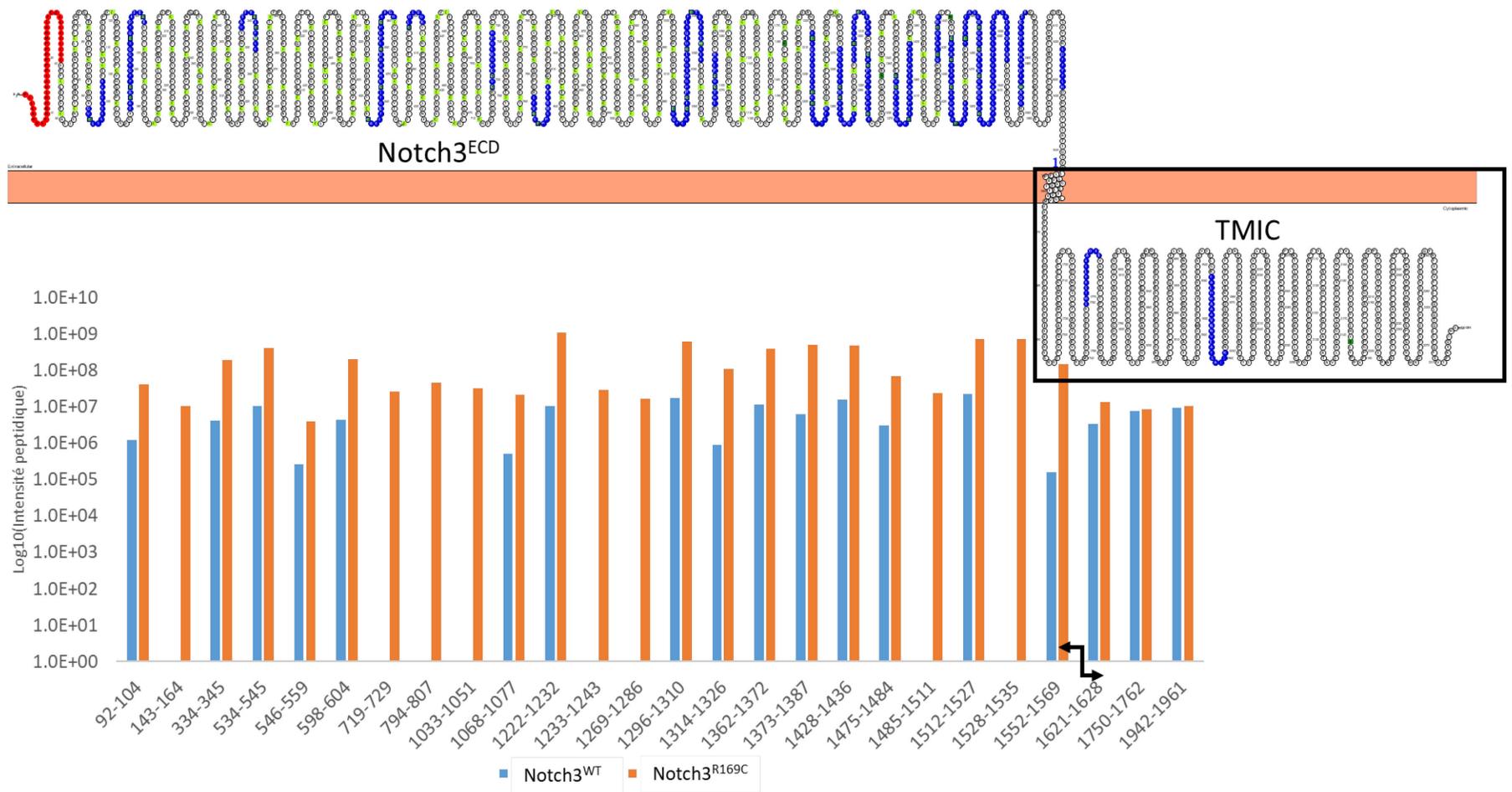
### III.2.4.1 Contrôles « qualité » dans les modèles génétiques de SVDs

Après les vérifications quantitatives en termes de nombres de MSMS, PSM, peptides et protéines identifiées lors des analyses réalisées, il était important également de s'assurer que notre protocole était capable de détecter certaines spécificités « protéiques » propres à chaque modèle. A notre connaissance, il n'existe pas de spécificités connues pour les modèles souris modélisant l'effet de l'âge ou de l'hypertension. En revanche, il en existe pour les modèles génétiques.

#### III.2.4.1.1 CADASIL

De nombreuses études ont été réalisées sur cette maladie des petits vaisseaux cérébraux, par conséquent plusieurs « contrôles positifs » sont observables sur les analyses accomplies.

- Le premier contrôle positif est incontestablement une surexpression de la protéine Notch3 dans CADASIL, produit du gène modifié responsable de la pathologie, qui s'accumule dans les vaisseaux des souris et des patients CADASIL. De plus, il a été également démontré que seul le domaine extracellulaire (Notch3<sup>ECD</sup>) qui contient les 34 domaines EGFR et les répétitions lin 12 (position sur la séquence de 1 à 1572) s'accumulent au niveau de la membrane plasmique<sup>17</sup> (Figure 16). Les peptides et leur position, dans l'ordre de la séquence, sont représentés ainsi que leur intensité respective en échelle logarithmique (log10) pour plus de visibilité. Les résultats démontrent que les intensités des peptides du domaine extracellulaire sont nettement plus élevées dans les artères de souris TgNotch3<sup>R169C</sup> en comparaison avec les artères de souris TgNotch3<sup>WT</sup> alors que dans le domaine TMIC, pour la partie TransMembranaire IntraCellulaire, elles sont très similaires.



Protter - visualize proteoforms Omasits et al., *Bioinformatics*. 2013 Nov 21.

Figure 16 : Distribution des intensités des peptides de Notch3 en fonction de leur position sur la séquence

- Comme décrit dans la première partie des résultats, dans le précédent chapitre, notre protocole a permis d'identifier les 4 protéines du matrisome (TIMP3, HTRA1, vitronectine et COL18A1) connues pour être en excès dans CADASIL (Figure17).

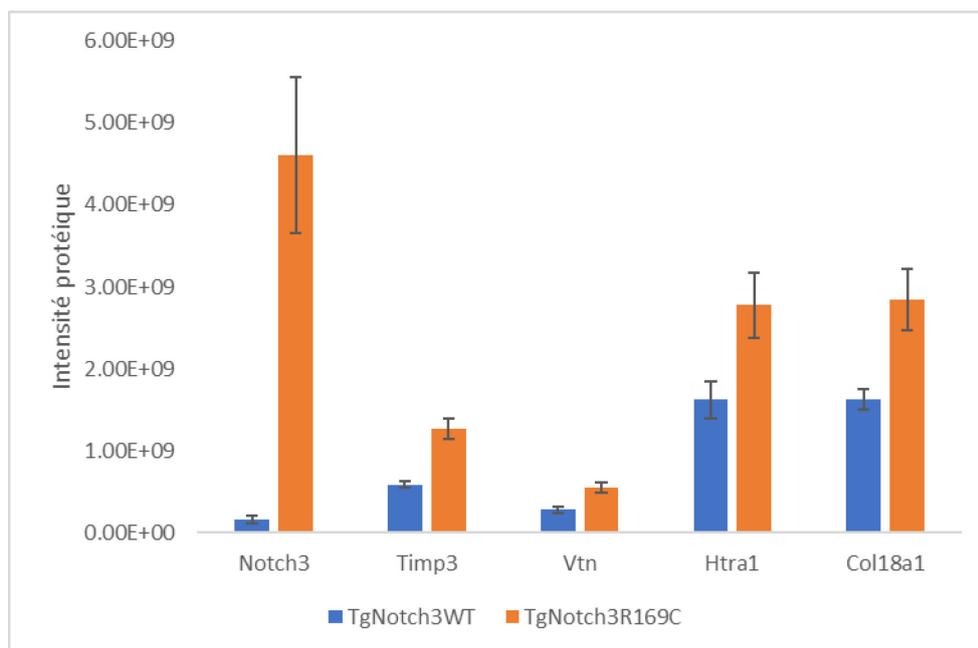
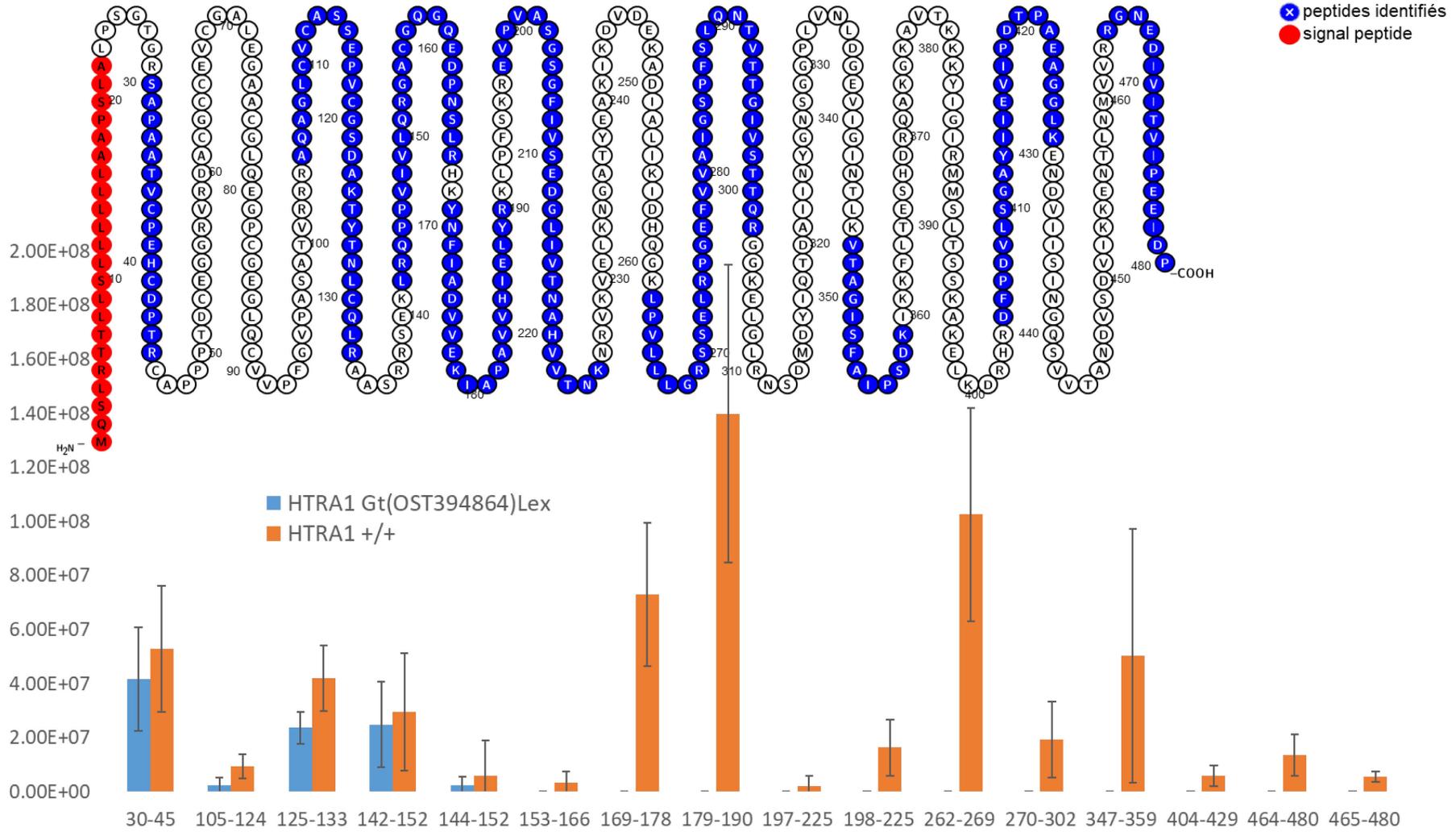


Figure 17 : Mesure des intensités normalisées pour tous les échantillons pour les protéines contrôles des protéines NOTCH3, TIMP3, VTN et HTRA1 dans les artères cérébrales des souris TgNotch3R169C et TgNotch3WT.

### III.2.4.1.2 CARASIL

Les études protéomiques sur CARASIL sont malheureusement encore rares et il y a peu de contrôle positif pour l'instant sur lequel s'appuyer<sup>115</sup>. Cependant, il y a un point de contrôle, dans les souris que nous avons analysées qui est lié à la méthode avec laquelle le gène *Htra1* a été invalidé. En effet, les souris que nous avons utilisées ont été générées par piégeage de gène. La modification génétique dans ces souris est censée conduire à la production d'une protéine tronquée instable de 157 acides aminés, à partir de l'exon 1. D'après les résultats obtenus (Figure 18), les peptides identifiés dans les échantillons des souris  $HTRA1^{Gt(OST394864)Lex}$ , sont d'une part d'une intensité très faible en comparaison avec les souris  $HTRA1^{+/+}$  et d'autre part, il y n'a pas d'identification au-delà de la position 152 de la séquence. Cette analyse nous a conduit à éliminer un échantillon (échantillon CAR2) identifié à tort comme provenant d'une souris invalidée pour HTRA1. Le domaine serine protease étant situé de la position 204 à 364, la protéine est bien inactivée au niveau de son activité serine protease.



Protter - visualize proteforms Omasits et al., Bioinformatics. 2013 Nov 21.

Figure 18 : Répartition des intensités peptidiques sur la séquence de la protéine HTRA1

### III.2.4.1.3 Maladie du collagène IV

Les maladies des petits vaisseaux cérébraux dues à des modifications du collagène de type IV, les études protéomique sur la maladie du collagène IV sont également encore peu documentés mais des publications ont mis en évidence une diminution de la sécrétion à la membrane basale de l'hétérotrimère COL4A1-COL4A1-COL4A2<sup>52</sup>. En effet, les cellules exprimant le mutant COL4A1 de manière stable ont une sécrétion réduite de la chaîne  $\alpha$  1 (IV) ainsi que de la chaîne  $\alpha$  2 (IV). L'effet est le même pour les mutants COL4A2. De nombreuses hypothèses du mécanisme sont à l'étude<sup>51,255</sup>. Une mesure de l'intensité normalisée des protéines COL4A1 et COL4A2 dans les analyses des souris COL4A1<sup>G1064D/+</sup> et des souris COL4A1<sup>+/+</sup> permet de valider cette diminution (Figure 19).

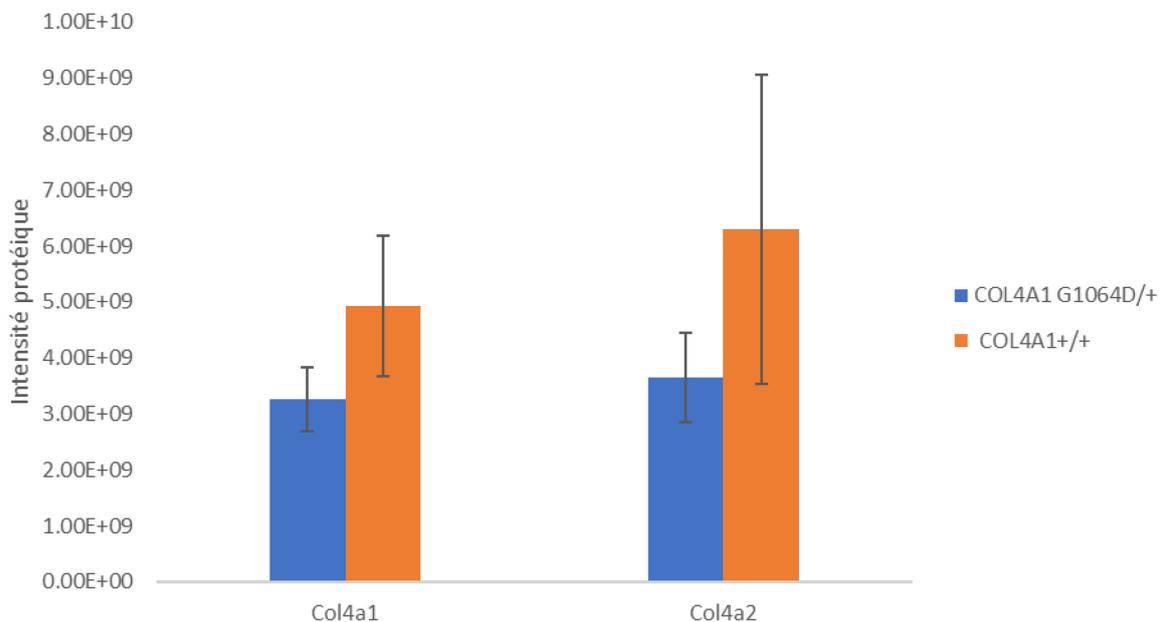


Figure 19 : Mesures des intensités protéiques normalisées des protéines COL4A1 et COL4A2 dans l'étude de la maladie du collagène IV dans les artères cérébrales des souris Col4a1G1064D/+ et Col4a1+/+

### III.2.4.2 Analyse des protéines du matrisome différemment exprimées dans les modèles génétiques de maladies des petits vaisseaux cérébraux

Une première comparaison est réalisée entre les analyses de protéomiques quantitatives dans les modèles génétiques de maladies de petits vaisseaux cérébraux afin d'identifier éventuellement un potentiel recouvrement des protéines différemment exprimées dans la maladie. Le tableau ci-dessous résume le nombre total de protéines ainsi que le nombre de protéines du matrisome quantifiées dans les échantillons des trois modèles génétiques, montre qu'ils sont relativement homogènes.

Table 6 : Tableau récapitulatif et quantitatif pour les 3 maladies génétiques

Facteur	Souris modèle	Nbre de protéines totales quantifiées	Nbre de protéines du matrisome quantifiées (Core/associé)	Nbre de protéines du matrisome différemment exprimées (mutant vs control) (fdr=0.05, q-value <0.05)
CADASIL	<i>TgNotch3</i> <sup>R169C</sup>	3522	203 (103-100)	<b>19 (18 up, 1 down)</b>
CARASIL	<i>Htra1</i> <sup>Gt(OST394864)Lex</sup>	3316	196 (101-95)	<b>9 (7 up, 2 down)</b>
Collagène IV	<i>Col4a1</i> <sup>G1064D</sup>	4257	215 (109-106)	<b>23 (19 up, 4 down)</b>

En moyenne, nous avons quantifié 200 protéines du matrisome dans les artères cérébrales de chacun de ces 3 modèles génétiques, ce qui représente entre 5 et 6 % des protéines quantifiées. Entre 5 et 10% de protéines sont différemment exprimées entre les souris sauvages et les souris mutées.

Nous avons ensuite représenté ci-dessous les protéines différentiellement exprimées dans chacun des 3 modèles sous forme de *Volcano plot* (de la figure 20 à 22). Les protéines dont l'expression est considérée significativement différente avec un FDR à 5% et une p-value ajustée (q-value) <0.05 sont colorées en rouge dans le graphe.

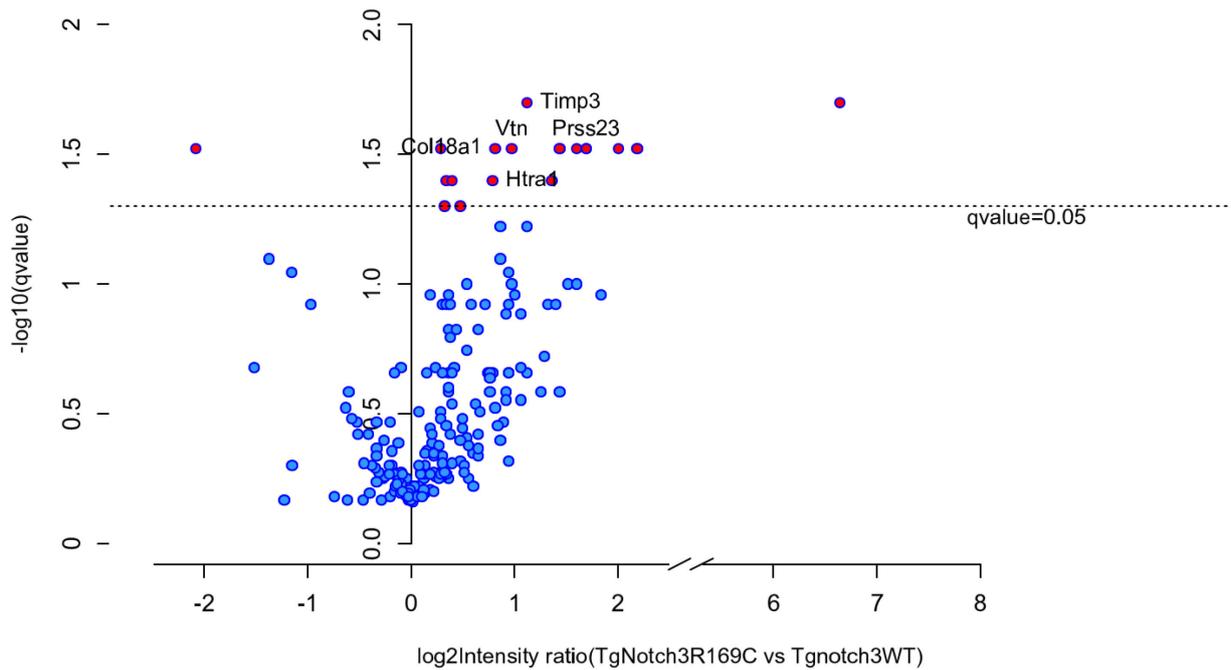


Figure 20 : représentation graphique de type Volcano plot pour l'étude CADASIL des rapports d'intensités des protéines  $\log_2(\text{ratio})$  entre les deux conditions avec leur significativité  $-\log_{10}(\text{qvalue})$ .

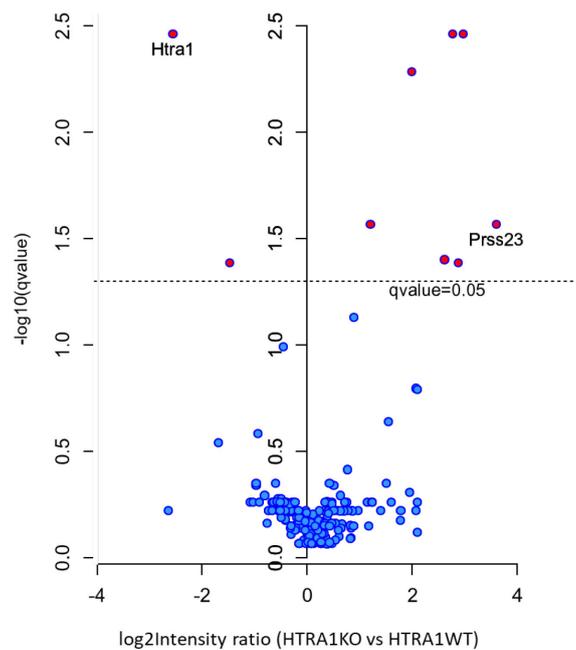


Figure 21 : représentation graphique de type Volcano plot pour l'étude CARASIL, des rapports d'intensités des protéines  $\log_2(\text{ratio})$  entre les deux conditions avec leur significativité  $-\log_{10}(\text{qvalue})$ .

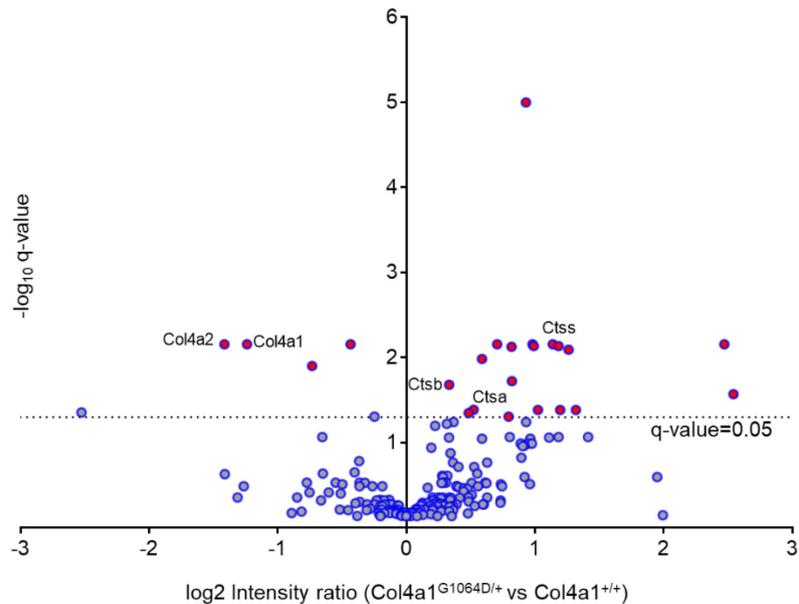


Figure 22 : représentation graphique de type Volcano plot pour l'étude de la maladie du collagène de type IV, des rapports d'intensités des protéines  $\log_2(\text{ratio})$  entre les deux conditions avec leur significativité  $-\log_{10}(q\text{-value})$ .

De manière générale, il s'agit le plus souvent de protéines surexprimées dans les artères cérébrales des souris mutées.

L'identité des protéines du matrisome différenciellement exprimées dans les souris CADASIL a été commenté dans le chapitre précédent.

Dans les artères cérébrales des souris CARASIL, déficientes en la sérine protéase HTRA1, figurent parmi les protéines significativement surexprimées, la protéines LTBP1, identifiée comme un substrat majeur de HTRA1<sup>256</sup>, ainsi que la protéine LTBP4, qui appartient à la même famille que LTBP1). En effet, les LTBP (latent transforming growth factor  $\beta$  binding proteins) comportent 4 isoformes numérotées de 1 à 4, elles interagissent avec les microfibrilles de fibriline qui est un constituant important des fibres élastiques. On trouve également la fibuline 5, également surexprimée, qui intervient dans la formation des fibres élastiques, en interagissant avec la tropo-élastine et le LTBP4<sup>257</sup>. **Il est intéressant de rappeler ici qu'il existe dans CARASIL, des anomalies majeures de la limitante élastique interne.**

Dans les artères des souris Col4a1 mutées, outre la diminution d'expression des protéines COL4A1 et COL4A2 déjà commentée ci-dessus, figurent parmi les protéines dont l'expression est différente de manière significative, HTRA1 impliquée dans CARASIL et la cathepsine A (CTSA) impliquée dans CARASIL<sup>258</sup>, qui sont toutes les deux surexprimées dans les artères mutantes. On trouve également 2 autres cathepsines (la cathepsine B—CTSB— et la cathepsine S—CTSS).

En utilisant la base de données d'expression des gènes de l'unité neurovasculaire établie par le laboratoire de Betsholtz (<http://betsholtzlab.org/VascularSingleCells/database.html>), on note que les gènes de ces 3 cathepsines sont essentiellement exprimés par les cellules microgliales périvasculaires.

Nous avons ensuite comparé entre elles les listes des protéines différentiellement exprimées dans ces trois modèles génétiques de cSVDs. Deux protéines sont partagées entre ces trois modèles. Il s'agit de la sérine protéase HTRA1, encore appelée PRSS11, qui est surexprimée dans CADASIL et la maladie COL4. La seconde protéine est PRSS23, une autre serine protéase qui est également surexprimée dans les trois maladies (Figure 23).

On note également qu'il existe plusieurs protéines partagées d'une part entre les modèles CADASIL et CARASIL, et d'autre part entre les modèles CADASIL et COL4.

Ensuite, pour identifier si un lien existe entre les différents cSVD, nous allons utiliser une méthode de type loi hypergéométrique, le même type de tests statistiques, parmi d'autres, utilisés pour l'analyse d'enrichissement en GO. Ce test permet de calculer la probabilité qu'un ensemble de protéines d'un même groupe GO soit présent, et plus la probabilité est basse, plus cela mettrait en évidence l'enrichissement d'une localisation subcellulaire, une fonction ou un mécanisme biologique. Dans cette même optique, au niveau de nos analyses, nous avons essayé de savoir s'il existait un enrichissement particulier de protéines entre nos différentes études.

Dans les modèles CADASIL et CARASIL, 179 protéines du matrisome ont été quantifiées en commun, 19 sont différentiellement exprimées dans les échantillons CADASIL et 9 dans les échantillons CARASIL. Le test hypergéométrique indique qu'avec le hasard une seule protéine ne devrait être commune aux 2 sets alors que nous trouvons 4 protéines communes entre les 2 sets. Selon le test hypergéométrique, il y a effectivement un recouvrement significatif entre la maladie CADASIL et CARASIL (enrichissement=4.2,  $p=0.008$ ). L'utilisation de ce test permet de mettre également en évidence un recouvrement significatif entre les protéines différentiellement exprimées dans CADASIL et celles différentiellement exprimées dans la maladie du collagène IV (enrichissement = 2.5,  $p=0.019$ ), mais différent de celui avec CARASIL, puisque qu'aucun recouvrement n'apparaît sur la maladie du collagène IV et CARASIL. Cependant, il existe une signature commune entre ces trois modèles génétiques, à savoir l'expression différentielle des protéines HTRA1 et PRSS23 (Figure 23).

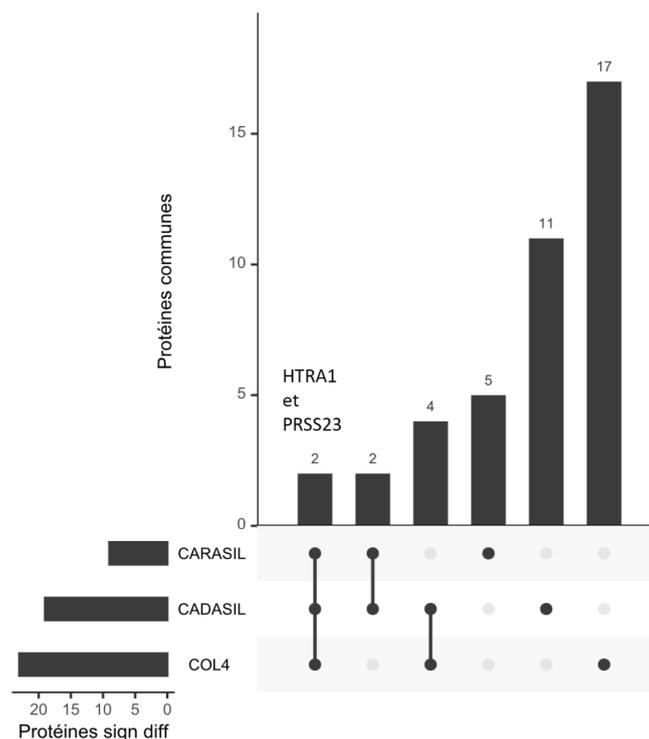


Figure 23 : Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les 3 maladies des cSVD génétiques

### III.2.4.3 Analyse des protéines du matrisome différemment exprimés dans les modèles sporadiques de maladies des petits vaisseaux cérébraux

Dans ce paragraphe sera présentée une brève description des résultats obtenus pour les maladies de type sporadique et ensuite une seconde comparaison est réalisée entre elles.

Le tableau ci-dessous résume le nombre de protéines totales, protéines du matrisome et protéines différemment exprimées dans les modèles souris mimant les deux principaux facteurs de risque des cSVDs.

Table 7 : Tableau récapitulatif et quantitatif pour les 2 maladies sporadiques

Facteur	Souris modèle	Nbre de protéines totales quantifiées	Nbre de protéines du matrisome quantifiées (Core/associated)	Nbre de protéines du matrisome différemment exprimées (mutant vs control) (fdr=0.05, q-value <0.05)
Age	20 mois vs 2 mois	3893	214 (104-110)	<b>129 (110 up et 19 down)</b>
Hypertension	BPH vs BPN	3364	194 (100-94)	<b>38 (9 up et 29 down)</b>

La couverture du protéome et du matrisome est très similaire à ce qui a été obtenu dans les modèles génétiques précédents. L'analyse protéomique quantitative entre les souris jeunes (2 mois) et les souris âgées (20 mois), ce qui correspond à l'échelle humaine d'étudier le protéome d'une personne âgée de 20 ans versus une personne âgée de 60 ans, a dévoilé des changements majeurs sur le protéome des petits vaisseaux cérébraux murins. On constate que l'âge induit des changements majeurs du matrisome, affectant 60% des protéines quantifiées et aussi bien le core matrisome (67 protéines) que le matrisome associé (62 protéines). La très grande majorité de ces protéines différentiellement exprimées (85%) voient leur abondance augmentée avec l'âge. De façon intéressante parmi les 19 protéines dont le niveau d'expression est diminué dans les artères cérébrales des souris âgées figurent des protéines très importantes pour la maturation et la stabilité des fibres élastiques, à savoir, l'élastine, la fibuline 5 ainsi que les protéines LTPB1 et LTPB4. Cette observation est cohérente avec la notion de la réduction avec l'âge du compartiment élastique, entraînant une augmentation de la rigidité des artères, décrite précédemment <sup>260</sup>. Parmi les protéines dont l'expression est augmentée avec l'âge, on note que celles dont l'expression est la plus élevée est HTRA1 et que parmi les 20 protéines dont l'expression augmente le plus avec l'âge figurent 5 protéines (HTRA1, TIMP3, vitronectine, MFGE8 et THBS1) dont l'expression est surexprimée dans les artères des souris CADASIL. L'analyse avec l'outil Gene Set Enrichment Analysis (GSEA)<sup>261</sup> met en évidence parmi les protéines différentiellement exprimées chez les souris de 20 mois une sur-représentation très significative des protéines de la voie du complément et de la coagulation ainsi que des protéines de la voie des intégrines ou impliquées dans les interactions ECM-récepteurs. Enfin, comme indiqué dans les résultats du précédent chapitre, l'analyse en cluster non supervisée de toutes les protéines du matrisome quantifiées dans les artères cérébrales des souris âgées de 20 mois et de 2 mois, en fonction de leur niveau d'expression permet de diviser les échantillons de manière robuste en deux groupes jeunes / vieux, indiquant que les artères cérébrales jeunes et âgées ont un profil du matrisome propre.

La figure 24, une carte de fréquentation ou plus communément une *heatmap* en anglais, représente très clairement une division entre les souris de 2 mois, dont l'expression protéique est majoritairement en jaune et les souris de 20 mois, qui tendent vers le bleu. Pour des questions de présentation, toutes les protéines n'ont pas pu être annotées sur cette cartographie, mais elle a été réalisée sur les 214 protéines identifiées du matrisome.

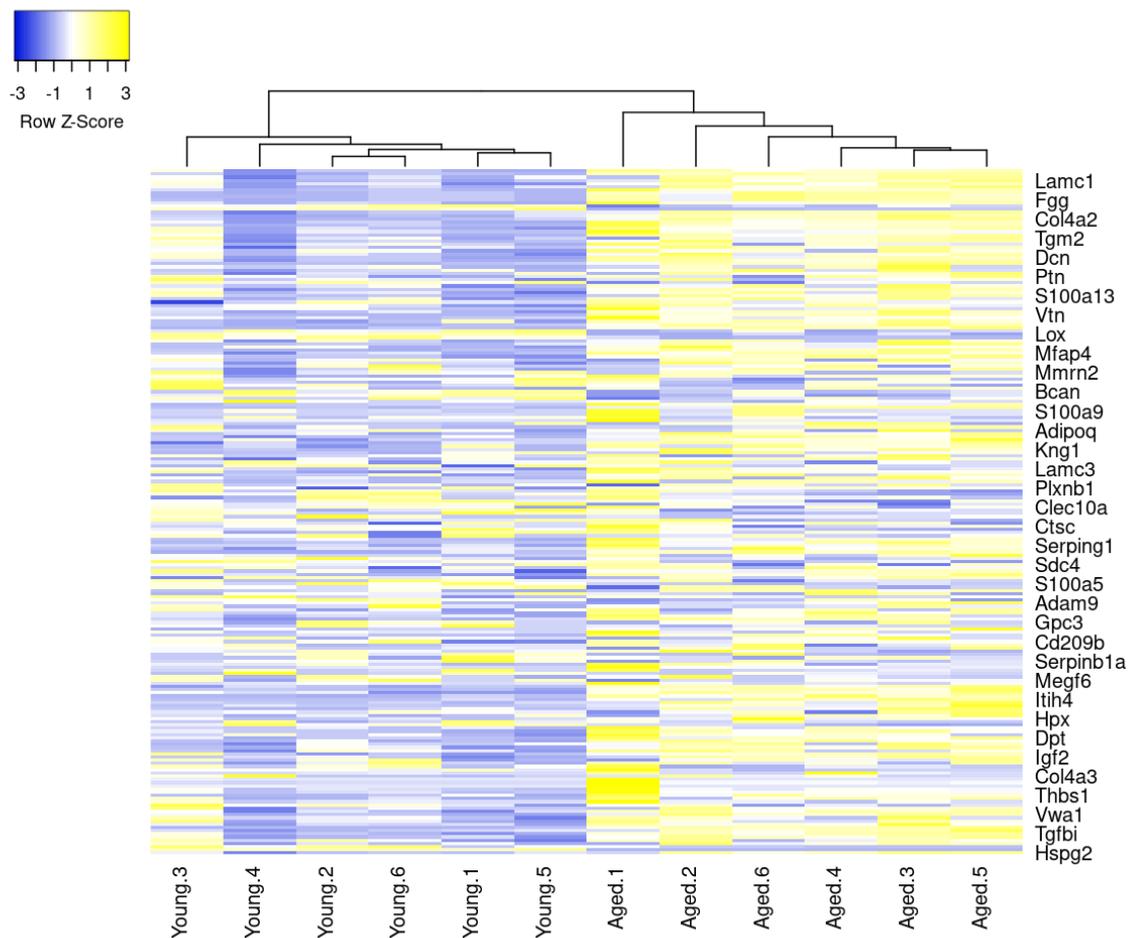


Figure 24 : Heatmap des analyses protéomiques pour l'étude de l'effet de l'âge

A la différence de l'âge, les effets de l'hypertension artérielle sur le matrisome sont plus modestes, avec « seulement » 20% de protéines dont l'abondance est modifiée entre les souris BPH et BPN de même âge. De plus, la majorité de ces protéines (76%) voit leur expression diminuée. Une des limites dans la comparaison du profil protéomique des artères cérébrales des souris BPH (hypertendues) et des souris BPN (normotendues) est que les BPN, même si elles sont couramment utilisées dans la littérature comme contrôles des souris BPH, n'ont pas exactement le même fond génétique. Ainsi les BPH ont un pelage blanc alors que les BPN ont un pelage foncé. Pour analyser plus spécifiquement l'effet de l'augmentation de

la pression artérielle sur les changements du matrisome, nous avons réalisé également une analyse protéomique du matrisome des souris BPH traitées avec un antihypertenseur administré de l'âge de 3 mois jusqu'au sacrifice de la souris, qui a permis de normaliser la pression artérielle. Cette analyse n'a permis d'identifier que 2 protéines dont l'expression est significativement différente (FDR=0.05 et q-value < 0.05) entre ces deux groupes de souris, à savoir Crlf1(Cytokine receptor-like factor 1) et MFGE8.

Nous avons ensuite comparé entre elles les listes des protéines différentiellement exprimées chez les souris âgées et celles différentiellement exprimées chez les souris hypertendues. Vingt-trois protéines sont communes entre les 2 listes (Figure 25). Cependant le test hypergéométrique indique que cela résulte du hasard. Au contraire, ce test suggère plutôt une divergence entre les 2 listes.

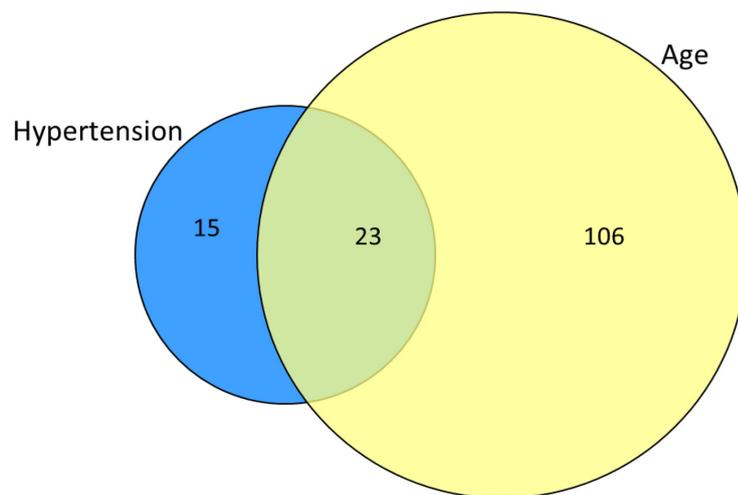


Figure 25 : Diagramme de Venn du recouvrement protéique entre l'effet de l'âge et l'hypertension

### III.2.4.4 Comparaison des différentes maladies et conclusions

Nous avons vérifié dans un premier temps s'il existait des similitudes dans les changements du matrisome induit par l'âge et ceux induits par une mutation génétique. L'utilisation du test hypergéométrique montre qu'il n'existe aucun chevauchement significatif entre chacune des formes génétiques de cSVD et l'effet de l'âge. Nous avons obtenu le même type de résultat en comparant l'effet de l'hypertension et les effets des mutations génétiques.

Nous avons analysé plus en détail la liste des protéines partagées entre les différentes formes sporadiques et génétiques de cSVD. Alors qu'il n'en existe aucune entre les 5 conditions, cette analyse fait apparaître que dans les 3 cSVDs génétiques et les souris âgées, 2 sérines protéases, HTRA1 et PRSS23 sont communes. HTRA1 et PRSS23 sont surexprimées dans les artères cérébrales des souris CADASIL, COL4 et âgées, PRSS23 est également surexprimée dans les artères des souris CARASIL mais HTRA1, comme attendu, est quasi-absente dans les artères des souris CARASIL (figure 26).

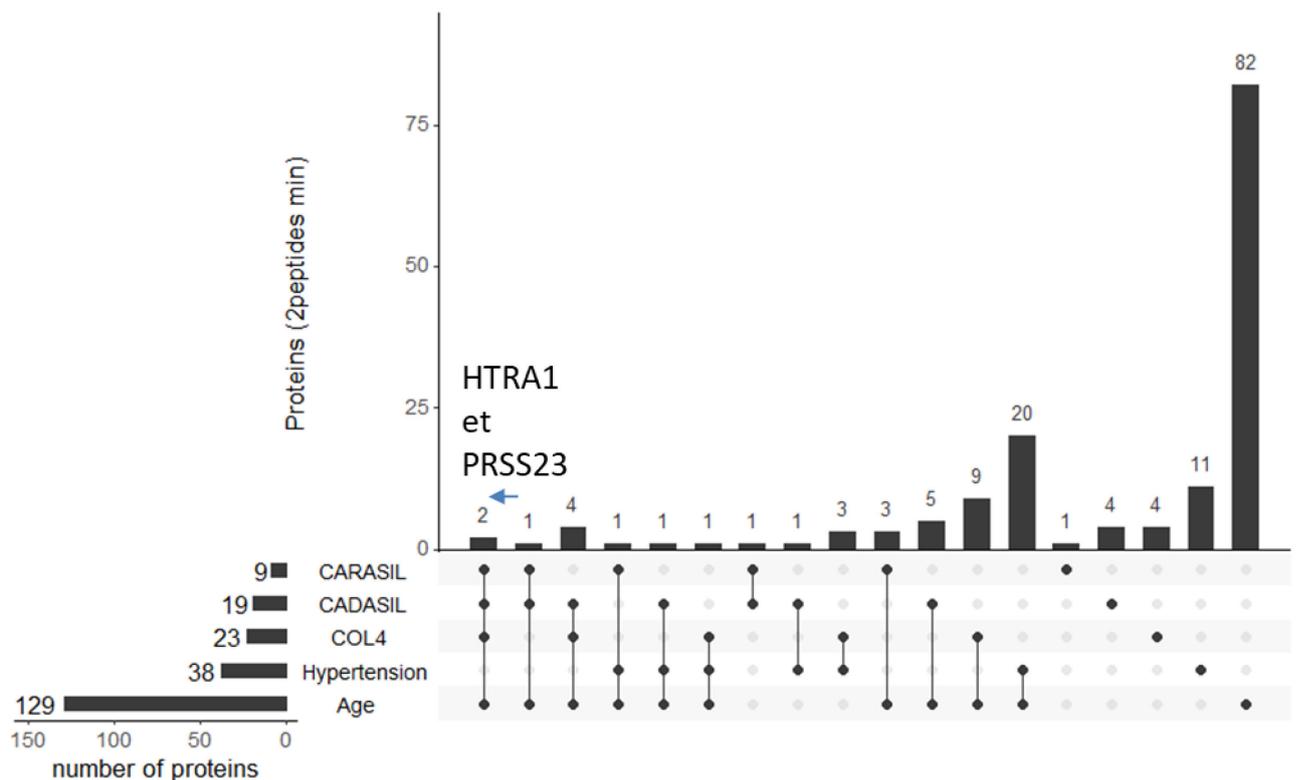


Figure 26 : Diagramme de Venn Upset <sup>259</sup> pour le recouvrement protéique des différentes maladies étudiées

Fait important, lorsqu'on analyse la liste des protéines différenciellement exprimées entre les souris hypertendues non-traitées et les hypertendues traitées, figure la protéine PRSS23 avec un seuil du FDR fixé à 0.05 mais dont la q-valeur est très légèrement supérieure à 0.05.

En conclusion, ces dernières analyses indiquent que l'augmentation d'expression de la protéine PRSS23 constitue une signature des formes sporadiques et génétiques de cSVDs.

### III.2.5 Discussion

Ce projet de thèse apporte de nouveaux résultats avec une étude qui n'avait jamais été entreprise ou très partiellement<sup>115</sup>. Nous avons réalisé une analyse protéomique quantitative du matrisome dans les artères cérébrales de modèles de souris de 3 formes génétiques et de modèles souris reproduisant les effets de l'âge ou de l'hypertension artérielle. Les résultats obtenus sont particulièrement informatifs :

- Selon la maladie, l'impact des modifications d'expression des protéines du matrisome est différent. En effet, les mutations génétiques n'induisent de modification « que » sur 5 à 10 % du matrisome, alors que l'âge en modifie plus de 50%. L'hypertension a un effet intermédiaire, puisque ces changements tournent autour de 20% des protéines du matrisome identifiées. Néanmoins, l'effet est équitable entre les protéines du core matrisome et les protéines du matrisome associé.
- CADASIL et CARASIL, d'après notre test d'enrichissement, et la liste des protéines du matrisome différentiellement significatives, possèdent des anomalies communes, un lien, moins prononcé cependant, existe également entre CADASIL et la maladie du collagène de type IV. A contrario, les changements provoqués par l'âge et ceux induits par l'hypertension sont très divergents.
- Nos analyses révèlent également un point commun entre les formes génétiques et sporadiques, une augmentation de l'expression des protéines HTRA1 (sauf pour CARASIL évidemment) et PRSS23.

Il faut souligner que cette étude a été réalisée dans des modèles de souris bien caractérisés, que les artères cérébrales ont été disséquées par le même expérimentateur dans le laboratoire de Anne Joutel. De même, l'ensemble des analyses protéomiques ont été réalisées dans des conditions identiques en utilisant un système de chromatographie identique sur toute la période de l'étude et sur une instrumentation définie au début du projet. Les contrôles qualité après optimisation du système ont été systématiquement validés de manière identique. C'est à notre sens une condition indispensable pour obtenir une bonne reproductibilité des résultats sur une période de 5 ans ce qui permet leur inter-comparabilité.

Si beaucoup de nos résultats sont confirmés dans la littérature, quelques aspects de notre étude pourraient être encore améliorés :

- Nous avons choisi d'analyser seulement des artères piales, il aurait été également intéressant d'analyser aussi des vaisseaux intracérébraux, même si nous avons considéré que ceci pourrait introduire davantage de variabilité biologique.
- L'état d'avancement du processus pathologique dans chacun de nos modèles souris varie. En effet, autant les souris Col4a1 mutées modélisent un stade avancé de la maladie, à l'autre extrême les souris déficientes en HTRA1 ne développent pas de phénotype apparent.
- Même si le nombre de protéines du matrisome quantifiées dans nos différents modèles est d'environ 200 protéines, il est possible que nos analyses ne couvrent pas totalement le répertoire des protéines du matrisome dans les artères cérébrales.
- Aucune validation des protéines identifiées comme différentiellement exprimées n'a pu être entamée. Une des difficultés est l'absence d'anticorps fonctionnant sur tissus de souris, pour une majorité de ces protéines.

Comme indiqué ci-dessus, nous avons montré que les changements du matrisome dans les modèles CADASIL et CARASIL partagent des similitudes très fortes. Il est intéressant de noter que Zellner et al, en analysant le protéome des vaisseaux cérébraux isolés à partir de tissu cérébral post-mortem provenant de patients CADASIL et celui de vaisseaux cérébraux isolés à partir d'un autre modèle de souris déficientes en HTRA1, ont également rapporté une similitude forte entre les changements du protéome dans ces deux maladies<sup>115</sup>. Cette similitude a été interprétée comme l'existence d'une perte de fonction de HTRA1 dans la maladie CADASIL. Néanmoins nous ne pouvons pas valider cette hypothèse. Notre hypothèse est qu'il existe des phénomènes de remodelage de la MEC, causaux ou non, communs dans ces deux pathologies, de même que nous avons identifié des similitudes dans les changements du matrisome entre CADASIL et la maladie COL4.

L'effet de l'âge, d'après les résultats, exerce un effet majeur sur le matrisome des petits vaisseaux cérébraux. Par exemple, on peut constater que les protéines impliquées dans la formation des fibres élastiques sont généralement sous-exprimées, effet constaté sur les gros vaisseaux dans plusieurs publications. Une augmentation de l'expression des collagènes est également notable, effet également déjà identifié par ailleurs<sup>262-264</sup>.

L'effet de l'hypertension est, contre toute attente, différent de l'effet du vieillissement sur le matrisome des petits vaisseaux cérébraux. Ce qui laisse supposer, si la perturbation du matrisome des vaisseaux cérébraux est pathologique, que les effets de l'âge et de l'hypertension artérielle pourraient être additifs lorsqu'ils sont présents. Le peu d'impact du traitement pour contrer l'hypertension peut s'interpréter de différentes façons. Dans la mesure où le traitement a effectivement corrigé l'augmentation de pression artérielle, confirmée par la mesure de la pression des souris après traitement, cela pourrait suggérer que les modifications du matrisome ne résultent pas directement de l'hypertension.

Un autre résultat particulièrement important de cette étude est la mise en évidence d'une signature commune à toutes ces formes de cSVDs avec la surexpression de PRSS23 dans les 5 modèles (son absence dans l'étude de l'hypertension est due à une  $q$ value à 0.06) et la surexpression de HTRA1 dans CADASIL, la maladie du collagène de type IV et l'âge, HTRA1 étant la protéine la plus augmentée chez les souris âgées. Bien que nous n'ayons pas encore confirmé la surexpression de ces deux protéines par une méthode orthogonale, nous sommes assez confiants que ces résultats sont bien réels. D'une part, si on regarde les résultats de l'étude protéomique réalisée par des collaborateurs sur des vaisseaux cérébraux de patients CADASIL et de souris HTRA1 déficientes, la protéine PRSS23 a été également surexprimée dans ces échantillons<sup>115</sup>. Il est assez remarquable que ces protéines soient toutes les deux des sérines protéases. Enfin, il est à noter que des résultats préliminaires indiquent que la surexpression de HTRA1 et PRSS23 dans nos modèles ne résulte pas d'une augmentation de la synthèse de leur transcrit.

Le PRSS23, protéine très peu étudiée, est une sérine protéase de 382 acides aminés. D'après la base de données MEROPS<sup>87</sup>, elle appartient à la famille des peptidases S1, dans laquelle il y a également la chymotrypsine, la trypsine mais également des élastases. La base de données transcriptomiques réalisée sur la BBB chez la souris confirme la présence de transcrits de *prss23* dans l'unité neurovasculaire, aussi bien dans les cellules endothéliales que les cellules musculaires lisses. Cette protéine serait impliquée dans la genèse et le maintien du bon fonctionnement de la valve cardiaque d'après une étude sur des poissons zèbres<sup>265</sup>, une autre publication suggère son implication dans les processus fibrotiques au niveau rénal<sup>266</sup>. Et enfin dans une étude sur les gènes dont l'expression est dérégulée dans du tissu provenant de cicatrices cutanées hypertrophiques versus de la peau normale chez l'homme ainsi que du tissu provenant de cicatrices cutanées prélevées chez la souris 8 semaines versus 6 semaines après la blessure, *PRSS23* fait partie de la liste des 16 gènes et des 5 sérine protéases dont l'expression est surexprimée aussi bien chez l'homme que chez la souris dans le processus cicatriciel.

HTRA1, une sérine protéase également et comme PRSS23, ses fonctions et mécanismes dans les vaisseaux cérébraux sont encore méconnus. Sont encore moins connus les mécanismes par lesquels les mutations à l'état homozygote ou hétérozygote qui diminuent la fonction de HTRA1 sont responsables de formes génétiques de cSVDs. Cependant, plusieurs publications évoquent l'impact délétère d'une accumulation de HTRA1, dans la dégénérescence maculaire liées à l'âge<sup>267,268</sup> ou encore dans diverses maladies musculosquelettiques (ostéoporose, arthrose,..)<sup>269</sup>.

# Synthèse et conclusion générale

---

Les maladies des petits vaisseaux cérébraux sont responsables de lésions de la substance blanche du cerveau et d'infarctus cérébraux profonds multiples. Elles sont à l'origine de plus de 25% des accidents vasculaires cérébraux et constituent la deuxième cause de démence après la démence d'Alzheimer. C'est un ensemble de processus pathologiques, qui affectent les petites artères, les artérioles, veinules ou capillaires cérébraux de moins de 400µm.

Depuis quelques années, de nombreuses maladies, tels que les cancers et les fibroses sont associées à des dysfonctionnements du matrisome. Et en effet, le matrisome cérébrovasculaire semble être une voie pathologique convergente entre les différentes maladies des petits vaisseaux de type génétique, mais également pour les maladies de type sporadique, les symptômes cliniques étant très similaires. En effet pour une multitude de maladies génétiques, la caractéristique commune est la présence de remaniements importants de la MEC des vaisseaux cérébraux. L'utilisation de la spectrométrie de masse aujourd'hui pour l'étude protéomique d'une pathologie dont on sait peu de chose, était une évidence.

Le cadre de ce projet a consisté à caractériser de manière quantitative et qualitative le matrisome microvasculaire dans les maladies des petits vaisseaux cérébraux, ainsi que d'identifier des anomalies communes ou spécifique à chaque maladie.

Pour cela nous avons développé une approche protéomique quantitative de type label-free sur des vaisseaux cérébraux et périphériques isolés et nous l'avons appliqué sur trois modèles murins précliniques de CADASIL, CARASIL et de la maladie du collagène de type IV pour leur caractère génétique et deux modèles murins pour l'hypertension et l'âge, des modèles pour le caractère sporadique.

Le premier point de cette étude a été donc de développer et de valider une nouvelle méthode de préparation et d'analyse des vaisseaux cérébraux de souris en protéomique, afin d'identifier et de quantifier de manière robuste et la plus complète possible le matrisome de ces vaisseaux. Pour chaque étude et chaque analyse nous avons établi des points de contrôle, afin de s'assurer que les données soient fiables et utilisables pour une quantification sans marquage (Annexe2).

La mise au point de la méthode du traitement de l'échantillon, qui a consisté à tester différents tampons de solubilisation, ainsi que leur concentration, différents supports de digestion et à optimiser la quantité d'échantillon minimale nécessaire nous a pris quelques mois (et quelques cerveaux !). Nous avons très vite compris que la manipulation à outrance d'un échantillon de cette dimension était très pénalisante en termes de rendement de digestion mais également et surtout en termes de robustesse pour obtenir une analyse quantitative sans marquage non biaisée. Nous avons donc simplifié au maximum le protocole de digestion,

en réduisant les étapes, en minimisant le nombre de contenants, notamment en ajoutant successivement les réactifs dans un unique et même tube jusqu'à l'étape de dessalage. Un travail considérable a été réalisé également sur le traitement des données obtenus après les analyses sur le spectromètre de masse. A l'image du protocole du traitement de l'échantillon, toutes les étapes, de la normalisation aux statistiques, sont contrôlées. L'utilisation de la programmation R a permis de réaliser tout le traitement dans un seul logiciel développé à façon, après l'écriture d'un programme contenant les divers modules nécessaires.

Le second point a consisté à étudier les changements du matrisome cérébrovasculaire dans les maladies des petits vaisseaux, en appliquant la nouvelle méthode mise au point pour chaque maladie, mais pour identifier les points communs et les spécificités potentiels des différentes maladies.

Nous avons eu la chance de pouvoir commencer à travailler sur une maladie relativement bien documentée, CADASIL, avec de nombreuses publications et surtout ayant un modèle clinique de souris très bien caractérisé. Ce qui a d'ailleurs permis de valider notre méthode de préparation des petits vaisseaux cérébraux murins.

Et cette méthode nous a permis d'identifier quelques pistes intéressantes pour chaque maladie de manière indépendante mais a surtout mis en évidence quelques signatures communes entre les différentes études. En effet, nous avons identifié une protéine commune à toutes ces formes de cSVDs, PRSS23, une serine protéase, dans les cinq modèles étudiés, bien qu'elle ne soit pas identifiée de manière significative dans les analyses des souris hypertendues avec une  $q$ value à 0.06. HTRA1, une autre serine protéase, est également surexprimée, dans CADASIL, la maladie du collagène de type IV et l'effet de l'âge. Si les fonctions de ces deux protéines sont encore très méconnues, il est à noter que ce sont toutes deux des serine protéases. Néanmoins aucune validation orthogonale n'a encore été effectuée pour cause d'absence d'anticorps.

Malgré le travail réalisé, il reste de nombreux points à améliorer aussi bien du point de vue analytique que sur la partie biologique.

Ce travail nous a permis d'identifier de nombreuses protéines, dont deux sur l'ensemble des résultats des différentes maladies étudiées et qui pour l'instant n'ont malheureusement pas été validées par une méthode alternative telle qu'un Western blot pour des raisons d'absence d'anticorps. Outre le fait que tous les anticorps en vente ne sont pas tous fiables<sup>270</sup>, de nombreuses autres raisons telles qu'une modification insoupçonné de la taille de la protéine (interaction protéine-protéine, lyse partielle) ou modification de la configuration de la structure par l'utilisation de détergent, peuvent rendre le transfert en Western blot difficile voir erroné. De plus, ce type de validation doit être réitéré pour **chaque** protéine. Quelles autres possibilités avons-nous quand il n'y a pas d'anticorps pour valider la présence d'une protéine ? Car c'est avant tout sa présence qui est validé, avant sa variation d'expression.

Les méthodes en spectrométrie de type MRM/PRM, plus sensibles et spécifiques, pourraient confirmer la modification d'expression des protéines d'intérêt mais pas leur présence puisque leur méthode est basée sur les identifications réalisées en DDA, elles ne sont donc pas indépendantes. Cependant, sur le même principe que la quantification absolue de protéine en MRM<sup>271</sup>, nous pourrions valider la présence de la protéine d'intérêt en utilisant des peptides synthétisés lourd de cette protéine. L'avantage de cette méthode, en comparaison de l'utilisation d'anticorps, est que d'une part, toutes les séquences sont, à priori, synthétisables<sup>272</sup>, c'est plutôt en termes choix de peptides pour représenter la protéine que nous serons limités (I.3.3.4.1). D'autre part, les validations et quantifications de l'ensemble des protéines d'intérêt peuvent être réalisées dans une même analyse.

Un autre point, totalement occulté durant ce projet, a été la recherche de modification post-traductionnelle (MPTs). En effet, mis à part les modifications principalement artéfactuelles, considérées classiquement pour une recherche dans les banques de données, nous n'avons pas du tout ciblé notre étude sur les modifications post-traductionnelles potentielles et ce pour différentes raisons :

- La préparation de l'échantillon avait pour but de quantifier de manière robuste et la plus complète possible le matrisome de ces vaisseaux, donc en minimisant la manipulation de l'échantillon, or de nombreuses modifications post-traductionnelles nécessitent des manipulations supplémentaires, tel un enrichissement, pour augmenter les possibilités de les identifier.
- La quantification a été réalisée dès le départ sur l'information de l'intensité totale de la protéine or, pour une étude des MPTs, il est nécessaire de travailler directement sur l'information de l'intensité des peptides.
- De plus, si nous souhaitons intégrer les MPTs, nous devons également les quantifier. Or la quantification des MPTs nécessite également un traitement des données de spectrométrie de masse relativement complexe. En effet, la quantification des MPTs doit toujours être liée au niveau d'expression de la protéine à laquelle elle est associée<sup>273</sup>.

Une étude des MPTs pour ce type d'étude demanderait à elle seule un autre projet de thèse.

Du point de vue biologique, si l'artère est un bon modèle pour caractériser la pathologie par son matrisome, l'information est incomplète. En effet, une analyse des vaisseaux intracérébraux serait un complément pour cette étude. Une autre validation essentielle à ce travail est bien évidemment une analyse sur des vaisseaux humains, et dans la mesure du possible, à différents stades de la maladie.

Ce projet de thèse m'a permis d'ajouter une ligne de plus dans mon expertise en protéomique par spectrométrie de masse. La prochaine étape serait très probablement de tester et sûrement d'optimiser ce protocole sur un autre type de tissu.



# Annexes

---



## **Annexe I : Article 1**

---





## Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain?

Anne Joutel<sup>1,2</sup>, Iman Haddad<sup>3</sup>, Julien Ratelade<sup>1,2</sup> and Mark T Nelson<sup>4,5</sup>

### Abstract

The term matrisome refers to the ensemble of proteins constituting the extracellular matrix (ECM) (core matrisome) as well as the proteins associated with the ECM. Every organ has an ECM with a unique composition that not only provides the support and anchorage for cells, but also controls fundamental cellular processes as diverse as differentiation, survival, proliferation, and polarity. The current knowledge of the matrisome of small brain vessels is reviewed with a focus on the basement membrane (BM), a specialized form of ECM located at the interface between endothelial cells, contractile cells (smooth muscle cells and pericytes), and astrocyte endfeet—a very strategic location in the communication pathway between the cerebral microcirculation and astrocytes. We discuss some of the most recent genetic data and relevant findings from experimental models of nonamyloid cerebral small vessel disease (SVD). We propose the concept that perturbations of the cerebrovascular matrisome is a convergent pathologic pathway in monogenic forms of SVD, and is likely relevant to the sporadic disease.

### Keywords

Basement membrane, cerebral small vessel disease, collagen type IV, extracellular matrix, proteomics

Received 14 January 2015; Accepted 17 March 2015

### Introduction

Non-amyloid cerebral small vessel disease (cSVD) is a heterogeneous group of disorders that refers collectively to the pathologic processes that affect the structure and function of small intracerebral vessels. The salient consequences of nonamyloid cSVD are small, deep infarctions or hemorrhages in the white and/or deep gray matter, and the development of widespread white-matter lesions. Cerebral SVD accounts for ~25% to 30% of strokes, and ample evidence now exists that it is a leading cause of cognitive decline and disability in adults.<sup>1</sup> Cerebral SVD is a mostly sporadic disease that appears to be driven by a complex mix of genetic and environmental factors, among which age and arterial hypertension are currently deemed the most important. However, during the past 20 years, rare Mendelian forms of cSVD, largely indistinguishable from sporadic SVD, have been characterized and the genes responsible for causing them have been identified, propelling an unprecedented leap forward in the effort to unravel the pathogenesis of these related diseases. To date,

highly penetrant mutations in six distinct genes have been linked to familial cSVD, namely *NOTCH3*, *COL4A1*, *COL4A2*, *HTRA1*, *TREX1*, and *FOXCI*.<sup>2,3</sup> These genetic findings led us to ask whether the genes associated with familial forms of cSVD represent distinct mechanisms for producing the cSVD phenotype or instead ultimately converge on a common pathologic

<sup>1</sup>Department of Genetics and Pathogenesis of Cerebrovascular Diseases, INSERM, U1161 and Univ Paris Diderot, Sorbonne Paris Cité, UMR S 1161, Paris, France

<sup>2</sup>DHU NeuroVasc Sorbonne Paris Cité, Paris, France

<sup>3</sup>Department of Mass Spectrometry, USR3149, ESPCI ParisTech, Paris, France

<sup>4</sup>Department of Pharmacology, College of Medicine, University of Vermont, Burlington, Vermont, USA

<sup>5</sup>Institute of Cardiovascular Sciences, University of Manchester, Manchester, UK

### Corresponding author:

Anne Joutel, Faculté de Médecine Paris Diderot, site Villemin, 10 av de Verdun, 75010 Paris, France.  
Email: anne.joutel@inserm.fr

pathway that might also be relevant to the sporadic disease.

The extracellular matrix (ECM) consists of a complex meshwork of highly crosslinked proteins that constitutes a fundamental component of the cellular microenvironment. The ECM proteins bind to cell-surface adhesion receptors, including the integrins, evoking intracellular transduction responses.<sup>4</sup> They also have the ability to bind and regulate the distribution, activation, and presentation of a myriad of growth and secretory factors to cells.<sup>5</sup> Hence, the ECM provides much more than support and anchorage for the cell; it also has crucial roles in the differentiation, survival, proliferation, polarity, and migration of cells.

Here, we review our current understanding of the ECM in general and in the context of cerebral blood vessels in particular, and critically examine the contribution of cerebrovascular ECM alterations in nonamyloid cSVD.

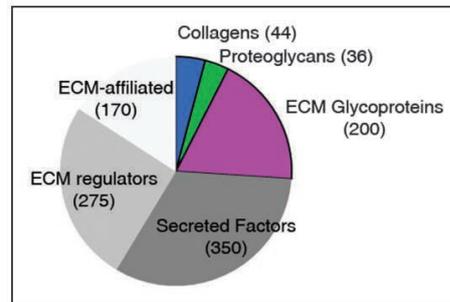
### The extracellular matrix: The core matrisome and matrisome-associated proteins

#### What Is the Matrisome?

The name 'matrisome' has been coined to designate the ensemble of proteins constituting the ECM (core matrisome) plus the proteins associated with the ECM (matrisome-associated proteins).<sup>6</sup> The ECM proteins classically contain repeats of a characteristic set of domains (e.g., EGF-like, fibronectin, laminin, and collagen triple helix repeat) that are highly conserved among different species, in both sequence and arrangement.<sup>7</sup> Using this feature, Hynes and coworkers recently developed a bioinformatics approach that they used in combination with manual annotation to define a comprehensive repertoire of genes encoding the matrisome in murine and human genomes.<sup>6</sup>

#### Matrisome Categories

The core matrisome comprises 44 collagen subunits, 36 proteoglycans, and about 200 glycoproteins in humans and mice (Figure 1). Collagens, the main structural proteins of the ECM, are classified into fibrillar (collagens I to III, V, and XI) and nonfibrillar forms. Collagen subunits are characterized by a unique sequence stretch in which every third amino acid is a glycine (Gly-X-Y motif), and the structural hallmark of all collagens is their triple helix.<sup>8</sup> Proteoglycans are glycoproteins containing attached glycosaminoglycans, which represent a significant fraction of the total mass. Proteoglycans fill the extracellular interstitial space and provide hydration by sequestering water within



**Figure 1.** Schematic representation of the *in silico* matrisome. Pie chart displaying the different matrisome categories with the approximate number of genes (brackets) determined by *in silico* analysis of human and murine genomes. ECM, extracellular matrix.

the tissue; they are also involved in binding many secreted factors, including growth factors, and sequestering them in the ECM.<sup>9</sup> Glycoproteins include a myriad of proteins with various functions; among these, laminins, fibrillins, fibronectin, and vitronectin have been extensively studied in the context of normal physiology or pathology.<sup>10–13</sup>

The matrisome-associated protein category comprises almost 800 proteins divided into three categories: (1) secreted factors, including growth factors, that are bound to the ECM; (2) ECM regulators, consisting of enzymes and other proteins that modify the ECM; and (3) ECM-affiliated proteins, which either are known to be associated with ECM proteins or share some architectural similarities with ECM proteins (Figure 1).<sup>6</sup> The ECM acts as a sink or a reservoir for almost 350 secreted factors, which bind either to glycosaminoglycans, especially heparan sulfates, or to specific domains of proteins of the core matrisome.<sup>14</sup> Among these factors are cytokines and many of the major regulator of physiologic angiogenesis, such as vascular endothelial growth factor, platelet-derived growth factor, fibroblast growth factor, and bone morphogenic proteins, among others.<sup>15</sup> The second category, comprising 250 to 300 proteins, includes modifiers of ECM structure and function, such as lysyl hydroxylase and lysyloxidase, which crosslink collagens, matrix metalloproteinases (MMPs), HTRA serine proteases, and many other proteolytic enzymes, as well as regulators of these modifiers, including members of the tissue inhibitor of metalloproteinases (TIMP) family.<sup>16</sup> The third category defines a heterogeneous set of about 170 ECM-affiliated proteins, considered ECM associated by some researchers, that includes annexins, complement components, lectins, and semaphorins, among others. In contrast with the core matrisome, the matrisome-associated

category is less firmly established, and the list of these proteins can be expected to change with subsequent experimental analyses.<sup>17</sup>

### Key Features

*In vivo* characterization of the matrisome from a given tissue or cell line is challenging because of the peculiar physical properties of ECM proteins. Specifically, these proteins generally have high molecular weights, are difficult to solubilize, and carry extensive post-translational modifications. Moreover, the presence of highly abundant cytoplasmic and mitochondrial proteins, which are easily solubilized, can hamper the detection of ECM proteins that may be present at comparatively low relative abundance. Recent proteomic profiling of ECMs in lung, liver, and colon revealed that each of these tissues contained about 100 to 150 ECM proteins.<sup>18</sup>

### The matrisome of cerebral blood vessels

The circle of Willis gives rise to three pairs of main arteries, which divide into progressively smaller arteries that run on the surface of the brain. These pial arteries branch extensively into smaller arterioles, eventually penetrating into the brain parenchyma and ultimately terminating as an extensive capillary network. Blood is then drained through venules and veins back into dural venous sinuses at the surface.<sup>19</sup>

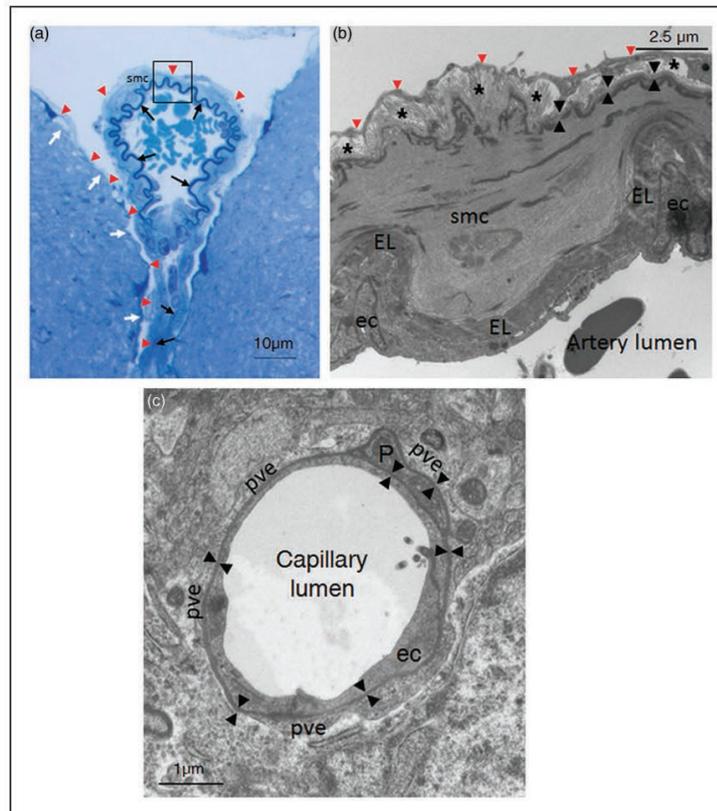
### An Histologic View

The ECM matrix of cerebral blood vessels differs with regard to the type, location within the vessel tree, and diameter of blood vessels. Pial arteries lie within the subarachnoid spaces on top of brain tissue and are covered with the pia mater. Unlike systemic arteries, these arteries are characterized by a well-developed internal elastica lamina (which separates the endothelium from the tunica media) but no external elastica lamina, a paucity of elastic fibers in the media and a very thin adventitia (Figures 2A and 2B).<sup>20</sup> Distinct basement membranes (BMs) cover the basal aspect of endothelial cells, facing the elastic lamina, and encircle smooth muscle cells (SMCs). The inner layer of the pia mater is anchored to the cortex via the glia limitans superficialis, which seals the entire surface of the brain. The glia limitans is composed of a meshwork of astrocyte processes covered by an outer BM that makes intimate contact with cells of the pia mater.<sup>21</sup> As blood vessels dip into the brain, they carry with them the extensions of the pia mater, forming a cuff around the vessel. They are separated from the parenchyma by the glia limitans perivascularis, which has a structure

nearly identical to that of the glia limitans superficialis (Figure 2A). The Virchow-Robin space containing pial processes corresponds to the perivascular space between the smooth muscle BM and the glial limitans. As the arteriole penetrates deeper into the parenchyma and morphs into the parenchymal arteriole, the elastica lamina becomes more and more discontinuous and the pial investment disappears; notably, the vascular and glial BMs fuse and the Virchow-Robin space disappears.<sup>22,23</sup> At the capillary level, one-third of the circumference of the endothelial tube surface is covered by pericytes, and both cell types are wrapped by astroglial endfeet.<sup>24</sup> Notably, there is a single, but composite, continuous BM, produced by endothelial cells, astrocytes, and possibly pericytes, that covers the basal aspect of endothelial cells and enwraps the pericytes (Figure 2C).<sup>23</sup> The ECM structure of venules and veins is similar to that of arterioles and arteries at the same location, although the SMC coat and elastica lamina of venules and veins are discontinuous and much thinner.<sup>20,22</sup>

### A Proteomic View

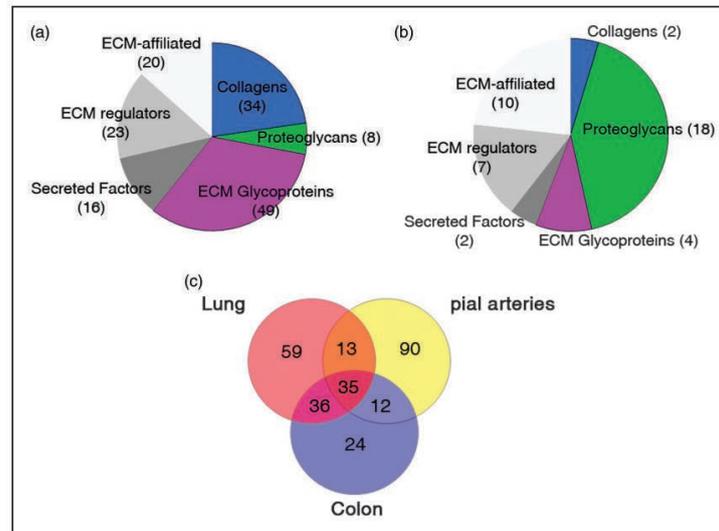
Two separate studies have analyzed the proteome of murine pial arteries and brain microvessels. The pial arteries used in the former study included arteries of the circle of Willis and their ramifications, surgically removed from the pia mater.<sup>25</sup> In the latter study, brain microvessels were isolated from the mouse cortex using low speed centrifugation in a Ficoll gradient, and lysates were subsequently processed by depleting cytosolic proteins.<sup>26</sup> Employing the commonly used gene ontology categories, 79 out of 2,188 proteins (identified by at least two distinct peptides) in the pial arteries were classified as ECM proteins, and only 27 out of 1,472 proteins (identified with a mean spectral count greater than 5) were so classified in the brain microvessels. As discussed recently by Hynes and colleagues,<sup>17</sup> gene ontology categories are inadequate for extracting a complete list of ECM proteins from a data set. Therefore, we reanalyzed these two data sets using the matrisome repertoire defined by Hynes and colleagues.<sup>6</sup> Consistent with the notion that every tissue contains about 100 to 150 ECM proteins, we found 150 ECM proteins in the pial artery data set, including 91 core matrisome proteins and 59 matrisome-associated proteins. Of interest, the majority of proteins (48 out of 59; 81%) annotated in this study as matrisome-associated proteins, according to Hynes and colleagues classification, had not been annotated as ECM proteins by gene ontology analysis in the original study (Figure 3A, Supplementary Table 1). A similar reanalysis of the brain microvessels data set yielded only 43 ECM proteins, including 24 core matrisome proteins



**Figure 2.** Histologic view of the extracellular matrix (ECM) along cerebral vessels in the mouse. (A) Semithin section of a pial artery and one of its branch penetrating into the brain parenchyma (toluidine blue staining). The elastic lamina (EL) (black arrows), the pia matter (red arrowheads), which enwraps the artery and covers the surface of the brain, and the glia limitans (white arrows) are shown. Brain tissue was fixed without prior transcardiac perfusion. (B) Electron micrograph of a pial artery (magnification of the boxed area in A). (From top to bottom) The pia matter (red arrowheads), a thin adventitia with collagen bundles (stars), a smooth muscle cell (SMC) with the basement membrane (BM) on its abluminal face (back arrowheads), the elastic lamina, and endothelial cells (ec) are shown. Note that BMs of the SMC luminal side and of ECs are not visible at this magnification. Brain tissue was fixed without prior transcardiac perfusion. (C) Electron micrograph of a capillary within the striatum. The single, continuous BM (black arrowheads) that covers the basal aspect of the EC and enwraps the pericyte (P) is shown. The BM is surrounded on its abluminal face by several perivascular astrocyte endfeet (pve). Brain tissue was postfixed after transcardiac perfusion of aldehydes. ec, endothelial cell; EL, elastic lamina; P, pericyte; smc, smooth muscle cell.

and 19 matrisome-associated proteins, suggesting that the proteomic coverage in this data set is incomplete (Figure 3B, Supplementary Table 2). A comparison of matrisome proteins from the two brain vessel data sets revealed that 24 ECM proteins were common to both pial arteries and brain microvessels. Interestingly, half of the 24 shared proteins are core components of the

BM (see next section). A comparison of matrisome proteins in pial arteries with those in lung and colon samples (identified by at least two peptides)<sup>17</sup> revealed that a majority of ECM proteins in the pial artery data set (90 out of 150; 60%) have a tissue-specific expression. Conversely, only 35 out of 269 (13%) ECM proteins are found in all three tissues (Figure 3C).



**Figure 3.** Characterization of the matrisome of cerebral blood vessels. (A, B) Pie charts displaying the number of matrisome proteins in murine pial arteries (A) and brain microvessels (B) determined by proteomic analysis<sup>25, 26</sup> and reclassified using the matrisome database.<sup>6</sup> (C) Venn diagram showing the number of matrisome proteins that overlap between pial arteries, lung, and colon samples. ECM, extracellular matrix.

### The Basement Membrane: A Key Specialized Form of Extracellular Matrix at the Interface Between the Vessel and Brain Parenchyma

On electron micrographs, all BMs are classically described as composed of a 50- to 100-nm-thick electron-dense layer (lamina densa) separated from the plasma membrane by an electron-lucent layer (lamina lucida). However, when tissues are processed by freeze substitution, the BM appears as a fairly uniform lamina densa that closely follows the contours of the plasma membrane, suggesting that the electron-lucent layer is a dehydration artifact.<sup>27</sup> All BMs are composed of a common set of interacting proteins, including at least one member of the laminin family, one or more variants of type IV collagens, heparan sulfate proteoglycans, and nidogens.<sup>28</sup>

**Laminins.** Laminins are large (400 to 800 kDa), cross-shaped heterotrimeric molecules consisting of  $\alpha$ ,  $\beta$ , and  $\gamma$  chains joined through a long coiled-coil domain. There are five  $\alpha$  chains, four  $\beta$  chains, and three  $\gamma$  chains; the laminin molecules are named according to their chain composition, and a total of 16 chain combinations have been characterized biochemically.<sup>29</sup> The laminin  $\alpha$  chains carry the major

domains that interact with cellular receptors and are considered the functionally active portion of the heterotrimer. Moreover,  $\alpha$  chains have tissue-specific expression patterns. In cerebral blood vessels, laminin  $\alpha 1$  is associated with the pia matter and the glia limitans, and is detected around pial and penetrating arteries but not around capillaries.<sup>30</sup> Laminin  $\alpha 2$  is synthesized by astrocytes and deposited at the glia limitans, although it is unclear whether it is also synthesized by SMCs.<sup>31</sup> Laminin  $\alpha 4$  and  $\alpha 5$  chains are detected in the endothelial and smooth muscle BM of all cerebral blood vessels, with a discontinuous distribution at the level of venules.<sup>32</sup> These  $\alpha$  chains can assemble with  $\beta 1$  or  $\beta 2$  and  $\gamma 1$  to form Laminins-111, -121, -211, -221, 411, -421, -511, and -521. Laminin  $\alpha 3$  is not expressed in cerebral blood vessels.<sup>10,30</sup>

**Type IV collagens.** Type IV collagens, the most abundant constituent of the BM, comprise up to six genetically distinct  $\alpha$  chains designated  $\alpha 1(\text{IV})$  to  $\alpha 6(\text{IV})$ ;  $\alpha$  chains assemble with a remarkable specificity to form only three distinct heterotrimers,  $[\alpha 1\alpha 1\alpha 2]$ ,  $[\alpha 3\alpha 4\alpha 5]$ , and  $[\alpha 5\alpha 5\alpha 6]$ , with  $[\alpha 1\alpha 1\alpha 2]$  being the predominant variant found in nearly all BMs, including those of cerebral blood vessels. The  $[\alpha 5\alpha 5\alpha 6]$  variant has been reported in association with the pia matter.<sup>33</sup> Type IV collagens

are triple-helical molecules that twist along their length and terminate in a globular noncollagenous (NC1) domain. The glycine residues within the characteristic Gly-X-Y repeats have a crucial role in helix formation and stabilization.<sup>34</sup>

**Heparan sulfate proteoglycans.** Heparan sulfate proteoglycans are glycoproteins with the common characteristic of covalently attached chains of heparan sulfate, a type of glycosaminoglycan. Heparan sulfate proteoglycans present in the BM include perlecan, agrin, and type XVIII collagen, which are large secreted proteins composed of multiple, functionally independent domains. They bind to many growth factors, cytokines, chemokines, enzymes, enzyme inhibitors, and ECM proteins, usually with sulfated domains within heparan sulfate chains but also with the protein core. Perlecan, agrin, and type XVIII collagen can also undergo proteolytic processing to release bioactive molecules.<sup>9</sup> Nidogen-1 and -2 are sulfated monomeric glycoproteins that interact with a range of ECM proteins, in particular, laminins, collagen type IV, and perlecan.<sup>35</sup>

**Basement membrane assembly.** The BM assembles through a multistep process that is initiated by the binding of laminin to cell-surface proteins such as integrins and dystroglycan, promoting laminin polymerization and allowing recruitment of other laminin binding components into the nascent laminin scaffold. Laminin is the only component that is critical for early embryonic BM assembly.<sup>36,37</sup> Type IV collagen polymerizes to form a second covalently stabilized network that is considered to provide structural stability to the BM. The laminin and type IV collagen networks are linked by heparan sulfate proteoglycans<sup>38</sup> and nidogens.<sup>29</sup> This scaffold then provides specific sites at which many other BM constituents interact to generate a fully functional BM.

### Familial small vessel disease caused by mutations in genes encoding matrisome proteins

#### Collagen Type IV-Related Small Vessel Disease

**Pathogenic mutations in COL4A1 and COL4A2.** As noted above, COL4A1 and COL4A2 are members of the core matrisome and major components of nearly all BMs, including those of pial and brain microvessels. Dominant mutations in COL4A1 or COL4A2 are associated with typical features of cSVD and a broad range of other disorders that most commonly involve the brain and eyes, but also affect the kidney and skeletal muscle, in both humans and mice.<sup>39</sup> Cerebral SVD manifestations consist of microhemorrhages and

macrohemorrhages that can occur at virtually every age, patchy to diffuse leukoaraiosis, dilated periventricular spaces and, less frequently, small deep infarcts. Other brain features are porencephaly and schizencephaly, which typically manifests in infants with epilepsy, motor deficits or psychomotor retardation.<sup>40</sup> Common eye features consist of dysgenesis of the anterior segment (cornea, iris, ciliary body, and lens), including Axenfeld-Rieger anomaly, congenital cataract, and juvenile-onset glaucoma.<sup>2</sup> Although there are no reported neuropathologic studies of patients with COL4A1 or COL4A2 mutations, electron microscopy analyses of skin or kidney biopsies have shown focal defects of the BMs of vascular and epithelial cells, including areas of fragmentation, duplication, and even herniation.<sup>41</sup> Similar defects have been described in the BM of cerebral blood vessels from Col4a1<sup>+/Δex41</sup> mice, a well-characterized mouse model that expresses a mutant collagen alpha-1(IV) chain with a 17-amino-acid in-frame deletion.<sup>42,43</sup> A large proportion of COL4A1 and COL4A2 mutations are missense mutations resulting in the substitution for one of the invariant glycine residues within the Gly-Xaa-Yaa repeats in the collagenous domain of  $\alpha 1$  or  $\alpha 2$  chains.<sup>39</sup> Expression of COL4A1 and COL4A2 mutations is extremely variable, even between members of the same family, with part of this variability apparently related to genetic-context and environmental factors. Accumulating evidence in humans and mice suggests that allelic heterogeneity also contributes to this variable expression.<sup>43,44</sup> Specifically, in humans, mutations clustered within 31 amino-acid residues in the N-terminus of the collagenous domain of COL4A1 are associated with a clinical subentity called HANAC (Hereditary Angiopathy with Nephropathy, Aneurysms, and Cramps) a syndrome characterized by co-association of arterial aneurysms and eye, kidney, and skeletal manifestations.<sup>41</sup> In mice, mutations within the triple-helical domain nearer the carboxy terminus are associated with more severe intracerebral hemorrhages than mutations nearer the amino terminus.<sup>43</sup>

**Collagen type IV-related small vessel disease: a basement membrane disease?** The  $\alpha$ -chain stoichiometry of collagen type IV predicts that heterozygous COL4A1 and COL4A2 mutations result in the production of 75% and 50% abnormal collagen IV molecules, respectively, a prediction that is consistent with the observation that COL4A2 mutations result in a disease of reduced severity compared with COL4A1 mutations.<sup>45</sup> Several lines of evidence suggest that the pathogenicity of COL4A1 and COL4A2 mutations arises from a reduced incorporation of collagen IV in the BM. First, cells stably expressing mutant COL4A1 have reduced secretion of

$\alpha 1$  (IV) chain as well as  $\alpha 2$  (IV) chain, and the same holds true for cells expressing mutant COL4A2.<sup>39</sup> Second, deposition of collagen  $\alpha 1$  and  $\alpha 2$  (IV) is almost absent in the Reichert's membrane, a BM of embryonic origin, in embryos homozygous for the Col4a1 $\Delta$ ex41 mutation.<sup>46</sup> Also, Col4a1<sup>+/ $\Delta$ ex41</sup> mice have reduced extracellular COL4A1 in the BM of retinal vessels, but increased intracellular retention of COL4A1 in vascular cells.<sup>43</sup> Third, an apparent lack of expression of mutated COL4A1 transcript has been documented in the fibroblasts of affected patients from two families carrying a frameshift and a splice site COL4A1 mutation, respectively, because of nonsense-mediated mRNA decay.<sup>47</sup> However, the lack of an overt clinical phenotype in mice heterozygous for null alleles of Col4a1 and Col4a2 suggests that haploinsufficiency may not solely account for the disease phenotype.<sup>48</sup> One possibility is that mutant molecules may reach the extracellular space and exert a dominant-negative effect over the various partners of collagen IV that could also compromise BM integrity and, in addition, promote abnormal cell matrix and cell cell interactions.

Collagen type IV interacts with all three main components of the BM, namely laminin, perlecan, and nidogen. Collagen type IV-like perlecan and nidogens are dispensable for initiation of BM assembly, but are required for maintaining BM integrity.<sup>48–50</sup> Immunohistochemical analyses of embryos completely lacking COL4A1 and COL4A2 show reduced deposition and a patchier distribution of laminin and nidogen.<sup>48</sup> Whether similar effects occur in patients and mice with COL4A1 or COL4A2 mutations has yet to be investigated. The role of the BM, especially that of collagen type IV, in brain vessels is essentially unknown. Nevertheless, a recent analysis of mice lacking laminin chains provides some clues. In mice with ablation of astrocytic laminin  $\gamma 1$ , astrocytic endfoot (parenchymal) BM formation is disrupted, the blood brain barrier (BBB) is profoundly altered and multifocal hemorrhages develop in deep brain regions. Mechanistic studies show that loss of astrocytic laminin  $\gamma 1$  impairs SMC differentiation, with decreased expression of contractile proteins in small arteries and arterioles in deep brain regions, impaired pericyte differentiation, loss of astrocytic endfeet polarity, and a reduction in endothelial tight junction protein expression.<sup>51,52</sup> Mice with constitutive inactivation of laminin  $\alpha 2$  chain (dy3K/dy3K),<sup>53</sup> which die at 4 weeks of age, exhibit breaches in the parenchymal and endothelial BM and have profound alterations in BBB development and function with limited pericyte coverage, poorly polarized astrocyte endfeet, and immature endothelial tight junctions.<sup>54</sup> It is tempting to speculate that COL4A1 mutations might similarly affect key components of the gliovascular unit.

Collagen type IV also participates in cell-matrix interactions via cell-surface receptors; specifically, it binds to integrins containing a  $\beta 1$  subunit, notably  $\alpha 1\beta 1$  and  $\alpha 2\beta 1$ .<sup>55</sup> Remarkably, HANAC syndrome-associated COL4A1 mutations are clustered within 30 amino-acid residues in the N-terminus of the collagenous domain, which contains a major integrin binding site.<sup>56</sup> Interestingly, loss of Integrin  $\beta 1$  in SMCs results in progressive smooth muscle degeneration and increased deposition of fibrillary collagens in the tunica media.<sup>57</sup> Moreover, mice with postnatal inactivation of Integrin  $\beta 1$  in endothelial cells exhibit junctional defects in the endothelium, loss of vascular integrity and brain hemorrhages.<sup>58</sup>

Another possibility besides defects of the BM and cell-matrix interaction, suggested by both *in vitro* and *in vivo* data, is that COL4A1 or COL4A2 mutations promote intracellular retention of unfolded or misfolded mutant  $\alpha 1$  and  $\alpha 2$  chains and activate the unfolded protein response.<sup>43,45,59</sup> Prolonged endoplasmic reticulum stress can eventually trigger cell death, oxidative stress, or potentially elicit an inflammatory reaction.<sup>60,61</sup> Interestingly, Gould and colleagues recently showed in a proof of concept study that treating Col4a1<sup>+/ $\Delta$ ex41</sup> mice with an FDA-approved chaperone (Sodium 4-phenylbutyrate) could decrease intracellular COL4A1 accumulation and, notably, reduce intracerebral hemorrhage severity.<sup>43</sup>

## CARASIL

### Pathogenic mutations in HTRA1

HTRA1 is a 50-kDa secreted protein composed of a signaling peptide, an insulin growth factor binding domain, a Kazal-like protease inhibitor domain, a conserved serine protease domain, and a PDZ domain. HTRA1, classified as an ECM regulator in the matrixome-associated category, has been reported to be widely expressed, although a detailed analysis of its expression pattern in the brain has not yet been performed. Proteomic studies indicate that HTRA1 is expressed in brain vessels, with higher expression in pial arteries than in brain microvessels.<sup>25,26</sup> Recessive mutations in HTRA1 cause CARASIL, a rare SVD initially reported in Japanese families. Clinical landmarks include cognitive deficit, motor dysfunction, and subcortical stroke leading to dementia and severe disability in the fourth decade, in association with diffuse alopecia and attacks of severe low back pain due to spondylosis deformans and disk degeneration.<sup>62</sup> Microvascular changes in CARASIL consist of extensive loss of SMCs and deposition of fibro-hyaline material in the media, fibrous intimal proliferation, and splitting and fragmentation of the internal elastica

lamina, with many vessels ultimately developing a double-barreled or split-wall appearance.<sup>63</sup> To date, very few mutations have been reported; these include homozygous nonsense mutations as well as homozygous missense mutations in the serine protease domain, resulting in the loss of HTRA1 protein expression or protease activity.<sup>64</sup>

#### **CARASIL: Dysregulation of transforming growth factor- $\beta$ signaling?**

Although it has been shown that HTRA1 is involved in various pathologic conditions in addition to CARASIL, its physiologic role remains unclear. HTRA1 possesses serine protease activity, and a variety of potential substrates have been identified.<sup>65</sup> Members of the transforming growth factor- $\beta$  (TGF- $\beta$ ) signaling pathway have received particular attention, but experimental studies have generated conflicting results. Early studies reported that HTRA1 inhibited TGF- $\beta$  family members, showing that the inhibitory effects on TGF- $\beta$  signaling required the proteolytic activity of HTRA1.<sup>66</sup> Subsequent studies suggested that HTRA1 antagonizes TGF- $\beta$  signaling by cleaving pro TGF- $\beta$ 1 intracellularly in the reticulum endoplasmic,<sup>67</sup> or by cleaving type II and type III TGF- $\beta$  receptors extracellularly.<sup>68</sup> These findings, combined with immunohistochemical analyses of postmortem brain samples from two CARASIL patients, led to the hypothesis that CARASIL is associated with upregulation of TGF- $\beta$  signaling in cerebral vessels.<sup>67</sup> However, these earlier studies were not without limitations, especially their use of cells that vastly overexpressed HTRA1. Recently, Beaufort and colleagues revisited this issue using embryonic fibroblasts from HTRA1-null mice and *in vitro* assays using physiologic quantities of HTRA1. Remarkably, they came to very different conclusions. Specifically, they provided evidence that HTRA1 has a facilitating role in TGF- $\beta$  pathway activation, and that LTBP1 (latent TGF- $\beta$  binding protein 1), an ECM protein with a key role in TGF- $\beta$  bioactivation,<sup>69</sup> is a physiologic target of HTRA1.<sup>70</sup> Further studies are warranted to confirm these results in brain vessels, especially since the levels of active TGF- $\beta$ , which is itself a matrisome protein, are tightly controlled by a number of matrisome proteins and the composition of the matrisome is unique to each tissue.

There is ample evidence that TGF- $\beta$  signaling has a key role in vessel development and maintenance, acting in both endothelial and SMCs. Moreover, both up- and downregulation of this pathway have been implicated in several vascular pathologies.<sup>71</sup> For example, hereditary hemorrhagic telangiectasia, or Osler Rendu Weber syndrome, is caused by loss-of-function mutations in members of the TGF- $\beta$  pathway.

However, enhanced TGF- $\beta$  signaling has emerged as the common final pathway in the pathogenesis of familial forms of thoracic aortic aneurysm.<sup>72</sup> Whether dysregulated TGF- $\beta$  signaling contributes to the cerebrovascular manifestations of CARASIL has yet to be investigated.

#### **Perturbations of the matrisome as a possible proximate cause in familial small vessel disease**

##### **CADASIL**

*Early perturbations of the cerebrovascular matrisome.* CADASIL is an archetypal SVD, considered the most frequent hereditary cSVD.<sup>73</sup> The disease is caused by dominant mutations in NOTCH3, a heterodimeric receptor predominantly expressed in vascular SMCs that is a critical regulator of developmental formation of small arteries.<sup>74</sup> The pathology of cerebral vessels is roughly similar to that of sporadic forms of SVD except that CADASIL is characterized by the presence of pathognomonic deposits of granular osmiophilic material within the BM of SMCs and pericytes that can protrude into the interstitial tissue around these cells.<sup>73</sup> The majority of CADASIL mutations are missense mutations that lead to the gain or loss of a cysteine residue in the NOTCH3 extracellular domain (Notch3<sup>ECD</sup>).<sup>75,76</sup> Regardless of whether they concomitantly impair NOTCH3 receptor activity, these mutations alter Notch3<sup>ECD</sup> itself in a way that promotes its multimerization and aberrant accumulation in granular osmiophilic material.<sup>77</sup> Importantly, vascular Notch3<sup>ECD</sup> and granular osmiophilic material deposits have emerged as the earliest pathologic features in patients and mice with CADASIL.<sup>78</sup> Recently, we showed that excess levels or multimerization of mutant Notch3<sup>ECD</sup> facilitates interactions with key components of the cerebrovascular ECM and promotes their accumulation and sequestration in Notch3<sup>ECD</sup>-containing deposits. Among these components are the metalloproteinase inhibitor TIMP3, a matrisome-associated regulator that modifies the ECM, and vitronectin, a glycoprotein of the core matrisome. Both TIMP3 and vitronectin have been shown to abnormally accumulate in the brain vessels not only of patients but also of preclinical models, at an early stage of the disease. Moreover, we found evidence for increased TIMP3 activity in brain microvessels from both TgNotch3<sup>R169C</sup> mice, a well-established preclinical model of CADASIL, and patients with CADASIL.<sup>77</sup> Other ECM proteins, such as clusterin, endostatin (a proteolytic fragment derived from collagen XVIII), and LTBP1, have been shown to accumulate in Notch3<sup>ECD</sup>-containing deposits in patients with CADASIL.<sup>79,80</sup>

However, these latter proteins appear to accumulate at a later stage of disease progression (AJ, unpublished).<sup>77</sup>

### **CADASIL: Cerebrovascular dysfunction arising from a change in the matrisome?**

These matrisome changes may affect the BBB; however, it is yet unclear whether BBB disruption has a role in the pathogenesis of CADASIL. Histopathologic studies on human postmortem tissues showed focal breaches of the BBB at the sites of microbleeds or infarcts, but no evidence of generalized BBB leakage at this end stage of the disease.<sup>81</sup> Moreover, immunohistochemical and transmission electron microscopy analyses, as well as systemic injections of fluorescent tracers failed to reveal alterations of the BBB integrity in TgNotch3<sup>R169C</sup> mice (JR and AJ, unpublished data).<sup>82,83</sup>

However, CADASIL patients exhibit alterations in cerebral hemodynamics and cerebrovascular reactivity.<sup>73</sup> Studies of CADASIL mouse models suggest that cerebrovascular dysfunction is an early disease mechanism in the pathophysiology of CADASIL.<sup>78</sup> Specifically, cerebral blood flow autoregulation and functional hyperemia are impaired and myogenic responses of pial and parenchymal arteries are reduced in TgNotch3<sup>R169C</sup> mice.<sup>82,84</sup> A recent study from Dabertrand and colleagues suggests that increased TIMP3 activity might be a proximate cause of cerebrovascular deficits in this preclinical model of CADASIL. TIMP3 is the only TIMP that becomes tightly bound to the ECM after secretion and is thus in a position to exert tissue-specific effects.<sup>85,86</sup> Notably, TIMP3 is the only TIMP that can inhibit TACE (tumor necrosis factor- $\alpha$ -converting enzyme), which mediates ectodomain shedding of transmembrane receptors and ligands, including tumor necrosis factor- $\alpha$  and ligands of the epidermal growth factor receptor family (EGFR).<sup>87</sup> Specifically, Dabertrand et al.<sup>84</sup> showed that reduced myogenic responses of pial and parenchymal arteries in the TgNotch3<sup>R169C</sup> mice were caused by functional upregulation of Kv1 channels at the plasma membrane of arterial SMC, and further showed that these defects could be rescued by applying the EGFR agonist HB-EGF. Given that activation of EGFR with the agonist HB-EGF has been reported to suppress KV1 channel activity in cerebral arteries by enhancing KV1 channel endocytosis,<sup>88</sup> we speculate that increased TIMP3 activity may lead to decreased myogenic tone in brain arteries by reducing TACE/HB-EGF/EGFR mediated K<sub>v</sub> channel endocytosis and increasing K<sub>v</sub> channel activity.

The TIMP3 dysregulation may also contribute to small vessel pathology through its well-known MMP inhibitory activity, and abnormally elevated TIMP3 activity may lead to vessel fibrosis. Also, several reports

have shown that high levels of TIMP3 are proapoptotic in a variety of cell types, including vascular SMCs, and the importance of TIMP3 has been shown in inflammatory pathologies and autoimmune diseases.<sup>86,89</sup> Additional studies are warranted to test these hypotheses and investigate whether and how other perturbations of the matrisome, such as the abnormal accumulation of vitronectin, contribute to CADASIL pathophysiology.

### *FOXC1-Associated Small Vessel Disease*

*Pathogenic mutations in a transcription factor expressed in blood vessels.* Autosomal-dominant mutations or gene-dosage changes, including deletions or duplications, in FOXC1 result in a variety of malformations in the anterior ocular segment that ultimately lead to glaucoma and malformations of the cerebellar posterior fossa.<sup>90,91</sup> Recently, French et al.<sup>3</sup> reported that patients with FOXC1-attributable Axenfeld-Rieger syndrome have a higher incidence of stroke and highly penetrant magnetic resonance imaging features of cSVD, including white-matter hyperintensities, dilated perivascular spaces, and lacunes.<sup>3</sup> FOXC1, a member of the forkhead family of transcription factors, has a broad expression pattern and regulates, in a dose-dependent manner, an array of fundamental processes, including the early development of the heart, blood vessels and somites, as well as ocular and genitourinary systems.<sup>92,93</sup> FOXC1 is expressed in the pericytes and endothelial cells of blood vessels in the brain, and in the periocular mesenchyme of the eye. In the mouse, conditional deletion of Foxc1 in pericytes produces multifocal brain microhemorrhages toward the end of gestation, preceded by focal breakdown of BBB integrity and hyperplasia of both endothelial cells and pericytes, suggestive of increased angiogenic activity.<sup>94</sup> Furthermore, conditional deletion of Foxc1 in neural crest cells, which give rise to most of the periocular mesenchyme, phenocopies the eye anomalies associated with the human disease Axenfeld-Rieger syndrome, notably including corneal angiogenesis; by contrast, the normal cornea is avascular.<sup>95</sup>

### *FOXC1 mutations and the matrisome?*

At first glance, there is no overt relationship between FOXC1 mutations and the matrisome. Nevertheless, studies in neural crest-deleted *Foxc1*<sup>-/-</sup> mice provide one possible mechanistic link between matrisome perturbations and the vascular phenotype induced by FOXC1 mutations. Specifically, Seo et al.<sup>95</sup> found that expression of the MMPs, MMP3, MMP9, and MMP19 all ECM regulators of the matrisome-associated category is increased in the cornea of

neural crest-deleted *Foxc1*<sup>-/-</sup> mice. Notably, these data suggest that the increased expression of MMPs, which are known to cleave and release ECM-sequestered vascular endothelial growth factor and enhance vascular endothelial growth factor bioavailability,<sup>96</sup> is likely a contributing factor to the enhanced angiogenesis in the cornea of mutant mice. A second clue comes from experiments in zebrafish, where knockdown of *Foxc1* produces similar eye abnormalities and brain hemorrhages.<sup>97</sup> Interestingly, loss of *foxC1* leads to focal disruptions of vascular BM integrity, with impaired expression of laminin 1-1-1. These findings have led us to speculate that FOXC1-related cSVD manifestations actually arise from vascular BM defects through deregulation of the availability of growth factors.

### **Matrisome changes in sporadic small vessel disease: Consequent pathologic findings or proximate cause?**

#### *Pathologic Findings*

Cerebrovascular changes associated with age and vascular risk factors are commonly referred to as hyaline, fibrohyaline, or fibrinoid necrosis, although the meanings of each of these terms are somewhat different.<sup>98</sup> A closer look at histopathologic data reveals that vessel changes consist of major alterations of the ECM associated with a more or less severe degeneration of SMCs from the tunica media, with or without exudation of plasma proteins within the vessel wall. Changes of ECM may range from thickening of the BMs, splitting and duplication of the elastic lamina and, notably, marked deposition of normally occurring types of collagens and possibly other components of the ECM in the intima, media, and adventitia.<sup>99,100</sup> That said, it is difficult to draw conclusions about the primary site and mechanism of these ECM changes since they are postmortem observations at the end stage of the disease. In particular, it is conceivable that all or part of the increased ECM deposition is a compensatory response to SMC degeneration that serves as a repair mechanism.

#### *Matrisome Changes: Proximate Cause?*

The microvascular changes in CARASIL or CADASIL are generally considered to be qualitatively similar to those observed in sporadic SVD; however, the reverse is also true, meaning that specific perturbations of the matrisome can produce such vascular lesions. Moreover, experimental studies examining the effects of aging and arterial hypertension the two most important risk factors for cSVD have highlighted vascular ECM changes at the early stage of the disease in

mouse or rat models. During aging, cerebral arterioles undergo SMC atrophy and exhibit an increase in the ratio of nondistensible (collagen and BM) ECM to distensible (smooth muscle and elastin) components.<sup>101</sup> Hypertension commonly produces narrowing of the arterial lumen (inward remodeling), especially at the level of small distal arteries, with or without medial thickening (hypertrophy). Inward remodeling has been extensively documented in both peripheral and cerebral arteries, and angiotensin II has been shown to have a key role in this effect.<sup>102,103</sup> Inward remodeling is thought to result from both changes in SMC contractility and in ECM composition, with a potential contribution of MMP activation.<sup>104</sup> Reorganization of the ECM, with an increase in deposition of collagens and fibronectin in the tunica media, an increase in collagen-to-elastin ratio, and disorganization and fragmentation of elastic fibers, has been clearly documented in the peripheral arteries of models with chronic hypertension.<sup>105</sup> However, evidence of ECM changes at the level of brain arteries remains scarce,<sup>106,107</sup> although it should be noted that this aspect has been so far poorly investigated.

The dominant view is that arterial remodeling during chronic hypertension represents both an adaptive process that reduces stress on the vessel wall and protects the downstream vascular bed from increased arterial pressure, and a deleterious process that ultimately results in arterial stenosis and stiffness, which may contribute to reductions in resting cerebral blood flow, impaired vasodilation, and increased susceptibility to ischemic injury.<sup>108</sup> Increased deposition of ECM in cerebral vessels, whether proximate cause or consequent pathologic finding, is likely to contribute to the long-term deleterious effects of hypertension.

### **Concluding remarks**

Cerebral SVD is a heterogeneous group of disorders that are likely multifactorial in nature with different ultimate causes acting through specific pathways. Nevertheless, with recent advances in identifying major genetic causes and pathophysiologic mechanisms of familial cSVD has come the emerging view that perturbation of the cerebrovascular matrisome could be a convergent pathway that drives the functional and structural alterations of small brain vessels and disease manifestations. Specifically, in collagen type IV-related SVD and CARASIL, genetic alterations of a matrisome protein are clearly the primary event. In CADASIL and FOXC1-related SVD, the data suggest that specific perturbations of the matrisome are the proximate cause. Importantly, studies in animals and humans support the possibility that this may also apply to age- and hypertension-related SVD. Despite these

provocative connections, a better characterization of matrisome alterations using more sensitive tools and further experimental studies are needed to strengthen the link between matrisome alterations and cSVD.

Though unexplained, it is not unusual for mutations in a gene with a vascular expression broader than the brain, for example NOTCH3, to lead to a brain-restricted phenotype. Also, within the brain, it is essentially unknown why lesions of cSVD predominantly affect small penetrating arteries and parenchymal arterioles, particularly in the deep cerebral white matter. The vessel-specific expression pattern of the matrisome may hold the key to these unanswered questions. Every organ has an ECM with a unique composition, and it is very likely that brain and peripheral arteries have distinct matrisomes. On the basis of a preliminary characterization of the matrisome of pial arteries and brain microvessels, it is tempting to speculate that penetrating arteries, intraparenchymal arterioles, and brain capillaries also harbor unique repertoires of matrisome proteins. Moreover, unlike pial arteries, penetrating and parenchymal arterioles are bottlenecks of flow in the cerebral angioarchitecture.<sup>109</sup> Therefore, it is conceivable that lesions could affect brain vessels with a specific spatial pattern if expression of a specific set of matrisome proteins is required for the pathogenic process. Accordingly, identification of those matrisome proteins with a vessel-specific expression pattern might provide insight into differential manifestations of vascular pathologies.

A significant proportion of matrisome proteins is presumably expressed within the vascular BM, especially at the capillary level. The BM is located at the interface between endothelial cells, contractile cells (SMCs and pericytes), and astrocyte endfeet a very strategic location within the gliovascular unit. Therefore, changes in the composition of the BM are in a position to exert profound effects on all components of the gliovascular unit as well as on the brain parenchyma itself that may account for the broad range of manifestations in cSVD.

In conclusion, the ECM of cerebral blood vessels is a key component of the gliovascular unit that warrants greater research attention in the future, in both physiologic and pathologic settings.

#### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported in part by a grant from the Fondation Leducq (Transatlantic Network of Excellence on the Pathogenesis of Small Vessel Disease of the Brain) to AJ and MTN, National Institutes of Health (P01-HL-095488, R01-HL-044455, R01-HL-098243) and Totman Medical Research Trust to MTN.

#### Acknowledgement

The authors gratefully acknowledge David Hill-Eubanks for editorial assistance.

#### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

#### Authors' contributions

AJ and MTN wrote the manuscript and generated the figures. IH and JR provided assistance in preparing the manuscript and IH performed the proteomic analysis.

#### Supplementary Material

Supplementary material for this paper can be found at <http://jebfm.sagepub.com/content/by/supplemental-data>

#### References

- Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol* 2010; 9: 689–701.
- Joutel A. Pathogenic aspects of hereditary small vessel disease of the brain. In: Pantoni L and Gorelick PB (eds) *Cerebral Small Vessel Disease*. Cambridge University Press, 2014, pp.64–81.
- French CR, Seshadri S, Destefano AL, Fornage M, Arnold CR, Gage PJ, et al. Mutation of FOXC1 and PITX2 induces cerebral small-vessel disease. *J Clin Invest* 2014; 124: 4877–4881.
- Hynes RO. Integrins: bidirectional, allosteric signaling machines. *Cell* 2002; 110: 673–687.
- Hynes RO. The extracellular matrix: not just pretty fibrils. *Science* 2009; 326: 1216–1219.
- Naba A, Clauser KR, Hoersch S, Liu H, Carr SA and Hynes RO. The matrisome: in silico definition and *in vivo* characterization by proteomics of normal and tumor extracellular matrices. *Mol Cell Proteomics* 2012; 11: M111.014647.
- Jung J, Ryu T, Hwang Y, Lee E and Lee D. Prediction of extracellular matrix proteins based on distinctive sequence and domain characteristics. *J Comput Biol* 2010; 17: 97–105.
- Ricard-Blum S. The collagen family. *Cold Spring Harb Perspect Biol* 2011; 3: a004978.
- Sarrazin S, Lamanna WC and Esko JD. Heparan sulfate proteoglycans. *Cold Spring Harb Perspect Biol* 2011; 3: pii:a004952.
- Hallmann R, Horn N, Selg M, Wendler O, Pausch F and Sorokin LM. Expression and function of laminins in the embryonic and mature vasculature. *Physiol Rev* 2005; 85: 979–1000.
- Ramirez F and Dietz HC. Fibrillin-rich microfibrils: Structural determinants of morphogenetic and homeostatic events. *J Cell Physiol* 2007; 213: 326–330.

12. Preissner KT and Reuning U. Vitronectin in vascular context: facets of a multitaled matricellular protein. *Semin Thromb Hemost* 2011; 37: 408–424.
13. Schwarzbauer JE and DeSimone DW. Fibronectins, their fibrillogenesis, and *in vivo* functions. *Cold Spring Harb Perspect Biol* 2011; 3: pii:a005041.
14. Hynes RO and Naba A. Overview of the matrisome—an inventory of extracellular matrix constituents and functions. *Cold Spring Harb Perspect Biol* 2012; 4: a004903.
15. Potente M, Gerhardt H and Carmeliet P. Basic and therapeutic aspects of angiogenesis. *Cell* 2011; 146: 873–887.
16. Bonnans C, Chou J and Werb Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* 2014; 15: 786–801.
17. Naba A, Hoersch S and Hynes RO. Towards definition of an ECM parts list: an advance on GO categories. *Matrix Biol* 2012; 31: 371–372.
18. The Matrisome Project <http://matrisomeproject.mit.edu/>
19. Cipolla MJ. The cerebral circulation. In: Granger DN and Granger J (eds) *Integrated Systems Physiology: From Molecule to Function*. San Rafael, CA: Morgan & Claypool Life Sciences, 2010, pp.1–59.
20. Frederickson RG and Low FN. Blood vessels and tissue space associated with the brain of the rat. *Am J Anat* 1969; 125: 123–145.
21. Peters A, Palay SL and de Webster HF. The meninges. *The Fine Structure of the Nervous System: Neurons and Their Supporting Cells*. Oxford University Press, 1991, pp.395–406.
22. Krisch B, Leonhardt H and Oksche A. Compartments and perivascular arrangement of the meninges covering the cerebral cortex of the rat. *Cell Tissue Res* 1984; 238: 459–474.
23. Zhang ET, Inman CB and Weller RO. Interrelationships of the pia mater and the perivascular (Virchow-Robin) spaces in the human cerebrum. *J Anat* 1990; 170: 111–123.
24. Mathiesen TM, Lehre KP, Danbolt NC and Ottersen OP. The perivascular astroglial sheath provides a complete covering of the brain microvessels: An electron microscopic 3D reconstruction. *Glia* 2010; 58: 1094–1103.
25. Badhwar A, Stanimirovic DB, Hamel E and Haqqani AS. The proteome of mouse cerebral arteries. *J Cereb Blood Flow Metab* 2014; 34: 1033–1046.
26. Chun HB, Scott M, Niessen S, Hoover H, Baird A, Yates J, et al. The proteome of mouse brain microvessel membranes and basal lamina. *J Cereb Blood Flow Metab* 2011; 31: 2267–2281.
27. Chan FL, Inoue S and Leblond CP. The basement membranes of cryofixed or aldehyde-fixed, freeze-substituted tissues are composed of a lamina densa and do not contain a lamina lucida. *Cell Tissue Res* 1993; 273: 41–52.
28. Yurchenco PD. Basement membranes: cell scaffoldings and signaling platforms. *Cold Spring Harb Perspect Biol* 2011; 3: pii:a004911.
29. Hohenester E and Yurchenco PD. Laminins in basement membrane assembly. *Cell Adh Migr* 2013; 7: 56–63.
30. Sixt M, Engelhardt B, Pausch F, Hallmann R, Wendler O and Sorokin LM. Endothelial cell laminin isoforms, laminins 8 and 10, play decisive roles in T cell recruitment across the blood-brain barrier in experimental autoimmune encephalomyelitis. *J Cell Biol* 2001; 153: 933–946.
31. Yousif LF, Di Russo J and Sorokin L. Laminin isoforms in endothelial and perivascular basement membranes. *Cell Adh Migr* 2013; 7: 101–110.
32. Wu C, Ivars F, Anderson P, Hallmann R, Vestweber D, Nilsson P, et al. Endothelial basement membrane laminin alpha5 selectively inhibits T lymphocyte extravasation into the brain. *Nat Med* 2009; 15: 519–527.
33. Urabe N, Naito I, Saito K, Yonezawa T, Sado Y, Yoshioka H, et al. Basement membrane type IV collagen molecules in the choroid plexus, pia mater and capillaries in the mouse brain. *Arch Histol Cytol* 2002; 65: 133–143.
34. Khoshnoodi J, Pedchenko V and Hudson BG. Mammalian collagen IV. *Microsc Res Tech* 2008; 71: 357–370.
35. Ho MSP, Böse K, Mokkaapati S, Nischt R and Smyth N. Nidogens-Extracellular matrix linker molecules. *Microsc Res Tech* 2008; 71: 387–395.
36. Smyth N, Vatansver HS, Murray P, Meyer M, Frie C, Paulsson M, et al. Absence of basement membranes after targeting the LAMC1 gene results in embryonic lethality due to failure of endoderm differentiation. *J Cell Biol* 1999; 144: 151–160.
37. Miner JH, Li C, Mudd JL, Go G and Sutherland AE. Compositional and structural requirements for laminin and basement membranes during mouse embryo implantation and gastrulation. *Development* 2004; 131: 2247–2256.
38. Behrens DT, Villone D, Koch M, Brunner G, Sorokin L, Robenek H, et al. The epidermal basement membrane is a composite of separate laminin- or collagen IV-containing networks connected by aggregated perlecan, but not by nidogens. *J Biol Chem* 2012; 287: 18700–18709.
39. Kuo DS, Labelle-Dumais C and Gould DB. COL4A1 and COL4A2 mutations and disease: insights into pathogenic mechanisms and potential therapeutic targets. *Hum Mol Genet* 2012; 21: R97–R110.
40. Yoneda Y, Haginoya K, Kato M, Osaka H, Yokochi K, Arai H, et al. Phenotypic spectrum of COL4A1 mutations: porencephaly to schizencephaly. *Ann Neurol* 2013; 73: 48–57.
41. Plaisier E, Gribouval O, Alamowitch S, Mougnot B, Prost C, Verpont MC, et al. COL4A1 mutations and hereditary angiopathy, nephropathy, aneurysms, and muscle cramps. *N Engl J Med* 2007; 357: 2687–2695.
42. Gould DB, Phalan FC, van Mil SE, Sundberg JP, Vahedi K, Massin P, et al. Role of COL4A1 in small-vessel disease and hemorrhagic stroke. *N Engl J Med* 2006; 354: 1489–1496.
43. Jeanne M, Jorgensen J and Gould DB. Molecular and genetic analysis of collagen type IV mutant mouse models of spontaneous intracerebral hemorrhage identify mechanisms for stroke prevention. *Circulation* 2015; 131: 1555–1565.
44. Kuo DS, Labelle-Dumais C, Mao M, Jeanne M, Kauffman WB, Allen J, et al. Allelic heterogeneity contributes to variability in ocular dysgenesis, myopathy and

- brain malformations caused by Col4a1 and Col4a2 mutations. *Hum Mol Genet* 2014; 23: 1709–1722.
45. Murray LS, Lu Y, Taggart A, Van Regemorter N, Vilain C, Abramowicz M, et al. Chemical chaperone treatment reduces intracellular accumulation of mutant collagen IV and ameliorates the cellular phenotype of a COL4A2 mutation that causes haemorrhagic stroke. *Hum Mol Genet* 2014; 23: 283–292.
  46. Gould DB, Phalan FC, Breedveld GJ, van Mil SE, Smith RS, Schimenti JC, et al. Mutations in Col4a1 cause perinatal cerebral hemorrhage and porencephaly. *Science* 2005; 308: 1167–1171.
  47. Lemmens R, Maugeri A, Niessen HWM, Goris A, Tousseyn T, Demaerel P, et al. Novel COL4A1 mutations cause cerebral small vessel disease by haploinsufficiency. *Hum Mol Genet* 2013; 22: 391–397.
  48. Pöschl E, Schlötzer-Schrehardt U, Brachvogel B, Saito K, Ninomiya Y and Mayer U. Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development. *Development* 2004; 131: 1619–1628.
  49. Costell M, Gustafsson E, Aszódi A, Mörgelein M, Bloch W, Hunziker E, et al. Perlecan maintains the integrity of cartilage and some basement membranes. *J Cell Biol* 1999; 147: 1109–1122.
  50. Bader BL, Smyth N, Nedbal S, Miosge N, Baranowsky A, Mokkapatil S, et al. Compound genetic ablation of nidogen 1 and 2 causes basement membrane defects and perinatal lethality in mice. *Mol Cell Biol* 2005; 25: 6846–6856.
  51. Chen Z-L, Yao Y, Norris EH, Kruyer A, Jno-Charles O, Akhmerov A, et al. Ablation of astrocytic laminin impairs vascular smooth muscle cell function and leads to hemorrhagic stroke. *J Cell Biol* 2013; 202: 381–395.
  52. Yao Y, Chen Z-L, Norris EH and Strickland S. Astrocytic laminin regulates pericyte differentiation and maintains blood brain barrier integrity. *Nat Commun* 2014; 5: 3413.
  53. Miyagoe Y, Hanaoka K, Nonaka I, Hayasaka M, Nabeshima Y, Arahata K, et al. Laminin alpha2 chain-null mutant mice by targeted disruption of the Lama2 gene: a new model of merosin (laminin 2)-deficient congenital muscular dystrophy. *FEBS Lett* 1997; 415: 33–39.
  54. Menezes MJ, McClenahan FK, Leiton CV, Aranmolate A, Shan X and Colognato H. The extracellular matrix protein laminin  $\alpha 2$  regulates the maturation and function of the blood-brain barrier. *J Neurosci* 2014; 34: 15260–15280.
  55. Leitinger B and Hohenester E. Mammalian collagen receptors. *Matrix Biol* 2007; 26: 146–155.
  56. Parkin JD, San Antonio JD, Pedchenko V, Hudson B, Jensen ST and Savage J. Mapping structural landmarks, ligand binding sites, and missense mutations to the collagen IV heterotrimer predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes. *Hum Mutat* 2011; 32: 127–143.
  57. Turlo KA, Scapa J, Bagher P, Jones AW, Feil R, Korthuis RJ, et al.  $\beta 1$ -integrin is essential for vasoregulation and smooth muscle survival *in vivo*. *Arterioscler Thromb Vasc Biol* 2013; 33: 2325–2335.
  58. Yamamoto H, Ehling M, Kato K, Kanai K, van Lessen M, Frye M, et al. Integrin  $\beta 1$  controls VE-cadherin localization and blood vessel stability. *Nat Commun* 2015; 6: 6429.
  59. Gould DB, Marchant JK, Savinova OV, Smith RS and John SWM. Col4a1 mutation causes endoplasmic reticulum stress and genetically modifiable ocular dysgenesis. *Hum Mol Genet* 2007; 16: 798–807.
  60. Tabas I and Ron D. Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress. *Nat Cell Biol* 2011; 13: 184–190.
  61. Chaudhari N, Talwar P, Parimisetty A, Lefebvre d'Hellencourt C and Ravanan P. A molecular web: endoplasmic reticulum stress, inflammation, and oxidative stress. *Front Cell Neurosci* 2014; 8: 213.
  62. Hara K, Shiga A, Fukutake T, Nozaki H, Miyashita A, Yokoseki A, et al. Association of HTRA1 mutations and familial ischemic cerebral small-vessel disease. *N Engl J Med* 2009; 360: 1729–1739.
  63. Oide T, Nakayama H, Yanagawa S, Ito N, Ikeda S-I and Arima K. Extensive loss of arterial medial smooth muscle cells and mural extracellular matrix in cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL). *Neuropathology* 2008; 28: 132–142.
  64. Nozaki H, Nishizawa M and Onodera O. Features of cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy. *Stroke* 2014; 45: 3447–3453.
  65. An E, Sen S, Park SK, Gordish-Dressman H and Hathout Y. Identification of novel substrates for the serine protease HTRA1 in the human RPE secretome. *Invest Ophthalmol Vis Sci* 2010; 51: 3379–3386.
  66. Oka C, Tsujimoto R, Kajikawa M, Koshiba-Takeuchi K, Ina J, Yano M, et al. HtrA1 serine protease inhibits signaling mediated by Tgfbeta family proteins. *Dev* 2004; 131: 1041–1053.
  67. Shiga A, Nozaki H, Yokoseki A, Nihonmatsu M, Kawata H, Kato T, et al. Cerebral small-vessel disease protein HTRA1 controls the amount of TGF- $\beta$ 1 via cleavage of proTGF- $\beta$ 1. *Hum Mol Genet* 2011; 20: 1800–1810.
  68. Graham JR, Chamberland A, Lin Q, Li XJ, Dai D, Zeng W, et al. Serine protease HTRA1 antagonizes transforming growth factor- $\beta$  signaling by cleaving its receptors and loss of HTRA1 *in vivo* enhances bone formation. *PLoS ONE* 2013; 8: e74094.
  69. Ge G and Greenspan DS. BMP1 controls TGF $\beta$ 1 activation via cleavage of latent TGF $\beta$ -binding protein. *J Cell Biol* 2006; 175: 111–120.
  70. Beaufort N, Scharrer E, Kremmer E, Lux V, Ehrmann M, Huber R, et al. Cerebral small vessel disease-related protease HtrA1 processes latent TGF- $\beta$  binding protein 1 and facilitates TGF- $\beta$  signaling. *Proc Natl Acad Sci USA* 2014; 111: 16496–16501.
  71. Jakobsson L and van Meeteren LA. Transforming growth factor  $\beta$  family members in regulation of vascular

- function: in the light of vascular conditional knockouts. *Exp Cell Res* 2013; 319: 1264–1270.
72. Gillis E, Van Laer L and Loeyls BL. Genetics of thoracic aortic aneurysm: at the crossroad of transforming growth factor- $\beta$  signaling and vascular smooth muscle cell contractility. *Circ Res* 2013; 113: 327–340.
  73. Chabriat H, Joutel A, Dichgans M, Tournier-Lasserre E and Boussier MG. CadASIL. *Lancet Neurol* 2009; 8: 643–653.
  74. Fouillade C, Monet-Lepretre M, Baron-Menguy C and Joutel A. Notch signalling in smooth muscle cells during development and disease. *Cardiovasc Res* 2012; 95: 138–146.
  75. Joutel A, Vahedi K, Corpechot C, Troesch A, Chabriat H, Vayssiere C, et al. Strong clustering and stereotyped nature of Notch3 mutations in CADASIL patients. *Lancet* 1997; 350: 1511–1515.
  76. Dichgans M, Ludwig H, Muller-Hocker J, Messerschmidt A and Gasser T. Small in-frame deletions and missense mutations in CADASIL: 3D models predict misfolding of Notch3 EGF-like repeat domains. *Eur J Hum Genet* 2000; 8: 280–285.
  77. Monet-Lepretre M, Haddad I, Baron-Menguy C, Fouillot-Panchal M, Riani M, Domenga-Denier V, et al. Abnormal recruitment of extracellular matrix proteins by excess Notch3 ECD: a new pathomechanism in CADASIL. *Brain* 2013; 136: 1830–1845.
  78. Joutel A. Pathogenesis of CADASIL: transgenic and knock-out mice to probe function and dysfunction of the mutated gene, Notch3, in the cerebrovasculature. *Bioessays* 2011; 33: 73–80.
  79. Arboleda-Velasquez JF, Manent J, Lee JH, Tikka S, Ospina C, Vanderburg CR, et al. Hypomorphic Notch 3 alleles link Notch signaling to ischemic cerebral small-vessel disease. *Proc Natl Acad Sci USA* 2011; 108: E128–E135.
  80. Kast J, Hanecker P, Beaufort N, Giese A, Joutel A, Dichgans M, et al. Sequestration of latent TGF- $\beta$  binding protein 1 into CADASIL-related Notch3-ECD deposits. *Acta Neuropathol Commun* 2014; 2: 96.
  81. Tikka S, Baumann M, Siitonen M, Pasanen P, Pöyhönen M, Myllykangas L, et al. CADASIL and CARASIL. *Brain Pathol* 2014; 24: 525–544.
  82. Joutel A, Monet-Lepretre M, Gosele C, Baron-Menguy C, Hammes A, Schmidt S, et al. Cerebrovascular dysfunction and microcirculation rarefaction precede white matter lesions in a mouse genetic model of cerebral ischemic small vessel disease. *J Clin Invest* 2010; 120: 433–445.
  83. Cognat E, Cleophax S, Domenga-Denier V and Joutel A. Early white matter changes in CADASIL: evidence of segmental intramyelinic oedema in a pre-clinical mouse model. *Acta Neuropathol Commun* 2014; 2: 49.
  84. Dabertrand F, Kroigaard C, Bonev AD, Cognat E, Dalsgaard T, Domenga-Denier V, et al. Potassium channelopathy-like defect underlies early-stage cerebrovascular dysfunction in a genetic model of small vessel disease. *Proc Natl Acad Sci USA* 2015; 112: E796–E805.
  85. Brew K and Nagase H. The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with structural and functional diversity. *Biochim Biophys Acta* 2010; 1803: 55–71.
  86. Khokha R, Murthy A and Weiss A. Metalloproteinases and their natural inhibitors in inflammation and immunity. *Nat Rev Immunol* 2013; 13: 649–665.
  87. Nagase H, Visse R and Murphy G. Structure and function of matrix metalloproteinases and TIMPs. *Cardiovasc Res* 2006; 69: 562–573.
  88. Koide M, Penar PL, Tranmer BI and Wellman GC. Heparin-binding EGF-like growth factor mediates oxy-hemoglobin-induced suppression of voltage-dependent potassium channels in rabbit cerebral artery myocytes. *Am J Physiol Heart Circ Physiol* 2007; 293: H1750–H1759.
  89. Baker AH, Zaltsman AB, George SJ and Newby AC. Divergent effects of tissue inhibitor of metalloproteinase-1, -2, or -3 overexpression on rat vascular smooth muscle cell invasion, proliferation, and death *in vitro*. TIMP-3 promotes apoptosis. *J Clin Invest* 1998; 101: 1478–1487.
  90. Nishimura DY, Searby CC, Alward WL, Walton D, Craig JE, Mackey DA, et al. A spectrum of FOXC1 mutations suggests gene dosage as a mechanism for developmental defects of the anterior chamber of the eye. *Am J Hum Genet* 2001; 68: 364–372.
  91. Aldinger KA, Lehmann OJ, Hudgins L, Chizhikov VV, Bassuk AG, Ades LC, et al. FOXC1 is required for normal cerebellar development and is a major contributor to chromosome 6p25.3 Dandy-Walker malformation. *Nat Genet* 2009; 41: 1037–1042.
  92. Kume T, Deng KY, Winfrey V, Gould DB, Walter MA and Hogan BL. The forkhead/winged helix gene Mfl is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. *Cell* 1998; 93: 985–996.
  93. Winnier GE, Kume T, Deng K, Rogers R, Bundy J, Raines C, et al. Roles for the winged helix transcription factors MF1 and MFH1 in cardiovascular development revealed by nonallelic noncomplementation of null alleles. *Dev Biol* 1999; 213: 418–431.
  94. Siegenthaler JA, Choe Y, Patterson KP, Hsieh I, Li D, Jaminet S-C, et al. Foxc1 is required by pericytes during fetal brain angiogenesis. *Biol Open* 2013; 2: 647–659.
  95. Seo S, Singh HP, Lacial PM, Sasman A, Fatima A, Liu T, et al. Forkhead box transcription factor FoxC1 preserves corneal transparency by regulating vascular growth. *Proc Natl Acad Sci USA* 2012; 109: 2015–2020.
  96. Ferrara N. Binding to the extracellular matrix and proteolytic processing: two key mechanisms regulating vascular endothelial growth factor action. *Mol Biol Cell* 2010; 21: 687–690.
  97. Skarie JM and Link BA. FoxC1 is essential for vascular basement membrane integrity and hyaloid vessel morphogenesis. *Invest Ophthalmol Vis Sci* 2009; 50: 5026–5034.
  98. Ogata J, Yamanishi H and Ishibashi-Ueda H. Pathology of cerebral small vessel disease. In: Pantoni L and Gorelick PB (eds) *Cerebral Small Vessel Disease*. Cambridge University Press, 2014, pp.4–15.
  99. Farkas E, de Vos RAI, Donka G, Jansen Steur EN, Mihály A and Luiten PGM. Age-related microvascular degeneration in the human cerebral periventricular white matter. *Acta Neuropathol (Berl)* 2006; 111: 150–157.

100. Zhang WW and Olsson Y. The angiopathy of subcortical arteriosclerotic encephalopathy (Binswanger's disease): immunohistochemical studies using markers for components of extracellular matrix, smooth muscle actin and endothelial cells. *Acta Neuropathol (Berl)* 1997; 93: 219–224.
101. Hajdu MA, Heistad DD, Siems JE and Baumbach GL. Effects of aging on mechanics and composition of cerebral arterioles in rats. *Circ Res* 1990; 66: 1747–1754.
102. Intengan HD and Schiffrin EL. Structure and mechanical properties of resistance arteries in hypertension: role of adhesion molecules and extracellular matrix determinants. *Hypertension* 2000; 36: 312–318.
103. Pires PW, Dams Ramos CM, Matin N and Dorrance AM. The effects of hypertension on the cerebral circulation. *Am J Physiol Heart Circ Physiol* 2013; 304: H1598–H1614.
104. Martínez-Lemus LA, Zhao G, Galiñanes EL and Boone M. Inward remodeling of resistance arteries requires reactive oxygen species-dependent activation of matrix metalloproteinases. *Am J Physiol Heart Circ Physiol* 2011; 300: H2005–H2015.
105. Lemarié CA, Tharaux P-L and Lehoux S. Extracellular matrix alterations in hypertensive vascular remodeling. *J Mol Cell Cardiol* 2010; 48: 433–439.
106. Izzard AS, Graham D, Burnham MP, Heerkens EH, Dominiczak AF and Heagerty AM. Myogenic and structural properties of cerebral arteries from the stroke-prone spontaneously hypertensive rat. *Am J Physiol Heart Circ Physiol* 2003; 285: H1489–H1494.
107. Baumbach GL, Walmsley JG and Hart MN. Composition and mechanics of cerebral arterioles in hypertensive rats. *Am J Pathol* 1988; 133: 464–471.
108. Iadecola C and Davisson RL. Hypertension and cerebrovascular dysfunction. *Cell Metab* 2008; 7: 476–484.
109. Nishimura N, Schaffer CB, Friedman B, Lyden PD and Kleinfeld D. Penetrating arterioles are a bottleneck in the perfusion of neocortex. *Proc Natl Acad Sci USA* 2007; 104: 365–370.



## Annexe 2 : Informations complémentaires

### III.3 Validation des analyses et identification des protéines dans les petits vaisseaux cérébraux des souris CADASIL

Table 8 : Table récapitulative des analyses MSMS pour l'étude CADASIL

Les valeurs pour chaque échantillon est l'ensemble des résultats pour deux analyses. Pour

Sample	# MS/MS	# PSM	% id.	# Peptides id.	# Proteins id.	# Proteins id. with 2pep
TgNotch3 <sup>WT</sup> 1	180951	58328	32%	20867	3123	2227
TgNotch3 <sup>WT</sup> 2	176903	55560	31%	20264	3063	2170
TgNotch3 <sup>WT</sup> 3	189452	52016	27%	24570	2851	2073
TgNotch3 <sup>WT</sup> 4	194974	60091	31%	22784	3205	2297
TgNotch3 <sup>R169C</sup> 1	192908	58216	30%	21211	3078	2181
TgNotch3 <sup>R169C</sup> 2	189577	56007	30%	20825	3165	2232
TgNotch3 <sup>R169C</sup> 3	193151	62807	33%	23112	3382	2456
TgNotch3 <sup>R169C</sup> 4	196367	56645	29%	21240	3194	2270

chaque échantillon, nous avons le nombre d'acquisitions MS / MS (189285 ± 3%), le nombre de spectres validés (PSM pour Peptide Match Spectrum) (57459 ± 5%), ce qui donne environ 30% d'identification (% id) avec un nombre d'identification de protéines avec au moins 2 peptides (2 238 ± 5%).

Le gain de protéines obtenu par l'utilisation du module de transfert d'identification *match between run* est représenté en figure 27

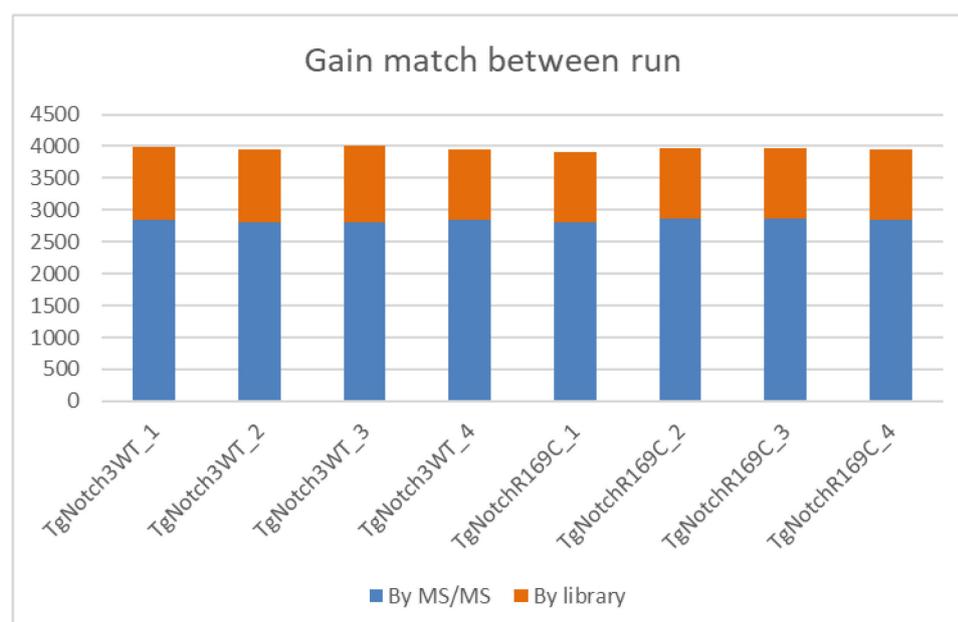


Figure 27: Nombre de protéines identifiées par MS/MS et par transfert d'identification pour l'étude de la maladie CADASIL

La distribution des  $\log_2$ (intensités des protéines), ainsi que les multiscatterplot et leur coefficient de corrélation de Pearson est construits pour chaque comparaison d'échantillons pour s'assurer l'homogénéité du groupe d'échantillons.

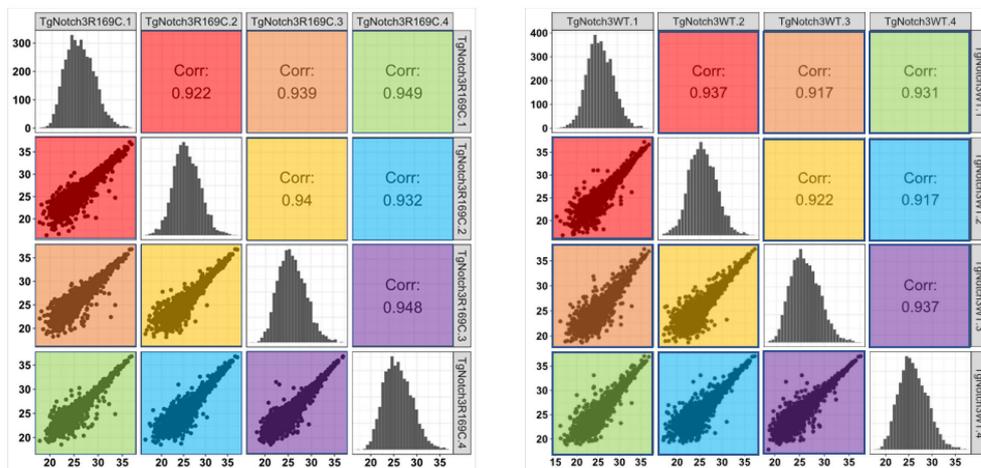


Figure 28 : Multiscatterplot et distribution de  $\log_2$  (intensités) des souris  $TgNotch3^{wt}$  et  $TgNotch3^{R169C}$

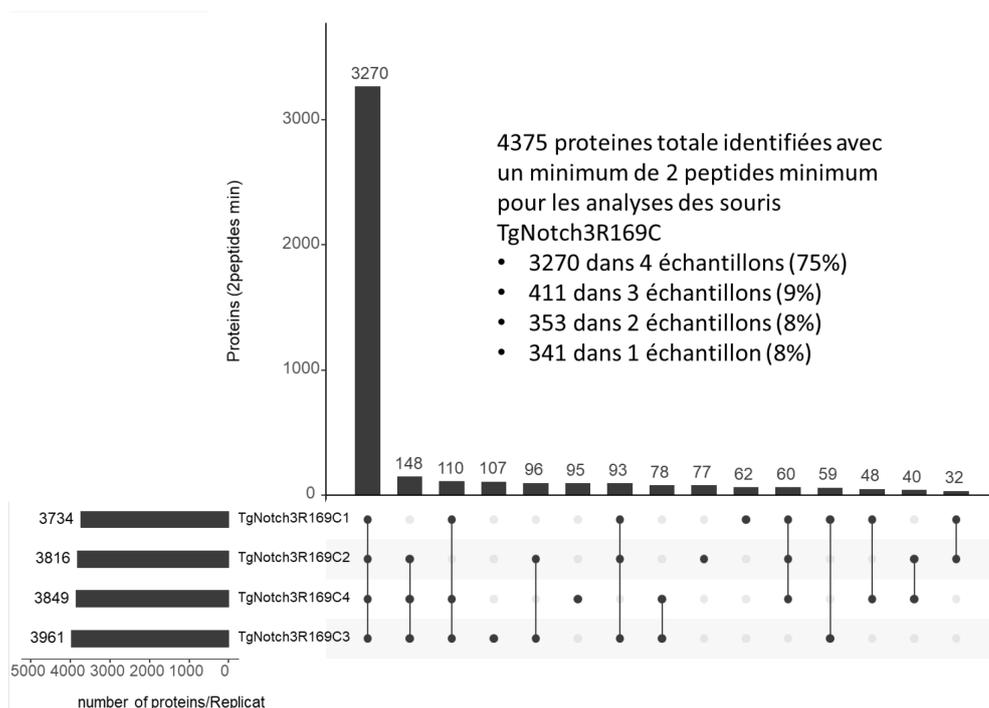


Figure 29 : Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris  $TgNotch3^{R169C}$

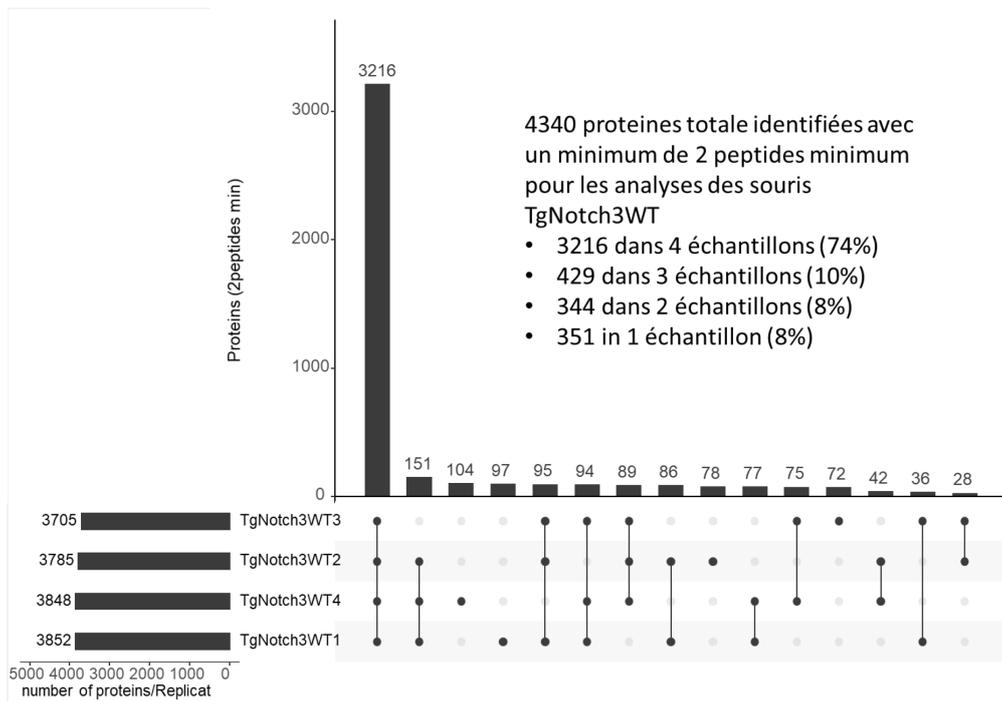


Figure 30 : Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris TgNotch3<sup>WT</sup>

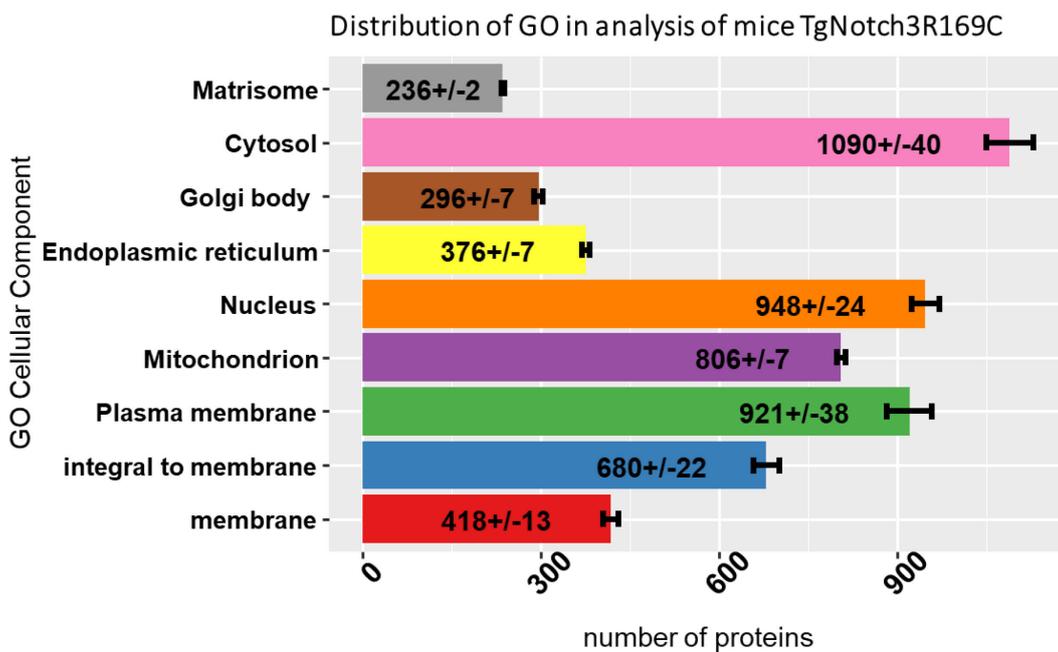


Figure 29 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris TgNotch3<sup>R169C</sup>

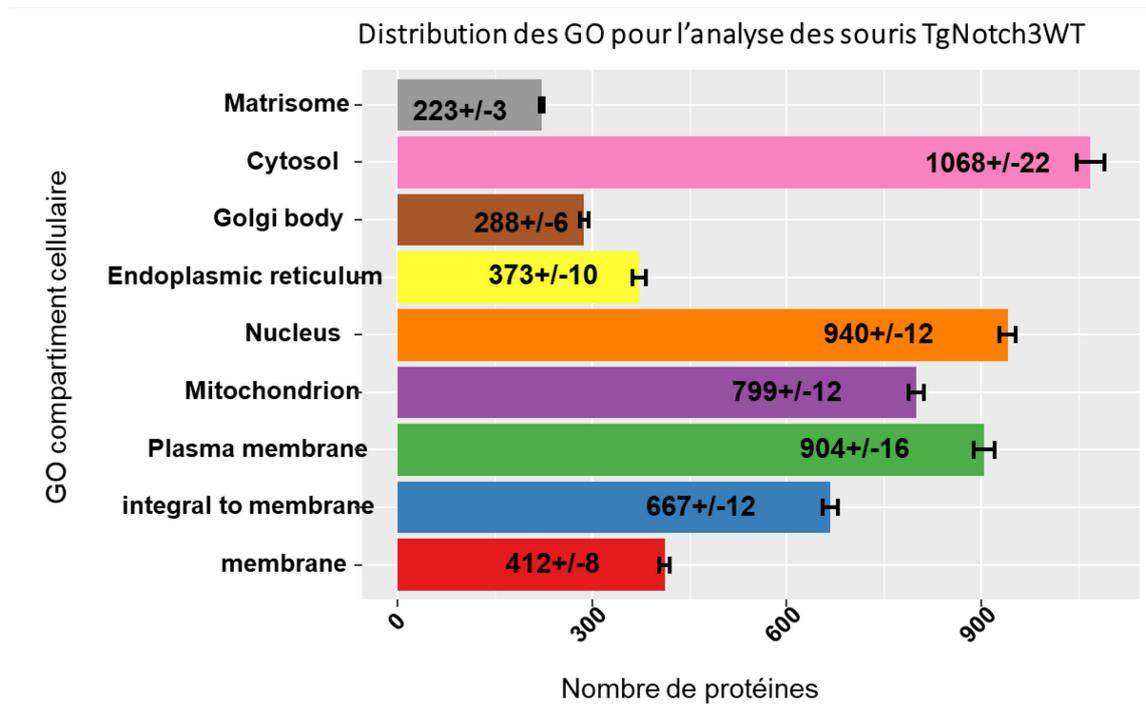


Figure 31 La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris gNotch3<sup>wt</sup>

### III.4 Validation des analyses et identification des protéines dans les petits vaisseaux cérébraux des souris CARASIL

Pour des raisons pratiques d'écriture pour cette étude, les échantillons provenant du modèle murin  $HTRA1^{Gt(OST394864)Lex}$  sera remplacé par CAR, et  $HTRA1^{+/+}$  par WT. L'échantillon CAR2 a par la suite été retiré du groupe CAR (erreur dans le génotype de cette souris). Les valeurs pour chaque échantillon est l'ensemble des résultats pour deux analyses. Pour chaque échantillon, nous avons le nombre d'acquisitions MS / MS ( $183168 \pm 5\%$ ), le nombre de spectres validés (PSM pour Peptide Match Spectrum) ( $59649 \pm 12\%$ ), ce qui donne environ 30% d'identification (% id.) avec un nombre d'identification de protéines avec au moins 2 peptides ( $2444 \pm 12\%$ ).

Table 9 : Table récapitulative des analyses MSMS pour l'étude CARASIL

Samples	# MS/MS	# PSM	% id.	# Peptides id.	# Proteins id.	# Proteins id. with 2pep
CAR1	172411	66049	38 %	23722	3271	2316
CAR2	191820	61619	32 %	22139	3542	2599
CAR3	191596	58016	30 %	20677	3458	2507
CAR4	173328	68555	40 %	25436	3548	2494
CAR5	187059	54551	29 %	19124	3224	2313
CAR6	185601	69252	37 %	23888	3767	2748
WT1	183312	60318	32 %	22147	3447	2597
WT2	191665	60617	31 %	22101	3594	2638
WT3	186861	54030	29 %	19523	3290	2433
WT4	182802	58483	32 %	20712	3409	2494
WT5	158954	39849	25 %	15115	2325	1567
WT6	192603	64446	33 %	22338	3664	2631

Le gain de protéines obtenu par l'utilisation du module de transfert d'identification *match between run* est représenté en figure 32

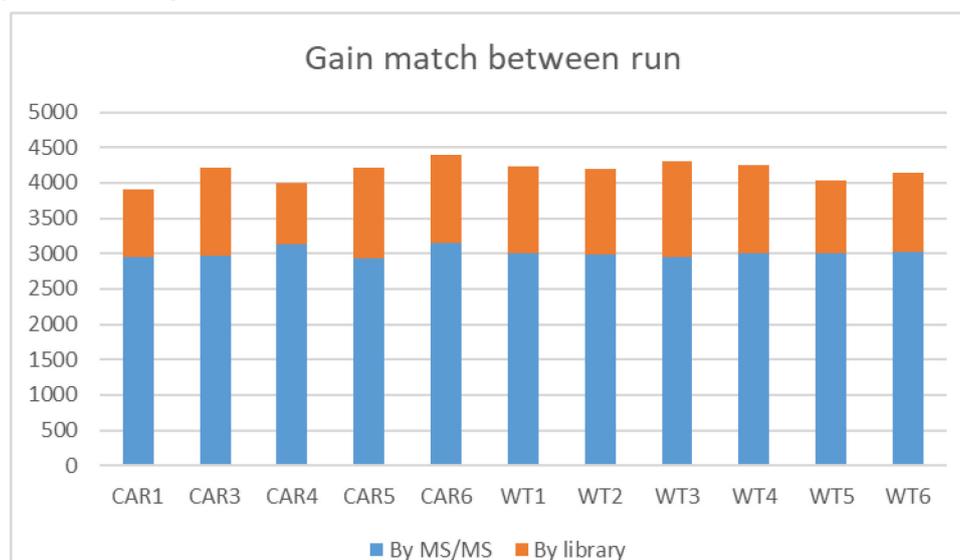


Figure 32: Nombre de protéines identifiées par MS/MS et par transfert d'identification pour l'étude de la maladie CARASIL

La distribution des  $\log_2$ (intensités des protéines), ainsi que les multiscatterplot et leur coefficient de corrélation de Pearson est construits pour chaque comparaison d'échantillons pour s'assurer l'homogénéité du groupe d'échantillons

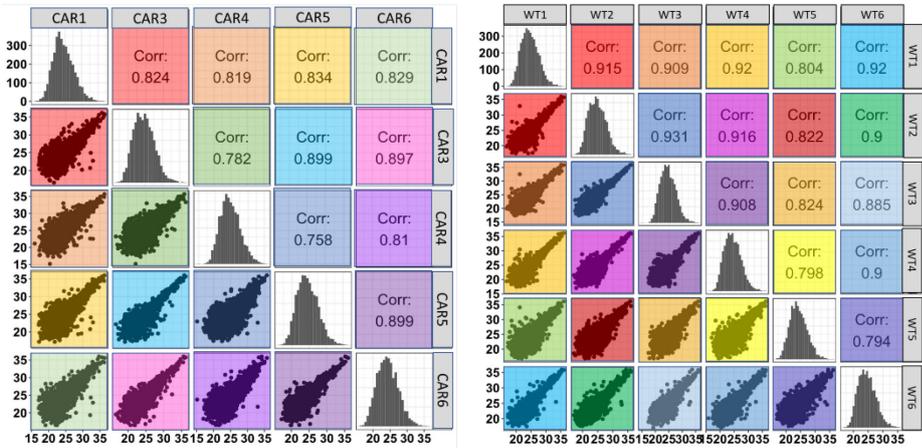


Figure 33 : Multiscatterplot et distribution de  $\log_2$  (intensités) des souris CAR et WT

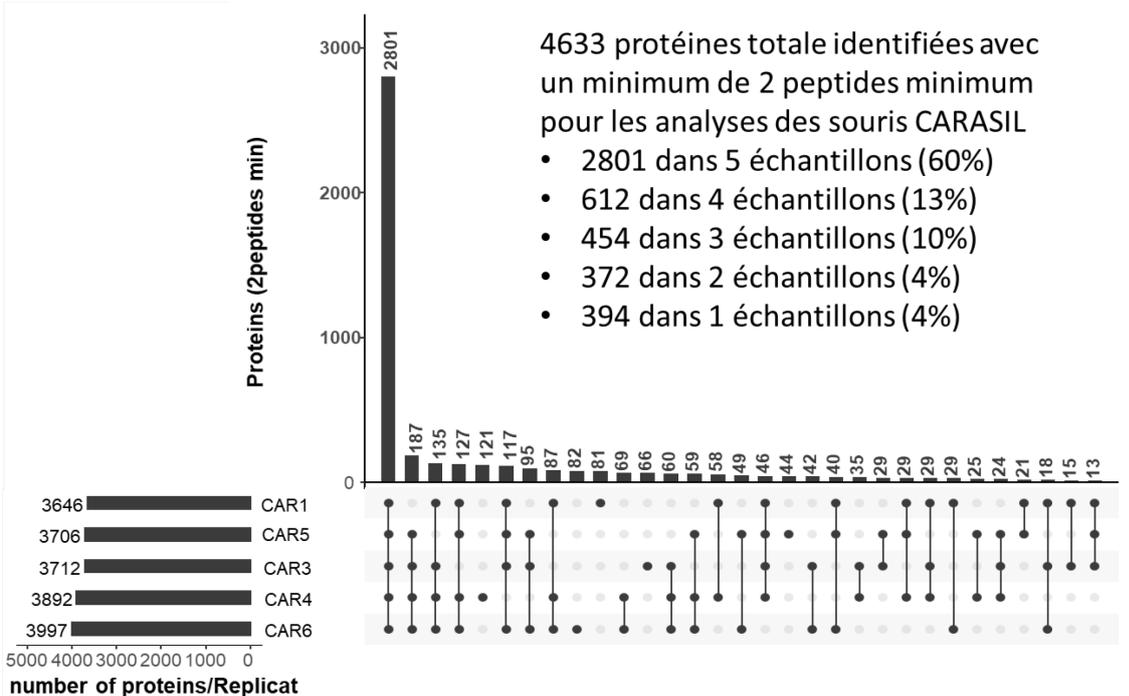


Figure 34: Diagramme de Venn de type Upset258 représentant le recouvrement protéique entre les échantillons du groupe de souris CARASIL

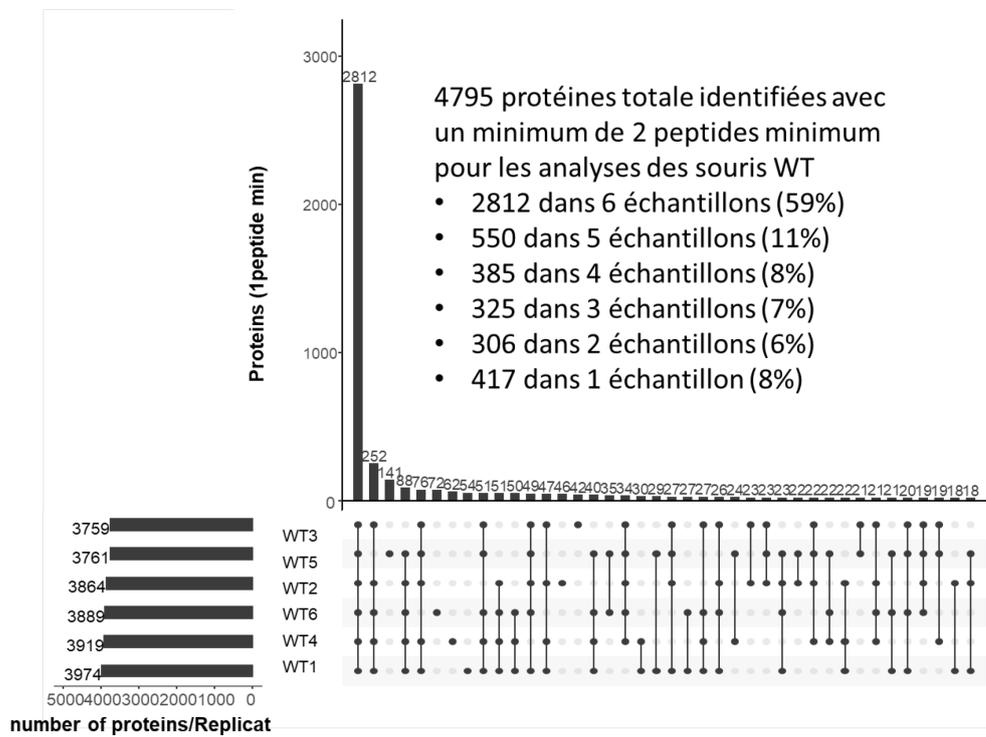


Figure 35: Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris WT

Distribution des GO pour l'analyse des souris CARASIL

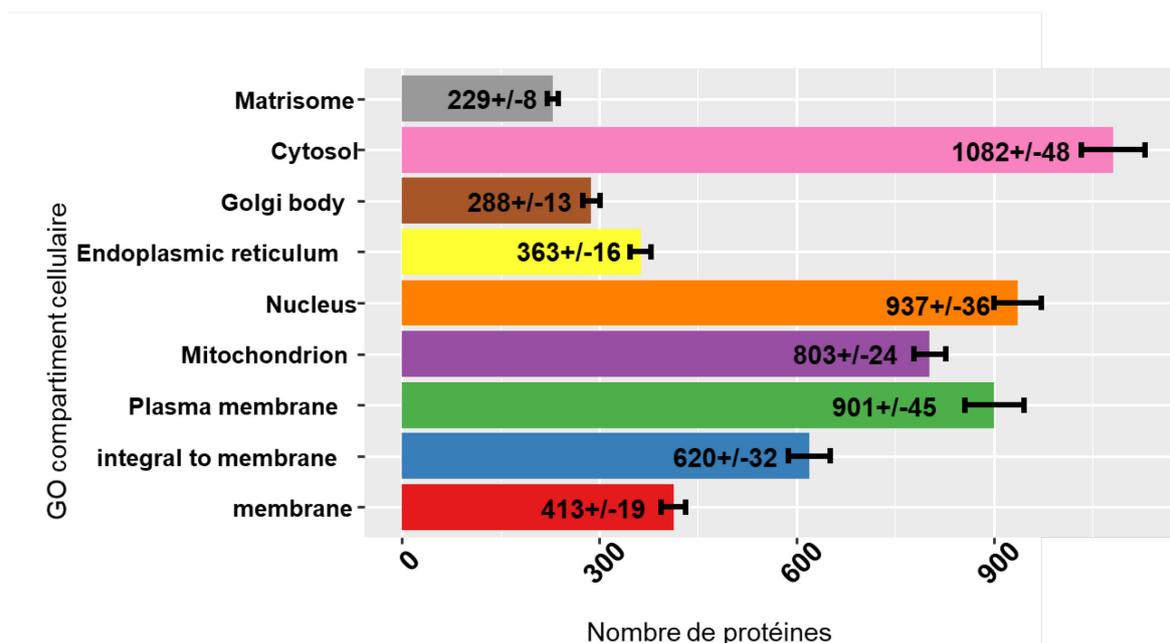


Figure 36 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris CARASIL

Distribution des GO pour l'analyse des souris WT

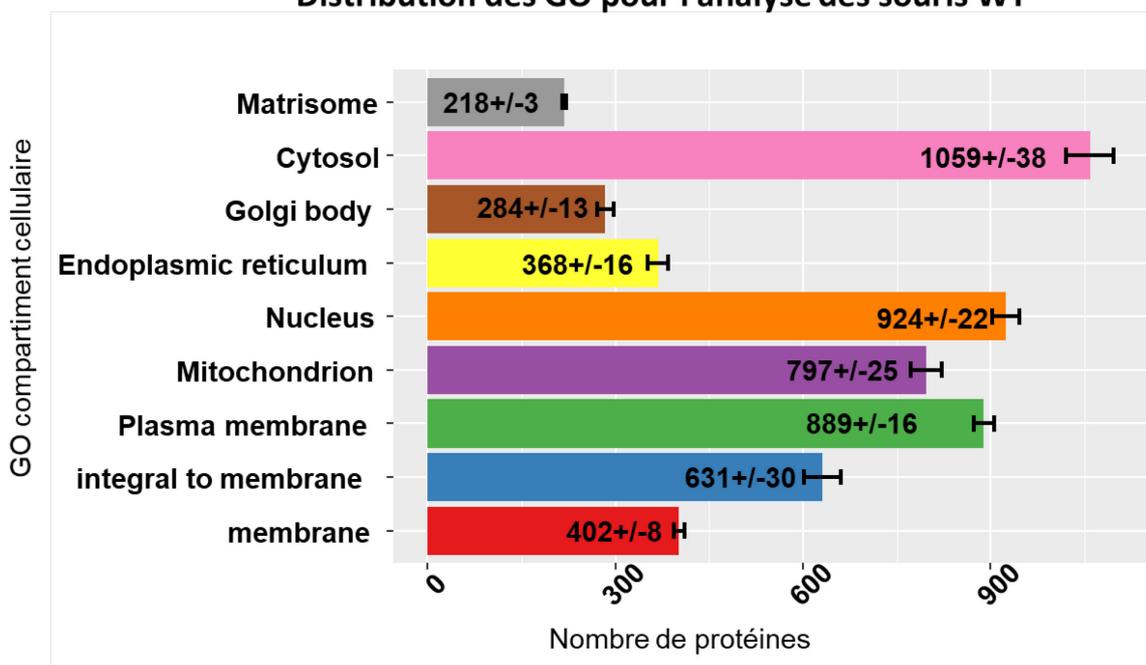


Figure 37 La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris WT

### III.5 Validation des analyses et identification des protéines dans les petits vaisseaux cérébraux des souris modèle pour la maladie du collagène de type IV

Pour des raisons pratiques d'écriture pour cette étude, les échantillons provenant du modèle murin Col4a1<sup>G1064D</sup> sera remplacé par COL4A1, et COL4A1<sup>+/+</sup> par WT. Les valeurs pour chaque échantillon est l'ensemble des résultats pour deux analyses. Pour chaque échantillon, nous avons le nombre d'acquisitions MS / MS (193686± 3%), le nombre de spectres validés (PSM pour Peptide Match Spectrum) (85870± 12%), ce qui donne environ 40% d'identification (% id.) avec un nombre d'identification de protéines avec au moins 2 peptides (3201±7%).

Table 10 : Table récapitulative des analyses MSMS pour l'étude de la maladie du collagène de type IV

Samples	# MS/MS	# PSM	% id.	# Peptides id.	# Proteins id.	# Proteins id. with 2pep
COL4-1	194616	91605	47%	31109	4027	2950
COL4-2	196803	94217	48%	31950	4143	3073
COL4-3	192770	89917	47%	30124	3976	2974
COL4-4	197117	75241	38%	29759	4309	3160
COL4-5	199045	77718	39%	31836	4568	3382
COL4-6	180733	67532	37%	30873	4560	3355
WT1	193272	99093	51%	33818	4194	3126
WT2	198632	100651	51%	34719	4450	3284
WT3	190688	88512	46%	29089	3986	2897
WT4	204756	85534	42%	33266	4703	3486
WT5	194353	88180	45%	35506	4831	3612
WT6	181451	72241	40%	29540	4272	3119

Le gain de protéines obtenu par l'utilisation du module de transfert d'identification *match between run* est représenté en figure 38.

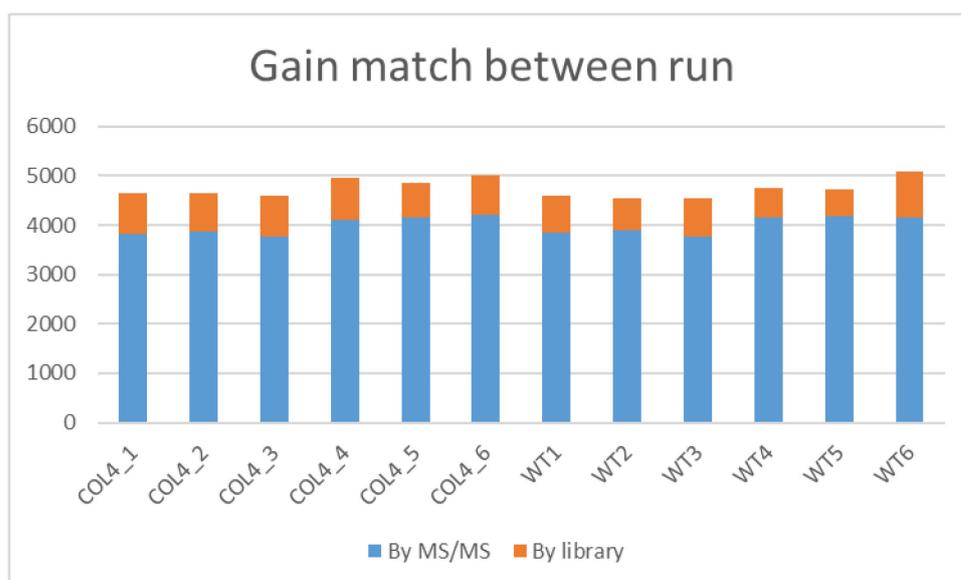


Figure 38: Nombre de protéines identifiées par MS/MS et par transfert d'identification pour l'étude de la maladie collagène IV

La distribution des  $\log_2$ (intensités des protéines), ainsi que les multiscatterplot et leur coefficient de corrélation de Pearson est construits pour chaque comparaison d'échantillons pour s'assurer l'homogénéité du groupe d'échantillons

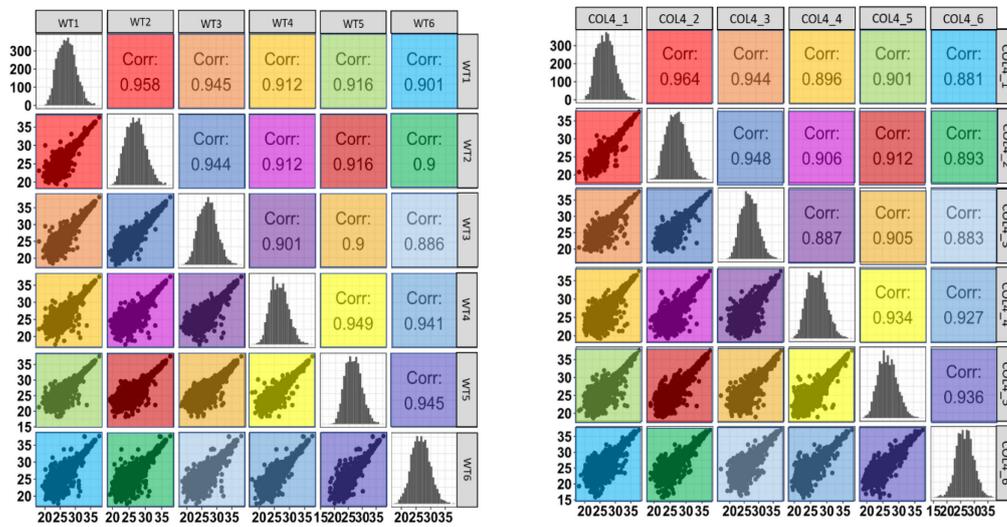


Figure 39 : Multiscatterplot et distribution de  $\log_2$  (intensités) des souris COL4A1 et WT

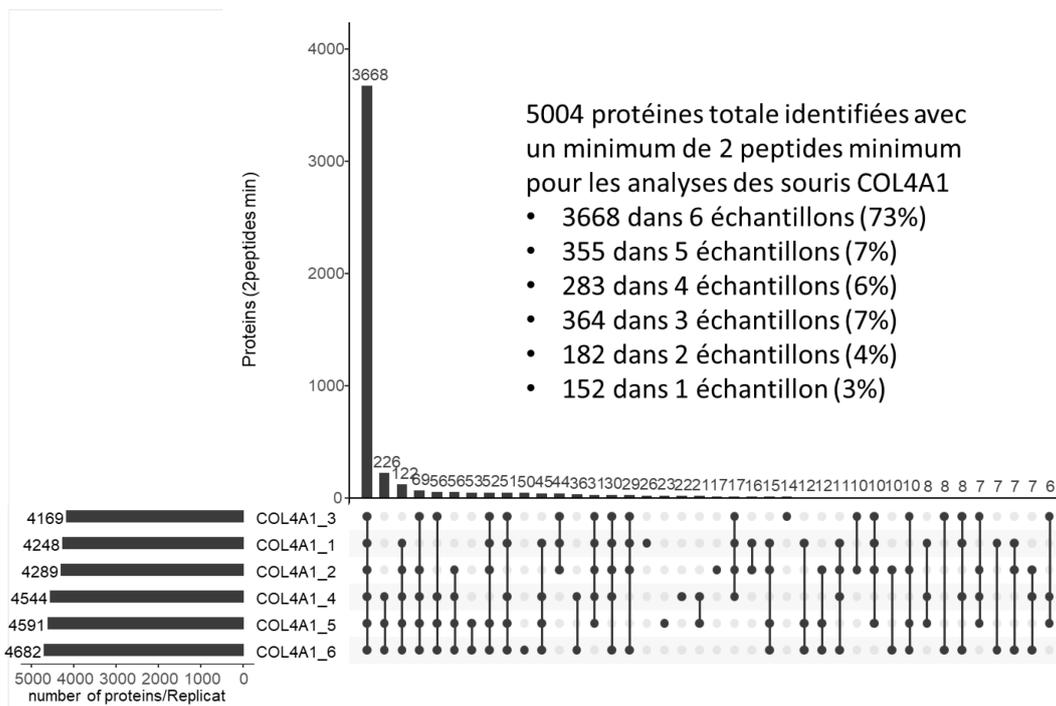


Figure 40 : Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris COL4

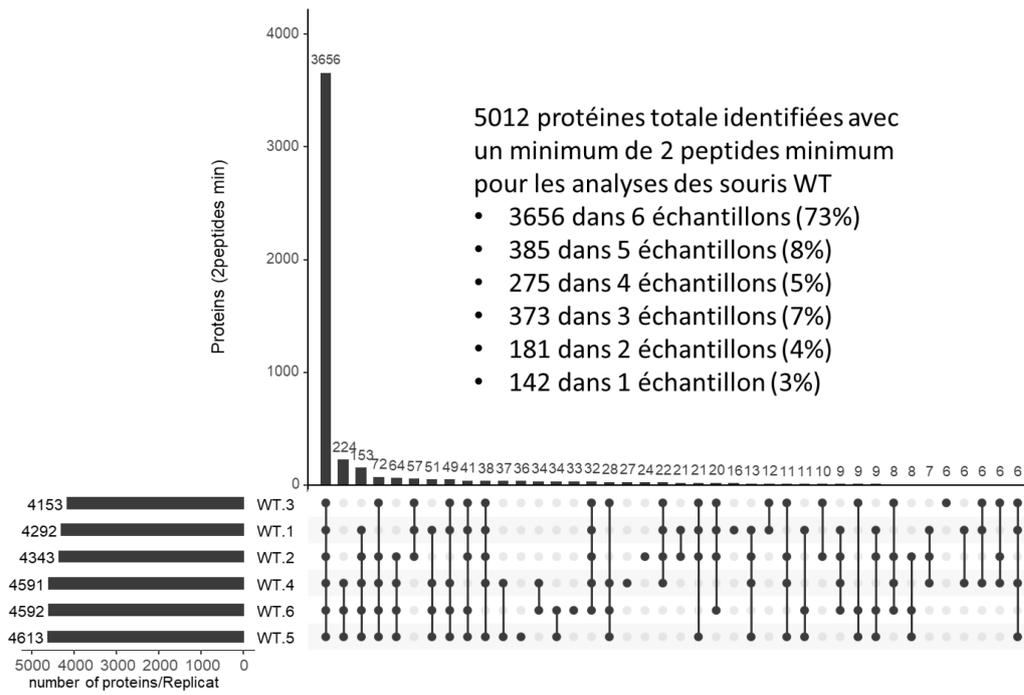


Figure 41 : Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris WT

Distribution des GO pour l'analyse des souris COL4

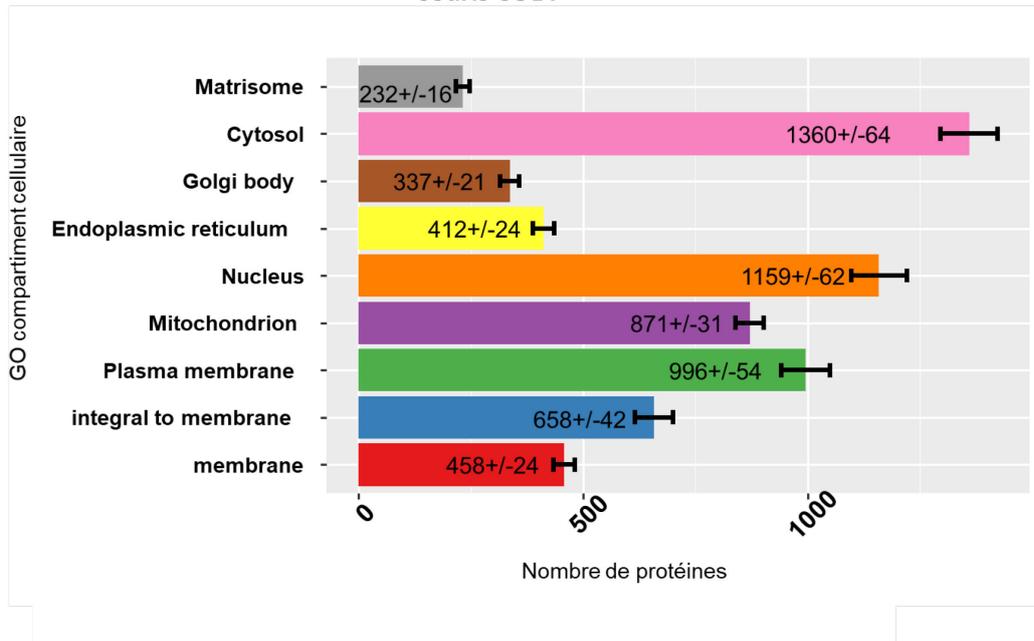


Figure 42 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris COL4

Distribution des GO pour l'analyse des souris WT (maladie COL4)

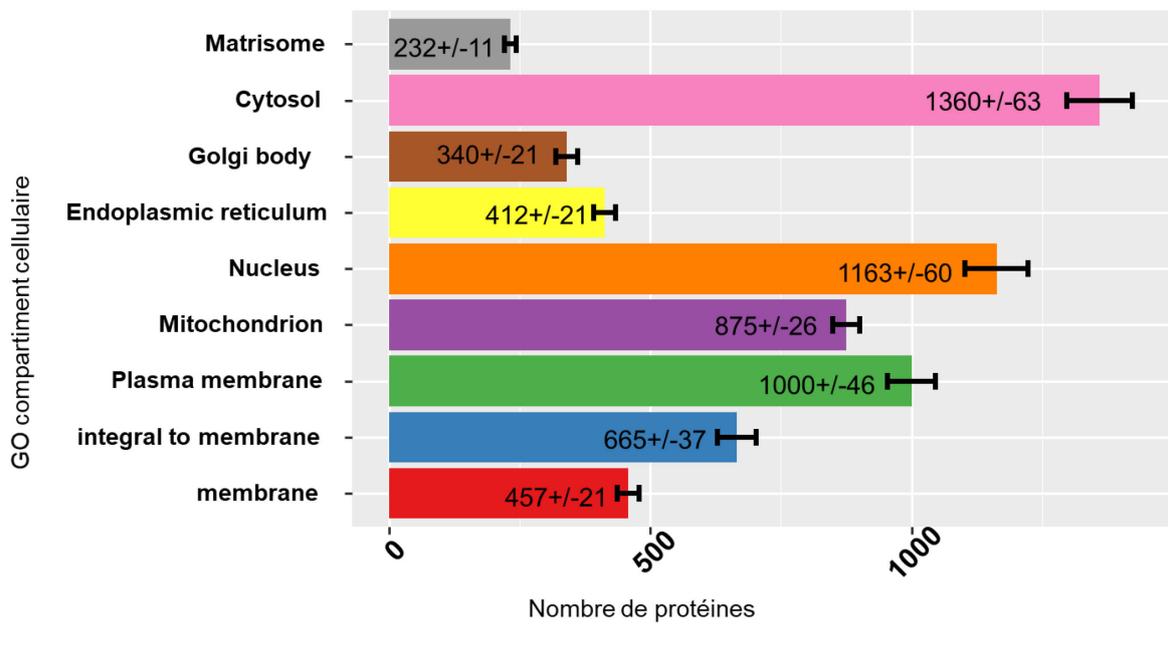


Figure 43 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris WT

### III.6 Validation des analyses et identification des protéines dans les petits vaisseaux cérébraux des souris modèle pour l'effet de l'âge

Pour des raisons pratiques d'écriture pour cette étude, les souris de 20 mois seront annotées 20M et celles de 2 mois 2M. Les valeurs pour chaque échantillon est l'ensemble des résultats pour deux analyses. Pour chaque échantillon, nous avons le nombre d'acquisitions MS / MS (195162± 1%), le nombre de spectres validés (PSM pour Peptide Match Spectrum) (59282± 12%), ce qui donne environ 30% d'identification (% id.) avec un nombre d'identification de protéines avec au moins 2 peptides (2483±7%).

Table 11 : Table récapitulative des analyses MSMS pour l'étude de la maladie de l'effet de l'âge

Samples	# MS/MS	# PSM	% id.	# Peptides id.	# Proteins id.	# Proteins id. with 2pep
20M1	194121	59544	31%	23426	3383	2419
20M2	194167	50555	26%	20415	3178	2208
20M3	195650	52312	27%	21372	3394	2361
20M4	196060	62078	32%	23719	3442	2417
20M5	193286	52892	27%	21118	3186	2231
20M6	195823	63211	32%	24653	3566	2528
2M1	195742	51147	26%	20673	3258	2254
2M2	195574	56648	29%	23287	3727	2615
2M3	195900	73560	38%	29229	4021	2899
2M4	194329	68233	35%	27067	3799	2745
2M5	194382	60710	31%	24596	3640	2552
2M6	196914	60490	31%	24249	3580	2570

Le gain de protéines obtenu par l'utilisation du module de transfert d'identification *match between run* est représenté en figure 44

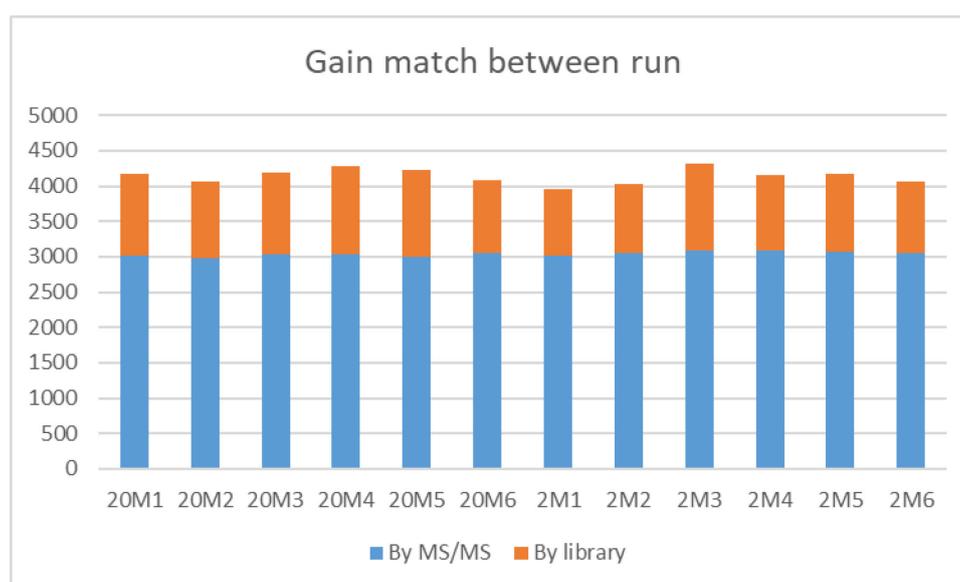


Figure 44 : Nombre de protéines identifiées par MS/MS et par transfert d'identification pour l'étude de l'effet de l'âge

La distribution des  $\log_2$ (intensités des protéines), ainsi que les multiscatterplot et leur coefficient de corrélation de Pearson est construits pour chaque comparaison d'échantillons pour s'assurer l'homogénéité du groupe d'échantillons

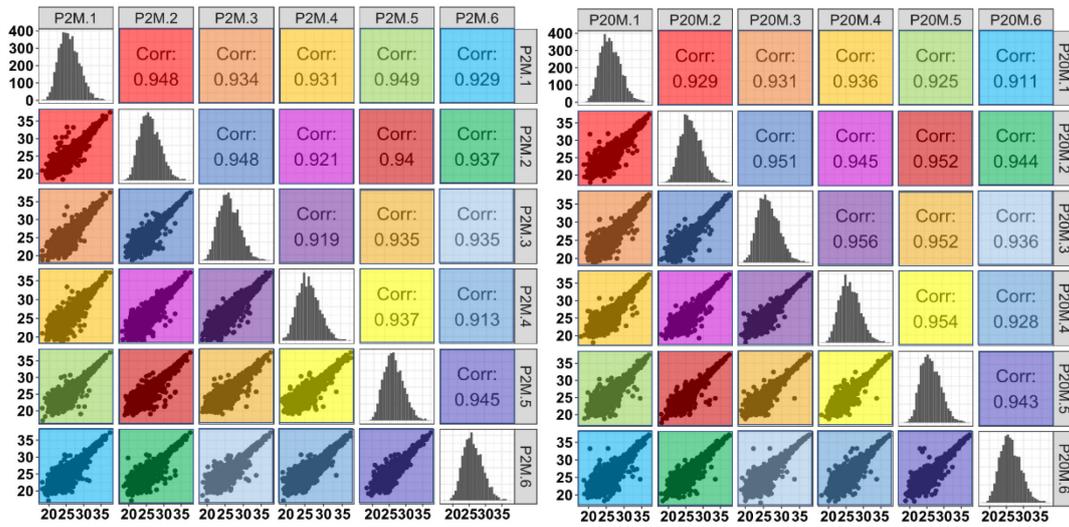


Figure 45 : Multiscatterplot et distribution de  $\log_2$  (intensités) des souris de 2 et 20 mois

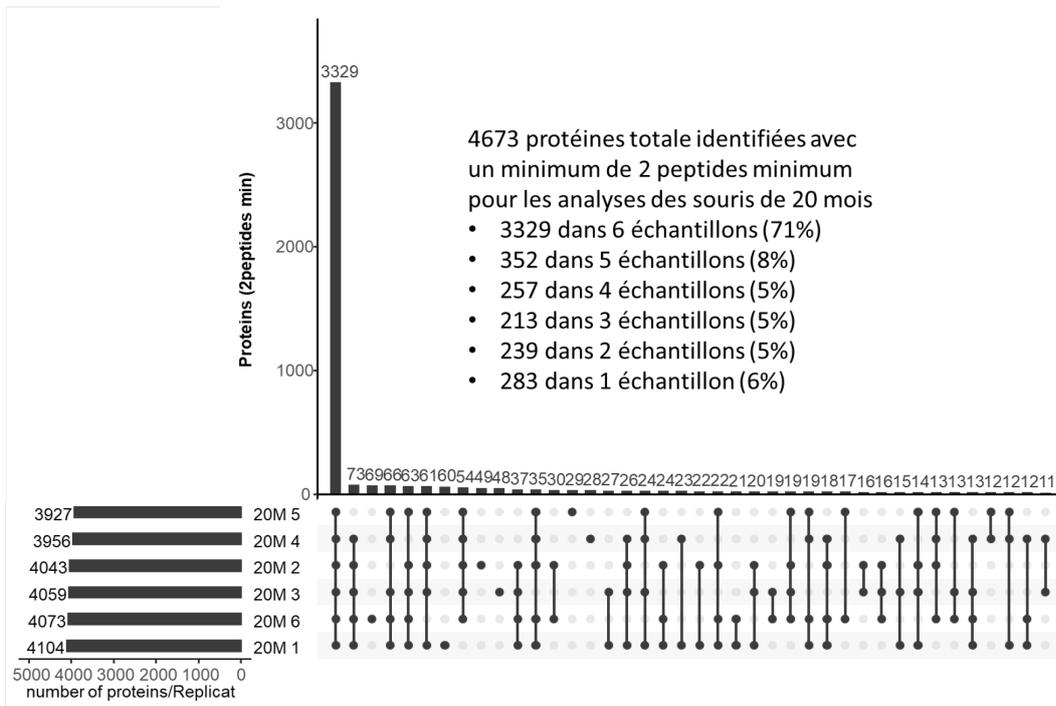


Figure 46 : Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris de 20 mois

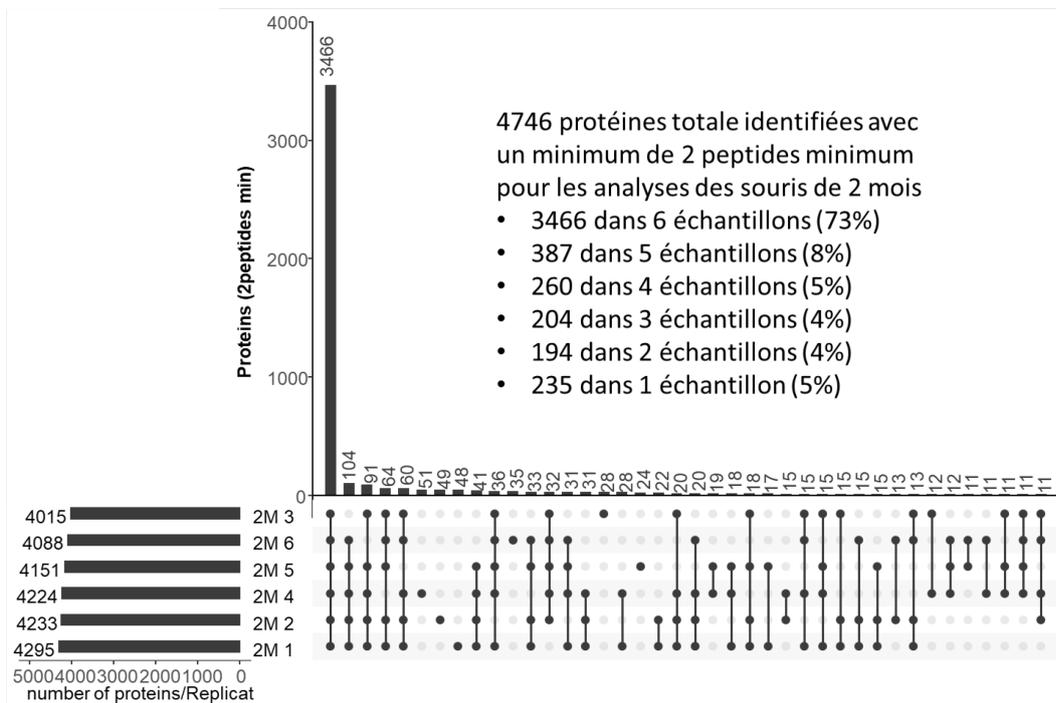


Figure 47 : Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris de 2 mois

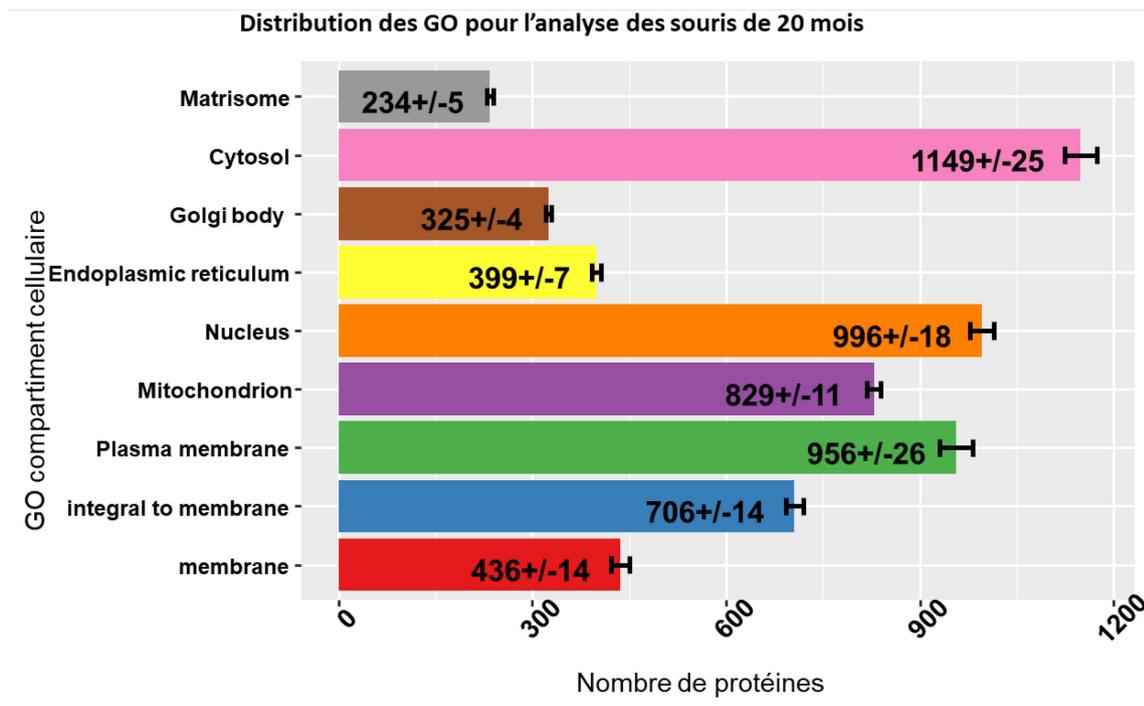


Figure 48 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris de 20 mois

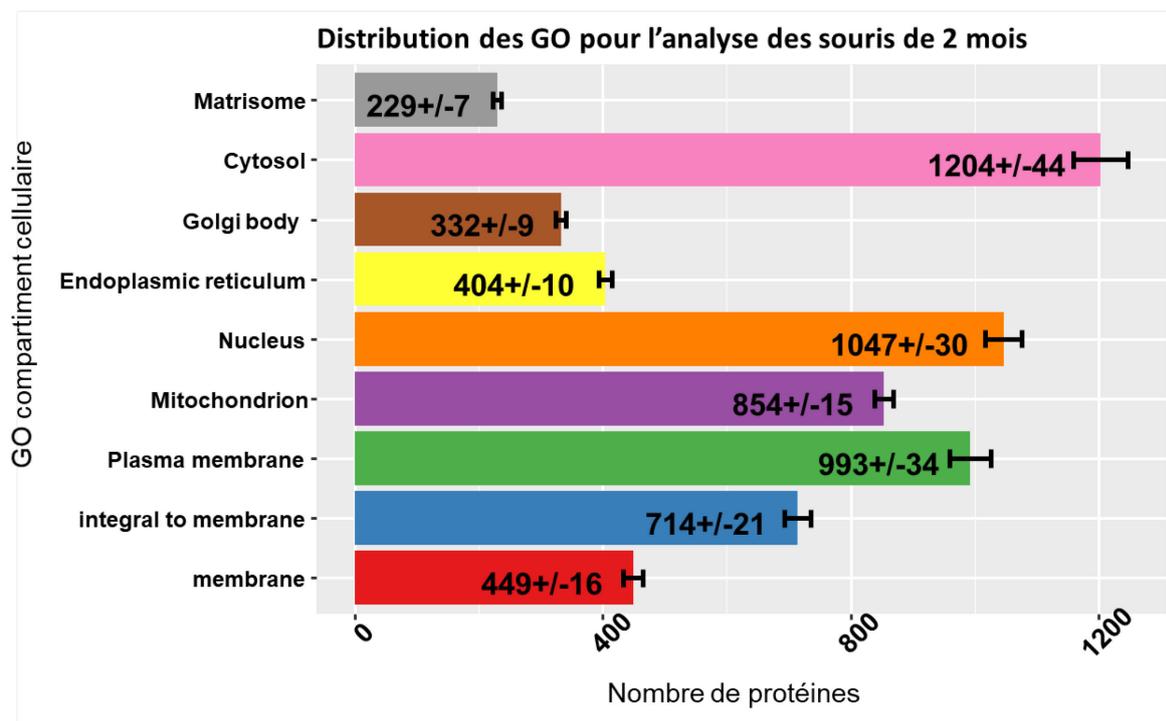


Figure 49 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris de 2 mois

### III.7 Validation des analyses et identification des protéines dans les petits vaisseaux cérébraux des souris modèle pour l'effet de l'hypertension

Pour des raisons pratiques d'écriture pour cette étude, les souris hypertendus sont annotées BPH, les souris normaux tendu, CTL et les souris hypertendues traitées BPht. Les valeurs pour chaque échantillon est l'ensemble des résultats pour deux analyses. Pour chaque échantillon, nous avons le nombre d'acquisitions MS / MS ( $180497 \pm 11\%$ ), le nombre de spectres validés (PSM pour Peptide Match Spectrum) ( $59969 \pm 13\%$ ), ce qui donne environ 30% d'identification (% id.) avec un nombre d'identification de protéines avec au moins 2 peptides ( $2224 \pm 6\%$ ).

Table 12 : Table récapitulative des analyses MSMS pour l'étude de la maladie de l'effet de l'hypertension

Samples	# MS/MS	# PSM	% id.	# Peptides id.	# Proteins id.	# Proteins id. with 2pep
BPH1	202058	60110	30%	19137	3110	2273
BPH2	184193	59181	32%	16267	2802	2005
BPH3	153920	53532	35%	18769	3000	2242
BPH4	189598	62189	33%	20095	3152	2338
BPH5	186744	59269	32%	18861	2943	2185
BPH6	182251	59680	33%	19922	3322	2453
CTL1	185203	51633	28%	17534	3140	2153
CTL2	123616	40095	32%	16309	2811	2083
CTL3	188389	58992	31%	17941	2985	2147
CTL4	190034	58959	31%	18675	3024	2204
CTL5	190376	54285	29%	17305	2819	2105
CTL6	186556	57579	31%	18787	2990	2237
BPht1	191610	62385	33%	20119	3212	2365
BPht2	142023	37786	27%	14239	2557	1942
BPht3	188764	66850	35%	21329	3330	2491
BPht4	186356	63275	34%	19697	3149	2343
BPht5	189771	60591	32%	18818	3019	2181
BPht6	187485	59059	32%	19213	3080	2277

Le gain de protéines obtenu par l'utilisation du module de transfert d'identification *match between run* est représenté en figure 50

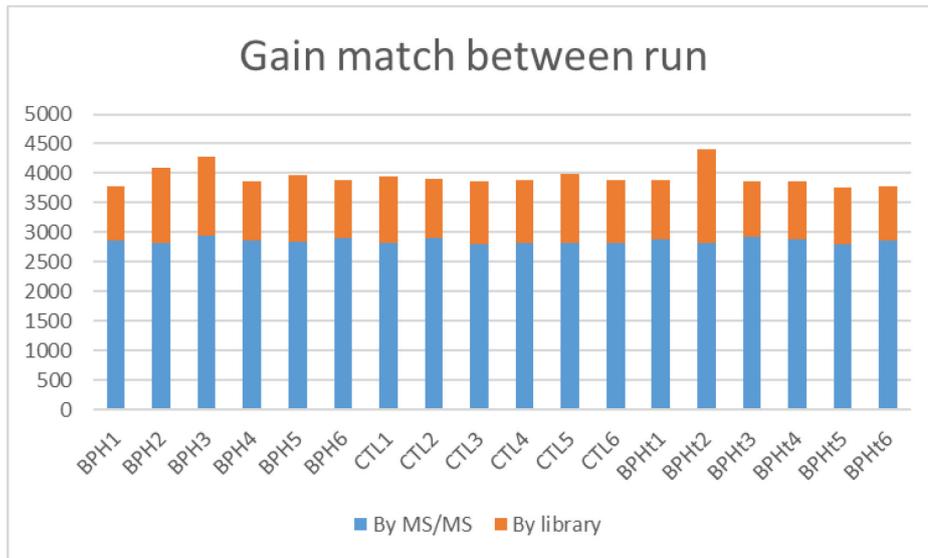


Figure 50: Nombre de protéines identifiées par MS/MS et par transfert d'identification pour l'étude de l'effet de l'hypertension

La distribution des  $\log_2$ (intensités des protéines), ainsi que les multiscatterplot et leur coefficient de corrélation de Pearson est construits pour chaque comparaison d'échantillons pour s'assurer l'homogénéité du groupe d'échantillons

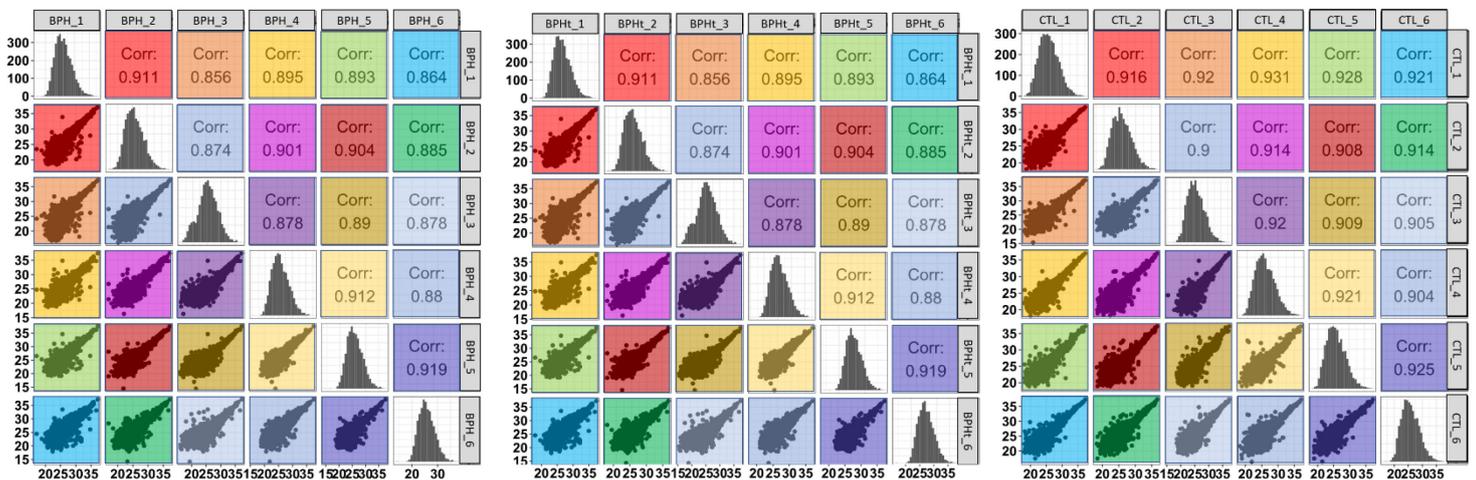


Figure 51 : Multiscatterplot et distribution de  $\log_2$  (intensités) des souris CTL, BPH et BPH traitées

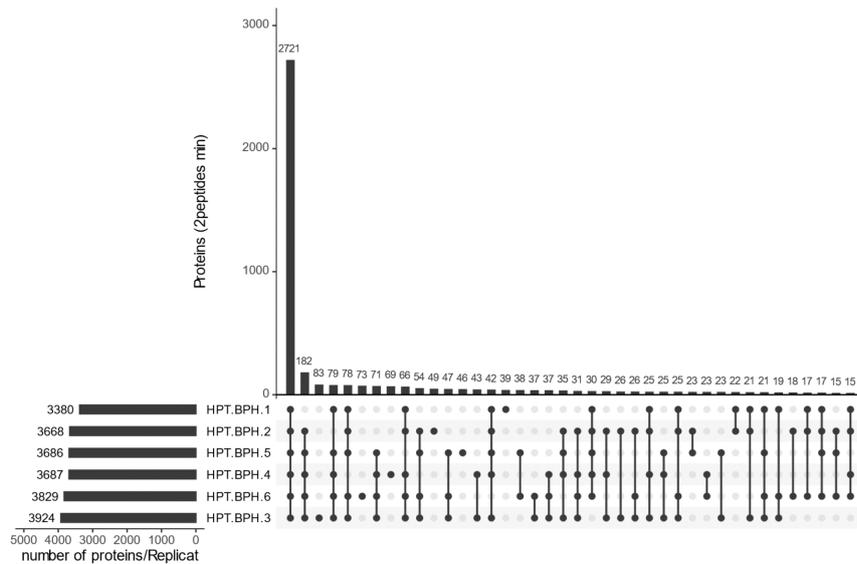


Figure 52 : Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris hypertendues BPH

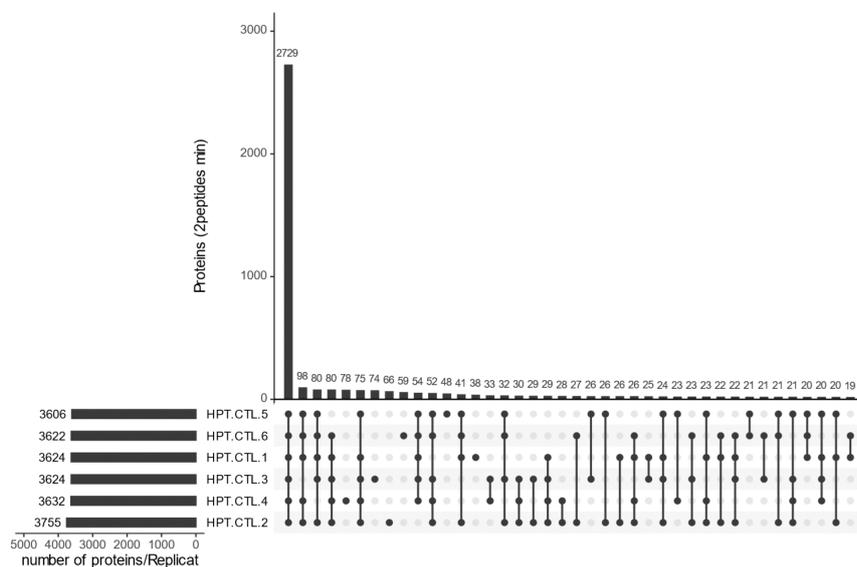


Figure 53 : Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris normaux tendues CTL

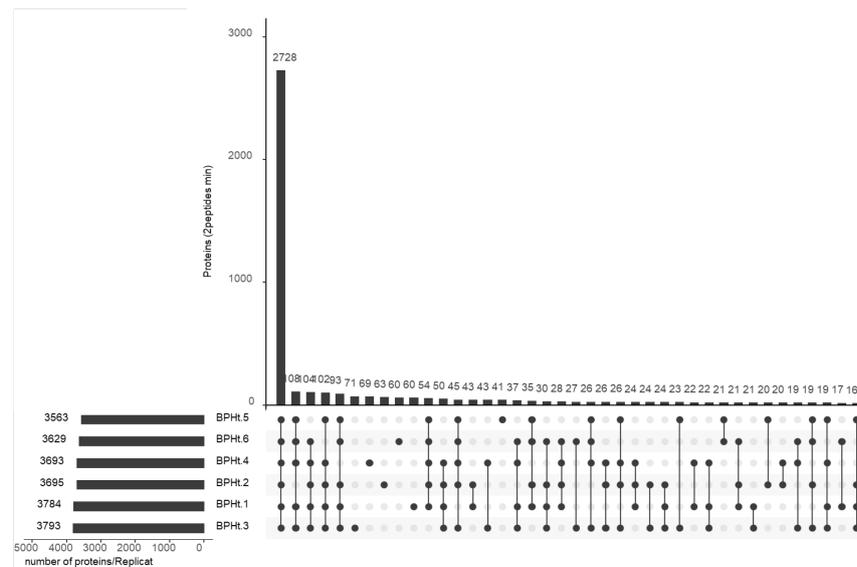


Figure 54 : Diagramme de Venn de type Upset<sup>259</sup> représentant le recouvrement protéique entre les échantillons du groupe de souris hypertendues traitées BPHt

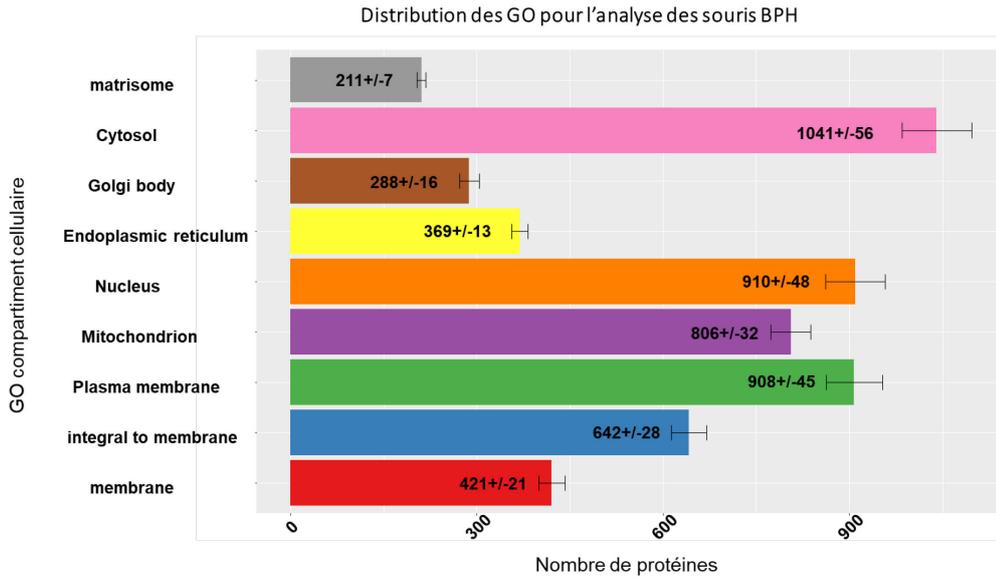


Figure 55 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris BPH

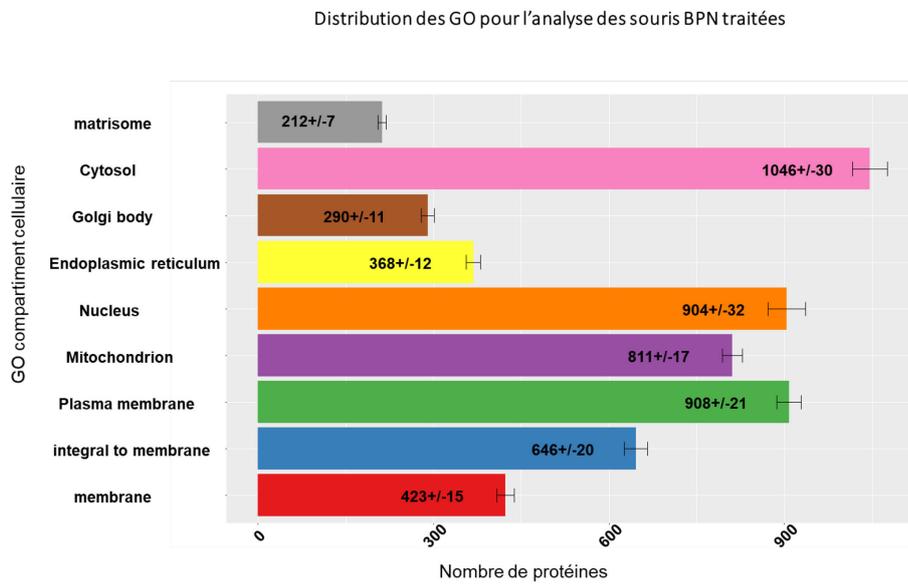


Figure 57 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris BPHt

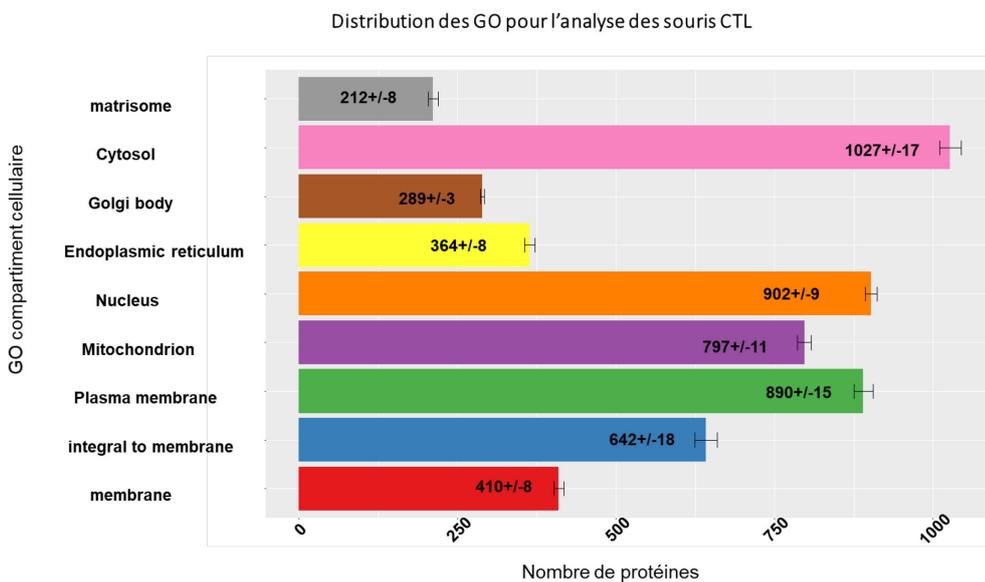


Figure 56 : La distribution des compartiments cellulaires GO des protéines identifiées pour le groupe de souris CTL

# I. Annexe 3 : Protocoles expérimentaux

---

## I.1 Préparation des artères piales pour la LC MS/MS

Les artères piales sont d'abord lavés avec une solution de PBS 1X. Ensuite, elles sont mises en suspension dans un tampon d'extraction (ABC 50 mM / DTT 10 mM / SDC 2%) et chauffés pendant 5 min à 95 ° C pour réduire et extraire les protéines. Une solution d'iodoacétamide IAM (50 mM final) a été ajouté pour alkyler les résidus de cystéine libres à température ambiante pendant 20 min. La solution protéique est ensuite diluée dans 50 mM d'ABC pour atteindre 0,5% en SDC de concentration finale avant d'ajouter le mélange trypsine-LysC et est incubée pendant une nuit à 37 °C sous agitation constante.

Les produits de digestion sont ensuite centrifugés pour récupérer le surnageant. Un volume de TFA à 10% est ajouté pour désactiver la trypsine et précipiter le SDC. Un cinquième du surnageant est utilisé pour quantifier les peptides par absorbance à 215 nm (DeNovix, modèle DS-11 et Thermo Fisher Scientific, modèle NanoDrop One/One) pour normaliser les quantités d'échantillon à injecter. Les peptides sont dessalés sur des Peptide Cleanup C18 (Agilent). Après élimination de l'acétonitrile avec un SpeedVac (Savant ISS110), les peptides dessalés sont remis alors en suspension dans le solvant d'injection, d'abord dans 20 µl pour la mesure de l'absorbance, ensuite dans un volume à ajuster selon l'absorbance mesurée avant l'analyse en LC-MS / MS.

## I.2 Analyses LC MS/MS

Environ 3,5 µg de chaque échantillon est injectée en double sur une colonne de piégeage à phase inversée C18 (C18 Acclaim PepMap100, 5 µm, 300 µm id × 5 mm) et éluée dans une colonne à phase inversée (C18 Acclaim PepMap100, 3 µm, 75 µm id × 50 cm) à un débit constant de 300 nL / min, avec un gradient de 2 à 50% de solvant B (H<sub>2</sub>O / ACN 10:90 (v / v), 0,1% FA) dans le solvant A ( H<sub>2</sub>O / ACN 98: 2 (v / v), 0,1% FA) sur 150 min. La chromatographie est couplé à un spectromètre de masse de type Q-Exactive HF (Thermo Fisher Scientific) équipé d'une source d'ions nano-électrospray en mode Data-Dependant-Acquisition DDA et un maximum de 20 ions les plus intenses peuvent être fragmentés et être analysés dans l'Orbitrap avec une résolution de 60000, une gamme de masse allant de 375-1500 m/z et un temps de remplissage d'injection maximum 60 ms. Les paramètres de la MS/MS sont les suivants : exclusion dynamique 30 s, pas de précurseurs monochargés, résolution 15000, NCE 28, cible AGC 1e4, temps d'injection maximum 60 ms, fenêtre d'isolement 2 m / z.

### I.3 Traitement et validation des données

Les données de spectrométrie de masse sont traitées dans un premier temps à l'aide du logiciel Proteome Discoverer (Thermo Scientific, v2.2) et recherchées à l'aide du logiciel MASCOT v2.6.0 (Matrix Science, Londres, Royaume-Uni) avec à la base de données Mouse SwissProt (publiée en avril 2018) avec les paramètres suivants: 10 ppm de tolérance de masse en MS, une tolérance de masse de 0,02 Da MS / MS, enzyme trypsine, carbamidométhylation de la cystéine (C), désamidation de l'asparagine (N) et de la glutamine (Q), oxydation de la méthionine (M) comme modifications des variables, maximum 2 erreurs de clivage et un taux de 1% de faux positif. Les annotations sur les protéines sont fournies par la base de données Uniprot et les annotations du matrisome sont obtenues par la base de données MatrisomeDB construite par le Hynes Lab (<http://matrisomeproject.mit.edu/>). Le nombre de MS / MS, PSM (peptide spectre-match) et le nombre de protéines ont été utilisés pour valider toute expérience.

### I.4 Quantification, validation et statistiques

La quantification des protéines est réalisée à l'aide de MaxQuant (version 1.5.3.8) avec à la base de données Mouse Swissprot (publiée en avril 2018). L'oxydation (M), la carbamidométhylation (C) et la désamidation de l'asparagine (N) et de la glutamine (Q) ont été définies comme des modifications variables, avec jusqu'à 2 erreurs de clivage, l'utilisation de l'option « Match between run » qui permet le transfert d'une identification MS/MS d'une analyse à l'autre et avec un taux de 1% de faux positifs pour les protéines et les peptides. Les intensités de protéines sont normalisées, les données manquantes sont triées et imputées si c'est nécessaire. Ensuite une analyse statistique, un test de Student suivie d'une correction par permutation multiple, est appliqué sur les résultats obtenus afin d'identifier les protéines dont l'expression est significativement différente entre les conditions. Toutes les étapes de traitement des données sont réalisées par des scripts R fait maison à l'aide des bibliothèques suivantes : DESeq pour la normalisation, imp4p pour l'imputation (impute.slsa) et Siggenes pour le test SAM.





## Annexe 4 : Script R pour le traitement des données

Ce script a été réalisé pour le traitement des fichiers de type proteinGroups.txt, résultat extrait de la suite logiciel de maxquant.

```
#supprimer objet en m?moire
rm(list=ls())
#librairies nécessaires
library(calibrate)
library("rJava")
library("Hmisc")
library(dplyr)
library(stringr)
library(readxl)
library(tidyr)
library(here)
library(xlsx)
library(siggenes)
library(DESeq)
library(imp4p)
library(forcats)
library(XLConnect)
options( java.parameters = "-Xmx16g" )
xlcFreeMemory()
jgc <- function()
{
  .jcall("java/lang/System", method = "gc")
}

#Le fichier extrait de maxquant doit comporter les Protein.IDs,
Protein.name...
#Les différents échantillons d'un meme groupe doivent avoir le meme nom +
espace + chiffre (pas accol?)
#Fonction pour sauvegarder plusieurs tableaux dans un fichier excel
#un onglet par tableau

#Fonction pour sauvegarder plusieurs onglets dans un fichier xlsx
save.xlsx <- function (file, ...)
{
  gc()
  jgc()
  require(xlsx, quietly = TRUE)
  objects <- list(...)
  fargs <- as.list(match.call(expand.dots = TRUE))
  objnames <- as.character(fargs)[-c(1, 2)]
  nobjects <- length(objects)
  for (i in 1:nobjects) {
    if (i == 1)
      write.xlsx(objects[[i]], file, sheetName = objnames[i])
    else write.xlsx(objects[[i]], file, sheetName = objnames[i],
                    append = TRUE)
  }
  print(paste("Workbook", file, "has", nobjects, "worksheets."))
}

#Aller chercher le fichier à traiter
```

```

file=file.choose()

if (str_detect(file,"xlsx") == TRUE)
{
  donnees=read_xlsx(file)
} else

{
  donnees=read.table(file,header=TRUE,sep="\t")
}

#Dossier où se trouve le fichier à traiter, nécessaire pour l'export des
résultats
dossier <- dirname(file)
#dossier <- "L:/Labo/2018/these/Campagne5VxCADASIL/txt" par exemple

#####
#####
#####
#####

#Eliminer les colonnes inutiles

#modifier le nom des certaines colonnes, trop long, trop compliqué
#colnames(donnees) <- gsub ("C..", "", colnames(donnees))
colnames(donnees) <- gsub ("N..", "", colnames(donnees))
colnames(donnees) <- gsub
("T..Protein.IDs", "Protein.IDs", colnames(donnees))
colnames(donnees) <- gsub ("T..Gene.name", "Gene.name", colnames(donnees))
colnames(donnees) <- gsub
("T..Protein.name", "Protein.name", colnames(donnees))
colnames(donnees) <- gsub
("Razor...unique.peptides", "RazorUniquePep", colnames(donnees))
colnames(donnees) <- gsub
("Only.identified.by.site", "OIBS", colnames(donnees))

#suppression des colonnes inutiles
donnees <- dplyr::select(donnees,
Protein.IDs,Protein.names,Gene.names,starts_with("RazorUniquePep"),starts_w
ith("Intensity"),OIBS,Reverse)
asup = which(colnames(donnees) == "Intensity.library")
if(length(asup)!=0)
{
  asup2 = which(colnames(donnees) == "Intensity")
  asup3 = which(colnames(donnees) == "Intensity.Library")
  donnees <- donnees[, -c(asup, asup2, asup3)]
}

#Filtrer sur reverse, modifier le nom du data.frame, donnees sera à
exporter

donnees2 <- dplyr::filter(donnees, Reverse != "+")

#enlever les lignes nulles pour compter le nb de protéines identifiés par
condition
coldata <- dplyr::select(donnees2,starts_with("Intensity"))
coldata[is.na(coldata)]<- 0
coldata$total <- apply(coldata,1,sum)
donnees2$total <- coldata$total
donnees2 <- filter(donnees2,donnees2$total!=0)

```

```

donnees2$total<-NULL
n1pep <- nrow(donnees2)

# chargement de la bdd matrisome et application sur la liste totale
matrisome_mm <-
read_excel("E:/Labo/R/Projet/GO/matrisome_mm masterlist.xls")
matrisome_mm <- dplyr::rename(matrisome_mm,Gene.names = 'Gene Symbol')
donnees2 <- merge(donnees2,matrisome_mm,by="Gene.names", all.x=TRUE) #a
exporter proteines totales
donnees2m <- merge(donnees2,matrisome_mm,by="Gene.names") #a exporter
proteines totales
nmatrix1pep <- nrow(donnees2m)

#supprimer colonne reverse
donnees2$Reverse <- NULL

#Normalisation par DESeq

#Créer les groupes
## format titre doit ?tre : Intensity.Tg88 1
##                               Intensity.HTRAWT 1
## !!!! pas d'underscore _ (tiret du 8), pas de tiret, rien...

coldata <- dplyr::select(donnees2,starts_with("Intensity"))
coldata <- coldata/1000
coldata <- apply(coldata,2,function(x) as.integer(as.numeric(x)))
titrecol <- colnames(coldata)
groupe <- gsub("Intensity.2.","",titrecol)
groupe <- gsub("[.] [0-9]", "",groupe)
#groupe <- gsub("[0-9]", "",groupe)
groupe <- as.factor(groupe)
designDESeq <- data.frame(titrecol,groupe)
cds = newCountDataSet(coldata,designDESeq)
cds=estimateSizeFactors(cds)
facteurs = sizeFactors(cds)
a= nrow(designDESeq)
designDESeq <- cbind(designDESeq,facteurs)
donnees3 <- donnees2

i <- for (i in 1:a)
{
  name = colnames(coldata)[i]
  donnees2[name] <- donnees2[name]/designDESeq$facteurs[i]
  donnees2[donnees2==0] <- NA
  donnees3[name] <- log2(donnees2[name])
}

#Filtrer les protéines sur les OIBS
donnees3 <- dplyr::filter(donnees3,OIBS!="+")

#Différents comptages sur le nb de peptide (1 et +), du matrisome ou pas,
permetts génération des graphes
data <- dplyr :: select(donnees3,Gene.names,starts_with("RazorUniquePep.))
data[is.na(data)]<- 0
fin <- ncol(data)
data$nbpep <- apply(data[,c(2:fin)],1,sum)
pepnul <- which(data$nbpep==0)
pepnul2 <- which(data$nbpep == 1)
data1pep = subset(data,data$nbpep>0)
data2pep = subset(data,data$nbpep>1)

```

```

n1pep <- nrow(data1pep)
n2pep <- nrow(data2pep)
donnees3$nbpep <- data$nbpep
supprime1pep <- subset(donnees3, donnees3$nbpep == 1)
donnees3 <- subset(donnees3, donnees3$nbpep>1)
n2pep <- nrow(donnees3)
donnees3m <- subset (donnees3, Category != "")
nmatrix2pep <- nrow(donnees3m)
donnees3mcore <- filter(donnees3m, Division=="Core matrisome")
donnees3mma <- filter(donnees3m, donnees3m$Division=="Matrisome-associated")
nmatrix2pepcore <- nrow(donnees3mcore)
nmatrix2pepma <- nrow(donnees3mma)
donnees3col <- filter(donnees3m, Category=="Collagens")
donnees3glyco <- filter(donnees3m, Category=="ECM Glycoproteins")
donnees3proteo <- filter(donnees3m, Category=="Proteoglycans")
donnees3regul <- filter(donnees3m, Category=="ECM Regulators")
donnees3aff <- filter(donnees3m, Category=="ECM-affiliated Proteins")
donnees3sec <- filter(donnees3m, Category=="Secreted Factors")
nmatrixcore <- nrow(donnees3mcore)
nmatrixma <- nrow(donnees3mma)
nmatrixcol <- nrow(donnees3col)
nmatrixglyco <- nrow(donnees3glyco)
nmatrixproteo <- nrow(donnees3proteo)
nmatrixregul <- nrow(donnees3regul)
nmatrixaff <- nrow(donnees3aff)
nmatrixsec <- nrow(donnees3sec)
donnees3$OIBS <- NULL

#####
#####

#Traitement des nan en appliquant la règle des "80%" (ici 75%), applicable
sur 2 conditions pour l'instant
#Les protéines supprimées sont sauvegardé dans une data.frame à part

#generer les informations du nb de groupe et de répliquats à partir du
fichier de départ
groupes= nlevels(groupe) #nb de groupe
replicat = designDESeq$groupe # les replicat et leur nom
replicat <- as.vector(replicat)
#suppression des nan totaux à partir du nb total de répliquats
data <- dplyr::select(donnees3, starts_with("Intensity."))
Nan =apply(is.na(data), 1, sum)
donnees3 <- cbind(donnees3, Nan)
data <- cbind(data, Nan)

#length(replicat)= nb de répliquat, si j'ai autant de NAN que de répliquat,
je peux supprimer la protéine
nbreplicat <- length(replicat)
donnees3 <- subset(donnees3, donnees3$Nan != nbreplicat)
data <- subset(data, data$Nan != nbreplicat)
nomgroupe=levels(groupe)

#Permet de calculer le min et max de nan en fonction du nb d'échantillon
par condition
a=1
echants=0
serie =c(0)
i <- for (i in 1:groupes)
{

```

```

b<-sum(groupe==nomgroupe[i])

data2 <- data[a:(a+b-1)]
nan2 = apply(is.na(data2),1,sum)
data=cbind(data,nan2)
echants[i]=paste("nan_",nomgroupe[i],sep='')
colnames(data) <- gsub("nan2",echants[i],colnames(data))

donnees3=cbind(donnees3,nan2)
echants[i]=paste("nan_",nomgroupe[i],sep='')
colnames(donnees3) <- gsub("nan2",echants[i],colnames(donnees3))
a=b+a

min=c(0)
max=c(0)
avoir=c(0)
echant_min=c(0)
echant_max=c(0)

serie[i] <- sum(replicat==replicat[i])
}

#boucle pour rejeter les protéines ayant trop de nan
arrondiplus <- function (x) ifelse(x-trunc(x)>=0.5,trunc(x)+1,trunc(x))
arrondimoins <- function (x) ifelse(x-trunc(x)>=0.5,trunc(x),trunc(x))
minimum= c(0)
maximum=c(0)
minima=c(0)
maxima =c(0)
asupprimer= c(0)

i <- for (i in 1:groupes)
{
  asupprimer[i]=paste("delete_",nomgroupe[i],sep='')
  b<-sum(groupe==nomgroupe[i])
  #calcul du nb de nan autorisé
  min [i] =b-(b*0.25)
  max [i]= b- (b*0.75)
  #fonction pour arrondir au supérieur ou inférieur en fonction

  max[i] <- arrondiplus(max[i]) #min arrondi pour chqe groupe
  min[i] <- arrondimoins(min[i]) #max arrondi pour chqe groupe

  echants[i]=paste("nan_",nomgroupe[i],sep='') #nom des colonnes
  avoir[i] = which(colnames(data) == echants[i]) #n°des colonnes
  verif= (data[,avoir[i]]>=min[i] | data[,avoir[i]]<=max[i])
  data <- cbind(data,verif)
  donnees3 <- cbind(donnees3,verif)
  colnames(donnees3) <- gsub("verif",asupprimer[i],colnames(donnees3))

}

supprime <- c(0)

i <- for (i in 1:groupes)
{
  supprime2<- filter (donnees3, donnees3[asupprimer[i]] == FALSE)

```

```

supprime <- rbind(supprime2,supprime)
#suppression des protéines present 1 * sur 2
donnees3 <- filter(donnees3,donnees3[asupprimer[i]] != FALSE)
}

#supprime les doublons dans le tableau des supprimés car 50% du nb
replicats
#doublonstest<-which(duplicated(supprime))
#supprime2<-supprime[-doublonstest,]

y=2

i=1

supprimeB1<- filter (donnees3, donnees3[echants[i]] >= min[i])
supprimeB2<- filter (donnees3, donnees3[echants[y]] >= min[y])

fin =ncol(supprimeB1)
supprimeB1<- cbind(supprimeB1[,c(2:1,3:fin)])
supprimeB2<- cbind(supprimeB2[,c(2:1,3:fin)])

#Liste des protéines à supprimer
supprimeB <- merge(supprimeB1,supprimeB2, by = "Protein.IDs")

#suppression des protéines de supprimeB
donnees3 <- donnees3 [ !donnees3$Protein.IDs %in% supprimeB$Protein.IDs,]

donnees3b <-subset (donnees3, donnees3$Category != "")

#Merge des protéines supprimées dans un seul fichier

supprime <- rbind(supprime,supprimeB1,supprimeB2)
doublonstest<-which(duplicated(supprime))
colajouter = ncol(supprime)-ncol(supprime1pep)

i <- for(i in 1:colajouter){
  col<- paste("colonne",i)
  supprime1pep[,col] <- c(0)
}

colnames(supprime1pep)<-colnames(supprime)
supprime <- rbind(supprime,supprime1pep)
supprime2<-supprime[-doublonstest,] # à exporter

#identifier les presents/absents
y=2

i=1
presentsabsent12 <- subset(donnees3,donnees3[echants[i]] >= min[i] &
donnees3[echants[y]]<= max[y])
presentsabsent21 <- subset(donnees3,donnees3[echants[y]] >= min[y] &
donnees3[echants[i]]<= max[i])

donnees3 <- donnees3 [ !donnees3$Protein.IDs %in%
presentsabsent12$Protein.IDs,]
donnees3 <- donnees3 [ !donnees3$Protein.IDs %in%
presentsabsent21$Protein.IDs,]

```

```

#Comptage des Nan

imputationcount <- dplyr::select(donnees3,starts_with("Intensity"))
nbnanbysample <- apply(is.na(imputationcount),2,sum)
nbnanbyprot <- apply(is.na(imputationcount),1,sum)
imputationcount$nan <- nbnanbyprot
pepmoyen <-dplyr::select(donnees3,starts_with("RazorUniquePep."))
nbpep <- apply(pepmoyen,1,mean)
nbpep <- round(nbpep,0)
pepmoyen$mean <- nbpep
calculnan <- cbind(imputationcount,pepmoyen)
graphe <- cbind(nbpep,nbnanbyprot)
graphe <- as.data.frame(graphe)
graphe <- filter (graphe,nbnanbyprot>0)
maxpep <- max(graphe$nbpep)
graphe2 <- table(factor(graphe$nbpep,levels=1:maxpep))
maxnan <- max(graphe2[1])
proportionNanSample <- mean(nbnanbysample)/nrow(calculnan)*100
proportionNanSample <- paste0(proportionNanSample,"%")

#graphe pour voir evolution du nombre de Nan en fonction du nb de peptide
pour identifier une prot
pdf(paste(file,"countNaN.pdf",sep=""))
p <- plot(graphe2,ylab="number of NaN",xlab="number of peptide to identify
protein",ylim=c(1,maxnan),type="o",col="#FFAAAA",lwd=2,pch=16)
print(p)
dev.off ()

#imputation de type lsa
#Formater le tableau de sorte à avoir une matrice seulement avec des
chiffres
data <-dplyr:: select(donnees3,starts_with("Intensity"))
data <- as.matrix(data)
testimp4p <- impute.slsa(data,conditions=groupe)
#Calcul foldchange
y=1
moyenne=c(0)
titre=c(0)
z=0

donnees4 <- dplyr::select(donnees3,
Protein.IDs,Protein.names,Gene.names,starts_with("RazorUniquePep"))
donnees4 <- cbind(donnees4,testimp4p)

i <- for (i in 1:groupes)
{
  z=z+serie[i]
  data <- testimp4p[,y:z]
  data2 <- 2^data
  moyenne <- apply(data2,1,mean)
  moyenne <- round(moyenne,2)
  donnees4 <- cbind(donnees4,moyenne)
  colnames(donnees4) <- gsub
("moyenne",paste("Moyenne_",nomgroupe[i],sep=" "),colnames(donnees4))
  if (y<=nbrelicat){
    y=y+serie[i]
  }
  titre [i]<- gsub ("Intensity.", "", nomgroupe[i])
}

```

```

#Calcul des moyennes des intensites pour calculer le ratio FC
data <- dplyr::select(donnees4,starts_with("Moyenne"))
FC12 <- data[,1]/data[,2]
FC12 <- round(FC12,2)
donnees4 <- cbind(donnees4,FC12)

#Annotation matrisome

donnees5 <- merge(donnees4,matrisome_mm,by="Gene.names", all.x=TRUE) #a
exporter proteines totales
donnees6 <- merge(donnees4,matrisome_mm,by="Gene.names") #à exporter
proteines du matrisome

#Annotation matrisome des protéines présente-absente
resentsabsent1ann <- merge(resentsabsent12,matrisome_mm,by="Gene.names",
all.x=TRUE)
resentsabsent2ann <- merge(resentsabsent21,matrisome_mm,by="Gene.names",
all.x=TRUE)

#####
#####

#Tests T SAM

#Matrisome
debut = which(colnames(donnees5) == titrecol[1])
fin = which(colnames(donnees5) == titrecol[length(titrecol)])
test_sam_m <- sam (donnees6[,debut:fin],groupe,gene.names= donnees6[,2] ,
method = d.stat, var.equal = TRUE,B=250,s0=0)
pvalue_matrisome <- test_sam_m@p.value
#pvalue_matrisome <- round(pvalue_matrisome,2)
donnees6 <- cbind(donnees6,pvalue_matrisome)
qvalue_matrisome <- test_sam_m@q.value
#qvalue_matrisome <- round(qvalue_matrisome,2)
donnees6 <- cbind(donnees6,qvalue_matrisome)
fdrm <- findDelta(test_sam_m,fdr=0.05)
deltam <- mean (fdrm[,1])
protSigMatrix<-list.siggenes(test_sam_m,deltam)
protSigMatrix <- as.data.frame(protSigMatrix)
significant <- c(rep("+",nrow(protSigMatrix)))
protSigMatrix <- cbind(protSigMatrix,significant)
donnees6 <-
merge(donnees6,protSigMatrix,by.x="Protein.IDs",by.y="protSigMatrix",all=TR
UE)
donnees6[,debut:fin] <- round(donnees6[,debut:fin],2)

nmatrixquant <- nrow(donnees6)
donnees6core <- filter(donnees6,Division=="Core matrisome")
donnees6ma <- filter(donnees6,Division=="Matrisome-associated")
donnees6col <- filter(donnees6,Category=="Collagens")
donnees6glyco <- filter(donnees6,Category=="ECM Glycoproteins")
donnees6proteo <- filter(donnees6,Category=="Proteoglycans")
donnees6regul <- filter(donnees6,Category=="ECM Regulators")
donnees6aff <- filter(donnees6,Category=="ECM-affiliated Proteins")
donnees6sec <- filter(donnees6,Category=="Secreted Factors")
nmatrixquantcore <- nrow(donnees6core)
nmatrixquantma <- nrow(donnees6ma)
nmatrixquantcol <- nrow(donnees6col)
nmatrixquantglyco <- nrow(donnees6glyco)
nmatrixquantproteo <- nrow(donnees6proteo)
nmatrixquantregul <- nrow(donnees6regul)

```

```

nmatrixquantaff <- nrow(donnees6aff)
nmatrixquantsec <- nrow(donnees6sec)

graphSAM <- paste0(file, "_SAMmatrix.jpeg")
jpeg(graphSAM,width=7 ,height=7, units="in",res = 300)
plot(test_sam_m,deltam, main=paste("SAM Plot",titre[1],"vs",titre[2],"in
Matrisome"))
dev.off()
#####
#####

#All
test_sam <- sam (donnees5[,debut:fin],groupe,gene.names= donnees5[,2] ,
method = d.stat, var.equal = TRUE,B=250,s0=0)
pvalue<- test_sam@p.value
#pvalue <- round(pvalue,2)
donnees5 <- cbind(donnees5,pvalue)
qvalue <- test_sam@q.value
#qvalue <- round(qvalue,2)
donnees5 <- cbind(donnees5,qvalue)
fdr <- findDelta(test_sam,fdr=0.05)
delta <- mean (fdr[,1])
protSig<-list.siggenes(test_sam,delta)
protSig <- as.data.frame(protSig)
significant <- c(rep("+",nrow(protSig)))
protSig <- cbind(protSig,significant)
donnees5 <-
merge(donnees5,protSig,by.x="Protein.IDs",by.y="protSig",all=TRUE)
donnees5[,debut:fin] <- round(donnees5[,debut:fin],2)
#Deplacer les colonnes pour la lecture pour les 2 derniers tableaux
#trouver les colonnes moyennes
graphSAM <- paste0(file, "_SAM.jpeg")
jpeg(graphSAM,width=7 ,height=7, units="in",res = 300)
plot(test_sam,delta, main=paste("SAM Plot",titre[1],"vs",titre[2],""))
dev.off()

nquant <- nrow(donnees5)

sign <- which(colnames(donnees5) == "significant")
fc <- which(colnames(donnees5) == "FC")
pv <- which(colnames(donnees5) == "pvalue")
qv <- which(colnames(donnees5) == "qvalue")
div <- which(colnames(donnees5) == "Division")
cat <- which(colnames(donnees5) == "Category")
test <- grep("^Moyenne", colnames(donnees5))
testT_all <- cbind(donnees5[,c(1:3,sign,fc,pv,qv,test,4:fin,div,cat)])
testT_matrisome <-
cbind(donnees6[,c(1:3,sign,fc,pv,qv,test,4:fin,div,cat)])

#information sur le nombre de Na
nbttotalabsent = nrow(presentsabsent12)+nrow(presentsabsent21)
nbttotalquant = nrow(donnees5)+nbttotalabsent
dataNA <-dplyr ::
select(donnees2,Protein.IDs,Gene.names,Protein.names,starts_with("Intensity
. "))
donnees2m [donnees2m==0]<- NA
dataNAM <- dplyr ::
select(donnees2m,Protein.IDs,Gene.names,Protein.names,starts_with("Intensit
y. "))
dataSansNa <- na.omit(dataNA)

```

```

SansNa <- nrow(dataSansNa)
dataSansNam <- na.omit(dataNAM)
inversedataSanNa <- dataNA[!complete.cases(dataNA), ]
propNatotal <- paste(round((nlpep-SansNa)/nrow(dataNA)*100,2), "%")
inversedataSanNam <- dataNAM[!complete.cases(dataNAM), ]
protsigNa <-
merge(inversedataSanNa, protSig, by.x="Protein.IDs", by.y="protSig")
protsigNam <-
merge(inversedataSanNa, protSigMatrix, by.x="Protein.IDs", by.y="protSigMatrix
")
Alldata <- nrow(dataNA)
Alldatam <- nrow(dataNAM)
SansNam <- nrow(dataSansNam)
nprotsigNA <- nrow(protsigNa)
nprotsigNAM <- nrow(protsigNam)
gainNA <- (nbttotalquant-SansNa)/SansNa*100
gainNA <- paste0(round(gainNA,2), "%")

#Renommage des data.frame et export

proteinGroups <- donnees
NormalisationDESeq <- donnees2
Traitement_Nan <- donnees3
Proteins_Deleted <- supprime2
presentAbsent1 <- presentsabsent1ann
presentsAbsent2 <- presentsabsent2ann
imputation <- donnees4

#####
#####

#Creation des tableaux pour faire les graphes de répartition des
différentes catégories
#Graphe nb protéines totales et matrisome

type=c("1 peptide", "2peptide+", "quantifiable")
nbprot <- data.frame(c(type), nbproteinTotal =
c(nrow(donnees), nrow(donnees2), nrow(donnees3)), nbproteinMatrisome
=c(nrow(donnees2m), nrow(donnees3m), nrow(donnees6)))
#supprimer les colonnes inutiles
ncoldonnees2m <- ncol(donnees2m)
donnees2m <- donnees2m[, -c((ncoldonnees2m-1):ncoldonnees2m)]
donnees2m$Reverse <- NULL
x=table(factor(donnees2m$Division.x))
y=table(factor(donnees2m$Category.x))
x <- data.frame(x)
y <- data.frame(y)
graphematrisome <- rbind(x,y)
graphematrisome$pourcentage <-
paste(round((graphematrisome$Freq/nlpep)*100,1), "%")

#Recapitulatif fichier traité et paramètre telle que nb de groupe + règle
nan + nb de proteine supprimée

nbttotalabsent = nrow(presentsabsent12)+nrow(presentsabsent21)
nbttotalquant = nrow(donnees5)+nbttotalabsent
nanmin <- min(nbnanbysample)
nanmean <- mean(nbnanbysample)
nanmax <- max(nbnanbysample)

```

```

recapitulatif <- data.frame(row.names = c("Nb total proteine 1pep", "Nb
total proteine 2pep", "Nb total proteine absent/present", "Nb total proteine
supprimees", "fichier initial", "nb de groupe", "nb moyen de Nan par
sample", "%Nan/sample", "nb min de NAn", "nb max de Nan", "nb de prot ident
sans NA", "nb prot significant NA ", "nb de prot ident matrix sans NA", "nb
prot significant matrix NA ", "%NA total", "Gain imputation", "Nb prot matrix
2pep", "Nb prot matrix 2pep Core", "Nb prot matrix 2pep MA", "nb prot matrix
core", "nb prot matrix ma", "nb prot matrix collagen", "nb prot matrix
glyco", "nb prot matrix proteo", "nb prot matrix regul", "nb prot matrix
aff", "nb prot matrix sec"), information
=c(nlpep, n2pep, nbtotalabsent, nrow(supprime2), file, groupes, nanmean, proportio
nNanSample, nanmin, nanmax, SansNa, nprotsigNA, SansNam, nprotsigNAM, propNatotal,
gainNA, nmatrix2pep, nmatrix2pepcore, nmatrix2pepma, nmatrixcore, nmatrixma, nmat
rixcol, nmatrixglyco, nmatrixproteo, nmatrixregul, nmatrixaff, nmatrixsec))
recapitulatifquant <- data.frame (row.names=c("Nb total proteine quant", "nb
prot matrix quant core", "nb prot matrix quant ma", "nb prot matrix quant
collagen", "nb prot matrix quant glyco", "nb prot matrix proteo", "nb prot
matrix regul", "nb prot quant matrix aff", "nb prot quant matrix sec"),
information
=c(nbtotalquant, nmatrixquantcore, nmatrixquantma, nmatrixquantcol, nmatrixquan
tglyco, nmatrixquantproteo, nmatrixquantregul, nmatrixquantaff, nmatrixquantsec
))
test <- gsub("(.) (\\.txt)", sprintf("\\1_%s\\2", format(Sys.time(),
"%Y_%m_%d")), file)
testinfo <-gsub("(.) (\\.txt)", sprintf("\\1_%s\\2", format(Sys.time(),
"%Y_%m_%d")), paste0(file, "info"))
fichier=paste(test, ".xls", sep="")
fichier2=paste(testinfo, ".xls", sep="")
fichier3=paste(testinfo, "volcanoData.xls", sep="")

#export des fichiers
options( java.parameters = "-Xmx16g" )
xlcFreeMemory()
save.xlsx(fichier3, volcanom, volcano)#data pour faire les volcanos
save.xlsx(fichier, proteinGroups, NormalisationDESeq, Traitement_Nan, Proteins_
Deleted, presentAbsent1, presentsAbsent2, imputation)#tableau des différentes
étapes
save.xlsx(fichier2, testT_all, testT_matrisome, graphematrisome, recapitulatif,
recapitulatifquant, calculnan, graphe2, protsigNam, protsigNa)#tableau résultats
et tests T

```



# Bibliographie

---

1. Levecque, K., Anseel, F., De Beuckelaer, A., Van der Heyden, J. & Gisle, L. Work organization and mental health problems in PhD students. *Res. Policy* **46**, 868–879 (2017).
2. Iadecola, C. The Neurovascular Unit Coming of Age: A Journey through Neurovascular Coupling in Health and Disease. *Neuron* **96**, 17–42 (2017).
3. Iadecola, C. Neurovascular regulation in the normal brain and in Alzheimer’s disease. *Nature Reviews Neuroscience* **5**, 347–360 (2004).
4. Plog, B. A. *et al.* Biomarkers of traumatic injury are transported from brain to blood via the glymphatic system. *J. Neurosci.* **35**, 518–526 (2015).
5. Wardlaw, J. M. *et al.* Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet. Neurol.* **12**, 822–38 (2013).
6. Pantoni, L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol.* **9**, 689–701 (2010).
7. Joutel, A. Prospects for Diminishing the Impact of Nonamyloid Small-Vessel Diseases of the Brain. *Annu. Rev. Pharmacol. Toxicol.* **60**, 437–456 (2020).
8. Joutel, A., Haddad, I., Ratelade, J. & Nelson, M. T. Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain? *J. Cereb. Blood Flow Metab.* **36**, 143–157 (2016).
9. Mulvany, M. J. Small Artery Remodelling in Hypertension. *Basic Clin. Pharmacol. Toxicol.* **110**, 49–55 (2012).
10. Søndergaard, C. B., Nielsen, J. E., Hansen, C. K. & Christensen, H. Hereditary cerebral small vessel disease and stroke. *Clinical Neurology and Neurosurgery* **155**, 45–57 (2017).
11. Mancuso, M. *et al.* Monogenic cerebral small-vessel diseases: diagnosis and therapy. Consensus recommendations of the European Academy of Neurology. *Eur. J. Neurol.* **27**, 909–927 (2020).
12. Fryxell, K. J., Soderlund, M. & Jordan, T. V. An Animal Model for the Molecular Genetics of CADASIL. *Stroke* **32**, 6–11 (2001).
13. Joutel, A. Pathogenesis of CADASIL: Transgenic and knock-out mice to probe function and dysfunction of the mutated gene, Notch3, in the cerebrovasculature. *BioEssays* **33**, 73–80 (2011).
14. Chabriat, H., Joutel, A., Dichgans, M., Tournier-Lasserre, E. & Bousser, M.-G. CADASIL. *Lancet Neurol.* **8**, 643–653 (2009).
15. Van Bogaert, L. Encéphalopathie sous-corticale progressive (Binswanger) à évolution rapide chez deux soeurs. *Med Hell.* **24**, 961–72 (1955).
16. Baudrimont, M., Dubas, F., Joutel, A., Tournier-Lasserre, E. & Bousser, M. G. Autosomal dominant leukoencephalopathy and subcortical ischemic stroke: A clinicopathological study.

- Stroke* **24**, 122–125 (1993).
17. Joutel, A. *et al.* The ectodomain of the Notch3 receptor accumulates within the cerebrovasculature of CADASIL patients. *J. Clin. Invest.* **105**, 597–605 (2000).
  18. Ehebauer, M., Hayward, P. & Martinez Arias, A. Notch, a universal arbiter of cell fate decisions. *Science* **314**, 1414–1415 (2006).
  19. Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signaling: Cell fate control and signal integration in development. *Science* **284**, 770–776 (1999).
  20. Baeten, J. T. & Lilly, B. Notch Signaling in Vascular Smooth Muscle Cells. in *Advances in Pharmacology* **78**, 351–382 (2017).
  21. Krebs, L. T. *et al.* Characterization of Notch3-Deficient Mice: Normal Embryonic Development and Absence of Genetic Interactions with a Notch1 Mutation. *Genesis* **37**, 139–143 (2003).
  22. Domenga, V. *et al.* Notch3 is required for arterial identity and maturation of vascular smooth muscle cells. *Genes Dev.* **18**, 2730–2735 (2004).
  23. Joutel, A. *et al.* Strong clustering and stereotyped nature of Notch3 mutations in CADASIL patients. *Lancet* **350**, 1511–1515 (1997).
  24. Federico, A., Bianchi, S. & Dotti, M. T. The spectrum of mutations for CADASIL diagnosis. *Neurol. Sci.* **26**, 117–124 (2005).
  25. Arboleda-Velasquez, J. F. *et al.* Hypomorphic Notch 3 alleles link Notch signaling to ischemic cerebral small-vessel disease. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E128-35 (2011).
  26. Joutel, A., Monet, M., Domenga, V., Riant, F. & Tournier-Lasserre, E. Pathogenic Mutations Associated with Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy Differently Affect Jagged1 Binding and Notch3 Activity via the RBP/JK Signaling Pathway. *Am. J. Hum. Genet.* **74**, 338–347 (2004).
  27. Monet-Lepretre, M. *et al.* Distinct phenotypic and functional features of CADASIL mutations in the Notch3 ligand binding domain. *Brain* **132**, 1601–1612 (2009).
  28. Baron-Menguy, C., Domenga-Denier, V., Ghezali, L., Faraci, F. M. & Joutel, A. Increased Notch3 Activity Mediates Pathological Changes in Structure of Cerebral Arteries. *Hypertension* **69**, 60–70 (2017).
  29. Neves, K. B. *et al.* ER stress and Rho kinase activation underlie the vasculopathy of CADASIL. *JCI Insight* **4**, (2019).
  30. Capone, C. *et al.* Mechanistic insights into a TIMP3-sensitive pathway constitutively engaged in the regulation of cerebral hemodynamics. *Elife* **5**, e17536 (2016).
  31. Capone, C. *et al.* Reducing Timp3 or vitronectin ameliorates disease manifestations in CADASIL mice. *Ann. Neurol.* **79**, 387–403 (2016).
  32. Maeda, S., Nakayama, H., Isaka, K., Aihara, Y. & Nemoto, S. Familial Unusual Encephalopathy of Binswanger's Type without Hypertension. *Psychiatry Clin. Neurosci.* **30**, 165–177 (1976).
  33. Hara, K. *et al.* Association of HTRA1 Mutations and Familial Ischemic Cerebral Small-Vessel Disease. *N. Engl. J. Med.* **360**, 1729–1739 (2009).

34. Ito, J. *et al.* Histopathologic features of an autopsied patient with cerebral small vessel disease and a heterozygous *HTRA1* mutation. *Neuropathology* **38**, 428–432 (2018).
35. Verdura, E. *et al.* Heterozygous *HTRA1* mutations are associated with autosomal dominant cerebral small vessel disease. *Brain* **138**, 2347–2358 (2015).
36. Oide, T. *et al.* Extensive loss of arterial medial smooth muscle cells and mural extracellular matrix in cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL). *Neuropathology* **28**, 132–142 (2008).
37. Clausen, T., Kaiser, M., Huber, R. & Ehrmann, M. HTRA proteases: Regulated proteolysis in protein quality control. *Nat. Rev. Mol. Cell Biol.* **12**, 152–162 (2011).
38. An, E., Sen, S., Park, S. K., Gordish-Dressman, H. & Hathout, Y. Identification of novel substrates for the serine protease *HTRA1* in the human RPE secretome. *Investig. Ophthalmol. Vis. Sci.* **51**, 3379–3386 (2010).
39. Yamamoto, Y. & Ihara, M. Disruption of transforming growth factor- $\beta$  superfamily signaling: A shared mechanism underlying hereditary cerebral small vessel disease. *Neurochemistry International* **107**, 211–218 (2017).
40. Shiga, A. *et al.* Cerebral small-vessel disease protein *HTRA1* controls the amount of TGF- $\beta$ 1 via cleavage of proTGF- $\beta$ 1. *Hum. Mol. Genet.* **20**, 1800–1810 (2011).
41. Graham, J. R., Chamberland, A., Lin, Q., Li, X. J. & Dai, D. Serine Protease *HTRA1* Antagonizes Transforming Growth Factor- $\beta$  Signaling by Cleaving its Receptors and Loss of *HTRA1* In Vivo Enhances Bone Formation. *PLoS One* **8**, 74094 (2013).
42. Oka, C. *et al.* *Htra1* serine protease inhibits signaling mediated by Tgf $\beta$  family proteins. *Development* **131**, 1041–1053 (2004).
43. Beaufort, N. *et al.* Cerebral small vessel disease-related protease *Htra1* processes latent TGF- $\beta$  binding protein 1 and facilitates TGF- $\beta$  signaling. *Proc. Natl. Acad. Sci.* **111**, 16496–16501 (2014).
44. Klose, R. *et al.* Loss of the serine protease *HTRA1* impairs smooth muscle cells maturation. *Sci. Rep.* **9**, 18224 (2019).
45. Smit, L. M., Barth, P. G., Valk, J. & Njiokiktjien, C. Familial porencephalic white matter disease in two generations. *Brain Dev.* **6**, 54–58 (1984).
46. Gould, D. B. *et al.* Mutations in *Col4a1* cause perinatal cerebral hemorrhage and porencephaly. *Science (80-. )*. **308**, 1167–1171 (2005).
47. Aguglia, U. *et al.* Suggestive evidence for linkage to chromosome 13qter for autosomal dominant type 1 porencephaly. *Neurology* **62**, 1613–1615 (2004).
48. Jeanne, M. *et al.* *COL4A2* mutations impair *COL4A1* and *COL4A2* secretion and cause hemorrhagic stroke. *Am. J. Hum. Genet.* **90**, 91–101 (2012).
49. Verbeek, E. *et al.* *COL4A2* mutation associated with familial porencephaly and small-vessel disease. *Eur. J. Hum. Genet.* **20**, 844–851 (2012).
50. Khoshnoodi, J., Pedchenko, V. & Hudson, B. G. Mammalian collagen IV. *Microsc. Res. Tech.* **71**, 357–370 (2008).

51. Jeanne, M., Jorgensen, J. & Gould, D. B. Molecular and genetic analyses of collagen type IV mutant mouse models of spontaneous intracerebral hemorrhage identify mechanisms for stroke prevention. *Circulation* **131**, 1555–1565 (2015).
52. Ratelade, J. *et al.* Severity of arterial defects in the retina correlates with the burden of intracerebral haemorrhage in COL4A1-related stroke. *J. Pathol.* **244**, 408–420 (2018).
53. Bordeu, T. de, Bordeu, F. de & Bordeu, A. de. *Recherches sur les maladies chroniques, leurs rapports avec les maladies aiguës, leurs périodes, leur nature ...* (Ed Ruault, 1775).
54. Vicq d’Azyr, F. *Traité d’anatomie et de physiologie avec des planches coloriées représentant au naturel les divers organes de l’Homme et des Animaux / vol. I. I*, (François Ambroise Didot l’aîné, 1786).
55. Bichat Xavier. *Traité des membranes en général et de diverses membranes en particulier /* (Gabon & Mequignon-Marvis, 1827).
56. Lewis, W. H. The adhesive quality of cells. *Anat. Rec.* **23**, 387–392 (1922).
57. Naba, A. *et al.* The Matrisome: In Silico Definition and In Vivo Characterization by Proteomics of Normal and Tumor Extracellular Matrices. *Mol. Cell. Proteomics* **11**, M111.014647 (2012).
58. Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **15**, 786–801 (2014).
59. Dupont, S. *Anatomie et physiopathologie en soins infirmiers. ANATOMIE ET PHYSIOPATHOLOGIE* (2018).
60. Theocharis, A. D., Skandalis, S. S., Gialeli, C. & Karamanos, N. K. Extracellular matrix structure. *Adv. Drug Deliv. Rev.* **97**, 4–27 (2016).
61. Alberts, B. *et al.* *Molecular biology of the cell.* (Garland Science, 2002).
62. Pearce, O. M. T. *et al.* Deconstruction of a Metastatic Tumor Microenvironment Reveals a Common Matrix Response in Human Cancers. *Cancer Discov.* **8**, 304–319 (2018).
63. Holland, A., Murphy, S., Dowling, P. & Ohlendieck, K. Pathoproteomic profiling of the skeletal muscle matrisome in dystrophinopathy associated myofibrosis. *Proteomics* **16**, 345–366 (2016).
64. Jung, J., Ryu, T., Hwang, Y., Lee, E. & Lee, D. Prediction of extracellular matrix proteins based on distinctive sequence and domain characteristics. *J. Comput. Biol.* **17**, 97–105 (2010).
65. Hohenester, E. & Engel, J. Domain structure and organisation in extracellular matrix proteins. *Matrix Biol.* **21**, 115–128 (2002).
66. Naba, A., Hoersch, S. & Hynes, R. O. Towards definition of an ECM parts list: An advance on GO categories. *Matrix Biology* **31**, 371–372 (2012).
67. Ricard-Blum, S. The Collagen Family. *Cold Spring Harb. Perspect. Biol.* **3**, a004978–a004978 (2011).
68. Mouw, J. K., Ou, G. & Weaver, V. M. Extracellular matrix assembly: A multiscale deconstruction. *Nature Reviews Molecular Cell Biology* **15**, 771–785 (2014).
69. Hulmes, D. J. S. Building Collagen Molecules, Fibrils, and Suprafibrillar Structures. *J. Struct.*

- Biol.* **137**, 2–10 (2002).
70. Page-McCaw, A., Ewald, A. J. & Werb, Z. Matrix metalloproteinases and the regulation of tissue remodelling. *Nat. Rev. Mol. Cell Biol.* **8**, 221–233 (2007).
  71. Martel-Pelletier, J., Boileau, C., Pelletier, J.-P. & Roughley, P. J. Cartilage in normal and osteoarthritis conditions. *Best Pract. Res. Clin. Rheumatol.* **22**, 351–384 (2008).
  72. Hjorten, R. *et al.* Type XXVII collagen at the transition of cartilage to bone during skeletogenesis. *Bone* **41**, 535–42 (2007).
  73. Hynes, R. O. & Naba, A. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harb. Perspect. Biol.* **4**, a004903 (2012).
  74. Elfenbein, A. & Simons, M. Auxiliary and autonomous proteoglycan signaling networks. in *Methods in Enzymology* **480**, 3–31 (Academic Press, 2010).
  75. Silbert, J. E. & Sugumaran, G. Intracellular membranes in the synthesis, transport, and metabolism of proteoglycans. *BBA - Reviews on Biomembranes* **1241**, 371–384 (1995).
  76. Prydz, K. & Dalen, K. T. Synthesis and sorting of proteoglycans. *J. Cell Sci.* **113 Pt 2**, 193–205 (2000).
  77. Iozzo, R. V. & Schaefer, L. Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biology* **42**, 11–55 (2015).
  78. Zhang, X., Lee, S. J., Young, M. F. & Wang, M. M. The Small Leucine-Rich Proteoglycan BGN Accumulates in CADASIL and Binds to NOTCH3. *Transl. Stroke Res.* **6**, 148–155 (2015).
  79. Brooks, S. A. Strategies for Analysis of the Glycosylation of Proteins: Current Status and Future Perspectives. *Mol. Biotechnol.* **43**, 76–88 (2009).
  80. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta - Gen. Subj.* **1473**, 4–8 (1999).
  81. Dwek, R. A. Glycobiology: Toward understanding the function of sugars. *Chem. Rev.* **96**, 683–720 (1996).
  82. Halper, J. & Kjaer, M. Basic components of connective tissues and extracellular matrix: Elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins and thrombospondins. *Adv. Exp. Med. Biol.* **802**, 31–47 (2014).
  83. Schwarzbauer, J. E. & DeSimone, D. W. Fibronectins, their fibrillogenesis, and in vivo functions. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
  84. Mosesson, M. W., Siebenlist, K. R. & Meh, D. A. The structure and biological features of fibrinogen and fibrin. in *Annals of the New York Academy of Sciences* **936**, 11–30 (2001).
  85. Mithieux, B. S. M. & Weiss, A. S. Elastin is a key extracellular matrix protein that is critical to the elasticity I . Elastic Fiber The extracellular matrix imparts structural integrity on the tissues and. *Advances* **70**, 437–461 (2006).
  86. Webb, E. C. Enzyme Nomenclature. Recommendations 1984. Supplement 2: corrections and additions. *Eur. J. Biochem.* **179**, 489–533 (1989).

87. Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. *MEROPS* : the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **42**, D503–D509 (2014).
88. Vallet, S. D. *et al.* Insights into the structure and dynamics of lysyl oxidase propeptide, a flexible protein with numerous partners. *Sci. Rep.* **8**, 11768 (2018).
89. Di Cera, E. Serine proteases. *IUBMB Life* **61**, 510–515 (2009).
90. Mohamed, M. M. & Sloane, B. F. multifunctional enzymes in cancer. *Nat. Rev. Cancer* **6**, 764–775 (2006).
91. Giebel, N. & Zigrino, P. A Disintegrin and Metalloprotease (ADAM): Historical Overview of Their Functions. *Toxins (Basel)*. **8**, 122 (2016).
92. Timpl, R. & Brown, J. C. Supramolecular assembly of basement membranes. *BioEssays* **18**, 123–132 (1996).
93. Mak, K. M. & Mei, R. Basement Membrane Type IV Collagen and Laminin: An Overview of Their Biology and Value as Fibrosis Biomarkers of Liver Disease. *Anat. Rec.* **300**, 1371–1390 (2017).
94. Kalluri, R. *Basement membranes: Structure, assembly and role in tumour angiogenesis.* *Nature Reviews Cancer* **3**, 422–433 (Nature Publishing Group, 2003).
95. Aumailley, M. *et al.* A simplified laminin nomenclature. *Matrix Biology* **24**, 326–332 (2005).
96. Timpl, R., Dziadek, M., Fujiwara, S., Nowack, H. & Wick, G. Nidogen: a new, self-aggregating basement membrane protein. *Eur. J. Biochem.* **137**, 455–465 (1983).
97. Carlin, B., Jaffe, R., Bender, B. & Chung, A. E. Entactin, a novel basal lamina-associated sulfated glycoprotein. *J. Biol. Chem.* **256**, 5209–14 (1981).
98. Murdoch, A. D., Dodge, G. R., Cohen, I., Tuan, R. S. & Iozzo, R. V. Primary structure of the human heparan sulfate proteoglycan from basement membrane (HSPG2/perlecan). A chimeric molecule with multiple domains homologous to the low density lipoprotein receptor, laminin, neural cell adhesion molecules, and epidermal growth factor. *J. Biol. Chem.* **267**, 8544–57 (1992).
99. Gubbiotti, M. A., Neill, T. & Iozzo, R. V. A current view of perlecan in physiology and pathology: A mosaic of functions. *Matrix Biol.* **57–58**, 285–298 (2017).
100. Granger, J., Cipolla, M. J. & Granger, D. N. *The Cerebral Circulation, Second Edition The Cerebral Circulation Second Edition Integrated Systems Physiology Integrated Systems Physiology.* (LSU Health Sciences, 2016).
101. Girouard, H. & Iadecola, C. Neurovascular coupling in the normal brain and in hypertension, stroke, and Alzheimer disease. *J. Appl. Physiol.* **100**, 328–335 (2006).
102. Haffner, C., Malik, R. & Dichgans, M. Genetic factors in cerebral small vessel disease and their impact on stroke and dementia. *J. Cereb. Blood Flow Metab.* **36**, 158–171 (2016).
103. Thomsen, M. S., Routhé, L. J. & Moos, T. The vascular basement membrane in the healthy and pathological brain. *J. Cereb. Blood Flow Metab.* **37**, 3300–3317 (2017).
104. Cox, T. R. & Erler, J. T. Remodeling and homeostasis of the extracellular matrix: Implications for fibrotic diseases and cancer. *DMM Disease Models and Mechanisms* **4**, 165–178 (2011).

105. Kopito, R. R. Aggresomes, inclusion bodies and protein aggregation. *Trends in Cell Biology* **10**, 524–530 (2000).
106. Haffner, C. Proteostasis in Cerebral Small Vessel Disease. *Frontiers in Neuroscience* **13**, (2019).
107. Kalluri, R. Basement membranes: structure, assembly and role in tumour angiogenesis. *Nat. Rev. Cancer* **3**, 422–433 (2003).
108. Pöschl, E. *et al.* Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development. *Development* **131**, 1619–1628 (2004).
109. Petittlerc, E. *et al.* New functions for non-collagenous domains of human collagen type IV. Novel integrin ligands inhibiting angiogenesis and tumor growth in vivo. *J. Biol. Chem.* **275**, 8051–8061 (2000).
110. Jakobsson, L. & van Meeteren, L. A. Transforming growth factor B family members in regulation of vascular function: In the light of vascular conditional knockouts. *Experimental Cell Research* **319**, 1264–1270 (2013).
111. Grau, S. *et al.* The role of human HtrA1 in arthritic disease. *J. Biol. Chem.* **281**, 6124–6129 (2006).
112. Chamberland, A. *et al.* Identification of a novel HtrA1-susceptible cleavage site in human aggrecan. Evidence for the involvement of HtrA1 in aggrecan proteolysis in vivo. *J. Biol. Chem.* **284**, 27352–27359 (2009).
113. Monet-Leprêtre, M. *et al.* Abnormal recruitment of extracellular matrix proteins by excess Notch3ECD: A new pathomechanism in CADASIL. *Brain* **136**, 1830–1845 (2013).
114. Kast, J. *et al.* Sequestration of latent TGF- $\beta$  binding protein 1 into CADASIL-related Notch3-ECD deposits. *Acta Neuropathol. Commun.* **2**, 96 (2014).
115. Zellner, A. *et al.* CADASIL brain vessels show a HTRA1 loss-of-function profile. *Acta Neuropathol.* 1–15 (2018). doi:10.1007/s00401-018-1853-8
116. Mathieson, T. *et al.* Systematic analysis of protein turnover in primary cells. *Nat. Commun.* **9**, 1–10 (2018).
117. Ihalainen, S. *et al.* Proteome analysis of cultivated vascular smooth muscle cells from a CADASIL patient. *Mol. Med.* **13**, 305–314 (2007).
118. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **11**, 1–11 (2012).
119. Barkovits, K., Tönges, L. & Marcus, K. CSF sample preparation for data-independent acquisition. in *Methods in Molecular Biology* **2044**, 61–67 (Humana Press Inc., 2019).
120. Vasilj, A., Gentzel, M., Ueberham, E., Gebhardt, R. & Shevchenko, A. Tissue Proteomics by One-Dimensional Gel Electrophoresis Combined with Label-Free Protein Quantification. *J. Proteome Res.* **11**, 3680–3689 (2012).
121. Dapic, I., Uwugiaren, N., Jansen, P. J. & Corthals, G. L. Fast and Simple Protocols for Mass Spectrometry-Based Proteomics of Small Fresh Frozen Uterine Tissue Sections. *Anal. Chem.* **89**, 10769–10775 (2017).

122. Lin, J. S. & Lai, E. M. Protein–protein interactions: Co-immunoprecipitation. in *Methods in Molecular Biology* **1615**, 211–219 (Humana Press Inc., 2017).
123. Luo, S. Y., Araya, L. E. & Julien, O. Protease Substrate Identification Using N-terminomics. *ACS Chem. Biol.* **14**, 2361–2371 (2019).
124. Larsen, M. R., Trelle, M. B., Thingholm, T. E. & Jensen, O. N. Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *BioTechniques* **40**, 790–798 (2006).
125. Lu, B., McClatchy, D. B., Kim, J. Y. & Yates, J. R. Strategies for shotgun identification of integral membrane proteins by tandem mass spectrometry. *Proteomics* **8**, 3947–3955 (2008).
126. Moore, S. M., Hess, S. M. & Jorgenson, J. W. Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics. *J. Proteome Res.* **15**, 1243–1252 (2016).
127. Malherbe, G., Humphreys, D. P. & Davé, E. A robust fractionation method for protein subcellular localization studies in *Escherichia coli*. *Biotechniques* **66**, 171–178 (2019).
128. Wang, H. *et al.* An off-line high pH reversed-phase fractionation and nano-liquid chromatography-mass spectrometry method for global proteomic profiling of cell lines. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **974**, 90–95 (2015).
129. Byron, A., Humphries, J. D. & Humphries, M. J. Defining the extracellular matrix using proteomics. *Int. J. Exp. Pathol.* **94**, 75–92 (2013).
130. Gilbert, T., Sellaro, T. & Badylak, S. Decellularization of tissues and organs. *Biomaterials* **27**, 3675–3683 (2006).
131. De Castro Brás, L. E. *et al.* Texas 3-Step decellularization protocol: Looking at the cardiac extracellular matrix. *J. Proteomics* **86**, 43–52 (2013).
132. Xu, H. *et al.* Comparison of decellularization protocols for preparing a decellularized porcine annulus fibrosus scaffold. *PLoS One* **9**, e86723 (2014).
133. Hellewell, A. L., Rosini, S. & Adams, J. C. A Rapid, Scalable Method for the Isolation, Functional Study, and Analysis of Cell-derived Extracellular Matrix. *J. Vis. Exp.* **2017**, 55051 (2017).
134. Krasny, L. *et al.* Comparative proteomic assessment of matrixome enrichment methodologies. *Biochem. J.* **473**, 3979–3995 (2016).
135. Johnson, T. D. *et al.* Quantification of decellularized human myocardial matrix: A comparison of six patients. *Proteomics - Clinical Applications* **10**, 75–83 (2016).
136. Naba, A., Clauser, K. R., Mani, D. R., Carr, S. A. & Hynes, R. O. Quantitative proteomic profiling of the extracellular matrix of pancreatic islets during the angiogenic switch and insulinoma progression. *Sci. Rep.* **7**, 40495 (2017).
137. Didangelos, A. *et al.* Extracellular Matrix Composition and Remodeling in Human Abdominal Aortic Aneurysms: A Proteomics Approach. *Mol. Cell. Proteomics* **10**, M111.008128 (2011).
138. Hill, R. C., Calle, E. A., Dzieciatkowska, M., Niklason, L. E. & Hansen, K. C. Quantification of Extracellular Matrix Proteins from a Rat Lung Scaffold to Provide a Molecular Readout for Tissue Engineering. *Mol. Cell. Proteomics* **14**, 961–973 (2015).
139. Kuljanin, M., Brown, C. F. C., Raleigh, M. J., Lajoie, G. A. & Flynn, L. E. Collagenase treatment enhances proteomic coverage of low-abundance proteins in decellularized matrix

- bioscaffolds. *Biomaterials* **144**, 130–143 (2017).
140. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
  141. Aird, W. C. Endothelial cell heterogeneity. *Cold Spring Harb. Perspect. Med.* **2**, (2012).
  142. Simpson, R. J. Homogenization of Mammalian Tissue. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5455-pdb.prot5455 (2010).
  143. Liu, D., Zeng, X. A., Sun, D. W. & Han, Z. Disruption and protein release by ultrasonication of yeast cells. *Innov. Food Sci. Emerg. Technol.* **18**, 132–137 (2013).
  144. Nanduri, B. *et al.* Use of focused ultrasonication in activity-based profiling of deubiquitinating enzymes in tissue. *Anal. Biochem.* **515**, 9–13 (2016).
  145. Wu, D. *et al.* Biological and conventional food processing modifications on food proteins: Structure, functionality, and bioactivity. *Biotechnol. Adv.* **40**, 107491 (2020).
  146. Heerklotz, H. Interactions of surfactants with lipid membranes. *Q. Rev. Biophys.* **41**, 205–264 (2008).
  147. Loo, R. R. O., Dales, N. & Andrews, P. C. Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein Sci.* **3**, 1975–1983 (1994).
  148. Botelho, D. *et al.* Top-down and bottom-up proteomics of sds-containing solutions following mass-based separation. *J. Proteome Res.* **9**, 2863–2870 (2010).
  149. Bosserhoff, A., Wallach, J. & Frank, R. W. Micropreparative separation of peptides derived from sodium dodecyl sulphate-solubilized proteins. *J. Chromatogr. A* **473**, 71–77 (1989).
  150. Wiśniewski, J. R. *et al.* Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
  151. Manza, L. L., Stamer, S. L., Ham, A. J. L., Codreanu, S. G. & Liebler, D. C. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **5**, 1742–1745 (2005).
  152. Jiang, L., He, L. & Fountoulakis, M. Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *J. Chromatogr. A* **1023**, 317–320 (2004).
  153. LAEMMLI, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680–685 (1970).
  154. Kachuk, C., Stephen, K. & Doucette, A. Comparison of sodium dodecyl sulfate depletion techniques for proteome analysis by mass spectrometry. *J. Chromatogr. A* **1418**, 158–166 (2015).
  155. Glatter, T., Ahrné, E. & Schmidt, A. Comparison of different sample preparation protocols reveals lysis buffer-specific extraction biases in gram-negative bacteria and human cells. *J. Proteome Res.* **14**, 4472–4485 (2015).
  156. Murgia, M. *et al.* Single Muscle Fiber Proteomics Reveals Fiber-Type-Specific Features of Human Muscle Aging. *Cell Rep.* **19**, 2396–2409 (2017).
  157. León, I. R. *et al.* Quantitative Assessment of In-solution Digestion Efficiency Identifies Optimal Protocols for Unbiased Protein Analysis. *Mol. Cell. Proteomics* **12**, 2992–3005 (2013).

158. Farsang, E., Gaál, V., Horváth, O., Bárdos, E. & Horváth, K. Analysis of non-ionic surfactant Triton X-100 using hydrophilic interaction liquid chromatography and mass spectrometry. *Molecules* **24**, (2019).
159. Liu, J. *et al.* High-sensitivity n-glycoproteomic analysis of mouse brain tissue by protein extraction with a mild detergent of N-dodecyl  $\beta$ -d-maltoside. *Anal. Chem.* **87**, 2054–2057 (2015).
160. Ujihara, T., Sakurai, I., Mizusawa, N. & Wada, H. A method for analyzing lipid-modified proteins with mass spectrometry. *Anal. Biochem.* **374**, 429–431 (2008).
161. Henningsen, R., Gale, B. L., Straub, K. M. & DeNagel, D. C. Application of zwitterionic detergents to the solubilization of integral membrane proteins for two-dimensional gel electrophoresis and mass spectrometry. *Proteomics* **2**, 1479–1488 (2002).
162. Lucche, S., Santoni, V. & Rabilloud, T. Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics* **3**, (2003).
163. Rabilloud, T. Detergents and chaotropes for protein solubilization before two-dimensional electrophoresis. *Methods Mol. Biol.* **528**, 259–67 (2009).
164. Brownridge, P. & Beynon, R. J. The importance of the digest: Proteolysis and absolute quantification in proteomics. *Methods* **54**, 351–360 (2011).
165. Swaney, D. L., Wenger, C. D. & Coon, J. J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J. Proteome Res.* **9**, 1323–1329 (2010).
166. Shi, Y., Xiang, R., Horváth, C. & Wilkins, J. A. The role of liquid chromatography in proteomics. *J. Chromatogr. A* **1053**, 27–36 (2004).
167. Shen, Y. & Smith, R. D. Proteomics based on high-efficiency capillary separations. *Electrophoresis* **23**, 3106–3124 (2002).
168. Kota, U. & Stolowitz, M. L. Improving proteome coverage by reducing sample complexity via chromatography. in *Advances in Experimental Medicine and Biology* **919**, 83–143 (Springer New York LLC, 2016).
169. Sandra, K. *et al.* Highly efficient peptide separations in proteomics. Part 2: Bi- and multidimensional liquid-based separation techniques. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **877**, 1019–1039 (2009).
170. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).
171. Tanaka, K. *et al.* Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151–153 (1988).
172. Hardman, M. & Makarov, A. A. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal. Chem.* **75**, 1699–1705 (2003).
173. Penning, F. M. Die glimmentladung bei niedrigem druck zwischen koaxialen zylindern in einem axialen magnetfeld. *Physica* **3**, 873–894 (1936).
174. Dehmelt, H. G. Radiofrequency Spectroscopy of Stored Ions I: Storage. in *National Bureau of Standards, Special Publication V*, 53–72 (1968).

175. Kingdon, K. H. A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phys. Rev.* **21**, 408–418 (1923).
176. Makarov, A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* **72**, 1156–1162 (2000).
177. Hu, Q. *et al.* The Orbitrap: A new mass spectrometer. *Journal of Mass Spectrometry* **40**, 430–443 (2005).
178. Olsen, J. V. *et al.* Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **4**, 709–712 (2007).
179. Olsen, J. V. *et al.* A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **8**, 2759–2769 (2009).
180. Michalski, A. *et al.* Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol. Cell. Proteomics* **10**, M111.011015 (2011).
181. Scheltema, R. A. *et al.* The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer. *Mol. Cell. Proteomics* **13**, 3698–3708 (2014).
182. Tabb, D. L. *et al.* Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography–Tandem Mass Spectrometry. *J. Proteome Res.* **9**, 761–776 (2010).
183. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: A tutorial. *Molecular Systems Biology* **4**, 222 (2008).
184. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* **11**, 1475–1488 (2012).
185. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39–45 (2004).
186. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. C. & Geromanos, S. J. Absolute quantification of proteins by LCMSE: A virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **5**, 144–156 (2006).
187. Palz, B. & Suhal, S. Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews* **24**, 508–548 (2005).
188. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biological Mass Spectrometry* **11**, 601–601 (1984).
189. Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biol. Mass Spectrom.* **16**, 99–111 (1988).
190. Sleno, L. & Volmer, D. A. Ion activation methods for tandem mass spectrometry. *Journal of Mass Spectrometry* **39**, 1091–1112 (2004).
191. Hart-Smith, G. A review of electron-capture and electron-transfer dissociation tandem mass spectrometry in polymer chemistry. *Analytica Chimica Acta* **808**, 44–55 (2014).

192. Verheggen, K., Martens, L., Berven, F. S., Barsnes, H. & Vaudel, M. Database search engines: Paradigms, challenges and solutions. *Adv. Exp. Med. Biol.* **919**, 147–156 (2016).
193. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
194. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467 (2004).
195. Cottrell, J. S. & London, U. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
196. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–805 (2011).
197. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
198. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
199. Elias, J. E., Haas, W., Faherty, B. K. & Gygi, S. P. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2**, 667–675 (2005).
200. Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *J. Proteome Res.* **7**, 40–44 (2008).
201. Bantscheff, M., Lemeer, S., Savitski, M. M. & Kuster, B. Quantitative mass spectrometry in proteomics: Critical review update from 2007 to the present. *Analytical and Bioanalytical Chemistry* **404**, 939–965 (2012).
202. Oda, Y., Huang, K., Cross, F. R., Cowburn, D. & Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6591–6596 (1999).
203. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
204. Paša-Tolic, L. *et al.* High throughput proteome-wide precision measurements of protein expression using mass spectrometry [10]. *Journal of the American Chemical Society* **121**, 7949–7950 (1999).
205. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
206. Chen, X., Smith, L. M. & Bradbury, E. M. Site-specific mass tagging with stable isotopes in proteins for accurate and efficient protein identification. *Anal. Chem.* **72**, 1134–1143 (2000).
207. Ross, P. L. *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169 (2004).
208. Thompson, A. *et al.* Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003).
209. Yao, X., Freas, A., Ramirez, J., Demirev, P. A. & Fenselau, C. Proteolytic <sup>18</sup>O labeling for

- comparative proteomics: Model studies with two serotypes of adenovirus. *Anal. Chem.* **73**, 2836–2842 (2001).
210. Dannenmaier, S. *et al.* Complete Native Stable Isotope Labeling by Amino Acids of *Saccharomyces cerevisiae* for Global Proteomic Analysis. *Anal. Chem.* **90**, 10501–10509 (2018).
211. Krüger, M. *et al.* SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell* **134**, 353–364 (2008).
212. Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E. & Yates, J. R. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal. Chem.* **76**, 4951–4959 (2004).
213. Gruhler, A., Schulze, W. X., Matthiesen, R., Mann, M. & Jensen, O. N. Stable isotope labeling of *Arabidopsis thaliana* cells and quantitative proteomics by mass spectrometry. *Mol. Cell. Proteomics* **4**, 1697–1709 (2005).
214. Olsen, J. V., Ong, S. E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **3**, 608–614 (2004).
215. Zee, B. M. & Garcia, B. A. Discovery of lysine post-translational modifications through mass spectrometric detection. *Essays Biochem.* **52**, 147–163 (2012).
216. Rose, K., Savoy, L. A., Simona, M. G., Offord, R. E. & Wingfield, P. C-terminal peptide identification by fast atom bombardment mass spectrometry. *Biochem. J.* **250**, 253–259 (1988).
217. Ramos-Fernández, A., López-Ferrer, D. & Vázquez, J. Improved method for differential expression proteomics using trypsin-catalyzed <sup>18</sup>O labeling with a correction for labeling efficiency. *Mol. Cell. Proteomics* **6**, 1274–1286 (2007).
218. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.* **389**, 1017–1031 (2007).
219. Chahrour, O., Cobice, D. & Malone, J. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *Journal of Pharmaceutical and Biomedical Analysis* **113**, 2–20 (2015).
220. Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231–1245 (2002).
221. Sanders, S. L., Jennings, J., Canutescu, A., Link, A. J. & Weil, P. A. Proteomics of the Eukaryotic Transcription Machinery: Identification of Proteins Associated with Components of Yeast TFIID by Multidimensional Mass Spectrometry. *Mol. Cell. Biol.* **22**, 4723–4738 (2002).
222. Liu, H., Sadygov, R. G. & Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201 (2004).
223. Davis, M. T. *et al.* Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry II. Limitations of complex mixture analyses. *Proteomics* **1**, 108–117 (2001).
224. Zhang, Y., Wen, Z., Washburn, M. P. & Florens, L. Effect of dynamic exclusion duration on spectral count based quantitative proteomics. *Anal. Chem.* **81**, 6317–6326 (2009).

225. Bondarenko, P. V., Chelius, D. & Shaler, T. A. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography - Tandem mass spectrometry. *Anal. Chem.* **74**, 4741–4749 (2002).
226. Matzke, M. M. *et al.* A comparative analysis of computational approaches to relative protein quantification using peptide peak intensities in label-free LC-MS proteomics experiments. *Proteomics* **13**, 493–503 (2013).
227. Blein-Nicolas, M. & Zivy, M. Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochim. Biophys. Acta - Proteins Proteomics* **1864**, 883–895 (2016).
228. Sandin, M., Teleman, J., Malmström, J. & Levander, F. Data processing methods and quality control strategies for label-free LC-MS protein quantification. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1844**, 29–41 (2014).
229. Park, T. *et al.* Evaluation of normalization methods for microarray data. *BMC Bioinformatics* **4**, 1–13 (2003).
230. Quackenbush, J. Microarray data normalization and transformation. *Nature Genetics* **32**, 496–501 (2002).
231. Karpievitch, Y. V., Dabney, A. R. & Smith, R. D. Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics* **13 Suppl 16**, S5 (2012).
232. Callister, S. J. *et al.* Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* **5**, 277–286 (2006).
233. Kultima, K. *et al.* Development and evaluation of normalization methods for label-free relative quantification of endogenous peptides. *Mol. Cell. Proteomics* **8**, 2285–95 (2009).
234. Chawade, A., Alexandersson, E. & Levander, F. Normalyzer: A Tool for Rapid Evaluation of Normalization Methods for Omics Data Sets. *J. Proteome Res.* **13**, 3114–3120 (2014).
235. Välikangas, T., Suomi, T. & Elo, L. L. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinform.* **19**, 1–11 (2016).
236. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
237. Bo., N., Little, R. J. A. & Rubin, D. B. Statistical Analysis with Missing Data. *Popul. (French Ed.* **43**, 1174 (1988).
238. Yang, J., Zhao, X., Lu, X., Lin, X. & Xu, G. A data preprocessing strategy for metabolomics to reduce the mask effect in data analysis. *Front. Mol. Biosci.* **2**, 4 (2015).
239. Webb-Robertson, B.-J. M. *et al.* Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics. *J. Proteome Res.* **14**, 1993–2001 (2015).
240. Wei, R. *et al.* Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Sci. Rep.* **8**, 663 (2018).
241. Badhwar, A., Stanimirovic, D. B., Hamel, E. & Haqqani, A. S. The proteome of mouse cerebral arteries. *J. Cereb. Blood Flow Metab.* **34**, 1033–46 (2014).

242. Chun, H. B. *et al.* The proteome of mouse brain microvessel membranes and basal lamina. *J. Cereb. Blood Flow Metab.* **31**, 2267–2281 (2011).
243. Hondius, D. C. *et al.* Proteomics analysis identifies new markers associated with capillary cerebral amyloid angiopathy in Alzheimer’s disease. *Acta Neuropathol. Commun.* **6**, 46 (2018).
244. Johnson, T. D. *et al.* Quantification of decellularized human myocardial matrix: A comparison of six patients. *Proteomics - Clinical Applications* **10**, 75–83 (2016).
245. Barallobre-Barreiro, J. *et al.* Extracellular matrix remodelling in response to venous hypertension: proteomics of human varicose veins. *Cardiovasc. Res.* **110**, 419–30 (2016).
246. Didangelos, A. *et al.* Proteomics Characterization of Extracellular Space Components in the Human Aorta. *Mol. Cell. Proteomics* **9**, 2048–2062 (2010).
247. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
248. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5116–21 (2001).
249. Bø, T. H. *et al.* LImpute: accurate estimation of missing values in microarray data with least squares methods. *Nucleic Acids Res.* **32**, e34–e34 (2004).
250. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nature Genetics* **25**, 25–29 (2000).
251. Tan, R., Traylor, M., Rutten-Jacobs, L. & Markus, H. New insights into mechanisms of small vessel disease stroke from genetics. *Clin. Sci.* **131**, 515–531 (2017).
252. Joutel, A. *et al.* Cerebrovascular dysfunction and microcirculation rarefaction precede white matter lesions in a mouse genetic model of cerebral ischemic small vessel disease. *J. Clin. Invest.* **120**, 433–445 (2010).
253. Ratelade, J. *et al.* Reducing Hypermuscularization of the Transitional Segment between Arterioles and Capillaries Protects against Spontaneous Intracerebral Hemorrhage. *Circulation* **141**, 2078–2094 (2020).
254. Schlager, G. Selection for blood pressure levels in mice. *Genetics* **76**, 537–549 (1974).
255. Mao, M., Alavi, M. V., Labelle-Dumais, C. & Gould, D. B. Type IV Collagens and Basement Membrane Diseases: Cell Biology and Pathogenic Mechanisms. *Curr. Top. Membr.* **76**, 61–116 (2015).
256. Kast, J. *et al.* Sequestration of latent TGF- $\beta$  binding protein 1 into CADASIL-related Notch3-ECD deposits. *Acta Neuropathol. Commun.* **2**, 96 (2014).
257. Noda, K. *et al.* Latent TGF- $\beta$  binding protein 4 promotes elastic fiber assembly by interacting with fibulin-5. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2852–2857 (2013).
258. Bugiani, M. *et al.* Cathepsin A-related arteriopathy with strokes and leukoencephalopathy (CARASAL). *Neurology* **87**, 1777–1786 (2016).
259. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).

260. Tsamis, A., Krawiec, J. T. & Vorp, D. A. Elastin and collagen fibre microstructure of the human aorta in ageing and disease: A review. *Journal of the Royal Society Interface* **10**, (2013).
261. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
262. Wuyts, F. L. *et al.* Elastic properties of human aortas in relation to age and atherosclerosis: A structural model. *Phys. Med. Biol.* **40**, 1577–1597 (1995).
263. Fonck, E. *et al.* Effect of elastin degradation on carotid wall mechanics as assessed by a constituent-based biomechanical model. *Am. J. Physiol. - Hear. Circ. Physiol.* **292**, 2754–2763 (2007).
264. Xu, X. *et al.* Age-related impairment of vascular structure and functions. *Aging and Disease* **8**, 590–610 (2017).
265. Chen, I. H. *et al.* PRSS23 is essential for the Snail-dependent endothelial-to-mesenchymal transition during valvulogenesis in zebrafish. *Cardiovasc. Res.* **97**, 443–453 (2013).
266. LeBleu, V. S. *et al.* Identification of human epididymis protein-4 as a fibroblast-derived mediator of fibrosis. *Nat. Med.* **19**, 227–231 (2013).
267. Jones, A. *et al.* Increased expression of multifunctional serine protease, HTRA1, in retinal pigment epithelium induces polypoidal choroidal vasculopathy in mice. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14578–14583 (2011).
268. Lin, M. K. *et al.* HTRA1, an age-related macular degeneration protease, processes extracellular matrix proteins EFEMP1 and TSP1. *Aging Cell* **17**, 12710 (2018).
269. Tjaden, A. N. & Richards, P. J. The emerging roles of HTRA1 in musculoskeletal disease. *American Journal of Pathology* **182**, 1482–1488 (2013).
270. Uhlen, M. *et al.* A proposal for validation of antibodies. *Nature Methods* **13**, 823–827 (2016).
271. DR, B. *et al.* Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. *Anal. Chem.* **75**, (2003).
272. Agouridas, V. *et al.* Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chemical Reviews* **119**, (2019).
273. Mahmood, S. S. Développement de nouvelles stratégies analytiques pour la caractérisation moléculaire des états d'oxydation à l'échelle protéomique. *HAL Id tel-01299792* (2016).

## RÉSUMÉ

---

Les maladies des petits vaisseaux cérébraux sont responsables de lésions de la substance blanche du cerveau et d'infarctus cérébraux profonds multiples. C'est un ensemble de processus pathologiques, qui affectent les petites artères, les artérioles, veinules ou capillaires cérébraux de moins de 400µm. Le matrisome cérébrovasculaire semble être une voie pathologique convergente entre les différentes maladies des petits vaisseaux de type génétique et de type sporadique. La diversité structurale et physico-chimique des protéines du matrisome rend cependant leur analyse particulièrement délicate. En effet, certaines protéines du *core* matrisome, protéines de haut poids moléculaire, sont particulièrement insolubles et d'autres protéines du matrisome associé sont généralement plus petites et moins abondantes. Dans le cadre de cette thèse nous nous proposons de caractériser de manière quantitative et qualitative le matrisome microvasculaire dans les maladies des petits de vaisseaux, ainsi que d'identifier des anomalies communes ou spécifiques à chaque maladie. Pour cela nous avons, dans un premier temps, développé une approche protéomique quantitative en spectrométrie de masse de type *label-free* sur des vaisseaux cérébraux et périphériques isolés, et nous l'avons validée dans une première étude sur un modèle murin préclinique de CADASIL. Ensuite nous avons appliqué cette approche sur deux autres modèles murins pour les maladies génétiques CARASIL et la maladie du collagène de type IV, et deux modèles murins pour l'hypertension et l'âge, modèle pour leur caractère sporadique. Nous avons développé et validé une nouvelle méthode robuste et sensible pour l'analyse quantitative sans marquage des changements du matrisome des artères cérébrales de souris. En effet, nous avons mis en évidence une protéine commune à toutes ces formes de cSVDs, PRSS23, une serine protéase dans les 5 modèles étudiée et HTRA1, une autre serine protéase, dans CADASIL, la maladie du collagène de type IV et l'âge.

## MOTS CLÉS

---

Protéomique quantitative, quantification label free, matrisome, matrice extracellulaire, maladie des petits vaisseaux

## ABSTRACT

---

Diseases of the small vessels of the brain are responsible for damage to the white matter of the brain and multiple deep brain infarctions. They are the cause of more than 25% of strokes and are the second cause of dementia after Alzheimer's dementia. It is a set of pathological processes, which affect small arteries, arterioles, cerebral venule or capillary of less than 400µm. The cerebrovascular matrisome seems to be a converging pathological pathway between the various diseases of the small vessels of the genetic type and of the sporadic type. The matrisome is the set of proteins constituting the extracellular matrix (ECM) as well as the associated proteins, their roles consist not only in the support and the anchoring of the cells but also in various fundamental processes such as differentiation, proliferation, Survival or migration of cells. The structural and physico-chemical diversity of these proteins, however, makes their analysis particularly delicate. Within the framework of this thesis we propose to characterize in a quantitative and qualitative way the microvascular matrisome in the diseases of the small vessels, as well as to identify common abnormalities or specific to each disease. For this we have developed a label-free quantitative proteomic approach on cerebral and peripheral vessels isolated from three preclinical genetic murine models and two murine models for the sporadic character. We have developed and validated a new robust and sensitive method for the quantitative non-labeling analysis of changes in the matrisome of mouse cerebral arteries and the application of our method on the arteries of the different mouse models studied has allowed us to identify some avenues. Interesting for each disease independently but also highlighted some common signatures between the different studies.

## KEYWORDS

---

Quantitative proteomics, label free, matrisome, extracellular matrix, small vessel disease