



**HAL**  
open science

# Rôle de la nouvelle isoforme d'épissage du gène von Hippel-Lindau dans la survenue d'érythrocytoses et de cancers du rein héréditaires

Marion Lenglet

► **To cite this version:**

Marion Lenglet. Rôle de la nouvelle isoforme d'épissage du gène von Hippel-Lindau dans la survenue d'érythrocytoses et de cancers du rein héréditaires. Biologie moléculaire. Université Paris sciences et lettres, 2019. Français. NNT : 2019PSLEP035 . tel-03767550

**HAL Id: tel-03767550**

**<https://theses.hal.science/tel-03767550v1>**

Submitted on 2 Sep 2022

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

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



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

Préparée à l'École Pratique des Hautes Études

**Rôle de la nouvelle isoforme d'épissage du gène  
von Hippel-Lindau dans la survenue  
d'érythrocytoses et de cancers du rein héréditaires**

Soutenue par

**Marion LENGLET**

Le 24 septembre 2019

École doctorale n° 472

**École doctorale de l'École  
Pratique des Hautes Études**

Spécialité

**Biologie cellulaire et  
moléculaire et sciences  
de la santé**

Composition du jury :

Yannick, ARLOT DR1, CNRS	<i>Président</i>
Isabelle, PLO DR2, INSERM	<i>Rapporteur</i>
Claude, HOUDAYER Pharm.D, Ph.D, M.B.A, INSERM	<i>Rapporteur</i>
Yannick, ARLOT DR1, CNRS	<i>Examineur</i>
Frédéric, Mazurier DR, INSERM	<i>Examineur</i>
Nathalie, Mazure DR2, CNRS	<i>Membre invité</i>
Franck, CHESNEL CR, CNRS	<i>Membre invité</i>
Betty, GARDIE Maitre de Conférence, EPHE	<i>Directeur de thèse</i>







*On ne fait jamais attention à ce qui a été fait ;  
on ne voit que ce qui reste à faire.*  
Marie Curie (1867 – 1934)

## Remerciements

Je tiens tout d'abord à remercier chaleureusement Mme Isabelle PLO et M Claude HOUDAYER pour avoir accepté d'être rapporteur de ma thèse ainsi que Mme Yannick ARLOT et M Frédéric MAZURIER pour avoir accepté de faire partie du jury de ma thèse. Je remercie également Mme Nathalie MAZURE et M Franck CHESNEL pour m'avoir suivie durant ces 3 années lors de mes comités de suivis de thèse. Je leur suis très reconnaissante pour les remarques et les conseils qui ont contribué à l'amélioration de ce travail de thèse.

Je souhaite remercier ma directrice de thèse, Betty Gardie, pour m'avoir poussée à vivre cette aventure dans son laboratoire ainsi que pour la confiance qu'elle m'a accordée durant mon master puis ma thèse. Son soutien permanent et sa patience ont été une aide précieuse et m'ont permis de mener à bien mon projet.

Je tiens aussi à remercier Stéphane Bézieau, chef de l'équipe émergente qui m'a permis de réaliser ma thèse dans les meilleures conditions. Je remercie également tous les membres de cette équipe et plus particulièrement Amandine Leroy pour toute l'aide technique qu'elle m'a apporté depuis son arrivée ainsi que Thomas Besnard pour tous les conseils techniques et sa bonne humeur à toute épreuve.

Je remercie également le groupe de Karim Si-Tayeb avec Amandine Caillaud, Aurore Girardeau et Meryl Roudaut pour leur collaboration, leurs nombreux conseils, leur aide et leur écoute. Un grand merci aussi à Virginie Forest pour sa bienveillance et sa bonne humeur quotidienne.

Je n'oublie pas l'institut du thorax, son équipe de direction, son équipe technique ainsi que tout le personnel de recherche qui ont pu participer de près ou de loin à cette aventure de part leur aide et leur soutien ou de part leur bonne humeur. Je voulais simplement vous témoigner tout le plaisir que j'ai eu à travailler avec vous : « Tout seul on va plus vite, ensemble on va plus loin ».

Merci à ma famille et à mes amis qui ont une oreille attentive à tout moment et qui ont su me soutenir. Merci pour votre patience, votre présence et pour vos sourires.

Un grand merci, je vous souhaite à tous et à toutes beaucoup de bonheur et de réussite dans tous vos projets professionnels et personnels.

## Résumé

La protéine von Hippel-Lindau (VHL), codée par le gène *VHL* constitué de 3 exons, est un facteur clé de la régulation de la voie de l'hypoxie dont l'acteur majeur est le facteur de transcription inductible en hypoxie (HIF). En présence d'oxygène, HIF est reconnu par VHL puis ubiquitinyllée et dégradée via le protéasome. En absence d'oxygène, HIF est stabilisé et induit l'expression d'une centaine de gènes impliqués notamment dans l'angiogenèse (VEGF), la prolifération (TGF), le métabolisme (GLUT1) et l'érythropoïèse (EPO). Quand *VHL* est muté, HIF n'est plus dégradé et entraîne la surexpression de ces gènes qui peuvent conduire à la formation de tumeurs via l'hypervascularisation et l'hyperprolifération cellulaire ou à une érythropoïèse incontrôlée. Deux pathologies sont associées à des mutations de *VHL*. La polyglobulie, caractérisée par un taux de globules rouges sanguins et d'EPO élevés, a pour origine une mutation homozygote ou hétérozygote composite de *VHL*. La maladie de VHL, caractérisée par le développement de tumeurs hypervascularisées (hémangioblastomes du système nerveux central ou de la rétine, phéochromocytomes, paragangliomes ou cancers du rein), a pour origine une mutation hétérozygote de *VHL* associée à une perte d'hétérozygotie dans les tumeurs. Il a été mis en évidence que les mutations dans le gène *VHL* suivent un modèle d'oncogenèse en gradient : plus la perte de fonction de pVHL est importante, plus la pathologie développée est sévère.

Cependant, certains patients présentent un tableau clinique caractéristique d'une anomalie *VHL* sans aucune mutation (maladie de VHL) ou seulement à l'état hétérozygote (polyglobulie). Les mécanismes moléculaires à l'origine de ces phénotypes n'étant pas encore compris, l'objectif de ma thèse s'est porté sur l'étude de mutations complexes de *VHL*, à savoir des mutations impactant l'épissage du gène. Une étude approfondie a été réalisée sur 12 familles : étude de ségrégation, analyse phylogénétique, tests minigènes, séquençage de l'ARN des tumeurs et lignées lymphoblastoïdes de patients, mesure de l'expression des ARNm et des protéines, études fonctionnelles des nouvelles isoformes VHL. L'épissage étant un mécanisme dépendant du type cellulaire, j'ai également participé au développement d'un modèle de cellules productrices d'EPO

à partir de cellules souches pluripotentes induites (iPS) de patients afin de réaliser des études adaptées.

Le début de mon projet s'est porté sur l'étude de 2 mutations synonymes de l'exon 2 du gène *VHL*, à l'origine de la maladie de VHL ou d'une polyglobulie. J'ai pu mettre en évidence que ces mutations entraînent le saut de l'exon 2 dans les transcrits avec une diminution de l'isoforme sauvage (contenant les exons 1, 2 et 3) en faveur de celle contenant les exons 1 et 3, décrite comme non fonctionnelle. J'ai ainsi pu montrer que des mutations exoniques de *VHL* pouvaient être à l'origine de défauts d'épissage et entraîner une perte de fonction de *VHL* expliquant les phénotypes observés.

L'étude approfondie des transcrits *VHL* m'a permis de mettre en évidence la présence de 2 exons cryptiques situés dans les introns 1 et 2. Ces 2 exons nommés E1' et E2' ont été séquencés et retrouvés mutés dans les cas inexplicables de la maladie de VHL ou de polyglobulie. J'ai pu montrer que les mutations dans E1' conduisent à une rétention excessive de cet exon associée à une diminution de l'expression protéique totale de VHL. Dans les deux cas (saut de l'exon 2 ou rétention de E1'), l'impact sur l'épissage est plus important chez les patients avec tumeurs qu'avec une polyglobulie.

L'analyse des mutations de E2' n'a pas encore permis de déterminer les mécanismes d'action en jeu.

Lors de ma thèse, j'ai donc découvert une structure et une expression complexes du gène *VHL*. Mon travail a permis d'identifier l'origine de la pathologie chez des patients jusqu'ici sans diagnostic génétique, leur permettant un meilleur suivi médical et ouvrant sur de nouvelles pistes thérapeutiques.

## Abstract

The von Hippel-Lindau protein (VHL), encoded by the *VHL* gene that contains three exons, is a key factor of the hypoxia pathway. The main player of this pathway is the hypoxia inducible factor (HIF). In presence of oxygen, HIF is recognized by VHL, subunit of an E3 ubiquitin ligase complex, then ubiquitinated and degraded by the proteasome. In absence of oxygen, HIF is stabilized and induces the expression of hundred of genes involved in angiogenesis (VEGF), proliferation (TGF), metabolism (GLUT1) and erythropoiesis (EPO). When *VHL* is mutated, HIF is not degraded and induces the overexpression of these genes which can led to the formation of tumors via the hypervascularisation and the hyperproliferation of the cells or to an excessive erythropoiesis. Two pathologies are associated with mutations in *VHL*. The erythrocytosis is characterized by an elevated level of red blood cells and EPO serum level associated with homozygous or compound heterozygous mutations of *VHL*. The VHL disease is characterized by the development of hypervascularized tumors (hemangioblastomas of the central nervous system or of the retina, pheochromocytomas, paragangliomas or renal cancers) associated with heterozygous mutations of *VHL* and a loss of heterozygosity in tumors. It has been described that *VHL* mutations followed a continuum model of tumor suppression: the more severe is the loss of pVHL function the more severe is the developed pathology.

However, some patients presented with a clinical aspects typical of *VHL* mutations but without any identified mutation (VHL disease) or with only heterozygous mutation (erythrocytosis). The molecular mechanisms at the origin of these phenotypes are not yet understood, the aim of my thesis was to study complex mutations of *VHL*, namely mutations that impact *VHL* splicing. A comprehensive study has been performed on twelve families: segregation study, phylogenetic analysis, minigene experiments and sequencing of RNA from tumors or lymphoblastoid cell lines of patients, measurement of the expression of mRNA and proteins, functional study of the new *VHL* isoforms. The splicing being specific to cell types, I also participated to the development of an EPO producing cellular model established from induced pluripotent stem cells (iPS) derived from patients.

The beginning of my project was focused on the study of two synonymous mutations located in the second exon of the *VHL* gene, one associated with a VHL disease and one associated with an erythrocytosis. I demonstrated that these mutations induce the *VHL* exon 2 skipping associated with a decreased expression of the normal transcript (containing exon 1, 2 and 3 spliced) in favor of transcripts that contain exon 1 spliced with exon 3, a non functional isoform. I showed that exonic mutations in *VHL* can induce splicing defects and led to a loss of function of *VHL* at the origin the observed phenotypes.

In-depth study of *VHL* transcripts allowed the discovery of two cryptic exons located in the introns 1 and 2 of *VHL*. These two exons named E1' and E2' have been sequenced and found mutated in unsolved cases of VHL disease or erythrocytosis. I showed that mutations in E1' led to an excessive retention of this exon associated with a decreased expression of VHL protein. In addition, comparative studies of synonymous and E1'- *VHL* mutations associated with von Hippel-Lindau or erythrocytosis confirmed the continuum model of tumor suppression.

Analyses of the mutations in the E2' exon have not yet permitted to identify the processes involved in the developed diseases.

During my thesis, I have therefore demonstrated that exon-skipping and cryptic-exon-retention are new *VHL* alterations and I identified a novel complex splicing regulation of the *VHL* gene. This study allowed the identification of the patients' disease origin who had so far no genetic diagnosis. These results provided access of patients to a better medical monitoring and paved the way of further therapeutic prospects.

## Lexique

ADN : acide désoxyribonucléique

ARN : acide ribonucléique

ARNT : aryl hydrocarbon receptor nuclear translocator

BFU-E : Burst forming unit erythroid

CARD9 : caspase recruitment domain-containing protein 9

CFU-E : colony forming unit erythroid

Chk2 : checkpoint kinase 2

COL4A2 : collagène IV alpha 2

CUL2 : culline 2

EBV : Epstein-Barr virus

EJC : exons junction complexes

EPO : erythropoïétine

EPOR : récepteur à l'érythropoïétine

ESE : exon splicing enhancer

ESI : exon splicing silencer

ESR : séquences régulatrices d'exons

FIH : facteur inhibiteur de HIF

HIF : facteur inductible en hypoxie (hypoxia inducible factor)

HRE : élément de réponse à l'hypoxie (hypoxia response element)

IPAS : inhibitory Per/Arnt/Sim

ISE : intronic splincing enhancer

ISI : intronic splincing inhibitor

JAK2 : janus kinase 2

KO : knock-out

LLB : lignée lymphoblastoïde

lncRNA : long ARN non codant

LOH : perte d'hétérozygotie (Loss of heterozygosity)

NAT : transcrit antisens naturel

NF $\kappa$ B : facteur nucléaire  $\kappa$ B (nuclear factor  $\kappa$ B)

NMD : nonsense mediated decay

ODD : domaine dépendant de l'oxygène (oxygen dependent domain)

ORF : cadre de lecture ouvert (open reading frame)

PDGF : facteur de croissance plaquettes-dérivé (platelet derived growth factor)

PHD : prolyl hydroxylase domain

qPCR : PCR quantitative (quantitative polymerase chain reaction)

Rbx1 : RING box protéine 1

RCC : carcinome rénal (renal cell carcinoma)

REPC : cellules rénales productrices d'érythropoïétine (renal EPO producing cells)

SASP : senescence associated secretory phenotype

SCF : facteur de cellules souches (stem cell factor)

SEM : erreur standard à la moyenne (standard error mean)

SOCS1 : suppressor of cytokine signalling 1

STAT5 : transducteur de signal et activateur de transcription 5 (signal transducer and activator of transcription)

TGF : facteur de croissance transformant (transforming growth factor)

UPF : up-frameshift protein

UTR : région non transcrites (untranslated region)

VEGF : facteur de croissance de l'endothélium vasculaire (vascular epithelial growth factor)

VHL : von Hippel-Lindau

VLP : VHL-like protéine

# Sommaire

<b>I.</b>	<b>INTRODUCTION</b>	<b>1</b>
<b>A.</b>	<b>L'HYPOXIE</b>	<b>1</b>
1.	GENERALITES SUR LA REPONSE A L'HYPOXIE	1
2.	HIF : FACTEUR CLE DE LA VOIE DE L'HYPOXIE	3
3.	ROLE ET ORIGINE DE L'ERYTHROPOÏÉTINE (EPO)	8
<b>B.</b>	<b>DESCRIPTION DU GENE VHL ET FONCTION DE PVHL</b>	<b>11</b>
1.	LE GENE VHL	11
2.	PROTEINES VHL ET FONCTIONS DEPENDANTES DE HIF	12
3.	PROTEINES VHL ET FONCTIONS INDEPENDANTES DE HIF	13
a.	<i>Régulation de la transcription</i>	14
b.	<i>Maintenance du cil primaire et stabilisation des microtubules</i>	14
c.	<i>Régulation de la formation de la matrice extracellulaire</i>	15
d.	<i>Régulation de l'apoptose</i>	16
e.	<i>Régulation de la survie et de la prolifération des progéniteurs érythroïdes</i>	17
<b>C.</b>	<b>MALADIES DUES A DES ANOMALIES DE VHL</b>	<b>18</b>
1.	LA MALADIE DE VHL	18
a.	<i>Généralités</i>	18
b.	<i>Tumeurs du spectre VHL<sup>102,88</sup></i>	21
i.	<i>Hémangioblastomes du système nerveux central</i>	21
ii.	<i>Hémangioblastomes rétinien</i>	22
iii.	<i>Carcinome rénal à cellules claires (RCC : renal cell carcinoma)</i>	22
iv.	<i>Phéochromocytomes</i>	23
v.	<i>Tumeurs neuroendocrines du pancréas</i>	24
vi.	<i>Tumeurs du sac endolymphatique</i>	24
vii.	<i>Tumeurs kystiques de l'épididyme</i>	24
2.	LA POLYGLOBULIE DE CHUVASH	25
a.	<i>Les différents types de polyglobulie</i>	25
b.	<i>La polyglobulie de Chuvash</i>	26
3.	COMPLEXITE DES CORRELATIONS GENOTYPE/PHENOTYPE	27
a.	<i>Modèle d'oncogenèse en gradient</i>	28
b.	<i>Mutations complexes de VHL</i>	31
4.	MODELE EXISTANTS : LIGNEES CELLULAIRES, IPS, MODELES ANIMAUX	32
a.	<i>Lignées cellulaires</i>	33
b.	<i>Modèle iPS</i>	33
c.	<i>Modèles animaux</i>	34
<b>D.</b>	<b>ISOFORMES D'EPISSAGE</b>	<b>35</b>
5.	EPISSAGE ALTERNATIF ET EPISSAGE PATHOLOGIQUE : ORIGINES, PRODUITS ET METHODES D'ANALYSE	35
a.	<i>Description</i>	35
b.	<i>Perturbations de l'épissage et impacts en pathologies</i>	38
c.	<i>Le mécanisme de Nonsense Mediated Decay (NMD)</i>	40
6.	LES LONGS ARN NON CODANT	42
<b>E.</b>	<b>OBJECTIFS DE THESE</b>	<b>44</b>
<b>II.</b>	<b>MATERIELS ET METHODES</b>	<b>46</b>
i.	Extraction d'acides nucléiques	46

ii.	Reverse transcription.....	46
iii.	Détection des isoformes de VHL.....	47
iv.	Clonage des produits de PCR et transformation bactérienne .....	47
v.	Séquençage d'ADN .....	47
vi.	Quantification des isoformes de VHL.....	47
vii.	Culture cellulaire de lignées.....	49
viii.	Analyses in silico des séquences d'épissage .....	50
ix.	Expériences de minigènes.....	50
x.	Construction des vecteurs codant X1 .....	51
xi.	Tests fonctionnels .....	52
xii.	Western Blot .....	52
xiii.	Culture iPS.....	53
xiv.	Immunomarquage .....	53
xv.	ELISA EPO .....	54
<b>III.</b>	<b>RESULTATS .....</b>	<b>55</b>
1.	ETUDE DE MUTATIONS SYNONYMES DE L'EXON 2 DE <i>VHL</i> ET DESCRIPTION D'UN NOUVEL EXON CRYPTIQUE E1' DE <i>VHL</i> – BLOOD, LENGLET ET AL., 2018 <sup>219</sup> .....	55
a.	<i>Introduction</i> .....	55
b.	<i>Résultats</i> .....	56
i.	Etude de variants synonymes de l'exon 2 de VHL .....	56
ii.	Découverte et analyse de l'exon cryptique E1'.....	58
c.	<i>Article et données supplémentaires</i> .....	61
d.	<i>Résultats complémentaires</i> .....	104
i.	Etude fonctionnelle dans E1' .....	104
ii.	Etude fonctionnelle des isoformes induites par les variants dans l'intron 1 et l'exon 2 par utilisation d'ARN interférent .....	106
iii.	Etude fonctionnelle des isoformes induites par les variants dans l'intron 1 par utilisation de la méthode CRISPR/CAS9.....	112
e.	<i>Discussion</i> .....	116
2.	SEQUENÇAGE DE L'EXON CRYPTIQUE E1' DANS UN PANEL DE PATIENTS AVEC AU MOINS UNE TUMEUR DU SPECTRE VHL 118 .....	118
a.	<i>Introduction</i> .....	118
b.	<i>Article et données supplémentaires</i> .....	118
c.	<i>Résultats complémentaires</i> .....	134
d.	<i>Discussion</i> .....	136
3.	ETUDES FONCTIONNELLES DE VARIANTS IDENTIFIES DANS L'EXON E2' .....	138
a.	<i>Introduction</i> .....	138
b.	<i>Résultats</i> .....	138
c.	<i>Article en préparation et données supplémentaires</i> .....	139
4.	MISE EN PLACE D'UN MODELE DE CELLULES PRODUCTRICES D'EPO A PARTIR DE CELLULES SOUCHES PLURIPOTENTES INDUITES DE PATIENTS .....	155
a.	<i>Introduction</i> .....	155
b.	<i>Résultats</i> .....	156
i.	Caractérisation des lignées iPS .....	156
ii.	Protocoles de différenciation des lignées iPS .....	157
iii.	Caractérisation des progéniteurs de la crête neurale.....	161
iv.	Caractérisation des différenciations hépatiques .....	164
c.	<i>Discussion</i> .....	166
<b>IV.</b>	<b>CONCLUSION ET DISCUSSION GENERALES .....</b>	<b>169</b>
<b>V.</b>	<b>BIBLIOGRAPHIE .....</b>	<b>187</b>

# I. INTRODUCTION

## A. L'hypoxie

### 1. Généralités sur la réponse à l'hypoxie

L'homme fait partie des espèces animales dont la survie dépend totalement de la présence d'oxygène ( $O_2$ ) en quantité suffisante dans l'air. Notre système respiratoire permet à notre organisme de capter l'oxygène dans nos poumons. Notre système circulatoire, via l'hémoglobine des globules rouges, permet de redistribuer cet oxygène aux différents tissus et organes dans notre organisme. Dans l'air que nous respirons, à la pression atmosphérique du niveau de la mer, le taux d'oxygène est d'environ 18 à 21%. Il est distribué dans l'organisme en fonction des besoins des organes qui, pour leur bon fonctionnement, demandent plus ou moins d'oxygène<sup>1</sup>. Au niveau des poumons qui captent l'oxygène, le taux d' $O_2$  est de 18% et il descend à 6% dans les tissus.

L'hypoxie se définit par un manque d'apport en oxygène dans les tissus. Cela se traduit par une baisse du taux d'oxygène tissulaire qui peut devenir trop bas par rapport aux besoins de l'organisme. Pour un organe où le taux d' $O_2$  nécessaire est de 6%, quand celui-ci atteint 3% dans les tissus, on parle d'hypoxie puis d'anoxie quand celui-ci atteint 0,1%. L'hypoxie peut avoir plusieurs origines et être générale via un manque d'oxygène dans l'environnement (en altitude élevée par exemple) ou via un problème d'acheminement de l'oxygène dans l'organisme (problème d'affinité de l'oxygène sur les globules rouges ou affection pulmonaire). Elle peut également être locale dans le cas de tumeurs peu vascularisées ou de vaisseau sanguin bouché. Le taux d'oxygène sanguin est détecté en continu grâce à des structures spécialisées de la crosse aortique, cela permet à l'organisme de réguler en premier lieu le rythme cardiaque et la respiration pour adapter le niveau d'oxygène à l'effort réalisé via l'envoi d'un signal dopaminergique au cerveau<sup>2</sup>.

Plus nous montons en altitude, plus la pression atmosphérique est faible, moins il y a d'oxygène dans l'air. En effet, quand la pression diminue, pour un même volume d'air, il y a moins de

molécules de gaz. Par exemple, en haut de l'Everest qui culmine à 8848 mètres, la pression atmosphérique est divisée par 3 par rapport à celle au niveau de la mer. En haut du mont Blanc qui culmine à 4810 mètres, la pression est 2 fois moins élevée qu'au niveau de la mer. Une seule respiration nous apporte donc 3 fois moins d'oxygène en haut de l'Everest qu'au niveau de la mer. C'est pourquoi, lorsqu'on est en altitude élevée, il faut faire une ascension lente avec des périodes d'adaptation pour laisser le corps s'habituer à un taux d'oxygène plus faible pour éviter de souffrir du mal des montagnes. Il correspond à une mauvaise adaptation de l'organisme à l'altitude et plus particulièrement à un taux faible d'oxygène dans l'air. Celui-ci se caractérise par des maux de tête, des nausées, des vomissements, des problèmes de sommeil, des vertiges, des palpitations, des problèmes de vue et dans les cas les plus graves, les personnes peuvent souffrir d'un œdème pulmonaire (altitude supérieure à 4000 mètres) ou d'un œdème cérébral (altitude supérieure à 5000 mètres) qui peut conduire à un coma voire à la mort<sup>3,4</sup>. Les personnes sont plus ou moins sensibles au mal des montagnes. A 2000 mètres d'altitude, moins de 20% de la population en souffre, en revanche, plus de 50% en développent les symptômes à plus de 4000 mètres d'altitude. Il a été montré que les populations qui vivent en hautes altitudes depuis au moins 10 000 ans, comme les tibétains ou la communauté Andine, ce sont adaptés à un faible taux d'oxygène et ne présentent pas le mal des montagnes<sup>5,6</sup>. Différents moyens d'adaptation pour acheminer plus d'oxygène dans les tissus ont été observés comme une augmentation de la ventilation, une augmentation de la saturation en oxygène de l'hémoglobine, une augmentation de la concentration en hémoglobine ou une augmentation de la densité en capillaires sanguins dans les muscles. Il a été montré que ces adaptations sont héréditaires et qu'elles proviennent de modifications de certains gènes au cours de l'évolution pour une meilleure adaptation à l'environnement.

Une réponse moléculaire à l'hypoxie a également lieu au niveau cellulaire. Pour répondre à un manque d'oxygène, les cellules vont activer la voie dite de l'hypoxie pour essayer de pallier au manque. Les cellules les plus importantes de ce processus sont les cellules productrices d'érythropoïétine (EPO) qui sont situées dans le foie fœtal ou dans le rein adulte qui, en réponse à un manque d'oxygène vont produire une hormone : l'érythropoïétine<sup>7,8</sup>. Celle-ci va se fixer à

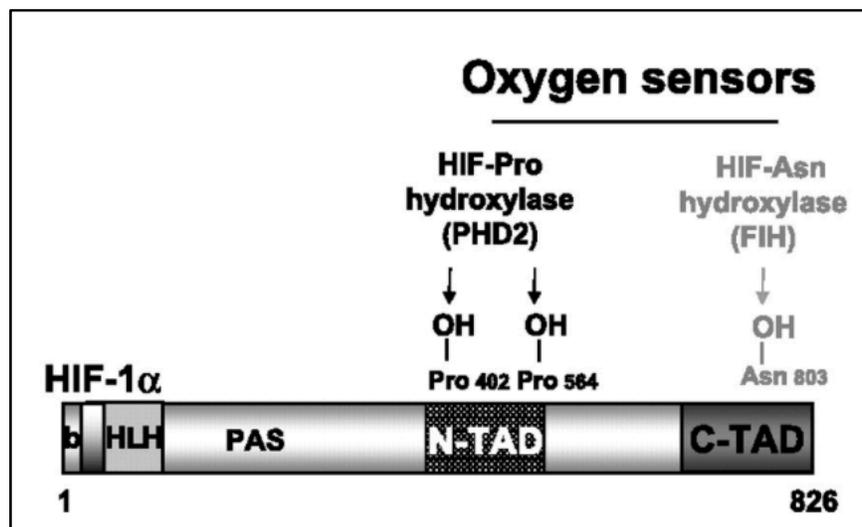
son récepteur : EPOR (Récepteur à l'Erythropoïétine) qui est présent à la surface des progéniteurs érythroïdes situés dans la moelle osseuse<sup>9</sup>. Ces cellules sont des progéniteurs qui peuvent, en réponse à un signal, proliférer et/ou se différencier en globules rouges. La fixation de l'EPO sur EPOR va entraîner une cascade de phosphorylation intracellulaire qui va notamment activer le facteur de transcription STAT5 (Transducteur de signal et activateur de transcription 5). Celui-ci va activer la transcription de gènes, ce qui va conduire à la multiplication et à la différenciation des progéniteurs érythroïdes en globules rouges. Il va ainsi y avoir plus de globules rouges dans le sang qui vont pouvoir transporter plus d'oxygène et l'acheminer via les vaisseaux aux tissus et aux organes. L'expression de l'EPO est régulée par un facteur de transcription qui est inductible en hypoxie : HIF (Hypoxia Inductible Factor) (Figure 3).

## 2. [HIF : facteur clé de la voie de l'hypoxie](#)

Le facteur clé de la réponse cellulaire à l'hypoxie est le facteur inductible en hypoxie : HIF<sup>10,11</sup>. Il s'agit d'un facteur de transcription hétérodimérique constitué de 2 sous-unités alpha et beta<sup>12</sup>. Il existe deux isoformes de la sous-unité HIF $\alpha$  : HIF-1 $\alpha$  et HIF-2 $\alpha$ . Ces deux sous-unités ont une structure bHLH-PAS : basic-helix-loop-helix-Per-ARNT-AHR-Sim du côté N-terminal de la protéine<sup>12</sup> (Figure 1). Le domaine HLH permet l'hétérodimérisation de la sous-unité HIF $\alpha$  avec la sous-unité HIF $\beta$ . Le domaine acide permet quant à lui la fixation et la spécificité de la liaison de HIF à l'ADN (Acide DésoxiriboNucléique). Le domaine bHLH contient également une séquence permettant l'import nucléaire de HIF $\alpha$ .

Les deux isoformes HIF1 $\alpha$  et HIF2 $\alpha$  s'hétérodimérisent avec une sous-unité  $\beta$  de HIF pour former un facteur de transcription actif qui peut activer différentes cibles de la voie de l'hypoxie<sup>13,14,15,16</sup>. HIF1 $\alpha$  stimule principalement l'expression des gènes glycolytiques alors que HIF2 $\alpha$  régule un panel de gènes plus important. Cependant, les deux sous-unités régulent des panels de gènes différents qui peuvent varier en fonctions des types cellulaires<sup>15,17</sup>.

La sous-unité beta de HIF est constitutivement exprimée dans le noyau des cellules, elle est aussi connue sous le nom d'ARNT (aryl hydrocarbon receptor nuclear translocator). La sous-unité alpha est quant à elle exprimée dans le cytoplasme dans lequel son expression est régulée via le protéasome en fonction de l'oxygène disponible. La sous-unité HIF $\alpha$  contient un domaine ODD (Oxygen-dependant Degradation) divisé en 2 principaux sites d'hydroxylation appelés CODD (C-Terminal ODD) et NODD (N-terminal ODD) qui est le plus sensible à l'hypoxie<sup>18,19</sup> (Figure 1). Ce domaine central est situé entre les acides aminés 401 et 603 de HIF $\alpha$ <sup>20</sup>.

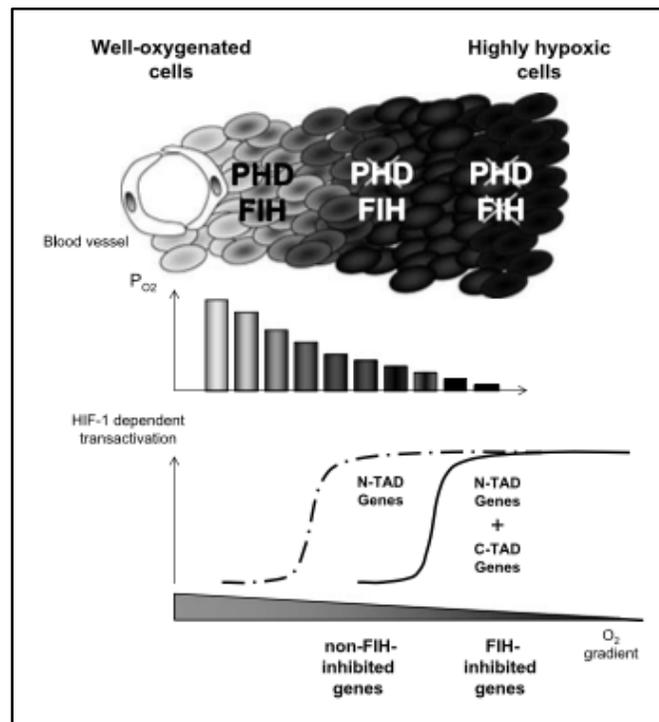


**Figure 1** : Structure de la protéine HIF1 $\alpha$ . (D'après Dayan et al.<sup>23</sup>) La protéine HIF1 $\alpha$  est constituée en sa partie N terminale d'un domaine bHLH suivis d'un domaine PAS. Elle contient également un domaine ODD séparé en 2 domaines : NODD (N-TAD) et C-ODD (C-TAD) hydroxylés respectivement par PHD2 ou FIH en fonction du taux d'oxygène tissulaire.

En conditions normoxiques, les protéines PHD (Prolyl Hydroxylase domain) (PHD1, PHD2 ou PHD3) qui sont des dioxygénases vont hydroxyler HIF1 $\alpha$  sur les prolines 402 et 564 (HIF2 $\alpha$  est hydroxylée sur ses résidus proline 405 et 531) de son domaine NODD en utilisant l'oxygène comme cofacteur<sup>21,22</sup> (Figure 2). Le domaine CODD est quand à lui hydroxylé sur le résidu proline 803 par une autre hydrolase FIH (factor inhibiting HIF) qui est active en conditions normoxiques ou en hypoxie modérée (conditions intermédiaires d'oxygène)<sup>23</sup>. Quand HIF $\alpha$  est hydroxylée, sur les prolines par les PHD, son affinité pour un complexe E3 ubiquitine ligase via la protéine von

Hippel-Lindau (VHL) est augmentée. Après sa liaison à pVHL, HIF $\alpha$  va être polyubiquitinylée et dégradée via le protéasome<sup>24,25</sup> (Figure 3).

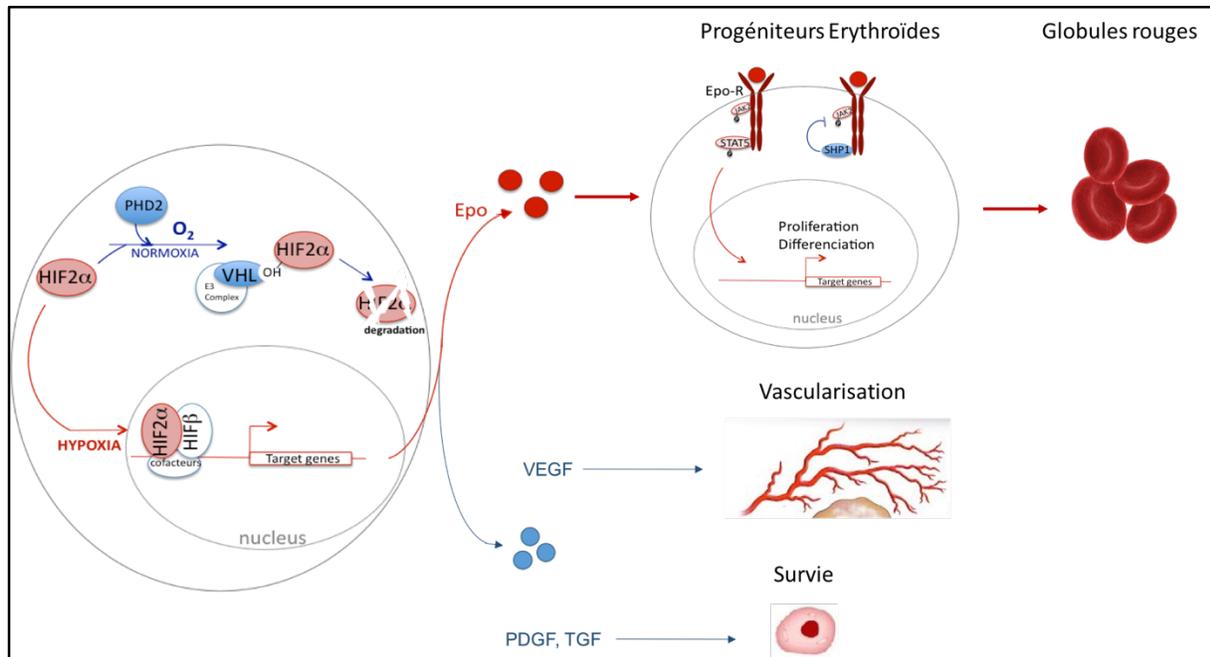
L'hydroxylation de HIF1 $\alpha$  sur son domaine CODD par FIH n'empêche pas la translocation de HIF1 $\alpha$  dans le noyau qui peut alors réguler l'expression de certains gènes cibles (appelés gènes non inhibés par FIH) (Figure 2).



**Figure 2 : Gradient d'hydroxylation de HIF par les PHD et FIH en fonction de l'oxygène disponible.** (D'après Dayan et al.<sup>23</sup>) En conditions normoxiques, HIF1 $\alpha$  est hydroxylée par les PHD sur son domaine N-TAD et par FIH sur son domaine C-TAD. En hypoxie modérée, HIF1 $\alpha$  est uniquement hydroxylée par FIH sur son domaine C-TAD et peut agir comme facteur de transcription pour l'expression de certains gènes cibles. En hypoxie plus sévère, HIF1 $\alpha$  n'est plus hydroxylée et va donc entraîner la transcription de tous ces gènes cibles.

Cette double action des PHD et de FIH dans différentes conditions d'oxygénation permet de constituer un gradient d'activation de HIF $\alpha$  qui est fortement dégradée en normoxie du fait de son hydroxylation via les deux hydrolases puis peu hydroxylée et non dégradée en début

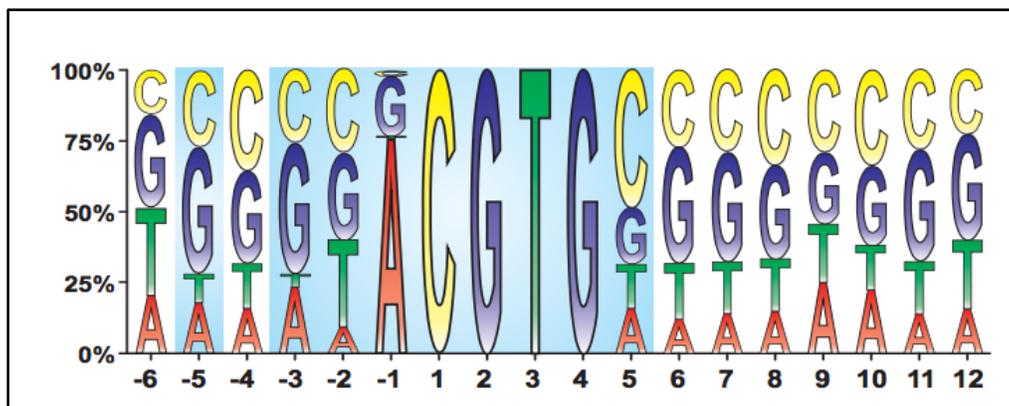
d'hypoxie par l'action seule de FIH et enfin non hydroxylée et totalement active en hypoxie sévère.



**Figure 3 : Représentation schématique de la voie de l'hypoxie.** En présence d'oxygène, la sous-unité HIF $\alpha$  est hydroxylée par une PHD. Elle est ensuite reconnue par VHL, sous-unité de reconnaissance d'un complexe E3 ligase. HIF $\alpha$  est alors ubiquitinylée et dégradée *via* le protéasome. En conditions hypoxiques, HIF $\alpha$  est stabilisée et va transloquer dans le noyau pour former un facteur de transcription actif avec HIF $\beta$ . Celui-ci va entraîner la transcription de nombreux gènes cibles en se fixant sur un élément de réponse à l'hypoxie (HRE). Une fois traduits, certains facteurs sont relargués dans la circulation et jouent un rôle dans l'érythropoïèse ou peuvent favoriser la vascularisation et/ou la survie cellulaire. L'EPO va se fixer à son récepteur EPOR à la membrane des progéniteurs érythroïdes. Cela va entraîner la phosphorylation de STAT5 *via* JAK2 et entraîner la prolifération et la différenciation des progéniteurs érythroïdes en globules rouges.

En condition hypoxique, la sous-unité alpha de HIF n'est pas hydroxylée par les PHD ou FIH. HIF $\alpha$  n'est donc pas capable de se lier à VHL. Elle n'est pas ubiquitinylée et n'est donc pas ciblée par le protéasome. Ainsi stabilisée, HIF $\alpha$  peut transloquer dans le noyau et s'hétérodimériser avec la sous-unité HIF bêta pour former le facteur de transcription inductible en hypoxie actif. Ce facteur va alors se fixer aux éléments de réponse à l'hypoxie, séquences consensus de type RCGTG appelées HRE (Hypoxia Response Element)<sup>26</sup> (Figure 4). Ces séquences sont le plus souvent

situées au niveau du promoteur des gènes cibles mais peuvent également être des promoteurs distaux comme cela est le cas pour l'érythropoïétine<sup>27</sup>. La fixation de HIF sur ces promoteurs va entraîner une augmentation de l'expression des gènes<sup>28,26</sup>. Il existe plus de 200 gènes cibles (plus de 2% des gènes humains) de HIF connus à ce jour. Leur expression permet une adaptation tissulaire à l'hypoxie. Les plus connus sont l'EPO qui va conduire à une production plus importante de globule rouge pour assurer le transport d'une quantité plus importante d'oxygène, le VEGF (vascular epithelial growth factor) qui va permettre la néoformation de vaisseaux sanguins pour l'acheminement d'oxygène dans des zones hypoxiques, le TGF $\beta$  (transforming growth factor) qui est une cytokine impliquée notamment dans l'hématopoïèse et le PDGF (platelet derived growth factor) qui permet de réguler la croissance et la division cellulaire pour augmenter le nombre de cellules hématopoïétiques en conditions hypoxiques. HIF est donc impliqué dans de multiples voies cellulaires, notamment l'érythropoïèse, la prolifération, l'angiogenèse, le métabolisme, l'autophagie ou la survie cellulaire (Figure 3).



**Figure 4 : Séquence consensus HRE (Hypoxia Respons Element)<sup>26</sup>. Fréquence de survenue de la séquence nucléotidique dans les séquences consensus HRE connues.**

Il existe une troisième sous-unité alpha de HIF $\beta$  : HIF3 $\beta$ . Le gène codant cette protéine est constitué de 16 exons qui peuvent être épissés en différentes isoformes (au moins 7 connues à ce jour). L'isoforme contenant tous les exons du gène *HIF3 $\alpha$*  correspond à une IPAS (Inhibitory PAS (Per/Arnt/Sim) domain protein)<sup>29,30</sup> qui agit comme un régulateur transcriptionnel. Cette protéine n'est exprimée qu'en hypoxie (à partir de 8% d'O<sub>2</sub>) et a été retrouvée surexprimée dans

les poumons et le cœur de souris placées en hypoxie. Elle a un rôle de rétrocontrôle négatif sur le facteur de transcription HIF. En effet l'isoforme IPAS de HIF3 $\alpha$  se lie à l'hétérodimère HIF $\alpha$ / $\beta$ , l'empêchant de se fixer sur les promoteurs des gènes cibles de HIF. Il a été montré que d'autres isoformes longues de HIF3 $\alpha$  sont localisées dans le noyau des cellules en hypoxie<sup>31</sup>. A elles seules, elles ne sont pas capables d'activer des gènes cibles de HIF. Cependant selon les isoformes et les gènes cibles, elles peuvent soit inhiber la fonction du facteur de transcription hétérodimérique HIF, soit agir comme un cofacteur pour l'expression des gènes. Les fonctions exactes des isoformes de HIF3 $\alpha$  sont encore assez méconnues et elles semblent avoir des rôles complexes dans la réponse à l'hypoxie.

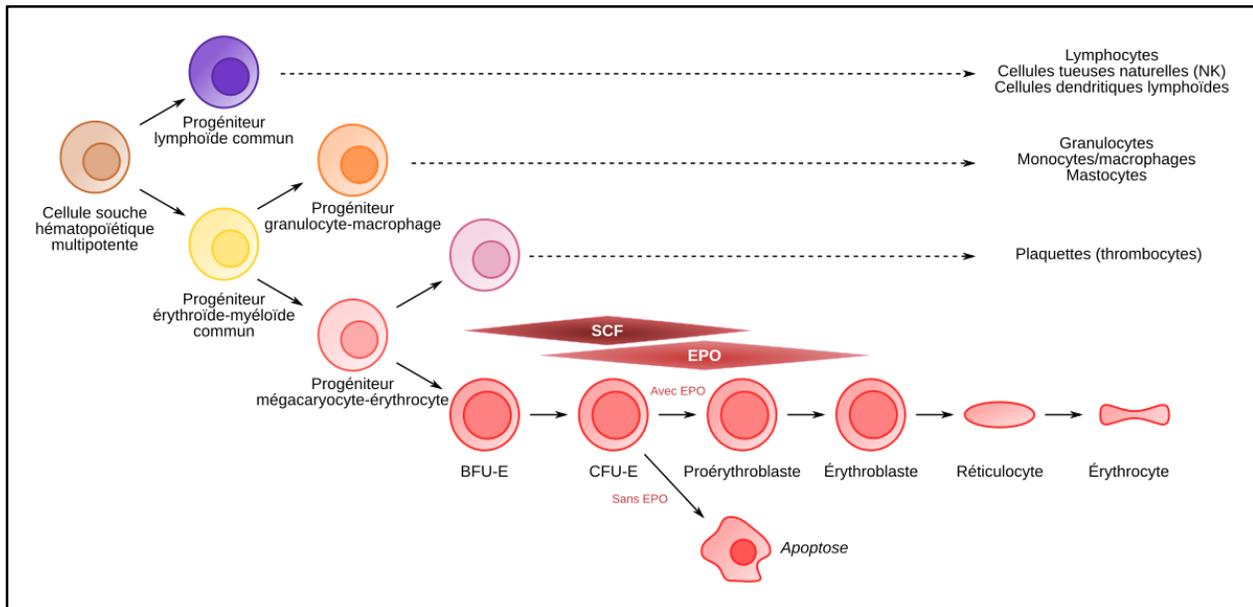
### 3. Rôle et origine de l'érythropoïétine (EPO)

L'érythropoïétine est une hormone glycoprotéique exprimée dans des types cellulaires spécifiques. Le gène de l'EPO est composé de différents sites régulateurs qui permettent cette tissu spécificité<sup>32</sup>. Elle est synthétisée chez le fœtus par la crête neurale puis par le foie<sup>33,8</sup>. Chez l'adulte, 90% de l'EPO circulante est produite par les cellules productrices d'EPO rénales (REPC)<sup>7</sup>. Les dix autres pourcents de l'EPO circulante sont produits par le foie (qui garde chez l'adulte sa capacité à produire de l'EPO), le cerveau (neurones, cellules gliales et cellules endothéliales), l'utérus, les testicules et les glandes adrénales qui ont également été rapportés comme capable de produire de l'EPO<sup>34,35,36</sup>. L'origine des cellules rénales productrices d'EPO est encore controversée mais de nombreuses hypothèses montrent qu'elles auraient pour origine les cellules de la crête neurale<sup>37,38,33,39</sup>. Certains travaux indiquent que les cellules productrices d'EPO sont des « neurones-like »<sup>40</sup>. A ce jour, il n'existe aucune lignée issue de cellules rénales productrices d'EPO, personne n'ayant encore réussi à les isoler et immortaliser en culture. Les seules lignées capables de produire de l'EPO sont une lignée de progéniteurs neuronaux : SH5-SY5Y, deux lignées d'hépatocarcinome : Hep-3b et HepG2, une lignée rénale cancéreuse SKRC44<sup>41</sup> et une lignée de cellules murines productrices d'EPO rénales récemment générée (FAIK)<sup>38</sup>.

Une surproduction d'EPO peut avoir lieu dans certaines tumeurs. En effet, pour certaines tumeurs peu vascularisées ou portant certains types de mutations, il y a une forte production d'EPO de la tumeur elle-même du fait d'une dérégulation de la voie de l'hypoxie. Cela a été montré pour des tumeurs du rein, du foie, du système nerveux central, de l'utérus, des glandes adrénales, des ovaires, des poumons, du thymus et pour les hémangioblastomes<sup>42</sup>. Cela conduit à un taux d'EPO circulant trop élevé qui peut causer une érythrocytose (taux de globule rouge sanguin trop élevé) associée à la tumeur chez les patients. Lors de la résection de la tumeur par chirurgie, le taux d'EPO sanguin redevient normal et l'érythrocytose disparaît.

Le gène codant l'EPO a été cloné en 1985, ce qui a permis les premières productions d'EPO de synthèse utilisée pour le traitement des anémies (diminution de la quantité ou de la qualité des globules rouges)<sup>43,44</sup>. La production d'EPO est sous le contrôle d'un promoteur qui contient une séquence HRE, elle est donc régulée par le facteur de transcription HIF. En cas d'hypoxie sévère, l'EPO peut être 1000 fois plus élevée que la normale dans le sang<sup>45</sup>. La quantité d'EPO circulant dans le sang est inversement proportionnelle à la quantité d'oxygène présente, cette proportionnalité est contrôlée par le facteur de transcription HIF<sup>46</sup>. L'EPO se fixe sur son récepteur EPOR (récepteur à l'érythropoïétine) qui est exprimé à la surface des progéniteurs érythroïdes à partir de leur stade « burst forming unit-erythroid » (BFU-E) qui sont localisés dans la moelle osseuse (Figure 5). Les BFU-E sont issues des cellules souches hématopoïétiques capables de donner tous les types cellulaires retrouvés dans le sang qui peuvent se spécialiser en réponse à différents signaux jusqu'à donner des progéniteurs mégacaryocytes-érythrocytes qui peuvent soit se différencier en plaquettes, soit en BFU-E. Les BFU-E ont besoin pour survivre de la cytokine SCF (stem cell factor) et pour poursuivre leur différenciation en « colony forming unit-erythroid » (CFU-E) d'EPO. La fixation de l'EPO sur les CFU-E va conduire à une cascade de réactions de phosphorylation notamment des protéines JAK2 (Janus kinase 2) puis STAT5 (transducteur de signal et activateur de transcription 5) (Figure 3). Une fois phosphorylé, STAT5 va s'hétérodimériser pour former un facteur de transcription actif qui va permettre la prolifération des progéniteurs érythroïdes (aux stades BFU-E puis CFU-E) et induire la différenciation des CFU-E en érythrocytes via quatre stades (Figure 5) : les proérythroblastes puis les érythroblastes (3

stades sont décrits : érythroblastes basophiles puis érythroblastses polychromatiques puis érythroblastes orthochromatiques) qui donnent naissance aux réticulocytes qui se transforment en érythrocytes (globules rouges)<sup>47, 48</sup>. Au stade CFU-E, si la quantité d'EPO disponible est trop faible, les progéniteurs vont entrer en apoptose sans se différencier en érythrocytes.



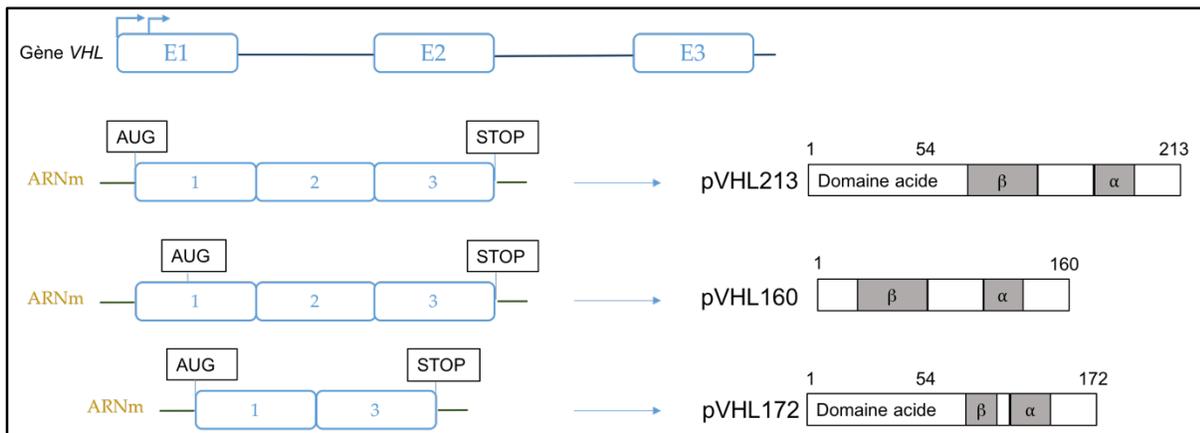
**Figure 5 : Différenciation des cellules souches hématopoïétiques multipotentes.** (D'après PlanetVie ENS éducol<sup>49</sup>) Les cellules souches hématopoïétiques sont capables de donner tous les types cellulaires retrouvés dans le sang. La différenciation en érythrocytes (globules rouges) nécessite le passage par le stade de progéniteur mégacaryocytes-érythrocytes qui va donner les BFU-E (burst forming unit-erythroid) puis les CFU-E (colony forming unit-erythroid). A ce stade, les progéniteurs ont besoin de la cytokine stem cell factor (SCF) pour survivre et de l'EPO pour poursuivre leur différenciation en érythrocytes. En absence de ces cytokines, les CFU-E entrent en apoptose.

Au cours de la vie, l'érythropoïèse est modérée et continue pour remplacer quotidiennement environ 200 milliards de globules rouges qui ont une durée de vie de 120 jours. L'expression de l'EPO est donc modérée et constante sauf en conditions hypoxiques<sup>50</sup>. L'érythropoïèse peut également être dérégulée dans le cas de mutation de gènes de la voie de l'hypoxie, les personnes porteuses de ces mutations développent des polyglobulies (ou érythrocytoses secondaires) associée à une activation de la voie de l'hypoxie et à une surproduction d'EPO.

## B. Description du gène *VHL* et fonction de pVHL

### 1. Le gène *VHL*

Le gène *VHL* est un gène suppresseur de tumeur qui a été isolé en 1993<sup>51,52,53</sup>. Il est situé sur le bras court du chromosome 3 au niveau du locus 3p25. Il est constitué de 3 exons composés de 852 nucléotides codant une protéine de 213 acides aminés (pVHL213) de 30kDa (Figure 6 et 7). Le transcrit peut également être traduit à partir d'un second site d'initiation de la traduction alternatif situé au codon 54 (nucléotide 162) après le premier ATG du cadre de lecture (ORF : Open Reading Frame) de *VHL*, ce qui donne naissance à une protéine de 160 acides aminés (pVHL160) de 19kDa. Cette deuxième protéine possède une activité de suppresseur de tumeur comparable à la protéine pVHL213<sup>54,55,56</sup>. Le gène *VHL* est impliqué dans plusieurs voies cellulaires mais la plus étudiée reste la régulation de la voie de l'oxygène<sup>57</sup>.



**Figure 6 : Gène *VHL*, ARN messager et protéines VHL.** *VHL* code 3 protéines de 213, 160 et 172 acides aminés contenant 2 domaines  $\beta$  et  $\alpha$ , le domaine  $\alpha$  étant tronqué dans la protéine pVHL172. Ces domaines permettent la liaison de pVHL à d'autres protéines.

L'exon 2 de *VHL*, codant les acides aminés 114 à 154, peut être épissé de façon alternative. Il en résulte une isoforme constituée des exons 1 et 3 de *VHL* (E1E3 ou  $\square$ E2) (Figure 6). Cette isoforme est présente en faible proportion dans les tissus<sup>58</sup>. Elle code une protéine de 172 acides aminés (pVHL172) exprimée de façon endogène<sup>59</sup>. Il a été montré que cette isoforme n'agit pas comme un suppresseur de tumeur et n'est pas capable de réguler la sous-unité alpha du facteur de

transcription HIF<sup>60</sup>. L'expression de pVHL172 conduit à une augmentation de l'expression de gènes pro-tumorigènes tels que *TGFB1* (Transforming Growth Factor Beta 1) qui code un facteur de croissance contrôlant la prolifération et la différenciation cellulaire, *MMP1* (Matrix Metallopeptidase 1) et *MMP3* (Matrix Metallopeptidase 3) codant des protéases qui ont un rôle dans le remodelage de la matrice extra-cellulaire.

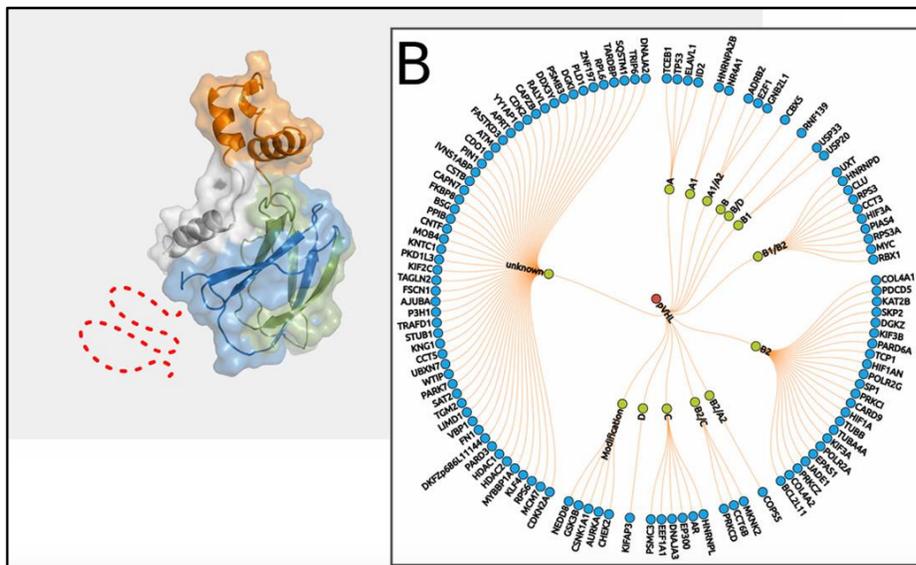
## 2. Protéines VHL et fonctions dépendantes de HIF

Les protéines pVHL213 et pVHL160 codées par les 3 exons du gène (E1E2E3) sont constituées de 2 domaines fonctionnels nommés  $\alpha$  et  $\beta$ <sup>61</sup> (Figure 6). Elles sont principalement localisées dans le cytosol<sup>62,63</sup> mais il a été montré que lorsque les cellules sont non adhérentes, pVHL est localisée dans le noyau<sup>64</sup>. Le domaine  $\alpha$  de pVHL est situé entre les acides aminés 157 et 189 dans l'exon 3 du gène *VHL*. Il est constitué de 3 hélices  $\alpha$  capables de lier l'élongine C et avec une importance moindre l'élongine B dans un complexe nommé complexe VBC<sup>62</sup>. Il a été montré que l'interaction de ces 3 protéines les stabilise mutuellement<sup>65</sup>. L'élongine C va former un complexe tertiaire avec pVHL, l'élongine B, la culline 2 (CUL2), la RING-box protéine 1 (Rbx1) et l'enzyme de conjugaison de l'ubiquitine E2<sup>61</sup>. Le complexe ainsi formé possède une activité E3 ubiquitine ligase. Le domaine  $\beta$  est situé entre les acides aminés 64 et 154, à cheval entre les exons 1 et 2 du gène. Il est formé de deux feuillets  $\beta$  arrangés en sandwich avec une hélice  $\alpha$  au sommet. Ce domaine constitue la sous-unité de reconnaissance spécifique du substrat du complexe E3 ligase, son substrat principal étant HIF $\alpha$  préalablement hydroxylé par les PHD<sup>25,66</sup>. HIF $\alpha$  est reconnu par pVHL *via* son domaine de transactivation (N-TAD) de 15 acides aminés qui est conservé dans toutes les isoformes de HIF $\alpha$ <sup>67</sup>. Une fois lié au complexe E3 ligase, HIF $\alpha$ -OH est polyubiquitinylé et conduit au protéasome où elle va être dégradée. Certains gènes régulés par HIF doivent être exprimés continuellement. Cela est possible grâce à la protéine VHL-like (VHL-like protein : VLP), codée par le gène *VHLL* situé sur le chromosome 1, qui a une homologie de 78% avec la protéine VHL<sup>68</sup>. Elle contient le domaine  $\alpha$  de VHL et peut donc fixer HIF $\alpha$ . Cependant elle ne contient pas le domaine  $\beta$  de VHL et ne peut donc pas former de complexe E3 ubiquitine ligase. VLP agit donc

comme un dominant négatif de VHL pour HIF $\alpha$  et lui sert donc de protecteur en conditions normoxiques.

La protéine VHL172, codée uniquement par les exons 1 et 3 du gène, contient le domaine  $\beta$  de la protéine, elle est donc capable de complexer avec l'élongine C et ses partenaires. Cependant, pVHL172 ne contient pas le domaine  $\alpha$  de VHL fonctionnel puisque celui-ci est tronqué. Elle n'est donc pas capable de fixer HIF $\alpha$  ni d'entraîner son ubiquitinylation. C'est pourquoi elle est pour le moment décrite comme non fonctionnelle dans la voie de dégradation de HIF $\alpha$ <sup>69</sup>. Elle est exprimée dans différents tissus et lignées cellulaires rénales embryonnaires ou cancéreuses. Au sein des tumeurs, son expression est variable, tout comme celle des protéines pVHL172 et pVHL160. Elle pourrait participer à l'initiation tumorale ou à la progression tumorale.

### 3. Protéines VHL et fonctions indépendantes de HIF



**Figure 7 : Interactome de pVHL.** (Daprès Tabaro et al.<sup>71</sup>) Représentation 3D de la protéine pVHL 213 (gauche). pVHL interagit avec 117 partenaires *via* ses différents domaines. Pour une partie des protéines interagissant avec pVHL, le domaine de liaison n'est pas encore connu. (B)

Même s'il est admis que la fonction dépendante de HIF reste pour le moment la fonction majeure la plus étudiée et la plus connue de *VHL*, il est clair que pVHL a des fonctions indépendantes de HIF<sup>70</sup>. En effet, pVHL est capable d'interagir avec au moins 117 autres protéines *via* ses différents domaines<sup>71</sup> (Figure 7). Dans certains cas, ces fonctions peuvent expliquer des phénotypes observés dans le cadre de maladies dont l'origine est une mutation du gène *VHL*.

#### a. Régulation de la transcription

Une des autres fonctions décrites de pVHL213 est liée à sa forte affinité pour l'élongine C. En effet, l'élongine B et C font également partie d'un complexe de régulation de l'élongation de la transcription SIII (ou élongine ABC) avec l'élongine A<sup>72</sup>. Cette dernière est une sous-unité transcriptionnellement active. Le complexe élongine ABC formé est capable d'activer l'élongation de la transcription par l'ARN polymérase II. Par sa forte affinité avec l'élongine C, pVHL la séquestre dans le cytoplasme, ce qui inhibe son activité de cofacteur de régulation de la transcription. pVHL peut également se lier à deux sous-unités du complexe ARN polymérase II *via* son domaine  $\square$ <sup>73</sup>. La première est la sous-unité 7 du complexe (hsRPB7), la deuxième est la grande sous-unité du complexe (Rpb1) lorsque celle-ci est hyperphosphorylée notamment à la suite de radiations UV ou de dommages à l'ADN *via* des agents chimiques<sup>74</sup>. Elles deviennent alors la cible d'une ubiquitinylation. Elles sont dégradées *via* le protéasome et leur quantité dans le noyau des cellules est diminuée. Cela conduit donc à une diminution de la transcription par le complexe ARN polymérase II.

Une baisse générale de la transcription dans le cas des cellules porteuses de mutations de *VHL* peut expliquer le développement souvent lent des tumeurs du spectre VHL.

#### b. Maintenance du cil primaire et stabilisation des microtubules

Dans les années 2000, il a été montré que la protéine VHL est indispensable à la polarisation cellulaire, et au développement du cil primaire, indépendamment de la quantité de HIF présente dans les cellules<sup>75</sup>. Il a été montré que cette capacité est spécifique de pVHL213 et que pVHL160 n'intervient pas dans le développement du cil. pVHL est capable de se lier au cil primaire *via* la

kinesine 2<sup>76</sup> qui est une protéine qui se déplace sur les microtubules. Cela permet de stabiliser les microtubules qui composent le cil primaire<sup>77</sup>. La fixation aux microtubules de pVHL se fait *via* les acides aminés 95 à 123 qui sont souvent retrouvés mutés dans le cadre de la maladie de VHL. Cette fixation inhibe la dépolymérisation des microtubules. Il a été montré que les mutants de *VHL* qui ne conduisent pas au développement de cancer du rein n'affectent pas le développement du cil primaire contrairement à ceux porté par les patients qui développent un cancer du rein<sup>75</sup>. En effet, il a été montré que la perte du cil primaire conduit au développement de tumeurs et de kystes du foie, du cerveau, du rein et du pancréas. En réponse au flux urinaire, les cils primaires ont un rôle de chimiotenseur et de mécanosenseur en transmettant un signal intracellulaire *via* un influx de calcium<sup>78,79</sup>. Ce flux de calcium a un rôle déterminant dans le développement et dans le maintien de l'intégrité des néphrons. Nous pouvons donc penser que la perte du cil primaire des cellules qui ont une mutation de *VHL* sur leur 2 allèles conduit à une perte d'intégrité des néphrons et donc d'une réduction de l'activité rénale qui participerait au développement tumoral.

### c. Régulation de la formation de la matrice extracellulaire

Une étude a montré *in vivo* chez la souris qu'un défaut de la protéine VHL dans les cellules endothéliales conduit à un défaut de vascularisation des embryons. Il peut être partiellement inversé *via* l'apport de fibronectine exogène qui est une glycoprotéine extracellulaire qui entre en jeu dans la formation de la matrice extra-cellulaire<sup>80</sup>. Plus récemment, il a été mis en évidence que pVHL est capable d'interagir avec la partie N-terminale de la protéine COL4A2 qui est la collagène IV alpha 2<sup>81</sup>. Cette interaction se fait au niveau du réticulum endoplasmique. Elle est nécessaire pour la formation du réseau extracellulaire de collagène 4 qui est spécifique de la lame basale des cellules épithéliales et des cellules musculaires.

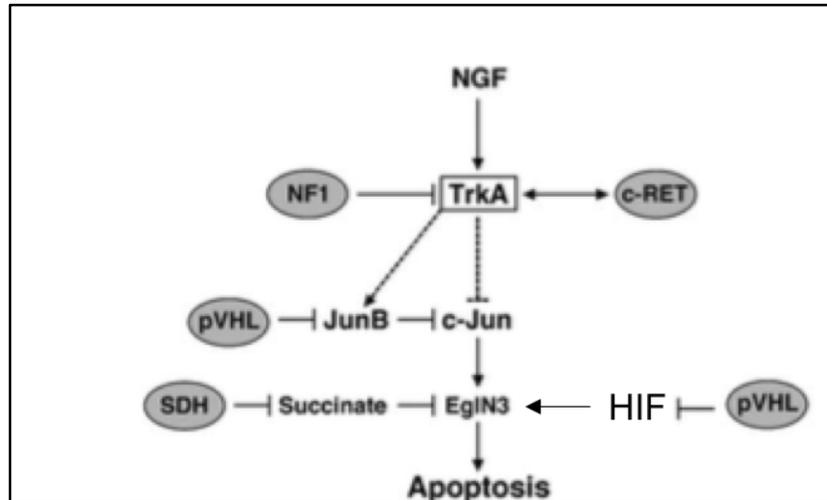
Quand *VHL* est muté, il y a donc des défauts de la structuration de la matrice extracellulaire, ce qui peut expliquer le caractère kystique de nombreuses tumeurs du spectre VHL.

#### d. Régulation de l'apoptose

pVHL est capable de moduler des régulateurs de l'apoptose tels que le facteur nucléaire  $\kappa$ B (NF $\kappa$ B) et P53. En effet, pVHL sert d'adaptateur pour promouvoir la phosphorylation de CARD9 (caspase recruitment domain-containing protein 9) sur son extrémité C terminale par la caséine kinase 2<sup>82</sup>. CARD9 est un agoniste de NF $\kappa$ B qui entraîne une activation de NF $\kappa$ B, un facteur de transcription qui est impliqué dans la régulation de la réponse immunitaire et de la réponse au stress cellulaire. Dans le cadre de la réponse au stress cellulaire, NF $\kappa$ B entraîne la transcription de gènes anti-apoptotiques. Quand CARD9 est phosphorylé, il ne peut plus activer NF $\kappa$ B, ce qui conduit à une diminution de l'expression de facteurs anti-apoptotiques. Quand *VHL* est muté, il n'est plus capable de conduire à la phosphorylation de CARD9, il y a donc forte expression de facteurs anti-apoptotiques, ce qui promeut la progression tumorale. Dans le cas de p53, il a été montré que pVHL est capable de s'y lier, ce qui conduit à sa stabilisation en bloquant son export nucléaire et son ubiquitination par MDM2 (une E3 ubiquitine ligase). Cela conduit à une augmentation de l'activité transcriptionnelle de p53<sup>83</sup>. pVHL se lie à p53 *via* son domaine  $\Omega$  uniquement après avoir été phosphorylée sur la sérine 111 par la checkpoint kinase 2 (Chk2) après un dommage à l'ADN<sup>84</sup>. En conditions « normales » (absence de stress cellulaire), il y a très peu de p53 dans les cellules, celle-ci étant dégradée par le protéasome après ubiquitination par MDM2. En cas de stress cellulaire, MDM2 ne se lie plus à p53, cette dernière va alors conduire à la transcription de facteurs qui induisent l'apoptose des cellules afin d'éliminer les cellules ayant eu une lésion de l'ADN. Quand *VHL* est muté, il ne peut plus stabiliser p53, ce qui conduit à une diminution de l'apoptose des cellules lésées qui pourront alors former des tumeurs.

Il a également été montré dans les phéochromocytomes que des mutations de *VHL* empêchent pVHL d'inhiber JunB, est un facteur de transcription qui régule l'activité des gènes impliqués dans la réponse primaire aux facteurs de croissance (Figure 8)<sup>85</sup>. Cette augmentation du niveau d'expression de JunB conduit à une diminution de l'induction de l'apoptose dans les phéochromocytomes. La régulation de l'apoptose par JunB est dépendante des protéines c-Jun (un facteur de transcription) et EglN3 (ou PHD3 qui est une hydroxylase permettant notamment

l'hydroxylation de HIF $\alpha$ ). Dans les phéochromocytomes, les mutations de *VHL* entraînent donc une diminution de l'induction de l'apoptose qui favorise le développement de la tumeur.



**Figure 8 : Régulation de JunB par pVHL.** (D'après Lee et al.<sup>85</sup>) Régulation de la voie de signalisation de l'apoptose via JunB par pVHL.

#### e. Régulation de la survie et de la prolifération des progéniteurs érythroïdes

La liaison de l'EPO sur son récepteur EPOR entraîne la phosphorylation de la protéine STAT5 *via* la protéine JAK2 (Janus kinase 2)<sup>86</sup>. Une fois phosphorylée, STAT5 va activer la transcription de gènes cibles des voies de prolifération, de différenciation et de survie cellulaire dans les progéniteurs érythroïdes. En réponse à l'hypoxie, pVHL permet une stabilisation de HIF et donc une production plus importante d'EPO qui conduit à une augmentation de la prolifération et de la survie des progéniteurs érythroïdes. Il a été montré que pVHL est également capable d'entraîner la dégradation de JAK2 *via* la formation d'un complexe E3 ligase actif avec la protéine SOCS1 (Suppressor of cytokine signalling 1)<sup>47</sup>. Le complexe ainsi formé reconnaît la protéine JAK2 phosphorylée sur les tyrosines 1007 et 1008 puis entraîne sa polyubiquitinylation qui est le signal de sa dégradation *via* le protéasome<sup>87</sup>. *VHL* serait donc un régulateur de la survie et de la prolifération des progéniteurs érythroïdes *via* la régulation de la production d'EPO mais également *via* la régulation de la dégradation de JAK2.

## C. Maladies dues à des anomalies de VHL

### 1. La maladie de VHL

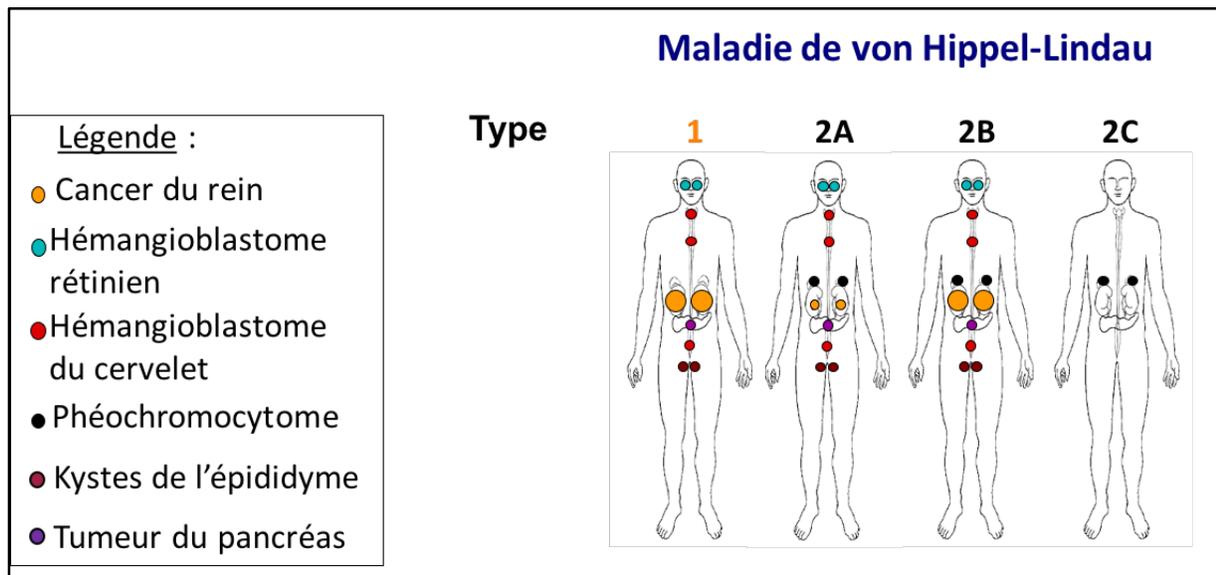
#### a. Généralités

Des mutations germinales hétérozygotes du gène *VHL* prédisposent les patients au développement de la maladie de von Hippel-Lindau (VHL) qui se caractérise par le développement de tumeurs hypervascularisées bénignes ou malignes telles que des hémangioblastomes du système nerveux central ou de la rétine, des phéochromocytomes, des tumeurs endocrines du pancréas ou des cancers du rein<sup>53,88</sup>. Les patients développent de nombreuses tumeurs, appelées tumeurs du spectre VHL, tout au long de leur vie et doivent donc avoir un suivi très régulier avec des imageries (échographies, IRM) du tronc et de la tête. Cela permet de détecter les tumeurs dès leur apparition et les surveiller avant qu'elles ne se développent. Lors de l'apparition de la 1<sup>ère</sup> tumeur, les patients ayant droit à un suivi sont ceux pour lesquels une mutation de *VHL* à l'origine de la pathologie a été identifiée. En France, dans les familles à risque, le suivi commence à l'âge de 5 ans quand aucune tumeur n'a été détectée avant alors qu'aux Etats Unis le suivi commence dès l'âge de 1 an. Un diagnostic prénatal ou préimplantatoire est également possible chez les familles à risque porteuses d'une mutation du gène *VHL*.

La maladie a été nommée VHL en 1936 à la suite de la description d'hémangioblastomes de la rétine, en 1904<sup>89</sup> par Eugen von Hippel, un ophtalmologiste allemand, qui pouvaient être associés à des hémangioblastomes du cervelet décrits en 1927 par Arvid Lindau, un pathologiste suédois<sup>90</sup>. Les critères de diagnostic ont été définis en 1964<sup>91</sup> avec au moins deux tumeurs du spectre VHL quand il n'y a pas d'historique familial de la maladie de VHL ou avec une tumeur du spectre VHL quand il y a un historique de la maladie. Pour environ 20% des diagnostics de la maladie, la mutation du gène *VHL* est *de novo*, c'est-à-dire qu'elle n'est pas retrouvée chez les parents.

Des cas de mosaïcismes qui restent plutôt rares sont répertoriés dans le cadre de la maladie<sup>92,93</sup>. Dans ces cas, certaines populations de cellules sont saines et d'autres portent une mutation de *VHL*. Il est donc recommandé d'utiliser des techniques sensibles telles que la FISH ou le

séquençage de l'ADN génomique après clonage pour pouvoir détecter de faibles proportions de lymphocytes du sang périphérique porteurs d'une mutation dans le cas de mosaïcisme. Cela permet aux patients mosaïques de pouvoir bénéficier du suivi spécifique de la maladie de VHL. Les patients mosaïques ont une expression variable de la protéine VHL mutée qui conduit à différents phénotypes qui vont de l'absence totale du développement de tumeurs du spectre VHL à un développement intense de tumeurs en fonction des types cellulaires porteurs d'une mutation de *VHL*.



**Figure 9 : La maladie de VHL.** En fonction des phénotypes observés, on classe la maladie en 2 types : type 1 pour les patients ne développant aucun phéochromocytome et type 2 pour les patients qui en développent associés ou non à un risque fort (2A) ou nul (2C) de développer un cancer du rein.

La maladie présente des hétérogénéités phénotypiques qui ont été montrées dans les années 1990<sup>94</sup>. Il s'en est suivi une classification de la maladie de VHL en 2 types : type 1 pour les patients ne développant pas de phéochromocytomes et type 2 pour les patients qui en développent<sup>95</sup> (Figure 9). Le type 1 est le type le plus sévère en terme de risque de développer un cancer du rein. Le type 2 est subdivisé en 3 classes : type 2A pour les patients qui développent des phéochromocytomes et des hémangioblastomes mais pas de cancers du rein, type 2B pour les patients qui développent des phéochromocytomes, des hémangioblastomes et des cancers du

rein et type 2C pour les patients qui développent uniquement des phéochromocytomes<sup>96</sup>. Il est difficile de classer les familles dans le bon sous-type de maladie étant donné que les symptômes se développent différemment et à des âges différents chez les patients. Il peut arriver que les phéochromocytomes ne se développent que chez les patients de la 2<sup>ème</sup> ou 3<sup>ème</sup> génération d'une famille, ce qui la fait alors passer du type 1 au type 2.

L'incidence de la maladie est de 1 naissance sur 36 000 et la pénétrance est complète à 65 ans (tous les porteurs d'une mutation ont développé la maladie à 65 ans). Elle est donc classée dans la catégorie des maladies rares. La morbidité et la mortalité de la maladie sont souvent dues à des complications neurologiques à la suite d'hémangioblastomes avec une mauvaise localisation ou dues à la progression tumorale et métastatique des cancers du rein. L'espérance de vie des patients qui développent la maladie de VHL est de 49 ans mais la surveillance peut permettre de rallonger leur durée de vie.

Les patients qui développent la maladie de VHL sont porteurs d'une mutation héréditaire du gène *VHL* à l'état hétérozygote (uniquement sur un des allèles du gène). La prédisposition au cancer est transmise sous le mode dominant. La survenue d'une tumeur dépend de l'inactivation de la 2<sup>ème</sup> copie du gène soit par une mutation ou deletion acquise soit par la perte d'expression totale de l'allèle sauvage suite à une méthylation du promoteur. La perte de l'allèle sauvage dans la tumeur est appelée perte d'hétérozygotie (LOH). Ce modèle dit « Two-hits » a été décrit en 1971 par Knudson lors de l'étude de rétinoblastomes. Il est défini par la prédisposition au cancer lors de la perte de fonction d'un allèle hérité et le développement d'un cancer lors la perte du 2<sup>ème</sup> acquise<sup>97</sup>. Ce modèle permet d'expliquer l'apparition plus précoce et multiple des tumeurs ainsi que les risques plus importants d'en développer chez les apparentés également porteur d'une mutation. Knudson a ensuite en 1972 validé ce modèle dans deux autres types de tumeurs du système nerveux, les neuroblastomes et les phéochromocytomes<sup>98</sup>. Dans le cas de la maladie de VHL, il est retrouvé dans les tumeurs une perte de l'allèle sauvage<sup>98,99,100</sup>. La perte du 2<sup>ème</sup> allèle sauvage conduit à une perte totale de la production de la protéine VHL sauvage. La protéine VHL exprimée est alors issue d'un allèle muté (le seul à s'exprimer), le plus souvent inactive pour la

dégradation du facteur de transcription HIF. En conditions normoxiques, il y a accumulation du facteur HIF qui ne peut plus se lier à pVHL. Il y a alors une surexpression des gènes régulés par HIF. Cela conduit à la formation de tumeurs hypervascularisées puisque ces gènes régulent la prolifération cellulaire et la néo-vascularisation qui deviennent excessives.

La maladie de VHL (maladie héréditaire rare) a permis la découverte du gène *VHL* qui est impliqué dans plus de 90% des cancers du rein à cellules claires sporadiques (qui touchent l'ensemble de la population). Les cancers du rein sporadiques apparaissent de façon plus tardive (en moyenne à l'âge de 62 ans) et sont unilatéraux. En effet, une cellule doit dans ce cas acquérir la perte des 2 allèles du gène *VHL* avant de devenir cancéreuse, *via* une mutation, délétion ou *via* son extinction par méthylation<sup>101</sup>. L'étude de la maladie de VHL a permis de mieux comprendre les mécanismes impliqués dans les cancers du rein sporadiques qui représentent la 8<sup>ème</sup> cause de cancer en France et ainsi d'aider au développement de traitement plus ciblés.

## b. Tumeurs du spectre VHL<sup>102,88</sup>

### i. *Hémangioblastomes du système nerveux central*

Les hémangioblastomes du système nerveux central sont les tumeurs les plus fréquentes du spectre VHL et affectent 60 à 80% des patients<sup>103,104</sup>. Elles sont bénignes et souvent multiples. Elles se développent le long de l'axe cranio-facial. Leur localisation peut conduire à une augmentation de la pression intracrânienne avec pour principaux symptômes des paresthésies (fourmillement, picotement, engourdissement) ou des douleurs. Les symptômes peuvent aller jusqu'à la perte de l'utilisation des neurones sensitifs ou moteurs. Dans 20% des cas, une polyglobulie secondaire est associée à ces tumeurs qui sécrètent de l'EPO. Leur croissance est lente et se caractérise par une alternance de périodes de développement et de périodes de quiescence. L'âge moyen de diagnostic est de 30 ans mais des cas ont été retrouvés de 9 à 70 ans. L'âge de survenue des tumeurs est donc très hétérogène entre les patients. Ces hémangioblastomes sont des tumeurs encapsulées composées de cellules stromales d'origine mésenchymateuse et d'un abondant réseau de capillaires sanguins. Le traitement utilisé est la

chirurgie pour retirer les tumeurs au fur et à mesure de leur apparition pour limiter les effets secondaires qui en découlent. Elle reste la principale cause de morbidité et de décès étant donné le nombre et la fréquence des tumeurs développées.

#### *ii. Hémangioblastomes rétiniens*

Les hémangioblastomes rétiniens affectent 50% des patients qui développent la maladie de VHL et peuvent aussi être appelés angiomes rétiniens<sup>105,106</sup>. L'âge moyen de diagnostic est de 25 ans mais des cas sont retrouvés de 1 à 68 ans, l'âge de survenue est donc pour ces tumeurs aussi très hétérogène. Les hémangioblastomes sont souvent multiples et bilatéraux et se développent en périphérie ou sur le nerf optique. Ils sont bénins mais leur localisation peut conduire à une perte partielle ou totale de la vue. En absence de détection précoce, ils peuvent conduire à un décollement de la rétine, des hémorragies, des glaucomes ou la cataracte. Plus de 25% des patients atteints d'hémangioblastomes rétiniens ont une perte de la vision permanente. Les traitements les plus efficaces pour ces tumeurs sont la photo-coagulation laser ou la cryothérapie. Ces tumeurs ont la même composition que les hémangioblastomes du système nerveux central à savoir des cellules stromales, qui sont d'origine mésenchymateuses, avec un réseau dense de capillaires sanguins<sup>107</sup>.

#### *iii. Carcinome rénal à cellules claires (RCC : renal cell carcinoma)*

La moyenne d'âge d'apparition des carcinomes rénaux à cellules claires dans le cadre de la maladie de VHL est de 39 ans mais ils peuvent se déclarer entre 13 et 70 ans. Passé 60 ans, 75% des patients qui développent la maladie de VHL déclarent un RCC qui est souvent bilatéral et multicentrique et souvent associé à des kystes rénaux<sup>108,109</sup>. Les tumeurs ont des aspects kystiques et sont constituées de cellules claires issues de l'épithélium du tubule proximal dans un stroma hypervascularisé. Il est à noter que ces cellules sont entourés de cellules productrices d'EPO qui peuvent envoyer des signaux aux cellules épithéliales. Ce carcinome est responsable de 50% des décès liés à la maladie de VHL. Il affecte autant les hommes et les femmes dans le cadre de la maladie de VHL contrairement aux RCC sporadiques liés notamment au tabac, qui affectent plus les hommes. Les cancers du rein les plus agressifs peuvent évoluer vers des

métastases dans le foie, les poumons et les os<sup>110</sup>. Le traitement le plus efficace reste aujourd'hui la chirurgie<sup>111</sup>. Elle doit avoir lieu assez tôt afin de pouvoir garder les néphrons et donc l'activité rénale intacte pour maintenir les fonctions rénales le plus longtemps possible étant donné que les tumeurs sont multiples. La chirurgie conservatrice évite aux patients de devoir être placés sous dialyse ou de devoir subir une transplantation rénale. Il faut également limiter le plus possible le développement des tumeurs à proximité des veines pour limiter la propagation des métastases. Dans les cas les plus avancés, la néphrectomie totale peut être la seule option pour les chirurgiens.

#### *iv. Phéochromocytomes*

Les phéochromocytomes sont des tumeurs des surrénales qui secrètent des catécholamines (adrénaline et noradrénaline) qui peuvent engendrer des effets secondaires tels que des palpitations, de la tachycardie, de la transpiration excessive, des maux de tête, de l'hypertension, des nausées ou une augmentation de la pression artérielle. Ils sont souvent multiples et bilatéraux et sont malins dans seulement 5% des cas. La moyenne d'âge d'apparition de la 1<sup>ère</sup> tumeur est 27 ans mais les tumeurs peuvent se développer entre 5 et 58 ans<sup>112</sup>. Les phéochromocytomes peuvent être révélateurs de la maladie de VHL et peuvent rester la seule manifestation de la maladie, notamment dans le cas du type 2C<sup>113</sup>. Le traitement recommandé est la chirurgie par laparoscopie quand la tumeur est peu développée, sinon, les chirurgiens procèdent à une surrénalectomie partielle pour préserver la fonction des surrénales (sécrétion de différentes hormones) et la qualité de vie des patients<sup>114</sup>. Les cellules à l'origine des tumeurs sont des cellules chromaffines (qui contiennent des catécholamines) de la glande surrénale qui dérivent de la crête neurale embryonnaire, elles font partie du système neuroendocrine.

D'autres tumeurs issues de cellules neuroendocrines regroupées en para-ganglions peuvent également se développer, il s'agit de paragangliomes qui sont des tumeurs qui peuvent sécréter des catécholamines comme les phéochromocytomes. Les para-ganglions et donc les paragangliomes se situent le long des axes vasculaires et nerveux de la tête et le long de la colonne vertébrale. Comme pour les phéochromocytomes, la sécrétion de catécholamines est responsable des effets de la tumeur tels que l'hypertension artérielle ou les maux de tête.

#### *v. Tumeurs neuroendocrines du pancréas*

Les tumeurs neuroendocrines du pancréas se développent chez environ 15% des patients affectés par la maladie de VHL<sup>115</sup>. Les plus agressives conduisent à la formation de métastases dans le foie, surtout lorsque la taille de la tumeur du pancréas dépasse une taille de 3 cm avant d'être réséquée. La moyenne l'âge d'apparition des tumeurs est de 38 ans mais des cas sont retrouvés entre 16 et 68 ans. Les tumeurs sont le plus souvent accompagnées par le développement de kystes multiples du pancréas qui sont retrouvés chez plus de 70% des patients<sup>116</sup>. Le traitement utilisé est la chirurgie pour retirer la tumeur et/ou les kystes. Comme pour les phéochromocytomes, les cellules à l'origine de la tumeur sont d'origine neuroendocrine et dérivent de la crête neurale.

#### *vi. Tumeurs du sac endolymphatique*

Les tumeurs du sac endolymphatique sont assez rares dans la population générale mais sont retrouvées chez 15% des patients qui développent la maladie de VHL. Elles sont bilatérales dans 30% des cas. Ces tumeurs entraînent une perte de l'audition qui est parfois détectée avant la tumeur étant donné que des tumeurs de très petite taille peuvent affecter l'audition<sup>117</sup>. L'âge moyen de diagnostic est de 22 ans mais des cas sont retrouvés de 12 à 50 ans. Les tumeurs ont pour origine l'épithélium endolymphatique et sont hypervascularisées. Elles peuvent conduire à l'érosion de l'os temporal, ce qui rend irréversible la perte d'audition. Dans le cas de tumeurs très développées, les patients peuvent présenter une paralysie faciale du fait de l'écrasement du nerf facial. Le traitement recommandé est la chirurgie pour réséquer la tumeur<sup>118</sup>.

#### *vii. Tumeurs kystiques de l'épididyme*

Les tumeurs kystiques de l'épididyme sont retrouvées chez plus de la moitié des hommes qui développent la maladie de VHL. La moyenne d'âge de diagnostic est de 36 ans avec des cas retrouvés entre 16 et 81 ans<sup>119</sup>. Ces tumeurs peuvent être uni ou bilatérales et elles dérivent de l'épididyme qui est formée à partir du conduit mesonéphrique (ou canal de Wolff) embryonnaire. Elles sont le plus souvent asymptomatiques mais les tumeurs bilatérales larges peuvent entraîner

des problèmes de fertilité et un inconfort. La chirurgie n'est pratiquée que dans les cas les plus importants avec des grosses tumeurs.

Chez les femmes, de rares cas de tumeurs du ligament large (qui correspondraient à des tumeurs de l'épididyme chez l'homme) sont répertoriées<sup>120</sup>. Ces tumeurs apparaissent entre 16 et 46 ans chez les femmes qui développent la maladie de VHL. Comme pour les tumeurs de l'épididyme, la chirurgie n'est pratiquée que lorsqu'elles sont trop développées.

## 2. [La polyglobulie de Chuvash](#)

### a. Les différents types de polyglobulie

La polyglobulie est une pathologie caractérisée par une production excessive de globules rouges qui peut avoir plusieurs origines<sup>121</sup>. Elle est classée en 2 types.

Elle est dite primaire lorsque la cause est liée à un défaut inhérent aux progéniteurs érythroïdes. La cascade de signalisation EPO/EPOR est activée de façon constitutive et les cellules prolifèrent de façon indépendante ou à très faibles concentrations d'EPO (hypersensibilité à l'EPO). Dans ce cas, le taux d'EPO sanguin est abaissé. Ce type d'érythrocytose peut être acquise suite à la survenue d'une mutation du gène JAK2 (la plus répandue étant la mutation de *JAK2 V617F*<sup>122</sup>) qui conduit à une activation constitutive de la kinase JAK2. On parle alors de polyglobulie de Vaquez qui est un syndrome myeloprolifératif d'occurrence tardive. Elle peut également être héréditaire *via* la transmission d'une mutation hétérozygote du gène codant le récepteur à l'EPO<sup>123,124</sup>. Les mutations décrites entraînent la perte du domaine C-terminal d'EPOR responsable du rétrocontrôle négatif. Le récepteur est alors activé de façon constitutive et stimule en excès la prolifération et la différenciation des progéniteurs en globules rouges.

La polyglobulie est dite secondaire quand elle a pour origine une surexpression de l'EPO. Dans ce cas, le taux d'EPO sanguin est supérieur à la normale. Elle peut être acquise *via* la sécrétion d'EPO par une tumeur (les symptômes disparaissent après résection de la tumeur en cause). Elle peut

également être héréditaire *via* la transmission d'une mutation d'un gène de la voie de l'hypoxie à l'état hétérozygote pour les gènes *PHD2* et *HIF2A* (pathologie autosomique dominante) et homozygote pour le gène *VHL* (pathologie autosomique récessive). Récemment, il a également été décrit une mutation gain de fonction dans le gène de l'*EPO*<sup>125</sup>. Il s'agit d'une courte délétion qui conduit à la surexpression d'une isoforme modifiée de l'*EPO*, ce qui conduit au développement d'une érythrocytose chez toutes les personnes affectées.

#### b. La polyglobulie de Chuvash

Les premiers cas d'érythrocytoses héréditaires liées au gène *VHL* ont été décrits dans la région de Chuvash qui est une région enclavée de l'ex URSS où le taux de consanguinité est très important. Les patients de cette communauté étaient porteurs d'une mutation du gène *VHL* c.598C>T, p.Arg200Trp (*VHL*-R200W) située à la fin de l'exon 3 du gène à l'état homozygote<sup>126</sup>. Il a été montré que cette mutation est apparue il y a 14000 à 62000 ans et qu'elle est transmise avec un haplotype core particulier dans cette région<sup>127</sup>. Cette même mutation a par la suite été retrouvée impliquée dans la pathologie notamment au Pakistan et sur l'île italienne de Ishia (sans l'haplotype core de Chuvash)<sup>128</sup>. Depuis, de nombreuses autres mutations du gène *VHL* ont été décrites à l'origine de polyglobulies à l'état homozygote, hétérozygote composite ou dans de rares cas inexplicables à l'état hétérozygote<sup>129,130</sup>. Les mutations à l'origine des polyglobulies sont dans la grande majorité différentes de celles qui prédisposent à la maladie de VHL, mais quelques cas inexplicables sont décrits pour des mutations impliquées dans les deux pathologies.

Les mutations homozygotes du gène *VHL* prédisposent les patients au développement d'une érythrocytose<sup>126,121</sup> qui comporte les caractéristiques des polyglobulies secondaires et primaires. Elle est caractérisée par une production excessive de globules rouges liée à une surexpression d'*EPO* associée à une hypersensibilité des progéniteurs érythroïdes pour l'*EPO*<sup>126</sup>, un taux sanguin élevé d'hémoglobine et de VEGF, une faible pression sanguine (dûe au sang plus visqueux) et une espérance de vie réduite<sup>131</sup>. La polyglobulie est associée à des varices, de rares cas d'hémangiomes vertébraux (tumeurs bénignes asymptomatiques) et augmente le risque d'accident vasculaire cérébral (AVC) ou de thromboses qui sont le plus souvent la cause de la mort

du patient. Cependant, les patients ne développent jamais de tumeurs malignes comme pour la maladie de VHL.

Dans le cadre des polyglobulies primaires, la surproduction d'EPO est due à l'expression d'une protéine VHL mutée uniquement (patients homozygotes ou hétérozygotes composites, quelques rares cas hétérozygotes décrits qui finissent le plus souvent par s'avérer être hétérozygotes composites). Cette protéine garde la capacité de se lier à HIF mais le temps de dégradation de HIF en conditions normoxiques est plus lent du fait d'une perte de fonctionnalité partielle de la protéine VHL. En conditions normoxiques, il y a alors une production plus importante que la normale d'EPO. Celle-ci va se lier aux progéniteurs érythroïdes qui vont se multiplier de façon anormale et se différencier en globules rouges en quantité plus importante que celle nécessaire pour le fonctionnement normal de l'organisme.

En ce qui concerne les caractéristiques de la polyglobulie primaire, il semblerait que la mutation *VHL-R200W* ait perdu sa capacité à fixer SOCS1, ce qui conduit à une perte d'ubiquitylation de JAK2<sup>47</sup>. Cela conduit à une stabilisation de JAK2 avec une sur-activation de la voie phospho-STAT5 dans les progéniteurs érythroïdes. Cela conduit à une hypersensibilité à l'EPO des progéniteurs. En plus de la production excessive d'EPO par les cellules productrices d'EPO, s'ajoute donc une hypersensibilité des progéniteurs à celle-ci, ce qui favorise la pathologie qui a donc les deux caractéristiques : primaire et secondaire.

Le seul traitement efficace à ce jour reste la saignée régulière des patients afin de diminuer leur nombre de globules rouges. Cette pathologie, même si elle affecte génétiquement autant les femmes que les hommes, est souvent moins sévère chez les femmes du fait des menstruations qui permettent une élimination naturelle des globules rouges en excès. On voit également, sur le même principe, une diminution des symptômes chez les donneurs de sang réguliers.

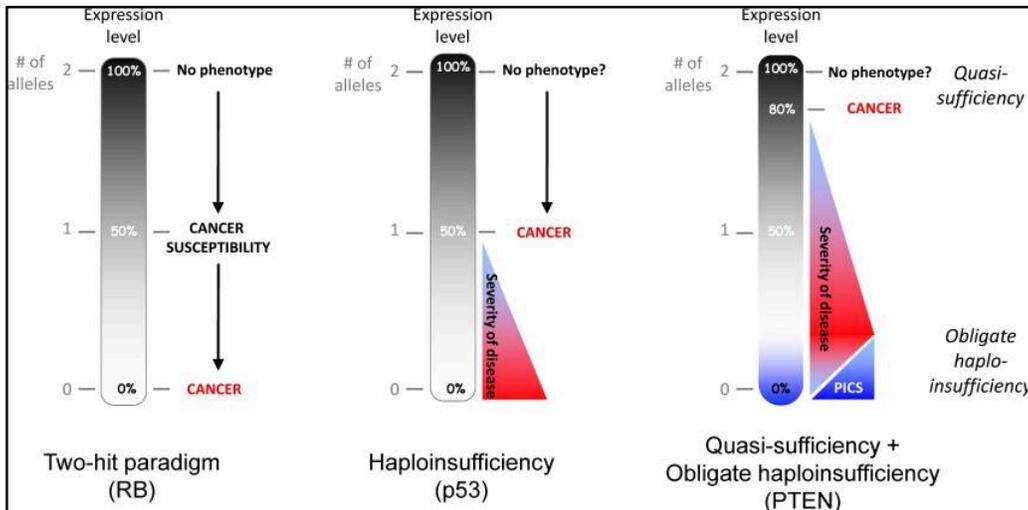
### [3. Complexité des corrélations génotype/phénotype](#)

### a. Modèle d'oncogénèse en gradient

Des mutations d'un seul gène, ici *VHL*, prédisposent au développement de pathologies de sévérité très différentes, soit une polyglobulie qui n'est associée au développement d'aucune tumeur, soit la maladie de VHL qui est caractérisée par le développement de nombreuses tumeurs au cours de la vie. La problématique des laboratoires est de comprendre la complexité qu'il existe entre les génotypes et les phénotypes observés.

Plus de 1500 mutations ont déjà été décrites à l'origine de ces deux pathologies, elles sont de différents types à savoir faux sens (52%), non sens (11%), délétion/insertion plus ou moins importante avec ou sans décalage du cadre de lecture (30%) ou mutation d'épissage (7%)<sup>132,133</sup>. Les mécanismes moléculaires à l'origine des différences phénotypiques observées sont en cours d'investigation mais les derniers résultats suggèrent qu'il existe une corrélation entre la sévérité de la pathologie et l'importance de la perte de fonction du gène *VHL*<sup>57</sup> (Figure 10). En 2011, un nouveau modèle graduel en "continuum" plus subtil que le modèle « Two-hit » décrit précédemment a été proposé<sup>134</sup> (Figure 10). Le modèle initial de Knudson établissait que la perte totale de fonction des deux allèles était nécessaire au développement de tumeur. Ici, le modèle tient compte de sa sensibilité au dosage génique qui peut entraîner des modifications visibles même lors d'une inactivation partielle d'un gène suppresseur de tumeur mais également de la spécificité de chaque type tissulaire. On parle d'haploinsuffisance lorsque la perte d'un allèle est suffisante pour l'induction d'un cancer. La quasi-insuffisance est décrite lorsque la perte d'un allèle n'est pas nécessaire pour le développement d'un cancer mais que la perte de fonction du suppresseur de tumeur est corrélée au risque de développer une tumeur. L'haploinsuffisance obligatoire correspond au stade où la perte de fonction totale du gène suppresseur de tumeur n'est pas viable. Cela a lieu dans certains tissus pour certains gènes suppresseurs de tumeurs (comme PTEN) lorsque les mécanismes cellulaires tels que la sénescence ou l'apoptose sont déclenchés lorsque la fonction des deux allèles du gène suppresseur de tumeur est perdue. Ce mécanisme est incompatible avec le développement de tumeurs. Il est associé à la sécrétion, notamment en sénescence de facteurs proinflammatoires nommés SASP (Senescence-Associated Secretory Phenotype)<sup>135</sup>. Cela conduit à une conversion des cellules sénescents en cellules

proinflammatoires qui peuvent promouvoir la progression tumorale des cellules voisines. Les mécanismes de protection mis en place par les cellules se transforment alors en mécanismes tumorigéniques pour les cellules environnantes.



**Figure 10 : Paradigme des gènes suppresseurs de tumeur.** (Berger et al.<sup>134</sup>) Le gradient central grisé représente l'expression du gène suppresseur de tumeur avec le nombre d'allèle présent. Le gradient rouge et bleu représente la sévérité de la pathologie développée. A gauche est représenté le modèle « Two-hit » de Knudson où la perte d'un allèle d'un gène suppresseur de tumeur prédispose au cancer et la perte des deux allèles conduit au développement du cancer. Au milieu est représenté le modèle de l'haploinsuffisance où il existe un gradient de sévérité lors de la perte du deuxième allèle. A droite sont représentés les modèles de quasi-insuffisance où le gradient de sévérité de la pathologie débute depuis la perte du 1<sup>er</sup> allèle jusqu'à la perte du 2<sup>ème</sup> allèle et d'haploinsuffisance obligatoire qui correspond à la perte totale des deux allèles qui est compensée par un autre mécanisme cellulaire tel que la sénescence et l'apoptose qui « protège » du développement de tumeurs.

134

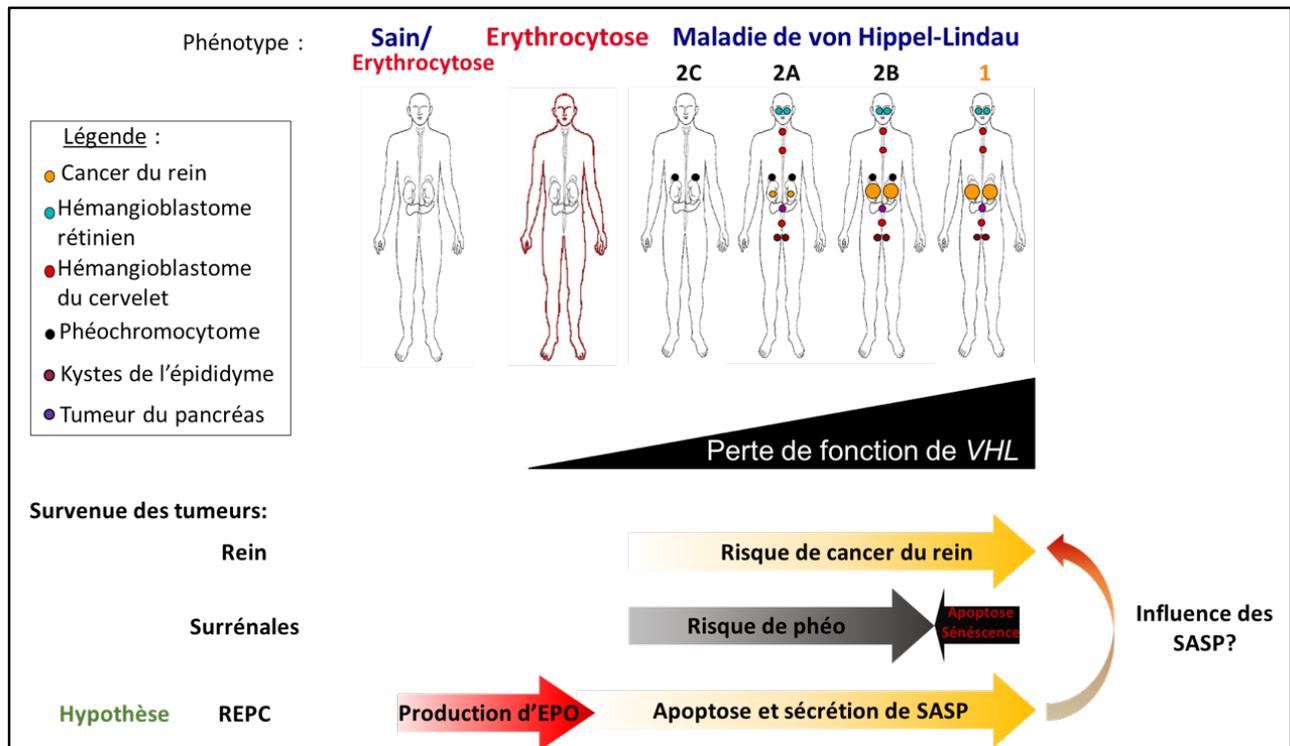
L'inactivation du gène suppresseur de tumeur *VHL* suivrait également un modèle d'oncogenèse en gradient (Figure 11). Une faible altération de *VHL* conduirait au développement d'érythrocytose bénigne. A l'inverse, une altération sévère de *VHL* entraînerait le développement de multiples tumeurs. Le type 1 de la maladie de VHL, qui est le plus sévère au niveau phénotypique est associé à des pertes quasi totales de fonction de *VHL* (délétions, mutations non sens), alors que le type 2 est plutôt associé à des mutations faux-sens qui entraînent des altérations sévères (type A ou B) ou moindres (type 2C) de la protéine VHL<sup>136,100</sup>. Pour le type 1 de la maladie, les patients ne développent pas de phéochromocytomes pour lesquels ont pourrait

parler d'haploinsuffisance obligatoire : une perte totale de fonction de VHL serait, dans le cadre des phéochromocytomes, le signal de l'induction de mécanismes de sécurité cellulaires tels que la sénescence ou l'apoptose. En effet, une mutation trop sévère de *VHL* pourrait être associée à l'entrée en apoptose des cellules à l'origine des phéochromocytomes lors de la survenue du 2<sup>ème</sup> événement sur *VHL*<sup>85</sup>.

Dans le cadre des cancers du rein pour lesquels l'origine cellulaire est épithéliale, notre hypothèse est que le mécanisme d'oncogenèse pourrait être indirect. Les cellules du micro-environnement rénal (cellules interstitielles productrices d'EPO) pourraient, dans le cas d'une perte sévère de *VHL*, entrer en sénescence ou en apoptose et sécréter des facteurs oncogéniques comme les SASP. D'ailleurs, les patients qui développent la maladie de VHL ont un taux d'EPO normal contrairement à ceux qui développent une polyglobulie liée à une mutation héréditaire de *VHL* chez qui le taux d'EPO est élevé. Les SASP seraient donc d'après notre hypothèse l'origine des tumeurs rénales développées par les patients dans le cadre de la maladie de VHL.

Aucune des mutations à l'origine de la maladie de VHL n'est jamais retrouvée à l'état homozygote. Il est probable que de fortes altérations de pVHL ne soit pas viable au niveau développemental.

Pour rappel, les mutations du gène *VHL* qui conduisent au développement d'une érythrocytose sont toujours différentes de celles à l'origine de la maladie de VHL. Il existe quelques rares exceptions comme par exemple les mutations Y175C et L188V associées à la fois à des érythrocytoses et à des phéochromocytomes<sup>66,137,138,139,140</sup>. Ces mutations sont donc considérées avec une perte de fonction insuffisante pour induire des cancers rénaux mais suffisantes pour induire des phéochromocytomes ou des érythrocytoses en fonction du contexte génétique.



**Figure 11 : Corrélations génotype/phénotype associées aux mutations du gène *VHL*<sup>57</sup>.** Les patients porteurs de mutations du gène *VHL* développent des phénotypes différents. L'hypothèse est que la sévérité de la pathologie est corrélée à la sévérité de la perte de fonction de *VHL* (« modèle d'oncogenèse en gradient »). Pour les phéochromocytomes, une perte de fonction trop importante de *VHL* conduirait à l'entrée en apoptose (ou sénescence) des cellules. Pour les cellules productrices d'EPO, une faible perte de fonction conduit à une surproduction d'EPO. Nous faisons l'hypothèse qu'une perte de fonction plus forte conduit à l'entrée en apoptose (ou sénescence) et à la sécrétion de facteurs proinflammatoires tels que les SASP (Senescence-Associated Secretory Phenotype), possiblement tumorigéniques pour les cellules rénales environnantes.

### b. Mutations complexes de *VHL*

Il a été montré que la protéine *VHL* sauvage (WT) est stabilisée grâce à son association avec le complexe élongine B et C<sup>65</sup>. En revanche, dès que p*VHL* est porteuse d'une mutation qui affecte son affinité pour l'élongine C, elle va alors être rapidement dégradée *via* le protéasome. Cette dégradation a lieu suite à la phosphorylation de la tyrosine 185 de p*VHL* par Src, une tyrosine kinase cytoplasmique, ce qui conduit à son ubiquitinylation et à sa dégradation *via* le protéasome<sup>141</sup>. A l'inverse, une mutation de la tyrosine 185 conduit à la stabilisation de p*VHL*. En effet, cette mutation ne permet plus la phosphorylation de p*VHL* qui est le signal de sa

dégradation, on observe alors une stabilisation de pVHL chez les patients porteurs d'une mutation en position 185.

Il est à noter que des cas de délétions totale du gène *VHL* sont décrites dans le cadre de la maladie de VHL. Elles sont hétérogènes et associées à des hémangioblastomes du système nerveux central et de la rétine avec un faible risque de développer des cancers du rein<sup>142,143,144,145</sup>. Ce risque est d'autant plus faible si la délétion du gène *VHL* comprend aussi la délétion du gène voisin *C3orf10*, aussi appelé *HSPC300*. Ce gène code une protéine très conservée appelé BRICK1 qui est nécessaire pour la prolifération cellulaire et la transformation cellulaire par les oncogènes<sup>146</sup>. Elle régule également l'organisation du cytosquelette et la dynamique de l'actine. C'est pourquoi la perte de ce gène ou son inhibition protège du développement tumoral du fait d'un défaut de prolifération et de motilité des cellules tumorales. Il semble donc que le développement de cancers du rein n'est pas uniquement lié à la taille de la délétion du gène *VHL* mais des gènes qui y sont associés.

Etonnement, des mutations différentes d'un même codon de VHL peuvent conduire à des manifestations de gravités différentes. Nous pouvons citer l'exemple du codon 138. La mutation c.413C>T, p.Pro138Leu (*VHL*-P138L) est associée à une polyglobulie<sup>137</sup> alors que la mutation c.414A>G, p.Pro138Pro (*VHL*-P138P) a été identifiée à l'origine de la maladie de VHL. Pour l'étude des mutations du gène *VHL*, même si celui-ci n'est constitué que de 3 exons, il est important de vérifier l'implication de la mutation dans les différents sites de régulation de l'épissage.

#### 4. Modèle existants : lignées cellulaires, iPS, modèles animaux

L'objectif du laboratoire est de comprendre la complexité qui existe entre les génotypes et les phénotypes associés à la maladie de VHL et aux polyglobulies. Pour cela, il faut disposer d'un modèle cellulaire adapté. Pour l'heure, les outils à disposition, listés ci-après, sont assez rares et ne permettent pas une analyse assez fine des mutations à l'origine des pathologies.

### a. Lignées cellulaires

Comme dit précédemment, seules quelques lignées cellulaires sont capables de produire de l'EPO. L'étude de la dérégulation de son expression est donc très délicate et les outils sont rares. De plus, ces lignées sont dérivées de tumeurs de patients qui produisent de l'EPO, qui contiennent des mutations relatives à la tumeur initiale qui peuvent interférer avec l'analyse de mutations de la voie de l'hypoxie. En effet, les cellules rénales productrices d'EPO ne sont à ce jour pas immortalisées puisqu'elles sont très difficiles à isoler et personne n'a encore réussi à les maintenir en culture avec leur capacité à produire de l'EPO. Une seule lignée avait été décrite mais l'article a été rétracté car il s'agissait en réalité d'une lignée commerciale (Hep3B) qui avait contaminé la culture<sup>147</sup>. D'autres modèles sont donc nécessaires pour étudier au mieux l'impact de mutations de la voie de l'hypoxie au niveau cellulaire.

Il est possible de dériver des lignées lymphoblastoïdes (LLB) à partir de lymphocytes de patients immortalisés grâce au virus de l'EBV (Epstein-Barr Virus). Cela permet d'étudier l'expression de gènes en amont de la régulation de HIF. En effet, la transformation par le virus de l'EBV entraîne des modifications transcriptionnelles en aval du facteur de transcription HIF, partie de la voie qu'il est donc impossible d'étudier avec ces lignées. Les LLB de patients peuvent donc être utilisées pour étudier l'expression de *VHL* et de ses différentes isoformes en présence ou non d'une mutation. Il est également possible d'y transférer des siRNA dirigés contre *VHL* pour en étudier les effets uniquement sur l'expression de *VHL*.

### b. Modèle iPS

Il n'existe à ce jour aucun modèle publié de cellules rénales productrices d'EPO dérivée d'iPS de patients. Ce modèle pourrait être très utile pour l'analyse de la dérégulation de la voie de l'hypoxie dans le cadre de la maladie de VHL ou de polyglobulie. En effet, ce modèle cellulaire pourrait permettre de recréer en culture un modèle de cellules productrices d'EPO à partir de cellules souches de patients porteurs de mutations de la voie de l'hypoxie pour étudier les impacts phénotypiques au niveau cellulaire ainsi que le transcriptome de ces cellules. Il serait également possible d'utiliser des cellules contrôles dans lesquelles des mutations seraient introduites grâce

à la technologie CRISPR. Le problème majeur pour obtenir ce modèle est l'origine des cellules productrices d'EPO qui reste aujourd'hui controversée. Beaucoup d'hypothèses penchent en faveur d'une origine de la crête neurale pour ces cellules. En effet, il a été montré chez la souris que durant l'embryogenèse, l'EPO est exprimée dans le foie mais également dans les cellules de la crête neurale<sup>33</sup>. Il a plusieurs fois été mis en évidence que le premier site de production d'EPO durant l'embryogenèse est la crête neurale puisque l'érythropoïèse est active avant les hépatocytes<sup>148</sup>. Il serait donc intéressant de se pencher sur la différenciation d'iPS en crête neurale qu'il faudrait ensuite réussir à faire dériver en cellules rénales productrices d'EPO pour obtenir un modèle robuste.

### c. Modèles animaux

A ce jour il n'existe aucun modèle animal qui mime parfaitement la maladie de VHL. Les essais réalisés de modèle de la maladie de VHL avec des knock out (KO) homozygotes de *VHL* conduisent à la mort des embryons *in utero* entre E11,5 et E12,5<sup>149</sup>. Des modèles de KO inducibles *via* le modèle Cre/Lox ont pu être générés afin d'étudier le rôle tissu spécifique de *VHL* au cours du développement et dans les organes fonctionnels. Cependant, aucun de ces modèles ne mime le développement de tumeurs différentes et tout au long de la vie de la maladie de VHL. Il existe uniquement des modèles qui développent un certain type de tumeur du spectre VHL. Les seuls modèles animaux mimant réellement une pathologie du spectre VHL sont des modèles d'érythrocytoses qui ont été construits chez la souris<sup>150</sup> et chez le poisson zèbre<sup>151</sup>. Pour la souris, le modèle réalisé est porteur d'une équivalente à la mutation *VHL*-R200W à l'état homozygote, ce qui conduit au développement d'une polyglobulie très similaire à celle observée chez l'homme entre 10 et 14 semaines après la naissance. En effet, cette souris présente des taux d'EPO et de VEGF élevés au niveau transcriptionnel et un nombre élevé de progéniteurs érythroïdes avec une augmentation de leur différenciation quand on les cultive *in vitro*. Il s'agit donc d'un très bon modèle pour l'étude de la polyglobulie et les éventuels tests de nouveaux traitements. Cependant, le modèle souris laisse une limite à savoir la taille plus restreinte des introns murins qui ne permet pas l'analyse des exons cryptiques qui peuvent être retrouvés chez l'homme. De

plus, chez la souris, il n'y a pas de lncRNA qui peuvent intervenir dans le développement de pathologies chez l'homme.

Chez le poisson zèbre, la voie régulée par le facteur de transcription HIF est très conservée<sup>152</sup>, de même que les mécanismes d'hématopoïèses<sup>153,154</sup> et de signalisation *via* l'EPO<sup>155</sup>. Deux mutations inactivatrices de l'orthologue de *VHL* chez le poisson zèbre ont été obtenues et analysées, l'une homozygote, l'autre hétérozygote<sup>151</sup>. Ces deux mutants conduisent au développement d'une polyglobulie avec une augmentation du taux d'EPO et de VEGF circulants. On note également une hyper-prolifération des tissus hématopoïétiques primitifs et définitifs. Il est à noter que dans le cas du mutant homozygote, les poissons survivent à l'embryogenèse et il est donc possible d'étudier l'inactivation complète de *VHL* chez le poisson zèbre au stade post-embryonnaire. Il s'agit donc d'un bon outil pour étudier l'impact des mutations de *VHL* pour l'hématopoïèse et l'homéostasie de l'oxygène mais comme pour le modèle murin, il n'est pas utile pour l'étude du développement tumoral dans le cadre de la maladie de VHL.

## D. Isoformes d'épissage

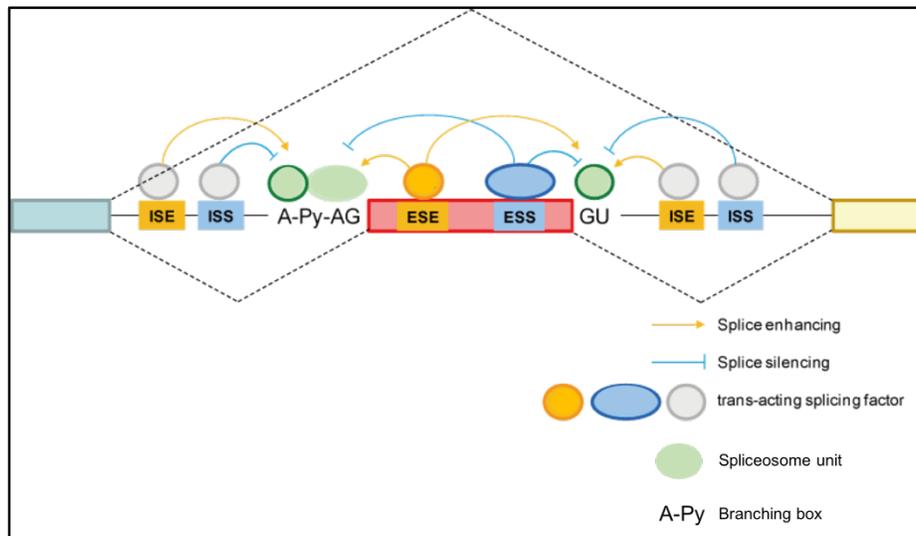
### 5. Epissage alternatif et épissage pathologique : origines, produits et méthodes d'analyse

#### a. Description

Les gènes sont constitués d'une alternance d'exons et d'introns qui sont éliminés lors de l'épissage des pré-ARN messagers (ARNm) en ARNm<sup>156,157</sup>. Il a été montré que 92 à 94% des gènes peuvent être épissés de façon alternative et qu'il en résulte différentes isoformes protéiques qui peuvent avoir des fonctions semblables, distinctes ou totalement opposées<sup>158,159</sup>.

Les sites donneurs (SD) et accepteurs (SA) d'épissage sont des séquences nucléotidiques situées en début et fin des exons qui vont être reconnues par les composants du spliceosome (ou particule d'épissage). Il s'agit d'un complexe ribonucléoprotéique, composé d'ARN et de plus de

200 protéines qui, grâce à deux réactions de trans-estérification, va permettre d'exciser les introns et de lier les exons entre eux<sup>160</sup>. Le splicéosome va dans un premier temps se fixer à l'ARN pré-messager sur une séquence appelée point de branchement qui est située quelques nucléotides avant le site accepteur d'épissage. L'utilisation d'un site accepteur ou d'un autre dépend de sa « force » qui est déterminée par la divergence de sa séquence par rapport à une séquence consensus reconnue par un ARN du spliceosome<sup>161</sup>. Plus un site est « fort », plus il sera utilisé en conditions dites normales. Les forces des sites sont également déterminées par les séquences régulatrices ou ESR (séquences régulatrices d'exons) qui favorisent ou inhibent un épissage donné en réponse à la fixation d'une protéine régulatrice enhancer ou silencer qui vont favoriser ou non la fixation du spliceosome sur le site d'épissage. Il existe différents types de séquences régulatrices : ESE (Site enhancer d'épissage exonique), ISE (Site enhancer d'épissage intronique), ESS (Site silencer d'épissage exonique) ou ISS (Site silencer d'épissage intronique) (Figure 12). Ces séquences permettent également la régulation de l'expression de différents transcrits de façon tissu spécifique<sup>162</sup>. En effet, certaines protéines qui se fixent sur ces séquences pour activer ou inhiber l'épissage d'un exon vont préférentiellement être exprimées dans certains types cellulaires et non dans d'autres. C'est pourquoi pour certaines protéines, des isoformes sont plus spécifiquement exprimées dans certains tissus et en réponse à certains signaux.

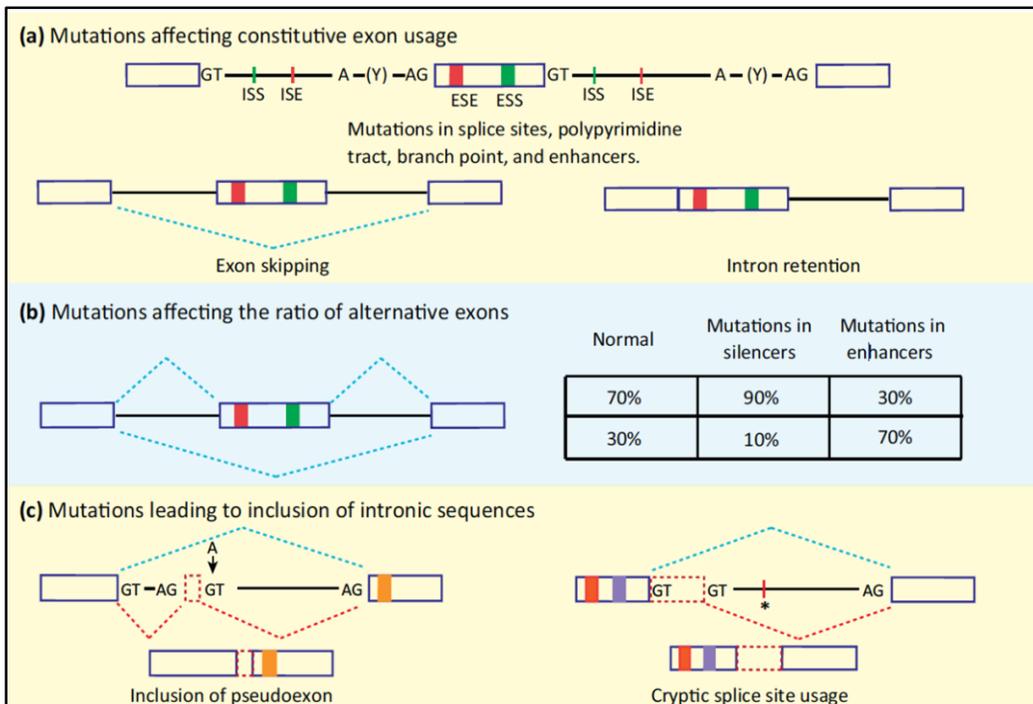


**Figure 12 : Séquences et facteurs régulateurs de l'épissage.** L'épissage des exons est régi par la présence de séquences donneurs (AG) et accepteurs (GU) d'épissage qui définissent le début et la fin d'un exon. Le splicéosome va se fixer sur la boîte de branchement et va être activé ou inhibé par les facteurs d'épissages se liant aux séquences ISE (Intronic Splicing Enhancer), ISS (Intronic Splicing Silencer), ESE (Exonic Splicing Enhancer) et ESS (Exonic Splicing Silencer).

De plus, il a été montré que des conditions hypoxiques favorisent le saut d'exon de gènes dont l'expression n'est pas régulée par la voie de l'hypoxie ou dont l'hypoxie diminue l'expression<sup>163,164</sup>. Dans le même temps, l'hypoxie inhibe l'inclusion d'exon dans les gènes réprimés en conditions d'hypoxie. A l'inverse, l'hypoxie favorise l'inclusion d'exons des gènes cibles de HIF. Cette inclusion d'exon a pour origine l'activité transcriptionnelle spécifique de HIF et non une réponse cellulaire au stress hypoxique. Il en résulte un changement des ratios des différentes isoformes de certaines protéines suite à une hypoxie où à une mutation de la voie de l'hypoxie qui mime une hypoxie constante. Cet épissage alternatif augmente également le nombre de transcrits qui ne codent pas de protéines<sup>165</sup>. Le shift hypoxique vers une production de transcrits non codant conduit à une augmentation des dommages à l'ADN, notamment dans les tumeurs *via* une diminution de l'expression de p53 qui est l'un des « gardien » du génome *via* la régulation de différentes voies de réparation du génome.

## b. Perturbations de l'épissage et impacts en pathologies

En plus de l'hypoxie, certaines mutations peuvent conduire à des défauts d'épissage (Figure 13). Il a été montré qu'environ 15% des mutations sont situées sur un site d'épissage et qu'environ 20% des mutations faux sens sont situées sur un site régulateur d'épissage, soit environ un tiers des mutations pourraient impacter l'épissage<sup>166,167</sup>. Beaucoup de défauts d'épissage ont été décrits pour les myopathies, notamment dans le gène codant la Desmine (*DES*), avec des mutations qui conduisent à des sauts d'exons<sup>168</sup>. Des défauts d'épissage ont déjà été décrits pour le gène *VHL*. Celui-ci n'étant constitué que de 3 exons, les mutations d'épissage conduisent le plus souvent à un saut de l'exon 2 dont plusieurs cas ont été décrits<sup>169,133,170</sup>. Les mutations doivent être situées dans un point de branchement ou dans des séquences régulatrices de l'épissage : ESE, ISE, ESS ou ISS (Figure 13a). Une mutation d'un site très conservé peut empêcher la fixation d'une protéine régulatrice ou créer un nouveau site de fixation d'une de ces protéines dans le gène et donc conduire à une altération de l'épissage. Une mutation d'une séquence silencer d'épissage (ESS ou ISS) conduit le plus souvent à l'augmentation de l'inclusion d'exon alors que des mutations des séquences enhanceurs d'épissage conduisent le plus souvent à la diminution de l'inclusion d'exon. En plus des exons constitutifs d'un gène, des exons dits alternatifs ou pseudo-exons (ou introns alternatifs) peuvent être inclus dans la séquence d'un ARNm à la suite d'une mutation ou d'une période d'hypoxie, on parle de rétention d'intron<sup>171</sup> (Figure 13c). Cela est possible grâce à la longueur des introns chez les vertébrés supérieurs qui est suffisamment importante pour contenir une séquences qui peut devenir un exon alternatif<sup>172</sup>. Le plus souvent, ces exons alternatifs peuvent être inclus en conditions dites normales mais le ratio de leur expression est très faible par rapport à l'isoforme majoritaire du gène. Une mutation modifiant l'épissage d'un gène conduit souvent à une modification des ratios des transcrits exprimés à partir d'un gène. La rétention d'intron est décrite comme un mécanisme courant de l'inactivation des gènes suppresseurs de tumeur<sup>173</sup>. En effet, l'exonisation intronique (insertion d'un exon alternatif) conduit à une perte de la fonction de la protéine sauvage, ce qui conduit à son inactivation.



**Figure 13 : Séquences permettant la régulation de l'épissage et différents types de défauts d'épissage. (D'après Singh et al.<sup>174</sup>)** Des mutations dans les séquences régulatrices de l'épissage peuvent conduire à des défauts d'épissage tels que l'inclusion de parties introniques, le saut d'exon ou le changement de ratio des différentes isoformes. ESE = Exonic splicing enhancer, ESS = Exonic splicing silencer, ISE = Intronic splicing enhancer et ISS = Intronic splicing silencer.

174

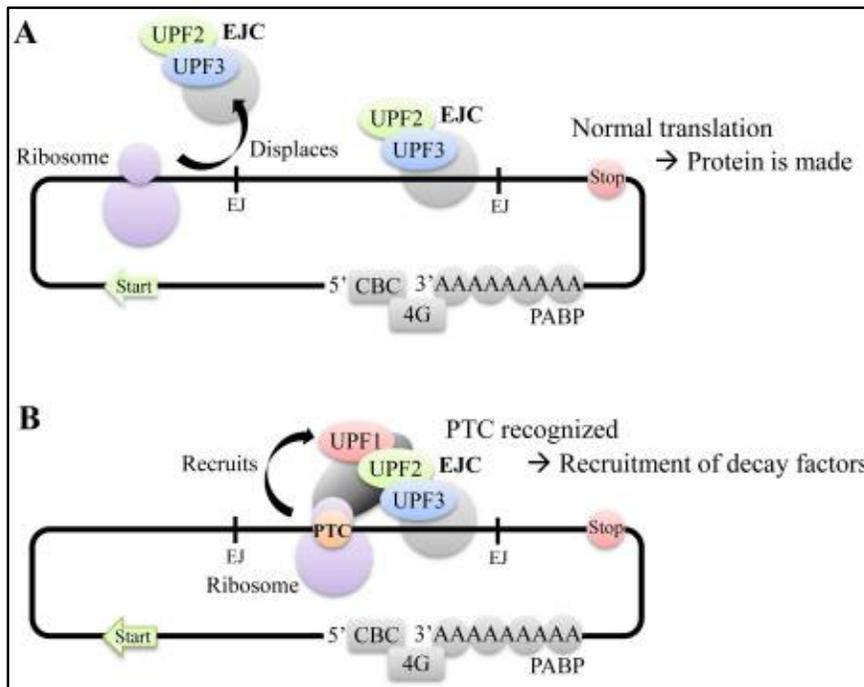
Des mutations des sites d'épissage conduisent le plus souvent à l'utilisation du site d'épissage le plus proche qui est souvent utilisé uniquement en 2<sup>ème</sup> intention. Cela ajoute ou retire donc quelques nucléotides aux transcrits. Les effets obtenus peuvent donc être très variés allant de l'ajout de codons n'influençant pas la fonction de la protéine et ne décalant pas le cadre de lecture à l'ajout d'un codon stop prématuré qui entraîne la dégradation du transcrit.

L'ajout ou le retrait d'un exon dans un transcrit peut conduire à un décalage du cadre de lecture ou à l'insertion d'un codon stop dit prématuré, c'est à dire qui est localisé dans la séquence avant le codon stop de la protéine sauvage. Ce codon prématuré classe l'ARNm dans les cibles du mécanisme de NMD (nonsense mediated decay) qui va conduire à sa dégradation<sup>175</sup>.

Des outils bio-informatiques permettent de prédire l'effet des mutations d'un gène sur l'épissage. Ces analyses *in-silico* permettent de déterminer les mutations qu'il est intéressant d'étudier en analyses fonctionnelles, notamment *via* une analyse en minigènes rapporteurs<sup>176</sup>. Ces outils informatiques ne sont pas fiables à 100% car pour prédire l'effet d'une mutation sur l'épissage, surtout lorsque celles-ci impactent des ESE car il faut tenir compte à la fois de la séquence de l'ARNm avec ces séquences régulatrices mais il faut également tenir compte de la structure secondaire des ARNm qui n'est pas toujours bien définie<sup>177</sup>. Le dernier outil diagnostique décrit en mai 2018 par Lemman et al., nommé SPicE : Splicing Prediction in Consensus Element prédit correctement l'impact sur l'épissage pour 98,8% des variants testés<sup>178</sup>. Il reste compliqué d'établir avec certitudes une relation de cause à effet entre une altération de l'épissage et une pathologie<sup>179</sup>. En effet, la tissu spécificité de l'épissage n'est pas prise en compte par les logiciels d'analyse *in silico*.

### c. Le mécanisme de Nonsense Mediated Decay (NMD)

Le mécanisme de NMD est un mécanisme de surveillance de l'ARNm qui protège les cellules des effets potentiellement dangereux des protéines tronquées qui pourraient être produites à partir d'ARNm qui contiennent un codon stop prématuré. Un codon de terminaison de la traduction (TAA, TAG ou TGA) est dit « prématuré » lorsqu'il est situé au moins 50 à 55 nucléotides avant la dernière jonction exon/exon de l'ARNm<sup>180,181,182</sup> (figure 14). Ces messagers, appelés « nonsense mRNA », doivent être dégradés par les cellules le plus vite possible après leur synthèse avant d'être traduits en protéine. Pour cela, la cellule possède un mécanisme de « screen » des ARNm quand ceux-ci sont encore liés au noyau (durant la transition noyau-cytoplasme)<sup>183</sup>. En effet, ces protéines tronquées peuvent agir comme des dominant négatifs ou avoir un gain de fonction délétère par rapport à la protéine sauvage. L'origine d'un codon stop prématuré peut être une mutation non-sens ou qui entraîne un décalage du cadre de lecture. Une autre origine peut être un épissage aberrant qui par la rétention d'un exon alternatif ou le saut d'un exon ajoute à la séquence protéique un codon stop prématuré<sup>184</sup>.



**Figure 14 : Schématisation du mécanisme de Nonsense Mediated Decay.** (D'après Nguyen et al.<sup>185</sup>) (A) Traduction d'une protéine sans codon stop prématuré. Le ribosome retire les complexes EJC (Exons Junction Complexes), associés aux protéines UPF3 et UPF2, situés juste avant les jonctions entre exons. La traduction s'arrête lorsque le ribosome rencontre le codon stop. (B) Traduction d'une protéine avec un codon stop prématuré (PTC). Lorsque le ribosome rencontre un codon stop alors qu'il reste un EJC sur l'ARNmessenger, il y a recrutement d'un complexe protéique dont la protéine UPF1 va être phosphorylée, ce qui conduit à la dissolution du ribosome.

185

La dégradation d'un ARNm avec un codon stop prématuré est basé sur la position du codon stop par rapport à l'extrémité 3'UTR (untranslated transcribed region) du messenger<sup>186</sup>. En effet, si la distance entre le codon stop et les éléments de la région 3'UTR est trop importante, l'arrangement spatial entre l'ARNm et les protéines qui s'y fixent pour le premier cycle de traduction n'est pas bon<sup>187</sup>. De plus, lors de l'excision des introns, des complexes protéiques appelés EJC (exons junction complexes) vont se lier aux ARNm. Lors de la première étape de traduction, un ribosome va lire l'ARNm et retirer les EJC. Quand le ribosome arrive à un codon stop mais qu'il reste un EJC sur l'ARNm, cela va déclencher le mécanisme de NMD. Cela consiste au recrutement de plusieurs protéines qui forment un complexe entre le ribosome et l'EJC. Les 3 principales sont les protéines « up-frameshift proteins » : UPF1, UPF2 et UPF3, ces deux dernières

faisant partie du complexe EJC. UPF1 va être phosphorylée par la protéine kinase SMG-1, ce qui va conduire à la dissociation du ribosome<sup>188,189</sup>. Va suivre le recrutement d'une enzyme dipeptidyl- carboxypeptidase, Dcp, qui va conduire à la perte de la coiffe de l'ARNm en 5'. Cela va conduire à la dégradation rapide de l'ARNm par l'exonucléase 5'-3' Xrn<sup>190</sup>.

Ce mécanisme est très peu étudié dans le cadre du gène *VHL* car celui-ci n'est constitué que de 3 exons, des mutations tronquantes ont donc peu de chance de se retrouver en phase avec le dernier exon de la protéine en étant situé au moins 50 nucléotides avant la jonction exon2/exon3. Cependant, une étude a été menée sur deux mutations frameshift (décalage du cadre de lecture) de *VHL*<sup>191</sup>. Il a été montré que la mutation frameshift de l'exon 1 c.163delG, à l'origine d'un hémangioblastome du système nerveux central, conduit à la dégradation du transcrit par le mécanisme de NMD. En revanche, la mutation c.239delG, à l'origine de la maladie de VHL de type 1, ne conduit pas à la dégradation du transcrit par le mécanisme de NMD alors qu'elle est également située dans l'exon 1 de *VHL*. Il en résulte une protéine tronquée avec un codon stop en position 158. Une autre étude a montré que des codons stop prématuré sur les acides aminés 104 et 147 ne conduisent pas à une dégradation forte des transcrits codés<sup>192</sup>. Dans cette étude, les auteurs concluent que le gène *VHL* n'a pas une taille suffisante pour que les transcrits qui contiennent un codon stop prématuré puissent être dégradés via le mécanisme de NMD.

## 6. Les longs ARN non codant

Les longs ARN non codant sont définis comme étant des ARN de plus de 200 bases non traduits en protéine<sup>193</sup> et font partie de la grande famille des ARN non codants (ncRNA) qui représentent à ce jour 70% des ARN transcrits depuis notre génome<sup>194,195</sup>. Ils ont été mis en évidence lors de la recherche de nouveaux gènes qui a permis la découverte de cette nouvelle classe d'ARN qui n'est pas classée dans la catégorie des gènes puisqu'ils ne sont pas traduits. Ils ont en premier lieu été décrits comme des ARN messagers « like » non codant<sup>196</sup>. Leur séquence ADN peut être intergénique (situé dans les régions non annotées du génome) ou intragénique (chevauchent les gènes codants) et peut notamment être le fruit d'un épissage alternatif d'un gène. Ils sont

faiblement conservés entre les espèces<sup>197</sup>. Ces ARN non codant peuvent être polyadénylés ou non<sup>198</sup> et sont le plus souvent caractérisés comme des ARN régulateurs du génome pouvant agir à toutes les étapes de l'expression génique *via* leur capacité à interagir avec l'ADN, l'ARN ou les protéines<sup>199,195</sup>. Il a été montré qu'ils peuvent intervenir dans de nombreuses fonctions cellulaires telles que la maintenance de la pluripotence des cellules souches embryonnaires<sup>194</sup>, la régulation du cycle cellulaire<sup>200</sup> ou encore le développement tumoral<sup>201</sup>. Seuls 10% des longs ARN non codant sont aujourd'hui décrits comme ayant une fonction<sup>202</sup>. Même s'ils semblent impliqués dans plusieurs mécanismes, leur mode de fonctionnement le plus décrit est le rôle d'adaptateur dans la formation des complexes ribo-nucléiques<sup>203,204</sup>. En effet, les longs ARN non codants peuvent permettre le maintien de 5 à 20 protéines à proximité quand ils ont une taille de 100 nucléotides alors qu'une protéine échafaudage (scaffold) de 100 acides aminés peut seulement maintenir à proximité 2 protéines<sup>205</sup>. De part leur capacité à pouvoir interagir avec des protéines ou des acides nucléiques, les longs ARN non codants peuvent maintenir ces deux types de constituants cellulaires à proximité<sup>206</sup>. Les mécanismes d'action des longs ARN non codant peuvent être variés comme la fonction de guide, d'enhancer, d'adaptateur ou de signal dans le promoteur d'un gène pour activer ou inhiber son expression<sup>207</sup>. Ils peuvent également servir « d'éponge » à miRNA, de régulateur d'épissage ou de cofacteur dans le remodelage chromatinien<sup>195</sup>.

Etant donné leur implication dans le développement tumoral, ils peuvent aujourd'hui servir de nouveaux marqueurs thérapeutiques du développement tumoral et métastatique<sup>201</sup>. C'est notamment le cas d'un long ARN non codant nommé  $\square$ HIF qui est décrit comme un marqueur de mauvais pronostic dans certains cancers du poumon<sup>208</sup> et qui est également surexprimé dans certains cancers du rein<sup>209</sup>.  $\square$ HIF est un transcrit antisens naturel (NAT) complémentaire de la région 3' non codante de HIF1 $\square$  qui le régule négativement<sup>210</sup>. Il est exprimé dans les tissus foetaux et adultes normaux ainsi que dans les tissus tumoraux. Il a été montré que le NAT  $\square$ HIF contient dans son promoteur une séquence HRE. Il est donc surexprimé en conditions hypoxiques, ce qui conduit à une boucle d'autorégulation de HIF1 $\square$  et de  $\square$ HIF<sup>211</sup>. La régulation des gènes par les longs ARN non codant rajoute donc un niveau de complexité dans la régulation cellulaire de l'expression des gènes qui est à prendre en compte. Cependant, à ce jour nous ne connaissons

pas et n'avons pas caractérisé tous les longs ARN non codant qui peuvent être exprimés. Aucun long ARN non codant n'est à ce jour associé à *VHL*, tant pour sa régulation *via* un lncRNA que pour un lncRNA codé par *VHL* ou *VHL like*.

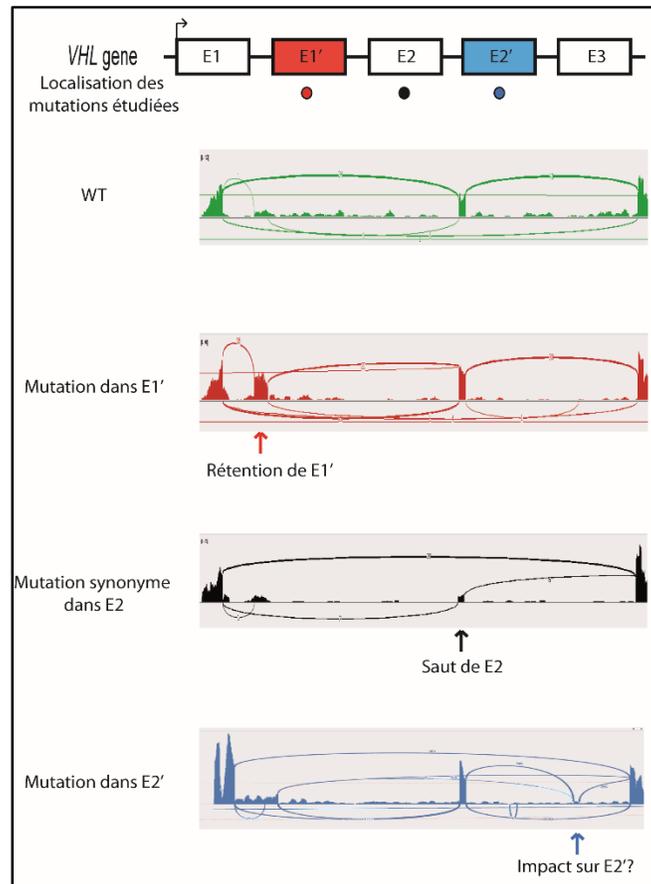
## E. Objectifs de thèse

L'objectif du laboratoire est d'identifier les mécanismes moléculaires à l'origine des différents phénotypes associés aux mutations *VHL*. Notre hypothèse est qu'il existe une corrélation entre la sévérité de l'altération de *VHL* et la gravité des symptômes. L'étude de protéines VHL mutées caractéristiques des différentes pathologies a permis de proposer un modèle d'oncogenèse en gradient<sup>57</sup> mais l'absence de tests fonctionnels suffisamment sensibles rend toute conclusion difficile. En effet, il s'est avéré très difficile d'étudier les effets, parfois subtils, engendrés par un simple changement d'acide aminé. Récemment, nous avons identifié des mutations synonymes dans l'exon 2 de *VHL* chez des patients atteints d'érythrocytose héréditaire (3 familles avec une mutation c.429C>T, p.Asp143Asp (*VHL*-D143D), suivies en Allemagne) et deux familles atteintes de la maladie de VHL (mutation c.414A>G, p.Pro138Pro (*VHL*-P138P) hétérozygote, suivies en France). Ces mutations nucléotidiques n'ont aucune influence sur la séquence protéique (en conséquence de la dégénérescence du code génétique) mais pourraient potentiellement impacter l'épissage et le ratio des deux isoformes de *VHL* connues : l'isoforme E1E2E3 active et l'isoforme E1E3 non fonctionnelle dans la voie de régulation de HIF.

L'objet de ma thèse porte sur l'étude génétique et fonctionnelle de ces cas de mutations synonymes dans le but de confirmer le modèle d'oncogenèse en gradient sans avoir à étudier l'effet subtil de mutations protéiques (Figure 15). Au cours de cette étude, nous avons découvert de façon fortuite de nouvelles isoformes d'épissage de *VHL* qui contiennent deux exons cryptiques situés dans les introns 1 et 2 de *VHL*. Un des objectifs de ma thèse a donc été d'étudier génétiquement et fonctionnellement ces deux exons cryptiques. Le séquençage de ces exons chez des patients qui développent des érythrocytoses ou la maladie de VHL nous a permis d'y découvrir

des mutations. Nous avons ainsi résolu des cas inexplicables de patients sans mutation du gène *VHL* ou avec une mutation à l'état hétérozygote pour des patients polyglobuliques. J'ai également participé à une étude de récurrence des variants de l'intron 1 de *VHL* dans une cohorte de malades de *VHL*, phéochromocytomes et paragangliomes.

Nous avons également développé un modèle nouvellement mis en place dans notre laboratoire de cellules productrices d'EPO dérivées d'iPS de patients qui m'a permis de pouvoir étudier plus finement l'impact des variants identifiés dans le gène *VHL*.



**Figure 15** : Schématisation des objectifs de thèse.

## II. MATERIELS ET METHODES

### *i. Extraction d'acides nucléiques*

L'extraction des ARN est réalisée sur colonnes avec le kit « Macherey Nagel DNA, RNA and protein purification » qui contient une étape de traitement à la DNase de 15 minutes. L'élution est réalisée dans 40µL d'H<sub>2</sub>O RNAses Free.

L'extraction d'ADN plasmidique après amplification dans des minicultures (5mL) de bactéries est réalisée avec le kit « Macherey Nagel Nucleospin® Plasmid » en partant de 2mL de culture bactérienne. L'élution a lieu dans 50µL de tampon d'élution. Celle des maxicultures (500mL) est réalisée avec le kit « Macherey Nagel Nucleobond Xtra Maxi Plus® Endotoxine Free » en partant d'un volume de 300mL de culture bactérienne. L'élution est réalisée dans 500µL de tampon d'élution.

L'extraction d'ADN génomique est réalisée avec le kit Qiagen « DNeasy Blood and Tissue » à partir de 5 millions de cellules. L'élution a lieu dans 200 µL de tampon AE.

La purification des produits de PCR après migration en gel d'agarose et excision de la bande d'intérêt est réalisée avec le kit « Macherey Nagel Nucleospin® Gel and PCR Clean Up ». L'élution est réalisée dans 15µL de tampon NE.

Le dosage des acides nucléiques est réalisé au Nanodrop par lecture de l'absorbance à 260nm.

### *ii. Reverse transcription*

Les RT-PCR sont réalisées avec un kit de reverse transcription « Fermentas » « ThermoFisher » (reverse transcription avec amorçage hexamère). Le mix de 20µL est constitué de tampon de réaction dilué à 1X et d'enzyme reverse transcriptase diluée au dixième et est complété avec de l'eau « RNase free ». La quantité d'ARN initiale utilisée est 500ng dans un volume de 8µL, ce qui nous permet d'obtenir 20µL d'ADN complémentaire à 25ng/µL. La quantité d'ARN est réduite et les volumes ajustés si la concentration initiale des échantillons n'est pas suffisante.

### *iii. Détection des isoformes de VHL*

Les PCR destinées à la détection des isoformes de *VHL* sont réalisées avec le kit « Taq DNA polymérase » de chez invitrogen avec une quantité initiale d'ADNc de 100ng amplifiée durant 40 cycles à 60°C. Cette polymérase a la propriété de rajouter une adénine en fin de produit de PCR, ce qui permettra son clonage dans un vecteur spécifique (TOPO-TA). Les amorces utilisées, localisées dans les exons 1 et 3 de *VHL* sont VHL-Exon1 forward : 5'-CAGGTCATCTTCTGCAATCG-3' et VHL-Exon3 reverse : 5'- TTCGTAGAGCGACCTGACGATGTC-3'. 10µL de produits de PCR sont ensuite déposés sur gel d'agarose à 2% avant que chacune des bandes ne soit extraite, clonée puis séquencée.

### *iv. Clonage des produits de PCR et transformation bactérienne*

Le clonage des bandes obtenues par PCR est réalisé avec le kit « TOPO®-TA Cloning Reaction » (Invitrogen) à l'aide du vecteur pCR®2.1TOPO®. 4µL de produits de PCR purifiés sur gel ont été utilisés pour chaque réaction de ligation, additionnés de 1µL de vecteur et de ligase et de 1µL de solution saline (200 mM NaCl; 10 mM MgCl<sub>2</sub>). Les transformations sont réalisées avec 5µL de produit de ligation dans des bactéries E.Coli TOP10 (25µL) avec une incubation de 30 min dans la glace, un choc thermique de 30 secondes à 42°C et une incubation de 45 min à 37°C dans du milieu SOC (Super Optimal broth with Catabolite repression) liquide. Celui-ci est ensuite étalé du milieu LB (lysogeny broth) et incubé 12h à 37°C.

### *v. Séquençage d'ADN*

Le séquençage de l'ADN est réalisé par séquençage capillaire Sanger sur une plateforme dédiée. Les produits de PCR sont purifiés par « ExoSAP » (Affimetrix) (Exonucléase 1 qui dégrade les amorces de PCR). Les séquences sont lues et alignées au laboratoire grâce au logiciel BioEdit®.

### *vi. Quantification des isoformes de VHL*

La quantification des différentes isoformes transcriptionnelles de *VHL* est réalisée après reverse transcription par PCR quantitative (qPCR) avec la méthode Taqman afin d'augmenter la spécificité

de la réaction (du fait de l'utilisation d'amorces encadrant la région à étudier et d'une sonde marquée d'un fluorochrome spécifique de la séquence amplifiée). Elles sont réalisées avec le kit « qPCR MasterMix No ROX » (Eurogentec) avec des couples amorces/sondes spécifiques des différents exons du gène *VHL* (Les séquences des différents couples sonde/amorces sont présentées dans le tableau suivant). 0.4µM de sonde et 0.6µM de chaque amorce sont utilisés pour chaque réaction durant laquelle 20ng d'ADNc sont amplifiés. La température d'hybridation utilisée est 60°C pour chaque couple. Chaque échantillon est quantifié 2 fois (duplicat) dans une même qPCR.

Exon 1	5'- CAGGTCATCTTCTGCAATCG-3' (Forward)
	5'- CAGCACGACGCGCGGACT-3' (Reverse)
	5'-CCGTCTGAAGTTGAGCCATA-3' (Sonde)
Exon 2	5'-CACCTTTGGCTCTTCAGAGATGCAG-3' (Forward)
	5'- GGCAGTGTGATATTGGCAAAAATAGGCTG-3' (Reverse)
	5'- GACACACGATGGGCTTCTGGTTAAC-3' (Sonde)
Exon 3	5'- ACTCTGAAAGAGCGATGCCTCCA-3' (Forward)
	5'- TTCGTAGAGCGACCTGACGATGTC-3' (Reverse)
	5'- GTTGTCCGGAGCCTAGTCAAGC-3' (Sonde)
Jonction exon1 – exon2	5'- ATCCACAGCTACCGAGGTCA-3' (Forward)
	5'-TAACCAGAAGCCCATCGTGT-3' (Reverse)
	5'- CCCTGCATCTCTCTGAAGAGCCAAAGG-3' (Sonde)
Jonction exon1 – exon3	5'- CGCATCCACAGCTACCGAGTGTATAC-3' (Forward)
	5'- TAGAGCGACCTGACGATGTCCAG-3' (Reverse)
	5'- GAAAGAGCGATGCCTCCAGGTTG-3' (Sonde)
Jonction exon1 pseudo exon E1'	5'-CATCCACAGCTACCGAGTCC-3' (Forward)
	5'-CAACATGACCCTGGCTTGTTTC-3'(Reverse)
	5'- CCAGTTCTGCGTAGTCCCTGCC-3' (Sonde)
Jonction pseudo exon E1' – exon3	5'-ACTAGCCAACCCTAGTGTATACTCTG-3' (Forward)
	5'- TCCAGTCTCCTGTAATTCTCAGGC-3' (Reverse)

	5'- ACAACCTGGAGGCATCGCTCTT-3' (Sonde)
Jonction exon2 – pseudo exon E2'	5'-CTTGCCAATATCACACTGCCAGAGTCTC-3' (Forward)
	5'-GGAGGCATCGCTCTTTCAGAGTATACTG-3' (Reverse)
	5'-CTGTGTCACCCAGGCTGGAGT-3' (Sonde)

La quantification des marqueurs de différenciation et de l'EPO est réalisée avec des sondes TaqMan commerciales (LifeTechnologies) dans un volume final de 7µL en plaque 384 puits. Elles sont réalisées avec le kit « qPCR MasterMix No ROX » (Eurogentec) avec 5ng d'ADNc et 3,5µM de mix de sonde/amorces. La température d'hybridation utilisée est 60°C pour chaque couple. Chaque échantillon est quantifié 2 fois (duplicat) dans une même qPCR.

#### *vii. Culture cellulaire de lignées*

Nous avons à disposition des lignées LLB (lignées lymphoblastoïdes) qui correspondent à des lymphocytes de patients immortalisés par le virus EBV (Virus d'Epstein-Barr). Ces cellules en suspension sont cultivées en milieu RPMI additionné de 10% de SVF, 1% d'antibiotiques penicilline et stréptomycine (100 unités/mL) et 1% de L-glutamine (2µmol/mL). Différentes lignées cellulaires ont été utilisées : HEK293T (cellules épithéliales de rein humains), 786-O (cellules épithéliales d'adénocarcinome rénal humain, *VHL*<sup>-/-</sup> qui surexpriment HIF2α), Hep3B (cellules de carcinome de foie humain), les RC1s<sup>212</sup> (cellules de carcinome rénal humain productrices d'EPO), HK2 (cellules épithéliales de tubule proximal rénal humain), HeLa (cellules épithéliales d'adénocarcinome utérin) et WT7 (cellules dérivées des 786-O transfectées de façon stable avec pRC-HAVHL). Ces cellules sont cultivées en milieu DMEM 4.5g/L de glucose additionné de 10% de sérum de veau fœtal (SVF), de 100 unités/mL d'antibiotiques penicilline et stréptomycine et 2µmol/mL de glutamine. Les PC12 (cellules de phéochromocytome de rat) sont cultivées en milieu DMEM 4.5g/L de glucose additionné de 10% de SVF, 5% de sérum de cheval, 100 unités/mL d'antibiotiques penicilline et stréptomycine et 2µmol/mL de glutamine. Les cellules UT7 (progéniteurs érythroïdes) sont cultivées en milieu MEM (Minimum Essential Medium) additionné de 10% de SVF, 1% de d'antibiotiques et 1% de L-glutamine. Les différentes

lignées cellulaires sont transfectées grâce à du « jetPRIME® » (Polyplus transfection) qui est un liant polyvalent sur une durée de 24h. Le ratio utilisé est de 3µL de JetPrime pour 1µg d'ADN.

#### *viii. Analyses in silico des séquences d'épissage*

L'analyse *in silico* de la dérégulation potentielle en réponse à une mutation des sites potentiels de fixation de facteurs d'épissage de l'exon 2 et des pseudo exons E1' et E2' de *VHL* ont été réalisés sur le site Human Splicing Finder<sup>176</sup>.

#### *ix. Expériences de minigènes*

Le meilleur moyen d'analyser un défaut d'épissage est de réaliser des séquençages des ARN dans le tissu dans lequel la mutation engendre une pathologie. Cependant, il n'est pas toujours possible d'obtenir ce tissu. Pour pallier à ce problème, il a été mis au point un test *in vitro* appelé test fonctionnel de rapporteur d'épissage par minigènes qui est un outil puissant d'analyse de l'impact de variants sur l'épissage<sup>213,214,215</sup>. Comme son nom l'indique, il s'agit d'une version simplifiée d'un gène qui contient uniquement 1 exon d'un gène flanqué d'une partie de ses introns. Quand les exons sont trop petits ou que la mutation implique le saut de plusieurs exons, il est possible d'intégrer plusieurs exons d'un gène dans cette version simplifiée, dans ce cas avec la totalité des introns situés entre les exons analysés. Ce vecteur peut être transfecté dans de nombreux types cellulaires, ce qui permet l'analyse de la tissu spécificité de l'épissage grâce à des RT-PCR sur l'ARN à la suite de l'expression du vecteur dans une lignée cellulaire.

Les expériences de « minigènes » sont réalisées en utilisant le plasmide pCas2 qui est un plasmide pcDNA3.1 dans lequel a été inséré le gène *SERPING1/C1NH* qui contient deux exons nommés « A » et « B » dont une partie de l'intron 1 a été remplacé par une séquence contenant les sites de restriction BamHI et MluI. Ces sites nous permettent d'insérer nos séquences d'intérêt (exon 2 et pseudo exon 1' de *VHL*) contenant ou non les mutations à étudier. Pour cela, les séquences d'intérêt sont amplifiées à partir d'ADN génomique de patients avec des amorces spécifiques contenant les séquences des enzymes de restriction BamHI et MluI à leurs extrémités. Les produits de PCR et le plasmide pCas2 sont ensuite digérés 3h à 37°C puis purifiés sur gel d'agarose. Après purification, une ligation est réalisée avec 50ng de plasmide et l'insert (ratio 1 : 3) durant

1h à température ambiante en utilisant la ligase « T7 DNA ligase » (New England Biolabs). 50µL de bactéries DH5α sont ensuite transformées avec 3µL de produit de ligation. Les clones obtenus sont amplifiés (minipréparation), séquencés et, après sélection, amplifiés à nouveau (maxipréparation).

Deux microgrammes des différentes constructions sont ensuite transfectés dans différentes lignées cellulaires ensemencées la veille en plaques 6 puits durant 24h. L'ARN est ensuite extrait et reverse transcrit en ADN complémentaire. Des PCR utilisant des amorces spécifiques des exons « A » et « B » de pCas2 sont réalisées en suivant les instructions du kit « PCR Go Taq G2 » de chez Promega avec une quantité initiale d'ADNc de 50ng. Les résultats sont observés en déposant 10µL de produits de PCR avec du bleu de charge dans un gel à 2% d'agarose et la migration a lieu à 75V durant 50 minutes.

#### *x. Construction des vecteurs codant X1*

La séquence codant la protéine VHL X1 (Exon 1 de *VHL* épissé avec l'exon cryptique E1') a été synthétisée dans un vecteur pENTR (pENTR-X1) (GeneArt, Life Technologies). Cette séquence, bordée des séquences des enzymes de restrictions BamHI et XHOI, a été extraite (après digestion et purification sur gel) et insérée dans pcDNA3.1-HA (après ligation avec l'enzyme « T7 DNA ligase »). Nous avons transformé, amplifié et séquencé des bactéries DH5α avec ces plasmides qui ont ensuite été amplifiés et séquencés.

Afin d'obtenir les différents mutants de X1, nous avons réalisé des mutagenèses dirigées. Les amorces de mutagenèse ont été commandées chez Eurogentec et les expériences ont été réalisées avec la Taq polymérase « Pfu Turbo DNA Polymerase » (Agilent). Les mutagenèses sont réalisées à partir de 50ng de plasmide pcDNA3.1-X1. Après 18 cycles d'amplification, les produits de PCR sont digérés avec l'enzyme Dpn1 1h à 37°C pour éliminer les plasmides initiaux (méthylés) puis sont déposés sur gel d'agarose à 2% afin de vérifier qu'il y a eu amplification. Trois microlitres de produits de mutagenèse sont utilisés pour transformer des bactéries DH5α. Les clones sont ensuite amplifiés et séquencés pour vérifier la présence de la mutation.

### *xi. Tests fonctionnels*

Un test rapporteur luciférase a été réalisé grâce au kit « Dual-Luciferase® Reporter Assay » (Promega). Des cellules 786-O ont été transfectées avec une quantité de plasmide finale de 400ng dont 200 ng de plasmides HRE-Luciférase, 20 ng de plasmides Renilla et différents plasmides codant VHL et X1. La Renilla est sous le contrôle d'un promoteur fort (SV40) et est donc exprimée de la même manière dans toutes les cellules. Elle va donc permettre de normaliser les résultats obtenus en fonction du taux de transfection des cellules. La luciférase est sous le contrôle d'un promoteur HRE (Hypoxia Response Element) qui va pouvoir être activé par HIF quand celui-ci n'est pas dégradé par pVHL. Ainsi, si pVHL est actif, il va entraîner la dégradation de HIF, il n'y aura donc qu'une très faible production de luciférase. Au contraire, si une mutation altère la protéine VHL au point que celle-ci ne puisse plus dégrader HIF, il y aura une forte expression de la luciférase. Après 24h de transfection, les cellules sont lysées avec du tampon PLB (Passive Lysis Buffer) afin d'en extraire les protéines. La luminescence émise par la luciférase en présence de son substrat LARII est lue grâce au lecteur « Mithras LB 940 Multimode Microplate ». Après lecture, un deuxième tampon STOP&Glo est ajouté, il entraîne l'arrêt de la réaction de la luciférase, la mesure de l'activité de la Renilla est alors lue grâce au substrat présent dans ce deuxième tampon.

### *xii. Western Blot*

Les Western Blot sont réalisés en gels NuPAGE® Bis-Tris Mini Gel 4-12% (Invitrogen). Dix-huit microlitres de lysats obtenus lors des tests fonctionnels (additionnés de bleu de dépôt et d'antioxydant) sont déposés sur le gel. La migration est réalisée en tampon MES, 15 min à 70V puis environ 2h à 100V. Le transfert des protéines est réalisé en tampon NuPage, 20% méthanol durant 1h à 30V sur membrane de nitrocellulose 0.2µM « Hybond ECL » (GE Healthcare). La membrane est ensuite saturée 1h à température ambiante avec du lait 5% en TBST-Tween 0.05%. Les protéines (contenant une étiquette hemagglutinine) sont détectées en utilisant un anticorps dirigé contre HA (Covance) dilué au 1/1000<sup>ème</sup> (incubation sur la nuit à 4°C). L'anticorps secondaire est un anticorps de chèvre anti-souris couplé à la peroxydase diluée au 1/10000<sup>ème</sup>

(Jackson) (incubation 1h à température ambiante). Le signal est révélé par chimio-luminescence après 5 min d'incubation avec le substrat « Clarity™ Western ECL Substrate » (BioRad).

### *xiii. Culture iPS*

Les lignées iPS utilisées sont soit obtenues à partir de PBMC de patients et sont reprogrammés grâce à des vecteurs viraux Sendai par la plateforme iPSC de Nantes soit obtenues à partir de cellules urinaires de patients<sup>216</sup> et reprogrammées au laboratoire par le kit EPI5 Episomal (Invitrogen). Ces cellules sont ensuite cultivées sur du Matrigel (Corning) dans du milieu StemMacs IPSBrew (Miltenyi) et maintenues dans un incubateur à 37°C à 5% de CO<sub>2</sub> et 4% d'O<sub>2</sub> (les conditions hypoxiques permettent le maintien de la pluripotence).

Les différenciations hépatiques sont réalisées en suivant le protocole publié par Si-Tayeb et al.<sup>217</sup>. Afin d'étudier l'impact de l'hypoxie, de jour 21 à jour 22, la moitié des puits de la différenciation sont placés en hypoxie à 1% d'O<sub>2</sub> dans une station hypoxie (Ruskinn).

Les différenciations en précurseurs de la crête neurale sont dérivées du protocole publié de Denham<sup>218</sup>. Les cellules sont plâtées sur de la rLaminine-521 (Corning) (coatée à 1mg/cm<sup>2</sup> de surface de culture) 3 jours avant le début de la différenciation. De J0 à J4, les cellules sont placées en milieu N2B27 (Neurobasal medium, DMEMF12, Insulin/transferrin/selenium (10X), N2 (100X), B27 retinol free (100X), Glucose (200g/L) et penicilline stréptomycine) avec du SB431542 StemMacs (Miltenyi) à 10µM et du CHIR99021 StemMacs (Miltenyi) à 3µM. A J4, le tapis cellulaire est découpé en carrés qui sont déposés un à un dans une plaque 96 puits faible attachement à fond plat (Corning) en milieu N2B27 additionné de Human BMP-2 à 5µg/µL et de Human FGF-2 à 20µg/µL et y restent jusque la fin de la différenciation à J11. Afin d'étudier l'effet de l'hypoxie, la moitié des cellules sont placées de J10 à J11 en hypoxie à 1% d'O<sub>2</sub> dans une station hypoxie.

### *xiv. Immunomarquage*

Les cellules à marquer sont fixées pendant 15 minutes à température ambiante dans du paraformaldéhyde (PFA) à 4% puis perméabilisées avec du triton X100 à 0,5% durant 15 minutes à température ambiante. Elles sont ensuite mises en présence avec l'anticorps primaire dilué en suivant les conseils du fournisseur pour 1h à température ambiante sous agitation. Elles sont

ensuite incubées durant 1 heure à température ambiante et à l'obscurité avec l'anticorps secondaire dilué au 1/1000<sup>ème</sup> dans du PBS, du DAPI (Mitenyi) au 1/500<sup>ème</sup> et si besoin avec de la Phalloïdine 568 Alexa Fluor (Cell signalling Technology) (marquage de l'actine en vert). Les sphéroïdes sont ensuite transparisés grâce à du RapiClear 1.9 (Sunjin Lab).

#### xv. *ELISA EPO*

Le dosage protéique de l'EPO est réalisé avec un kit ELISA Quantikine IVD ELISA (R&D systems) directement sur le surnageant de culture quand celui-ci est assez concentré. Dans le cas contraire, les dosages sont réalisés au centre anti-dopage de Chatenay Malabry (AFLD : association française de lutte contre le dopage) qui a une technique de concentration de l'EPO<sup>219</sup>.

### III. RESULTATS

#### 1. Etude de mutations synonymes de l'exon 2 de *VHL* et description d'un nouvel exon cryptique E1' de *VHL* – Blood, Lenglet et al., 2018<sup>220</sup>

##### a. Introduction

Ma thèse a débuté par l'étude de mutations synonymes de l'exon 2 de *VHL* chez 3 familles qui développent une polyglobulie (Familles 1, 9 et 10 de l'article 1) et 2 familles qui développent la maladie de VHL (Familles 11 et 12 de l'article 1). Chez ces dernières, toutes les personnes qui développent des tumeurs sont porteuses d'une mutation synonyme c.414A>G, p.Pro138Pro (*VHL*-P138P) à l'état hétérozygote. Cette mutation ne touchant pas la séquence protéique de pVHL, il fallait trouver un autre mode d'altération de pVHL à l'origine de la pathologie, la piste étant un défaut d'épissage qui conduirait au saut de l'exon 2 de *VHL*. Pour deux familles développant une polyglobulie (Familles 9 et 10 de l'article 1), les patients atteints sont porteurs d'une mutation synonyme c.429C>T, p.Asp143Asp (*VHL*-D143D) à l'état homozygote. Les porteurs de la mutation à l'état hétérozygote sont sains. Pour une famille (Famille 1 de l'article 1), le père et le fils sont porteurs de la mutation synonyme *VHL*-D143D à l'état hétérozygote, cependant, seul le fils développe une polyglobulie. C'est en étudiant cette famille et en réalisant une RT-PCR des ARNm de *VHL* avec des amorces situées dans les exons 1 et 3 du gène que nous avons remarqué la présence de bandes de tailles supérieures à celles attendues (Figure 1B de l'article). Après clonage et séquençage de ces bandes, nous avons pu montrer que pour l'une d'entre elle il y avait insertion d'une partie de l'intron 1 de *VHL* entre les exons 1 et 2 du gène que nous avons nommé « exon cryptique E1' ». Cet exon de 248 paires de bases est localisé entre les positions c.340+576 et c.340+834. Nous avons donc réalisé une étude approfondie de ce nouvel exon cryptique et de son isoforme protéique potentiellement associée. Nous avons également séquencé E1' chez les patients qui développent une polyglobulie avec une mutation de *VHL* à l'état hétérozygote et chez les patients qui développent la maladie de VHL sans altérations de *VHL* connues. Nous avons retrouvé des variants de cet exon chez 7 familles qui développent une

polyglobulie et chez 1 famille pour la maladie de VHL. Cela nous a permis de mettre en évidence que des mutations de l'exon cryptique E1' peuvent être à l'origine de polyglobulie ou de la maladie de VHL.

## b. Résultats

### i. *Etude de variants synonymes de l'exon 2 de VHL*

Dans un premier temps, mes recherches se sont portées sur des patients atteints d'une polyglobulie porteurs de mutations synonymes dans l'exon 2 du gène *VHL*. Deux patients homozygotes pour une mutation synonyme c.429C>T, p.Asp143Asp (*VHL*-D143D) ont développé la pathologie respectivement à l'âge de 5 mois et 4 ans et demi avec des taux d'EPO sanguin de 186 et 264 mU/mL pour un taux normal compris entre 3 et 17 mU/mL (Figure 5A). Ces mutations posaient question puisque, contrairement à l'ensemble des mutations répertoriées (hormis les délétions du gène), elles n'impactent pas directement la séquence protéique de VHL. Nous avons donc fait l'hypothèse qu'elles pourraient avoir un impact sur l'épissage du gène et le ratio des deux types de transcrits *VHL* connus : les transcrits contenant les exons 1, 2 et 3 (E1E2E3) et ceux contenant les exons 1 et 3 (E1E3).

Des études *in silico* de la mutation synonyme *VHL*-D143D ont prédit une modification d'un site exonique enhancer d'épissage (ESE) situé dans l'exon 2 de *VHL* (Figure supplémentaire 14). Afin d'étudier le potentiel impact sur l'épissage de cette mutation, nous avons réalisé des tests rapporteurs d'épissage de type « minigènes » (Figure 6B). Pour cela, l'exon 2 du gène *VHL*, sauvage ou muté, a été inséré avec une partie des introns qui l'encadrent dans un plasmide (pCas2) contenant deux exons artificiels A et B. Ce plasmide a ensuite été transfecté muté ou non dans différentes lignées cellulaires. Cela nous a permis de montrer que la mutation *VHL*-D143D entraîne un saut de l'exon 2 de *VHL*. Nous avons confirmé ces résultats par RT-PCR et des RT-qPCR sur des lignées lymphoblastoïdes (LLB) de patients, des leucocytes ou de l'ARN tumoral, dans lesquels nous avons observé une surexpression de l'isoforme E1E3 chez les patients porteurs de la mutation et une diminution de la quantité d'isoforme « fonctionnelle » E1E2E3 (Figure 5C et

D). Nous avons montré que cette perturbation de l'épissage est associée à une diminution de l'expression protéique totale de VHL (Figure 5E).

En parallèle, nous avons étudié deux familles qui ont développé la maladie de VHL associée à une autre mutation synonyme dans l'exon 2 de *VHL* à savoir la mutation c.414A>G, p.Pro138Pro (*VHL*-P138P) (Figure 5B). Comme pour la mutation *VHL*-D143D, les études *in silico* prédisent la modification d'un site ESE (Figure supplémentaire 14). Après analyse par test rapporteur d'épissage en « minigène », nous avons pu montrer que cette mutation entraîne un saut de l'exon 2 de *VHL* qui est plus important que pour la mutation *VHL*-D143D (Figure 6B). Ce résultat est en adéquation avec le modèle d'oncogenèse en gradient qui montre une corrélation entre l'importance de la perte de fonction de pVHL et la sévérité des symptômes observés. Nous avons réalisé un séquençage des ARN haut débit (RNAséq) sur une tumeur de patient porteur de la mutation *VHL*-P138P (Figure 6C). Cela nous a permis de confirmer dans la tumeur le saut de l'exon 2. Les résultats des RNAséq ont mis en évidence que les transcrits contenant l'exon 2 étaient majoritairement sauvages (très peu de transcrits contiennent l'exon 2 muté). L'analyse du profil d'expression de la tumeur a été comparé à des données transcriptomiques de 400 tumeurs de même type associées à des mutations germinales dans divers gènes (collection COMETE). Cela nous a permis de classer cette tumeur avec celles de patients porteurs de mutations *VHL* (associées à une surexpression de gènes de la voie de l'hypoxie) (Figure 6D). Cela nous a permis de démontrer que la mutation *VHL*-P138P est bien à l'origine du développement des tumeurs dans le cadre de la maladie de VHL. A noter qu'il pourrait être intéressant de réaliser des tests rapporteurs d'épissage pour toutes les mutations de *VHL* décrites dans l'exon 2. En effet, certaines mutations décrites dans le cadre de la maladie de VHL ou de polyglobulies n'ont aucun effet sur la fonction même de pVHL (Figure 6A). Ces mutations pourraient en fait entraîner un défaut d'épissage à l'origine de la maladie, comme nous avons pu le montrer pour les mutations *VHL*-P138L et *VHL*-G144R (Figure 6B).

Dans un second temps, nous avons réalisé l'étude d'un patient polyglobulique porteur de la mutation synonyme *VHL*-D143D, mais à l'état hétérozygote. Sachant que les mutations du gène

*VHL* associées aux polyglobulies sont toujours à l'état homozygote ou hétérozygote composite, nous avons recherché une anomalie supplémentaire dans le gène (Figure 1A).

ii. *Découverte et analyse de l'exon cryptique E1'*

Le patient porteur de la mutation *VHL*-D143D à l'état hétérozygote a été diagnostiqué pour une polyglobulie à l'âge de 15 ans avec un taux d'EPO sanguin de 163 mU/mL. Lors de l'étude de l'expression des différents transcrits *VHL* exprimés dans une lignée LLB de ce patient par RT-PCR, nous avons observé sur gel d'agarose l'apparition de bandes de tailles supérieures à celles attendues (Figure 1C). Après clonage et séquençage de ces bandes, nous avons pu mettre en évidence un nouveau transcrit de *VHL* contenant un exon cryptique situé dans l'intron 1 que nous avons nommé E1' (Figure 1D). Cet exon E1' est épissé en phase entre les exons 1 et 2 ou entre les exons 1 et 3 de *VHL*. Dans les deux cas, les transcrits possèdent la capacité de produire une protéine contenant les acides aminés codés par l'exon 1 (capable de fixer HIF) et 79 acides aminés sans homologie avec aucune protéine connue. Dans le même temps, un transcrit contenant cet exon E1' (épissé entre E1 et E2) a été découvert par séquençage haut débit et décrit sur le site NCBI, confirmant ainsi nos résultats. Cette protéine a été dénommée « X1 ». Cette isoforme a aujourd'hui été supprimée au profit d'un transcrit initié à partir d'un promoteur alternatif situé en amont de l'exon E1' de *VHL*, que nous avons nommé « upstream E1' » et qui est décrit comme ne codant pas une protéine. Après avoir mis au point des RT-qPCR spécifiques de chacune des isoformes contenant E1', nous avons pu mettre en évidence que les isoformes (épissées ou initiées à partir d'un nouveau promoteur) sont exprimées de façon endogène en faible quantité dans tous les tissus chez l'homme (Figure supplémentaire 1). Elles sont également exprimées dans toutes les lignées cellulaires que nous avons testées et nous n'avons pas observé de différence majeure d'expression entre la normoxie et l'hypoxie.

Nous avons réalisé des analyses *in silico* de conservation de E1' dans les espèces au niveau transcriptionnel et protéique (Figure supplémentaire 2A à D). Cela nous a permis de mettre en évidence que l'exon E1' est retrouvé uniquement chez les primates supérieurs.

Nous avons séquencé cet exon cryptique chez le patient hétérozygote *VHL*-D143D, ainsi que chez des patients porteurs de mutations hétérozygotes de *VHL* ayant développé une polyglobulie et chez de rares cas de patients ayant développé la maladie de VHL sans cause génétique connue (Figure 1 B). De façon très intéressante, nous avons trouvé une mutation (ponctuelle ou courte duplication) dans E1' chez 8 patients atteints de polyglobulie (7 diagnostiqués entre l'âge de 4 mois et 15 ans et 1 diagnostiqué à l'âge de 21 ans) et chez une grande famille atteinte de la maladie de VHL (apparition de la première tumeur à l'âge 10 et 15 ans chez 2 patients et entre 16 et 24 ans chez 4 patients) (Figure 2). Dans cette famille, une étude préalable de ségrégation de marqueurs microsatellites avait mis en évidence une ségrégation des marqueurs de la région du gène *VHL* avec la pathologie (Figure supplémentaire 7). En revanche, le séquençage des 3 exons de *VHL* et du promoteur s'était révélé négatif.

Nous avons ensuite réalisé une étude exhaustive de cet exon et de ces mutations.

Dans la famille atteinte de la maladie de VHL, nous avons démontré la ségrégation des mutations dans E1' uniquement chez les patients atteints (6 patients malades porteurs, 5 non porteurs sains). Seuls les patients porteurs d'une mutation dans E1' développent des tumeurs.

Chez les familles polyglobuliques, seuls les patients porteurs d'une mutation sur les deux allèles (hétérozygotes composites avec une mutation dans E1' et une mutation dans un autre exon de *VHL* ; ou mutation homozygote dans E1') développent la pathologie. Une étude de ségrégation nous a permis de montrer que les mutations retrouvées sont portées par deux allèles différents. Quand l'ADN des parents n'était pas disponible, nous avons réalisé une étude de polymorphismes de l'haplotype core Chuvash situés à proximité de E1', ce qui est le cas pour les patients des familles 2, 3 et 4 hétérozygotes pour la mutation *VHL*-R200W.

Le séquençage haut débit en RNAséq des LLB des patients ainsi que des tumeurs développées dans la famille atteinte de la maladie de VHL nous a permis de montrer que les mutations de l'exon cryptique E1' sont associées à une rétention importante de cet exon dans les transcrits VHL (Figure 3A). Ce résultat a été validé par RT-qPCR sur différentes lignées LLB de patients porteurs de mutation dans E1' ainsi que sur 2 tumeurs de la famille atteinte par la maladie de VHL (Figure 3B et C).

Sachant que la traduction des transcrits contenant E1' épissé entre E1 et E2 ou E1 et E3 est théoriquement capable de produire une protéine (X1) capable de fixer HIF (sans pour autant pouvoir le dégrader suite à l'absence de fixation au complexe E3 ubiquitine ligase), nous avons réalisé des tests fonctionnels rapporteurs de l'activité de HIF (Figure 4C). Nous avons mis en évidence que X1 n'est pas capable d'inhiber directement l'activité de HIF ni d'entrer en compétition avec VHL pour l'inhibition de HIF. Nous ne connaissons donc pas, pour le moment, la fonction de cette éventuelle protéine X1.

Il est important de noter que les transcrits contenant E1' pourraient être la cible du mécanisme de nonsense mediated decay (NMD) en charge d'éliminer les transcrits possédant un codon stop prématuré car l'introduction en phase de E1' entraîne la création d'un codon stop dans l'exon E1', situé en amont du codon STOP naturel de pVHL. Nos expériences ont mis en évidence qu'en présence d'un inhibiteur (puromycine) de la voie NMD<sup>221</sup> on observait une augmentation de l'expression des isoformes qui contiennent E1' (Figure 3C et D). Cette augmentation est observée dans les mêmes proportions chez les patients. L'expression que nous avons mesurée par RT-qPCR correspondrait donc à un taux résiduel d'isoformes qui ont échappé au NMD (saturation possible de la machinerie).

Nous avons ensuite réalisé une étude de l'expression protéique des différentes isoformes de pVHL par western blot à l'aide d'un anticorps dirigé contre l'exon 1 de VHL (décrit par Chesnel et al.<sup>59</sup>) (Figure 4A). Nous n'avons pas observé de bande endogène correspondant à la taille d'une hypothétique protéine X1, ni en condition normale, ni chez les patients surexprimant l'isoforme transcriptionnelle correspondante. Cependant, nous avons observé, dans toutes les lignées LLB de patients porteurs d'une mutation dans l'exon E1', une diminution de l'expression de l'ensemble des isoformes normales de pVHL (pVHL213 et pVHL160) (Figure 4 A et B).

Nous avons également réalisé une étude en minigène rapporteur des différentes mutations retrouvées dans l'exon cryptique E1' (Figure 4D). Cela nous a permis de mettre en évidence que les mutations dans E1' entraînent toutes une anomalie d'épissage avec une rétention de l'exon E1'. Comme dans le cas des mutations synonymes, les mutations à l'origine de la maladie de VHL entraînent une anomalie d'épissage plus importante que celles à l'origine de la polyglobulie.



RED CELLS, IRON, AND ERYTHROPOIESIS

## Identification of a new *VHL* exon and complex splicing alterations in familial erythrocytosis or von Hippel-Lindau disease

Marion Lenglet,<sup>1,3,\*</sup> Florence Robriquet,<sup>2,3,\*</sup> Klaus Schwarz,<sup>4,5</sup> Carne Camps,<sup>6,7</sup> Anne Couturier,<sup>8</sup> David Hoogewijs,<sup>9</sup> Alexandre Buffet,<sup>10-12</sup> Samantha J. L. Knight,<sup>6,7</sup> Sophie Gad,<sup>1,13</sup> Sophie Couvé,<sup>1,13</sup> Franck Chesnel,<sup>8</sup> Mathilde Pacault,<sup>2,14</sup> Pierre Lindenbaum,<sup>3</sup> Sylvie Job,<sup>15</sup> Solenne Dumont,<sup>2</sup> Thomas Besnard,<sup>3,14</sup> Marine Cornec,<sup>3</sup> Helene Dreau,<sup>16</sup> Melissa Pentony,<sup>6,7</sup> Erika Kvikstad,<sup>6,7</sup> Sophie Deveaux,<sup>17-20</sup> Nelly Burnichon,<sup>10-12,18,19,21</sup> Sophie Ferlicot,<sup>17,22</sup> Mathias Vilaine,<sup>2</sup> Jean-Michaël Mazzella,<sup>10-12,18,19,21</sup> Fabrice Airaud,<sup>14</sup> Céline Garrec,<sup>14</sup> Laurence Heidet,<sup>23</sup> Sabine Irtan,<sup>24</sup> Elpis Mantadakis,<sup>25</sup> Karim Bouchireb,<sup>23</sup> Klaus-Michael Debatin,<sup>26</sup> Richard Redon,<sup>3</sup> Stéphane Bezieau,<sup>3,14</sup> Brigitte Bressac-de Paillerets,<sup>13,27</sup> Bin Tean Teh,<sup>28</sup> François Girodon,<sup>29-31</sup> Maria-Luigia Randi,<sup>32</sup> Maria Caterina Putti,<sup>33</sup> Vincent Bours,<sup>34</sup> Richard Van Wijk,<sup>35</sup> Joachim R. Göthert,<sup>36</sup> Antonis Kattamis,<sup>37</sup> Nicolas Janin,<sup>38</sup> Celeste Bento,<sup>39</sup> Jenny C. Taylor,<sup>6,7</sup> Yannick Arlot-Bonnemains,<sup>8</sup> Stéphane Richard,<sup>1,13,17-20,†</sup> Anne-Paule Gimenez-Roqueplo,<sup>10-12,18,19,21,†</sup> Holger Carlo,<sup>26,‡</sup> and Betty Gardie<sup>1-3,31,‡</sup>

<sup>1</sup>École Pratique des Hautes Études, PSL Research University, Paris, France; <sup>2</sup>Centre de Recherche en Cancérologie et Immunologie Nantes-Angers, INSERM, Université de Nantes and Université d'Angers, Nantes, France; <sup>3</sup>L'Institut du Thorax, INSERM, Centre National de la Recherche Scientifique (CNRS), Université de Nantes, Nantes, France; <sup>4</sup>Institute for Transfusion Medicine, University of Ulm, Ulm, Germany; <sup>5</sup>Institute for Clinical Transfusion Medicine and Immunogenetics Ulm, German Red Cross Blood Service Baden-Württemberg-Hessen, Ulm, Germany; <sup>6</sup>Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; <sup>7</sup>Oxford NIHR Biomedical Research Centre, Oxford, United Kingdom; <sup>8</sup>Institut de Génétique et Développement de Rennes, UMR 6290, CNRS, Université de Rennes, Rennes, France; <sup>9</sup>Department of Medicine/Physiology, University of Fribourg, Fribourg, Switzerland; <sup>10</sup>Paris Cardiovascular Research Center, INSERM UMR 970, Hôpital Européen Georges Pompidou, Paris, France; <sup>11</sup>Faculté de Médecine, Université Paris Descartes, Paris, France; <sup>12</sup>Equipe Labellisée, Ligue Nationale Contre le Cancer, Paris, France; <sup>13</sup>Institut Gustave Roussy, INSERM UMR 1186, Université Paris-Saclay, Villejuif, France; <sup>14</sup>Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) de Nantes, Nantes, France; <sup>15</sup>Programme Cartes d'Identité des Tumeurs, Ligue Nationale Contre le Cancer, Paris, France; <sup>16</sup>Molecular Diagnostics Laboratories, Molecular Haematology Department, Oxford University Hospitals Trust, Oxford, United Kingdom; <sup>17</sup>Faculté de Médecine, Université Paris-Sud, Le Kremlin-Bicêtre, France; <sup>18</sup>Predispositions aux Tumeurs du Rein, Réseau Expert National pour Cancers Rares de l'Adulte, PREDIR labellisé par l'Institut National du Cancer (INCa), and <sup>19</sup>Maladie de VHL et Predispositions au Cancer du Rein, Réseau d'Oncogénétique, Institut National du Cancer, Le Kremlin-Bicêtre, France; <sup>20</sup>Service d'Urologie, Hôpital Bicêtre, Assistance Publique Hôpitaux de Paris (AP-HP), Le Kremlin-Bicêtre, France; <sup>21</sup>Service de Génétique, Hôpital Européen Georges Pompidou, AP-HP, Paris, France; <sup>22</sup>Pathology Department, Hôpitaux Universitaires Paris-Sud, AP-HP, Le Kremlin-Bicêtre, France; <sup>23</sup>Service de Néphrologie Pédiatrique, Centre de Référence des Maladies Rénales Hérititaires de l'Enfant et de l'Adulte, AP-HP, Paris, France; <sup>24</sup>Département de Chirurgie Pédiatrique, Hôpital Universitaire Necker-Enfants Malades, AP-HP, Université Paris Descartes-Sorbonne Paris Cité, Paris, France; <sup>25</sup>Faculty of Medicine Alexandroupolis, Democritus University of Thrace, Thrace, Greece; <sup>26</sup>Department of Pediatrics and Adolescent Medicine, University Medical Center Ulm, Ulm, Germany; <sup>27</sup>Département de Biologie et Pathologies Médicales, Université Paris-Saclay, Villejuif, France; <sup>28</sup>SingHealth/Duke-NUS Institute of Precision Medicine, National Heart Centre Singapore, Singapore; <sup>29</sup>Service d'Hématologie Biologique, Pôle Biologie, CHU Dijon, Dijon, France; <sup>30</sup>Equipe Protéines de Stress et Cancer, Lipides Nutrition Cancer, INSERM UMR 1231, "FCS Bourgogne Franche Comté, LipSTIC Labex, Dijon, France; <sup>31</sup>Laboratory of Excellence GR-Ex, Paris, France; <sup>32</sup>First Medical Clinic, Department of Medicine, and <sup>33</sup>Clinic of Pediatric Hemato-Oncology, Department of Woman's and Child's Health, University of Padua, Padua, Italy; <sup>34</sup>Service de Génétique Humaine du CHU Sart Tilman, Liège, Belgium; <sup>35</sup>Department of Clinical Chemistry and Haematology, University Medical Center Utrecht, Utrecht, The Netherlands; <sup>36</sup>Department of Hematology, West German Cancer Center, University Hospital Essen, Essen, Germany; <sup>37</sup>First Department of Pediatrics, National and Kapodistrian University of Athens, Athens, Greece; <sup>38</sup>Centre de Génétique Humaine, Cliniques Universitaires Saint-Luc, Brussels, Belgium; and <sup>39</sup>Department of Hematology, Centro Hospitalar e Universitario de Coimbra, Coimbra, Portugal

KEY POINTS

- Mutations in a *VHL* cryptic exon may be found in patients with familial erythrocytosis or *VHL* disease.
- Synonymous mutations in *VHL* exon 2 may induce exon skipping and cause familial erythrocytosis or *VHL* disease.

Chuvash polycythemia is an autosomal recessive form of erythrocytosis associated with a homozygous p.Arg200Trp mutation in the von Hippel-Lindau (*VHL*) gene. Since this discovery, additional *VHL* mutations have been identified in patients with congenital erythrocytosis, in a homozygous or compound-heterozygous state. *VHL* is a major tumor suppressor gene, mutations in which were first described in patients presenting with *VHL* disease, which is characterized by the development of highly vascularized tumors. Here, we identify a new *VHL* cryptic exon (termed E1') deep in intron 1 that is naturally expressed in many tissues. More importantly, we identify mutations in E1' in 7 families with erythrocytosis (1 homozygous case and 6 compound-heterozygous cases with a mutation in E1' in addition to a mutation in *VHL* coding sequences) and in 1 large family with typical *VHL* disease but without any alteration in the other *VHL* exons. In this study, we show that the mutations induced a dysregulation of *VHL* splicing with excessive retention of E1' and were associated with a downregulation of *VHL* protein expression. In addition, we demonstrate a pathogenic role for synonymous mutations in *VHL* exon 2 that altered splicing through E2-skipping in 5 families with erythrocytosis or *VHL* disease. In all the studied cases, the mutations differentially affected splicing, correlating with phenotype severity. This study demonstrates that cryptic exon retention and

## Introduction

Congenital erythrocytosis represents a heterogeneous group of rare disorders. Genetic changes affecting all parts of the regulatory pathway of erythropoiesis, including oxygen sensing, erythropoietin sensitivity, and hemoglobin oxygen affinity, have been described in patients with congenital erythrocytosis. The detection of underlying genetic changes in patients with presumed hematological pathology may have important implications for adequate clinical management. However, even with the use of next-generation sequencing panel diagnostics, the underlying genetic cause of presumed congenital erythrocytosis has been identified in fewer than one third of patients in most published cohorts.

The molecular basis of von Hippel-Lindau (*VHL*)-related congenital erythrocytosis was first described in the autonomous Russian Republic of Chuvashia, where this condition is an endemic disorder.<sup>1</sup> Chuvash polycythemia is frequently associated with rubor, vertebral hemangiomas, varicose veins, and low blood pressure. Chuvash patients have reduced survival rates associated with a higher prevalence of arterial and venous thromboses and pulmonary hypertension in addition to hemorrhagic events.<sup>2</sup>

Chuvash polycythemia arises from a homozygous c.598C>T p.Arg200Trp (R200W) mutation in the *VHL* gene. This specific *VHL* R200W mutation has also been identified in combination with other *VHL* mutations (compound heterozygosity), and other missense *VHL* mutations in both alleles have been described in patients with congenital erythrocytosis.<sup>3,4</sup> Interestingly, it has been described in some unexplained cases of patients with erythrocytosis carrying only one heterozygous mutation.<sup>4,6</sup>

*VHL* is located on 3p25-26 and has been reported to contain 3 exons (E1, E2, and E3). The commonly described *VHL* transcript contains the 3 spliced exons that encode a 213 amino acid (aa) protein (pVHL213; also termed pVHL30) and a smaller isoform (pVHL160 or pVHL19) initiated from an in-frame internal translation start site.<sup>7</sup> A naturally occurring splice variant, expressed at low levels in some tissues, comprises E1 directly spliced to E3 and is translated into a protein product termed pVHL172 (pVHLΔE2), the functions of which are still under investigation.<sup>8-12</sup> pVHL213 and pVHL160 are involved in a variety of functions, the most studied being the regulation of the cellular oxygen-sensing pathway. The main player of this pathway is the hypoxia inducible factor (HIF). Under normal oxygen supply, the  $\alpha$  subunits of HIF (HIF-1 $\alpha$ , -2 $\alpha$ , and -3 $\alpha$ ) are hydroxylated by the prolyl-4 hydroxylase domain enzymes (PHD1, -2, and -3) and subsequently targeted by pVHL, a subunit of an E3 ubiquitin-ligase complex that promotes HIF- $\alpha$  ubiquitination and subsequent proteasomal degradation.<sup>13,14</sup> Under hypoxic conditions or when *VHL* is mutated, HIF- $\alpha$  remains stable and heterodimerizes with HIF- $\beta$ , constituting a functional HIF. HIF transcriptionally activates a variety of genes involved in adaptation to reduced oxygen supply (eg, erythropoiesis, angiogenesis, metabolism, and

cell survival). Dysregulation of the hypoxia pathway<sup>13,14</sup> is central to the development of erythrocytosis<sup>1,4</sup> (via upregulation of erythropoietin, a HIF-2 $\alpha$  target gene), but also in the development of tumors.<sup>15</sup> Indeed, *VHL* is a tumor suppressor gene, heterozygous mutations of which are associated with *VHL* disease (Figure 1A).<sup>16,17</sup> *VHL* disease, described in 1936, is an autosomal dominant disorder with high penetrance characterized by the development of highly vascularized tumors like central nervous system and retinal hemangioblastomas, pancreatic neuroendocrine tumors, pheochromocytomas, and clear cell renal cell carcinomas.<sup>18-20</sup>

In patients carrying *VHL* mutations, the precise mechanistic aspects that underpin the different phenotypes remain obscure. Although most patients carry mutations in the *VHL* gene that induce a partial or complete loss of protein function, some cases remain unsolved. Indeed, some patients with erythrocytosis have been found to be heterozygous rather than homozygous for the expected alteration<sup>4,6</sup> or carry homozygous synonymous mutations that leave the amino acid sequence intact. In addition, some patients present with *VHL* disease in the absence of identified mutations or deletions in *VHL* or carry heterozygous synonymous mutations. Here, we report an investigation of 12 families linked to unexplained disease, including 9 families with erythrocytosis and 3 with *VHL* disease (Figure 1A). This study led to the discovery of a novel cryptic exon in the *VHL* gene and complex regulation of *VHL* splicing.

## Methods

Complete materials and methods are detailed in the supplemental Material and methods (available on the *Blood* Web site).

### Study approval

Informed consent for medical diagnosis and research was obtained from the patients and their relatives. This study was agreed by the Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale (French ethical committee) Paris-Sud at Bicêtre Hospital.

### Sanger sequencing

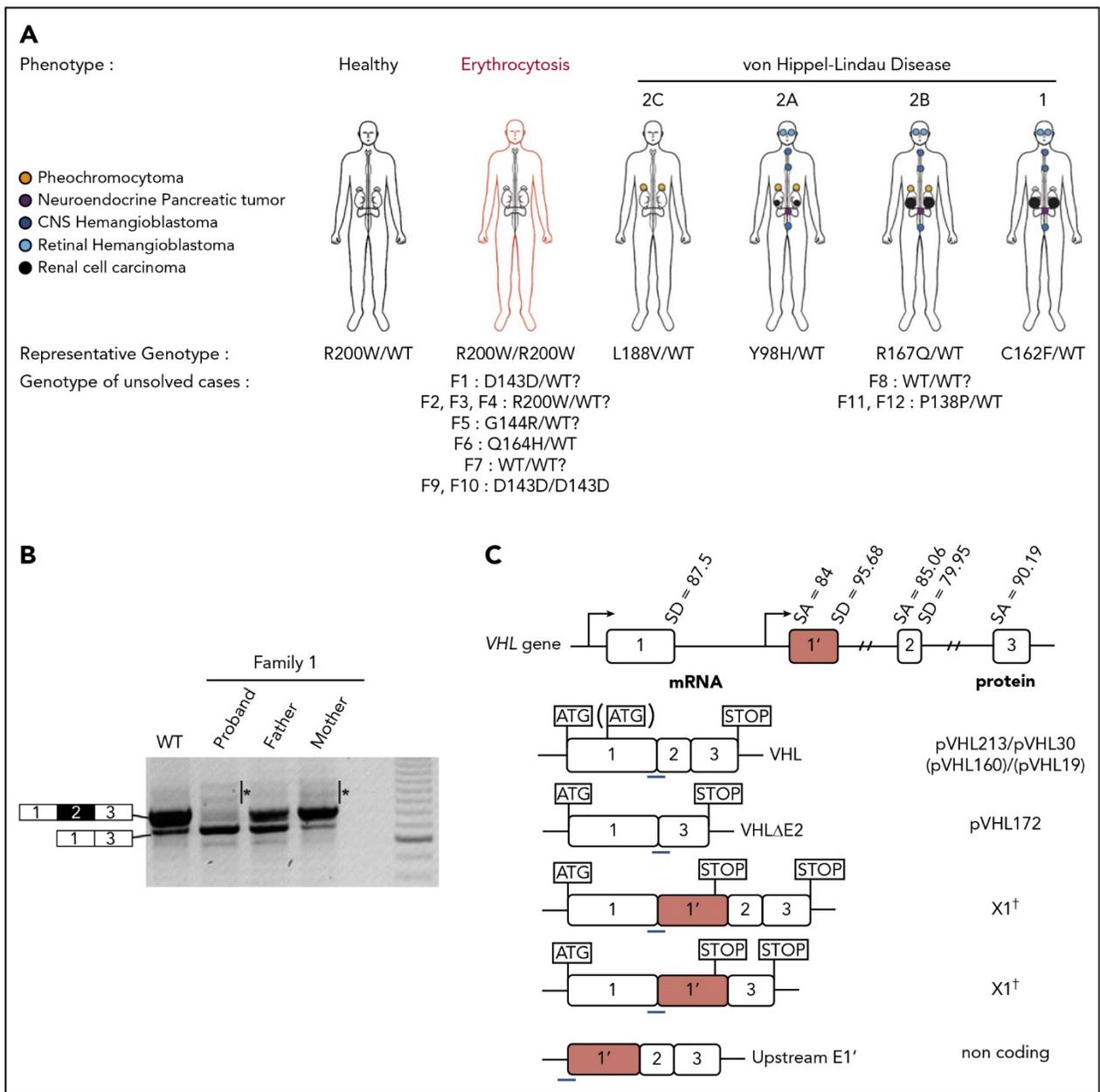
Exons and exon-intron junctions of the *VHL* gene were sequenced from DNA extracted from whole blood, as previously described.<sup>21</sup>

### WGS

Whole-genome sequencing (WGS; for families F2, F3, and F7) was performed at the clinically accredited Molecular Diagnostics Laboratory at the John Radcliffe Hospital using the Hi-Seq4000 platform (Illumina Inc., San Diego, CA) in high-throughput mode.<sup>22,23</sup> Analysis of single-nucleotide variants, short insertions/deletions, and copy number variants was conducted and is explained in detail in the supplemental Methods.

### Transcript detection and quantification

After reverse-transcription reactions (ThermoScientific), exon-specific polymerase chain reaction (PCR) was performed using



**Figure 1. Clinical manifestations of patients carrying VHL mutations and identification of a new VHL spliced isoform containing a cryptic exon.** (A) Mutations in the VHL gene predispose to different phenotypes. VHL disease is characterized by the development of central nervous system (CNS) and retinal hemangioblastomas, neuroendocrine pancreatic tumors, pheochromocytomas, and clear cell renal cell carcinomas. Chuvash polycythemia (erythrocytosis) is characterized by elevated red blood cell numbers. This study describes families with typical VHL-related phenotypes associated with an unexpected VHL status (ie, either synonymous mutations or no alterations in VHL). (B) Reverse transcription polymerase chain reaction (RT-PCR) using primers specific for E1 and E3 was performed on messenger RNA (mRNA) extracted from lymphoblastoid cell lines (LCLs) established from controls and patients of family 1 (F1). (C) Schematic representation of the VHL gene and its products. The different VHL exons are represented on a scale: E1, 340 bp from the ATG initiation codon; E1', 259 bp; E2, 123 bp; E3, 179 bp to the Stop termination codon. The full-length VHL mRNA isoform encodes pVHL213 (also named pVHL30). VHL E1 contains an internal translation initiation codon that initiates the production of pVHL160 protein (pVHL19). The isoform containing E1 spliced in phase with E3 encodes pVHL172 that lacks E2. The isoforms containing E1 spliced with E1' may theoretically encode a protein termed X1 of 193 aa (114 aa encoded by E1, and 79 aa encoded by E1'). Consensus values of donor (SD) and acceptor (SA) splice site sequences are indicated above the VHL gene, as calculated by the human splicing finder in silico tool. Horizontal blue lines indicate the location of probes used in TaqMan assays. \*Denotes larger fragments that were subsequently cloned and sequenced. †The isoforms containing E1' spliced with VHL exons were identified by cloning and sequencing in the laboratory but were described later by the National Center for Biotechnology Information as transcripts able to produce a protein termed X1. WT, wild type.

primers localized to flanking E1 and E3 exons. TaqMan real-time PCR were performed on 20 ng of complementary DNA (cDNA) with the quantitative PCR Mastermix (Eurogentec). Quantification of RPLP0 transcripts was used as internal control. The thresholds were determined using dilutions of plasmids containing coding sequences of each gene.

### RNA sequencing

Library construction was performed with SureSelect Strand-Specific RNA Library Prep for the Illumina Multiplexed kit (Agilent Technologies). After purification (Macherey-Nagel), the fragment size of libraries was controlled using the 2200 TapeStation system (Agilent Technologies). Ten pM of each

**Table 1. Description of variants identified in VHL E1' and VHL E2 with associated clinical manifestations**

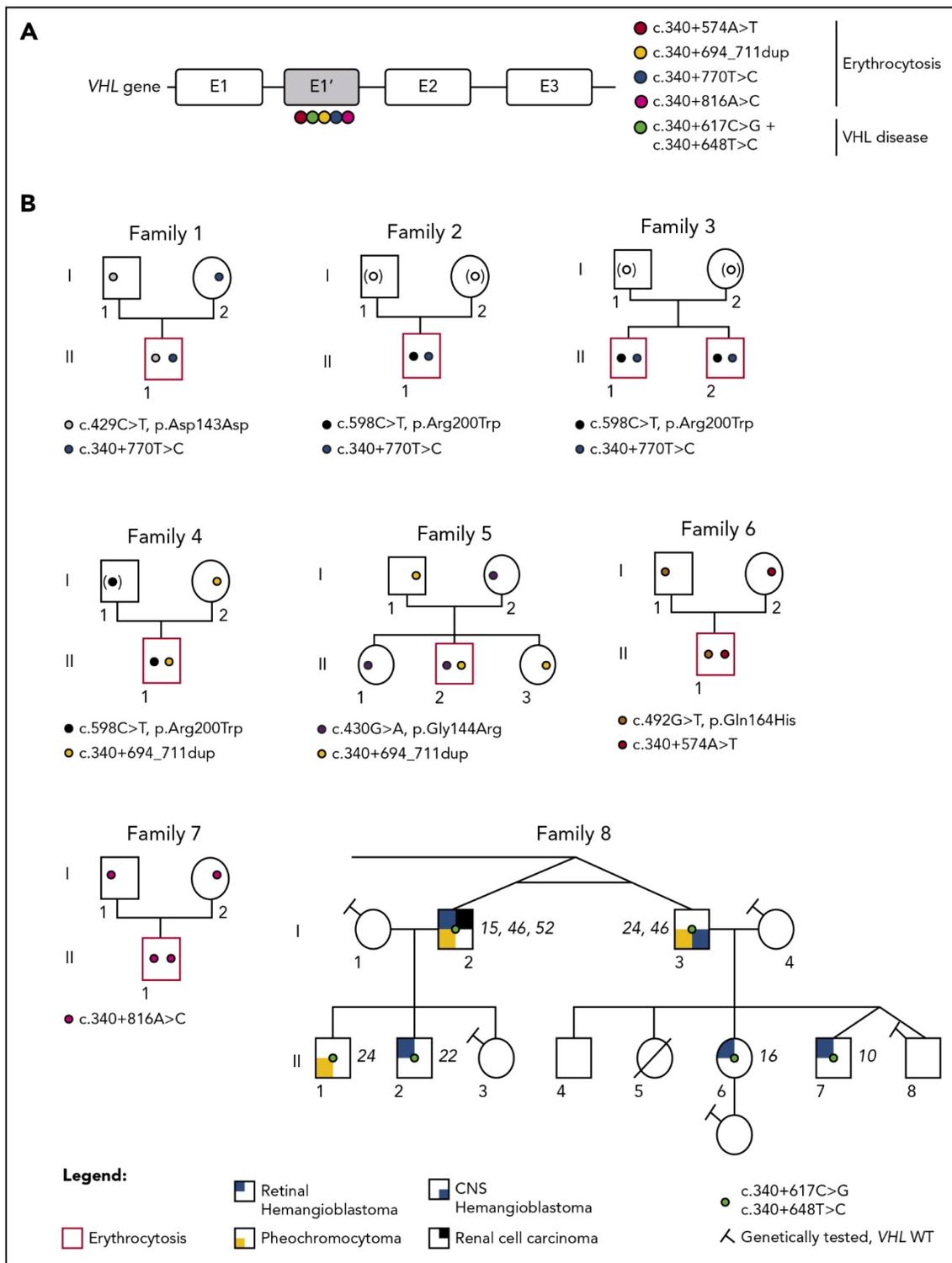
Family no., patient	Nucleotide variant on allele 1/allele 2	Impact on protein	Year of birth/age at diagnosis	Sex	Hb, g/dL	Ht, %	Red cells, ×10 <sup>9</sup> /L	EPO, mU/mL	Phenotype	Other mutation
F1, II.1	c.429C>T/c.340+770T>C	VHL p.Asp143Asp/X1 p.Ser179Pro?	1991/15 y	M	16.6	57	6.9	163	Erythrocytosis, kidney ischemic infarct	—
F2, II.1	c.598C>T/c.340+770T>C	VHL p.Arg200Trp/X1 p.Ser179Pro?	1957/21 y	M	19.9	67		22.2	Erythrocytosis	—
F3, II.1	c.598C>T/c.340+770T>C	VHL p.Arg200Trp/X1 p.Ser179Pro?	2003/4 mo	M	17.9	54	7.5	60	Erythrocytosis	—
F3, II.2	c.598C>T/c.340+770T>C	VHL p.Arg200Trp/X1 p.Ser179Pro?	2005/6 mo	M	14.6	45			Erythrocytosis	—
F4, II.1	c.598C>T/c.340+694_711dup	VHL p.Arg200Trp/X1 p.Trp159X?	1990/7 y	M	20.6	64	8.25	49.9	Erythrocytosis, deep vein thrombosis, intracerebral hemorrhage	FV Leiden
F5, II.2	c.430G>A/c.340+694_711dup	VHL p.Gly144Arg/X1 p.Trp159X?	1975/13 y	M	16.2	60	5.9		Erythrocytosis	—
F6, II.1	c.492G>T/c.340+574A>T*	VHL p.Gln164His/SA	2014/6 mo	M	15.6	49	6.9	1167	Splenomegaly	—
F7	c.340+816A>C/c.340+816A>C	X1 p.*194Serext*24/X1 p.*194Serext*24	7 y	M	20	64	8.09	33	Erythrocytosis	—
F8	c.340+617C>G + c.340+648T>C†/WT	X1 p.Leu128Val + X1 p.Leu138Pro/WT			N	N	N	N	VHL disease	—
F9, II.1	c.429C>T/c.429C>T	VHL p.Asp143Asp/VHL p.Asp143Asp	1997/5 mo	M	16.2	49.3	5.2	186	Erythrocytosis	Hb sitia
F10, II.1	c.429C>T/c.429C>T	VHL p.Asp143Asp/VHL p.Asp143Asp	2002/4.5 y	F	22.5	64-77.4	12.6	264	Erythrocytosis, splenomegaly	β thalassemia
F11	c.414A>G/WT	VHL p.Pro138Pro/WT			N	N	N	N	VHL disease	—
F12	c.414A>G/WT	VHL p.Pro138Pro/WT			N	N	N	N	VHL disease	—

The second column indicates the position of the nucleotide variants identified in the VHL gene regarding the current nomenclature (sequence encoded by the VHL E1-E2-E3). The third column indicates the impact either on the VHL protein (for variants located in E1, E2, or E3) or on the potential X1 protein, if encoded by E1-E1' (for variants located in E1). Normal values correspond to: Hb: M, 13 to 18 g/dL and F, 12 to 15 g/dL; Ht: M, 40% to 52% and F, 37% to 47%; red cells: M, 4.2 to 5.7 × 10<sup>9</sup>/L and F, 4.2 to 5.2 × 10<sup>9</sup>/L; EPO: 5 to 25 mU/mL.

EPO, erythropoietin; Hb, hemoglobin; Ht, hematocrit; N, normal value.

\*The nucleotide change was reported as rs982745672 with a global minor allele frequency that corresponds to T = 0.00007/2 (TOPMed).

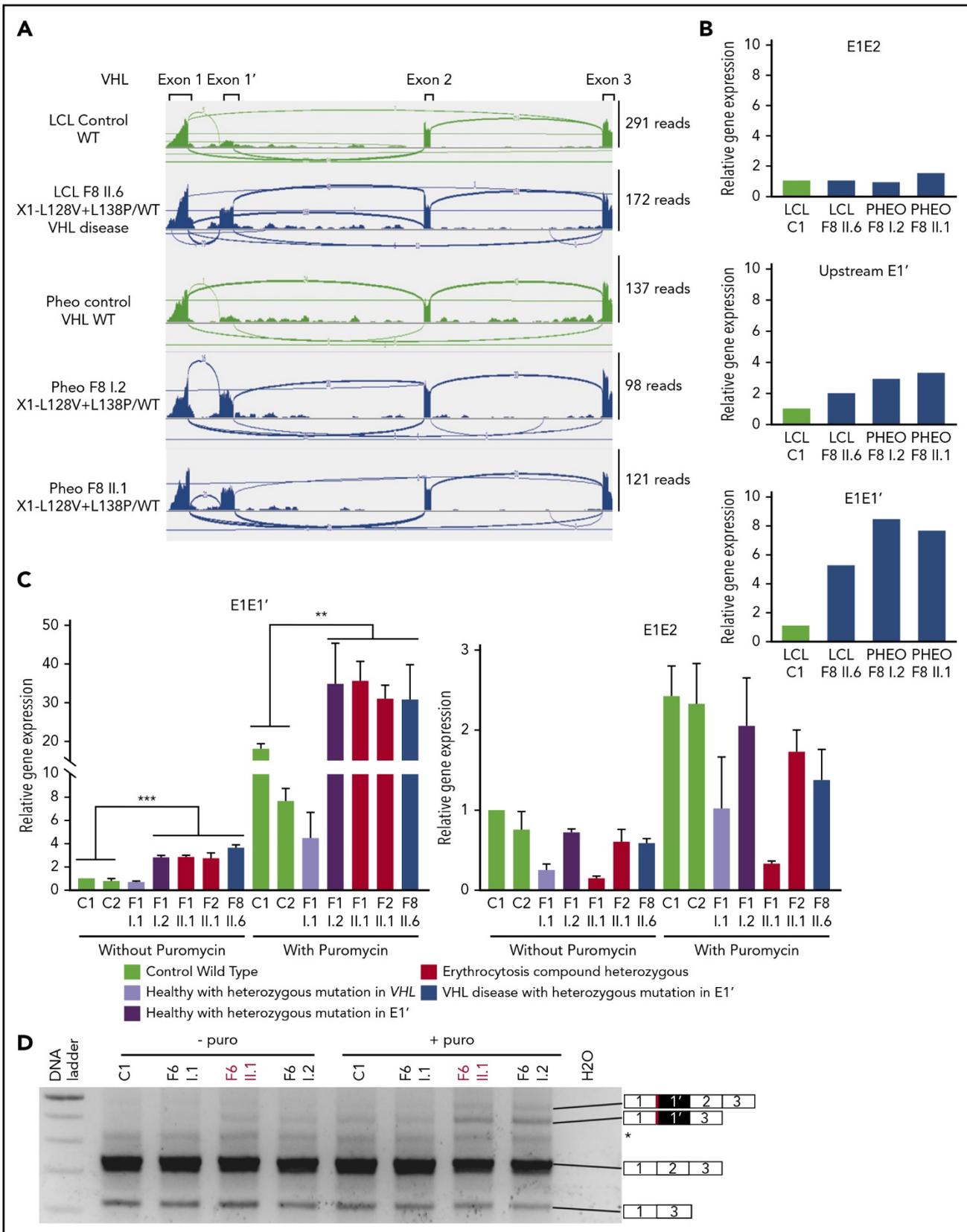
†The nucleotide change was reported as rs73024533 with a global minor allele frequency that corresponds to C = 0.0026/13 (1000 Genomes) and C = 0.0050/147 (TOPMed).



**Figure 2. Identification of mutations in the new VHL cryptic exon in 7 families with erythrocytosis and a large family with VHL disease.** (A) Schematic representation of the VHL gene and location of the identified mutations in the new VHL cryptic exon E1'. (B) Pedigree of families with erythrocytosis or VHL disease. The genotypes were elucidated by sequencing both parents and proband (F1, F5, and F6), deduced by sequencing of 1 parent and proband (F4; the mutation deduced in brackets), or deduced from allele cloning of proband carrying the conserved Chuvash mutation and core haplotype (F2 and F3), confirming the transmission of the mutations by 1 of each parent (for F2 and F3, the identity of the transmitting parent being unknown, the mutation is represented by a white circle in brackets). The genotype of parents from F7 was elucidated from WGS data. The numbers in italics (F8) indicate the age of the patient at tumor diagnosis. CNS, central nervous system.

library was pooled and prepared according to the denaturing and diluting library protocol for HiSeq and GAIx (Illumina) for cluster generation on the cBot system. Paired-end sequencing

was carried out in a single lane on the HiSeq 2500 system (Illumina). Processing of reads is detailed in the supplemental Methods.



**Figure 3. Expression study of the new VHL transcript isoforms in patient cells.** (A) Sashimi plots from RNA-seq data obtained from samples (LCLs or pheochromocytomas [Pheo]) of 3 patients from F8. The positions of the different VHL exons are indicated, with the maximal number of reads indicated at the right. Splice junctions are denoted by the horizontal links, with details provided in supplemental Figure 8. (B) TaqMan quantification of the different VHL isoforms from samples of F8 performed using probes specific to VHL E1-E2, the translated sequence upstream E1', or E1-E1' junctions. Relative gene expression was normalized to LCL control (C1) fixed at 1 (mean results of technical duplicates). (C) TaqMan quantification of the different VHL isoforms in LCLs (established from 2 independent controls and from patients of F1, F2, and F8) cultured in the absence

## Reporter assays

The plasmid encoding the WT hemagglutinin-tagged VHL was kindly provided by William G. Kaelin Jr. An expression plasmid for the hypothetical X1 protein was constructed by subcloning the coding sequence after its synthesis (Life Technologies). Mutations of interest were introduced by site-directed mutagenesis (New England Biolabs); 786-O cells were transfected (Ozyme) with constructs encoding WT or mutant proteins, firefly luciferase under the control of hypoxia response elements, and *Renilla* luciferase for normalization.<sup>24</sup> After 24 hours, luciferase assays were performed using the Dual-Luciferase Reporter Assay System (Promega).

## Western blotting

Cell lysates from the luciferase reporter assay were loaded into a Bis-Tris Mini Gel (4% to 12%; Invitrogen). After transfer, the membrane (GE Healthcare) was subsequently incubated with a mouse anti-hemagglutinin antibody (BioLegend) and then a goat anti-mouse horseradish peroxidase-conjugated antibody (Jackson Immuno Research). Western blot using the mouse monoclonal antibody JD-1956 (patent #14305925.1-1402-2014; Centre National de la Recherche Scientifique Etablissement Français du Sang) raised against human VHL was performed as described.<sup>8</sup>

## Minigene experiments

Minigene constructs were prepared in pCas2 plasmid, containing 2 artificial exons A and B (as described by Gaildrat et al<sup>25</sup>), between which VHL E1' or VHL E2 with intronic flanking sequences were cloned. Cells were transfected (Ozyme) or nucleofected (Lonza). RNA was extracted 24 hours after transfection and reverse transcribed. PCR amplification was performed using the PCR GoTaqQ2 kit (Promega) with primers against artificial (A and B) exons. PCR products were resolved in a 2% agarose gel.

## Results

### Identification of new VHL spliced isoforms containing a cryptic exon

We first focused our study on a patient with erythrocytosis in whom a synonymous VHL c.429C>T p.Asp143Asp (D143D) mutation in the heterozygous state had previously been identified (F1; Table 1). No other mutations in the 3 VHL canonical exons had been identified in genomic DNA of this patient. RT-PCR using primers in E1 and E3 was performed using mRNA samples extracted from LCLs established from different family members. The results showed a strong decrease of the E1E2E3 isoform and an upregulation of the E1E3 isoform (Figure 1B) compared with WT LCLs. Some minor extra fragments of larger size were observed in this patient and his mother's sample. Subcloning and sequencing of these fragments allowed us to identify new VHL transcripts that contained intronic sequences. This intronic sequence, which we termed the E1' cryptic exon, was spliced to E1 at its 5' end and to either E2 or E3 at its 3' terminus (Figure 1C). We show that these isoforms were expressed in a variety of tissues and cell lines (supplemental Figure 1). Their translation may theoretically lead to

the production of a protein of 193 aa that contains the first 114 aa encoded by E1<sup>26</sup> and additional 79 aa of unknown function encoded by E1'. During the course of our study, data from an automated computational analysis for an isoform containing the E1' cryptic exon (E1E1'E2 isoform) were deposited in the National Center for Biotechnology Information. This isoform was predicted to encode a protein of 193 aa, named X1 (XP\_011532380.1), with the same sequence (114 aa from E1 in addition to 79 aa encoded by E1') (Figure 1C). The analysis of this region revealed strong conservation in primates, but a moderate to low conservation in more distant species (supplemental Figure 2A-C); notably, the splice sites were identical to canonical sites (ttcag/TC, AG/gtaag) and were highly conserved. In silico analysis of the donor and acceptor splice sites of E1' showed similar consensus values compared with other VHL exons (Figure 1C, upper panel). The capacity to translate a potential X1 protein was only conserved in higher primates (supplemental Figure 2D). The deposited sequence was removed and replaced with a noncoding isoform containing E1' spliced with VHL E2 and E3 (ENST00000477538.1; Figure 1C, isoform on the bottom). This isoform may be initiated by an alternative promoter. Indeed, the sequence located at the 5' end of E1' represents a transcriptionally active region, as illustrated by epigenetic marks (supplemental Figure 3). We confirmed the expression of an additional transcript initiated from the upstream region of E1' (that we termed upstream E1') in different tissues and cell lines (supplemental Figure 1).

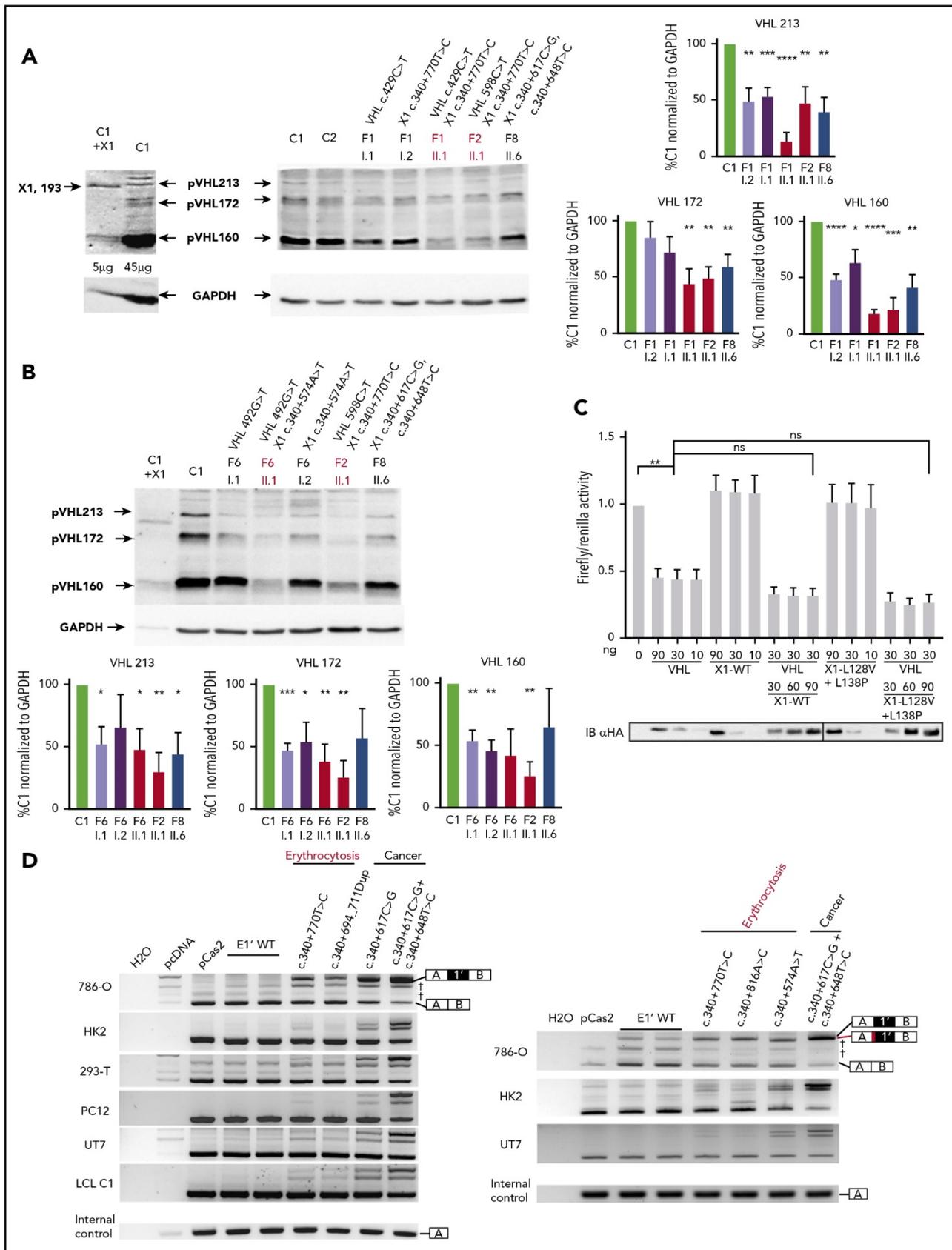
### New E1' cryptic exon is mutated in patients with erythrocytosis or VHL disease

Sanger sequencing of this new cryptic exon in the proband (F1-II.1) identified a variant that had not been reported in databases: c.340+770T>C (Figure 2B; Table 1; supplemental Figure 4). Sequencing of the germline DNA of the mother revealed the same variant, indicating that the proband F1-II.1 was compound heterozygous. This result prompted us to sequence additional patients with erythrocytosis described as heterozygous for VHL mutations.

We investigated patients with Chuvash polycythemia for whom VHL R200W had been found in a heterozygous state. Sanger sequencing identified the identical intronic variant c.340+770T>C in a singleton (F2)<sup>5</sup> and 2 affected brothers (F3; Figure 2B). In addition, we identified a duplication c.340+694\_711dup in the proband of F4 and F5<sup>6</sup> previously described as VHL R200W/WT and VHL G144R/WT, respectively (Figure 2B). In F6, previously diagnosed as VHL Q164H/WT, we identified a genetic variant c.340+574A>T that altered the consensus SA site ag/TC of E1'. This variant is described as a rare polymorphism (rs98274567) in the National Center for Biotechnology Information database (Table 1).

Because biological material from the parents of F2 and F3 was not available, we cloned the intronic region containing the E1' variant and the Chuvash core haplotype single-nucleotide polymorphism rs779808 associated with the VHL R200W mutation (described<sup>27</sup> to be located downstream of E1', in position

**Figure 3 (continued)** or presence of puromycin (puro), an inhibitor of nonsense-mediated mRNA decay. TaqMan probes were specific to the VHL E1-E1' or E1-E2 junction. The graph resumes experiments performed on LCLs cultured independently 3 times and quantified in duplicate. Data are represented as the mean  $\pm$  standard error of the mean. (D) RT-PCR using primers specific for E1 and E3 was performed on mRNA extracted from LCLs established from controls and patients of F6 (carrying the mutation c.340+574A>T that targets the SA site of E1'). Patient with erythrocytosis is indicated in red. On the right, the spliced isoforms are schematically represented. \*\**P* < .05, \*\*\**P* < .01 based on Student *t* test. \*Denotes larger fragments that contained E1' spliced with E1, but with 15 nucleotides deleted (represented in red) by the use of an alternative SA site (sequences of the cloned bands are detailed in supplemental Figure 10).



**Figure 4. Functional studies of mutations in VHL E1'.** (A-B) Immunoblot analysis of patient LCLs. A representative immunoblot (A) and quantification of 3 different immunoblots (B) are presented. Relative gene expression was normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression, and results obtained with LCL control (C1) were fixed at 100%. Data are represented as the mean  $\pm$  standard error of the mean. A VHL antibody that recognized VHL E1 was able to detect the hypothetical

c.340+1150). Segregation analysis (F4, F5, and F6) or Chuvash core haplotype analysis (F2 and F3) demonstrated that all patients were compound heterozygous, with 1 mutation being inherited from each parent (supplemental Figure 5).

In a parallel independent study, WGS was used to investigate patients with Chuvash polycythemia, heterozygous for the *VHL* R200W mutation (F2 and F3). The intronic variant c.340+770T>C was the only rare variant identified in the *VHL* gene in these patients. Other WGS filtering strategies (supplemental Methods) did not identify any additional significant mutations in biologically relevant genes (supplemental Figure 6). This study also conducted WGS for a trio with congenital erythrocytosis for which prior whole-exome sequencing had not identified any mutations (F7). No rare biologically relevant variants were identified using the filtering strategy described in supplemental Methods. However, further inspection of the new cryptic exon led to the identification of a c.340+816A>C variant in the homozygous state in the proband, with both parents being heterozygous for this mutation.

In addition to these patients with erythrocytosis, in which no variants in *VHL* were initially detected, screening of patients with VHL disease may also fail to detect mutations in canonical *VHL* exons. An example of such a family (F8) with hereditary hemangioblastoma, clear cell renal cell carcinoma, and pheochromocytoma was studied. Microsatellite analysis demonstrated a cosegregation of markers surrounding the *VHL* region with the disease (supplemental Figure 7). Sequencing of the new E1' cryptic exon in this family identified 2 heterozygous variants in E1': a previously unreported c.340+617C>G variant and a c.340+648T>C variant described as a rare polymorphism in databases (Table 1). These variants segregated in 6 patients who developed the disease and were absent in 4 healthy tested descendants, indicating their presence on a single disease-associated allele (Figure 2B). Sequencing of tumor DNA did not display loss of heterozygosity, suggesting that a WT *VHL* deletion (as observed in classic VHL disease) may not be a prerequisite in cells of patients with this specific *VHL* genotype (supplemental Figure 4).

### Expression study of the new *VHL* isoforms in patients' cells

RNA sequencing (RNA-seq) analysis of patients' LCLs and tumors demonstrated that genomic variants in E1' were associated with an upregulation of transcripts containing the E1' cryptic exon compared with controls (Figure 3A; supplemental Figure 8). In addition, the mutated allele was preferentially expressed (>70%), suggesting a causal relationship between the genetic variants and E1' retention (supplemental Figure 8C).

We performed quantitative RT-PCR using TaqMan probes specific for the different splicing junctions or for the region upstream of E1' (location of probes shown in Figure 1C), which showed a strong upregulation of a unique isoform containing E1' spliced with E1 in LCLs and tumors from patients of F8 compared with controls (Figure 3B). The study therefore focused on the isoform E1-E1'. Transcripts that retained E1' spliced with E1 contained (in-frame with E1) a premature termination codon (Figure 1C) and could likely be targeted for degradation according to nonsense-mediated mRNA decay (NMD) mechanisms. We therefore investigated the expression of transcripts that contained E1 spliced with E1' in the absence or presence of puromycin, an inhibitor of NMD. Without puromycin, higher levels of expression were seen in samples with mutated E1' (Figure 3C) compared with samples containing WT E1'. Exposure to puromycin resulted in a profound induction of isoforms containing E1 spliced with E1' vs other isoforms (Figure 3C; supplemental Figure 9). These results indicated that isoforms with E1-E1' junction were degraded by NMD and may have failed to produce proteins. We performed RT-PCR on LCL samples of F6 carrying the mutation in the SA site, using primers in *VHL* exons E1 and E3. We observed fragments of larger size in samples from patients carrying the mutation, which were highly visible in the presence of puromycin (Figure 3D). Cloning and sequencing of these fragments demonstrated a dysregulated splicing of E1' with the use of an alternative SA site located 15 nucleotides downstream (supplemental Figure 10).

To measure the impact of the splicing dysregulation on pVHL expression, we performed an immunoblot with an antibody able to recognize the amino acids encoded by E1 and therefore able to detect the different pVHL isoforms.<sup>8</sup> The antibody detected overexpressed exogenous X1 protein after transfection of the plasmid encoding X1 (Figure 4A). However, it failed to detect the endogenous protein, even in patients' samples that overexpressed the mRNA containing E1-E1'. Instead, the immunoblots showed a lower expression of all the VHL protein isoforms in patients with mutated E1' (Figures 4A-B).

### Functional characterization of the mutated *VHL* E1'

Because the hypoxia pathway represents the major pathway involved in the genesis of secondary erythrocytosis and VHL disease,<sup>13-15,28</sup> we performed functional studies of the hypothetical WT or mutated X1 proteins using hypoxia response element-dependent reporter assays. These functional tests failed to reveal any substantial effects of X1 on this pathway, either alone or in competition with pVHL (Figure 4C; supplemental Figure 11). We next focused our study on a potential impact of the E1' variants on splicing by performing splicing

**Figure 4 (continued)** X1 protein. In the most left lane, control LCLs were transfected with an expression vector containing the coding sequence for X1; 5 µg of proteins were blotted vs 45 µg for other samples to avoid signal saturation with overexpressed X1. Patients with erythrocytosis are indicated in red. (C) Functional hypoxia response element (HRE)-dependent reporter assays were performed in 786-O cells (ie, *VHL*<sup>-</sup> cells that constitutively express HIF-2α). The results are expressed as relative firefly luciferase activity with *Renilla* luciferase as an internal control; 1.0 unit denotes the basal activity of endogenous HIF-2α using the HRE luciferase reporter plasmid. The ability of WT and mutated X1 to downregulate firefly luciferase activity (related to HIF activity) was compared with pVHL and in competition with pVHL. An immunoblot using an antibody specific to the hemagglutinin tag (HA) was used to detect HA-VHL and HA-X1. X1-L128V\*L138P corresponds to a potential impact of the c.340+648T>C and c.340+617C>G mutations on the hypothetical X1 protein. Three independent experiments were performed. (D) Characterization of *VHL* E1' retention by the minigene experiment (representative picture of agarose gel; n = 3). RT-PCR was performed on mRNA obtained from cell lines transfected with a minigene construct containing *VHL* E1' (WT or mutated) flanked by large intronic sequences cloned between the *SERPINC1* exons (exons A and B, targeted by the RT-PCR primers). The plasmids were transfected, and the expression of the spliced chimeric transcripts (containing EA and EB ± E1') was analyzed. Two WT constructs containing E1' were used; 1 contained the single-nucleotide polymorphism c.340+1150T>C (rs779808). Bands corresponding to EA and EB spliced together or with *VHL* E1' are indicated on the right. \*P < .05, \*\*P < .01, \*\*\*P < .001, \*\*\*\*P < .0001 based on Student t test. †Corresponds to unspecific bands verified by sequencing. The minigene experiment performed with the construction carrying the mutation c.340+574A>T (that targets the SA site of E1' in F6) confirmed the use of an alternative SA site (right panel) with the deletion of 15 nucleotides (represented in red). ns, not significant.



reporter minigene assays in various cell lines. The experiments showed that splicing of WT E1' was barely detected (Figure 4D). Interestingly, higher molecular weight bands corresponding to the expected spliced isoforms containing E1' appeared in the presence of the mutations. Cloning and sequencing of the isoforms confirmed a retention of E1' at the expected splicing sites. The level of expression of this upper isoform, specific to E1' inclusion, was higher for mutations associated with cancers than for those with erythrocytosis, independently of the cell lines. Notably, the combination of both variants associated with cancer had a more pronounced effect than each variant individually (the single-nucleotide polymorphism showing very low effect), suggesting a synergistic effect of the variants on splicing (supplemental Figure 12). In silico analysis of c.340+617C>G consistently predicted a severe alteration of splicing by the creation of an exonic splicing enhancer site (supplemental Figure 13).

The minigene experiment performed with mutated E1' at the SA site (F6) confirmed the inclusion of E1' during splicing (Figure 4D, right panel), using the same downstream alternative SA site identified in mRNA extracted from patients' LCLs (supplemental Figure 10).

### Synonymous mutations in *VHL* E2 induce exon skipping

We then focused our study on the synonymous heterozygous D143D mutation in E2, identified in the proband of F1. This mutation was also identified in the homozygous state in 2 patients with erythrocytosis (F9 and F10; Figure 5A; Table 1). The mRNA extracted from LCLs established from the different members of these 2 families was reverse transcribed and sequenced. Comparison of chromatograms obtained by sequencing of DNA vs cDNA displayed a weaker peak of the mutated allele in cDNA, reflecting a decreased expression of the mRNA transcripts carrying the mutation.

A different synonymous mutation in E2, c.414A>G p.Pro138Pro (P138P), was identified in 2 families (F11 and F12) with *VHL* disease (Figure 5B). This heterozygous mutation segregated with the disease in 3 generations. Sequencing of DNA extracted from the pheochromocytoma of F11-III.1 showed a loss of the WT allele in the tumor, demonstrating loss of heterozygosity as currently described in classical *VHL* disease.

Suspecting an effect of the synonymous mutations on splicing, we next assessed the expression of the *VHL* transcripts in patients' samples by RT-PCR (Figure 5C). We found a significant change in the ratio of expressed *VHL* isoforms, with a higher expression of the E1-E3 transcripts in LCLs of patients homozygous for D143D and in the pheochromocytoma with P138P mutation. These results suggest an effect of the mutation on splicing regulatory elements, leading to E2 skipping. In silico analysis of

the synonymous mutations indicated a potential effect on exonic splicing enhancer motifs (supplemental Figure 14).

To specifically quantify the different *VHL* isoforms, we performed quantitative RT-PCR on LCLs using TaqMan probes complementary to the *VHL* E1-E2 or E1-E3 junction (Figure 5D). All mutated samples displayed an increase of the isoform with skipped E2 (E1-E3). The patients carrying a homozygous D143D mutation presented a severe decrease of the WT isoform expression level. These results demonstrated that synonymous mutations in E2 affected mRNA isoform production.

To evaluate a potential impact on protein expression, we performed a western blot analysis. We did not observe any overexpression of the pVHL172 isoform as expected from mRNA quantification studies. However, in patients homozygous for D143D, we detected a strong downregulation of all the pVHL isoforms, which was also observed to a lesser extent in LCL samples from heterozygous patients (Figure 5E).

### Proportion of *VHL* E2 skipping is correlated with disease severity in minigene experiments

To study the impact of the synonymous E2 mutations on splicing in different cell lines, we performed minigene assays. We tested the implications of the D143D and P138P mutations on splicing in addition to the nearby mutations described in patients with erythrocytosis (P138L and G144R).<sup>4,6,29</sup> Indeed, these mutations may also affect splicing rather than *VHL* protein function. We first evaluated the loss of function of these mutants by reporter assays. We found that these could downregulate HIF in a manner similar to the WT protein, in contrast to *VHL* proteins lacking E2 (pVHLΔE2), which are described as nonfunctional in terms of their regulation of HIF activity<sup>8-10,12</sup> (Figure 6A). Minigene experiments were performed in a variety of cell lines relevant to the studied diseases, using LCLs as control (Figure 6B). We demonstrated that all mutations caused E2 skipping in LCLs. In other cell lines, these experiments demonstrated a major impact on the *VHL* E2 splicing of the P138P mutation, which is associated with cancer. The mutations P138L, D143D, and G144R, which are all associated with erythrocytosis, displayed a weaker effect on splicing, with slight variations among cell lines but with a stronger effect in the erythroid cell line (UT7 cultured with erythropoietin).

### Splicing dysregulation of *VHL* is causal in the development of disease

RNA-seq confirmed that the synonymous mutations were not silent but instead induced potent E2 skipping (Figure 6C; supplemental Figure 15A). Additional transcriptome analyses of the pheochromocytoma carrying P138P showed an upregulation of HIF target genes (typically seen in *VHL*-related pheochromocytomas) compared with the pheochromocytoma carrying a *RET* mutation (supplemental Figure 15B).<sup>30</sup> We then reanalyzed

**Figure 5. Genetic and expression study of synonymous mutations in *VHL* E2.** (A-B) Pedigree and sequence chromatograms of germline DNA, tumor (pheochromocytoma) DNA, or cDNA prepared from 2 families (F9 and F10) with erythrocytosis (A) and 2 families (F11 and F12) with *VHL* disease (B). (C) Results of RT-PCR using mRNA extracted from LCLs (F9 and F10) and leukocytes and tumor material (pheochromocytoma) (F11 and F12). (D) TaqMan quantification of the different *VHL* isoforms in LCLs (established from patients of F9, F10, and F11) cultured in the absence or presence of puromycin. TaqMan probes are specific to the *VHL* E1-E2 or E1-E3 junction. Relative gene expression was normalized to LCL control (C1). (E) Immunoblot analysis of patient LCLs. Patients with erythrocytosis are indicated in red. A representative immunoblot (upper panel) and quantification of 4 different immunoblots are displayed (lower panel). Relative gene expression was normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression, and results obtained with C1 were fixed at 100%. Data are represented as the mean  $\pm$  standard error of the mean. \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ , \*\*\*\* $P < .0001$  based on Student *t* test. CNS, central nervous system.

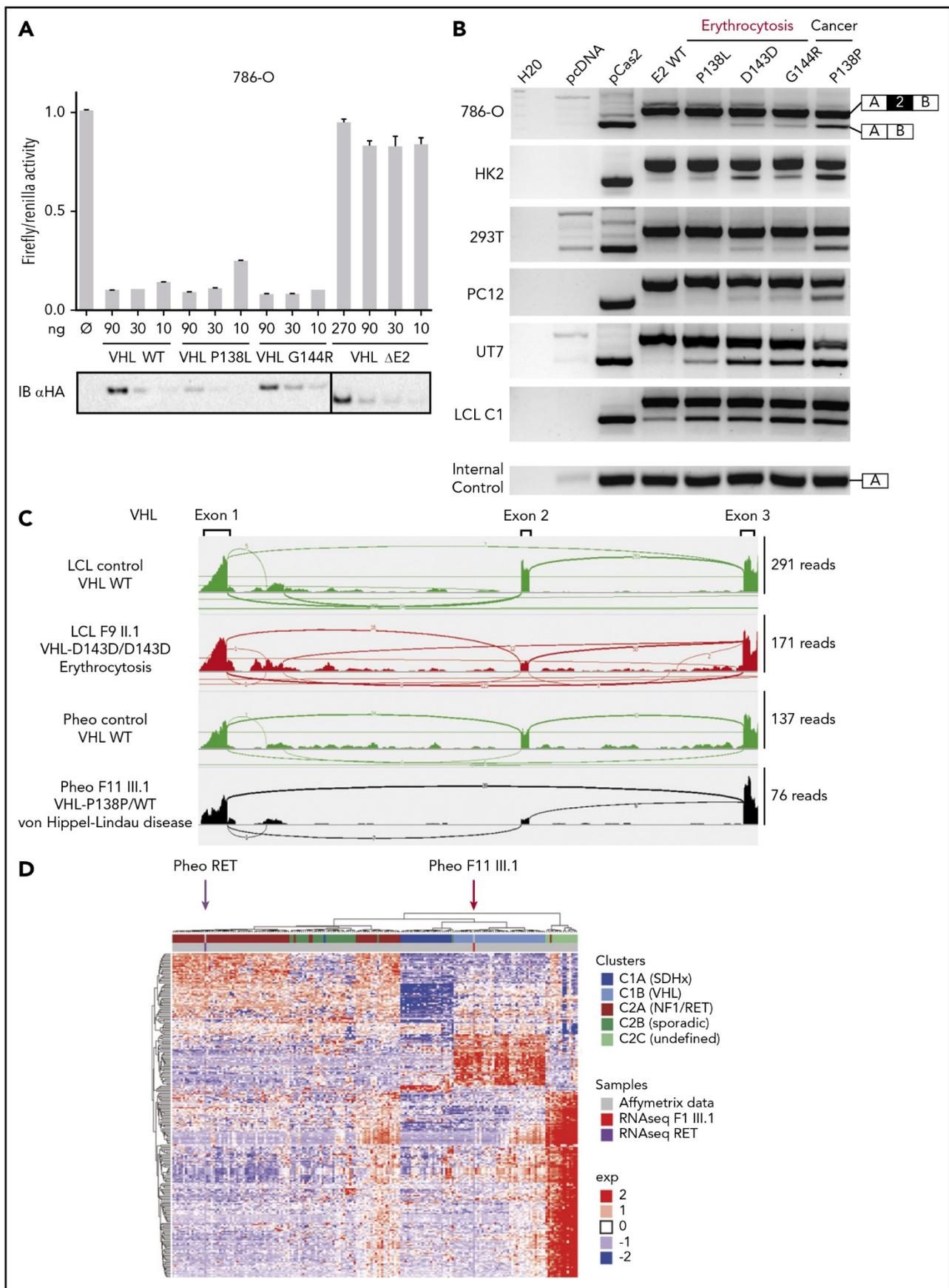


Figure 6.

our independent cohort of pheochromocytomas<sup>30</sup> using RNA-seq data from both tumors. After unsupervised classification, we observed a segregation of the P138P pheochromocytoma with other *VHL*-related tumors (C1B cluster), whereas the control pheochromocytoma bearing the *RET* mutation was grouped with other *RET*-related tumors (Figure 6D).

## Discussion

The hypoxia pathway plays a central role in erythrocytosis or tumors developed by patients carrying *VHL* mutations. Nevertheless, the full molecular mechanisms at the origin of these different phenotypes remain to be elucidated. To date, functional studies of *VHL* mutants have been performed on missense mutations. We describe here, for the first time, functional studies of *VHL* mutations that did not have an impact on the coding sequence but that influenced *VHL* splicing. We discovered a complex regulation of *VHL* splicing that may help to explain the complexity of genotype/phenotype correlations observed in *VHL*-related disorders. Notably, we demonstrated that synonymous variants (D143D or P138P) could have an impact on *VHL* splicing and should be considered as pathogenic mutations. Our study points to a particular region in E2 that may be considered as a splicing regulatory domain. Therefore, it would be interesting to evaluate the impact of all the nucleotide changes described in *VHL* E2<sup>17</sup> on splicing, in the same way as described for 2 missense mutations (P138L and G144R). We observed that, depending on the mutation in this region, the impact on splicing could be moderate (D143D, G144R, and P138L) or severe (P138P), correlating with the severity of the disease seen in individuals carrying these *VHL* mutations (erythrocytosis vs cancers). This observation confirms the hypothesis of a continuum model of tumor suppression by *VHL*.<sup>21,31</sup>

Regarding the erythrocytosis developed in patients homozygous for D143D, both probands (F9 II.1 and F10 II.1) presented mutations in the  $\beta$ -globin gene (*HBB*) that induced hemoglobin instability or thalassemia (Table 1). This may have compensated for the strong erythropoiesis associated with a very high serum erythropoietin level associated with the D143D mutation.

More importantly, we discovered a new *VHL* cryptic exon, E1', expressed in healthy tissues. Our study led to the identification of E1' heterozygous mutations occurring in the second allele of 6 families with erythrocytosis previously associated with a heterozygous mutation in *VHL* rather than a homozygous mutation. Our investigations further confirmed that polycythemia associated with *VHL* mutation was definitely an autosomal recessive disease.

In addition, we identified an E1' homozygous mutation in a patient with erythrocytosis of unknown origin. This result demonstrated the causal role of the alterations in this new cryptic exon in the pathophysiology of erythrocytosis. Importantly, we also identified E1' mutations in patients with unexplained *VHL* disease.

This *VHL* E1' exon remained unidentified until this point because of its low expression and the fact that this deep intronic region was never explored or represented in whole-exome sequencing data. Our data showed that these newly described E1'-containing transcripts may have been polyadenylated (because they were captured by polydT in RNA-seq) but were also likely to have been subjected to NMD and may therefore have failed to produce a protein. However, we cannot exclude the possibility that they were translated into a new protein (X1) not expressed in sufficient quantity to be detected by western blot. This potential X1 isoform would contain the *VHL* E1 that encodes the NH2-terminal part of pVHL, including 16 residues involved in HIF binding from the 17 described.<sup>26</sup> Nonetheless, the hypoxia-dependent reporter assays failed to identify a potential direct role of this isoform in the HIF pathway. Instead, our results provided compelling evidence that mutations in E1' induced a severe retention of this E1' cryptic exon, which correlated with a defect in global *VHL* protein expression. Therefore, our study strongly favors a dysregulation of splicing with a consequent downregulation of pVHL expression as the underlying cause of the diseases observed. Here, insufficient pVHL levels but not reduced HIF binding by the mutant pVHL (as seen in the Chuvash polycythemia mutation *VHL* R200W<sup>1</sup>) may have led to an impairment of HIF degradation. Of note is the fact that the functional study of mutations identified in E1' in association with erythrocytosis resulted in a less severe impact on splicing than the mutations associated with cancer, confirming that polycythemia was associated with *VHL* hypomorphic mutations.

Our findings may have broad implications for patients with presumed congenital erythrocytosis. First, the underlying cause of congenital erythrocytosis has been identified in only approximately one third of patients in most published studies. However, previous studies have focused on missense and nonsense changes in coding regions and known regulatory domains of candidate genes. This has also been the case in patients with suspected *VHL* disease. Our study shows that synonymous exonic changes, as well as changes within intronic sequences affecting exon splicing, may be responsible for these disorders and should be considered during the diagnostic process. Notably, the *VHL* E1' exon should be added to the list of regions routinely sequenced in patients with congenital erythrocytosis. Second, the detection of molecular changes

**Figure 6. Functional study of synonymous mutations in *VHL* E2.** (A) Functional hypoxia response element (HRE)-dependent reporter assays were performed in 786-O cells to evaluate the impact of *VHL* mutations in E2 (P138L and G144R) on *VHL* protein activity. The *VHL* protein lacking E2 (pVHL172/*VHL*ΔE2) was used as a negative control. The results are expressed as firefly luciferase activity relative to *Renilla* luciferase as an internal control; 1.0 unit denotes the basal activity of endogenous HIF-2 $\alpha$  using the HRE luciferase reporter plasmid. Immunoblots using an antibody specific for the hemagglutinin tag (HA) were used to detect HA-VHL. (B) Characterization of *VHL* E2 skipping by minigene analyses. Minigene experiments were performed in a variety of cell lines relevant to the studied diseases: renal (293T, 786-O, and HK2), pheochromocytoma (pheo; PC12), erythroid cell line (UT7 cultured with erythropoietin), and LCL. RT-PCR was performed on mRNA obtained from cell lines transfected with a minigene construct containing *VHL* E2 flanked by intronic sequences cloned between the *SERP1* exons (exons A and B; targeted by the RT-PCR primers). Bands corresponding to EA and EB spliced together or with *VHL* E2 are indicated on the right. (C) Sashimi plots from RNA-seq data. The positions of the different *VHL* exons are indicated, with the maximum number of reads for each exon indicated at the right. Splice junctions are denoted by the horizontal links, with details provided in supplemental Figure 15A. (D) Heatmap of pheochromocytoma transcriptome data. A comparison of transcriptome data for the pheochromocytoma from patient F11 III.1 (with P138P mutation) vs Affymetrix data from the largest available cohort of paragangliomas/pheochromocytomas (recruited by the French COMETE network) that identified homogeneous molecular subgroups associated with susceptibility genes (Burnichon et al<sup>30</sup>).

has implications for the clinical management of patients. For example, phlebotomy in patients with *VHL*- or *HIF2A*-related erythrocytosis may worsen the clinical situation by increasing the risk and severity of pulmonary hypertension in these patients.<sup>32</sup> Third, the confirmation of the continuum model of tumor suppression by *VHL* helps us to understand the very low frequency of secondary tumors in patients with *VHL*-related erythrocytosis.<sup>33</sup> However, this also means that later occurrence of such neoplasms cannot be definitely excluded. Therefore, the clinical management of these patients should include regular follow-up to assess the risk for thromboembolic complications, pulmonary hypertension, and cardiovascular disease, as well as regular examinations to check for the presence of typical *VHL*-related tumors. Patients with *VHL* disease resulting from E1' mutations will also benefit from regular screening for tumors. Finally, the detection of these new genetic changes will also allow for appropriate genetic counseling of affected patients and their families.

In conclusion, *VHL* is a major tumor suppressor gene that plays a pivotal role in the oxygen-sensing pathway, which is involved in multiple physiological (eg, angiogenesis and erythropoiesis) and pathological (eg, cancer) processes. Our findings regarding the complex splicing regulation of this gene in erythrocytosis and tumorigenesis may therefore open new avenues for diagnosis of these conditions as well as research on the biological mechanisms related to the hypoxia-signaling pathway. Notably, we suggest further targeted exploration of the *VHL* E1' region in unresolved cases of congenital erythrocytosis, inherited kidney cancers, hereditary paraganglioma/pheochromocytoma syndrome, and hemangioblastomas, in addition to all types of sporadic tumors with altered hypoxia signaling.

## Acknowledgments

The authors thank Richard Breathnach, Jean Feunteun, Judith Favier, and Sylvie Hermouet for scientific discussions; Helena M. Silva, Caroline Abadie, Sophie Giraud, Florence Fellmann, Pascal Pigny, Vinciane Dideberg, and Segers Karine for patient recruitment and medical advice; and Amandine Le Roy, Isabelle Barbieux, Christophe Simian, Marie-Laure Clenet, and Stéphanie Lebeau for technical assistance, in addition to the Genomics and Bioinformatics Core Facility of Nantes (GenoBiRD; Biogenouest) for its technical support. The authors are grateful to Mario Tosi, Pascaline Gaildrat, and Alexandra Martins (Institute for Research and Innovation in Biomedicine, INSERM U1079, University of Rouen, Rouen, France) for kindly providing the pCAS2 minigene plasmid. The authors thank the European Cooperation in Science and Technology (COST) Networks Molecular Diagnosis of Myeloproliferative Neoplasms (MPNs) and MPN-Related Congenital Diseases Euronet (COST action BM0902) and Hypoxia Sensing, Signaling, and Adaptation Hypoxianet (COST action TD0901). The authors also thank the Patients' Association *VHL* France for its constant support.

## REFERENCES

1. Ang SO, Chen H, Hirota K, et al. Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia. *Nat Genet.* 2002; 32(4):614-621.
2. Gordeuk VR, Sergueeva AI, Miasnikova GY, et al. Congenital disorder of oxygen sensing: association of the homozygous Chuvash polycythemia *VHL* mutation with thrombosis and vascular abnormalities but not tumors. *Blood.* 2004;103(10): 3924-3932.
3. Bento C, Cario H, Gardie B, Hermouet S, McMullin MF. Congenital Erythrocytosis

and Hereditary Thrombocytosis. MPN & MPN EuroNet COST Action; Brussels, Belgium: European Cooperation in Science and Technology; 2015.

4. Bento C, Percy MJ, Gardie B, et al; ECE-Consortium. Genetic basis of congenital erythrocytosis: mutation update and online databases. *Hum Mutat.* 2014;35(1):15-26.
5. Cario H, Schwarz K, Jorch N, et al. Mutations in the von Hippel-Lindau (*VHL*) tumor suppressor gene and *VHL*-haplotype analysis in patients with presumable congenital erythrocytosis. *Haematologica.* 2005;90(1): 19-24.

6. Randi ML, Murgia A, Putti MC, et al. Low frequency of *VHL* gene mutations in young individuals with polycythemia and high serum erythropoietin. *Haematologica.* 2005;90(5): 689-691.
7. Schoenfeld A, Davidowitz EJ, Burk RD. A second major native von Hippel-Lindau gene product, initiated from an internal translation start site, functions as a tumor suppressor. *Proc Natl Acad Sci USA.* 1998;95(15): 8817-8822.
8. Chesnel F, Hascoet P, Gagné JP, et al. The von Hippel-Lindau tumour suppressor gene: uncovering the expression of the pVHL172 isoform. *Br J Cancer.* 2015;113(2):336-344.

This study was supported by grants from the Région Pays de la Loire, Project EryCan; the Agence Nationale Recherche (Programme de Recherche Translationnelle en Santé 2015 GenRED); the Laboratory of Excellence GR-Ex (reference #ANR-11-LABX-0051); and the Ligue Nationale Contre le Cancer (Comités de la Loire Atlantique et des Côtes d'Armor). In addition, we acknowledge funding from the Oxford National Institute for Health Research Biomedical Research Centre and the Health Innovation Challenge Fund scheme. The views expressed in this manuscript are those of the authors and not necessarily those of the Wellcome Trust and Department of Health. We also acknowledge funding from the Wellcome Trust Core Award (grant #203141/Z/16/Z). A.B. received financial support from L'Institut Thématique Multi-Organisme Cancer Alliance Nationale pour les Sciences de la Vie et de la Santé (National Alliance for Life Sciences and Health) within the framework of the Cancer Plan. D.H. is supported by the National Center of Competence in Research Kidney Control of Homeostasis.

## Authorship

Contribution: M.L., F.R., K.S., A.C., D.H., A.B., S.G., S.C., F.C., M.P., T.B., S.F., M.V., Y.A.-B., and B.G. performed experiments; C.C., S.J.L.K., J.C.T., M.P., E.K., and H.D. conducted genome sequencing or data analysis and interpretation; P.L., M.C., S. Dumont, S.J., and M.L. performed bioinformatic analyses; H.C., K.S., S. Deveaux, N.B., J.-M.M., F.A., C.G., L.H., S.I., E.M., K.B., K.-M.D., B.B.-d.P., F.G., M.-L.R., M.C.P., V.B., R.V.W., J.R.G., A.K., N.J., C.B., A.-P.G.-R., and S.R. conducted the medical study; B.G., D.H., M.L., and F.R. wrote the manuscript; B.G. and H.C. designed the study; B.G. directed the study; and all authors contributed to the research and approved the final manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Betty Gardie, Ecole Pratique des Hautes Etudes, L'Institut du Thorax, INSERM, CNRS, Université de Nantes, Nantes, France; e-mail: betty.gardie@inserm.fr.

## Footnotes

Submitted 9 March 2018; accepted 23 May 2018. Prepublished online as *Blood* First Edition paper, 11 June 2018; DOI 10.1182/blood-2018-03-838235.

\*M.L. and F.R. contributed equally to this work.

†S.R. and A.-P.G.-R. contributed equally to this work.

‡H.C. and B.G. contributed equally to this work.

The online version of this article contains a data supplement.

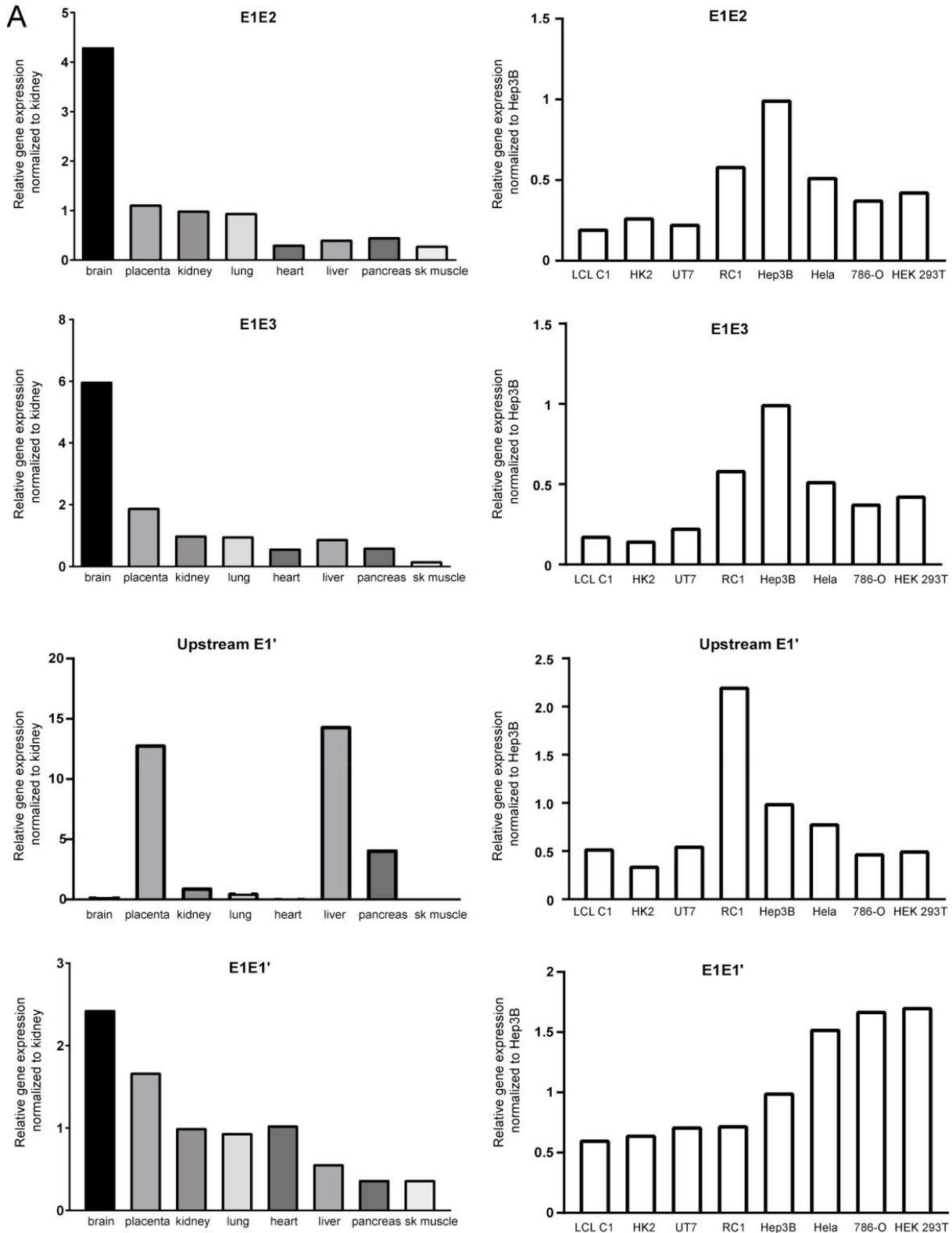
The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

9. Clifford SC, Cockman ME, Smallwood AC, et al. Contrasting effects on HIF-1alpha regulation by disease-causing pVHL mutations correlate with patterns of tumorigenesis in von Hippel-Lindau disease. *Hum Mol Genet.* 2001;10(10):1029-1038.
10. Gnarr JR, Tory K, Weng Y, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nat Genet.* 1994;7(1):85-90.
11. Hascoet P, Chesnel F, Jouan F, et al. The pVHL<sub>172</sub> isoform is not a tumor suppressor and up-regulates a subset of pro-tumorigenic genes including *TGFB1* and *MMP13*. *Oncotarget.* 2017;8(44):75989-76002.
12. Richards FM, Schofield PN, Fleming S, Maher ER. Expression of the von Hippel-Lindau disease tumour suppressor gene during human embryogenesis. *Hum Mol Genet.* 1996;5(5):639-644.
13. Maxwell PH, Wiesener MS, Chang GW, et al. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature.* 1999;399(6733):271-275.
14. Stebbins CE, Kaelin WG Jr, Pavletich NP. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science.* 1999;284(5413):455-461.
15. Iliopoulos O, Kibel A, Gray S, Kaelin WG Jr. Tumour suppression by the human von Hippel-Lindau gene product. *Nat Med.* 1995;1(8):822-826.
16. Latif F, Tory K, Gnarr J, et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science.* 1993;260(5112):1317-1320.
17. Nordstrom-O'Brien M, van der Luijt RB, van Rooijen E, et al. Genetic analysis of von Hippel-Lindau disease. *Hum Mutat.* 2010;31(5):521-537.
18. Kaelin WG Jr. Molecular basis of the VHL hereditary cancer syndrome. *Nat Rev Cancer.* 2002;2(9):673-682.
19. Maher ER, Kaelin WG Jr. von Hippel-Lindau disease. *Medicine (Baltimore).* 1997;76(6):381-391.
20. Richard S, Gardie B, Couvé S, Gad S. Von Hippel-Lindau: how a rare disease illuminates cancer biology. *Semin Cancer Biol.* 2013;23(1):26-37.
21. Couvé S, Ladroue C, Laine E, et al. Genetic evidence of a precisely tuned dysregulation in the hypoxia signaling pathway during oncogenesis. *Cancer Res.* 2014;74(22):6554-6564.
22. Camps C, Petousi N, Bento C, et al; WGS500 Consortium. Gene panel sequencing improves the diagnostic work-up of patients with idiopathic erythrocytosis and identifies new mutations. *Haematologica.* 2016;101(11):1306-1318.
23. Rimmer A, Phan H, Mathieson I, et al; WGS500 Consortium. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet.* 2014;46(8):912-918.
24. Ladroue C, Hoogewijs D, Gad S, et al. Distinct deregulation of the hypoxia inducible factor by PHD2 mutants identified in germline DNA of patients with polycythemia. *Haematologica.* 2012;97(1):9-14.
25. Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol.* 2010;653:249-257.
26. Czyzyk-Krzeska MF, Meller J. von Hippel-Lindau tumor suppressor: not only HIF's executioner. *Trends Mol Med.* 2004;10(4):146-149.
27. Liu E, Percy MJ, Amos CI, et al. The worldwide distribution of the VHL 598C>T mutation indicates a single founding event. *Blood.* 2004;103(5):1937-1940.
28. Kondo K, Kim WY, Lechpammer M, Kaelin WG Jr. Inhibition of HIF2alpha is sufficient to suppress pVHL-defective tumor growth. *PLoS Biol.* 2003;1(3):E83.
29. Lanikova L, Lorenzo F, Yang C, et al. Novel homozygous VHL mutation in exon 2 is associated with congenital polycythemia but not with cancer. *Blood.* 2013;121(19):3918-3924.
30. Burnichon N, Vescovo L, Amar L, et al. Integrative genomic analysis reveals somatic mutations in pheochromocytoma and paraganglioma. *Hum Mol Genet.* 2011;20(20):3974-3985.
31. Berger AH, Knudson AG, Pandolfi PP. A continuum model for tumour suppression. *Nature.* 2011;476(7359):163-169.
32. Sable CA, Aliyu ZY, Dham N, et al. Pulmonary artery pressure and iron deficiency in patients with upregulation of hypoxia sensing due to homozygous VHL(R200W) mutation (Chuvash polycythemia). *Haematologica.* 2012;97(2):193-200.
33. Woodward ER, Wall K, Forsyth J, Macdonald F, Maher ER. VHL mutation analysis in patients with isolated central nervous system haemangioblastoma. *Brain.* 2007;130(Pt 3):836-842.

# Supplemental data

## SUPPLEMENTAL FIGURES

### Supplemental Figure 1



## Supplemental Figure 1

**B**

	E1E2	E1E3	Upstream E1'	E1E1'	RPLP0
brain	26.76	32.31	28.16	32.90	21.425
placenta	25.54	30.90	26.125	30.25	25.2
kidney	25.10	31.10	26.595	30.38	22.03
lung	25.31	31.42	26.85	30.54	21.51
heart	26.76	31.96	26.74	30.39	19.155
liver	24.12	29.09	24.745	28.94	24.025
pancreas	23.98	29.57	24.045	29.47	21.44
skeletal muscle	29.48	36.34	30.775	34.41	19.105

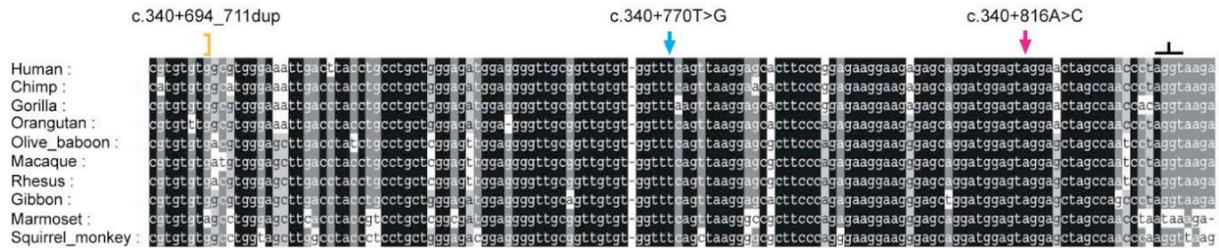
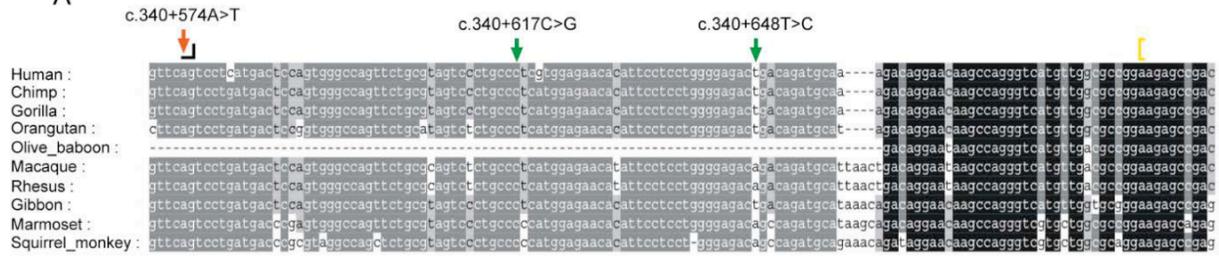
	E1E2	E1E3	Upstream E1'	E1E1'	RPLP0
LLB C1	22.90	28.44	25.12	27.47	18.6
HK2	22.73	28.93	25.58	27.68	18.4
UT7	22.37	26.75	24.62	26.99	18.2
RC1	22.70	28.00	25.77	28.67	21.3
Hep3B	21.10	26.49	24.32	27.31	19.1
Hep3B	21.52	27.01	24.495	27.80	18.9
Hela	22.93	28.99	25.225	27.66	19.27
786-O	22.77	28.64	25.74	26.92	19.065
HEK 293	22.25	27.83	24.99	26.54	18.385

### Supplemental Figure 1: Expression of transcripts containing Exon 1' in tissues and cell lines.

(A) Quantitative RT-PCR was carried out with a Taqman probe specific to junctions of the intron1-E1' (upstream E1'), E1-E1', E1-E2 or E1-E3 using 5 ng of cDNA from a collection of human tissues (left panel) and 20 ng of cDNA from cell lines (right panel). Relative gene expression (mean of duplicates) has been normalized to kidney (left panel) or Hep3B (right panel) fixed at 1. (B) Table of cycle threshold (Ct) numbers obtained by qPCR (mean of duplicates) in different tissues and cell lines with a threshold of 0.02 (UpE1' and E1E1'), 0.025 (E1E2), 0.0208 (E1E3), (0.015) RPLP0.

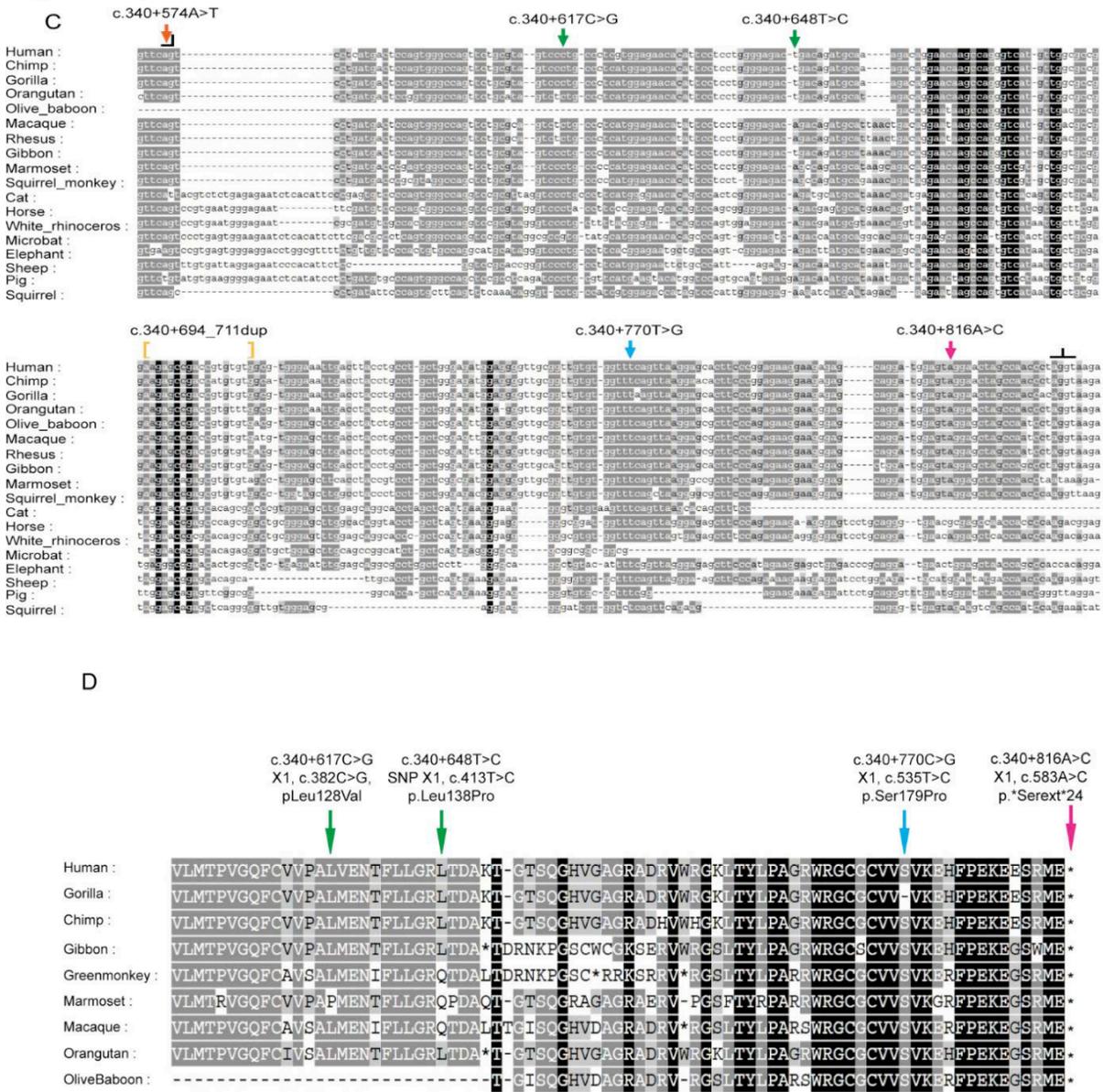
Supplemental Figure 2

A



B

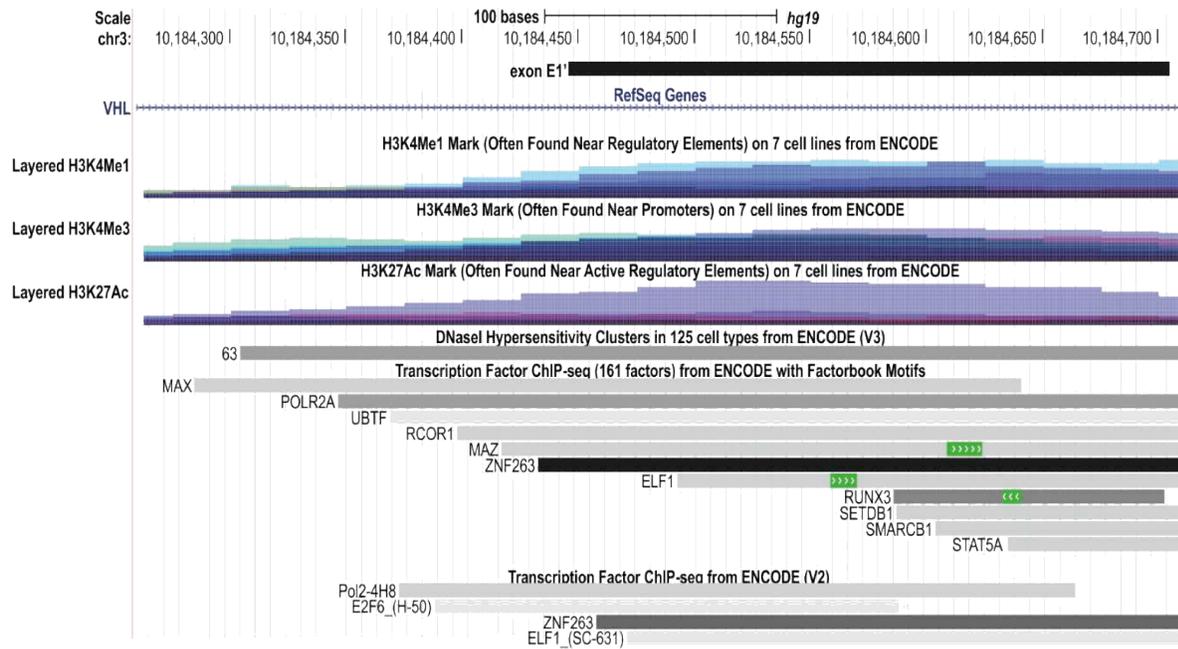
	Human :	Chimp :	Gorilla :	Orangutan :	Olive_baboon :	Macaque :	Rhesus :	Gibbon :	Marmoset :	Squirrel_monkey :
Human :	271	97%	97%	95%	61%	90%	90%	93%	85%	85%
Chimp :	265	271	97%	95%	60%	90%	90%	93%	86%	85%
Gorilla :	265	265	271	95%	60%	90%	90%	93%	86%	85%
Orangutan :	259	259	259	270	61%	91%	91%	93%	85%	85%
Olive_baboon :	166	164	164	167	180	64%	65%	60%	56%	56%
Macaque :	249	249	249	251	178	275	99%	92%	86%	85%
Rhesus :	250	250	250	252	179	274	275	93%	87%	86%
Gibbon :	256	256	256	256	166	255	256	275	88%	88%
Marmoset :	235	237	237	235	154	239	240	243	274	90%
Squirrel_monkey :	234	235	235	234	155	236	237	243	248	274



**Supplemental Figure 2: Conservation of the genomic region containing Exon 1' and the X1 protein.**

Multiple sequence alignment of *VHL* DNA sequences in primates (A) and mammals (C). Grey-scale shading indicates conservation of nucleotides. Splice sites and disease-associated *VHL-E1'* mutations are labeled at the top. (B) Sequence conservation. The percentage conservation relative to total alignment positions is displayed (D) Conservation of *VHL-X1* protein sequences encoded by *VHL E1-E1'* in various primates. Grey-scale shading indicates conservation of iso-functional residues; stars indicate stop codons. Disease-associated X1 mutations are labeled at the top.

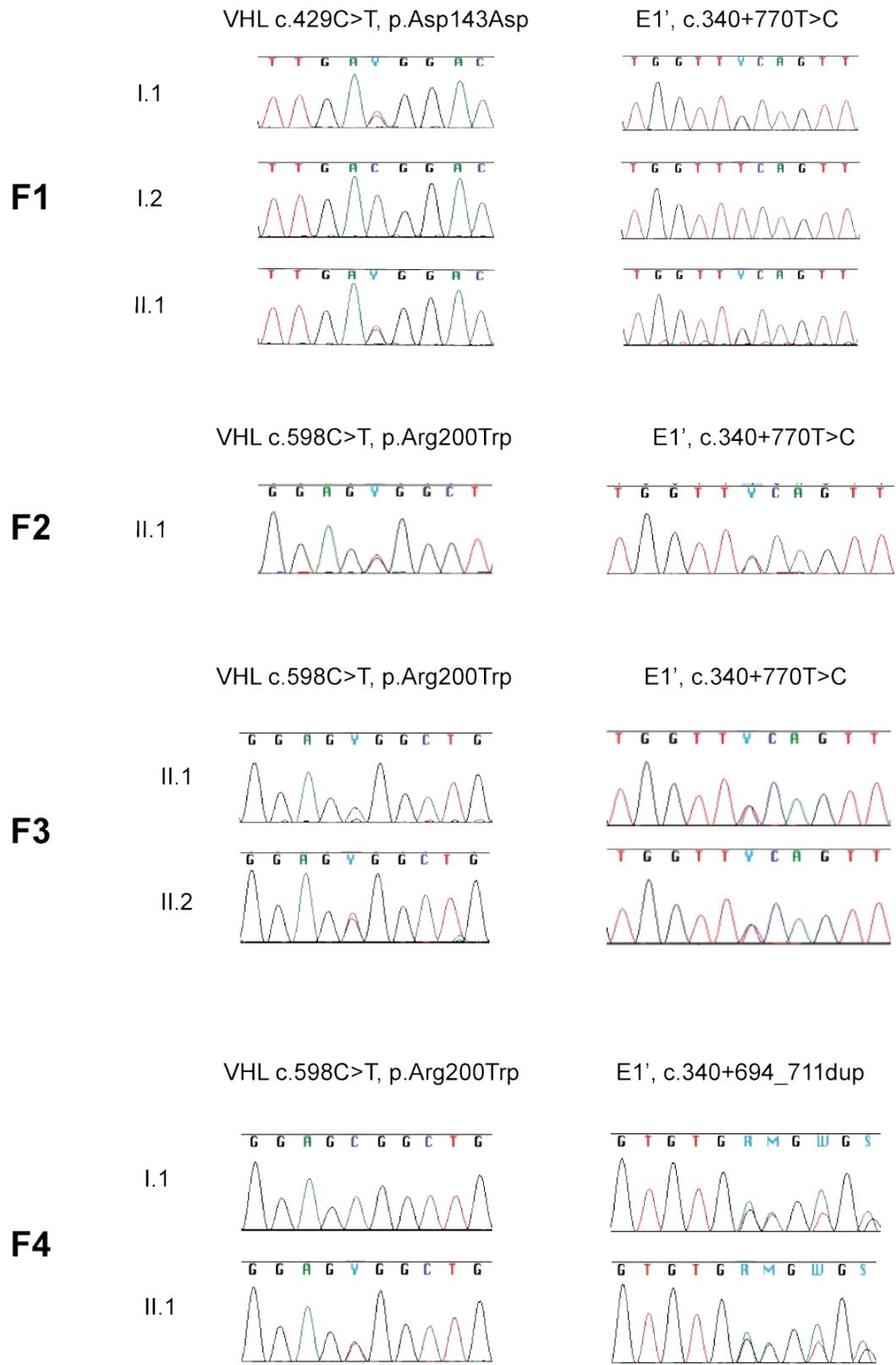
Supplemental Figure 3

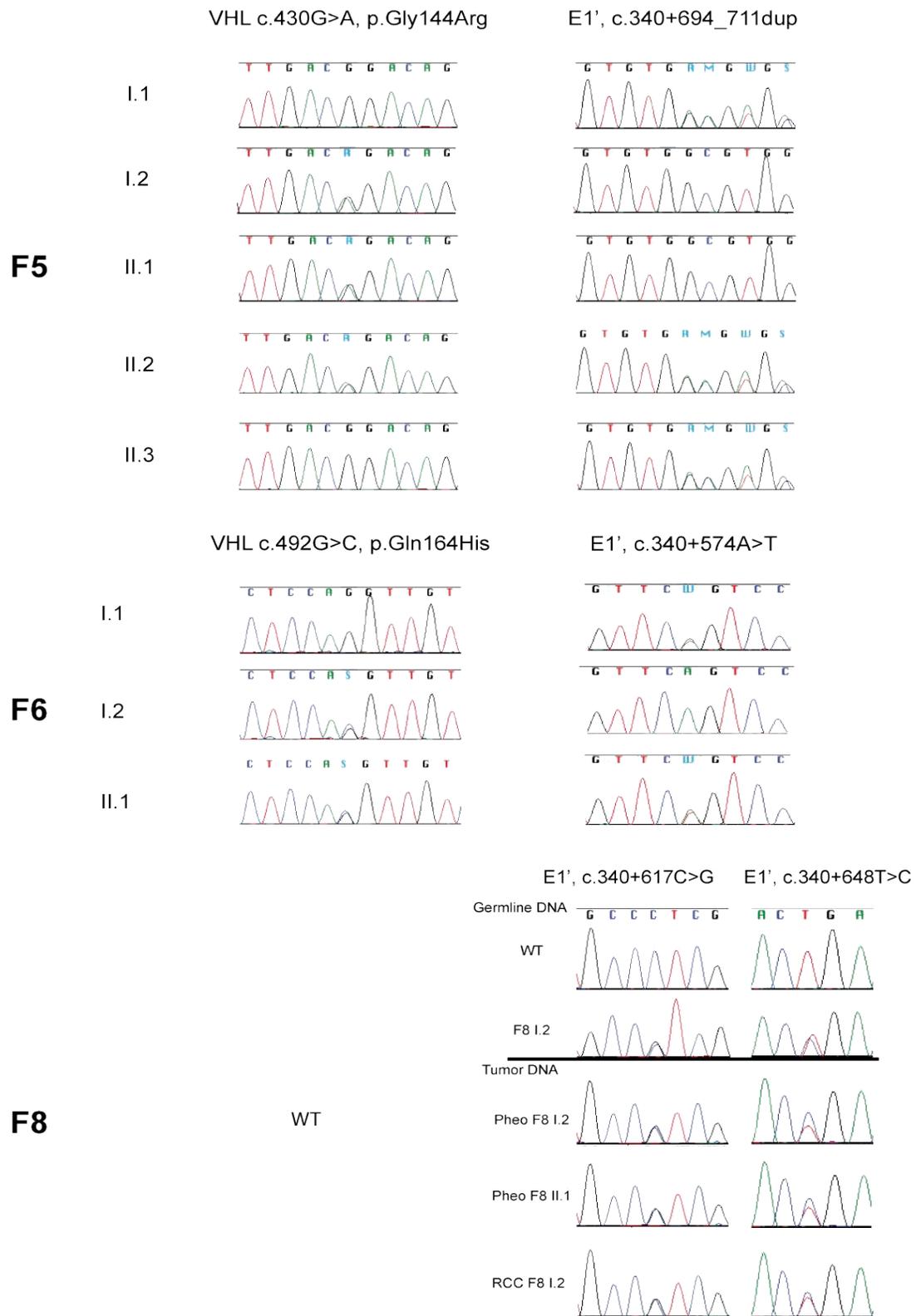


**Supplemental Figure 3: UCSC Genome Browser output (*hg19*) of the *VHL* exon E1' genomic region.**

The ENCODE integrated regulation track containing H3K4Me1/3 marks, H3K27Ac marks, DNaseI hypersensitivity clusters and transcription factor ChIP-seq data are displayed. Colors in the histone mark tracks represent conventional ENCODE codes used to illustrate different cell lines. The sequence located in 5' of E1' represents a transcriptionally active region as illustrated by epigenetic marks associated with promoters (H3K4Me3), enhancers (H3K4Me1), activated enhancers and promoter regions (H3K27Ac), in addition to DNase hypersensitivity commonly associated with open chromatin as well as high transcription factor occupancy.

Supplemental Figure 4





**Supplemental Figure 4: Results of genomic DNA sequencing.**

Chromatograms sequences of *VHL* Exon 2 or 3 are presented on the left, and *VHL* Exon 1' data on the right.

Supplemental Figure 5

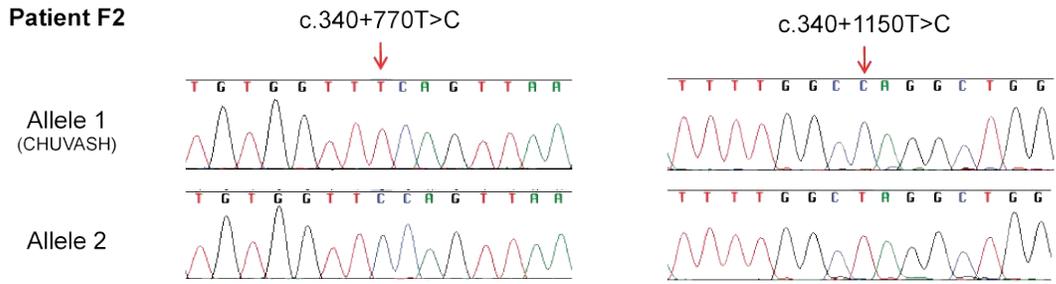
**ATGCCCGGAGGGCGGAGAACTGGGACGAGGCCGAGGTAGGCCGGAGGAGGCAGGCCGTCGAAGAGT**  
**ACGGCCCTGAAGAAGACGGCGGGGAGGAGTCGGGCGCCGAGGAGTCCGGCCCGAAGAGTCCGGCCCG**  
**GAGGAACTGGGCGCCGAGGAGGAGATGGAGGCCGGGCGCCGCGGCCCGTGTGCGCTCGGTGAACTC**  
**GCGCGAGCCCTCCCAGGTCATCTTCTGCAATCGCAGTCCGCGCGTGTGCTGCCGTATGGCTCAACT**  
**TCGACGGCGAGCCGAGCCCTACCCAACGCTGCCGCCTGGCACGGGCCGCCGCATCCACAGCTACCGA**  
**Ggtacgggcccggcgcttaggccgaccagcagggacgatagcaaggctctgaagccctctaccgcc**  
**ccggggtccatcttgcagacggggaactgagggcccttgaggcaggacacatccagggtgacgctgct**  
**cgtaagcgtcagagcatctctttttttttttttttttttttctgagacggagtctcgctctgtcgcccag**  
**gctggagtgacagtggcgcatctcgactcaactgcagcctccgctcccgggttcaagcgattctcctg**  
**cctcagcctcctgagtagctgggattacaggcgtgcccaccgcgcccggctgatttttatattttta**  
**gtagagacggggtttcaccatggttggtcaggctggtctcgaactgctgacctcgtgatccgcccgcct**  
**cggcctcccaaagtgctgggcttatgggcatgagcctccgcgcccggcccagagcattctttataagg**  
**ccgaatagtttgcatctgaaggtggctccccccagtcccccaccccacgtgtattttcccctcaaaga**  
**aaagctgcatccttaacaccccacatctgttcagtcctcatgactccagtgggccagttctgcgtagtcc**  
**ctgccctcgtggagaacacattcctcctggggagactgacagatgcaaagacaggaacaagccagggt**  
**catgttgggcggcgaagagccgaccgtgtgtggcgtgggaaattgacttacctgctgtgggagatg**  
**c.340+770T>C gaggggttgcggttggtggttccagttaaggagcacttcccggagaaggaagagagcaggatggagt**  
**aggaactagccaaccctag**gtaagaggttctagacatgcgtgcttgagacctggagtcttgggagag  
gatgcttaaaaggtgatttaccocctaggaatatgggggcaactgaaatTTTTTTTTTTTTTTTgagacg  
ggagtctgtcctgcaagctggagtgacgtggcccacgctagaatgcagtgggcgattgcggtcat  
tgcaacatctgccacctggggtcaagtggttctcttgccctcagcctcccaggagcggggattacagg  
cgtgcgccaccactcctggctaattTTTTTTTTTTtagtagagacgggggttctgctcattttggc**taggct**  
**c.340+1150T>C**ggtctcgaactcctgacctcagatgatccaccgccttggcctcccaaagtgctgagattacagggtg  
aagccactgcgcccagccctttgaaagttttcagtatattatgtatataatTTTTTgagttggagctc  
ggtctgtcggccagactggagtgctgttgac

Exon 1'

c.340+770T>C

c.340+1150T>C  
Chuvash SNP

SNP Chuvash (Liu et al., 2004) : [rs779808](#) (Chuvash Allele = C)  
*[Homo sapiens]* gagacgggggttctgctcattttggc[C/T]aggctggtctcgaactcctgacctc



Supplemental Figure 5: Families 2 and 3 alleles cloning and sequencing.

As biological material from the parents of patients F2 II.1, F3 II.1 and F3 II.2 were unavailable, we used patients genomic DNA to clone the intronic region containing E1' variant c.340+770T>C and the Chuvash core haplotype SNP rs779808 (described in Liu et al., 2004 and located upstream E1' in position c.340+1150), associated with the *VHL*-R200W mutation: *[Homo sapiens]* gagacgggggttctgctcattttggc[C/T]aggctggtctcgaactcctgacctc

We identified a "T" associated with the variant c.340+770T>C in E1' and a "C" (characteristic of the Chuvash core haplotype) associated with the wild type E1' sequence. This patient is therefore a compound heterozygote, with one mutation in each allele.

## Supplemental Figure 6

### A) Analysis of single nucleotide variants and small insertions/ deletions

#### A.1) Initial filtering steps common to both WGS datasets (patients F2 II.1, F3 II.1, F3 II.2; family 7 trio)

Filtering step	Number of variants	
	F2 II.1, F3 II.1, F3 II.2	Family 7 trio
<b>Import of multi-vcf file into IVA</b> (PASS variants with freq ≤ 3% in 1000G, NHLBI ESP6500 and ExAC)	694,445	661,545
<b>High confidence variants</b> (Quality ≥ 20; Read depth ≥ 10; Allele freq ≥ 0.05)	684,034	653,857
<b>Rare variants in general population</b> (Freq ≤ 1% in 1000G, NHLBI ESP, ExAC and gnomAD)	543,445	419,479

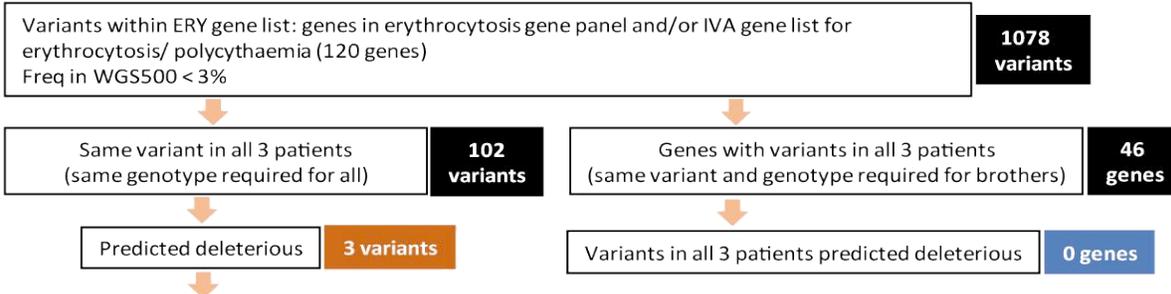
#### A.2) Downstream filtering for patients F2 II.1, F3 II.1 and F3 II.2 (all heterozygous for the *VHL* p.R200W mutation)

High confidence variants which are rare in general population **543,445 variants**

##### A.2.1) Variants within *VHL* gene +/- 5Kb

Genomic change	Gene Region	Transcript Variant	Protein Variant	Genotype			CADD	1000G Freq	ESP Freq	ExAC Freq	WGS500 Freq
				F2 II.1	F3 II.1	F3 II.2					
chr3:10184641T>C	Intronic	NM_000551.3: c.340+770T>C		Het	Het	Het	12.26	NR	NR	NR	NR
chr3:10191605C>T	Exonic	NM_000551.3: c.598C>T	p.R200W	Het	Het	Het	35	NR	0.015	0.023	NR

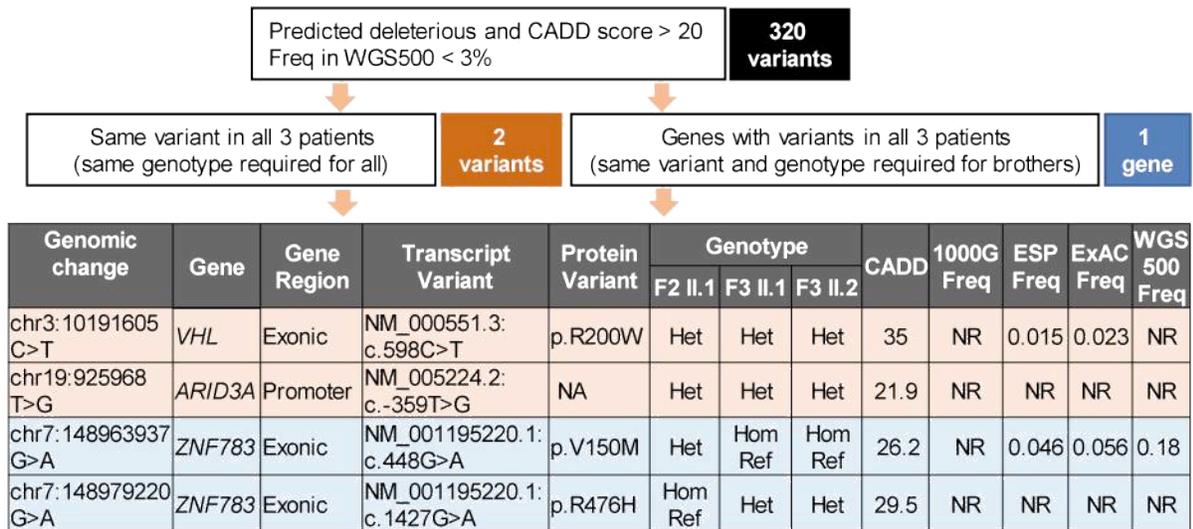
##### A.2.2) Variants in genes related or suspected to be related to erythrocytosis



Genomic change	Gene	Gene Region	Transcript Variant	Protein Variant	Genotype			CADD	1000G Freq	ESP Freq	ExAC Freq	WGS 500 Freq
					F2 II.1	F3 II.1	F3 II.2					
chr3:10191605 C>T	<i>VHL</i>	Exonic	NM_000551.3: c.598C>T	p.R200W	Het	Het	Het	35	NR	0.015	0.023	NR
chr9:135853324 T>C	<i>GFI1B</i>	Promoter Intronic	NM_004188.6: c.-925T>C; c.-355-570T>C		Het	Het	Het	7.34	0.26	NR	NR	0.55
chr11:33758819 delCACG	<i>CD59</i>	Promoter	NM_203331.2: c.-958_-955delCGTG;		Hom	Hom	Hom	NA	NR	NR	NR	NR

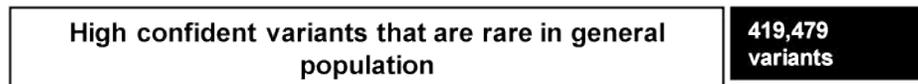
Literature search does not support candidacy for any of these variants

### A.2.3) Variants predicted deleterious with high functional impact

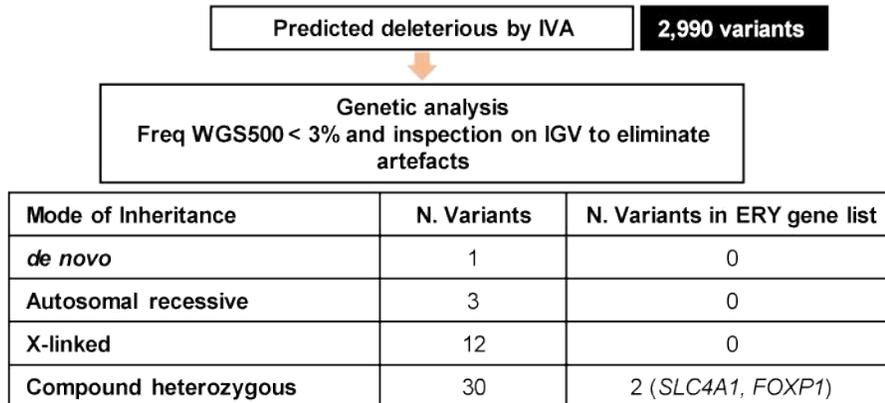


Literature search does not support candidacy for any of these variants

### A.3) Downstream filtering for family 7 (trio: proband affected, parents unaffected)



#### A.3.1) Analysis of single nucleotide variants and small insertions/ deletions



Genomic change	Gene	Gene Region	Transcript Variant	Genotype			CADD	1000G Freq	ESP Freq	ExAC Freq	WGS 500 Freq
				Proband	Father	Mother					
chr3:71114069 G>T	FOXP1	Intronic 5'UTR	NM_032682.5: c.283-11145C>A NM_001244813.1: c.-113C>A	Het	Hom Ref	Het	17.6	NR	NR	NR	NR
chr3:71448771 G>A	FOXP1	Intronic	NM_032682.5: c.-167-40376C>T	Het	Het	Hom Ref	1.4	0.04	NR	NR	NR
chr17:42343875 C>T	SLC4A1	Intronic	NM_000342.3: c.-69+1547G>A	Het	Het	Hom Ref	9.5	0.08	NR	NR	NR
chr17:42344297 A>G	SLC4A1	Intronic	NM_000342.3: c.-69+1125T>C	Het	Hom Ref	Het	4.9	0.99	NR	NR	0.36

Literature search does not support any of the variants as candidates

**A.3.2) Variants within *VHL* new cryptic exon**

Genomic change	Gene Region	Transcript Variant	Genotype			CADD	1000G Freq	ESP Freq	ExAC Freq	WGS500 Freq
			Proband	Father	Mother					
chr3:10184687A>C	Intronic	NM_000551.3:c.340+816A>C	Hom	Het	Het	9.95	NR	NR	NR	NR

**B) Analysis of copy number variants**

**B.1) Patients F2 II.1, F3 II.1 and F3 II.2**

(all heterozygous for the *VHL* p.R200W mutation)

**chrX:47829755-47834837 : Hemizygous loss in F3 II.1 and F3 II.2.**

- In *ZNF182* 3'UTR. This region has highly conserved POLR2A TFBS.
- Literature search on this gene does not support candidacy for this variant

**B.2) Family 7 (trio: proband affected, parents unaffected)**

No relevant copy number variants were identified

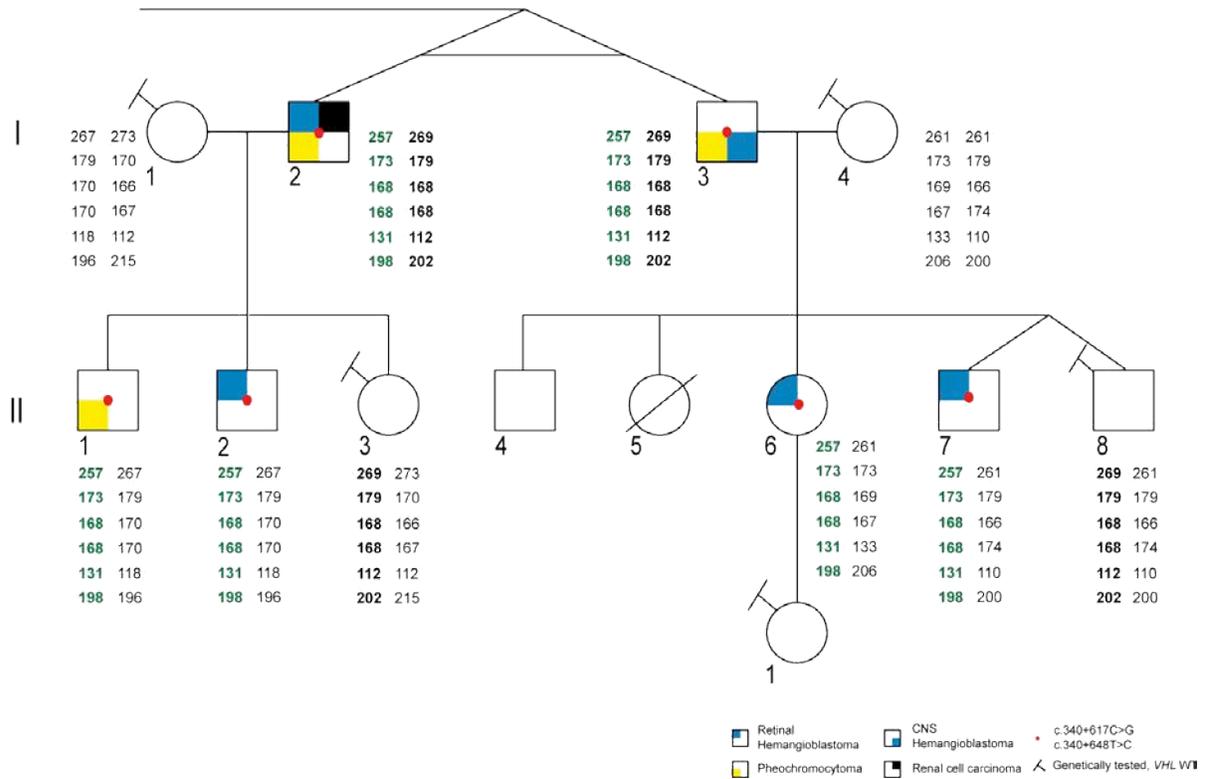
**Supplementary Figure 6: Analysis and interpretation of Whole Genome Sequencing (WGS) conducted in four patients with erythrocytosis.**

Sequencing data from patients F2 II.1, F3 II.1 and F3 II.2, all carrying a known heterozygous *VHL* p.Arg200Trp mutation, were analyzed together as a cohort. Data from Patient F7 II.1 and his unaffected parents (F7 I.1 and F7 I.2) were analyzed as a trio. (A) Filtering steps and variants detected using Ingenuity Variant Analysis Software (IVA) prioritizing single nucleotide variant (SNV) and small insertions/deletions according to potential clinical relevance. (A.1) Filtering steps common to both WGS data sets (A.2) Downstream filtering for patients F2 II.1, F3 II.1 and F3 II.2 using three strategies for variant identification: (A.2.1) Variants within the *VHL* gene (NM000551.3 transcript) and 5kb flanking; (A.2.2) Variants within 120 genes related or suspected to be related to erythrocytosis; (A.2.3) Variants predicted deleterious according to IVA parameters and with high functional impact defined as having a CADD score  $\geq 20$ . (A.3) Downstream filtering for family 7 trio using two strategies for variant identification: (A.3.1) Variants of clinical or biological relevance and mode of inheritance consistent with the affected/unaffected statuses of the proband and parents; (A.3.2) Variants within the *VHL* cryptic exon and 100bp flanking (B) Summary of copy number variants analyses (B.1) for patients F2 II.1, F3 II.1 and F3 II.2 and (B.2) for family 7 trio.

WGS500: WGS500 project union file of variants detected; 1000G: 1000 genomes project;

NHLBI ESP or ESP: NHLBI ESP6500 exomes project; ExAC: Exome Aggregation

Consortium; gnomAD: genome aggregation database; CADD: Combined Annotation Dependent Depletion; Het: heterozygous; Hom: homozygous; Hom Ref: homozygous for the reference allele.



### Supplemental Figure 7: Disease segregation analysis in family 8 using microsatellite markers segregation analysis.

Family 8 (F8) developed a von Hippel-Lindau disease but without any apparent mutations identified in canonical *VHL* exons E1, E2 or E3, nor in the 5' and 3' untranslated (UTR) regions. Chromosomal rearrangement analysis (by SYBRgreen-based quantitative PCR, Multiplex Ligation-dependent Probe Amplification and Quantitative Multiplex PCR of Short Fluorescent fragments) of *VHL* failed to identify any alteration. Multiplex PCR was performed using forward primers coupled with fluorescent dye designed to recognize 6 microsatellite markers that cover a 4.6 Mb region surrounding the *VHL* gene in 3p25-26. Haplotypes were reconstituted and are indicated for each patient. The green haplotype corresponds to the disease-associated haplotype.

## Supplemental Figure 8

**A**

Position on Chromosome 3 (hg19) :  
 Junction E1 on 3' : 10183872  
 Junction E1' on 5' : 10184447  
 Junction E1' on 3' : 10184706  
 Junction E2 on 5' : 10188198  
 Junction E2 on 3' : 10188321  
 Junction E3 on 5' : 10191471

**B**

LCL

	E1E2 (%)	E2E3 (%)	E1E3 (%)	E1E1' (%)	E1'E2 (%)	E1'E3 (%)	Total reads
Control	<b>47.8</b>	<b>33.9</b>	2.8	2	13.5	0	251
F11 II.6	26.8	20.3	5.9	<b>13.7</b>	<b>33.3</b>	0	153

Pheochromocytoma

	E1E2 (%)	E2E3 (%)	E1E3 (%)	E1E1' (%)	E1'E2 (%)	E1'E3 (%)	Total reads
Control	<b>41</b>	<b>48.2</b>	7.2	1.2	1.2	1.2	83
F11 II.1	22.8	17.2	5	<b>11.7</b>	<b>42.2</b>	1.1	180
F11 I.2	28.6	30.6	0	<b>16.3</b>	<b>20.4</b>	4.1	98

**C**

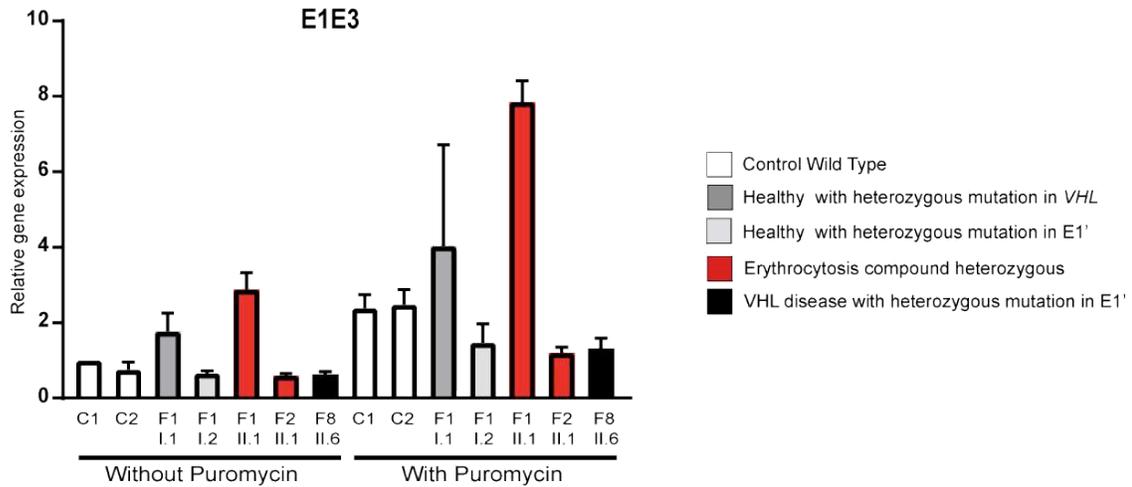
	c.340+617C>G		c.340+648T>C	
	C	G	T	C
Pheo F11 I.2	26%	<b>74%</b>	25%	<b>75%</b>
Pheo F11 II.1	26%	<b>74%</b>	26%	<b>74%</b>
LCL F11 II.6	28%	<b>72%</b>	31%	<b>69%</b>

### Supplemental Figure 8: RNA-Seq analysis of samples carrying mutations in *VHL*-Exon 1'.

(A) Position on chromosome 3 (*hg19*) of the different *VHL* splice junctions. (B) Total number and percentage of reads identified for specific junctions by RNA-Seq. (C) Percentage of wild type and mutant alleles of transcripts containing E1' detected by RNAseq. Most alleles expressed in the patient samples were mutated.

## Supplemental Figure 9

**A**



**B**

		E1E2	E1E3	E1E1'	RPLP0
- puro	LCL C1	23.15	28.70	27.60	19.37
	LCL C2	23.46	29.07	27.98	19.15
	LCL F1 I.2	24.39	27.51	27.73	19.37
	LCL F1 I.1	23.75	29.50	26.30	19.56
	LCL F1 II.1	26.05	27.28	26.17	19.45
	LCL F2 II.1	23.22	29.08	25.75	18.93
	LCL F8 II.6	23.88	29.33	25.69	19.32
+ puro	LCL C1	22.25	27.83	23.76	19.71
	LCL C2	21.85	27.28	24.54	19.22
	LCL F1 I.2	23.20	27.32	25.66	19.73
	LCL F1 I.1	22.63	28.71	22.99	19.71
	LCL F1 II.1	25.35	26.41	23.14	20.05
	LCL F2 II.1	22.32	28.73	22.84	19.61
	LCL F8 II.6	23.50	29.03	23.49	20.05

### Supplemental Figure 9: TaqMan quantification.

(A) Isoforms containing E1 spliced with E3 have been quantified by TaqMan using a probe specific to exon-exon junction. LCLs have been cultured in the absence or presence of puromycin, an inhibitor of Nonsense-Mediated mRNA Decay. (B) Table of cycle threshold (Ct) numbers obtained by qPCR in LCL, treated or not with puromycin, with a threshold of 0.02 (UpE1' and E1E1'), 0.025 (E1E2), 0.0208 (E1E3), (0.015) RPLP0. Results are expressed in mean of technical duplicates of three independent culture experiments.

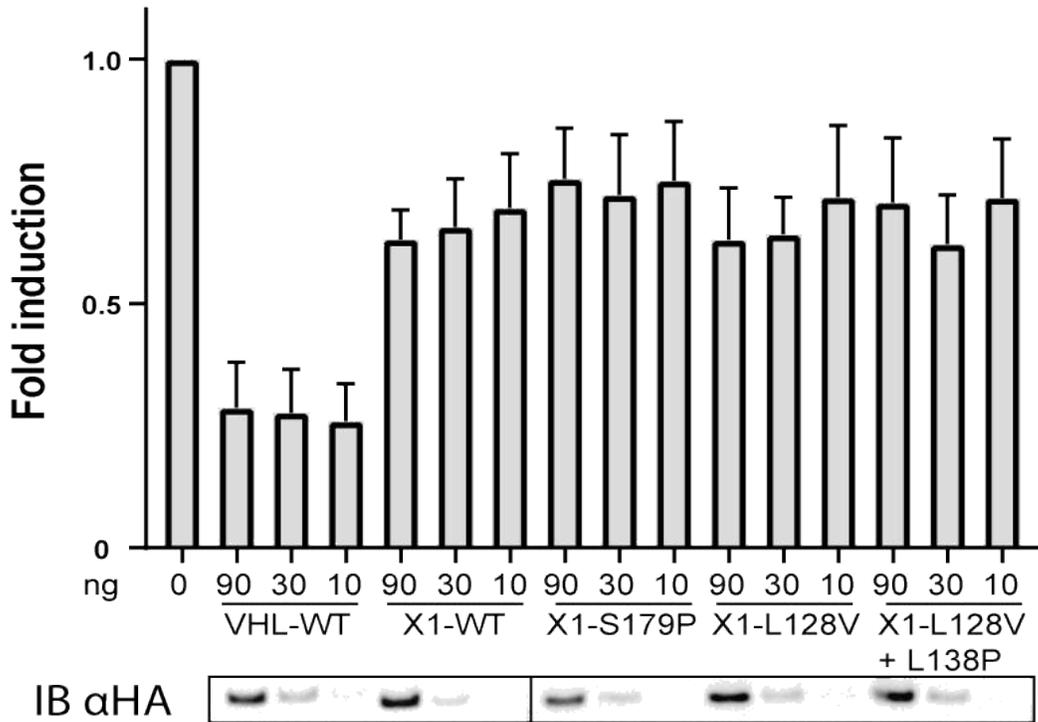
Supplemental Figure 10



Supplemental Figure 10: Sequence of *VHL* transcripts expressed in patient's F6 II.1 LCL.

Cloning and sequencing of *VHL* transcripts expressed in LCL of patient F5 II.2 carrying a mutation in the E1' splice acceptor site (c.340+574A>T). A RT-PCR has been performed on mRNA extracted from patient's LCL using primers in *VHL*-E1 and E3. PCR fragments obtained have been cloned and sequenced. Transcripts containing E1 correctly spliced with E2 and E3 have been obtained. Transcripts containing E1' contain sequences deleted to 15 nucleotides as a consequence of alternative splicing using a splice acceptor site located 3' of the consensus SA, mutated in the patient. Exon 1 is represented in black, Intron 1 in gray, Exon 1' in green, Exon 2 in red and Exon 3 in blue. The c.340+574A>T mutation is underlined in orange on the bottom sequence.

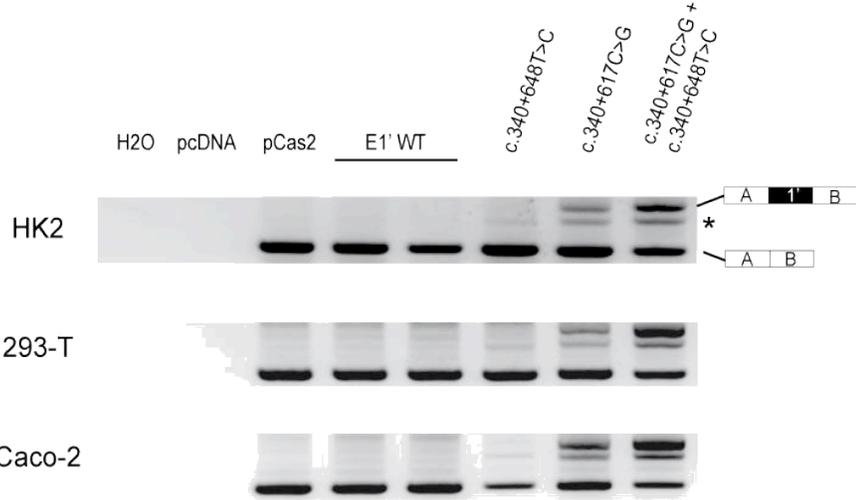
**Supplemental Figure 11**



**Supplemental Figure 11: Functional study of X1.**

Functional HRE-dependent reporter assays were performed in 786-O cells (deleted for *VHL* on one allele, and mutated in *VHL*-E1 on the other allele, with a consequent frameshift mutation in pVHL and X1). The results are expressed as relative Firefly luciferase activity with Renilla luciferase as an internal control. 1.0 unit denotes the basal activity of endogenous HIF-2 $\alpha$  using the HRE-luciferase reporter plasmid. The ability of wild-type and mutated X1 to downregulate Firefly luciferase activity (related to HIF activity) was compared to pVHL. By using the nomenclature of a cDNA containing VHL-E1 and E1' that would potentially encodes for a X1 protein, the X1-S179P mutant corresponds to the mutation c.340+770T>C (F1, F2, F3), the X1-L128V to c.340+617C>G (F8), X1-L138P to c.340+648T>C (F8) (Table 1).

**Supplemental Figure 12**

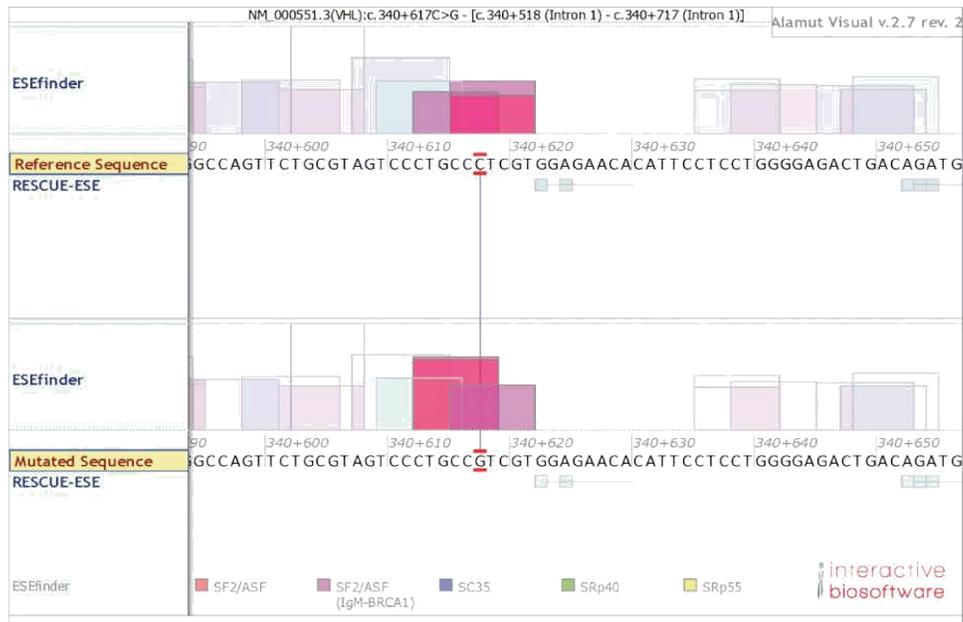


**Supplemental Figure 12: Minigene experiments using pCAS2 constructions carrying independent *VHL*-E1' variants.**

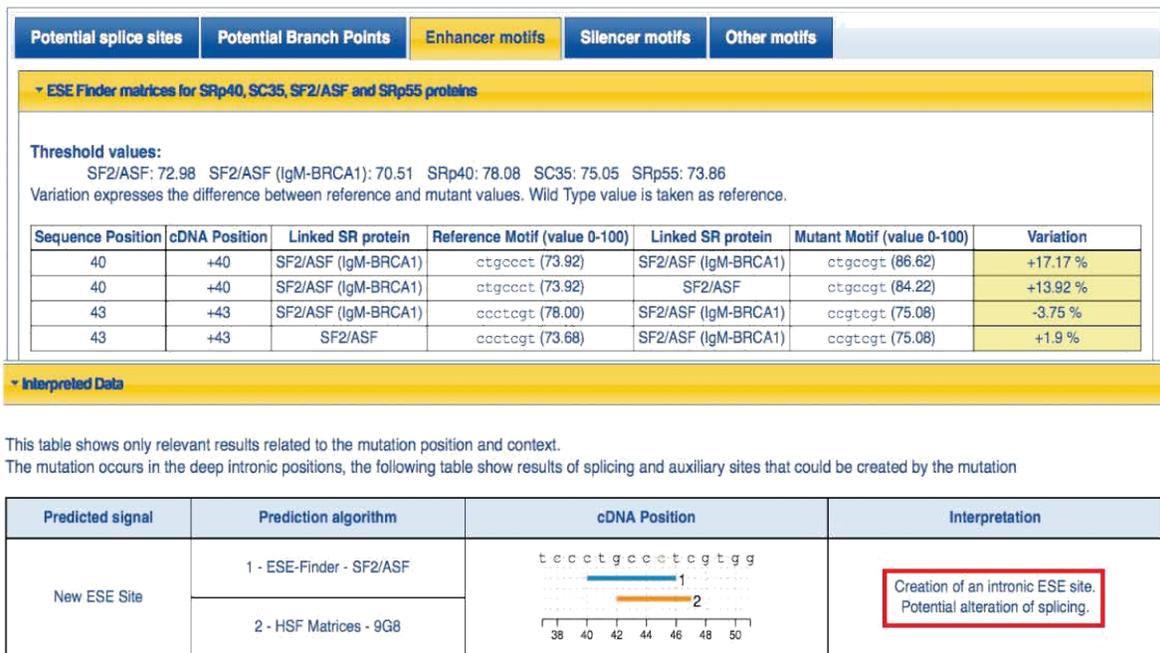
The ability to impact splicing of both E1' variants (taken independently or together) has been tested.

## Supplemental Figure 13

A



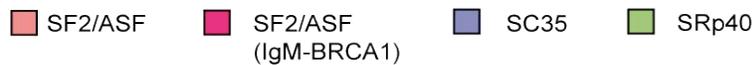
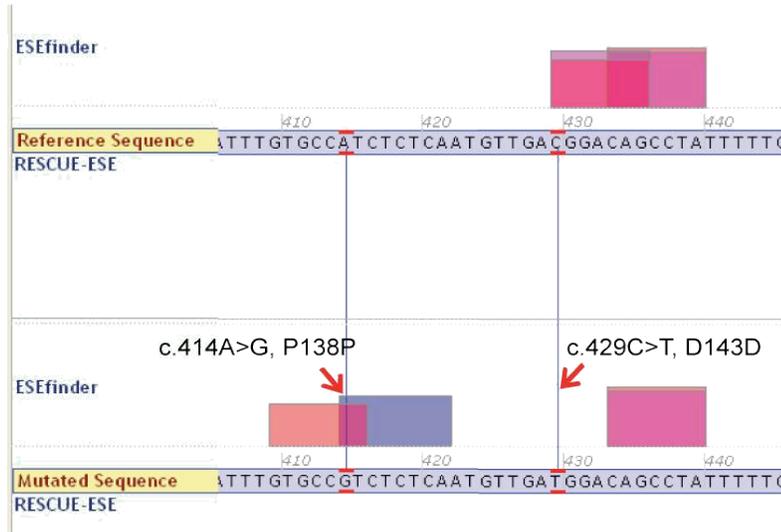
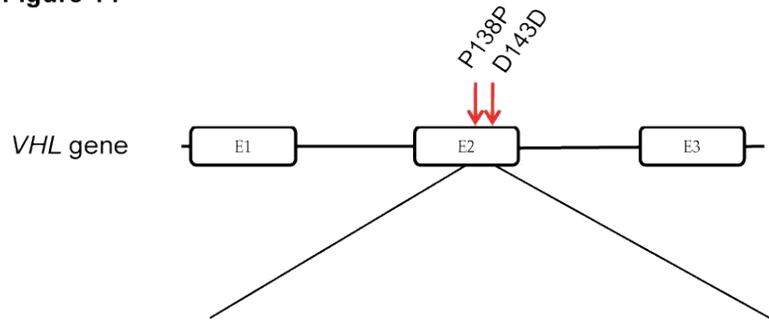
B



### Supplemental Figure 13: *In silico* analysis of exonic splicing enhancer (ESE).

The impact of the variant c.340+617C>G on splicing has been analyzed using web-based resources that identify putative ESEs responsive to the human SR proteins involved in splicing (SF2/ASF, SC35, SRp40 and SRp55) and predict whether mutations modify such elements. ESEfinder *via* Alamut program (A) and Human Splicing Finder (B) have been used.

Supplemental Figure 14



Supplementary Figure 14: *In-silico* analysis of the potential impact on splicing conferred by synonymous mutations in VHL E2.

*In silico* analysis by ESEfinder of the D143D and P138P mutations indicated differential effects on the Exonic Splicing Enhancer (ESE), with loss of the ESE in D143D, appearance of an ESE for P138P and G144R, and no effect for P138L.

**Supplemental Figure 15**

**A**

Total number and percentage of reads identified for specific *VHL* Exons splicing junctions

LCL

	E1E2 (%)	E2E3 (%)	E1E3 (%)	E1E1' (%)	E1'E2 (%)	E1'E3 (%)	Total reads
Control	<b>47.8</b>	<b>33.9</b>	2.8	2	13.5	0	251
F9 II.1	10.7	22.8	<b>55.7</b>	0.7	23	8.1	149

Pheochromocytoma

	E1E2 (%)	E2E3 (%)	E1E3 (%)	E1E1' (%)	E1'E2 (%)	E1'E3 (%)	Total reads
Control	<b>41</b>	<b>48.2</b>	7.2	1.2	1.2	1.2	83
F11 III.1	13.2	11.3	<b>73.6</b>	1.9	0	0	53

**B**

Genes	Fold change
<i>HK2</i>	7.80113
<i>EGFR</i>	6.38668
<i>EGLN3</i>	5.21586
<i>PRKCA</i>	4.67887
<i>FLT1</i>	3.89176
<i>BCL2</i>	3.75322
<i>HGF</i>	3.73589
<i>ANGPT2</i>	3.68337
<i>ENO1</i>	3.47601
<i>PFKFB3</i>	3.44818
<i>TEK</i>	3.42705
<i>SLC2A1</i>	3.34149
<i>TFRC</i>	3.10283
<i>PGK1</i>	2.95899
<i>LDHA</i>	2.91298
<i>HK3</i>	2.32731
<i>GAPDH</i>	2.18349
<i>VEGFA</i>	2.17118
<i>HK1</i>	2.02362
<i>PFKL</i>	1.9217
<i>EGLN1</i>	1.86894

**Supplementary Figure 15: RNA sequencing analyses of samples carrying mutations in *VHL*-Exon 2.**

(A) The total number and percentage of reads identified for specific junctions. (B). Transcriptome data obtained from RNA-Seq analysis of the *VHL* mutated pheochromocytoma (patient with *VHL*-P138P mutation) vs. a control pheochromocytoma (patient with *RET* mutation). Results are given in fold change for a subset of HIF target genes.

## ONLINE MATERIAL AND METHODS

**Study approval:** Informed consent for medical diagnosis and research was obtained from the patients and their relatives. This study was agreed by the CCPPRB (French Ethical Committee) Paris-Sud at Bicêtre Hospital (Le Kremlin-Bicêtre, France).

### Sanger sequencing and microsatellite analyses

The three exons and exon-intron junctions of the *VHL* gene were sequenced from DNA extracted from whole blood, as previously described.<sup>1</sup> Microsatellite analyses were performed by Multiplex PCRs (Qiagen Master Mix) using forward primers coupled with fluorescent dye designed to recognize 6 microsatellite markers across a 4.6 Mb region surrounding the *VHL* gene in 3p25-26: tel – D3S1304 – D3S1597 – D3S1317 – D3S1435 – D3S1038 – D3S1263 - cen. Primer sequences and a complete description of the protocols are available on request.

### Whole Genome Sequencing

DNA samples from four male patients with erythrocytosis (F2 II.1, F3 II.1, F3 II.2 and F7 II.1) were sequenced. For three of these patients (F2 II.1, F3 II.1, F3 II.2), a heterozygous *VHL* p.Arg200Trp mutation had been identified previously, but a second variant that might account for pathogenicity had not been found. For family 7, the patient and unaffected parental samples were sequenced. Previous tests of this trio, including Whole Exome Sequencing, did not reveal any clinically relevant variants.

Whole genome sequencing (WGS) was performed at the clinically accredited Molecular Diagnostics Laboratory at the John Radcliffe Hospital using the Hi-Seq 4000 platform (Illumina Inc., San Diego, CA) in high throughput mode.<sup>2,3</sup> The 150-bp reads generated were aligned and mapped to hs37d5 using BWA v.0.7.10 and Stampy v1.0.23. Samtools v1.3 was used to sort and merge the BAM files and Picard markduplicates v1.111 was used to remove duplicates from the bam files. The mean read depth was 30X.

### Analysis of single nucleotide variants and short insertions/deletions

Single nucleotide variants (SNV) and short insertions/deletions (InDels) were called using Platypus v0.8.1 (Rimmer et al., 2014) using default settings except for minFlank=0. Variant calling was performed together for the three patients with the heterozygous *VHL* p.Arg200Trp mutation (F2 II.1, F3 II.1, F3 II.2) and also together for the Family 7 trio. The two multi-vcf files generated were uploaded subsequently into Ingenuity Variant Analysis (IVA) software (<sup>59</sup>Qiagen, Redwood City, CA, USA), where PASS variants with

frequency of 3% or lower in 1000 genomes project, NHLBI ESP6500 exomes, ExAC and gnomAD were imported for further filtering. The initial two filtering steps used for both the three patient group and the trio were:

- 1) Selection of high confidence variants: variants with a quality score  $\geq 20$ , a read depth  $\geq 10$  and an allele frequency  $\geq 0.05$  in all samples.
- 2) Selection of rare variants in general population: variants present at an allele frequency of  $\leq 1\%$  in either the 1000 genomes project or the NHLBI ESP6500 exomes or the ExAC database or the gnomAD database.

The number of variants retained in each filtering step is given in Supplementary Figure 6A.1.

Subsequently, the filtering strategy diverged for each dataset.

#### *Subsequent filtering strategy for patients F2 II.1, F3 II.1 and F3 II.2*

For the variants selected as high confidence and rare in the general population, three filtering cascades were applied in order to capture different modes of pathogenicity:

- 1) Identification of a second variant in the *VHL* gene that could cause pathogenicity in combination with *VHL* p.R200W.
  - Selection of variants within the *VHL* gene and the 5kb region immediately upstream and downstream (see details for these variants in Supplementary Figure 6A.2.1).
- 2) Identification of variants in genes related to, or suspected to be related to, erythrocytosis.
  - Selection of variants in erythrocytosis related genes (ERY gene list): genes in erythrocytosis gene panel (Camps *et al.*, 2015) or genes classified by IVA as related to erythrocytosis and/or polycythaemia (see below for full 120 gene list).
  - Exclusion of variants with  $<3\%$  frequency in the WGS500 project union file (ie likely artefacts).
  - Selection of identical variants found in all three patients or variants found within the same gene in all three patients (same variant and genotype required for F3 II.1 and F3 II.2 but different variant allowed for F2 II.1).
  - Selection of variants predicted deleterious by IVA according to the following criteria: Pathogenic, Possibly Pathogenic OR Disease-associated according to HGMD OR established gain of function in the literature OR gene fusions OR in a microRNA binding site OR Frameshift, in-frame indel, or stop codon change OR Missense OR predicted to disrupt splicing by MaxEntScan OR deleterious to a microRNA OR structural variant OR in an ENCODE transcription factor binding site overlapping promoter region OR in enhancer OR within 6 bases into intron. (Supplementary figure 6A.2.2)

- Final evaluation of candidate variants was based on supporting literature.
- 3) Identification of variants predicted to be deleterious with high functional impact.
- Selection of variants predicted deleterious by IVA (criteria as above) and with a CADD phred score  $\geq 20$ .
  - Exclusion of variants with  $< 3\%$  frequency in the WGS500 project union file (ie likely artefacts).
  - Selection of identical variants found in all three patients or variants found within the same gene in all three patients (same variant and genotype required for F3 II.1 and F3 II.2 but a different variant allowed for F2 II.1) (criteria as above). Details for these variants are given in Supplementary figure 6A.2.3.
  - Final evaluation of candidate variants was based on supporting literature.

#### *Subsequent filtering strategy for family 7 trio*

From the variants selected as high confidence and rare in general population, variants predicted to be deleterious by IVA were selected. These variants were filtered further using the IVA Genetic analysis feature, taking into consideration inheritance models consistent with the affected and unaffected statuses of the proband and parents respectively (*de novo*, autosomal recessive, compound heterozygous or X-linked). Variants with  $\leq 3\%$  frequency in the WGS500 project union file were excluded as likely artefacts.

Visual examination and curation of the remaining variants using Integrative Genomics Viewer (IGV) (<http://software.broadinstitute.org/software/igv/>) was performed, discarding any possible artefacts observed in the BAM files. The final selection of candidate variants was based on supporting literature (see supplementary figure 6A.3.1 for variants within genes in ERY gene list). The final evaluation of candidate variants was based on supporting literature.

The identification of the cryptic exon in *VHL* gene and the lack of relevant variants from our previous global analysis prompted also a new screen for variants within this region of the *VHL* gene and the 100bp region immediately upstream and downstream, without restriction to predicted deleterious variants (see Supplementary figure 6A.3.2).

#### *Details on software and database versions*

All the analysis above was performed on Ingenuity Variant Analysis version 5.1.20171220. Content versions: CADD (v1.3), CentoMD (-), EVS (ESP6500SI-V2), Allele Frequency Community (2017-10-02), JASPAR (2013-11), Ingenuity Knowledge Base (Oz 180209.001), Vista Enhancer (2012-07), OMIM (May 26, 2017), gnomAD (2.0.1), Clinical Trials (Oz 180209.001), BSIFT (2016-02-23), TCGA (2013-09-05), PolyPhen-2 (v2.2.2), 1000 Genome

Frequency (phase3v5b), Clinvar (2017-09-07), DGV (2016-05-15), COSMIC (v82), ExAC (0.3.1), HGMD (2017.3), PhyloP (2009-11), DbSNP (150(2017-07-10)), TargetScan (6.2), SIFT4G (2016-02-23).

### *ERY gene list*

*BHLHE41, BPGM, EGLN1, EGLN2, EGLN3, EPAS1, EPO, EPOR, ADD2, APC, APOE, ARG2, ARID4A, ARNTL, ASXL1, ATN1, BARD1, BNIP3L, CALR, CBL, CBLB, CD44, CD59, CDC25A, CHEK2, CUX1, DLGAP3, DNMT3A, EZH2, F2, F5, FAT4, FOXP1, GATA1, GATA2, HAMP, HBA1/HBA2, HBB, HBZ, HIF1A, IDH1, IDH2, IFNAR1, IFNAR2, IL1RL1, IRAK2, JAK1, JAK2, JAK3, KDR, KIF17, KIT, KRAS, LCN2, LIG1, LRRC3, MECOM, MPL, MTHFR, MYBL2, MYOCD, NF1, NFE2, NFE2L2, NPRL3, NR3C1, NRAS, NTRK1, PDGFRA, PDGFRB, PIAS2, PLA2G4A, POLG, PRDX1, PRDX2, PRMT5, PTGS1, PTGS2, PTPN6, PTPRT, RAC1, RAC2, RHOA, RRM1, RRM2, RRM2B, RUNX1, SCARB1, SCRIB, SF3B1, SH2B3, SLC11A2, SLC14A1, SLC14A2, SLC30A10, SLC4A1, SOCS3, SPTA1, SPTB, SRSF2, STAT5A, STAT5B, STK11, TCF12, TET2, TP53, TRNT1, TYK2, U2AF1/U2AF1L5, VHL, ZXDC, GFI1B, HIF1AN, HIF3A, JAK2, KDM6A, OS9, ZNF197.*

### **Analysis of Copy Number Variants**

Copy Number Variants (CNVs) were identified using bam read account information and employing the ngCGH tool (<http://github.com/seandavi/ngCGH>) with a configured window size of 300 reads and the ngbin tool (Sun Microsystems Inc.) with a window size of 1000bp and a minimum read depth of 10 reads. B-allele frequency (BAF) information was obtained from vcf files generated by Platypus 0.5.2-py2.7.<sup>3</sup> Log<sub>2</sub>R and BAF outputs were imported, processed, calls flagged and visualised using Nexus Discovery Edition 8.0 (BioDiscovery, Inc., El Segundo, California, USA). The statistical approach was SNPRank Segmentation (based on Circular Binary Segmentation) and the settings used for flagging CN events and copy neutral, homozygous regions of interest (eg due to uniparental isodisomy and identity by descent) from imported ngCGH Log<sub>2</sub>R and BAF outputs were: Significance Threshold= $1 \times 10^{-6}$ ; Max Contiguous probe spacing (Kbp) = 5.0; Min number of probes per segment = 2; High Gain=1.5; Gain=0.13; Loss=-0.12; Big Loss=-1.5; homozygous Frequency threshold=0.88; Homozygous value threshold=0.8; Heterozygous Imbalance Threshold = 0.4; Minimum LOH length (KB) = 1000; Minimum SNP probe Density (Probes/MB)=0.0; Percent outliers to remove=3.0. The calls were curated to remove CNVs occurring commonly in the Database of Genomic Variants (<http://dgv.tcag.ca/dgv/app/home>), regions of homozygosity <2Mb (ie within the normal population size range distribution) and frequently encountered signal ‘artefacts’ (eg due to segmental duplication and repeat sequence regions). The remaining calls were annotated

comprehensively and used to assess pathogenicity either individually or in combination with sequence level variants.

### **Cell culture and puromycin test**

Cell lines 786-O, HEK293-T, and HK2 were maintained in DMEM supplemented with 10% FCS, 1 % L-Glutamine, and 1% penicillin/streptomycin. PC12 cells were maintained in DMEM containing 10% FCS, 1 % L-Glutamine, 1% penicillin/streptomycin, and 5% horse serum. UT-7 cells were maintained in  $\alpha$ -MEM supplemented with 20% FCS and 1 $\mu$ g/mL EPO. Cells cultured with EPO were used for minigene experiments following validation of EPO-R expression by flow cytometry.

Lymphoblastoid cell lines (LCL) were maintained in RPMI supplemented with 10% FCS, 1% L-Glutamine, and 1% penicillin/streptomycin. LCLs from each patient were grown and harvested independently three times for triplicate experiments.

A puromycin test was performed on LCLs. For each LCL, two separate flasks were seeded, with puromycin (InvivoGen) added to one flask at a final concentration of 200 $\mu$ g/ml with an overnight incubation. Cells were harvested the following day. The experiment has been repeated three independent times.

### **Transcript detection and quantification**

mRNAs were extracted from cells or tumor tissue using the NucleoSpin® RNA kit (Macherey-Nagel) according to the manufacturer's instructions. The extraction steps include a DNase treatment. Reverse transcription reactions were carried out with 1 $\mu$ g of RNA using the Maxima First Strand cDNA Synthesis Kit (ThermoScientific). Human Multiple panel cDNAs (Human MTC Panel I, Ozyme) were used to quantify different *VHL* isoforms in human tissue. Exon-specific PCR was performed using primers localized to flanking exons (Primer in E1: CAGGTCATCTTCTGCAATCG, Primer in E3: TTCGTAGAGCGACCTGACGATGTC). Taqman real-time PCR were performed on 20 ng of cDNA with the qPCR Mastermix (Eurogentec) and run on Rotor-Gene Q (Qiagen, Santa Clara, CA, USA). Primer sequences are available on request. The RPLP0 probe was used as internal control. The thresholds have been determined by using dilutions of plasmids containing coding sequences of each gene.

### **RNA sequencing**

Total RNA was quantified using a Nanodrop® ND-1000 spectrophotometer (Thermo Scientific). The quality and integrity of RNA samples were assessed using the 2100 Bioanalyzer and RNA

6000 Nano LabChip kit series II (Agilent Technologies). Library construction was performed with 500ng of total RNA with SureSelect Strand-Specific RNA Library Prep for Illumina Multiplexed kit (Ref 5190-6410, Agilent Technologies) according to the Agilent\_PrepLib\_G9691-90010\_juillet2015\_vD protocol. Purifications were carried out with NucleoMag NGS Clean-up and Size Select (Ref 744970.50, Macherey-Nagel). The fragment size of libraries was controlled using D1000 ScreenTape with the 2200 TapeStation system (Agilent Technologies). Libraries with P5-P7 adaptors were specifically quantified using a LightCycler<sup>®</sup> 480 Instrument II (Roche Life Science) and normalized with DNA Standards (Ref KK4903, KAPABIOSYSTEMS - CliniSciences). Ten pM of each library was pooled and prepared according to the denaturing and diluting libraries protocol for the HiSeq and GAIIx (part#15050107 v02 (Illumina)) for cluster generation on the cBot<sup>™</sup> system. Paired-end sequencing (2x100 cycles) was carried out in a single lane on the HiSeq<sup>®</sup> 2500 system (Illumina) using TruSeq v3 chemistry according to the manufacturer's instructions for the HiSeq<sup>®</sup> 2500 (System Guide, part#15035786 v01 (Illumina)).

Short reads were processed with cutadapt and then mapped to the human genome GRCh37 with STAR<sup>4</sup> (STAR\_2.5.1b\_modified) using a modified Ensembl Human Gene transfer format file (GTF v. GRCh37.82) to which some records were added (for the new potential junctions).

## Reporter assays

The plasmid encoding wild type HA-VHL was a kind gift of Prof William G. Kaelin Jr. An expression plasmid for the hypothetical X1 protein was constructed by subcloning the coding sequence following its synthesis by GeneArt (LifeTechnologies). Mutations of interest were introduced with either a custom protocol using Taq Phusion (ThermoScientific) or the Q5 Site-directed Mutagenesis Kit<sup>®</sup> (New England Biolabs), used according to the manufacturer's instructions. Following transformation into chemically competent bacteria, mutated plasmids were extracted with the NucleoSpin<sup>®</sup> Plasmid kit (Macherey-Nagel), sequenced, amplified and harvested with the Nucleobond<sup>®</sup> Xtra Maxi Plus EF kit (Macherey-Nagel).

Transfections were performed in 786-O cells (deleted for *VHL* on one allele, and mutated in *VHL*-E1 on the other allele, with a consequent frameshift mutation in pVHL and X1) using the jetPRIME<sup>®</sup> reagent (Polyplus), with constructs encoding wild-type or mutant protein, 200ng Firefly luciferase plasmid with luciferase under the control of Hypoxia Response Elements (3xHRE-luciferase plasmid), 20ng Renilla luciferase for normalization and empty pcDNA3 vector to give a total of 400ng DNA.<sup>5,6</sup> Three independent assays were performed. Empty vector was transfected in each assay for normalization. After 24h transfection, luciferase assays were

performed using a Dual-Luciferase® Reporter Assay System (Promega) on a single automatic injection Mithras luminometer (Berthold technologies, Bad Wildbad) according to the manufacturer's protocol. Ratios of Firefly luciferase readings to Renilla luciferase readings were normalized to the activity of the empty construct.

### **Western Blotting**

Western blotting was performed with the NuPAGE® kit from Invitrogen, according to the manufacturer's instructions. Briefly, 18µL of cell lysates from luciferase reporter were added to 4X Sample Loading Buffer and 10X Reducing Agent, then loaded into a Bis-Tris Mini Gel (4-12%). Migration was performed in MES SDS Running Buffer. Protein transfer to a nitrocellulose membrane (0.2 µM Hybond ECL; GE Healthcare) was in Invitrogen Transfer Buffer with 10% ethanol at 30V for 1h. The membrane was subsequently incubated with a mouse anti-HA antibody (dilution 1:1000, clone 16B12, BioLegend), and then a goat anti-mouse HRP-conjugated secondary antibody (dilution 1:10000, Jackson Immuno Research). Bound antibody was then revealed using the Clarity™ Western ECL Substrate from BioRad. Western blot using the mouse monoclonal antibody JD-1956 (Patent No. 14305925.1-1402-2014 ;CNRS-EFS) raised against human VHL was performed as described by Chesnel et al., 2015.<sup>7</sup>

### **Minigene experiments**

Minigene constructs were prepared according to Cooper and Gaildrat.<sup>8,9</sup> Briefly, mutant fragments containing the exon of interest, along with flanking intron sequences, were generated by PCR amplification from patient genomic DNA, using oligonucleotides engineered to contain the BamHI and MluI restriction enzyme sites at their 5' end. Following a double digestion and gel purification, the PCR product was ligated into a pCas2 plasmid containing two artificial exons from *SERPING* gene (A and B) that had been cut using BamHI/MluI and dephosphorylated. DH5α-competent bacteria were then transformed with the plasmids. Clones from each transformation were collected and amplified, with plasmid DNA harvested using the NucleoSpin® Plasmid kit (Macherey-Nagel). Plasmid DNA was then sequenced, and clones carrying the mutated plasmids selected. The cloned intron sequences flanking *VHL*-Exon 1' or *VHL*-Exon 2 contain approximately 480 base pairs (bps) and 120 bps respectively.

Cells were transfected with 2 µg of each pCas2 plasmid containing the exon of interest, wild type sequence, or mutant. Chemical transfection was performed with the jetPRIME® reagent (Polyplus) according to the manufacturer's instructions.

UT-7 human megakaryoblastic leukemia cells and LCLs were transfected by nucleofection using the Amaxa technology (Lonza). Transfection efficiency was assessed using 2 $\mu$ g of a GFP-containing plasmid. 24 hours after transfection, cells were harvested and total RNA extracted using the NucleoSpin® RNA kit (Macherey-Nagel) according to the manufacturer's instructions. RT-PCR was performed using the Maxima First Strand cDNA Synthesis Kit (ThermoScientific), with 0.5 $\mu$ g RNA. PCR amplification was performed with 50ng cDNA, using the PCR GoTaqQ2 kit (Promega) with primers against both artificial (A and B) exons. PCR products were resolved on a 2% agarose gel.

## References

1. Couve S, Ladroue C, Laine E, et al. Genetic evidence of a precisely tuned dysregulation in the hypoxia signaling pathway during oncogenesis. *Cancer Res.* 2014;74(22):6554-6564.
2. Camps C, Petousi N, Bento C, et al. Gene panel sequencing improves the diagnostic work-up of patients with idiopathic erythrocytosis and identifies new mutations. *Haematologica.* 2016;101(11):1306-1318.
3. Rimmer A, Phan H, Mathieson I, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet.* 2014;46(8):912-918.
4. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.
5. Dayan F, Roux D, Brahim-Horn MC, Pouyssegur J, Mazure NM. The oxygen sensor factor-inhibiting hypoxia-inducible factor-1 controls expression of distinct genes through the bifunctional transcriptional character of hypoxia-inducible factor-1 $\alpha$ . *Cancer Res.* 2006;66(7):3688-3698.
6. Ladroue C, Hoogewijs D, Gad S, et al. Distinct deregulation of the hypoxia inducible factor by PHD2 mutants identified in germline DNA of patients with polycythemia. *Haematologica.* 2012;97(1):9-14.
7. Chesnel F, Hascoet P, Gagne JP, et al. The von Hippel-Lindau tumour suppressor gene: uncovering the expression of the pVHL172 isoform. *Br J Cancer.* 2015;113(2):336-344.
8. Cooper TA. Use of minigene systems to dissect alternative splicing elements. *Methods.* 2005;37(4):331-340.
9. Gaildrat P, Killian A, Martins A, Tournier I, Frebourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol.* 2010;653:249-257.

#### d. Résultats complémentaires

##### i. Etude fonctionnelle dans E1'

Dans l'article du journal *Blood*, nous avons décrit une famille porteuse d'une mutation homozygote de E1' c.340+816A>C qui développe une polyglobulie (Famille 7). Les deux parents sont des porteurs hétérozygotes sains. Nous avons seulement réalisé une étude de cette mutation par rapporteur minigène pour lequel nous avons généré le plasmide par mutagenèse dirigée car nous ne disposions pas de matériel issu des patients. Depuis, nous avons généré une lignée lymphoblastoïde du patient porteur homozygote.

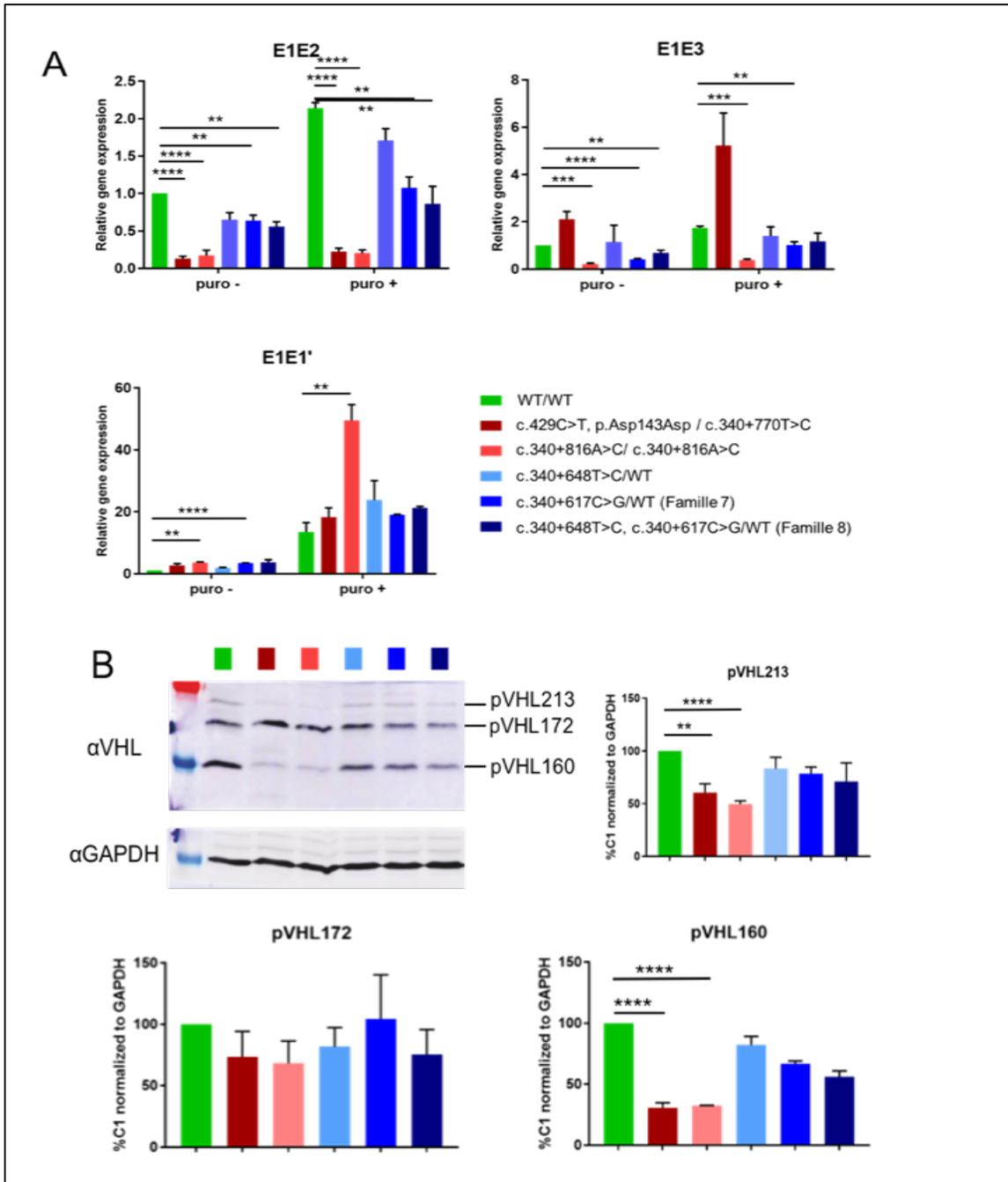
Nous avons également décrit une famille porteuse des mutations hétérozygotes de E1' c.340+617C>G et c.340+648T>C (Famille 8) sur le même allèle et qui développe la maladie de VHL. Nous avons décrit brièvement un cas de patient porteur uniquement de la mutation c.340+617C>G qui développe un hémangioblastome isolé mais pour lequel nous n'avons pas de matériel. Depuis, nous avons obtenu des LLB générées à partir de PBMC de ce patient ainsi que des LLB d'un patient porteur du variant c.340+648T>C seule. L'étude de ces trois cas nous permet de mieux comprendre les mécanismes à l'origine de la pathologie, notamment pour la famille porteuse des deux mutations.

Nous avons réalisé une analyse de l'expression des différentes isoformes de *VHL* ainsi qu'une analyse d'expression protéique avec un anticorps dirigé contre l'exon 1 de VHL (Figure 16).

L'étude transcriptionnelle des différentes isoformes de *VHL* dans les LLB de patients nous permet de voir que la diminution de E1E2 observée pour la famille 8 est due à une accumulation des effets des 2 mutations (Figure 16A). En effet, nous observons un léger gradient de diminution de l'isoforme sauvage pour les 3 mutations. Il en est de même pour l'accumulation de E1' en absence de puromycine qui inhibe le NMD. Au niveau protéique, les différences observées pour les 3 patients ne sont pas significatives, une seule mutation suffit pour obtenir la même perte de fonction que lorsque le patient est porteur des deux mutations de E1' simultanément (Figure 16B). Un léger gradient est tout de même observé pour l'isoforme pVHL160.

Nous avons également étudié un patient polyglobulique porteur d'une mutation homozygote de E1' VHL-c.340+816A>C. Nous pouvons voir qu'à la différence d'une mutation hétérozygote de E1',

cette mutation homozygote conduit à une forte diminution des transcrits contenant E1E2, E1E3 et E2E2' au profit des transcrits contenant l'exon E1', comme nous avons pu le constater avec les autres patients homozygotes de l'article (Figure 16A). Cette modification de l'expression des transcrits conduit à une forte diminution de l'expression protéique totale de VHL, significative pour pVHL213 et 160 (Collaboration Y.Arlot, A. Couturier, Rennes<sup>59</sup>) (Figure 16B).



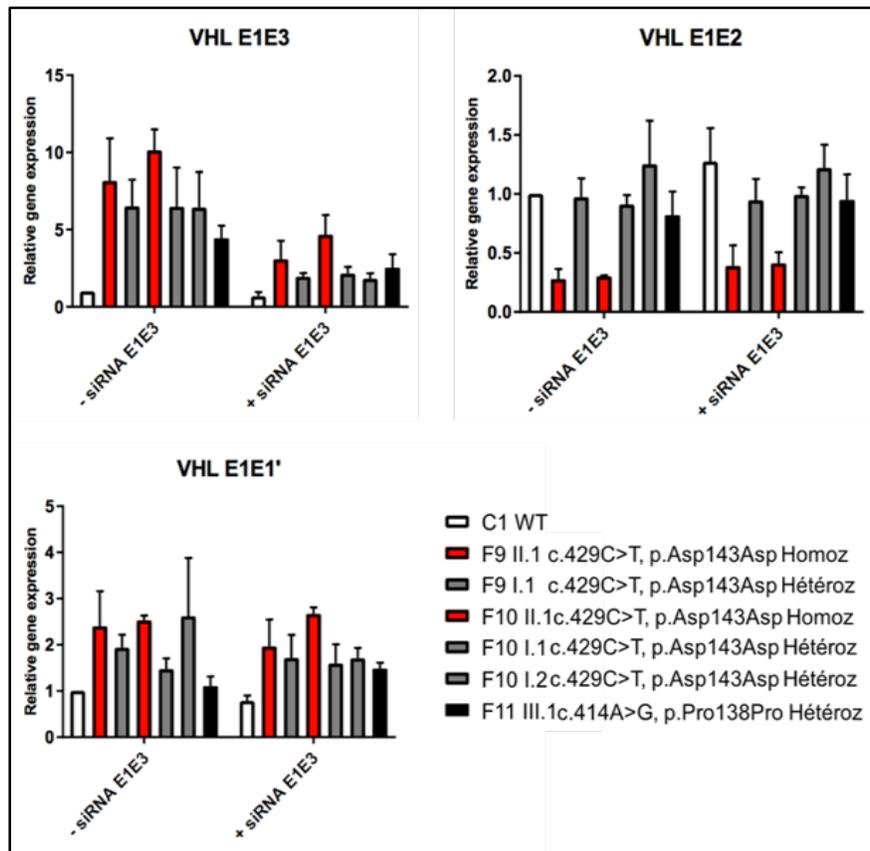
**Figure 16 : Etude d'expression des mutations de E1'.** Les patients en rouge et rouge pâle développent une polyglobulie et ceux en bleu la maladie de VHL, le vert correspond au contrôle. (A) Quantification par RT-qPCR TaqMan des différents transcrits de *VHL* avec des sondes spécifiques des jonctions E1E2, E1E3 et E1E1' dans les LLB de patients cultivées en absence (puro -) ou en présence (puro +) de puromycine, un inhibiteur du nonsense mediated decay. Les quantifications ont été réalisées en duplicat sur les cellules cultivées trois fois indépendamment. Les données sont représentées sous la forme de la moyenne +/- l'erreur standard à la moyenne (SEM) et sont normalisées par rapport aux LLB contrôles (WT) fixées à 1. (B) Analyse par western blot des extraits protéiques des LLB de patients cultivées 3 fois de façon indépendantes. Un blot représentatif des trois western blots réalisés est présenté et les quantifications sont représentées sous forme de la moyenne +/- l'erreur standard à la moyenne (SEM) avec les LLB contrôles fixées à 100% pour chaque isoforme. Les statistiques sont réalisées avec des *t.tests* : \*\*<0,001, \*\*\*<0,0001, \*\*\*\*<0,00001.

ii. *Etude fonctionnelle des isoformes induites par les variants dans l'intron 1 et l'exon 2 par utilisation d'ARN interférent*

Nos résultats nous ont permis de montrer deux nouvelles altérations de *VHL* à savoir un saut de l'exon 2 ou une rétention de l'exon cryptique E1' dans les transcrits. Nous avons pu voir qu'un taux de transcrits E1E3 élevé entraîne une diminution de l'expression de l'isoforme sauvage E1E2E3, ce qui conduit à une diminution de l'expression protéique totale de VHL. De la même manière, un taux élevé de l'isoforme contenant E1E1' est aussi associé à une diminution de l'expression de l'isoforme E1E2E3 au niveau transcriptionnel qui est associé à une diminution de l'expression protéique totale de VHL.

Nous avons voulu tester si une diminution des transcrits E1E3 et E1E1' grâce à des siRNA pourrait permettre un rétablissement de l'expression de E1E2E3 qui conduirait à une expression normale des protéines VHL totales.

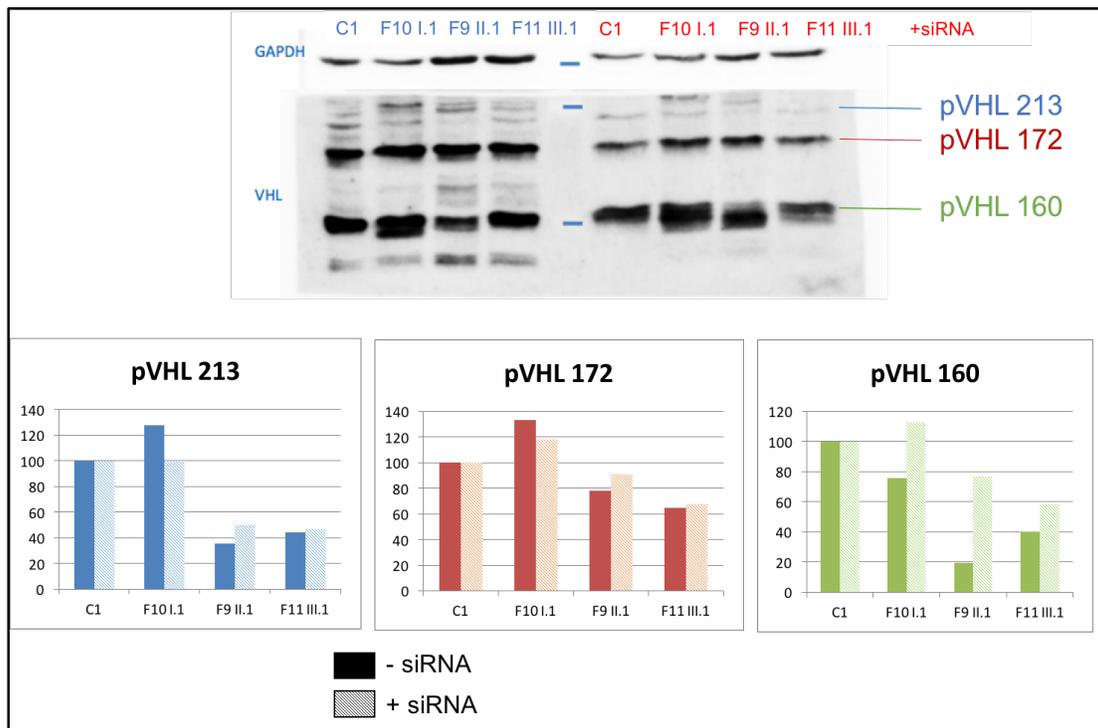
Nous avons utilisé un siRNA dirigé contre la jonction E1E3 pour cibler l'isoforme  $\Delta E2$  de *VHL* chez les patients porteurs d'une mutation synonyme de *VHL* dans l'exon 2 de l'isoforme E1E2E3 (Figure 17).



**Figure 17** : Expression des isoformes de *VHL* dans des LLB de patients transfectées avec des siRNA dirigés contre la jonction Exon1/Exon3 de *VHL*. Des lignées lymphoblastoïdes contrôles et de différents patients porteurs de mutations synonymes de l'exon 2 de *VHL* ont été transfectées avec un siRNA dirigé contre la jonction E1E3 de *VHL*. Des RT-qPCR Taqman spécifiques ont été réalisées pour quantifier les différentes isoformes de *VHL*. Les barres blanches correspondent aux LLB contrôles, les rouges correspondent à des LLB de patients qui développent une polyglobulie, les barres grises à des LLB de patients porteurs d'une mutation hétérozygote mais ne développant aucune pathologie et les barres noires à des LLB de patients qui développent la maladie de *VHL*. Les résultats sont exprimés en expression relative normalisées avec les LLB C1 non traitées (NT) fixées à 1.

Comme nous pouvons le voir sur la figure 2, le siRNA dirigé contre E1E3 diminue l'expression de l'isoforme  $\Delta E2$  de *VHL*. En revanche, cela ne permet pas de restaurer l'expression de l'isoforme sauvage de *VHL* dans les LLB de patients. Nous avons également réalisé des western blots sur des extraits protéiques de LLB transfectées ou non avec les siRNA dirigés contre E1E3 afin de vérifier que les résultats observés au niveau des ARNm sont également observables au niveau protéique (Figure 18). Pour cela, nous avons utilisé un anticorps dirigé contre l'exon 1 de *VHL* qui peut ainsi

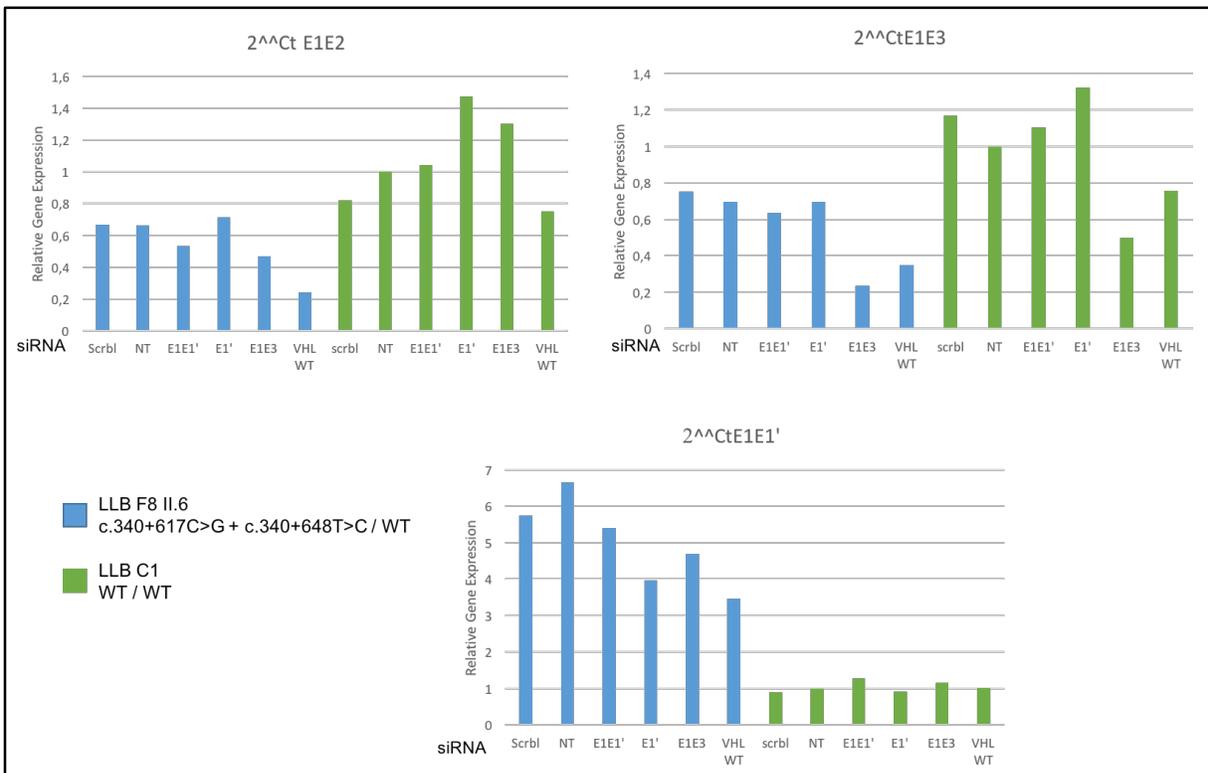
reconnaitre toutes les isoformes de pVHL<sup>59</sup>. Nous pouvons observer que dans les LLB de patients qui développent soit une polyglobulie (F9 II.1), soit la maladie de VHL (F11 III.1), il y a une diminution de l'expression des isoformes pVHL213, pVHL160 et également pVHL172 par rapport aux LLB contrôle et de porteur sain d'une mutation (F10 I.1) comme décrit dans l'article Blood<sup>220</sup>. Nous pouvons voir que dans les LLB transfectées avec un siRNA E1E3, il n'y a pas de modification significative de l'expression des isoformes pVHL 213 et 172 mais y a une augmentation de l'expression de l'isoforme pVHL160 par rapport aux cellules non traitées avec un siRNA.



**Figure 18 : Expression protéique des isoformes de VHL dans des LLB de patients.** Western blot réalisé sur des extraits protéiques de LLB contrôle, d'un patient sain porteur d'une mutation synonyme hétérozygote de l'exon 2 de *VHL* (F10 I.1), d'un patient porteur d'une mutation synonyme homozygote de l'exon 2 de *VHL* et développant une polyglobulie (F9 II.1) et d'un patient porteur d'une mutation hétérozygote de l'exon 2 de *VHL* et développant la maladie de VHL (F11 III.1). Les quantifications sont présentées sous forme de pourcentage par rapport aux LLB contrôle traitées (NT) ou non avec le siRNA E1E3 pour les isoformes pVHL 213 (WT), 172 et 160.

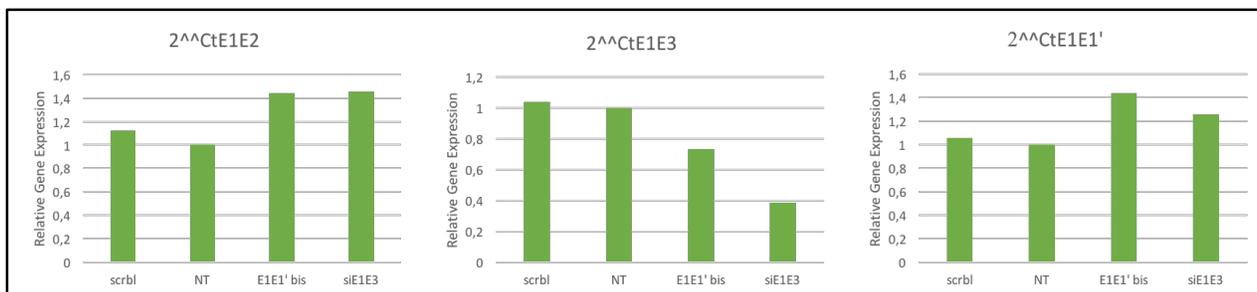
Nous avons construit un siRNA dirigé contre l'exon cryptique E1' (situé au milieu de l'exon, appelé siRNA E1') et un dirigé contre la jonction entre l'exon 1 et l'exon cryptique E1' (appelé siRNA

E1E1'). Nous avons transfecté ces siRNA dans des lignées lymphoblastoïdes contrôles (LLB C1) et d'un patient d'une famille porteuse de la mutation c.340+617C>G dans l'exon cryptique E1' associée au SNP c.340+648T>C dans E1' (Famille 8 de l'article Blood patient II.6). Des siRNA dirigés contre VHL WT (situé dans l'exon 2 de l'isoforme E1E2E3) et VHL E1E3 (situé à la jonction entre l'exon 1 et l'exon 3) ont également été utilisés pour valider la manipulation. 24h après transfection, les ARN des cellules ont été récoltés et des RT-qPCR spécifiques des différentes isoformes de VHL ont été réalisées avec des sondes Taqman dirigées contre des jonctions exon/exon (Figure 19). Nous pouvons voir que les 2 siRNA dirigés contre l'exon E1' n'entraînent pas de diminution de l'expression de l'isoforme de VHL contenant la jonction E1E1' alors que le siRNA dirigé contre la jonction E1E3 entraîne une forte diminution de l'expression de l'isoforme  $\Delta E2$  de *VHL*. Le siRNA dirigé contre l'isoforme WT montre des résultats mitigés avec une faible diminution de l'isoforme E1E2E3 de VHL comparée au siRNA contrôle (Scrb1).



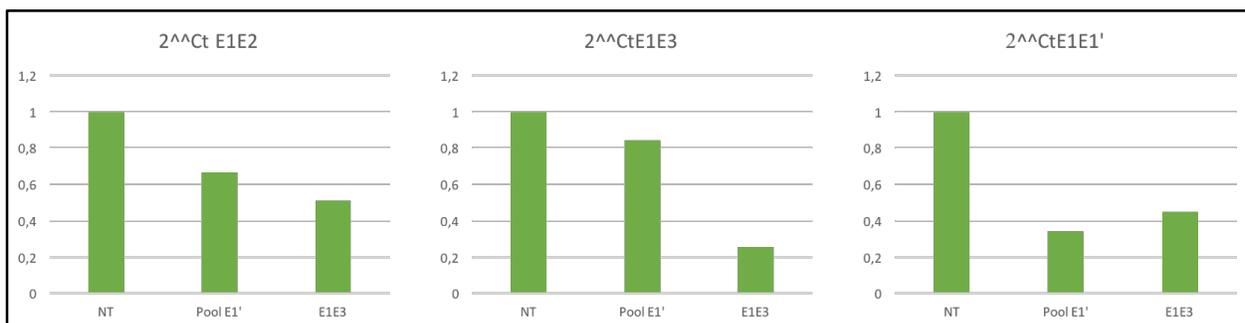
**Figure 19 : Expression des isoformes de *VHL* dans des LLB transfectées avec des siRNA.** Des lignées lymphoblastoïdes contrôles C1 et issues d'un patient de la famille 8 porteuse d'une mutation hétérozygote de E1' ont été transfectées avec des siRNA dirigés contre la jonction E1E1', E1', la jonction E1E3 ou l'exon 2 de *VHL* (*VHL* WT) ainsi qu'avec un siRNA contrôle (*scrbl*). Les résultats des RT-qPCR réalisées avec des sondes Taqman dirigées contre la jonction E1E3, la jonction E1E2 la jonction E1E1' sont représentés en expression relative normalisées avec les LLB contrôle C1 non traitées (NT) fixées à 1.

Nous avons donc cherché à construire un autre siRNA spécifique de l'exon cryptique E1' de *VHL* que nous avons nommé siRNA E1' bis. De la même manière, nous l'avons transfecté dans des LLB contrôles C1 et avons étudié l'expression des différentes isoformes de *VHL* par des RT-qPCR spécifiques (Figure 20). Ce siRNA ne permet pas non plus de diminuer l'expression de l'isoforme de *VHL* contenant l'exon cryptique E1'.



**Figure 20 : Expression des isoformes de *VHL* dans des LLB contrôles transfectées avec différents siRNA.** Des lignées lymphoblastoïdes contrôles C1 ont été transfectées avec des siRNA dirigés contre E1', la jonction E1E3 ou contrôle (Scrbl). Des RT-qPCR spécifiques des jonctions E1E1' et E1E3 ont ensuite été réalisées avec des sondes Taqman situées sur les jonctions exon/exon. Les résultats sont exprimés en expression relative normalisées avec les LLB C1 non traitées (NT) fixées à 1.

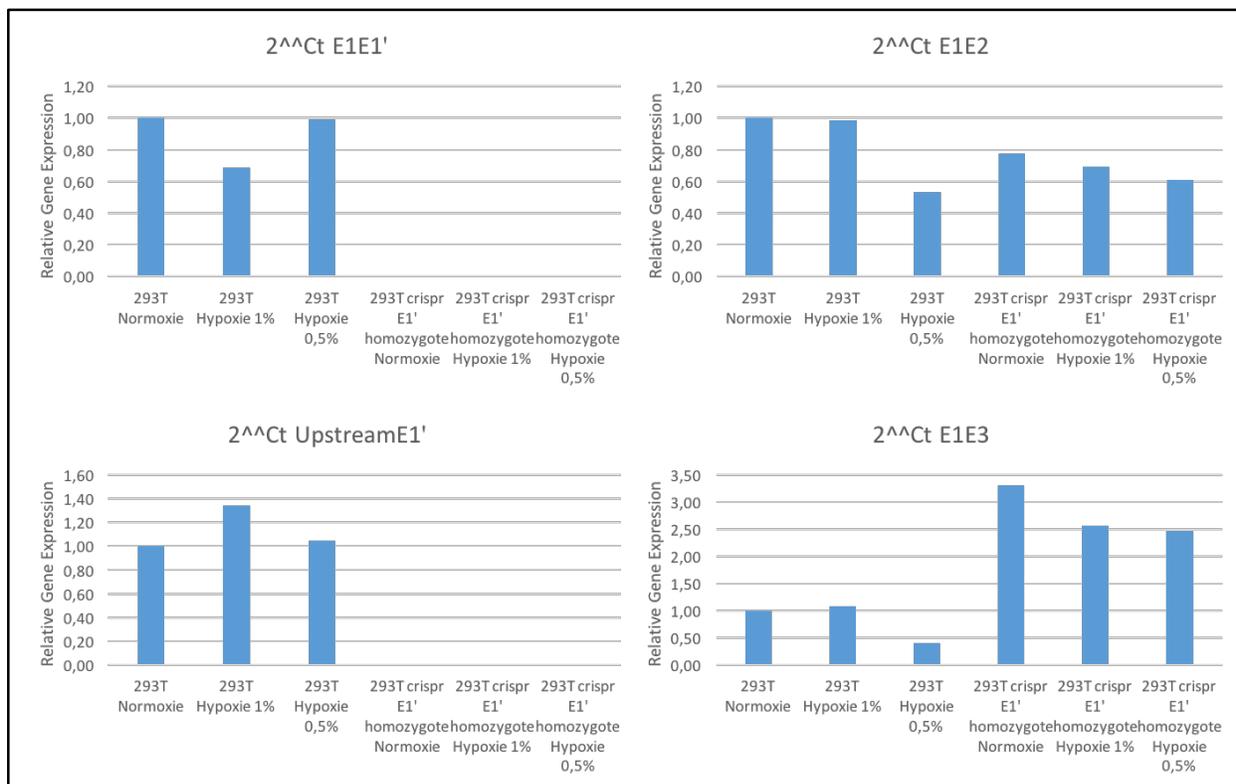
Nous avons alors essayé de transfecter un pool des 3 différents siRNA dirigés contre E1' designés (Figure 21). Nous pouvons voir que ce pool de siRNA conduit à une diminution de l'expression de l'isoforme de *VHL* contenant l'exon E1'. Cependant, nous pouvons voir que cette inhibition n'est pas spécifique puisque ce pool de siRNA conduit également à une inhibition de l'isoforme WT et  $\Delta E2$  de *VHL*



**Figure 21 : Expression des isoformes de *VHL* dans des LLB contrôles transfectées avec des siRNA.** Des lignées lymphoblastoïdes contrôles C1 ont été transfectées avec des siRNA dirigés contre E1' ou la jonction E1E3. Des RT-qPCR spécifiques des jonctions E1E1', E1E2 et E1E3 ont ensuite été réalisées avec des sondes Taqman situées sur les jonctions exon/exon. Les résultats sont exprimés en expression relative normalisées avec les LLB C1 non traitées (NT) fixées à 1.

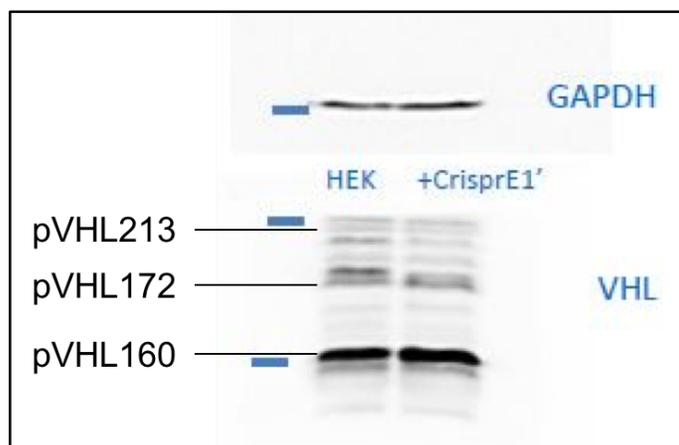
*iii. Etude fonctionnelle des isoformes induites par les variants dans l'intron 1 par utilisation de la méthode CRISPR/CAS9*

Afin de comprendre au mieux le rôle de l'exon cryptique E1', nous avons généré des lignées cellulaires Knock Out (KO) pour cet exon. Nous avons dans un premier temps réalisé cette manipulation dans une lignée de cellules embryonnaires rénales, les HEK293T (collaboration avec S. Couvé, Institut Gustave Roussy). Nous avons obtenu un clone KO homozygote pour E1' dans lequel nous avons analysé l'expression des isoformes de *VHL* par des RT-qPCR TaqMan spécifiques des jonctions E1E1', E1E2 et E1E3 après culture des cellules en normoxie ou après 24h d'incubation à 1% ou à 0,5% d'O<sub>2</sub>. Nous pouvons voir que dans les CRISPR KO homozygotes E1', il y a perte totale d'expression des transcrits contenant E1' (Figure 22). Cela conduit à une légère diminution de l'expression des transcrits contenant la jonction E1E2 et conduit à une augmentation de l'expression des transcrits contenant la jonction E1E3.



**Figure 22** : Analyse transcriptionnelle des isoforme de *VHL* dans les HEK293T sauvages ou KO homozygote pour *E1'*. Les quantifications ont été réalisées par RT-qPCR TaqMan avec des sondes spécifiques des jonctions des exons de *VHL*. Les cellules ont été cultivées en normoxie ou incubées 24h à 1 ou 0,5% d'O<sub>2</sub>.

Nous avons également réalisé un western blot sur un extrait protéique des HEK293T KO homozygote pour l'exon *E1'* (Figure 23). Nous pouvons voir qu'il n'y a pas de modification de l'expression protéique totale de *VHL* par rapport aux cellules sauvages.



**Figure 23 : Expression protéique de VHL dans les HEK293T sauvages ou KO homozygotes pour E1'.** Le western blot a été réalisé avec un anticorps dirigé contre l'exon 1 de VHL capable de reconnaître les protéine pVHL213, 172 et 160. La normalisation est faite grâce à un anticorps dirigé contre la GAPDH.

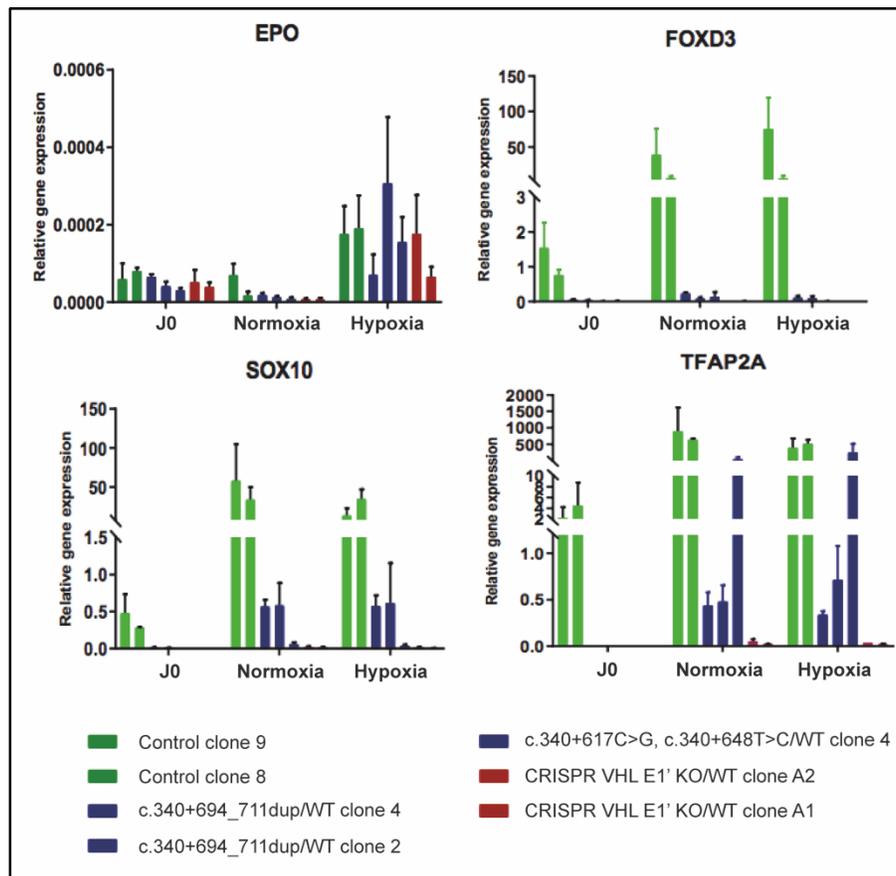
Les analyses réalisées sur les HEK293T nous ont permis de valider le guide utilisé pour la génération des KO pour E1'. Nous avons donc réalisé cette manipulation dans des iPS contrôles (Collaboration avec K. Si Tayeb, A. Caillaud, l'institut du thorax, Nantes). Après tri des clones qui ont insérés le guide CRISPR et la Cas9, puis amplification et séquençage, nous avons obtenu 23 clones (Figure 24) : des lignées KO hétérozygotes et des lignées KO homozygotes pour E1'. Les clones hétérozygotes ont pu être utilisés pour réaliser des différenciations en progéniteurs de la crête neurale (modèle de cellules productrices d'EPO décrit en partie 4 des résultats) Nous n'avons pas réussi pour le moment à différencier les lignées homozygotes en progéniteurs de la crête neurale. L'obtention de nombreux clones pour un premier essai de CRISPR pour le gène *VHL* est de bon présage pour la génération d'autres CRISPR pour ce gène, que ce soit des KO ou de KI (insertion d'une mutation homozygote ou hétérozygote).

	Clone	Allèle 1	Allèle 2	Ccl
CRISPR 1	A1	WT	E1' KO	Hétérozygote
	A2	WT	E1' KO	Hétérozygote
	A4	à reconfirmer		
	A5	séq non lisible		
	A6	E1' KO	E1' KO	Homozygote KO E1'
	A8	à reconfirmer		
CRISPR 2	A1	WT	?	
	A2	E1' KO	E1' KO	Hétérozygote composite
	A5	E1' KO	E1' KO	Homozygote KO E1'
	A6	WT	?	
	A8	E1' KO	E1' KO	Hétérozygote composite
	A9	E1' KO	?	
	A10	E1' KO	E1' KO	Hétérozygote composite
	A11	E1' KO	E1' KO	Homozygote KO E1'
	A12	E1' KO	E1' KO	Homozygote recombiné avec?
	A16	E1' KO	Pas dans E1'	
	A17	WT	?	
	A19	WT	E1' KO	Hétérozygote
	A20	WT	WT	Homozygote WT
	A21	E1' KO	WT + 2 A	Hétérozygote muté
	A23	WT	?	
A26	E1' KO	E1' KO	Hétérozygote composite	
A28	?	?		

**Figure 24 : Séquençage des clones iPS CRISPRE1' obtenus.** 23 clones CRISPR pour l'exon cryptique E1' ont été obtenus. Les résultats du séquençage de chaque clone sont présentés dans le tableau pour chaque allèle. Les ? correspondent à une séquence non identifiée. Seuls les clones avec une conclusion claire ont été sauvegardés et amplifiés.

Nous avons différencié en progéniteurs de la crête neurale deux clones CRISPR hétérozygotes KO pour E1' (CRISPR1 cA1 et A2) ainsi qu'un clone d'une lignée iPS dérivée d'un patient qui développe la maladie de VHL porteur des mutations de E1' c.340+617C>G et c.340+648T>C à l'état hétérozygote sur le même allèle, de deux clones d'un patient sain porteur de la mutation hétérozygote de E1' c.340+694\_711dup (le patient porteur homozygote développe une polyglobulie) et de deux clones contrôles (Figure 25). Nous espérons pouvoir étudier l'impact des mutations de E1' sur le développement des pathologies grâce à ce modèle. Nous avons eu plus

de difficultés à différencier ces cellules, ce qui est visible au niveau de l'expression des marqueurs de différenciation, principalement pour les iPS CRISPR KO hétérozygotes pour E1' (Figure 25). Nous n'observons aucune différence significative pour la production d'EPO entre les cellules porteuses d'une mutation E1' ou KO E1' et les cellules contrôles.



**Figure 25** : Quantification des marqueurs spécifiques des progéniteurs neuronaux et de l'EPO. Une quantification par RT-qPCR TaqMan des transcrits des marqueurs spécifiques des progéniteurs de la crête neurale est réalisée grâce à des sondes spécifiques pour chaque gène. L'EPO exprimée dans ces cellules est également quantifiée. Les résultats représentent la moyenne des  $2^{(-\Delta\Delta Ct)}$  de 3 différenciations réalisées indépendamment et quantifiées en duplicats +/- l'erreur standard à la moyenne (SEM). A la fin de la différenciation, les cellules sont cultivées 24h en normoxie ou en hypoxie à 1% d'O<sub>2</sub>.

#### e. Discussion

Nous avons pu montrer que les siRNA E1' testés ne permettent pas de diminuer de façon spécifique l'expression de l'isoforme de *VHL* contenant l'exon cryptique E1'. Même si l'approche

semble intéressante pour une potentielle thérapie ciblée chez les patients, le design du siRNA dirigé spécifiquement contre cet exon reste compliqué. Comme nous n'arrivons pas à détecter l'isoforme contenant l'exon cryptique E1' au niveau protéique, nous ne pouvons pas voir si le pool de siRNA qui semble avoir un effet non spécifique au niveau messager pourrait avoir un effet positif au niveau protéique. De même, nous ne connaissons pas la fonction de l'isoforme de *VHL* contenant l'exon cryptique E1', nous ne pouvons donc pas voir si le pool de siRNA dirigé contre E1' a un effet sur la fonction de cette isoforme qui serait bénéfique dans le cadre du développement des pathologies.

En revanche, nous avons pu montrer dans le cadre des mutations synonymes de l'exon 2 de *VHL*, qui conduisent à un saut de cet exon, que les siRNA permettent une augmentation de l'expression de l'isoforme pVHL160 en réponse à une diminution de l'isoforme  $\Delta E2$  au niveau transcriptionnel. Cependant, nous ne pouvons pas utiliser ce modèle cellulaire pour étudier les effets sur la régulation de HIF. En effet, se trouvant dans des lignées LLB modifiées par le virus EBV, exprimant l'oncoprotéine virale LMP1 qui stabilise HIF, nous ne pouvons pas valider que le siRNA conduit à une réactivation de la protéine VHL dans les cellules. Pour cela, il faudrait introduire les mutations par CRISPR/CAS9 dans des cellules productrices d'EPO.

Nous avons pu constater qu'il est plus difficile de différencier en progéniteurs de la crête neurale des cellules porteuses de mutations de E1' à l'origine de la maladie de VHL. L'une des hypothèses envisagée est que l'exon cryptique E1' pourrait jouer un rôle dans le développement, et plus particulièrement dans celui de la crête neurale. En effet, lors de l'étude d'expression de E1', nous avons pu remarquer qu'il était plus exprimé dans le placenta et dans le cerveau où il pourrait avoir une fonction importante.

## 2. Séquençage de l'exon cryptique E1' dans un panel de patients avec au moins une tumeur du spectre VHL

### a. Introduction

Après la découverte de l'exon cryptique E1' et de son implication dans la survenue de la maladie de VHL et de polyglobulie, cet exon a été ajouté dans la liste des exons à séquencer dans le cadre du diagnostic moléculaire de ces deux pathologies. Des patients ont donc été séquencés rétrospectivement pour cet exon, patients pour lesquels les diagnostics moléculaires n'avaient pas encore été établis. Un panel de 1167 patients a été séquencé dont 221 avec au moins une tumeur du spectre VHL sans mutations *VHL* identifiée et 946 avec un phéochromocytome ou un paragangliome uniquement. Cette étude fait l'objet d'un article actuellement en cours de soumission. Six mutations différentes ont pu être trouvées dans E1' chez 12 familles. Cela représente 0,45% des patients avec 1 à 3 tumeurs du spectre VHL et 0,32% des patients avec un paragangliome ou un phéochromocytome uniquement.

### b. Article et données supplémentaires

#### **Germline mutations in the new E'1 cryptic exon of the *VHL* gene in patients with tumors of von Hippel-Lindau disease spectrum or with paraganglioma**

A Buffet<sup>1</sup> (MD, PhD), B Calsina<sup>2\*</sup> (M.Sc.), SK Flores<sup>3\*</sup> (M.Sc.), S Giraud<sup>4\*</sup>, M Lenglet<sup>5,6,7\*</sup> (PhD), P Romanet<sup>8</sup>, E Deflorenne<sup>1</sup>, J Aller<sup>9</sup> (MD), I Bourdeau<sup>10</sup> (MD), B Bressac de Paillerets<sup>11</sup>, M Calatayud<sup>12</sup> (MD), C Dehais<sup>13</sup> (MD), E De Mones Del Pujol<sup>14</sup> (MD), A Elenkova<sup>15</sup> (MD), P Herman<sup>16</sup> (MD), P Kamenicky<sup>17</sup> (MD, PhD), S Lejeune<sup>18</sup>, JL Sadoul<sup>19</sup>, A Barlier<sup>8</sup> (MD, PhD), S Richard<sup>6,20</sup> (MD), J Favier<sup>1</sup> (PhD), N Burnichon<sup>1</sup>, B Gardie<sup>5,6,7</sup>, PLM Dahia<sup>3</sup> (MD, PhD), M Robledo<sup>2</sup> (PhD), AP Gimenez-Roqueplo<sup>1,21,22</sup> (MD, PhD)

\* contribute equally to this work

<sup>1</sup> INSERM, UMR970, Paris-Centre de Recherche Cardiovasculaire, Paris ; équipe labellisée par la Ligue contre le Cancer

<sup>2</sup> Hereditary Endocrine Cancer Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain; and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Madrid, Spain.

<sup>3</sup> Division of Hematology and Medical Oncology, Department of Medicine, Mays Cancer Center, University of Texas Health Science Center at San Antonio, San Antonio, TX, 78229

<sup>4</sup> Genetics Department, Hospices Civils de LYON (HCL), University Hospital, East Pathology Center, B-A3, 59 Bld Pinel, 69677, Bron Cedex, France.

<sup>5</sup> Centre de Recherche en Cancérologie et Immunologie Nantes-Angers, INSERM, Université de Nantes and Université d'Angers, Nantes, France.

<sup>6</sup> École Pratique des Hautes Études, PSL Research University, Paris, France.

<sup>7</sup> L'Institut du Thorax, INSERM, Centre National de la Recherche Scientifique (CNRS), Université de Nantes, Nantes, France.

<sup>8</sup> Aix Marseille Univ, APHM, INSERM, MMG, Laboratory of Molecular Biology Hospital La Conception, Marseille, France.

<sup>9</sup> Endocrinology and Nutrition Service. Hospital Universitario Puerta de Hierro, 28222 Majadahonda, Spain

<sup>10</sup> Division of Endocrinology, Department of Medicine, Centre de recherche du Centre hospitalier de l'Université de Montréal (CRCHUM), Université de Montréal, Montréal, Canada.

<sup>11</sup> Département de Biologie et Pathologies Médicales, Université Paris-Saclay, Villejuif, France.

<sup>12</sup> Department of Endocrinology and Nutrition, Hospital Universitario 12 de Octubre, 28041 Madrid, Spain

<sup>13</sup> Assistance Publique-Hôpitaux de Paris, Hôpitaux Universitaires La Pitié Salpêtrière-Charles Foix, Service de Neurologie 2-Mazarin, Paris, France.

<sup>14</sup> Service d'ORL et de Chirurgie Cervico-faciale, Hôpital Pellegrin, CHU de Bordeaux, Bordeaux

<sup>15</sup> Clinical Center of Endocrinology and Gerontology, Medical Faculty, Sofia Medical University, Sofia, Bulgaria

<sup>16</sup> Assistance Publique, Hôpitaux de Paris, Service ORL-CCF, hôpital Lariboisière, université Paris VII, AP-HP, Paris

<sup>17</sup> Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital de Bicêtre, Service d'Endocrinologie et des Maladies de la Reproduction, Centre de Référence des Maladies Rares de l'Hypophyse, Le Kremlin-Bicêtre, France.

<sup>18</sup> Department of Clinical Genetics, Centre Hospitalier Régional et Universitaire de Lille, Lille, France

<sup>19</sup> Service d'Endocrinologie, Hôpital de L'Archet, CHU de Nice, Nice

<sup>20</sup> Department of Genetic Oncology, Ecole Pratique des Hautes Etudes, PSL Research University, INSERM U1186, Gustave Roussy Cancer Campus, Villejuif, France; National Expert Center for Rare Cancers (PREDIR), Bicêtre Hospital, Le Kremlin-Bicêtre, France.

<sup>21</sup> Université Paris Descartes, PRES Sorbonne Paris Cité, Faculté de Médecine, F-75006 Paris, France

<sup>22</sup> Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, Service de Génétique, F-75015 Paris, France

**Corresponding author**

Anne-Paule Gimenez-Roqueplo

Service de Génétique,

Hôpital Européen Georges Pompidou,

20 rue Leblanc, 75908 Paris cedex 15, France.

Tel: 33 1 56 09 38 81

Fax: 33 1 56 09 38 84

**Manuscript word count:** 1185 (1200 words)

## **KEYS POINTS**

### **Question**

Should germline mutations in the newly described cryptic exon (E'1) of the *VHL* gene be screened for in patients with clinical suspicion of von Hippel-Lindau disease (VHL)?

### **Finding**

In 1,167 patients with a former negative genetic test, a germline mutation of E1' was identified in 0.45% of patients with one to three tumors of the VHL spectrum and in 0.32 % of those with paraganglioma or pheochromocytoma only.

### **Meaning**

The E1' cryptic exon of the *VHL* gene should be screened in patients with VHL suspicion and added to panels for NGS diagnostic testing of hereditary paragangliomas and pheochromocytomas.

## **ABSTRACT**

### **Importance**

For improving the diagnosis of von Hippel-Lindau (VHL) disease and associated tumors and cancers, the mutation rate in the newly described cryptic exon (E1') of the *VHL* gene should be assessed.

### **Objective**

To determine the incidence of germline mutations in E1' in patients with tumors associated with VHL disease or with pheochromocytoma or paraganglioma (PPGL) only, but without an identified mutation by the former genetic tests.

### **Design, setting and patients**

This was a large international multicenter study based on a cohort of 1,167 patients with a negative genetic test. Germline DNA from 221 patients with one to three tumors of the VHL spectrum ("VHL-like" cohort) and 946 patients with a PPGL ("PPGL" cohort) were screened for E'1 genetic variations. Secondary genetic events and loss of heterozygosity (LOH) were further investigated at tumor level.

### **Results**

Twelve genetic variations in E'1 were detected. Two were classified as pathogenic, 2 as variant of unknown significance and 8 as benign or likely benign. The rs139622356 was found in 7 unrelated patients but described in only 16 patients out of the 31,390 of the gnomAD database ( $p < 0.0001$ ) suggesting that this variation could be a modifier mutation conferring a risk for the development of tumors and cancers of the VHL spectrum.

### **Conclusions and relevance**

A germline mutation of the E1' cryptic exon of *VHL* gene contributes to 0.45% of "VHL-like" cohort and to 0.32 % of "PPGL" cohort. Due to the significant impact of the diagnosis of von Hippel Lindau disease for affected patients and their relatives, the E1' cryptic exon of *VHL* gene should be screened in patients with clinical suspicion of VHL and added to panels for NGS diagnostic testing of hereditary paragangliomas and pheochromocytomas. Our study underlines the importance of studying variations identified in deep intronic sequences, which would have been missed by examining only coding sequences of genes/exomes, and that will be more frequently detected and studied with the upcoming arrival of whole genome sequencing into clinical practice.

## INTRODUCTION

Von Hippel-Lindau (VHL) disease is an autosomal dominant neoplastic syndrome<sup>1</sup> responsible for the development in affected patients of retinal or central nervous system hemangioblastomas, renal cysts or clear cell carcinomas, pancreatic cysts or neuroendocrine tumors, endolymphatic sac tumors and pheochromocytomas and/or paragangliomas (PPGLs). A germline mutation (including gross deletion) is identified in one of the 3 exons of *VHL* in almost all affected patients<sup>2, 3</sup>. Nevertheless, some patients with a VHL disease clinically diagnosed, but without identified *VHL* germline mutation, have been reported<sup>4</sup>. One of the tumor types of the VHL tumor spectrum, PPGL are rare, neuroendocrine tumors with a great genetic heterogeneity and the most impacted by the genetics with about 40% of genetically determined forms<sup>5, 6</sup>. Indeed, to date, approximately 17 susceptibility genes have been reported but two third of identified mutations are found in *SDHB*, *SDHD* and *VHL* genes<sup>7, 8</sup>.

Recently, a novel cryptic exon of *VHL* gene, named E1', has been discovered. A germline mutation in the first intronic region which results in creation of a cryptic exon designated E1', has been detected in one large family with a typical VHL disease and without any alteration in the other *VHL* exons<sup>9</sup>. *VHL* gene is one of the major PPGL susceptibility genes but, to date, E1' exon has not been included in PPGL target gene panels.

So, our objective was to assess the prevalence of E1' germline mutations in two international cohort of patients in, first 221 patients with a single or multiple tumors suggesting a von Hippel-Lindau disease and then in 946 with a PPGL but without an identified mutation in the three *VHL* exons or in other PPGL susceptibility genes, respectively.

## **METHODS**

### **Patients**

A total of 1,167 patients were analyzed, 221 patients with single or multiple tumors of the clinical spectrum of von Hippel-Lindau disease but without *VHL* mutation ("VHL-like" cohort) and 946 patients with PPGL but without germline mutation in one PPGL susceptibility genes ("PPGL" cohort) (**Table 1**). "VHL-like" cohort had *VHL* gene screening on germline DNA by Sanger sequencing or Next Generation Sequencing and large rearrangements were eliminated by MLPA or QMPSF. The procedures used for PPGL diagnosis were in accordance with international guidelines<sup>10, 11</sup>. Each patient signed a written informed consent for genetic analyses.

### **Direct sequencing of the E1' cryptic exon of *VHL* on germline DNA**

Sanger sequencing on germline DNA of E1' was performed as previously described<sup>9</sup>.

## RESULTS

In our cohort of 1,167 patients, we identified a rare germline genetic variation (minor allele frequency < 1%) in the E1' *VHL* cryptic exon in 12 patients (1%); 4 of these patients were from the “VHL-like” cohort (4 of 221 patients, 1.8%), and 8 from the “PPGL” cohort (8 of 946, 0.8%) (**Table 2**). Seven patients (patients #2 to #8) (0.6%) carried the same rare variation of uncertain significance (VUS), c.340+578C>T which is referenced in dbSNP as rs139622356 and has been previously reported in the Genome Aggregation Database (gnomAD), but not in ClinVar. The five remaining patients carried different E1' variations. One of them (patient #9) harbouring the c.340+617C>G mutation, was previously described in the original paper<sup>9</sup>. Patient #12 carried the c.340+866C>A VUS, which is referenced in dbSNP (rs536631685) and 1000 genomes, but not in gnomAD. Finally, three novel VUS of the E1' *VHL* cryptic exon were discovered in the three remaining patients. None of the four patients with a PPGL and an E1' VUS have developed VHL spectrum tumor(s) during their follow-up and none of them had family history of VHL disease; however, genetic familial screening was only performed in patient #9. The proband's mother was found to carry the variant, but no VHL spectrum manifestation.

Among the remaining 11 patients, three tumors were available, two as paraffin embedded tumors (patients #1, #10) and one as a frozen tumor (patient #3). None of them presented a loss of heterozygosity at *VHL* locus, and in tumor #1 the mutated allele was lost as determined by Sanger sequencing (**Supplemental methods**). We identified in tumor DNA of patient #10, which harbors the c.340+682T>C variation, a second variation in the exon 3 of *VHL* (c.482G>A; p.Arg161Gln) well known to be pathogenic (**Figure 1**). In the absence of LOH, this exon 3 variation may function as the second *VHL* hit in this tumor. Unfortunately RNA was not available from this tumor to determine whether these two variants were in the same or distinct alleles. No other mutation of the *VHL* gene was identified in tumor DNA of patients #1 and #10.

## DISCUSSION

E1' mutations were previously described in Lenglet et al. in 8 families, either with erythrocytosis or VHL disease. These mutations led to an abnormal *VHL* mRNA with the insertion of the E1' in the transcript and to a global defect in VHL protein expression<sup>9</sup>.

In our large international study, we identified 4 new germline variants in E1' *VHL* gene, representing 0.45% of "VHL-like" cohort and 0.32% of "PPGL" cohort. Our patients do not have all the manifestation of VHL disease. In the single patient in whom and familial genetic screening was performed (patient #9) the proband's mother is an asymptomatic carrier of the germline variant. These data suggest incomplete penetrance of these mutations, as it was previously described for *SDHA* gene mutations that show a relatively high allele frequency in gnomAD<sup>12</sup>. Moreover, we identified in 0.6 % of our cohort the variation rs139622356, which is 10 times more frequent in our cohort than in reference database. Indeed, this variation is described in 0.05 % of gnomAD patients (7/1,167 versus 16/31,390,  $p < 0.0001$ ). This result suggests that rs139622356 is a potential modifier variant leading to an 8.5 time (95% CI: 4.4-14.3,  $p < 0.0001$ , **Supplemental methods**) increasing risk of development of PPGL or VHL tumors.

## LIMITATIONS

Our study identifies rare germline variations in VHL E1' in an exceptional international population of patients with such rare diseases. The pathogenic significance of some of these variants remains to be further determined and functional assay need to be developed.

## CONCLUSIONS

Our study demonstrates that E1' *VHL* variants are rare events in "VHL-like" and "PPGL" patients, but nearly as frequent as the *VHL* mutation rate in other exons in patients with PPGL (in our molecular genetic laboratory *VHL* mutation rate in exon 1: 1.2% ( $p=0.022$ ), exon 2: 0.68% ( $p=0.22$ ), exon 3: 2.8 % ( $p < 0.001$ ) respectively, (personal communication) or as frequent as in exons of other PPGL susceptibility genes (for instance the mutation rate in exon 1 of *SDHD* is 0.43%). Furthermore, at least one of the variants is present at a higher frequency in the PPGL/VHL cohorts than in control populations, suggesting that it may play a modifier role in these diseases. As the identification of variants has important implications for management and follow-up of

patients and relatives, we suggest that E'1 cryptic exon should be added to NGS diagnostic panels. Considering the genetic heterogeneity of PPGLs and the high rate of detectable driver mutations in these tumors<sup>10</sup>, a low frequency of variants in any given new gene/exons is not unexpected. Finally, our study underlines the importance of studying variations identified in deep intronic sequences, which would have been missed by examining only coding sequences of genes/exomes.

## REFERENCES

1. Latif F, Tory K, Gnarr J, et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science*. May 28 1993;260(5112):1317-1320.
2. Lonsler RR, Glenn GM, Walther M, et al. von Hippel-Lindau disease. *Lancet*. Jun 14 2003;361(9374):2059-2067.
3. Maher ER, Neumann HP, Richard S. von Hippel-Lindau disease: a clinical and scientific review. *Eur J Hum Genet*. Jun;19(6):617-623.
4. Binderup ML, Galanakis M, Budtz-Jorgensen E, Kosteljanetz M, Luise Bisgaard M. Prevalence, birth incidence, and penetrance of von Hippel-Lindau disease (vHL) in Denmark. *Eur J Hum Genet*. Feb;25(3):301-307.
5. Favier J, Amar L, Gimenez-Roqueplo AP. Paraganglioma and pheochromocytoma: from genetics to personalized medicine. *Nat Rev Endocrinol*. Feb;11(2):101-111.
6. Gimenez-Roqueplo AP, Dahia PL, Robledo M. An update on the genetics of paraganglioma, pheochromocytoma, and associated hereditary syndromes. *Horm Metab Res*. May;44(5):328-333.
7. Amar L, Bertherat J, Baudin E, et al. Genetic testing in pheochromocytoma or functional paraganglioma. *J Clin Oncol*. Dec 1 2005;23(34):8812-8818.
8. Curras-Freixes M, Pineiro-Yanez E, Montero-Conde C, et al. PheoSeq: A Targeted Next-Generation Sequencing Assay for Pheochromocytoma and Paraganglioma Diagnostics. *J Mol Diagn*. Jul;19(4):575-588.
9. Lenglet M, Robriquet F, Schwarz K, et al. Identification of a new VHL exon and complex splicing alterations in familial erythrocytosis or von Hippel-Lindau disease. *Blood*. Aug 2;132(5):469-483.
10. Toledo RA, Burnichon N, Cascon A, et al. Consensus Statement on next-generation-sequencing-based diagnostic testing of hereditary pheochromocytomas and paragangliomas. *Nat Rev Endocrinol*. Apr;13(4):233-247.
11. Lenders JW, Duh QY, Eisenhofer G, et al. Pheochromocytoma and paraganglioma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. Jun;99(6):1915-1942.
12. van der Tuin K, Mensenkamp AR, Tops CMJ, et al. Clinical Aspects of SDHA-Related Pheochromocytoma and Paraganglioma: A Nationwide Study. *J Clin Endocrinol Metab*. Feb 1;103(2):438-445.

## **ACKNOWLEDGEMENTS**

We thank Catherine Tritscher for administrative assistance, and Caroline Travers, Valentin Adamus and Astrid Ramahefasolo for technical assistance. We also thank the PREDIR and TENGEN networks for their help in patient inclusion.

A. Buffet received a financial support from ITMO Cancer AVIESAN (Alliance Nationale pour les Sciences de la Vie et de la Santé, National Alliance for Life Sciences & Health) within the framework of the Cancer Plan and from la Fondation pour la Recherche Médicale (FDT20170436955). S.K.F. is currently supported by an NIH-NIGMS individual predoctoral fellowship grant (F31-GM131634-01) and, previously, by an NIH NRSA Institutional Predoctoral Training Grant T32CA148724 and NRSA F31-GM131634-01. P.L.M.D. receives funding support from NIH-GM114102, Alex's Lemonade Stand Cancer Foundation (Innovation Award) and the NCATS- UL1 TR002645. The Genomic Sequencing Facility at the GCCRI is supported by the P30-CA54174 (CTRC at UTHSCSA) and NIH Shared Instrument grant 1S10OD021805-01 (S10 grant). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. B.C. is supported by Rafael del Pino Foundation. M.R. receives funding support from Instituto de Salud Carlos III (ISCIII), through the “Acción Estratégica en Salud” (AES) (projects PI17/01796), cofounded by the European Regional Development Fund (ERDF)), and the Paradifference Foundation.

## TABLES AND FIGURES

**Table 1:** Main clinical and tumor characteristics of “VHL-like” and “PPGL” cohorts.

**Table 2:** Patients with a genetic variant in the cryptic E1' exon of *VHL* gene.

**Figure 1:** E'1 variations identified in our cohort: mutations identified in patient #9 on the germline DNA and somatic DNA (A); Location in E1' of the germline mutations of E1' identified (B).

**Table 1:** Main clinical and tumor characteristics of “VHL-like” and “PPGL” cohorts.

<i><b>Total patients n= 1167</b></i>	
<i><b>VHL-like</b></i> n= 221 patients	
Age at first diagnosis mean, (min-max)	43 (11-81)
-----	
Single VHL spectrum tumor	105 (47.5%)
Two VHL spectrum tumors	73 (33%)
Three or more VHL spectrum tumors	43 (19.5%)
<i><b>PPGL</b></i> n= 946 patients	
Age at first diagnosis mean, (min-max)	43 (8-94)
-----	
Benign PPGL	869 (92%)
Single benign PPGL	771 (82%)
Multiple benign PPGL	98 (10%)
-----	
Metastatic PPGL	77 (8%)
Single metastatic PPGL	67 (7%)
Multiple metastatic PPGL	10 (1%)
-----	
Familial PPGL	17 (2%)

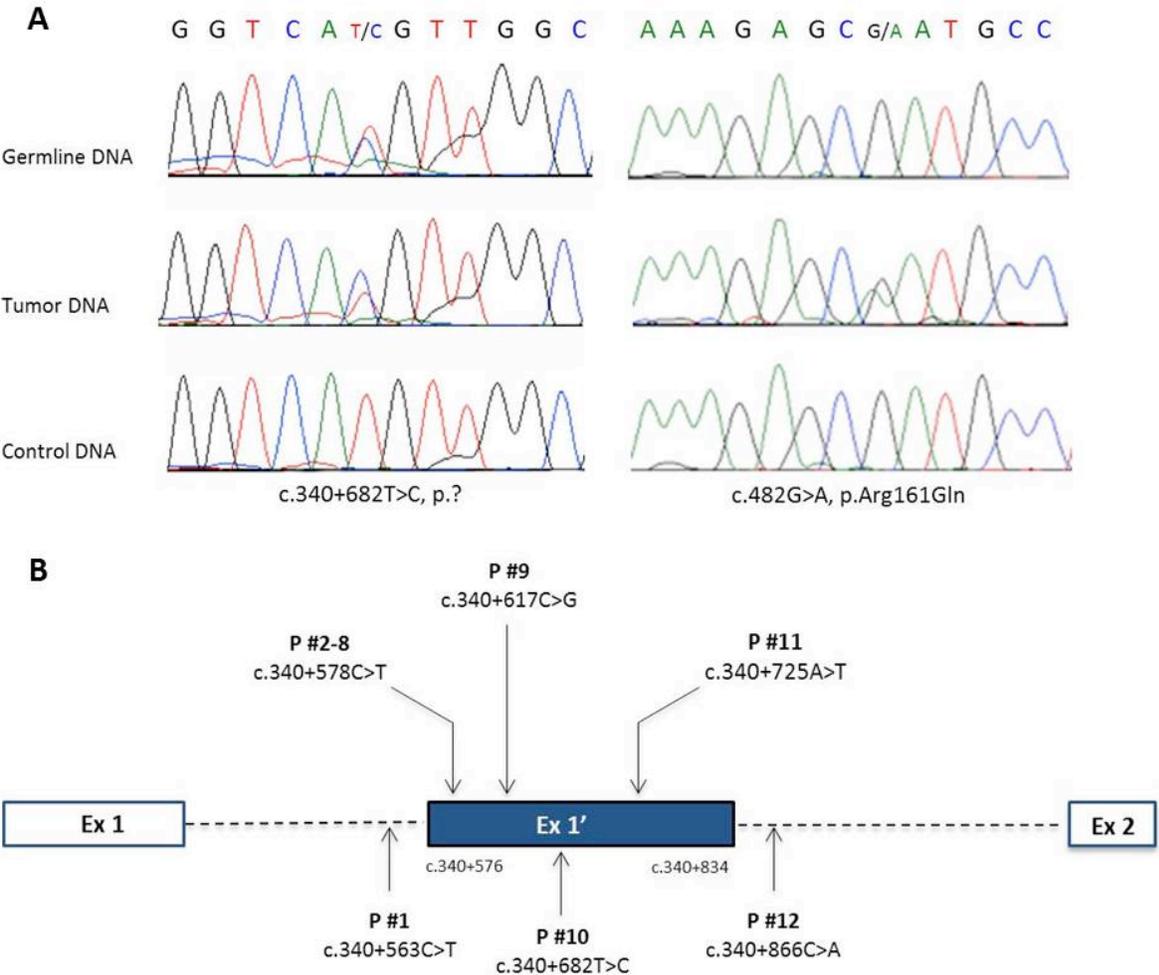
**Table 2:** Patients with a genetic variant in the cryptic E1' exon of *VHL* gene

Patient number	Phenotype (age at first diagnosis)	Germline genetic variant	SNP ID in ExAC or 1000 Genomes	Allele freq. <sup>b</sup> (%)	LOH	Somatic <i>VHL</i> mutation	Classification of VUS
#1	Carotid body PGL (47)	c.340+563C>T		NA	No <sup>c</sup>	ND	Benign
#2	Carotid body PGL (34)	c.340+578C>T	rs139622356	0.05	NA		Probably Benign
#3	PCC (38)	c.340+578C>T	rs139622356	0.05	No	No	Probably Benign
#4	Carotid body PGL (74)	c.340+578C>T	rs139622356	0.05	NA		Probably Benign
#5	Carotid body PGL (56)	c.340+578C>T	rs139622356	0.05	NA		Probably Benign
#6	PGL + ccRCC (81, 82)	c.340+578C>T	rs139622356	0.05	NA		Probably Benign
#7	bilccRCC (49, 51)	c.340+578C>T	rs139622356	0.05	NA		Probably Benign
#8	ccRCC (39)	c.340+578C>T	rs139622356	0.05	NA		Probably Benign
#9	Multiple retinal HMB (36)	c.340+617C>G <sup>a</sup>		NA	NA		Pathogenic
#10	PCC (11)	c.340+682T>C		NA	No	c.482G>A p.R161Q	Pathogenic
#11	Multiple HN PGL (12)	340+725A>T		NA	NA		VUS
#12	PCC (34)	c.340+866C>A	rs536631685	0.02	NA		VUS

<sup>a</sup> Mutation described in Lenglet et al., <sup>b</sup> Frequency in gnomAD or 1000 Genomes, <sup>c</sup> loss of the mutated allele.

PCC: pheochromocytoma, PGL: paraganglioma, HN: head and neck, RCC: renal cells carcinoma, ccRCC: clear cell renal cell carcinoma, bilccRCC: bilateral clear cell renal cell carcinoma, HMB: hemangioblastoma, NA: none available, ND: not done, VUS: variant of uncertain significance.

**Figure 1:** E'1 variations identified in our cohort: mutations identified in patient #9 on the germline DNA and somatic DNA (A); Location in E1' of the germline mutations of E1' identified (B).



## SUPPLEMENTAL METHODS

### ***VHL* gene analysis in tumor**

Tumor DNA was extracted from frozen or paraffin embedded tumor by the QIAamp DNA minikit (Qiagen). Loss of heterozygosity was evaluated by 1) sanger sequencing of the E1' cryptic exon of *VHL* by specific primers of the mutation 2) microsatellite analysis on D3S1537, D3S1038, D3S1317 D3S3547, D3S3727 as previously described<sup>1, 2</sup>. *VHL* deletion on tumor DNA was assessed with the *SALSA MLPA* P016 *VHL* probemix (MRC-Holland).

### **Statistical analysis**

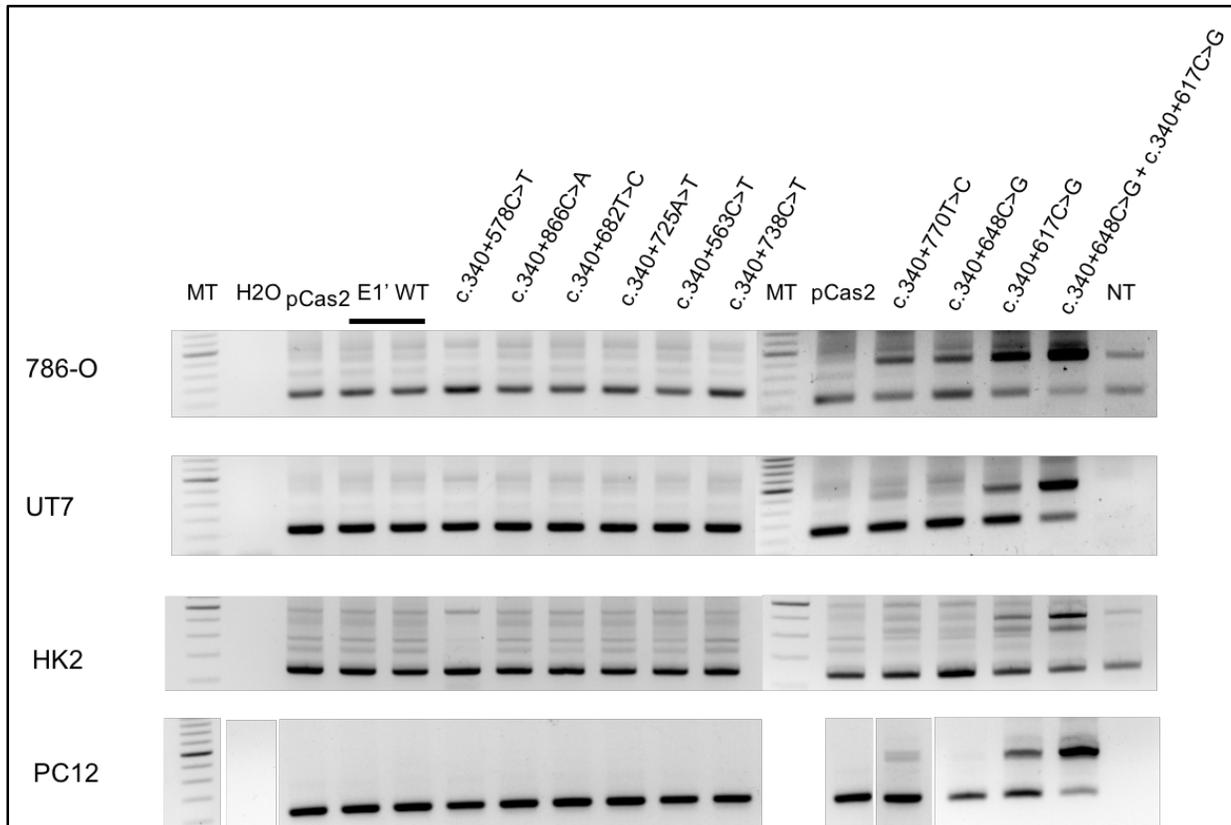
Statistical analysis was carried out with GraphPad software. Differences between allele's frequency in gnomAD and our cohort of patients and relative risk were assessed by Chi-squared tests. A *P*-value <0.05 was considered significant.

## REFERENCES

1. Hoefling C, Schmidt H, Meinhardt M, et al. Comparative evaluation of microsatellite marker, AP-PCR and CGH studies in primary renal cell carcinoma. *Int J Mol Med*. Jun 2004;13(6):835-842.
2. Girolami F, Passerini I, Gargano D, et al. Microsatellite analysis of chromosome 3p region in sporadic renal cell carcinomas. *Pathol Oncol Res*. 2002;8(4):241-244.

### c. Résultats complémentaires

Cette étude a permis de trouver 6 variants génétiques différents chez 12 patients ayant développé des tumeurs du spectre VHL. Nous avons donc réalisé une étude de minigènes rapporteurs d'épissage sur ces mutants afin de voir leur implication sur la rétention de l'exon cryptique E1' de *VHL* dans les transcrits (Figure 26).



**Figure 26 : Caractérisation de la rétention de *VHL*-E1' par expérience de minigènes rapporteurs.** Des RT-PCR ont été réalisées sur de l'ARNm extrait de lignées cellulaires transfectées avec des rapporteurs minigènes de E1' sauvages ou mutés. L'exon E1' est inséré dans un vecteur avec une large partie de ses introns (environ 300pb de chaque côté de l'exon cryptique) entre deux exons A et B du gène *SERPING1* qui sont la cible des amorces utilisées pour la RT-PCR.

Une liste des variants retrouvés dans l'exon cryptique E1' est présentée ci-après (Tableau 1).

Patient	Phenotype	Mutation in VHL	Mutation in E1'	Consequence on X1 protein	SNP in E1'	MAF
80 II.1 (Famille 1 article Blood)	Erythrocytosis	c.429C>T, p.Asp143Asp	c.340+770T>C	c.535T>C, p.Ser179Pro		
80 II.3 (Famille 1 article Blood)	healthy	WT	c.340+770T>C	c.535T>C, p.Ser179Pro		
20 (Famille 2 article Blood)	Erythrocytosis	c.598C>T, p.Arg200Trp	c.340+770T>C	c.535T>C, p.Ser179Pro		
53 + 92 (Famille 3 article Blood)	Erythrocytosis	c.598C>T, p.Arg200Trp	c.340+770T>C	c.535T>C, p.Ser179Pro		
E274 (Famille 5 article Blood)	Erythrocytosis	c.430G>A, p.Gly144Arg	c.340+694_711dup	c.458_476dup, p.Trp159X		
WC90 (famille 4 article Blood)	Erythrocytosis	c.598C>T, p.Arg200Trp	c.340+694_711dup	c.458_476dup, p.Trp159X		
TSEboy (Famille 6 article Blood)	Erythrocytosis	c.492G>T, p.Gln164His	c.340+574A>T	splicing (positon -2 on E1')	rs982745672	T=0.00007/2 (TOPMED)
816 (Famille 7 article Blood)	Erythrocytosis	WT	c.340+816+A>C, Homozygous	c.581A>C, dans le codon stop de E1' si X1	rs1031288121	no frequency submission
922 (Famille 8 article Blood)	VHL disease	WT	c.340+617C>G/c.340+648T>C*	c.382C>G, p.Leu128Val/c.413T>C, p.Leu138Pro		
1552 (Patient #9 article 2)	Hemangioblastoma	WT	c.340+617C>G	c.382C>G, p.Leu128Val		
FUNchr	Hemangioblastoma CNS	WT	c.340+648T>C*	c.413T>C, p.Leu138Pro	rs73024533	C=0.0026/13 (1000 Genomes) C=0.0050/147 (TOPMED)
DERmar		WT	c.340+648T>C*	c.413T>C, p.Leu138Pro	rs73024533	C=0.0026/13 (1000 Genomes) C=0.0050/147 (TOPMED)
683	RCC at 40 years	WT	c.340+738C>T	c.503C>T, p.Ala168Val		
162002616			c.340+578C>T	c.342C>T, p.Val114Val		
1078 (Patient #10 article 2)	PHEO at 11 years	WT	c.340+682T>C	c.447T>C, p.His149His		
923 (Patient #11 article 2)	multiple cervical PGL at 12 years	WT	c.340+725A>T	c.490A>T, p.Thr164Ser		
707 (Patient #2 article 2)	cervical PGL at 34 years	WT	c.340+578C>T	c.343C>T, p.Leu115Phe	rs139622356	T=0.0006/3 (1000 Genomes) T=0.0005/15 (TOPMED)
1191 (patient #3 article 3)	PHEO at 38 years, familial ATCD (sudden death on her 26 years old sister)	WT	c.340+578C>T	c.343C>T, p.Leu115Phe	rs139622356	T=0.0006/3 (1000 Genomes) T=0.0005/15 (TOPMED)

**Tableau 1 : Liste des variants de l'exon cryptique E1'.** Positions des variants identifiés dans E1' au niveau nucléotidique et protéique avec les phénotypes associés, les prédictions des conséquences sur l'épissage, la fréquence observée et la référence éventuelle du SNP. PHEO : phéochromocytome, PGL : paragangliome, SNP : polymorphisme nucléotidique, MAF : fréquence allélique.

#### d. Discussion

A l'issue des tests de minigènes rapporteurs, nous pouvons voir que dans les 4 lignées cellulaires testées, seule la mutation c.340+617C>G à l'origine d'hémangioblastomes multiples de la rétine (retrouvée chez le patient #9 de l'article 2) conduit à la rétention de l'exon cryptique E1' de *VHL*. Il faudrait vérifier cette observation dans les ARNm des tumeurs de patients afin de valider que les mutations conduisent à la rétention de E1' à l'état physiologique. Il est possible que les variants analysés n'aient pas d'effet sur la rétention de E1' dans les lignées cellulaires utilisées. Il est également possible que la méthode de minigène rapporteur utilisée ne soit pas assez sensible pour détecter de faibles rétentions de l'exon E1'. En effet, nous pouvons voir que pour E1', il y a une rétention de bandes non spécifiques qui peuvent diminuer la sensibilité de la technique pour cet exon.

Nous pouvons aussi faire l'hypothèse que l'isoforme E1E1'E2E3 de *VHL* a une fonction physiologique à son état basal d'expression et que les mutations conduisent à une perte de fonction de cette isoforme qui est à l'origine du développement des tumeurs chez les patients. Ceci pourrait expliquer que nous n'observons pas pour le moment de rétention de E1' pour ces mutations. Les éventuels effets de cette perte de fonction seront discutés en conclusion et discussion générale de ma thèse.

Pour les mutations hétérozygotes c.340+648C>G, c.340+617C>G et c.340+648C>G associée à c.340+617C>G, nous pouvons observer un effet graduel de rétention de E1' dans les quatre lignées cellulaires utilisées pour les tests rapporteurs minigène. En effet, il y a une faible rétention pour la mutation c.340+648C>G, une rétention plus importante pour c.340+617C>G et une forte rétention pour l'association des deux mutations. Dans ce modèle, nous observons donc une addition des deux mutations que nous avons observée de façon moins importante lors de la quantification des transcrits et des protéines dans les lignées LLB des patients. Ce modèle nous permet d'établir un gradient entre la rétention de E1' observée et la sévérité du phénotype développé à savoir des hémangioblastomes du système nerveux central pour le patient porteur de la mutation c.340+648C>G, des hémangioblastomes de la rétine pour le patient porteur de la mutation c.340+617C>G et la maladie de VHL pour les patients porteurs des deux mutations pour

lesquels la rétention de E1' est plus importante en test rapporteur d'épissage minigène. Ce gradient de rétention est à confirmer dans les tumeurs des patients.

### 3. Etudes fonctionnelles de variants identifiés dans l'exon E2'

#### a. Introduction

Nous avons pu mettre en évidence que dans certains cas, il y a rétention d'une partie de l'intron 2 du gène *VHL* entre les exons 2 et 3 du messager que nous avons nommé E2'. Cet exon de 111 paires de bases est situé entre les positions c.464-1083 et c.464-973 de l'intron 2. Ce transcrit E1E2'E3 semble être encore plus faiblement exprimé que le transcrit dans lequel l'exon cryptique E1' est retenu, c'est pourquoi nous avons dans un premier temps axé nos recherches sur l'exon E1'. Nous avons ajouté l'exon cryptique E2' dans les exons à séquencer chez les patients atteints d'érythrocytoses, de la maladie de VHL ou de phéochromocytomes, sans altération génétique décrite. Cela nous a permis d'identifier des variants génétiques ainsi qu'une délétion de cet exon, ce qui nous a poussé à étudier l'exon E2' et les éventuelles fonctions des transcrits et de la protéine qui y serait associée.

#### b. Résultats

L'exon cryptique E2' est exprimé, au niveau transcriptionnel, dans tous les tissus en faible quantité (Figure 1D et supp 1 de l'article 3 (page 133)). Il est épissé en phase avec l'exon 2 puis l'exon 3 du gène *VHL* (Figure 1C). L'analyse *in silico* de l'exon cryptique E2' de *VHL* nous a permis de voir que celui-ci a une homologie importante avec des domaines transmembranaires de l'hème oxygénase 2 et de la protéine TMEM116 (Figure 1A). Cependant, nous avons pu montrer par immunohistochimie sur des cellules HEK293T transfectées avec des plasmides codant VHL ou X2 que l'exon cryptique E2' ne confère pas à X2 une localisation transmembranaire (Figure 1B). En effet, cette dernière est retrouvée dans le cytoplasme, comme la protéine VHL.

La fonction de X2 a été analysée grâce à un test fonctionnel cellulaire (test rapporteur) luciférase et comparé avec celles de VHL WT et de VHL $\Delta$ E2. Nous avons utilisé une construction d'un gène luciférase sous le contrôle d'un promoteur contenant une séquence HRE. Cette construction est co-transfectée dans des cellules avec un vecteur codant X2 ou VHL. En absence de VHL, HIF va pouvoir se fixer sur la séquence HRE du promoteur luciférase qui sera alors exprimée dans les cellules et dont nous pourrions mesurer l'expression par luminescence. Quand VHL est présent et

fonctionnel, il va entraîner la dégradation de HIF $\alpha$  qui ne sera alors plus disponible pour se fixer sur la séquence HRE du promoteur de la luciférase. Cela se traduit par une diminution de l'expression de la luciférase que nous mesurons par une diminution de la luminescence émise par les lysats cellulaires. Au niveau fonctionnel, la protéine X2 possède une faible activité de dégradation de HIF, mais celle-ci est moindre par rapport à VHL sauvage (Figure 2). Les tests de biotinylation des partenaires de X2 comparés à ceux de pVHL213 et pVHL172 nous permettent de voir que X2 n'est capable de fixer la culine 2 que à 40% par rapport à pVHL213 et l'élongine C à 20% (Figure 1E) (Collaboration avec Y. Arlort, F. Chesnel, Rennes). Cela pourrait expliquer la perte de fonction de X2 par rapport à pVHL213 observée pour la dégradation de HIF en test fonctionnel luciférase.

Nous avons pu montrer sur des LLB de patients, que la délétion d'une copie de E2' ne conduit à une diminution de l'expression d'aucune des isoformes de VHL, que ce soit au niveau transcriptionnel ou protéique (Figure 3A et B).

Nous n'avons pas encore caractérisé le rôle cette isoforme ni l'impact des variants identifiés (Tableau complet dans l'article). Cette étude est en cours (voir bilan ci-dessous).

### c. Article en préparation et données supplémentaires

## **Caractérisation of a new cryptic exon E2' identified in the VHL gene.**

### **Introduction**

Congenital erythrocytosis is due to alterations of the proteins involved in the regulatory pathway of erythropoiesis. Erythrocytosis is associated with rubor, vertebral hemangiomas, varicose veins, low blood pressure, thrombosis, pulmonary hypertension and hemorrhagic events (1). Detection of the genetic alteration is important for adequate clinical management. The first genetic

alteration described in congenital erythrocytosis is the c.598C>T, p.Arg200Trp homozygous mutation in the *VHL* (von Hippel-Lindau) gene. It has been discovered in the autonomous Republic of Chuvashia, a landlocked region with a high level of consanguinity, where the pathology is an endemic disorder (2). Since, this mutation has been found in combination with other *VHL* mutations (compound heterozygosity) and other missenses *VHL* mutations in both alleles (at homozygous or compound heterozygous states) have been described (3), (4). Even with the use of next-generation of sequencing panel diagnostics, the underlying cause of erythrocytosis is found in only one third of patients. Therefore, there is still cases of unexplained erythrocytosis with no mutation found or with mutations found at heterozygous states. Recently, a novel cryptic exon named E1' was described in *VHL* and has been found mutated in some cases of erythrocytosis (compound heterozygosity or homozygosity) (4).

The von Hippel-Lindau disease, described in 1936 after the description of the different tumors by the doctors von Hippel and Lindau, is an autosomal dominant disorder caused by heterozygous mutations in the *VHL* gene identified in 1993 (5). Patients develop hypervascularised tumors like retinal or central nervous hemangioblastomas, pheochromocytomas, clear cell renal carcinomas or neuroendocrine tumors. In some cases, patients develop tumor of the spectrum of tumors of VHL disease without any *VHL* alteration found. Recently, few cases have been resolved with the discovery of the new E1' cryptic exon of *VHL* found mutated as a part of VHL disease (4).

The *VHL* gene is located on 3p25-26 (6). It has been reported firstly to contain 3 exons and recently, a new cryptic exon named E1' has been reported (4). The commonly described transcript contain the 3 spliced exons (E1E2E3) which encode a protein of 213 amino acid (pVHL213). An alternative in-frame translation start site can be used and result a 160 amino acid protein (pVHL160) (7). The functions of this protein is very similar to the pVHL213 protein. A naturally spliced variant is also described; it is transcribed at low level in some tissues. It contains only the E1 and E3 exons which code a protein of 172 amino acid (pVHL172 or  $\Delta$ E2). This protein has been described to be non-functional and this function is still under investigation (8), (9). The cryptic exon E1' is spliced after the first exon of *VHL*. Consequent isoform contains a premature stop

codon at the end of the E1' exon. The transcript is recognized by the non mediated decay (NMD) mRNA mechanism which conduce to an early degradation of this messenger. It has been found expressed in all tissues at low level but its function is still under investigation.

VHL proteins are involved in many cellular functions but the most studied is the regulation of the oxygen cellular pathway. The main player of this pathway is the transcriptional factor HIF (hypoxia inducible factor) which govern the transcription of more than 200 genes involved in many cellular processes (erythropoiesis, angiogenesis, metabolism and cell survival). Under normal oxygen supply, the  $\alpha$  subunits of HIF (HIF1, 2 or 3) are hydroxylated by the prolyl-4-hydrolase domain enzymes (PHD1, 2 or 3). HIF $\alpha$  is also recognized by pVHL, a subunit of an E3 ubiquitin ligase complex that promotes its degradation by the proteasome (10), (11). Under hypoxic conditions, or when *VHL* is mutated, HIF $\alpha$  remain stable and can translocate to the nucleus to form an heterodimer with the  $\beta$  subunit of HIF. This heterodimer forms an active transcription factor. Dysregulation of the hypoxia pathway is central to the development of tumors as a part of the VHL disease (12) and to the development of erythrocytosis *via* up regulation of the erythropoietin (EPO), a target of HIF, which led to hyper-proliferation of the erythroid progenitors and differentiation into red blood cells (2), (13). It as been demonstrated that the complex pattern of disease manifestations observed in VHL syndrome is correlated with a gradient of pVHL dysfunction in hypoxia signaling pathway (13), (4).

It has been found that splicing mutation can be at the origin of the two pathologies associated with *VHL* mutations. Two processes have been described: the exon 2 skipping or the inclusion of the cryptic exon E1' in consequence of mutations (14), (4). This study pursues the study of splicing mutations and led to the discovery of a second cryptic exon in *VHL* gene that we termed E2'. We report an investigation of functional study of E2'.

## Methods

### Study approval

Informed consent for medical diagnosis and research was obtained from the patients and their relatives. This study was agreed by the Comité Consultatif de Protection des Personnes dans la Recherche Biomédical (French ethical committee) Paris-Sud at Bicêtre Hospital.

### **Sanger sequencing**

Exons and exon-intron junctions of the VHL gene were sequenced from DNA extracted from whole blood, as previously described (13).

### **Transcript detection and quantification**

After reverse-transcription reactions (Life Technologies), TaqMan real time exon-specific polymerase chain reaction (PCR) was performed using 20ng of complementary DNA (cDNA) for the VHL transcripts and 5ng of cDNA for the others transcripts with the quantitative PCR MasterMix (Eurogentec). Quantification of *RPLPO* was used as internal control. For VHL transcripts, the thresholds were determined using dilutions of plasmids containing coding sequences of each isoform.

### **Reporter assays**

The plasmid encoding the WT hemagglutinin-tagged VHL was kindly provided by Professor William G. Kaelin Jr. The plasmid encoding the  $\Delta E2$  hemagglutinin-tagged VHL was kindly provided by Professor Arlot-Bonnemains Y. An expression plasmid for the hypothetical X2 protein was constructed by sub-cloning the coding sequence after its synthesis (Life Technologies). Mutations of interest were introduced by site-directed mutagenesis (New England Biolabs). 786-O cells were transfected (Ozyme) with constructs encoding WT or mutant proteins, firefly luciferase under the control of hypoxia response element, and *Renilla* luciferase for normalization, as previously described (15).

### **Western blotting**

Cell lysates from the luciferase reporter assay were loaded into a Bis-Tris Mini Gel (4% to 12%; Invitrogen). After transfer, the membrane (Biorad) was subsequently incubated with a mouse

antihemagglutinin antibody (BioLegend) and then, a goat anti-mouse horseradish peroxidase-conjugated antibody (Jackson Immuno Research). Western blot using the mouse monoclonal antibody JD-1956 raised against the exon 1 of VHL (patent #14305925.1-1402-2014; Centre National de la Recherche Scientifique, Etablissement Français du Sang) was performed as described (8).

### **Minigene experiments**

Minigene constructs were prepared in pCas2 plasmid, containing 2 artificial exons A et B (as described (16)), between which VHL exon 2' with intronic flanking sequences were cloned. Cells were transfected (HEK293T, Ozyme) or nucleofected (UT7, Lonza). RNA was extracted 24 hours after transfection and reverse transcribed. PCR amplification was performed using the PCR Go TaqQ2 kit (Promega) with primers against artificial exons A and B. PCR products were resolved in a 2% agarose gel and compared to a 100pb size marker (Invitrogen).

### **Immunostaining**

Immunostaining was performed in HEK293 transfected with VHL-HA or X2-HA plasmids. After 24 hours of transfection, cells were fixed with formaldehyde at 4% and then permeabilized in 0,5% Triton X-100. Cells were blocked with PBS/BSA at 3% solution and then incubated overnight at 4°C with the primary antihemagglutinine antibody (BioLegend). After an incubation with the donkey anti mouse 568 secondary antibody (red labelling of HA-protein) and the phalloidin 488 (green labelling of cytoskeleton of actin), the images are acquired by a confocal microscopy system.

### **Pulldown**

pcDNA3-BirA(HA) (a biotin protein ligase) and VHL(HA) isoforms were transfected in cells (Ozyme). 24h after, cells were lysate and amount of total protein were measured with Lowry method (Biorad Protein Assay). Same amount of denatured protein (=input) for each assay were bound in a resin with high affinity for the biotin. Biotinylated proteins were eluted with Laemmli buffer (=bound). Input and bound were then analyzed by western blotting with specific antibodies.

### **Generation of iPS CRISPR cells**

iPS CRISPR cells were obtained from iPS control transfected by nucleofection (Lonza) with a specific guide for E1' and the CRISPR Cas9 endonuclease. Cells are sorted 24 or 48 hours after transfection by FACS and scattered in single cell. Clones of interest were selected by Sanger sequencing before their amplification.

### **Differentiation of iPS in EPO producing cells**

Differentiation has been done as described (paper Robriquet et al, in publication).

## **Results**

### **- Identification of new VHL spliced isoform containing a cryptic exon**

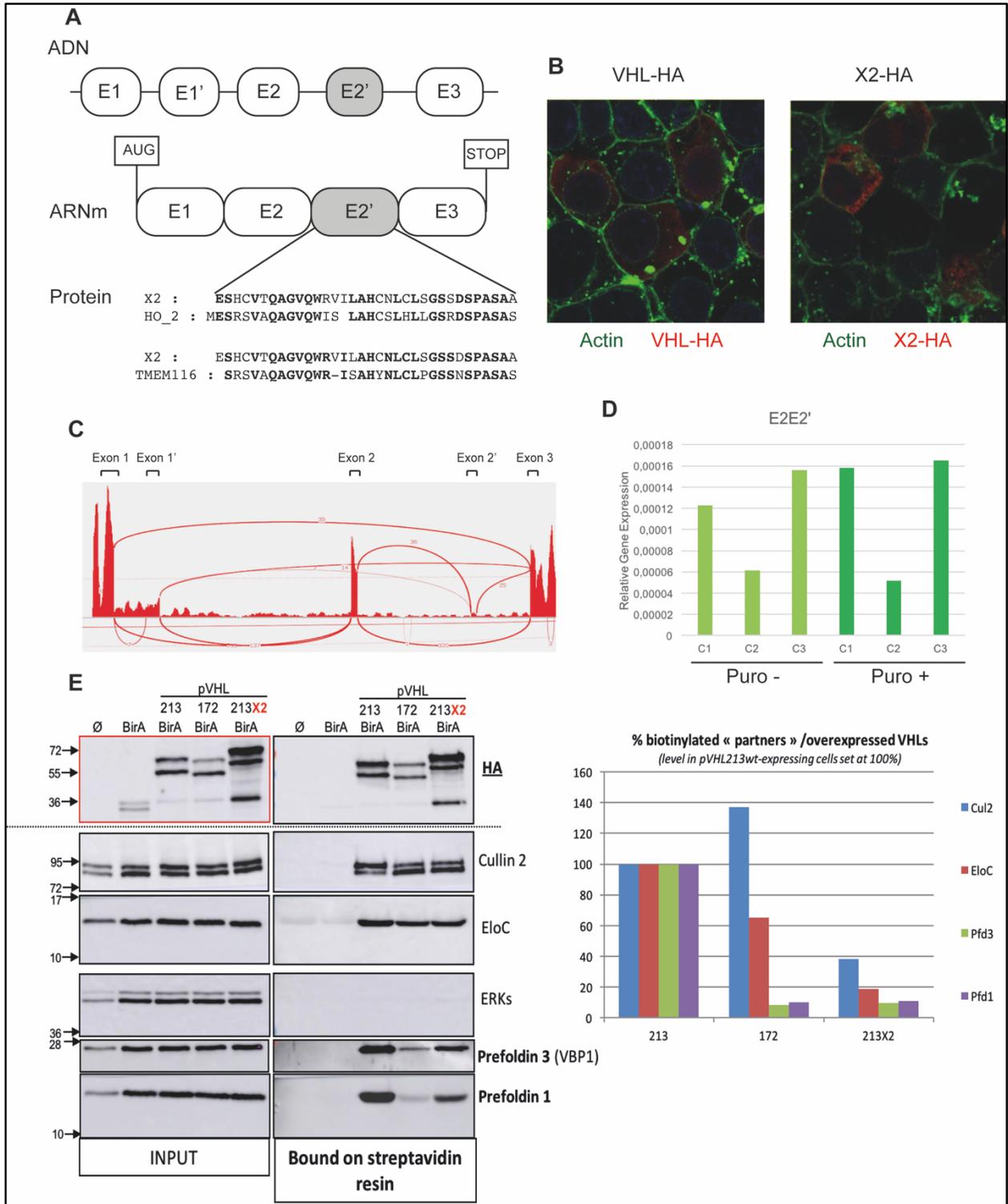
Our study began with the cloning of all the bands obtained after a RT-PCR with primers located in exon 1 and 3 of *VHL*. After the sequencing of all the clones obtained, we identified in addition to the E1' cryptic exon, another cryptic exon, that we named E2' because of its localization in the second intron of *VHL* (Figure 1 C and supplemental figure 1A). This exon contained 111 base-paired (bp) insert in phase between the exons 2 and 3 of the gene and led to an insertion of 37 amino acids in the protein pVHL213 (E1E2E3) which led to a hypothetical protein named X2 (Figure 1 A and C).

As hypothesized, the expression of the isoform that contains E2' is not regulated by the NMD mechanism because of the absence of a premature stop codon (contrary to the transcripts containing E1'). We have also investigated the expression of this transcript in a bank of cDNA of different human tissues (supplemental figure 1B). We showed a weak expression of this isoform (more expressed in lung, kidney and brain).

We investigated the expression of transcripts that contain E2 spliced with E2' in the presence or absence of puromycin in 3 LCLs controls (Figure 1D and supplemental figure 1C). We showed that this isoform is weakly expressed in the LCLs and that this expression is not induced by the puromycin.

The E2' domain of the protein X2, described in figure 1A, as an important homology to a transmembrane domain of 2 proteins: the heme-oxygenase 2 and the TMEM116 (transmembrane protein 116) (Figure 1A, below). We transfected VHL WT or X2 WT in HEK293T cells to show the localization of the two proteins. The E2' domain does not induce a relocation at the membrane of pVHL (Figure 1B).

To identify new partners of X2 and to study the interaction of E2' with known partners, pcDNA3-BirA(HA), a biotin protein ligase and VHL(HA) isoforms were cotransfected in different cells (Figure 1E). The biotin protein ligase conducted to the biotinylation of VHL and of each protein which are in contact with VHL during the 24h of transfection. After denaturation of the cells, biotinylated proteins were detected with a resin with a high affinity for the biotin. Each of these proteins were in contact with the isoform of transfected VHL. Some of the partners of pVHL213 were analyzed for its interaction with X2 and pVHL 172 in HEK293T cells. We showed that pVHLX2 have a profile close to pVHL172 for the interaction with the pfoldin 1 and 3. The interaction with these 2 proteins is very weak in comparison to the interaction between pVHL213 and these two pfoldins. We observed that pVHLX2 can only bind at 40% with the cullin 2 and at 20% with the ElonginC in comparison to pVHL213. It would therefore appear that pVHLX2 may failed to form an active E3 ubiquitin ligase complex in comparison to pVHL213. (Identification of new partners of X2 in process)



**Figure 1: Expression of the new VHL transcript isoform and functional study of the associated protein X2.** (A) Schematic representation of the VHL gene with the E2' cryptic exon and its associated protein sequence compared to two homologous sequences of Heme Oxygenase 2 and TMEM116. (B) Immunostaining of HEK293 transfected with VHL-HA or X2-HA. The images are acquired by a confocal microscopy system and reveal in green the cytoskeleton of actin and in red the HA tagged protein. (C) Sashimi plot from RNA-seq data obtained in the ENCODE database originated from Homo Sapiens neural progenitor cell differentiated from H9 (embryonic stem cells). The positions of the different VHL exons are indicated. Spliced junctions are denoted by the horizontal link, with details provided in supplemental figure 1A. (D) Quantitative RT-PCR was carried out with a Taqman probe specific to junctions of the exon2-exon2' using 20 ng of cDNA from 3 LCL control. Relative gene expression has been normalized to C1 fixed at 1. Cycle Threshold (Ct) are presented in supplemental figure 1C. (E) Detection of protein-protein interaction using nonimmune IgG and BirA-mediated biotinylation. Each protein which binds to VHL is biotinylated and bound on streptavidin resin. Five partners of VHL have been tested: cullin 2, Elongin C (EloC), ERKs (extra-cellular signal-regulated kinases), Prefoldin 3 (VBP1) and Prefoldin 1. Quantification of western blot is expressed on percentage of biotinylated partners normalized to VHL213 WT fixed at 100%.

- **Identification of genetic variants in the new E2' cryptic exon in patients with erythrocytosis**

After the identification of the new *VHL* cryptic exon E2', we have sequenced this exon in patients with polycythemia with heterozygous mutation or without mutation and in patients who develop pheochromocytomas only (COMET collection). We identified 2 variants in patients with polycythemia: a synonymous mutation c.464-1016G>C, *VHL-X2*-p.Leu177Leu in one patient (functional study with splicing reporter assay are in process) and a total deletion of E2' exon in another one. One mutation: c.464-1014C>T, *VHL-X2* p.Ser178Leu was also found in one patient with a pheochromocytoma developed at 18 years. This mutation led to a modification of the amino acid sequence of X2 with a leucine to replace a conserved serine in the transmembrane domain of the heme-oxygenase 2 and the TMEM116.

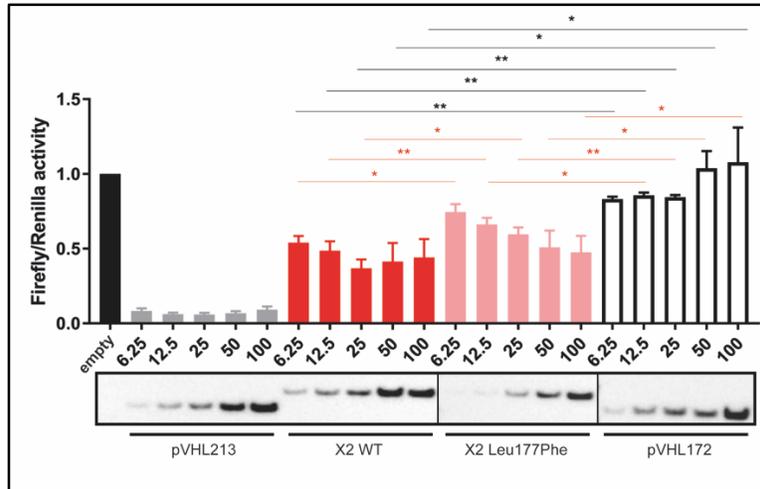
LISTE VARIANTS IN VHL-E2' (c.464-973 to c.464-1083)							
Patient ID	Phenotype	Nucleotide position	Consequence on X2 protein	Consequence on splicing	SNP	MAF	Localisation
16C5482	Polycythemia. EPO: low	c.464-1016C>G	Synonymous VHL-X2-p.Leu177Leu	New Donor site, modification of ESE, ESS	/	/	
17C1813	Hb: 18.6g/dL, Ht: 57%, GR: 6.5M/mm, EPO: 14-2mU/mL	del E2'	/	/	/	/	
PGL577001 PGL1060001 PGL1110001 PHEO793001 PHEO1439001 KREN060001 KREN127001 VHL014001 PHEO1148001	Cervical PGL multiple at 45 yrs Cervical PGL multiple at 50 yrs Cervical PGL at 31 yrs PHEO at 27 yrs PHEO at 26 yrs RCC at 35 yrs RCC at 2	c.464-1013 G>C	Synonymous VHL-X2-p.Ser178Ser		rs189039866	C=0.0018/9 (1000Genomes) C=0.0020/9 (Estonian) C=0.0042/129 (GnomAD) C=0.0062/773 (TOPMED)	chr3:10148774
PHEO915001	PHEO at 18 yrs	c.464-1014 C>T	VHL-X2-p.Ser178Leu	Modification of ESE, ESS	rs184276409	T=0.0001/3 (GnomAD) T=0.0001/9 (TOPMED) T=0.0008/4 (1000Genomes)	chr3: 10148773

Normal  
Ht, %, Nm =40–52; Nf =37–47  
Hb, g/dL, Nm =13–18; Nf =12–15  
RBC, million (M)/mm<sup>3</sup>; Nm =4.2–5.7,  
Nf =4.2–5.2  
EPO, mU/mL, N=5–25

**Figure 2 : Liste of variants in exon E2'.** Positions of variants identified in E2' SNP : nucleotidique polymorphisme, MAF : allelique frequence, Ht : hematocrit, Hb : hemoglobin, RBC : Red Blood Cells, Nf : complete blood count.

- **Functional study of VHL E2'**

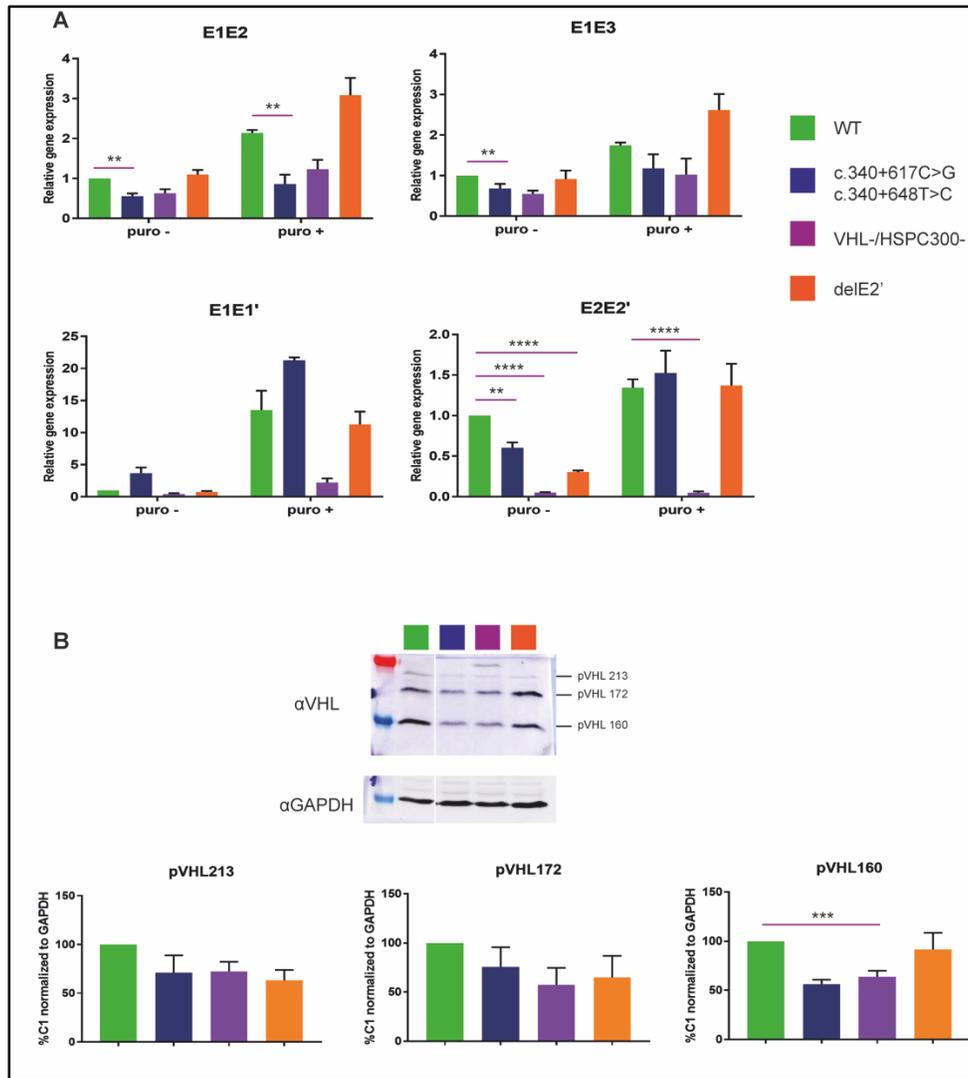
We performed functional study of the hypothetical X2 WT in the hypoxia pathway using hypoxia response element-dependent reporter assay (Figure 3A). We compared the function of X2 with the function of pVHL 213 (WT) and pVHL172 ( $\Delta$ E2) in addition to a mutant targeting a conserved amino acid of the HO-2 domain (c.464-1018C>T, *VHL-X2* Leu177Phe). The functional test revealed that X2 have an intermediate effect in the hypoxia pathway. X2 is less capable to degraded HIF that pVHL213 but significantly more capable to degrade HIF that pVHL172. In addition, we show that the variant c.464-1018C>T, *VHL-X2* Leu177Phe conferred a loss of function of the X2 protein to degrade HIF.



**Figure 3: Functional study of synonymous mutation in *VHL* E2'.** Functional hypoxia response element (HRE)-dependent reporter assays were performed in 786-O cells (*VHL*<sup>-</sup> cells that constitutively express HIF-2 $\alpha$ ). The results are expressed as relative firefly luciferase activity with *Renilla* luciferase as an internal control; 1.0 unit denotes the basal activity of endogenous HIF-2 $\alpha$  using the HRE luciferase reporter plasmid. The ability of WT and mutated X2 to downregulate firefly luciferase activity (related to HIF activity) was compared with pVHL. An immunoblot using an antibody specific to the hemagglutinin tag (HA) was used to detect HA-VHL, HA-X2 and HA- $\Delta$ E2. Three independent experiments were performed. \*P<0,099, \*\*P<0,01, based on Student *t* test.

#### - Expression study of the new *VHL* isoform E2' in patient's cells

We performed quantitative RT-PCR using TaqMan probes specific for the different splicing junctions of *VHL*: E1E2, E1E3, E1E1' and E2E2' (Figure 4A). We studied LCL of the patient with deletion of E2' of the *VHL* gene at the origin of erythrocytosis (1 patient in orange). As control we used LCL with mutations in E1' or deletion of *VHL*, at heterozygous state, at the origin of *VHL* disease (2 patients respectively in blue and in purple). In the LCL carrying the deletion of E2', we do not observe any significant difference for the expression of the different isoforms of *VHL*. For the E2E2' isoform, we show a decreased expression of this isoform in many LCLs of patient with a mutation compared to WT excepted for mutation c.340+648T>C associated with c.340+617C>G and surprisingly for the heterozygous deletion of E2'. We performed western blots with protein extracts of the LCLs of patients with an antibody against the exon 1 of *VHL* which can detect all the isoforms of *VHL* (Figure 4B). There is no significant difference for the isoform pVHL213 and 172. For the isoform pVHL172, there is a significant decrease of the expression only for the patient with a total deletion of one allele of *VHL*.



**Figure 4: Expression study of mutations in VHL E1' and E2' cryptic exons.** Patient in orange green develop an erythrocytosis and patients in blue and purple develop the VHL disease, green is the WT control. (A) Taqman quantification of the different VHL isoforms in LCLs cultured in the absence or presence of puromycin (puro), an inhibitor of nonsense-mediated mRNA decay. Taqman probes were specific to the VHL E1-E1, E1E3, E1E1' and E2E2' junctions. The graph resumes experiments performed on LCLs cultured independently 3 times and quantified in duplicate. Data are represented as the mean +/- standard error of the mean (SEM) and normalized to LCL WT (C1) fixed at 1 for each isoform. (B) Immunoblot analysis of patient's LCLs. A representative immunoblot and quantification of 3 different immunoblot are presented. A VHL antibody that recognized VHL E1 was able to detect pVHL213, 172 and 160. Relative gene expression was normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression, and results obtained with LCL control (WT) were fixed at 100%. Data are represented as the mean +/- standard error of the mean (SEM).

## Discussion

The full molecular mechanism at the origin of the two pathologies associated with *VHL* mutations are not elucidated. There are complex correlations between phenotype and genotype in *VHL*-related disorders. To date, the hypoxia pathway plays a central role in the development of the pathologies and is the most studied pathway. Functional studies are designed for this pathway and for missense mutations. It has been shown that there is a continuum model of tumor suppression by *VHL* (13). The more the loss of function of *VHL* is important, the more the pathology is severe and *vice versa*. Recently, synonymous mutations with no impact on the sequence of the protein have been studied (4), (14). It has been shown that these mutations are at the origin of a splicing defect of the *VHL* gene. On the same time, it has been shown that the structure of the *VHL* gene is more complex than presumed. In addition to the three described exons of the gene, we have discovered a new cryptic exon named E1' into the first intron of the gene. The cryptic exon has been found mutated in the context of erythrocytosis or *VHL* disease. The mutations led to a retention of this exon in the transcripts. The mechanism by which this new exon impacts the hypoxia pathway is not yet elucidated. During the investigation of the regulation of the splicing of the *VHL* gene, we discovered a second cryptic exon named E2', localized into the second intron of the gene. This exon is weakly expressed in tissues.

If this isoform is transduced into a protein, (termed X2), the domain encoded by the E2' exon has an important homology to a domain of two transmembrane proteins: the TMEM116 and the heme oxygenase 2. We showed that this domain does not induce a delocalization of the *VHL* protein at the membrane of the cells. X2 is localized, like pVHL, in the cytoplasm.

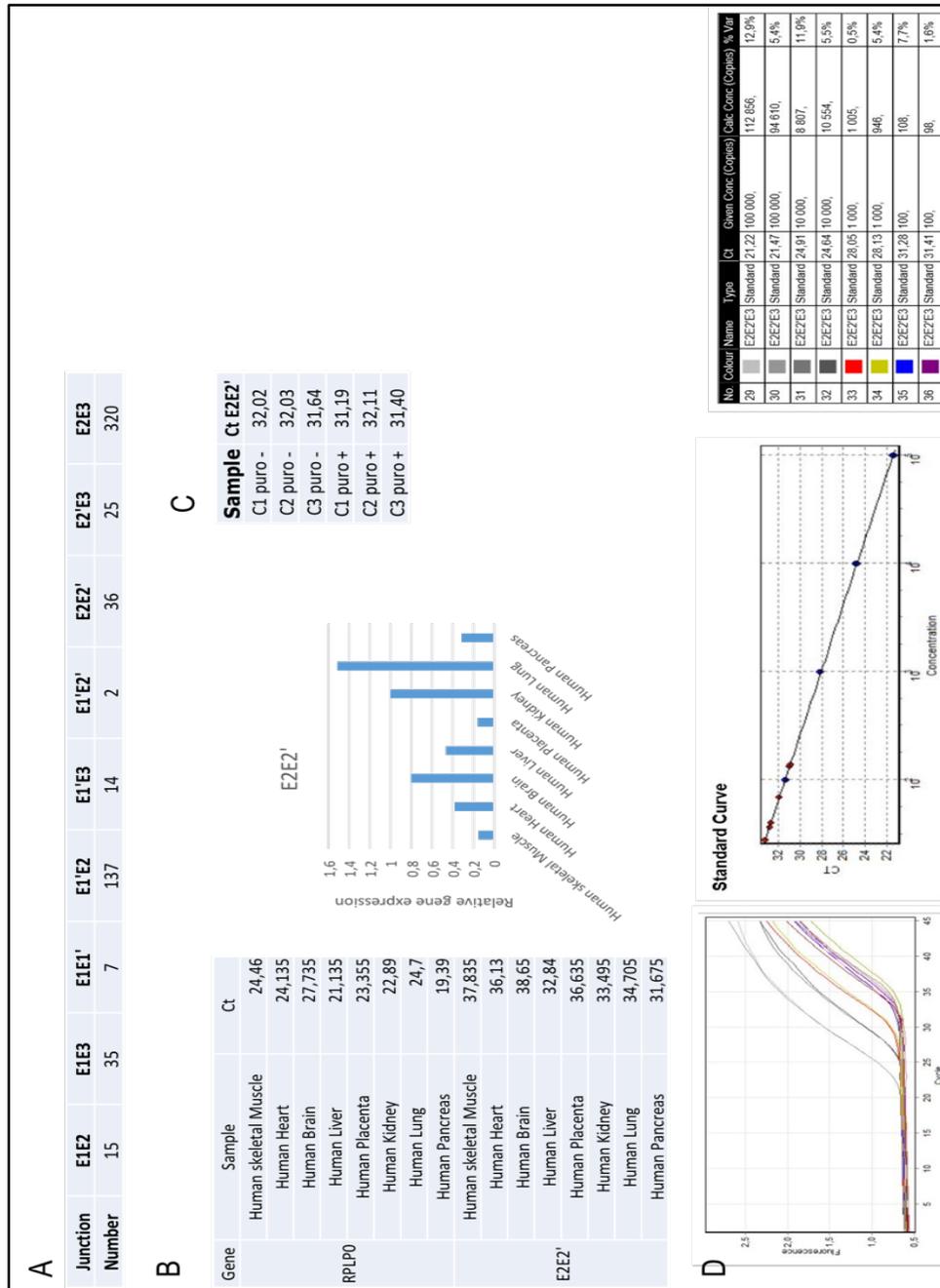
We have shown that X2 can bind to the cullin 2 and the elongin C but far less than pVHL213. This default of binding to the two partners of the E3 ubiquitin ligase complex may explain the difference of activity between pVHL213 and X2 showed in the functional study. Indeed, this test showed that X2 is less able to pVHL213 to inhibit HIF activity. If there is less E3 ubiquitin ligase complex formed, this may explain this difference.

The E2' cryptic exon has been found mutated in patients with erythrocytosis or with pheochromocytoma. The first results showed that deletion of this exon does not induce dysregulation of the other VHL isoforms (neither of transcriptional level, nor the proteins). LCL with mutations in E2' need to be studied. If the mutations do not impact splicing, and if they play a causal role in the occurrence of the related diseases, it could be, by *via* an impact on X2 function. Therefore, the role of X2 need to be identified. It would be very interesting to add this new cryptic exon, like E1', in the list of genes investigated in NGS panel for patients with unsolved cases of congenital erythrocytosis, inherited kidney cancer, hereditary pheochromocytoma/ paraganglioma syndrome and hemangioblastomas, in addition to all types of sporadic tumors with altered hypoxia signaling.

## References

1. Gordeuk, V. R. et al. Congenital disorder of oxygen sensing: association of the homozygous Chuvash polycythemia VHL mutation with thrombosis and vascular abnormalities but not tumors. *Blood* 103, 3924–3932 (2004).
2. Ang, S. O. et al. Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia. *Nat. Genet.* 32, 614–621 (2002).
3. Bento, C. Genetic basis of congenital erythrocytosis. *Int. J. Lab. Hematol.* 40, 62–67 (2018).
4. Lenglet, M. et al. New lessons from an old gene: complex splicing and a novel cryptic exon in VHL gene cause erythrocytosis and VHL disease. *Blood* (2018). doi:10.1182/blood-2018-03-838235
5. Maher, E. R. et al. Clinical features and natural history of von Hippel-Lindau disease. *Q. J. Med.* 77, 1151–1163 (1990).
6. Latif, F. et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* 260, 1317–1320 (1993).
7. Schoenfeld, A., Davidowitz, E. J. & Burk, R. D. A second major native von Hippel-Lindau gene product, initiated from an internal translation start site, functions as a tumor suppressor. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8817–8822 (1998).
8. Chesnel, F. et al. The von Hippel-Lindau tumour suppressor gene: uncovering the expression of the pVHL172 isoform. *Br. J. Cancer* 113, 336–344 (2015).
9. Hascoet, P. et al. The pVHL172 isoform is not a tumor suppressor and up-regulates a subset of pro-tumorigenic genes including TGFB1 and MMP13. *Oncotarget* 8, 75989–76002 (2017).
10. Stebbins, C. E., Kaelin, W. G. & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* 284, 455–461 (1999).
11. Maxwell, P. H. et al. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* 399, 271–275 (1999).
12. Iliopoulos, O., Kibel, A., Gray, S. & Kaelin, W. G. Tumour suppression by the human von Hippel-Lindau gene product. *Nat. Med.* 1, 822–826 (1995).
13. Couvé, S. et al. Genetic evidence of a precisely tuned dysregulation in the hypoxia signaling pathway during oncogenesis. *Cancer Res.* 74, 6554–6564 (2014).
14. Flores, S. K. et al. A synonymous VHL variant in exon 2 confers susceptibility to familial pheochromocytoma and von Hippel-Lindau disease. *J. Clin. Endocrinol. Metab.* (2019). doi:10.1210/jc.2019-00235
15. Ladroue, C. et al. Distinct deregulation of the hypoxia inducible factor by PHD2 mutants identified in germline DNA of patients with polycythemia. *Haematologica* 97, 9–14 (2012).
16. Gaildrat, P. et al. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol. Biol. Clifton NJ* 653, 249–257 (2010).

Supplemental data



**Supplemental Figure 1: Expression study of the new VHL isoform.** (A) Number of spliced junction of VHL transcripts in a RNAseq of H9 cells differentiated in neural progenitor cells. (B) TaqMan quantification of the transcript contain the E2E2' junction in a bank of cDNA of human tissues. Cycle threshold is presented at the left. Results have been presented in relative gene expression normalized to the expression in the kidney fixed at 1. (C) Cycle threshold of quantitative PCR of LCLs control for the quantification of the E2E2' junction in VHL transcripts. (D) Expression Curve, Standard curve and table of dilution of plasmids containing the E2E2' junctions used for the validation of the probes specific to E1E2' in TaqMan quantification.

## 4. Mise en place d'un modèle de cellules productrices d'EPO à partir de cellules souches pluripotentes induites de patients

### a. Introduction

Comme évoqué précédemment, les outils permettant la modélisation de la maladie de VHL ou la polyglobulie de Chuvash sont très limités. En effet, il n'existe à ce jour que des modèles animaux spécifiques de certaines tumeurs pour la maladie de VHL. En revanche, les modèles pour la polyglobulie de Chuvash sont plus robustes. De plus, il est très coûteux de produire un modèle murin pour chaque mutation que l'on souhaite étudier. Au niveau cellulaire, seules quelques lignées sont capables de produire de l'EPO mais il s'agit de lignées dérivées de cellules cancéreuses qui ont donc acquis de nombreuses modifications génétiques. Ces mutations peuvent impacter des voies de signalisation proches de la voie de l'hypoxie que l'on cherche à étudier et fausser les résultats obtenus. De plus, il faudrait générer des lignées stables par CRISPR d'une lignée pour chaque mutation que l'on souhaite étudier. Les lignées lymphoblastoïdes de patients ne permettent, quant à elles, l'analyse de la voie de l'hypoxie en amont de HIF du fait de la modification de celles-ci par le virus EBV. En effet, après transformation, les lignées expriment l'oncoprotéine virale LMP1 qui stabilise HIF.

Afin d'étudier au mieux les mécanismes moléculaires à l'origine de ces pathologies, il est important de disposer d'un meilleur modèle d'étude qui permettrait de mimer le développement de la maladie. Pour cela, il est possible d'utiliser l'un des derniers outils cellulaires, à savoir les cellules souches pluripotentes induites (iPS) reprogrammées grâce aux quatre facteurs de Yamanaka : Oct3/4, Sox2, KLF4 et c-Myc<sup>222</sup> à partir de cellules patients. Ces facteurs sont des protéines exprimées uniquement dans les cellules souches et qui permettent le maintien de leur caractère souche, notamment en empêchant l'expression de gènes qui caractérisent les cellules différenciées. Quelques lignées de cellules issues de patients porteurs de mutations dans le gène *VHL* ont déjà été décrites<sup>223</sup>. Cependant, pour le moment, ces cellules sont uniquement utilisées à l'état iPS et ne permettent pas de voir l'impact des mutations dans les types cellulaires en cause dans les pathologies.

Comme évoqué en introduction, l'origine embryonnaire des cellules responsables du développement des pathologies VHL reste aujourd'hui très controversée. Cependant, plusieurs hypothèses convergent vers une origine de la crête neurale. Cela serait en accord avec le développement des deux pathologies distinctes à partir de mutations du gène *VHL*. Les patients atteints de la maladie de VHL développant des tumeurs du système nerveux central, il ne serait pas incohérent que les cellules productrices d'EPO, en cause dans la polyglobulie, aient la même origine, à savoir la crête neurale. Les seules cellules impliquées dans la maladie de VHL qui n'ont pas pour origine la crête neurale sont les cellules épithéliales en cause dans les cancers du rein. Cependant, celles-ci sont entourées par les cellules rénales productrices d'EPO. Notre hypothèse serait que ces cellules productrices d'EPO sécrètent des facteurs dans le microenvironnement qui favorisent l'oncogenèse des cellules épithéliales voisines.

Nous avons généré des lignées iPS de patients porteurs de la mutation *VHL*-R200W à l'état homozygote. L'un des patients (Patient 1) est originaire de Chuvashie et l'autre du Pakistan (Patient 2). Nous avons ensuite différencié ces lignées en cellules productrices d'EPO de la crête neurale ou en hépatocytes. Nous avons pu comparer les résultats obtenus avec ces cellules et ceux obtenus avec des cellules contrôles.

## b. Résultats

### i. *Caractérisation des lignées iPS*

Avant d'utiliser une lignée iPS, il est important de la caractériser pour valider sa bonne reprogrammation et donc sa capacité de pluripotence (différenciation dans les 3 types de feuilletts embryonnaires : ectoderme, mésoderme et endoderme). Pour cela, plusieurs techniques sont utilisées mais la plus robuste reste la différenciation dans les 3 feuilletts grâce à des kits commerciaux, ce que nous n'avons pas encore réalisé. Nous avons pour le patient 1, dont les iPS ont été générées à partir de cellules urinaires du patient, réalisé des marquages en immunofluorescence et réalisé du FACS (Figure 27 A). Pour les 2 clones, nous avons marqué par immunofluorescence les marqueurs de pluripotence Oct4 (Octomer-binding transcription factor 4) et TRA1-60 (T-cell receptor alpha locus). Comme nous pouvons le voir, ces marqueurs sont bien exprimés chez les 2 clones (1 et 4) qui vont ensuite être utilisés. Par FACS, nous avons validé

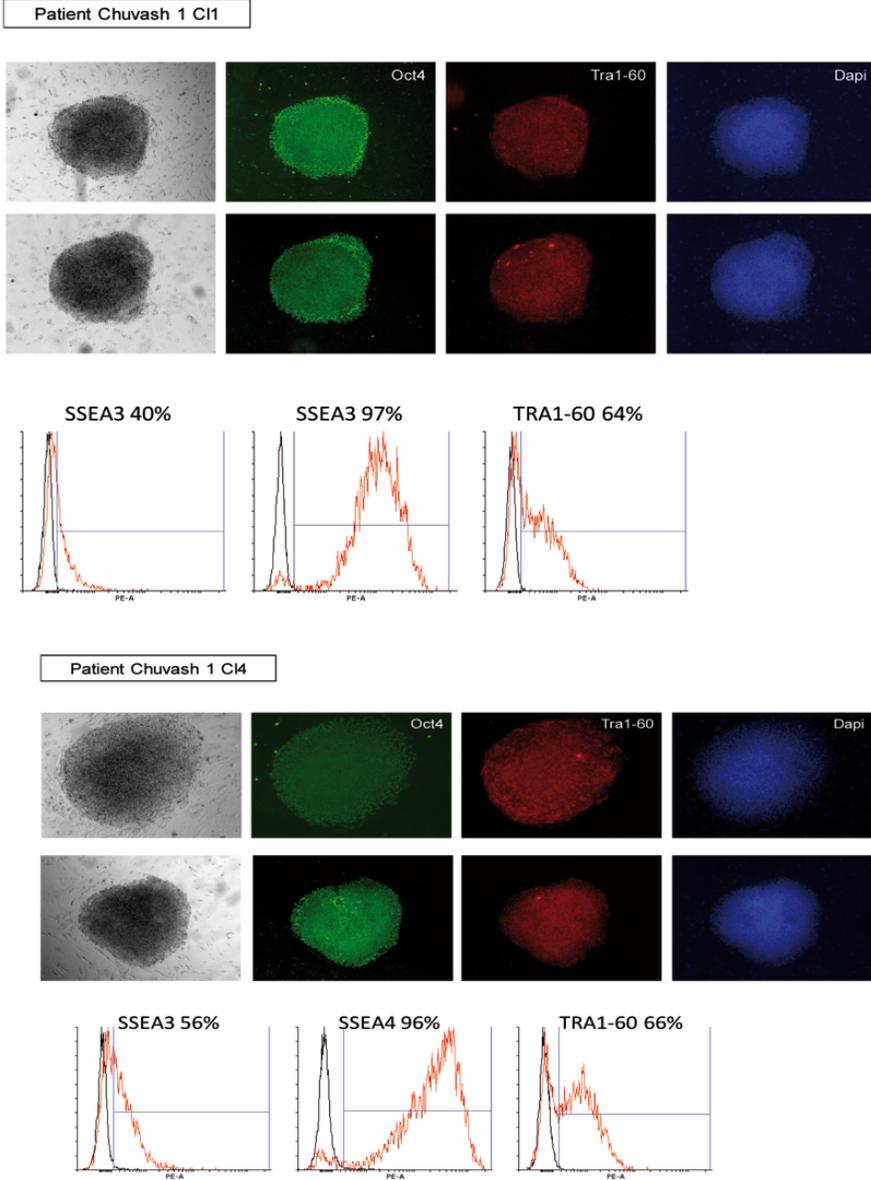
l'expression des marqueurs SSEA3 (stage specific embryonic antigen 3), SSEA4 (stage specific embryonic antigen 4) et TRA1-60. Nous voyons que pour les 2 clones, le marqueur SSEA4 est très fortement exprimé et que les marqueurs TRA1-60 et SSEA3 sont moyennement exprimés. Pour le patient 2, pour qui les cellules iPS ont été générées par la plateforme iPS de Nantes à partir de PBMC de patients, seules des qPCR ont été réalisées (Figure 27B). Elles nous permettent de voir que pour les 3 clones obtenus, le virus Sendai utilisé pour la reprogrammation des cellules a bien été éliminé des iPS (Log2 SeV). Nous voyons également que les 3 clones expriment bien les marqueurs de pluripotence Sox2, Oct4 et Nanog.

Les 2 clones de la lignées iPS contrôle utilisée ont été caractérisés et publiés précédemment<sup>224,225</sup>.

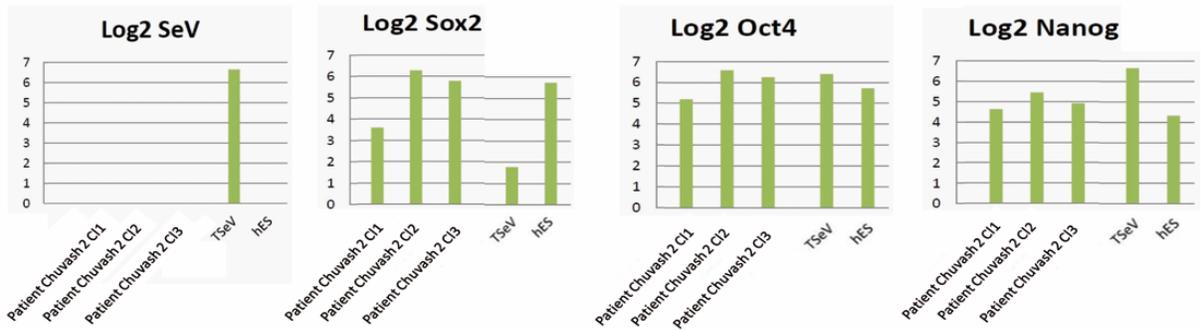
#### *ii. Protocoles de différenciation des lignées iPS*

Chez l'adulte, les cellules productrices d'EPO sont des cellules interstitielles du rein dont l'origine semble être la crête neurale<sup>33</sup>. Après plusieurs essais, nous avons pu voir qu'un certain type de progéniteurs de la crête neurale générés à partir d'iPS étaient capables de produire de l'EPO après une incubation de 24h en hypoxie (1% d'O<sub>2</sub>). Le protocole que nous avons utilisé est dérivé d'un protocole publié en 2015 par Denham<sup>218</sup>, les progéniteurs de la crête neurale sont obtenus sous forme de sphéroïdes après 11 jours de différenciation (Figure 28A). Entre J0 et J4, les cellules sont placées sur un support en laminine 521 spécifique de la différenciation neurale et cultivées en milieu Neurobasal et DMEMF12 supplémenté de N2 et de B27 ainsi que de SB431542 (inhibiteur de la voie du TGF- $\beta$ , il promeut la différenciation des cellules souches) et de CHIR99021 (inhibiteur de l'enzyme glycogen synthase kinase 3 (GSK-3), il favorise la transformation d'un type cellulaire en un autre) avec changement total à jour 2. Au jour 4, des « morceaux » de tapis cellulaires sont récupérés et placés dans des plaques 96 puits à faible attachement dans le même milieu contenant du FGF2 (Fibroblast Growth Factor) et du BMP2 (Bone Morphogenic Protein-2) jusqu'au jour 11, avec un changement partiel à jour 8. Entre jour 10 et jour 11, les cellules sont soit gardées en normoxie (condition normoxie) soit placées à 1% d'O<sub>2</sub> (condition hypoxie).

A

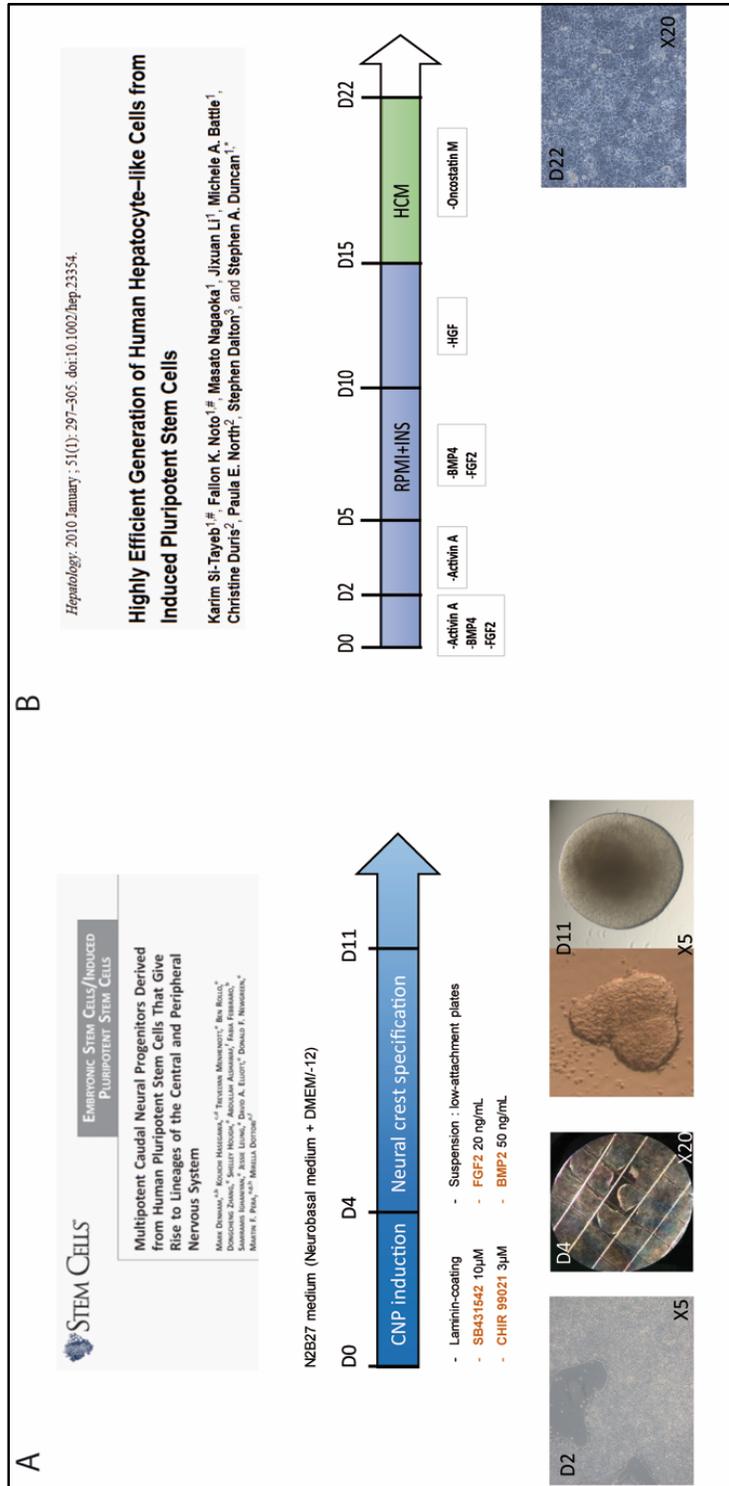


B



**Figure 27 : Caractérisation des lignées iPS.** (A) Caractérisation des lignées Uhips (cellules iPS dérivées de cellules urinaires) du patient Chuvash 1 pour les clones 1 et 4 par marquages immunofluorescents des marqueurs de pluripotence Oct4, Tra1-60 et du noyau par DAPI. Les cellules sont platées sur des MEFs (fibroblastes d'embryons de souris) 72 heures avant le marquage. Les cellules sont également caractérisées par cytométrie en flux pour les marqueurs de pluripotence SSEA3, SSEA4 et TRA1-60. (B) Caractérisation des lignées iPS dérivées de fibroblastes du patient Chuvash 2 pour les clones 1, 2 et 3 par qPCR pour les marqueurs de pluripotence Sox2, Oct4 et Nanog et pour le marqueur viral SeV. Les résultats sont présentés en quantité relative.

Les hépatocytes sont le 2<sup>ème</sup> site de production d'EPO chez l'adulte (production d'environ 10% de l'EPO circulante) et le 1<sup>er</sup> site de production d'EPO durant l'embryogenèse (production de 100% de l'EPO avant le fonctionnement des cellules rénales productrices d'EPO). Nous avons réalisé des essais et vu que les hépatocytes dérivés d'iPS obtenus avec le protocole de K. Si-Tayeb publié en 2010<sup>217</sup> étaient capables de produire de l'EPO après une incubation de 24h à 1% d'O<sub>2</sub> (Figure 28B). Ce protocole dure 22 jours, les cellules sont déposées sur un support de matrigel. Le milieu RPMI de J0 à J15 et HCM (Hepatocytes Culture Medium) de J15 à J22 est changé tous les jours avec addition de différentes cytokines selon le protocole publié. Entre J21 et J22, les cellules sont laissées en normoxie (condition normoxie), ou placées à 1% d'O<sub>2</sub> (condition hypoxie).



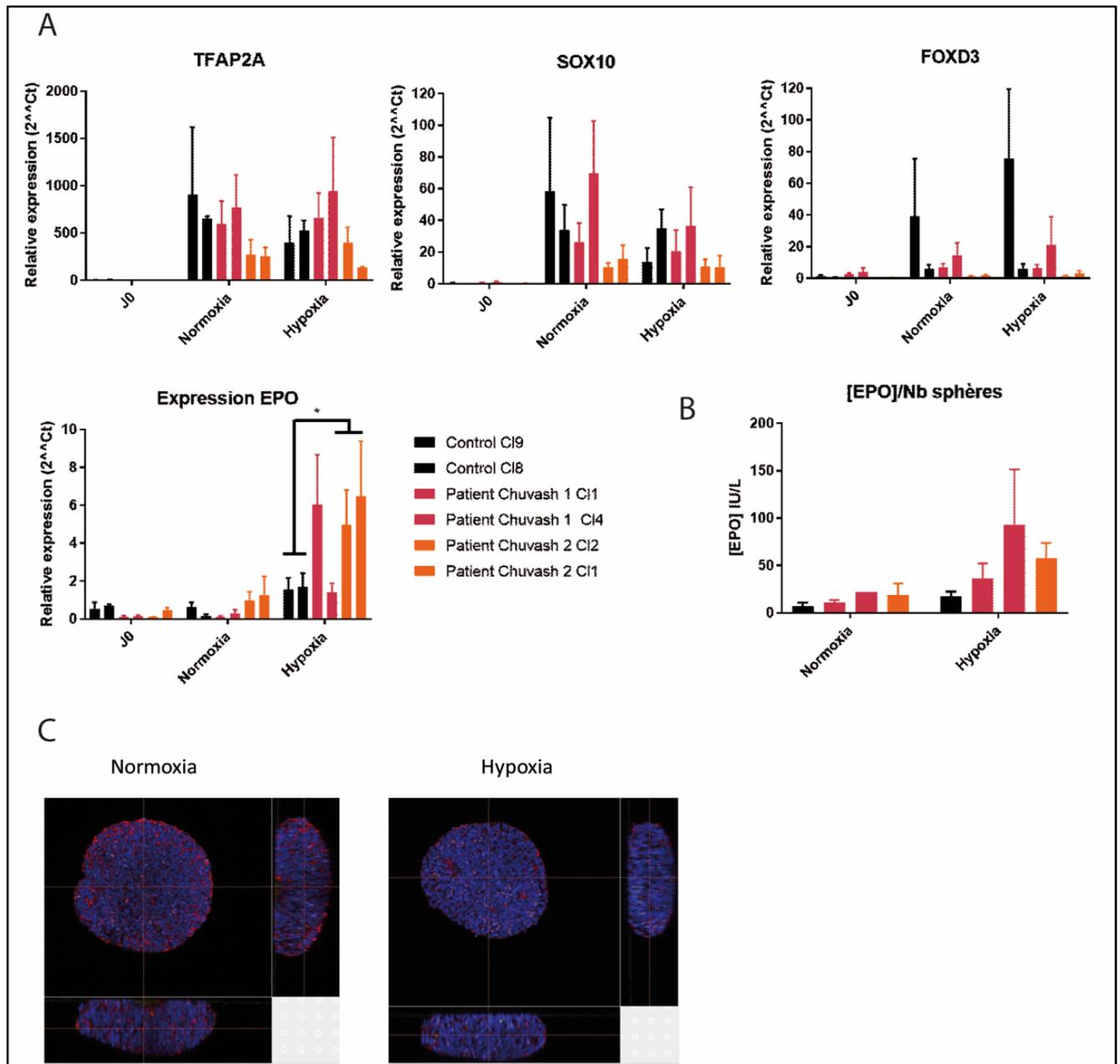
**Figure 28 : Protocoles de différenciation.** (A) Protocole de différenciation en progéniteurs de la crête neurale producteurs d'EPO. Entre J10 et J11, les cellules sont soit laissées en conditions normoxiques, soit placées en hypoxie. (B) Protocole de différenciation en hépatocytes. Entre J21 et J22, les cellules sont soit laissées en conditions normoxiques, soit placées en hypoxie.

### *iii. Caractérisation des progéniteurs de la crête neurale*

Après les 11 jours de différenciation en progéniteurs de la crête neurale, le surnageant cellulaire est récolté pour quantifier l'EPO secrétée et les cellules sont lysées afin d'en extraire l'ARN. On réalise ensuite des RT-qPCR TaqMan afin de quantifier les transcrits des marqueurs de différenciation, cela nous permet de valider cette dernière en comparant les expressions à J0 et à J11 (Figure 29A). Nous pouvons voir qu'à J0, les marqueurs de différenciation utilisés ne sont pas exprimés alors qu'ils le sont à J11 en normoxie et en hypoxie. Les facteurs utilisés sont le facteur de transcription AP-2 Alpha (TFAP2A) qui joue notamment un rôle important dans le développement du tube neural, SOX10 qui est un facteur de transcription jouant un rôle dans le développement embryonnaire et la myélinisation du système nerveux central et FOXD3 qui est un facteur de transcription qui joue un rôle dans le développement de la crête neurale à partir du tube neural. Au vue de l'expression des différents facteurs à J11, les différenciations sont validées. On quantifie également la quantité de transcrits de l'EPO dans les cellules en normoxie ou en hypoxie (Figure 29B). Nous pouvons voir qu'il n'y a pas de différence significative de production d'EPO par les cellules avant différenciation et après différenciation en normoxie, ni entre les contrôles et les patients en conditions normoxiques. En revanche, après une incubation en hypoxie, nous pouvons voir que les cellules progénitrices de la crête neurale des patients produisent significativement plus d'EPO que les contrôles, sauf pour le clone 4 du patient Chuvash 1 qui paradoxalement semble être le mieux différencié des clones de patients au vue de l'expression des marqueurs de différenciation.

Nous avons ensuite quantifié l'EPO secrétée par les cellules après différenciation en rapportant cela par rapport au nombre de sphères obtenues à l'issue de la différenciation (Figure 29B). Nous pouvons voir que malgré un taux plus faible de transcrits, le clone 4 du patient Chuvash 1 est celui qui secrète le plus d'EPO après une incubation de 24h en hypoxie. Nous voyons en normoxie une faible augmentation de la sécrétion d'EPO des cellules de patients par rapport aux cellules contrôle. En hypoxie, nous pouvons voir qu'il y a une augmentation de la production d'EPO pour toute les cellules et que celle-ci est plus importante pour les cellules de patients que pour les contrôles.

Des marquages par immunofluorescence ont été réalisés sur des cellules progénitrices de la crête neurale du patient Chuvash 1 clone 1 (Figure 29C). Pour cela, les sphéroïdes ont été transparisés afin de pouvoir visualiser l'EPO dans les cellules situées au centre des sphères. Un co-marquage des noyaux par DAPI et des cytosquelettes d'actine par la Phalloïdine ont été réalisés. L'EPO étant marquée par un fluorochrome situé dans le rouge lointain, il apparaît en violet sur les images et est donc peu visible sur l'image combiné des 3 fluorochromes. En revanche, lorsque l'on visualise l'EPO seul des 2 sphères, nous pouvons voir que la sphère placée en hypoxie produit plus d'EPO que la sphère qui est restée en normoxie durant les 24 dernières heures de la différenciation.

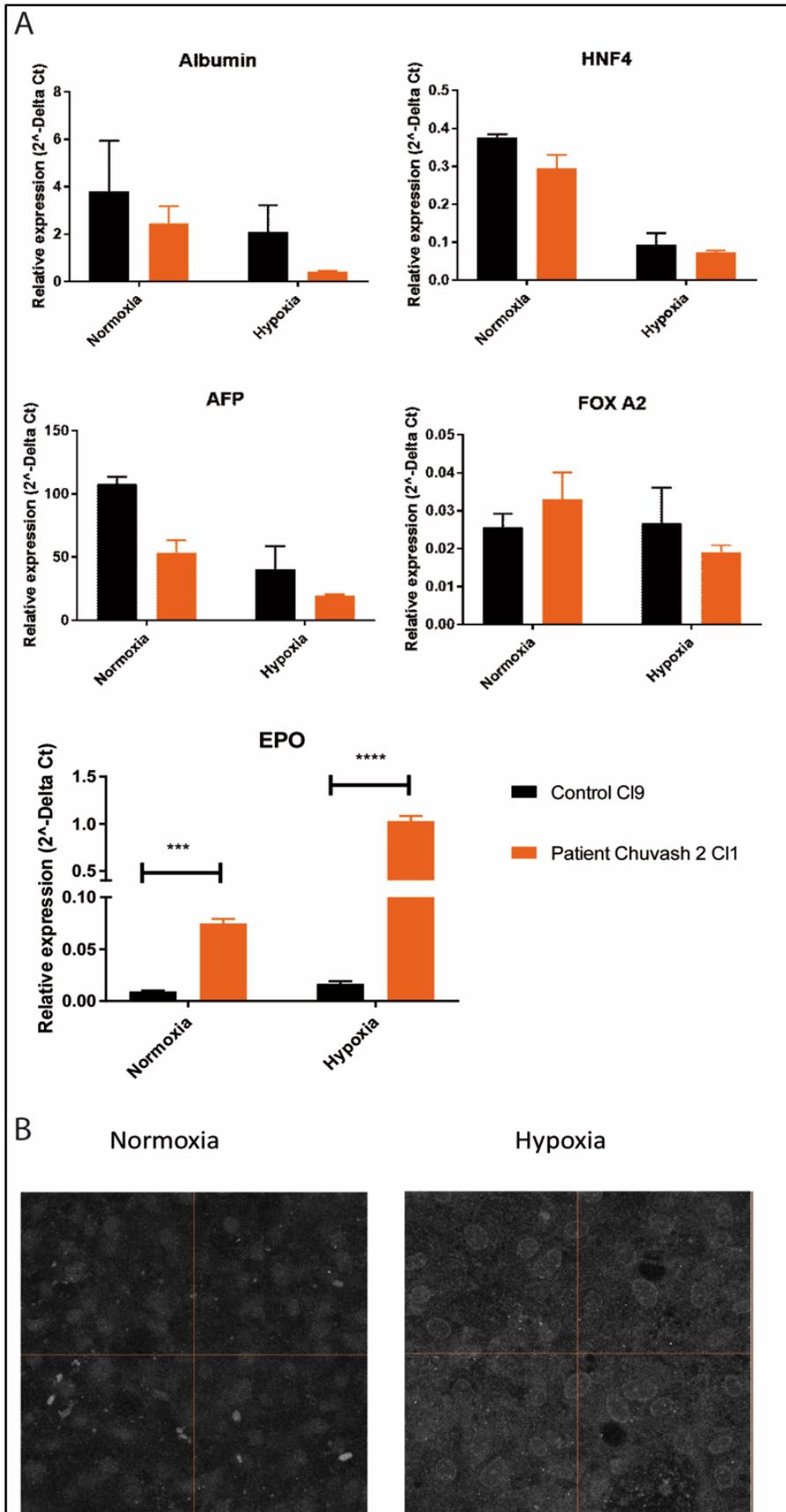


**Figure 29 : Caractérisation des progéniteurs de la crête neurale différenciés à partir d'iPS contrôle et de patients.** (A) Quantification par RT-PCR TaqMan de 3 marqueurs de différenciation spécifiques des progéniteurs neuraux TFAP2A, SOX10 et FOXD3 et quantification des transcrits d'EPO produits par ces progéniteurs en normoxie ou en hypoxie durant les 24 dernières heures de différenciation. Résultats présentés en expression relative normalisées par rapport à la moyenne des triplicats du contrôle clone 9 fixée à 1. (B) Dosage de l'EPO sécrétée dans le milieu de culture des sphères après concentration de l'EPO et rapporté par rapport au nombre de sphères. Résultats présentés en IU d'EPO par litre de milieu de culture. (C) Marquage par immunofluorescence de l'EPO (violet) contenu dans les sphères du patient Chuvash 1 en Normoxie ou en Hypoxie en référence aux noyaux (DAPI en bleu) et à l'actine (rouge).

#### *iv. Caractérisation des différenciations hépatiques*

Après 21 jours de différenciation, les surnageants des cellules hépatiques sont récupérés et celles-ci sont lysées pour en extraire l'ARN. Des RT-qPCR TaqMan sont ensuite réalisées pour quantifier les marqueurs de différenciation (Figure 30A). Les marqueurs utilisés sont l'albumine qui est une protéine de transport du sang qui est spécifiquement produite par des hépatocytes différenciés, HNF4A qui est un facteur nucléaire hépatocytaire, FOXA2 qui est un facteur de transcription spécifique des hépatocytes et l'AFP qui est l'alpha fétoprotéine du plasma qui est produite spécifiquement par le foie fœtal. Ici, nous pouvons voir que ces 4 marqueurs sont bien exprimés dans la lignée contrôle et dans la lignée de patient. Nous pouvons noter une diminution de ces marqueurs en hypoxie.

La quantification des transcrits de l'EPO nous montre pour les contrôles qu'il y a une légère augmentation d'EPO en hypoxie par rapport à la normoxie. En revanche, en normoxie, nous pouvons voir que les cellules dérivées du patient Chuvash 2 clone 1 produisent significativement plus d'EPO que les contrôles. Cette observation est validée en hypoxie par une augmentation encore plus significative de la production d'EPO par les cellules de patients que par les contrôles. Des marquages par immunofluorescence de l'EPO ont ensuite été réalisés sur des cellules contrôles clone 9 après 21 jours de différenciation placées 24h en normoxie ou en hypoxie pour voir l'expression protéique de l'EPO (Figure 30B). Nous pouvons voir que les cellules placées en hypoxie produisent plus d'EPO que les cellules restées en normoxie, ce qui reflète les observations faites au niveau des transcrits.



**Figure 30 : Caractérisation des cellules de foie différenciées à partir d'iPS contrôle et du patient Chuvash 2.** (A) Quantification par RT-PCR TaqMan des marqueurs de différenciation hépatique Albumine, HNF4, AFP et FOXA2 et des transcrits de l'EPO produits par les cellules de foie immature en normoxie ou en hypoxie. Les résultats sont présentés en expression relative ( $2^{(-\Delta Ct)}$ ) normalisée avec RPL13A pour les marqueurs de différenciation et RPLPO pour l'EPO. La significativité est exprimée avec un t.test : \*\*\* $P < 0,0001$ , \*\*\*\* $P < 0,00001$ . (B) Marquage immunofluorescent de l'EPO des cellules hépatiques de contrôle clone 9 placées 24h en normoxie ou en hypoxie.

### c. Discussion

Les différents clones d'iPS des deux patients générés au laboratoire doivent être mieux caractérisés avant de pouvoir publier des données sur ces cellules. Il faut notamment réaliser un test de pluripotence par différenciation dans les 3 feuillets embryonnaires soit grâce à des kits de différenciation commerciaux dits « tri-lineage » soit par la réalisation d'un teratome sur des souris à partir de ces cellules. Il faut également réaliser un caryotype afin de vérifier qu'il n'y a pas eu de réarrangements chromosomiques majeurs au moment de la reprogrammation.

Nous avons utilisé deux modèles de cellules productrices d'EPO à savoir du foie non mature, permettant de mimer du foie fœtal, et des progéniteurs de la crête neurale pour essayer de se rapprocher des cellules productrices d'EPO rénales.

Lorsque *VHL* est muté, cela peut conduire à une perte de fonction de la protéine pour la régulation de HIF $\alpha$ , qui est visible en normoxie *via* une augmentation de l'expression des gènes régulés par HIF. Dans nos modèles, cette observation est faite pour les différenciations hépatiques pour lesquelles les cellules du patient Chuvash 2 produisent significativement plus d'EPO que les cellules contrôles en normoxie et en hypoxie. En revanche, pour le modèle de cellules progénitrices de la crête neurale, nous ne voyons pas de différence significative entre les patients Chuvash et les contrôles en normoxie. Le taux d'EPO produit en normoxie par ces cellules étant plus faible que celui produit par les cellules hépatiques, nous pouvons penser que même avec une perte de fonction, la protéine pVHL-R200W est encore suffisamment active pour réguler HIF $\alpha$  dans ces cellules. Pour valider cette hypothèse, il faudrait quantifier HIF dans ces deux types

cellulaires afin de voir s'il est plus exprimé dans les cellules hépatiques que dans les cellules productrices d'EPO.

En revanche, les deux modèles nous permettent d'observer une différence entre les patients Chuvash et les contrôles en hypoxie, là où HIF $\alpha$  est fortement stabilisé du fait de l'absence de l'oxygène, cofacteur des PHD pour son hydroxylation. VHL n'intervenant pas dans la régulation de HIF $\alpha$  en hypoxie puisque cette dernière n'est pas hydroxylée, la différence observée pourrait être due à une adaptation des cellules à la baisse d'O<sub>2</sub>. En effet, HIF induit l'expression des PHD qui permettent au bout d'un certain temps d'hydroxyler HIF avec le peu d'O<sub>2</sub> présent. HIF redevient alors la cible de VHL. Nous avons pu noter que durant les 24h à 1% d'O<sub>2</sub> les cellules sont capables de s'adapter à ce taux d'oxygène. En effet, lorsque l'on réalise des cycles successifs normoxie/hypoxie, nous observons une production d'EPO beaucoup plus importante. Cela signifie que les cellules sont capables d'utiliser le peu d'oxygène présent pour dégrader HIF.

La deuxième hypothèse serait que la différence observée pourrait être due à une autre fonction de VHL, par exemple la régulation de l'apoptose. Un test de viabilité cellulaire pourrait être réalisé dans un premier temps pour normaliser les résultats obtenus par rapport aux cellules viables. *VHL* est un suppresseur de tumeurs qui permet de réguler la sécrétion de facteurs anti-apoptotiques ou pro-apoptotiques dans les cellules tels que P53 ou NF $\kappa$ B. Quand *VHL* est muté, comme ici pour les 2 patients Chuvash, il est possible qu'il y ait une inhibition de l'apoptose en réponse au stress cellulaire. Placées en hypoxie, les cellules sont soumises à un stress, ce qui conduit, en conditions normales, à une mortalité cellulaire importante. La forte augmentation de l'EPO observée en hypoxie pour les mutants peut en partie être due à une inhibition de l'expression de facteurs anti-apoptotiques et à une augmentation de l'expression de facteurs pro-apoptotiques en réponse au stress hypoxique qui ne peuvent pas être régulés par la protéine VHL-R200W, alors qu'ils le sont par pVHL WT. Pour valider l'hypothèse, il faudrait quantifier au niveau protéique le facteur P53 serait diminué chez les patients Chuvash alors qu'il est stabilisé par VHL WT. Cette comparaison devrait être faite pour des cellules à J0 ainsi qu'à J11 placées toutes deux en normoxie et en hypoxie, ce qui nous permettrait de voir si les cellules non différenciées subissent le même stress hypoxique que les cellules différenciées.

Pour le patient Chuvash 1 différencié en crête neurale, nous pouvons noter une forte différence d'expression de l'EPO pour le clone 1, qui se comporte comme le patient Chuvash 2, et le clone 4, qui se comporte comme les clones contrôles. Cela pourrait s'expliquer par un défaut de reprogrammation de ce clone. En effet, des gènes ont pu être altérés au moment de la reprogrammation. Les caryotypes des lignées iPS établies au laboratoire n'ayant pas encore été réalisés, nous ne savons pas si des réarrangements chromosomiques ont eu lieu au cours de la reprogrammation. Il est donc important de mieux caractériser les lignées de patients Chuvash avant d'interpréter les résultats obtenus.

Cette étude reste la première étude de modèle iPS de cellules productrices d'EPO connue à ce jour étant donné qu'aucune donnée n'a encore été publiée sur un modèle de ce genre. Même s'il reste des caractérisations à faire avant de les valider, il semblerait que ces modèles soient prometteurs pour l'étude du développement des pathologies du spectre VHL avec l'obtention de cellules dérivées de patients qui sont donc porteuses des mutations que l'on souhaite étudier dans le fond génétique du patient. Pour l'heure, il semble ne pas y avoir de différence entre le patient porteur de l'haplotype core de Chuvash (patient 1) et celui ne le portant pas (patient 2). Nous envisageons également l'obtention de lignée iPS CRISPR pour une étude plus fine des différentes pertes de fonctions de *VHL*. Cela pourrait nous permettre de valider le gradient de perte de fonction corrélé avec la sévérité de la pathologie. Pour cela, des analyses de transcriptomique par RNAséq ou DGEséqu seront réalisées sur les cellules différenciées.

## IV. Conclusion et discussion générales

La problématique du laboratoire est de comprendre les mécanismes moléculaires à l'origine de deux pathologies de la voie de l'hypoxie : la polyglobulie et la maladie de VHL. Dans le cadre de ma thèse, je me suis intéressée à des mutations particulières du gène *VHL* retrouvées chez des patients atteints par l'une ou l'autre de ces pathologies. Jusqu'alors, la grande partie des études fonctionnelles réalisées sur des mutants *VHL* étaient pour des mutations impactant directement la séquence protéique. Cependant, dans certains cas, la modification seule de la séquence protéique ne suffisait pas pour comprendre les différences phénotypiques observées chez les patients.

Le premier objectif de ma thèse a été de déterminer si des mutations synonymes de l'exon 2 de *VHL* pouvaient être à l'origine des pathologies développées en modifiant, notamment, l'épissage des transcrits de *VHL*. L'étude s'est portée sur 3 familles avec une mutation c.429T>C, p.Asp143Asp (*VHL*-D143D) atteintes d'érythrocytose héréditaire (Familles 1, 9 et 10 de l'article de la partie 1 des résultats (article 1)) et sur 2 familles atteintes de la maladie de VHL et porteuses de la mutation c.414A>G, p.Pro138Pro (*VHL*-P138P) (Familles 11 et 12 de l'article 1). Nous avons pu mettre en évidence que ces deux mutations synonymes entraînent un saut de l'exon 2 dans les transcrits *VHL*, ce qui conduit à une inversion de la proportion des différents transcrits de *VHL*, à savoir une expression plus importante des isoformes contenant uniquement les exon 1 et 3 du gène en défaveur de l'isoforme sauvage contenant les 3 exons. Cette observation a pu être faite grâce à des RT-PCR dans les lignées lymphoblastoïdes (LLB) pour les patients atteints d'érythrocytose et dans les leucocytes et les tumeurs pour les patients atteints par la maladie de VHL. Une quantification exacte de ces deux transcrits a été faite dans les LLB et dans les leucocytes des patients porteurs de la mutation *VHL*-P138P. Nous avons ainsi pu montrer que chez les patients polyglobuliques porteurs de la mutation *VHL*-D143D à l'état homozygote, l'isoforme *VHL* sauvage est 4 à 5 fois moins exprimée que dans les cellules contrôle en faveur de l'isoforme *VHL*ΔE2 qui est environ 10 à 15 fois plus exprimée que chez les contrôles. Au niveau protéique,

nous avons identifié que les patients porteurs de la mutation *VHL*-D143D expriment seulement 10 à 20% d'isoforme pVHL213 par rapport aux contrôles et environ 25% de l'isoforme pVHL160. Contrairement à ce que nous avons pu voir au niveau des transcrits, l'isoforme pVHL172 est moins exprimée chez les patients (50% par rapport aux contrôles). Cela signifie que l'expression de moins de 25% des isoformes pVHL213 et 160 avec la mutation *VHL*-D143D est suffisante pour être viable. Cette expression dans les REPC conduit au développement d'une polyglobulie. En revanche, la mutation *VHL*-P138P est uniquement retrouvée à l'état hétérozygote chez les patients qui développent la maladie de VHL. Nous avons pu montrer que dans les LLB cela conduit à une augmentation de l'expression du transcrits *VHL*ΔE2 sans conduire à une diminution significative de l'isoforme sauvage. Au niveau protéique, nous observons une diminution d'environ 50% des 3 isoformes pVHL213, 160 et 172 contre seulement environ 30 à 40% pour la mutation *VHL*-D143D à l'état hétérozygote (patients porteurs sains). Dans les tumeurs, la mutation *VHL*-P138P est associée à une perte d'hétérozygotie (LOH) de l'allèle sauvage, seul l'allèle muté est donc exprimé. La clusterisation des résultats de RNAséq de cette tumeur, en comparaison avec une banque de phéochromocytomes porteurs de mutations de *VHL* ou d'autres gènes nous a permis de classer le phéochromocytome d'un patient porteur de la mutation *VHL*-P138P dans la catégorie des tumeurs spécifiques liées à une mutation dans *VHL*. Cela nous a permis de confirmer que cette mutation était bien l'origine de la pathologie chez les patients. Cette conclusion a également été confirmée par une étude menée sur la mutation *VHL*-P138P publiée récemment par une équipe américaine<sup>226</sup>.

Etant situées dans le même locus, nous nous sommes également intéressés à un éventuel défaut d'épissage causé par les mutations *VHL*-P138L et *VHL*-G144R identifiées précédemment chez les patients avec polyglobulie<sup>137, 227</sup>. Nous avons pu montrer pour *VHL*-P138L qu'en plus de la faible perte de fonction de la protéine décrite, la mutation conduit à un faible saut de l'exon 2 de *VHL*. Ce saut d'exon semble expliquer le phénotype observé pour la mutation *VHL*-G144R. En effet, au laboratoire, l'étude fonctionnelle réalisée pour cette mutation n'a jamais pu détecter de perte de fonction significative de ce mutant (données non publiées). Cela signifie qu'en plus de la perte de fonction de la protéine à l'origine de la pathologie, s'ajoute la dérégulation de l'épissage. Il serait

intéressant d'étudier en plus de l'impact sur la fonction de pVHL, l'impact sur l'épissage des mutations de l'exon 2 de *VHL* pour lesquelles on observe des corrélations génotype/phénotype complexes. En effet, certaines mutations qui ne présentent pas de perte de fonction de pVHL lors des tests fonctionnels peuvent en réalité entraîner un saut de l'exon 2. Ces perturbations de l'épissage pourraient conduire à une diminution de l'expression de pVHL qui explique les symptômes observés.

Nous avons donc pu montrer pour la première fois que des mutations de l'exon 2 de *VHL* pouvaient conduire à un défaut d'épissage qui se traduit par un saut de l'exon 2 du gène dans les transcrits. En effet, seules des mutations situées dans les sites consensus donneurs ou accepteurs d'épissage de l'exon 2 avaient déjà été rapportées et conduisent également à un saut de cet exon<sup>170</sup>. L'étude des mutants par minigènes rapporteurs nous a permis de montrer que les défauts d'épissage suivent également le modèle d'oncogenèse en gradient décrit pour *VHL*<sup>57</sup>. En effet, les mutations à l'origine de la maladie de VHL entraînent un saut de l'exon 2 plus important que celles à l'origine des polyglobulies, remarqué dans les tests rapporteurs minigène dans différents types cellulaires ainsi que dans le séquençage des ARN des LLB d'un patient polyglobulique comparé à celui d'une tumeur d'un patient qui a la maladie de VHL (Figure 3A article 1). Cela pourrait avoir pour conséquence une diminution plus importante de l'expression protéique de VHL dans certains tissus dans le cadre de la maladie de VHL que dans celui des polyglobulies. Il serait intéressant de quantifier l'expression de *VHL* dans les tumeurs des patients porteurs d'une mutation synonyme de *VHL*.

L'utilisation de siRNA dirigés contre la jonction E1E3 des transcrits *VHL* dans les LLB de patients porteurs d'une mutation impactant l'épissage de E2 nous a permis de rétablir partiellement l'expression de l'isoforme pVHL160. Une thérapie potentielle est donc envisageable avec des oligonucléotides antisens qui induiraient la dégradation des transcrits contenant uniquement les exons 1 et 3, comme cela a déjà été décrit pour d'autres pathologies<sup>228</sup>. Cette approche permettrait de favoriser l'expression d'une isoforme fonctionnelle de VHL à défaut d'une isoforme non fonctionnelle qui se retrouve majoritaire chez les patients.

Durant cette étude, nous avons décrit un patient qui développe une érythrocytose porteur d'une mutation *VHL*-D143D à l'état hétérozygote (Famille 1 de l'article 1), alors que son père, porteur de la même mutation, comme les parents hétérozygotes sains des familles 9 et 10 de l'article 1, ne développe pas la pathologie. Nous nous sommes donc intéressé à ce patient et avons pu, en réalisant une RT-PCR avec des amorces situées dans les exons 1 et 3 de *VHL*, révéler la présence d'isoformes de tailles supérieures à celle des isoformes sauvages exprimées en condition normale. Une première isoforme contenait un exon cryptique situé dans l'intron 1 du gène, nommée E1' et une deuxième isoforme contenait un exon cryptique situé dans l'intron 2, nommée E2'. Peu de temps après cette découverte, l'exon E1' a été décrit dans des bases de données (NCBI, ENSEMBL). Nous avons alors approfondi nos recherches sur cet exon cryptique.

Après avoir réalisé une étude d'expression qui nous a permis de montrer que cet exon est exprimé en faible proportion dans tous les tissus à l'état normal (majoritairement dans le placenta et le cerveau), nous avons pu montrer qu'il est conservé uniquement chez les primates supérieurs. Cette conservation uniquement chez les primates supérieurs peut s'expliquer par le fait que de part et d'autre de cet exon cryptique, on retrouve des séquences ALU, spécifiques des primates. Ces séquences ont pu permettre l'insertion de cet exon par translocation d'une autre partie du génome humain<sup>229</sup>. En effet, les séquences ALU ont permis l'insertion d'exons dans de nombreux gènes chez les primates, ce qui a affecté leur évolution par rapport aux autres espèces<sup>171</sup>. Il s'agit des séquences transposables les plus abondantes dans le génome humain. Il est très probable que ces pseudo exons soient régulés de façon négative dans la plupart des cas mais qu'il soient exprimés soit en réponse à un stimuli extérieur, soit uniquement au cours du développement ou dans un tissu précis. Cela explique que de plus en plus de ces pseudos exons sont actuellement retrouvés dans des gènes très connus. Dans le cadre des pathologies, plusieurs cas de création de site cryptique pour des pseudos exons ont déjà été rapportés<sup>230,213</sup>. Dans notre cas, nous avons identifié des mutations de cet exon cryptique chez des patients qui développent une polyglobulie, soit à l'état hétérozygote composite associé avec une autre mutation, soit à l'état homozygote. et chez des patients qui développent la maladie de VHL à l'état hétérozygote (Tableau 1 des résultats complémentaires de l'article 2)

L'insertion, en phase avec l'exon 1, de l'exon cryptique E1' conduit à l'insertion d'un codon stop prématuré à la fin de ce dernier. Cet exon entre donc dans la catégorie des « exons non-sens » d'après Dhir et Buratti<sup>171</sup>. Ce codon stop prématuré conduit à la dégradation rapide des transcrits dans les cellules *via* le mécanisme de nonsense mediated decay (NMD). Nous avons pu le montrer grâce à des tests puromycine, un inhibiteur du NMD, dans des LLB. Nous aurions également pu confirmer cette hypothèse en utilisant des siRNA dirigés contre un des acteurs du NMD afin de bloquer ce mécanisme comme cela a déjà pu être décrit<sup>231</sup>. Cela nous a également permis de montrer que les mutations portées par les patients conduisent à une rétention importante de cet exon en absence et en présence de puromycine. Cette rétention est d'autant plus accentuée pour la mutation homozygote c.340+816A>C. Nous pouvons penser que la rétention de E1' causée par les mutations conduit à une saturation du mécanisme de NMD qui n'arrive pas à dégrader tous les transcrits contenant E1' produits. En absence de puromycine, il est donc probable que nous quantifions les transcrits qui ne peuvent pas être dégradé par le NMD après la saturation de cette machinerie.

De plus, le taux observé est en accord avec le gradient d'oncogenèse décrit pour *VHL* puisque nous avons pu montrer, notamment via des tests minigènes rapporteurs, que la rétention de E1' est plus importante chez les patients qui développent des tumeurs que chez les patients qui développent une polyglobulie. Ce gradient n'est pas visible dans toutes les lignées cellulaires, notamment dans les LLB. Il y a donc une tissu spécificité de l'expression des mutants *VHL*. La rétention de cet exon conduit au niveau protéique à une diminution de l'expression totale de pVHL. En revanche, nous ne sommes pas en mesure de détecter l'hypothétique protéine codée par l'exon 1 de *VHL* en phase avec E1', que nous avons nommée X1, au niveau de son expression cellulaire endogène. Nous sommes cependant en mesure de détecter X1 avec l'anticorps dirigé contre l'exon 1 de *VHL* quand nous le surexprimons dans des lignées cellulaires (Figure 4A de l'article 1). Il semblerait donc que les transcrits contenant E1' ne soient pas traduits en protéines dans le modèle cellulaire utilisé, à savoir les LLB, ou alors en quantité trop faible pour être détectée par western blot. Nous ne pouvons pas conclure si X1 n'est pas détectée car elle est absente ou si elle est exprimée en dessous du seuil de détection du western blot réalisé.

Au moins deux hypothèses sont envisageables, soit la protéine X1 est traduite uniquement dans un type cellulaire précis dans lequel les transcrits ne sont pas totalement dégradés *via* le mécanisme de NMD, soit les transcrits sont en fait des longs ARN non codant (lncRNA). Ces derniers peuvent agir à différents niveaux dans les cellules, comme par exemple dans la régulation de l'expression des gènes<sup>232</sup>. Ils peuvent être poly-adenylés, ce qui est en adéquation avec le fait que nous les retrouvons dans les résultats de RNAséq (qui séquent les ARN poly-A par le système de capture poly-dT). Il a été montré que des lncRNA jouent un rôle important dans le développement du système nerveux central *via* différents mécanismes<sup>233</sup>. Cela est en adéquation avec les deux pathologies développées par les patients porteurs de mutations de l'exon cryptique E1'. En effet, si cet exon conduit à la formation d'un lncRNA ayant un rôle dans le développement du système nerveux central, une mutation de celui-ci pourrait conduire à une dérégulation du développement qui pourrait expliquer les symptômes des pathologies associées à *VHL*. Cela va dans le sens des observations faites avec les cellules productrices d'EPO. En effet, nous avons pu constater qu'il est plus difficile de différencier des cellules de patients porteurs de mutations E1' en cellules productrices d'EPO et plus particulièrement en progéniteurs de la crête neurale. Cependant, nous n'avons pas encore réalisé de différenciation de cellules de patients développant la maladie de VHL avec des mutations des autres exons de *VHL*, ni de différenciation de cellules de patients atteints de polyglobulie avec une mutation de E1'. Nous ne pouvons donc pas dire avec certitude que seules les mutations de l'exon E1' entraînent un défaut de différenciation des cellules en progéniteurs de la crête neurale *in vitro* ou si cela est observé pour les mutations de *VHL* associées la maladie de VHL.

L'hypothèse d'une fonction précise de X1 ou d'un potentiel lncRNA codé par E1' au niveau cellulaire pourrait également expliquer l'absence de rétention de E1' pour les variants génétiques à l'origine des tumeurs de la partie 2 des résultats. En effet, nous avons pu voir grâce à des tests de minigène rapporteurs, que la plupart des variants retrouvés pour la première étude retrospective de patients avec phéochromocytomes, paragangliomes ou cancers du rein, ne conduisent pas à une rétention de l'exon E1', contrairement à ce que nous avons précédemment observé. L'étude génétique doit être approfondie pour établir un lien de causalité avec la

pathologie (étude des apparentés, bilan clinique et ségrégation du variant avec la pathologie). Si les résultats sont en faveur d'un rôle de ces variants, nous pouvons envisager que ces mutations conduisent à une perte de fonction de X1 ou du lncRNA associé à E1', ce qui favoriserait le développement tumoral. Nous n'avons pas encore pu réaliser d'analyses protéiques de VHL pour ces mutations. En effet, il serait important de vérifier si ces mutations sont aussi à l'origine d'une diminution de l'expression protéique totale de VHL. Il est déjà programmé une analyse des transcrits de *VHL* dans les tumeurs de ces patients. Cependant, celles-ci sont fixées et conservées dans de la paraffine, ce qui va compliquer l'analyse par qPCR puisque cette technique de conservation détériore les transcrits qui font souvent moins de 200 paires de bases après extraction.

Afin d'appuyer l'hypothèse d'une fonction de la protéine X1, autre que sur la régulation de HIF pour laquelle nous avons réalisé des tests fonctionnels qui ont révélé l'incapacité de X1 à réguler HIF (Figure 4C article 1), il faudrait dans un premier temps réaliser une modélisation 3D de X1. Cela nous permettrait de visualiser les éventuels sites d'interactions avec d'autres protéines. En réalisant cette même modélisation pour chacune des mutations retrouvées dans E1', nous pourrions ainsi voir si les mutations ont un impact sur la structure protéique de X1 qui pourrait en modifier la fonction. En plus de cette modélisation, nous pourrions réaliser des tests de biotinylation artificiels des protéines interagissant avec BirA<sup>234</sup>. Cela nous permettrait de d'identifier les partenaires potentiels de X1 dans différents types cellulaires. Le problème de ce modèle est que la protéine X1 serait sur-exprimée de façon artificielle dans les cellules, ce qui peut entraîner la modification de certains mécanismes. De façon plus simple, nous pourrions également regarder l'impact des mutations de X1 sur les fonctions indépendantes de HIF. Par exemple, nous pourrions regarder la formation du cil primaire (pVHL se lie à la kinesine 2 pour stabiliser les microtubules du cil primaire), de la matrice extra-cellulaire, et celle des microtubules, dans des cellules exprimant VHL WT en comparaison avec X1 WT ou muté, comme cela a pu être fait pour des mutants de VHL<sup>235,77</sup>. En effet, il a pu être montré que des mutations comprises entre les acides aminés 95 et 123 peuvent conduire à une destabilisation des microtubules du cil primaire, ce qui est le cas des mutations VHL Y98H et Y112H. L'exon E1' s'insérant dans la protéine

VHL au niveau de l'acide aminé 113, il est possible que la protéine X1 ne permette pas la stabilisation des microtubules et que sa surexpression due aux mutations entraîne un défaut dans l'organisation cellulaire et dans la stabilisation du cil primaire.

L'hypothèse d'un lncRNA peut, elle, être complexe à étudier. En effet, nous avons évoqué l'hypothèse d'un lncRNA constitué des exons 1 et 1' de *VHL*, cependant, les bases de données « Ensembl » et « NCBI » décrivent une isoforme de VHL commençant 217 nucléotides en amont du début de E1' avec lequel s'épissent les exons 2 et 3 de *VHL*, que nous avons nommée upstream E1'. Il existerait donc un promoteur alternatif en amont de E1' pour réguler l'expression de cette isoforme. Afin de valider cette hypothèse, nous pourrions par exemple réaliser des analyses par ATAC-séq (Assay for Transposase-Accessible Chromatin with highthroughput sequencing) qui permettent de séquencer les parties ouvertes de la chromatine qui correspondent à des séquences régulatrices. Cette isoforme est décrite comme non codante pour une éventuelle protéine. Nous avons mis au point une RT-qPCR spécifique de cette isoforme avec une amorce située dans la séquence en amont de l'exon cryptique E1'. Nous avons pu montrer que cette isoforme est également présente dans tous les tissus chez l'homme, mais en quantité moindre par rapport à l'isoforme E1E1'. Nous pouvons, comme pour l'autre isoforme contenant E1', imaginer qu'elle est exprimée uniquement dans certains tissus spécifiques et/ou au cours du développement. Nous pourrions facilement imaginer que si cette autre isoforme a une fonction, les mutations retrouvées dans E1' pourraient avoir un impact sur celle-ci. D'autres explorations de cette isoforme sont nécessaires afin d'en comprendre le rôle et l'éventuelle implication des formes mutées dans les pathologies.

Dans les tumeurs, les mutations hétérozygotes de *VHL* sont associées à une perte d'hétérozygotie (LOH) pour le 2<sup>ème</sup> allèle dans la majorité des cas<sup>236</sup>. Cependant, une absence de LOH a déjà été décrite, notamment pour des cas d'hémangioblastomes<sup>237</sup>. Il a été montré que la perte d'hétérozygotie pouvait avoir lieu suite à une hyperméthylation du gène, qui n'est pas détectée lors du séquençage classique des gènes dans les tumeurs<sup>238</sup>. Les mutations de l'exon

cryptique E1' que nous avons analysées dans le cadre de la maladie de VHL ne sont pas associées à une LOH dans les tumeurs. Nous avons pu observer que l'allèle muté est exprimé à 75% dans les tumeurs de patients contre 25% pour l'allèle sauvage (Figure supplémentaire 8 de l'article 1). Nous avons également pu valider qu'aucune autre mutation de la voie de l'hypoxie n'est exprimée dans la tumeur. Cependant, nous n'avons pu classer ces tumeurs à la suite de l'analyse transcriptomique du fait d'une contamination trop importante par les tissus avoisinants. Nous ne savons donc pas si l'expression déséquilibrée des allèles sauvages (25%) et des allèles mutés (75%) observée est spécifique de la tumeur ou si nous observons un mélange avec du tissu sain. Il faudrait réaliser une analyse de la méthylation de VHL dans ces tumeurs afin de voir si l'inactivation du 2<sup>ème</sup> allèle se fait *via* ce mécanisme.

La première famille, qui développe la maladie de VHL (Famille 8 de l'article 1), que nous avons retrouvée mutée pour E1' est porteuse de 2 mutations sur un même allèle : c.340+617C>G et c.340+648T>C décrite comme un SNP dans les bases de données. Nous avons pu par la suite montrer (partie 3 des résultats) que les deux mutations sont capables, prises isolément, d'induire une rétention de l'exon E1' dans les transcrits et une diminution de l'expression protéique totale de *VHL* dans les LLB (résultats complémentaires de l'article 1). Nous avons pu montrer que lorsque les deux mutations sont associées, les effets sont légèrement plus importants que pour les mutations prises indépendamment. Nous avons également pu montrer qu'au niveau transcriptionnel et protéique, une mutation de l'exon E1' dans le cadre de la maladie de VHL conduit au même niveau de diminution de l'expression des isoformes E1E2E3 et E1E3 qu'une délétion totale d'un allèle de *VHL* décrite à l'origine de la pathologie<sup>144</sup>. Cet allèle doublement muté dans E1' peut donc être considéré comme une perte de fonction de *VHL* à l'origine de la pathologie.

Comme nous avons pu le décrire, les observations faites pour la rétention de E1' peuvent varier en fonction du type cellulaire dans lequel est réalisée l'expérience mingène. Il est donc important d'utiliser le type cellulaire adapté pour conclure sur l'impact des mutations de E1' ainsi que pour les autres mutations de *VHL*.

Les modèles cellulaires iPS actuellement utilisés sont encore en cours de mise au point. En effet, l'idéal serait l'obtention de cellules rénales productrices d'EPO. Nous avons fait l'hypothèse d'un modèle de cellules progénitrices de la crête neurale qui, en conditions hypoxiques, sont capables de produire de l'EPO. Cependant, nous pouvons penser que ce modèle n'est pas optimal puisque pour les patients porteurs d'une mutation *VHL*-R200W homozygote qui développent une polyglobulie, nous observons une augmentation de la production d'EPO uniquement en hypoxie par rapport aux contrôles alors que cette différence devrait être observée en normoxie, condition dans laquelle les patients présentent un excès de globules rouges. Grâce à un modèle de cellules hépatiques immatures, nous avons pu observer cette différence entre les patients et les contrôles en normoxie et en hypoxie, tant au niveau transcriptionnel que protéique. Il faudrait en effet pouvoir poursuivre la différenciation vers des cellules plus spécifiques qui correspondraient à des cellules rénales productrices d'EPO. Pour cela, il serait intéressant de pouvoir utiliser un système de co-culture de cellules productrices d'EPO et de cellules épithéliales rénales. Les différenciations en organoïdes rénaux sont déjà, elles, très avancées et très utilisées pour l'étude de pathologies rénales<sup>239</sup>. En ajoutant des cellules productrices d'EPO aux organoïdes rénaux, nous pourrions observer les effets des cellules les unes sur les autres au plus proche des interactions qui ont lieu dans les reins des patients. Il en est de même pour la maladie de VHL qui se développe tout le long du système nerveux central et certains organes tels que la rétine, les phéochromocytomes ou le rein. Comme nous avons pu le voir en introduction, beaucoup de tissus touchés ont pour origine la crête neurale. Seuls les cancers du rein ont pour origine des cellules épithéliales qui sont cependant entourées par les cellules productrices d'EPO rénales qui pourraient leur envoyer des signaux oncogéniques tels que des SASP pro-inflammatoires qui conduiraient à l'acquisition du caractère cancérigène des cellules épithéliales. Afin de prouver cette hypothèse, il faudrait réaliser des différenciations en cellules productrices d'EPO d'iPS de patients porteurs d'une mutation de *VHL* et plus spécifiquement, dans notre cas, de l'exon cryptique E1', et d'en analyser le sécrétome en présence ou non de cellules épithéliales rénales porteuses ou non de la même mutation que les iPS utilisées afin de se remettre dans le même contexte que chez les patients. Nous allons optimiser le protocole de différenciation afin d'avoir une expression semblable des marqueurs des progéniteurs de la crête neurale utilisés pour les

contrôles et pour les patients porteurs d'une mutation E1'. Afin d'analyser au mieux les différences acquises par les cellules de patients par rapport aux cellules contrôles, nous allons dans un premier temps réaliser des analyses transcriptomiques par DGEseq des cellules progénitrices de la crête neurale obtenues. Ces analyses vont nous permettre d'obtenir les différentiels d'expression entre les patients et les contrôles. Par la suite, nous pourrons alors uniquement réaliser des études transcriptomiques par RT-qPCR ciblées sur les gènes les plus impactés au niveau transcriptomique en réponse aux différentes mutations analysées.

La génération de lignées iPS CRISPR KO pour l'exon cryptique E1' va également permettre de comprendre au mieux le rôle de cet exon cryptique dans le développement embryonnaire et également dans le développement des tumeurs dans le cadre de la maladie de VHL. La génération de lignées KO ou KI pourrait également permettre de nous affranchir du prélèvement de cellules chez les patients et de leur reprogrammation via l'utilisation de cellules contrôles modifiées.

Nous avons pu montrer que des mutations de l'exon cryptique E1' sont à l'origine de la maladie de VHL ou de polyglobulie. Cet exon a donc été ajouté dans la liste des exons à séquencer dans le cadre du diagnostic de ces deux pathologies. Cependant, la fonction exacte en condition normale de E1' en tant que protéine ou lncRNA reste à déterminer. Les modèles iPS mis en place pourront, quand ils seront finalisés, nous permettre une analyse plus fine des différences observées entre les mutations de *VHL* à l'origine des deux pathologies associées, ainsi que pour affiner le gradient d'oncogenèse décrit pour la maladie de VHL<sup>57</sup>. Il est également envisageable que X1 n'est aucune fonction spécifique mais que la pathologie soit uniquement due à la perturbation de l'épissage (inclusion d'un fragment d'intron due aux mutations) qui entraîne une perte de fonction de la protéine VHL.

Lors de la découverte de l'exon cryptique E2', nous l'avons ajouté dans le panel des exons à séquencer dans le cadre du diagnostic génétique des érythrocytoses. Cela nous a permis d'y trouver des variants, à savoir deux variants synonymes (dont un référencé comme un SNP), un

variant faux sens et une délétion totale (Figure 2 de l'article 3). Des expériences de minigènes nous permettront d'étudier l'impact des mutations synonymes de E2' sur la rétention de celui-ci dans les transcrits. Nous avons pu montrer que la délétion totale de E2' ne conduit pas à une diminution de son expression ni à celle des isoformes sauvages de VHL.

L'exon E2' ajoute, en phase avec les exons 2 et 3 de *VHL*, un domaine qui est homologue à un domaine transmembranaire de l'hème oxygénase 2 et de la protéine transmembranaire TMEM116. L'hème oxygénase 2 est une isoforme constitutionnellement exprimée dont l'une des fonctions est la dégradation des érythrocytes (globules rouges). Nous pourrions donc imaginer que l'hypothétique protéine X2 joue un rôle dans cette dégradation et que les mutations de E2' entraînent une perte de fonction de X2 qui conduit à une mauvaise dégradation des globules rouges qui restent plus longtemps dans la circulation sanguine. Cela expliquerait la polyglobulie développée par les patients dont le taux d'EPO circulant est normal, voir abaissé. A l'inverse, nous pouvons imaginer que X2 agit comme un compétiteur avec l'hème oxygénase 2 pour la dégradation des globules rouges. En effet, si les mutations conduisent à une rétention de E2', il y a alors plus de X2, ce qui bloque la dégradation des globules rouges et explique là encore la polyglobulie. Ces hypothèses seront à tester pour les confirmer.

Nous avons pu montrer grâce à des tests de localisation cellulaire par immuno-histochimie que l'insertion de cet exon n'entraîne pas de relocalisation de la protéine X2 surexprimée. En effet, il est envisageable que l'absence d'un peptide signal envoyant la protéine à la membrane ne permette pas un changement de localisation de X2 par rapport à VHL. Là aussi, il est probable que son expression soit très tissus spécifique, et notamment au cours du développement embryonnaire. Deux hypothèses sont envisageables, soit ce transcrit code une protéine faiblement exprimée dans les cellules étudiées, soit ce transcrit est utilisé comme un lncRNA qui serait polyadénylé (puisque retrouvé dans le séquençage par RNAséq avec capture des queues poly-A).

Pour l'éventuelle protéine X2, nous avons débuté des tests de biotinylation artificielle des protéines interagissant avec X2 qui est surexprimé dans les cellules HEK293T grâce à la

transfection d'un plasmide. Nous pouvons pour le moment voir que le profil des partenaires de X2 est plus proche de celui de pVHL172 que de celui de pVHL213 et pVHL160. Ces observations iraient donc dans le sens d'une perte de fonction de X2 pour la régulation de HIF par rapport à VHL WT. Cette étude est à approfondir pour mieux connaître les interactions de cette éventuelle protéine afin d'identifier son rôle précis dans les cellules. Une modélisation 3D de la protéine pourrait également nous aider à identifier les domaines fonctionnels de celle-ci. Cela nous permettrait de mettre en évidence une éventuelle perte de fonction de la protéine en réponse aux mutations *via* une modification majeure d'un domaine fonctionnel.

La génération de cellules iPS des patients porteurs de mutations de X2 pourrait nous permettre de déterminer le rôle des mutations dans le développement de la polyglobulie. En effet, les cellules hépatiques ou les cellules progénitrices de la crête neurale productrices d'EPO que nous réussissons à obtenir peuvent permettre d'observer des différences entre les iPS contrôles et celles dérivées de patients, dans un premier temps pour la production d'EPO. Des différenciations en progéniteurs érythroïdes permettraient également d'étudier l'érythropoïèse et l'impact des mutations sur les voies autres que la production d'EPO. Du RNAseq de ces cellules différenciées sera ensuite réalisé afin de mettre en évidence les différences d'expression des différentes voies cellulaires en réponse aux mutations de E2'. Nous pourrons ainsi voir si les mécanismes moléculaires impliqués dans le développement de la maladie dont l'origine est une mutation de E2' sont identiques à ceux observés dans les cellules de patients porteurs de mutations dans les exons « classiques » de VHL (E1, E2 et E3). Cette étude pourra être approfondie grâce à la génération de cellules iPS CRISPR KO homozygote et hétérozygote pour l'exon E2'. Ces cellules nous permettront de voir si le transcrite contenant E2' a un rôle dans le développement, que ce soit en codant une protéine ou un lncRNA. Là encore, une étude transcriptomique par RNAseq nous permettra de mettre en lumière les différences entre les cellules contrôles et celles KO pour E2' à l'état hétérozygote ou homozygote.

Une étude de ségrégation reste nécessaire afin de déterminer la pathogénicité des mutations de E2'. Quand l'étude de cet exon sera plus approfondie, nous pourrons le publier, ce qui permettra

son inclusion dans la liste des exons à séquencer dans le cadre du diagnostic génétique des pathologies du spectre VHL (polyglobulie, maladie de VHL, phéochromocytomes, paragangliomes, cancers du rein sporadiques). Une étude fonctionnelle est également nécessaire avant de conclure sur le rôle des mutations dans les pathologies car nous avons ouvert une nouvelle piste quand à l'origine de la pathologie.

L'étude menée durant ma thèse a permis de réaliser le diagnostic génétique de certaines familles qui n'avaient pour le moment qu'un diagnostic basé sur les manifestations phénotypiques de leur pathologie. L'un des autres aspects d'une pathologie est l'aspect psychologique. Ici, l'annonce du diagnostic de la maladie de VHL ou d'une polyglobulie peut être un choc chez les patients. En effet, cela signifie un suivi médical très intense, avec un risque élevé de développer des tumeurs et des complications associées (telles que perdre la vue ou l'audition). En fonction de l'importance de la polyglobulie, cette maladie peut demander des saignées régulières et les patients risquent de mourir prématurément d'un accident vasculaire cérébral ou de thrombose.

Il y a en réalité deux diagnostics différents de ces pathologies. Le premier est le diagnostic médical qui est basé sur les symptômes des patients et possiblement leur histoire familiale dans les cas familiaux. Le deuxième est le diagnostic génétique qui est utilisé pour confirmer le diagnostic médical et qui permet également de réaliser un suivi de la descendance ou des apparentés du patient qui peuvent ne pas encore avoir déclaré de symptômes. Tant que le diagnostic génétique n'a pas été établi, les familles ne peuvent pas bénéficier d'un suivi médical poussé et régulier et ne peuvent pas bénéficier du diagnostic prénatal pour celles désireuses d'avoir un enfant.

L'étude menée durant ma thèse a permis de déterminer le diagnostic génétique pour 13 familles atteintes par la maladie de VHL ou des tumeurs qui y sont associées et qui sont être porteuses d'une mutation de l'exon cryptique E1'. Pour la 1<sup>ère</sup> famille retrouvée mutée pour cet exon (Famille 8 de l'article 1), il a fallu attendre la 3<sup>ème</sup> génération et la naissance de la petite fille de

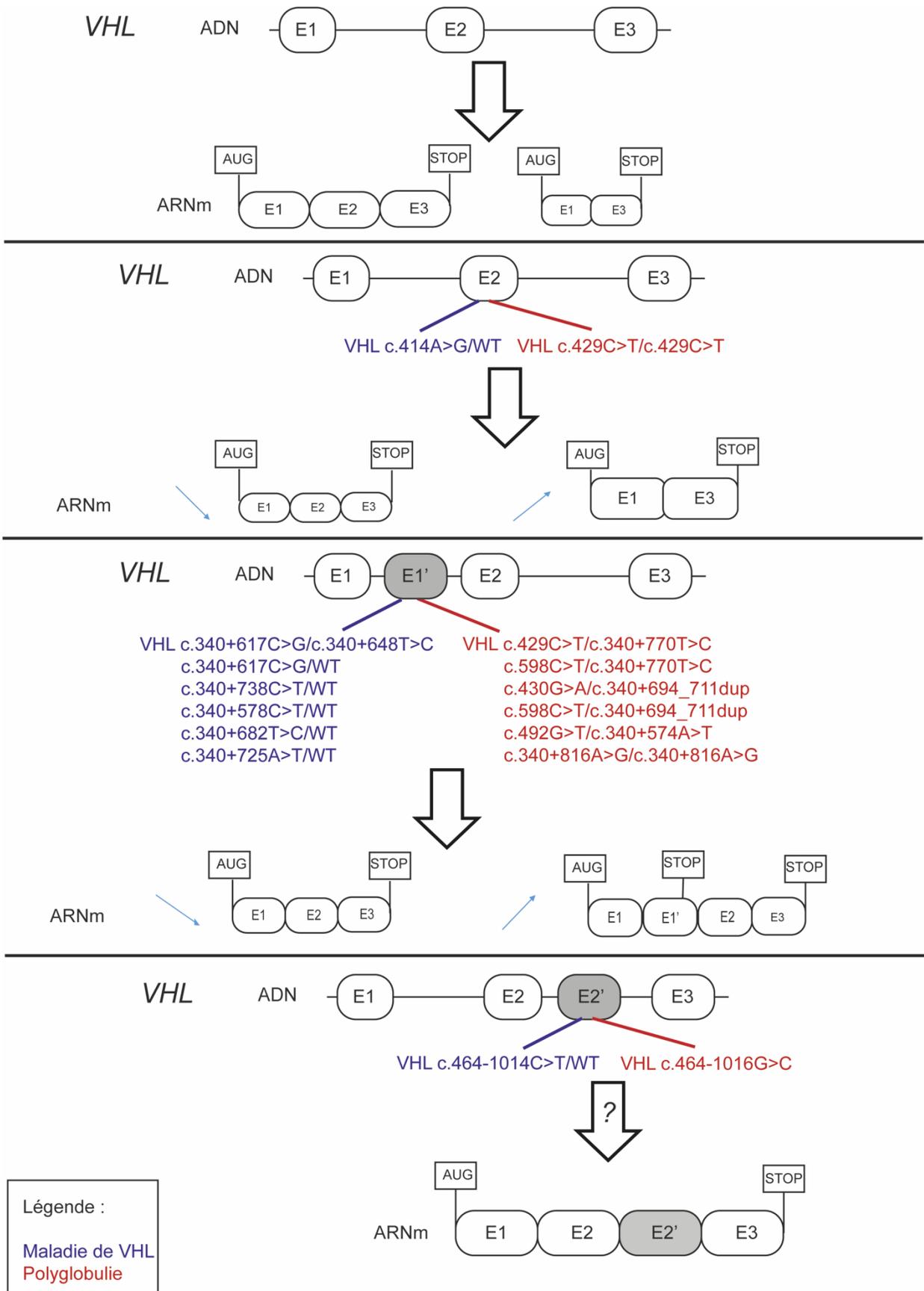
l'un des premiers proposants à la maladie pour pouvoir poser un diagnostic génétique. Nous pouvons donc imaginer le soulagement de cette famille de pouvoir enfin savoir si les futurs enfants à venir seront porteurs ou non de la pathologie mais également le stress retomber chez les personnes qui n'avaient pas encore développé de tumeur et qui viennent de savoir qu'elles n'en développeront pas dans le cadre de la maladie puisqu'elles ne sont pas porteuses de la mutation en cause dans leur famille. Dans cette famille, seule une personne qui n'a pas encore développé de tumeur n'a pas voulu être testée génétiquement. Pour toutes les autres, soit elles avaient déjà développé une tumeur, soit elles ne portent pas la mutation familiale, ce qui est le cas de la dernière petite fille née de la famille.

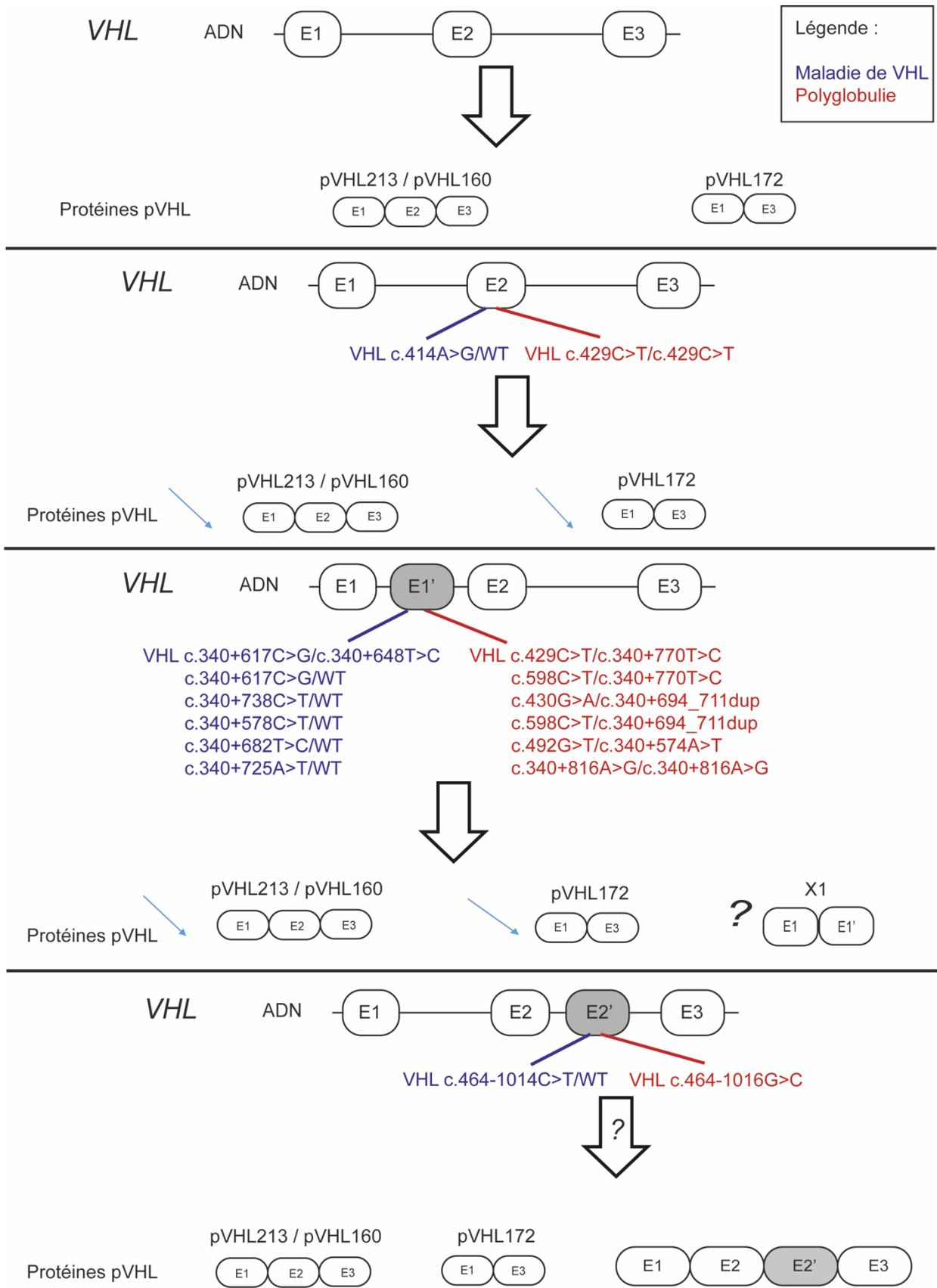
Quelques études se sont penchées sur l'aspect psychologique du diagnostic génétique de la maladie de VHL<sup>170,240</sup>. Le diagnostic génétique peut parfois durer plusieurs mois voire plusieurs années (comme dans le cas de la première famille pour laquelle nous avons trouvé une mutation de l'exon cryptique E1', 10 ans après la première consultation d'oncogénétique). Une étude de Rochette et al. a montré que le temps d'attente moyen du diagnostic génétique a été de 7,3 mois pour les patients de leur cohorte. Ils ont pu montrer que pour les 12 patients qui avaient développé un hémangioblastome mais qui n'avaient pas été dépistés pour la maladie de VHL, l'attente n'a pas affecté négativement l'aspect psycho-social de la maladie, basé sur plusieurs critères tels que l'anxiété, la dépression, la qualité de vie et l'inquiétude. La place du généticien est importante dans cette approche pour conseiller le patient et diminuer son anxiété. Elle est aussi importante pour le médecin traitant du patient qui ne connaît pas forcément la maladie ni les suivis à prescrire à son patient tout au long de sa vie. Les études portant sur les impacts psychologique d'une maladie héréditaire sont très difficiles à mener, c'est pourquoi elles sont rares. Cependant, il serait intéressant d'avoir accès à l'aspect psychologique chez une famille qui découvre l'origine de sa pathologie après 3 générations et de voir l'impact de ce diagnostic sur les générations futures (feront-ils des tests pré-natals de diagnostic préimplantatoire sur leurs futurs enfants, s'ils sont porteurs de la mutation, les feront-ils naître, etc).

En conclusion, ma thèse a permis une avancée majeure dans le diagnostic de la polyglobulie et de la maladie de VHL. Elle a permis à des familles de sortir d'une errance de diagnostic génétique, ce qui leur permet d'accéder à un meilleur suivi médical et thérapeutique familial. Nous avons pu montrer que la structure du gène *VHL* avec ses 3 exons et ses 2 exons cryptiques est plus complexe que ce qui avait été décrit jusqu'alors. Nous avons mis à jour de nouveaux mécanismes d'altération de *VHL* à savoir le saut de l'exon 2 ou la rétention de l'exon cryptique E1'. Nous avons pu mettre en évidence, pour des altérations de E1', qu'il existe une corrélation entre la sévérité des altérations (rétention de E1') et la sévérité des phénotypes associés. Ces données constituent un argument supplémentaire du modèle d'oncogenèse en gradient pour *VHL* décrit par Couvé et al.<sup>57</sup>. Cependant, les mécanismes moléculaires exacts à l'origine des pathologies restent encore à déterminer. Pour cela, il faudrait élucider la fonction des protéines ou lncRNA codés par les exons E1' et E2' de *VHL*. Une fois ces fonctions connues, il sera peut-être possible de donner aux patients des traitements plus ciblés pour leur type de pathologie et d'améliorer au mieux leur confort de vie. Pour le moment, ces deux exons cryptiques ont été ajoutés dans la liste des exons à séquencer pour le diagnostic génétique des polyglobulies et de la maladie de VHL ou des tumeurs sporadiques qui y sont associées avec une dérégulation de la voie de l'hypoxie (phéochromocytomes, hémangioblastomes, cancers du rein).

#### **Schémas récapitulatifs des résultats obtenus :**

Les schémas suivant récapitulent les conséquences des mutations retrouvées à l'origine de la maladie de VHL ou d'une polyglobulie sur l'expression des ARNm (premier schéma) et sur l'expression des protéines VHL (deuxième schéma).





## V. BIBLIOGRAPHIE

1. Zhang, H. & Barralet, J. E. Mimicking oxygen delivery and waste removal functions of blood. *Adv. Drug Deliv. Rev.* **122**, 84–104 (2017).
2. Guillemin, K. & Krasnow, M. A. The Hypoxic Response: Huffing and HIFing. *Cell* **89**, 9–12 (1997).
3. Honigman, B. Acute Mountain Sickness in a General Tourist Population at Moderate Altitudes. *Ann. Intern. Med.* **118**, 587 (1993).
4. Hackett, P. H. & Roach, R. C. High-Altitude Illness. *N. Engl. J. Med.* **345**, 107–114 (2001).
5. Beall, C. M. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8655–8660 (2007).
6. Gilbert-Kawai, E. T., Milledge, J. S., Grocott, M. P. W. & Martin, D. S. King of the mountains: Tibetan and Sherpa physiological adaptations for life at high altitude. *Physiol. Bethesda Md* **29**, 388–402 (2014).
7. Krantz, S. B. Erythropoietin. *Blood* **77**, 419–434 (1991).
8. Bachmann, S., Le Hir, M. & Eckardt, K. U. Co-localization of erythropoietin mRNA and ecto-5'-nucleotidase immunoreactivity in peritubular cells of rat renal cortex indicates that fibroblasts produce erythropoietin. *J. Histochem. Cytochem. Off. J. Histochem. Soc.* **41**, 335–341 (1993).
9. Wojchowski, D. M., Sathyanarayana, P. & Dev, A. Erythropoietin receptor response circuits. *Curr. Opin. Hematol.* **17**, 169–176 (2010).
10. Gothié, E. & Pouysségur, J. HIF-1 : régulateur central de l'hypoxie. *médecine/sciences* **18**, 70–78 (2002).
11. Lee, F. S. & Percy, M. J. The HIF Pathway and Erythrocytosis. *Annu. Rev. Pathol. Mech. Dis.* **6**, 165–192 (2011).
12. Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O<sub>2</sub> tension. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5510–5514 (1995).
13. Ratcliffe, P. J. HIF-1 and HIF-2: working alone or together in hypoxia? *J. Clin. Invest.* **117**, 862–865 (2007).
14. Mole, D. R. *et al.* Genome-wide Association of Hypoxia-inducible Factor (HIF)-1 $\alpha$  and HIF-2 $\alpha$  DNA Binding with Expression Profiling of Hypoxia-inducible Transcripts. *J. Biol. Chem.* **284**, 16767–16775 (2009).
15. Hu, C.-J., Wang, L.-Y., Chodosh, L. A., Keith, B. & Simon, M. C. Differential Roles of Hypoxia-Inducible Factor 1 $\alpha$  (HIF-1 $\alpha$ ) and HIF-2 $\alpha$  in Hypoxic Gene Regulation. *Mol. Cell. Biol.* **23**, 9361–9374 (2003).
16. Wang, V., Davis, D. A., Haque, M., Huang, L. E. & Yarchoan, R. Differential Gene Up-Regulation by Hypoxia-Inducible Factor-1 $\alpha$  and Hypoxia-Inducible Factor-2 $\alpha$  in HEK293T Cells. *Cancer Res.* **65**, 3299–3306 (2005).
17. Carroll, V. A. & Ashcroft, M. Role of hypoxia-inducible factor (HIF)-1 $\alpha$  versus HIF-2 $\alpha$  in the regulation of HIF target genes in response to hypoxia, insulin-like growth factor-I, or loss of von Hippel-Lindau function: implications for targeting the HIF pathway. *Cancer Res.* **66**, 6264–6270 (2006).

18. Chan, D. A., Sutphin, P. D., Yen, S.-E. & Giaccia, A. J. Coordinate regulation of the oxygen-dependent degradation domains of hypoxia-inducible factor 1 alpha. *Mol. Cell. Biol.* **25**, 6415–6426 (2005).
19. Tian, Y.-M. *et al.* Differential sensitivity of hypoxia inducible factor hydroxylation sites to hypoxia and hydroxylase inhibitors. *J. Biol. Chem.* **286**, 13041–13051 (2011).
20. Huang, L. E., Gu, J., Schau, M. & Bunn, H. F. Regulation of hypoxia-inducible factor 1 $\alpha$  is mediated by an O<sub>2</sub>-dependent degradation domain via the ubiquitin-proteasome pathway. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 7987–7992 (1998).
21. Appelhoff, R. J. *et al.* Differential Function of the Prolyl Hydroxylases PHD1, PHD2, and PHD3 in the Regulation of Hypoxia-inducible Factor. *J. Biol. Chem.* **279**, 38458–38465 (2004).
22. Kaelin, W. G. & Ratcliffe, P. J. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol. Cell* **30**, 393–402 (2008).
23. Dayan, F., Roux, D., Brahimi-Horn, M. C., Pouyssegur, J. & Mazure, N. M. The Oxygen Sensor Factor-Inhibiting Hypoxia-Inducible Factor-1 Controls Expression of Distinct Genes through the Bifunctional Transcriptional Character of Hypoxia-Inducible Factor-1 $\alpha$ . *Cancer Res.* **66**, 3688–3698 (2006).
24. Jaakkola, P. *et al.* Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O<sub>2</sub>-regulated prolyl hydroxylation. *Science* **292**, 468–472 (2001).
25. Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).
26. Wenger, R. H., Stiehl, D. P. & Camenisch, G. Integration of Oxygen Signaling at the Consensus HRE. *Sci STKE* **2005**, re12–re12 (2005).
27. Storti, F. *et al.* A novel distal upstream hypoxia response element regulating oxygen-dependent erythropoietin gene expression. *Haematologica* **99**, e45–e48 (2014).
28. Schödel, J. *et al.* High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. *Blood* **117**, e207–e217 (2011).
29. Makino, Y., Kanopka, A., Wilson, W. J., Tanaka, H. & Poellinger, L. Inhibitory PAS Domain Protein (IPAS) Is a Hypoxia-inducible Splicing Variant of the Hypoxia-inducible Factor-3 $\alpha$  Locus. *J. Biol. Chem.* **277**, 32405–32408 (2002).
30. Maynard, M. A. *et al.* Dominant-Negative HIF-3 $\alpha$ 4 Suppresses VHL-Null Renal Cell Carcinoma Progression. *Cell Cycle* **6**, 2810–2816 (2007).
31. Heikkilä, M., Pasanen, A., Kivirikko, K. I. & Myllyharju, J. Roles of the human hypoxia-inducible factor (HIF)-3 $\alpha$  variants in the hypoxia response. *Cell. Mol. Life Sci.* **68**, 3885–3901 (2011).
32. Semenza, G. L., Dureza, R. C., Traystman, M. D., Gearhart, J. D. & Antonarakis, S. E. Human erythropoietin gene expression in transgenic mice: multiple transcription initiation sites and cis-acting regulatory elements. *Mol. Cell. Biol.* **10**, 930–938 (1990).
33. Suzuki, N., Hirano, I., Pan, X., Minegishi, N. & Yamamoto, M. Erythropoietin production in neuroepithelial and neural crest cells during primitive erythropoiesis. *Nat. Commun.* **4**, 2902 (2013).
34. Tan, C. C., Eckardt, K. & Ratcliffe, P. J. Organ distribution of erythropoietin messenger RNA in normal and uremic rats. *Kidney Int.* **40**, 69–76 (1991).
35. Kumral, A. *et al.* Erythropoietin in neonatal brain protection: the past, the present and the future. *Brain Dev.* **33**, 632–643 (2011).

36. Lönnberg, M., Garle, M., Lönnberg, L. & Birgegård, G. Patients with anaemia can shift from kidney to liver production of erythropoietin as shown by glycoform analysis. *J. Pharm. Biomed. Anal.* **81–82**, 187–192 (2013).
37. Gerl, K. *et al.* Erythropoietin production by PDGFR- $\beta$ + cells. *Pflüg. Arch. - Eur. J. Physiol.* **468**, 1479–1487 (2016).
38. Imeri, F. *et al.* Generation of renal Epo-producing cell lines by conditional gene tagging reveals rapid HIF-2 driven Epo kinetics, cell autonomous feedback regulation, and a telocyte phenotype. *Kidney Int.* **95**, 375–387 (2019).
39. Pan, X. *et al.* Isolation and Characterization of Renal Erythropoietin-Producing Cells from Genetically Produced Anemia Mice. *PLoS ONE* **6**, e25839 (2011).
40. Nolan, K. A. & Wenger, R. H. Source and microenvironmental regulation of erythropoietin in the kidney. *Curr. Opin. Nephrol. Hypertens.* **27**, 277–282 (2018).
41. Ebert, T., Bander, N. H., Finstad, C. L., Ramsawak, R. D. & Old, L. J. Establishment and characterization of human renal cancer and normal kidney cell lines. *Cancer Res.* **50**, 5531–5536 (1990).
42. Hammond, D. & Winnick, S. Paraneoplastic Erythrocytosis and Ectopic Erythropoietins\*. *Ann. N. Y. Acad. Sci.* **230**, 219–227 (1974).
43. Lin, F. K. *et al.* Cloning and expression of the human erythropoietin gene. *Proc. Natl. Acad. Sci.* **82**, 7580–7584 (1985).
44. Jacobs, K. *et al.* Isolation and characterization of genomic and cDNA clones of human erythropoietin. *Nature* **313**, 806–810 (1985).
45. Ebert, B. L. & Bunn, H. F. Regulation of the Erythropoietin Gene. *Blood* **94**, 1864–1877 (1999).
46. Wenger, R. H. & Kurtz, A. Erythropoietin. in *Comprehensive Physiology* 1759–1794 (American Cancer Society, 2011). doi:10.1002/cphy.c100075
47. Russell, R. C. *et al.* LOSS OF JAK2 REGULATION VIA VHL-SOCS1 E3 UBIQUITIN HETEROCOMPLEX UNDERLIES CHUVASH POLYCYTHEMIA. *Nat. Med.* **17**, 845–853 (2011).
48. Nandakumar, S. K., Ulirsch, J. C. & Sankaran, V. G. Advances in Understanding Erythropoiesis: Evolving Perspectives. *Br. J. Haematol.* **173**, 206–218 (2016).
49. EPO et dopage. *Planet-Vie* Available at: <https://planet-vie.ens.fr/article/2090/epo-dopage>. (Accessed: 10th July 2019)
50. Franco, R. S. The measurement and importance of red cell survival. *Am. J. Hematol.* **84**, 109–114 (2009).
51. Clark, P. E. & Cookson, M. S. The von Hippel-Lindau Gene. *Cancer* **113**, 1768–1778 (2008).
52. Iliopoulos, O., Kibel, A., Gray, S. & Kaelin, W. G. Tumour suppression by the human von Hippel-Lindau gene product. *Nat. Med.* **1**, 822–826 (1995).
53. Latif, F. *et al.* Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* **260**, 1317–1320 (1993).
54. Schoenfeld, A., Davidowitz, E. J. & Burk, R. D. A second major native von Hippel-Lindau gene product, initiated from an internal translation start site, functions as a tumor suppressor. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8817–8822 (1998).

55. Iliopoulos, O., Ohh, M. & Kaelin, W. G. pVHL19 is a biologically active product of the von Hippel-Lindau gene arising from internal translation initiation. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11661–11666 (1998).
56. Blankenship, C., Naglich, J. G., Whaley, J. M., Seizinger, B. & Kley, N. Alternate choice of initiation codon produces a biologically active product of the von Hippel Lindau gene with tumor suppressor activity. *Oncogene* **18**, 1529–1535 (1999).
57. Couvé, S. *et al.* Genetic evidence of a precisely tuned dysregulation in the hypoxia signaling pathway during oncogenesis. *Cancer Res.* **74**, 6554–6564 (2014).
58. Richards, F. M., Schofield, P. N., Fleming, S. & Maher, E. R. Expression of the Von Hippel—Lindau Disease Tumour Suppressor Gene during Human Embryogenesis. *Hum. Mol. Genet.* **5**, 639–644 (1996).
59. Chesnel, F. *et al.* The von Hippel-Lindau tumour suppressor gene: uncovering the expression of the pVHL172 isoform. *Br. J. Cancer* **113**, 336–344 (2015).
60. Hascoet, P. *et al.* The pVHL172 isoform is not a tumor suppressor and up-regulates a subset of pro-tumorigenic genes including TGFB1 and MMP13. *Oncotarget* **8**, 75989–76002 (2017).
61. Stebbins, C. E., Kaelin, W. G. & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).
62. Kibel, A., Iliopoulos, O., DeCaprio, J. A. & Kaelin, W. G. Binding of the von Hippel-Lindau tumor suppressor protein to Elongin B and C. *Science* **269**, 1444–1446 (1995).
63. Kaelin, W. G. Molecular basis of the VHL hereditary cancer syndrome. *Nat. Rev. Cancer* **2**, 673 (2002).
64. Duan, D. R. *et al.* Characterization of the VHL tumor suppressor gene product: localization, complex formation, and the effect of natural inactivating mutations. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6459–6463 (1995).
65. Schoenfeld, A. R., Davidowitz, E. J. & Burk, R. D. Elongin BC complex prevents degradation of von Hippel-Lindau tumor suppressor gene products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8507–8512 (2000).
66. Ohh, M. *et al.* Ubiquitination of hypoxia-inducible factor requires direct binding to the beta-domain of the von Hippel-Lindau protein. *Nat. Cell Biol.* **2**, 423–427 (2000).
67. Cockman, M. E. *et al.* Hypoxia inducible factor- $\alpha$  binding and ubiquitylation by the von Hippel-Lindau tumor suppressor protein. *J. Biol. Chem.* **275**, 25733–25741 (2000).
68. Qi, H. *et al.* Molecular cloning and characterization of the von Hippel-Lindau-like protein. *Mol. Cancer Res. MCR* **2**, 43–52 (2004).
69. Clifford, S. C. *et al.* Contrasting effects on HIF-1 $\alpha$  regulation by disease-causing pVHL mutations correlate with patterns of tumourigenesis in von Hippel-Lindau disease. *Hum. Mol. Genet.* **10**, 1029–1038 (2001).
70. Kaelin Jr, W. G. The von Hippel—Lindau tumour suppressor protein: O<sub>2</sub> sensing and cancer. *Nat. Rev. Cancer* **8**, 865–873 (2008).
71. Tabaro, F. *et al.* VHLdb: A database of von Hippel-Lindau protein interactors and mutations. *Sci. Rep.* **6**, 31128 (2016).
72. Duan, D. R. *et al.* Inhibition of transcription elongation by the VHL tumor suppressor protein. *Science* **269**, 1402–1406 (1995).

73. Na, X. *et al.* Identification of the RNA polymerase II subunit hsRPB7 as a novel target of the von Hippel–Lindau protein. *EMBO J.* **22**, 4249–4259 (2003).
74. Kuznetsova, A. V. *et al.* von Hippel–Lindau protein binds hyperphosphorylated large subunit of RNA polymerase II through a proline hydroxylation motif and targets it for ubiquitination. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2706–2711 (2003).
75. Lutz, M. S. & Burk, R. D. Primary cilium formation requires von hippel-lindau gene function in renal-derived cells. *Cancer Res.* **66**, 6903–6907 (2006).
76. Lolkema, M. P. *et al.* The von Hippel–Lindau tumour suppressor interacts with microtubules through kinesin-2. *FEBS Lett.* **581**, 4571–4576 (2007).
77. Hergovich, A., Lisztwan, J., Barry, R., Ballschmieter, P. & Krek, W. Regulation of microtubule stability by the von Hippel–Lindau tumour suppressor protein pVHL. *Nat. Cell Biol.* **5**, 64 (2003).
78. Singla, V. & Reiter, J. F. The Primary Cilium as the Cell’s Antenna: Signaling at a Sensory Organelle. *Science* **313**, 629–633 (2006).
79. Davenport, J. R. & Yoder, B. K. An incredible decade for the primary cilium: a look at a once-forgotten organelle. *Am. J. Physiol.-Ren. Physiol.* **289**, F1159–F1169 (2005).
80. Tang, N., Mack, F., Haase, V. H., Simon, M. C. & Johnson, R. S. pVHL function is essential for endothelial extracellular matrix deposition. *Mol. Cell. Biol.* **26**, 2519–2530 (2006).
81. Kurban, G. *et al.* Collagen matrix assembly is driven by the interaction of von Hippel–Lindau tumor suppressor protein with hydroxylated collagen IV alpha 2. *Oncogene* **27**, 1004–1012 (2008).
82. Yang, H. *et al.* pVHL acts as an adaptor to promote the inhibitory phosphorylation of the NF-kappaB agonist Card9 by CK2. *Mol. Cell* **28**, 15–27 (2007).
83. Roe, J.-S. *et al.* p53 Stabilization and Transactivation by a von Hippel–Lindau Protein. *Mol. Cell* **22**, 395–405 (2006).
84. Roe, J.-S. *et al.* Phosphorylation of von Hippel–Lindau protein by checkpoint kinase 2 regulates p53 transactivation. *Cell Cycle* **10**, 3920–3928 (2011).
85. Lee, S. *et al.* Neuronal apoptosis linked to EglN3 prolyl hydroxylase and familial pheochromocytoma genes: developmental culling and cancer. *Cancer Cell* **8**, 155–167 (2005).
86. Rawlings, J. S., Rosler, K. M. & Harrison, D. A. The JAK/STAT signaling pathway. *J. Cell Sci.* **117**, 1281–1283 (2004).
87. Bellucci, S. & Michiels, J. J. The role of JAK2 V617F mutation, spontaneous erythropoiesis and megakaryocytopoiesis, hypersensitive platelets, activated leukocytes, and endothelial cells in the etiology of thrombotic manifestations in polycythemia vera and essential thrombocythemia. *Semin. Thromb. Hemost.* **32**, 381–398 (2006).
88. Lonser, R. R. *et al.* von Hippel–Lindau disease. *The Lancet* **361**, 2059–2067 (2003).
89. Hippel, E. Über eine sehr seltene Erkrankung der Netzhaut. *Albrecht Von Graefes Arch. Für Ophthalmol.* **59**, 83–106
90. Lindau, A. Zur Frage der Angiomatosis retinae und ihrer Hirnkomplikationen. *Acta Ophthalmol. (Copenh.)* **4**, 193–226 (1926).
91. Melmon, K. L. & Rosen, S. W. LINDAU’S DISEASE. REVIEW OF THE LITERATURE AND STUDY OF A LARGE KINDRED. *Am. J. Med.* **36**, 595–617 (1964).
92. Santarpia, L. *et al.* Mosaicism in von Hippel–Lindau disease: an event important to recognize. *J. Cell. Mol. Med.* **11**, 1408–1415 (2007).

93. Sgambati, M. T. *et al.* Mosaicism in von Hippel–Lindau Disease: Lessons from Kindreds with Germline Mutations Identified in Offspring with Mosaic Parents. *Am. J. Hum. Genet.* **66**, 84–91 (2000).
94. Neumann, H. P. & Wiestler, O. D. Clustering of features of von Hippel-Lindau syndrome: evidence for a complex genetic locus. *Lancet Lond. Engl.* **337**, 1052–1054 (1991).
95. Chen, F. *et al.* Germline mutations in the von Hippel-Lindau disease tumor suppressor gene: correlations with phenotype. *Hum. Mutat.* **5**, 66–75 (1995).
96. Hoffman, M. A. *et al.* von Hippel-Lindau protein mutants linked to type 2C VHL disease preserve the ability to downregulate HIF. *Hum. Mol. Genet.* **10**, 1019–1027 (2001).
97. Knudson, A. G. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 820–823 (1971).
98. Knudson, A. G. & Strong, L. C. Mutation and cancer: neuroblastoma and pheochromocytoma. *Am. J. Hum. Genet.* **24**, 514–532 (1972).
99. Tory, K. *et al.* Specific genetic change in tumors associated with von Hippel-Lindau disease. *J. Natl. Cancer Inst.* **81**, 1097–1101 (1989).
100. Crossey, P. A. *et al.* Molecular genetic investigations of the mechanism of tumourigenesis in von Hippel-Lindau disease: analysis of allele loss in VHL tumours. *Hum. Genet.* **93**, 53–58 (1994).
101. Linehan, W. M., Srinivasan, R. & Schmidt, L. S. The genetic basis of kidney cancer: a metabolic disease. *Nat. Rev. Urol.* **7**, 277–285 (2010).
102. Richard, S., Gardie, B., Couvé, S. & Gad, S. Von Hippel-Lindau: how a rare disease illuminates cancer biology. *Semin. Cancer Biol.* **23**, 26–37 (2013).
103. Wanebo, J. E., Lonser, R. R., Glenn, G. M. & Oldfield, E. H. The natural history of hemangioblastomas of the central nervous system in patients with von Hippel–Lindau disease. *J. Neurosurg.* **98**, 82–94 (2003).
104. Neumann, H. P. *et al.* Central nervous system lesions in von Hippel-Lindau syndrome. *J. Neurol. Neurosurg. Psychiatry* **55**, 898–901 (1992).
105. Dollfus, H. *et al.* Retinal Hemangioblastoma in von Hippel-Lindau Disease: A Clinical and Molecular Study. *Invest. Ophthalmol. Vis. Sci.* **43**, 3067–3074 (2002).
106. Webster, A. R., Maher, E. R. & Moore, A. T. Clinical Characteristics of Ocular Angiomas in von Hippel-Lindau Disease and Correlation With Germline Mutation. *Arch. Ophthalmol.* **117**, 371–378 (1999).
107. Chan, C.-C. *et al.* VHL Gene Deletion and Enhanced VEGF Gene Expression Detected in the Stromal Cells of Retinal Angioma. *Arch. Ophthalmol.* **117**, 625–630 (1999).
108. Neumann, H. P. H. *et al.* PREVALENCE, MORPHOLOGY AND BIOLOGY OF RENAL CELL CARCINOMA IN VON HIPPEL-LINDAU DISEASE COMPARED TO SPORADIC RENAL CELL CARCINOMA. *J. Urol.* **160**, 1248–1254 (1998).
109. Chauveau, D. *et al.* Renal involvement in von Hippel-Lindau disease. *Kidney Int.* **50**, 944–951 (1996).
110. Maher, E. R. *et al.* Clinical features and natural history of von Hippel-Lindau disease. *Q. J. Med.* **77**, 1151–1163 (1990).
111. Roupert Morgan *et al.* Nephron Sparing Surgery for Renal Cell Carcinoma and von Hippel-Lindau’s Disease: A Single Center Experience. *J. Urol.* **170**, 1752–1755 (2003).

112. WALTHER McCLELLAN M. *et al.* Clinical and genetic characterization of pheochromocytoma in von hippel-lindau families: comparison with sporadic pheochromocytoma gives insight into natural history of pheochromocytoma. *J. Urol.* **162**, 659–664 (1999).
113. Richard, S. *et al.* Pheochromocytoma as the first manifestation of von Hippel-Lindau disease. *Surgery* **116**, 1076–1081 (1994).
114. Baghai, M. *et al.* Pheochromocytomas and Paragangliomas in von Hippel–Lindau Disease: A Role for Laparoscopic and Cortical-Sparing Surgery. *Arch. Surg.* **137**, 682–689 (2002).
115. Libutti, S. K. *et al.* Pancreatic neuroendocrine tumors associated with von Hippel Lindau disease: Diagnostic and management recommendations. *Surgery* **124**, 1153–1159 (1998).
116. Hough, D. M., Stephens, D. H., Johnson, C. D. & Binkovitz, L. A. Pancreatic lesions in von Hippel-Lindau disease: prevalence, clinical significance, and CT findings. *Am. J. Roentgenol.* **162**, 1091–1094 (1994).
117. Manski, T. J. *et al.* Endolymphatic Sac Tumors: A Source of Morbid Hearing Loss in von Hippel-Lindau Disease. *JAMA* **277**, 1461–1466 (1997).
118. Wind, J. J. & Lonser, R. R. Management of von Hippel–Lindau disease-associated CNS lesions. *Expert Rev. Neurother.* **11**, 1433–1441 (2011).
119. Wernert, N., Goebbels, R. & Prediger, L. Papillary Cystadenoma of the Epididymis Case Report and Review of the Literature. *Pathol. - Res. Pract.* **181**, 260–262 (1986).
120. Funk, K. & Heiken, J. Papillary cystadenoma of the broad ligament in a patient with von Hippel-Lindau disease. *Am. J. Roentgenol.* **153**, 527–528 (1989).
121. Messinezy, M. & Pearson, T. C. REVIEWThe classification and diagnostic criteria of the erythrocytoses (polycythaemias). *Clin. Lab. Haematol.* **21**, 309–316 (1999).
122. James, C. *et al.* A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* **434**, 1144–1148 (2005).
123. Pasquier, F. *et al.* New pathogenic mechanisms induced by germline erythropoietin receptor mutations in primary erythrocytosis. *Haematologica* **103**, 575–586 (2018).
124. de la Chapelle, A., Träskelin, A. L. & Juvonen, E. Truncated erythropoietin receptor causes dominantly inherited benign human erythrocytosis. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 4495–4499 (1993).
125. Zmajkovic, J. *et al.* A Gain-of-Function Mutation in EPO in Familial Erythrocytosis. *N. Engl. J. Med.* (2018). doi:10.1056/NEJMoa1709064
126. Ang, S. O. *et al.* Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia. *Nat. Genet.* **32**, 614–621 (2002).
127. Liu E. The worldwide distribution of the VHL 598C>T mutation indicates a single founding event.
128. Perrotta, S. *et al.* Von Hippel-Lindau-dependent polycythemia is endemic on the island of Ischia: identification of a novel cluster. *Blood* **107**, 514–519 (2006).
129. Bento, C. *et al.* Genetic Basis of Congenital Erythrocytosis: Mutation Update and Online Databases. *Hum. Mutat.* **35**, 15–26 (2014).
130. Bento, C. Genetic basis of congenital erythrocytosis. *Int. J. Lab. Hematol.* **40**, 62–67 (2018).

131. Gordeuk, V. R. *et al.* Congenital disorder of oxygen sensing: association of the homozygous Chuvash polycythemia VHL mutation with thrombosis and vascular abnormalities but not tumors. *Blood* **103**, 3924–3932 (2004).
132. Zbar, B. *et al.* Germline mutations in the Von Hippel-Lindau disease (VHL) gene in families from North America, Europe, and Japan. *Hum. Mutat.* **8**, 348–357 (1996).
133. Nordstrom-O’Brien, M. *et al.* Genetic analysis of von Hippel-Lindau disease. *Hum. Mutat.* **31**, 521–537 (2010).
134. Berger, A. H., Knudson, A. G. & Pandolfi, P. P. A Continuum Model for Tumour Suppression. *Nature* **476**, 163–169 (2011).
135. Coppé, J.-P., Desprez, P.-Y., Krtolica, A. & Campisi, J. The Senescence-Associated Secretory Phenotype: The Dark Side of Tumor Suppression. *Annu. Rev. Pathol.* **5**, 99–118 (2010).
136. Li, L. *et al.* Hypoxia-Inducible Factor Linked to Differential Kidney Cancer Risk Seen with Type 2A and Type 2B VHL Mutations. *Mol. Cell. Biol.* **27**, 5381–5392 (2007).
137. Lanikova, L. *et al.* Novel homozygous VHL mutation in exon 2 is associated with congenital polycythemia but not with cancer. *Blood* **121**, 3918–3924 (2013).
138. Pastore, Y. *et al.* Mutations of von Hippel-Lindau Tumor-Suppressor Gene and Congenital Polycythemia. *Am. J. Hum. Genet.* **73**, 412–419 (2003).
139. Astapova, O., Biswas, A., DiMauro, A., Moalem, J. & Hammes, S. R. Bilateral Pheochromocytomas in a Patient with Y175C Von Hippel-Lindau Mutation. *Case Rep. Endocrinol.* **2018**, (2018).
140. Bento, C. *et al.* Molecular study of congenital erythrocytosis in 70 unrelated patients revealed a potential causal mutation in less than half of the cases (Where is/are the missing gene(s)?). *Eur. J. Haematol.* **91**, 361–368 (2013).
141. Chou, M. T.-H., Anthony, J., Bjorge, J. D. & Fujita, D. J. The von Hippel-Lindau Tumor Suppressor Protein Is Destabilized by Src. *Genes Cancer* **1**, 225–238 (2010).
142. Richards, F. M. *et al.* Detailed mapping of germline deletions of the von Hippel-Lindau disease tumour suppressor gene. *Hum. Mol. Genet.* **3**, 595–598 (1994).
143. Maranchie, J. K. *et al.* Solid renal tumor severity in von Hippel Lindau disease is related to germline deletion length and location. *Hum. Mutat.* **23**, 40–46 (2004).
144. Cascón, A. *et al.* Loss of the actin regulator HSPC300 results in clear cell renal cell carcinoma protection in Von Hippel-Lindau patients. *Hum. Mutat.* **28**, 613–621 (2007).
145. McNeill, A. *et al.* Genotype-phenotype correlations in VHL exon deletions. *Am. J. Med. Genet. A.* **149A**, 2147–2151 (2009).
146. Escobar, B. *et al.* Brick1 is an essential regulator of actin cytoskeleton required for embryonic development and cell transformation. *Cancer Res.* **70**, 9349–9359 (2010).
147. Frede, S., Freitag, P., Geuting, L., Konietzny, R. & Fandrey, J. Oxygen-regulated expression of the erythropoietin gene in the human renal cell line REPC. *Blood* **117**, 4905–4914 (2011).
148. Hirano, I. & Suzuki, N. The Neural Crest as the First Production Site of the Erythroid Growth Factor Erythropoietin. *Front. Cell Dev. Biol.* **7**, 105 (2019).
149. Haase, V. H. The VHL tumor suppressor in development and disease: functional studies in mice by conditional gene targeting. *Semin. Cell Dev. Biol.* **16**, 564–574 (2005).
150. Hickey, M. M., Lam, J. C., Bezman, N. A., Rathmell, W. K. & Simon, M. C. von Hippel-Lindau mutation in mice recapitulates Chuvash polycythemia via hypoxia-inducible factor-2alpha signaling and splenic erythropoiesis. *J. Clin. Invest.* **117**, 3879–3889 (2007).

151. van Rooijen, E. *et al.* Zebrafish mutants in the von Hippel-Lindau tumor suppressor display a hypoxic response and recapitulate key aspects of Chuvash polycythemia. *Blood* **113**, 6449–6460 (2009).
152. Kajimura, S., Aida, K. & Duan, C. Understanding Hypoxia-Induced Gene Expression in Early Development: In Vitro and In Vivo Analysis of Hypoxia-Inducible Factor 1-Regulated Zebra Fish Insulin-Like Growth Factor Binding Protein 1 Gene Expression. *Mol. Cell. Biol.* **26**, 1142–1155 (2006).
153. de Jong, J. L. O. & Zon, L. I. Use of the zebrafish system to study primitive and definitive hematopoiesis. *Annu. Rev. Genet.* **39**, 481–501 (2005).
154. Carradice, D. & Lieschke, G. J. Zebrafish in hematology: sushi or science? *Blood* **111**, 3331–3342 (2008).
155. Paffett-Lugassy, N. *et al.* Functional conservation of erythropoietin signaling in zebrafish. *Blood* **110**, 2718–2726 (2007).
156. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci.* **74**, 3171–3175 (1977).
157. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8 (1977).
158. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
159. Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457–463 (2010).
160. Grabowski, P. J., Seiler, S. R. & Sharp, P. A. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* **42**, 345–353 (1985).
161. Kornblihtt, A. R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**, 153–165 (2013).
162. Taliaferro, J. M., Alvarez, N., Green, R. E., Blanchette, M. & Rio, D. C. Evolution of a tissue-specific splicing network. *Genes Dev.* **25**, 608–620 (2011).
163. Sena, J. A., Wang, L., Heasley, L. E. & Hu, C.-J. Hypoxia regulates alternative splicing of HIF and non-HIF target genes. *Mol. Cancer Res. MCR* **12**, 1233–1243 (2014).
164. Weigand, J. E., Boeckel, J.-N., Gellert, P. & Dimmeler, S. Hypoxia-induced alternative splicing in endothelial cells. *PLoS One* **7**, e42697 (2012).
165. Memon, D. *et al.* Hypoxia-driven splicing into noncoding isoforms regulates the DNA damage response. *NPJ Genomic Med.* **1**, 16020 (2016).
166. Krawczak, M. *et al.* Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.* **28**, 150–158 (2007).
167. Lim, K. H., Ferraris, L., Filloux, M. E., Raphael, B. J. & Fairbrother, W. G. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11093–11098 (2011).
168. Park, K.-Y. *et al.* Desmin splice variants causing cardiac and skeletal myopathy. *J. Med. Genet.* **37**, 851–857 (2000).
169. Maher, E. R. *et al.* Phenotypic expression in von Hippel-Lindau disease: correlations with germline VHL gene mutations. *J. Med. Genet.* **33**, 328–332 (1996).

170. Sexton, A., Rawlings, L., McKavanagh, G., Simons, K. & Winship, I. A Novel von Hippel Lindau Gene Intronic Variant and Its Reclassification from VUS to Pathogenic: the Impact on a Large Family. *J. Genet. Couns.* **24**, 882–889 (2015).
171. Dhir, A. & Buratti, E. Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. *FEBS J.* **277**, 841–855 (2010).
172. Fox-Walsh, K. L. *et al.* The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl. Acad. Sci.* **102**, 16176–16181 (2005).
173. Jung, H. *et al.* Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* **47**, 1242–1248 (2015).
174. Singh, R. K. & Cooper, T. A. Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* **18**, 472–482 (2012).
175. Green, R. E. *et al.* Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. *Bioinforma. Oxf. Engl.* **19 Suppl 1**, i118-121 (2003).
176. Desmet, F.-O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).
177. Houdayer, C. *et al.* Evaluation of in silico splice tools for decision-making in molecular diagnosis. *Hum. Mutat.* **29**, 975–982 (2008).
178. Leman, R. *et al.* Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic Acids Res.* **46**, 7913–7923 (2018).
179. Baralle, D., Lucassen, A. & Buratti, E. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep.* **10**, 810–816 (2009).
180. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* **23**, 198–199 (1998).
181. Leeds, P., Peltz, S. W., Jacobson, A. & Culbertson, M. R. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev.* **5**, 2303–2314 (1991).
182. Brogna, S. & Wen, J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat. Struct. Mol. Biol.* **16**, 107–113 (2009).
183. Carter, M. S., Li, S. & Wilkinson, M. F. A splicing-dependent regulatory mechanism that detects translation signals. *EMBO J.* **15**, 5965–5975 (1996).
184. Lejeune, F. & Maquat, L. E. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* **17**, 309–315 (2005).
185. Nguyen, L. S., Wilkinson, M. F. & Gecz, J. Nonsense-mediated mRNA decay: inter-individual variability and human disease. *Neurosci. Biobehav. Rev.* **46 Pt 2**, 175–186 (2014).
186. Muhlrad, D. & Parker, R. Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* **5**, 1299–1307 (1999).
187. Chang, Y.-F., Imam, J. S. & Wilkinson, M. F. The Nonsense-Mediated Decay RNA Surveillance Pathway. *Annu. Rev. Biochem.* **76**, 51–74 (2007).
188. Perlick, H. A., Medghalchi, S. M., Spencer, F. A., Kendzior, R. J. & Dietz, H. C. Mammalian orthologues of a yeast regulator of nonsense transcript stability. *Proc. Natl. Acad. Sci.* **93**, 10928–10932 (1996).

189. Applequist, S. E., Selg, M., Raman, C. & Jäck, H.-M. Cloning and characterization of HUPF1, a human homolog of the *Saccharomyces cerevisiae* nonsense mRNA-reducing UPF1 protein. *Nucleic Acids Res.* **25**, 814–821 (1997).
190. Lejeune, F., Li, X. & Maquat, L. E. Nonsense-Mediated mRNA Decay in Mammalian Cells Involves Decapping, Deadenylation, and Exonucleolytic Activities. *Mol. Cell* **12**, 675–687 (2003).
191. Micale, L. *et al.* VHL frameshift mutation as target of nonsense-mediated mRNA decay in *Drosophila melanogaster* and human HEK293 cell line. *J. Biomed. Biotechnol.* **2009**, 860761 (2009).
192. Ong, K. R. *et al.* Genotype–phenotype correlations in von Hippel-Lindau disease. *Hum. Mutat.* **28**, 143–149 (2007).
193. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and Functions of Long Noncoding RNAs. *Cell* **136**, 629–641 (2009).
194. Dinger, M. E., Pang, K. C., Mercer, T. R. & Mattick, J. S. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput. Biol.* **4**, e1000176 (2008).
195. Gutschner, T. & Diederichs, S. The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biol.* **9**, 703 (2012).
196. Erdmann, V. A., Szymanski, M., Hochberg, A., de Groot, N. & Barciszewski, J. Collection of mRNA-like non-coding RNAs. *Nucleic Acids Res.* **27**, 192–195 (1999).
197. Pang, K. C., Frith, M. C. & Mattick, J. S. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet. TIG* **22**, 1–5 (2006).
198. Cheng, J. *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149–1154 (2005).
199. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, (2012).
200. Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**, 621–629 (2011).
201. Tsai, M.-C., Spitale, R. C. & Chang, H. Y. Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res.* **71**, 3–7 (2011).
202. Ulitsky, I. & Bartel, D. P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
203. Ribeiro, D. M. *et al.* Protein complex scaffolding predicted as a prevalent function of long non-coding RNAs. *Nucleic Acids Res.* **46**, 917–928 (2018).
204. Spitale, R. C., Tsai, M.-C. & Chang, H. Y. RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* **6**, 539–543 (2011).
205. Chujo, T., Yamazaki, T. & Hirose, T. Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochim. Biophys. Acta* **1859**, 139–146 (2016).
206. Schmitz, K.-M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* **24**, 2264–2269 (2010).
207. Devaux, Y. *et al.* Long noncoding RNAs in cardiac development and ageing. *Nat. Rev. Cardiol.* **12**, 415–425 (2015).
208. Cayre, A., Rossignol, F., Clottes, E. & Penault-Llorca, F. aHIF but not HIF-1 $\alpha$  transcript is a poor prognostic marker in human breast cancer. *Breast Cancer Res. BCR* **5**, R223–230 (2003).

209. Thrash-Bingham, C. A. & Tartof, K. D. aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia. *J. Natl. Cancer Inst.* **91**, 143–151 (1999).
210. Rossignol, F., Vaché, C. & Clottes, E. Natural antisense transcripts of hypoxia-inducible factor 1alpha are detected in different normal and tumour human tissues. *Gene* **299**, 135–140 (2002).
211. Uchida, T. *et al.* Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha. *J. Biol. Chem.* **279**, 14871–14878 (2004).
212. Sherwood, J. B. & Shouval, D. Continuous production of erythropoietin by an established human renal carcinoma cell line: development of the cell line. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 165–169 (1986).
213. Baralle, D. & Baralle, M. Splicing in action: assessing disease causing sequence changes. *J. Med. Genet.* **42**, 737–748 (2005).
214. Cooper, T. A. Use of minigene systems to dissect alternative splicing elements. *Methods San Diego Calif* **37**, 331–340 (2005).
215. Gaildrat, P. *et al.* Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol. Biol. Clifton NJ* **653**, 249–257 (2010).
216. Zhou, T. *et al.* Generation of human induced pluripotent stem cells from urine samples. *Nat. Protoc.* **7**, 2080–2089 (2012).
217. Si-Tayeb, K. *et al.* Highly Efficient Generation of Human Hepatocyte-like Cells from Induced Pluripotent Stem Cells. *Hepatol. Baltim. Md* **51**, 297–305 (2010).
218. Denham, M. *et al.* Multipotent Caudal Neural Progenitors Derived from Human Pluripotent Stem Cells That Give Rise to Lineages of the Central and Peripheral Nervous System. *Stem Cells Dayt. Ohio* **33**, 1759–1770 (2015).
219. Martin, L., Audran, M. & Marchand, A. Combined immuno-purification and detection of recombinant erythropoietins and activin receptor type II-Fc fusion proteins by isoelectric focusing for application in doping control. *Drug Test. Anal.* **11**, 168–172 (2019).
220. Lenglet, M. *et al.* New lessons from an old gene: complex splicing and a novel cryptic exon in VHL gene cause erythrocytosis and VHL disease. *Blood* (2018). doi:10.1182/blood-2018-03-838235
221. Andreutti-Zaugg, C., Scott, R. J. & Iggo, R. Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an in vivo fusion protein assay and conventional techniques. *Cancer Res.* **57**, 3288–3293 (1997).
222. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
223. Schuster, J. *et al.* Generation of human induced pluripotent stem cell (iPSC) lines from three patients with von Hippel-Lindau syndrome carrying distinct VHL gene mutations. *Stem Cell Res.* **38**, 101474 (2019).
224. Jouni Mariam *et al.* Toward Personalized Medicine: Using Cardiomyocytes Differentiated From Urine-Derived Pluripotent Stem Cells to Recapitulate Electrophysiological Characteristics of Type 2 Long QT Syndrome. *J. Am. Heart Assoc.* **4**, e002159
225. Si-Tayeb, K. *et al.* Urine-sample-derived human induced pluripotent stem cells as a model to study PCSK9-mediated autosomal dominant hypercholesterolemia. *Dis. Model. Mech.* **9**, 81–90 (2016).

226. Flores, S. K. *et al.* A synonymous VHL variant in exon 2 confers susceptibility to familial pheochromocytoma and von Hippel-Lindau disease. *J. Clin. Endocrinol. Metab.* (2019). doi:10.1210/jc.2019-00235
227. Randi, M. L. *et al.* Low frequency of VHL gene mutations in young individuals with polycythemia and high serum erythropoietin. *Haematologica* **90**, 689–691 (2005).
228. Crooke, S. T. Antisense strategies. *Curr. Mol. Med.* **4**, 465–487 (2004).
229. Shen, S. *et al.* Widespread establishment and regulatory impact of Alu exons in human genes. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2837–2842 (2011).
230. Faustino, N. A. & Cooper, T. A. Pre-mRNA splicing and human disease. *Genes Dev.* **17**, 419–437 (2003).
231. Usuki, F. *et al.* Specific inhibition of nonsense-mediated mRNA decay components, SMG-1 or Upf1, rescues the phenotype of Ullrich disease fibroblasts. *Mol. Ther. J. Am. Soc. Gene Ther.* **14**, 351–360 (2006).
232. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
233. Hart, R. P. & Goff, L. A. Long noncoding RNAs: Central to nervous system development. *Int. J. Dev. Neurosci.* doi:10.1016/j.ijdevneu.2016.06.001
234. Huang, C. & Jacobson, K. Detection of protein-protein interactions using nonimmune IgG and BirA-mediated biotinylation. *BioTechniques* **49**, 881–886 (2010).
235. Li, M. & Kim, W. Y. Two sides to every story: the HIF-dependent and HIF-independent functions of pVHL. *J. Cell. Mol. Med.* **15**, 187–195 (2011).
236. Lott, S. T. *et al.* High frequency loss of heterozygosity in von Hippel-Lindau (VHL)-associated and sporadic pancreatic islet cell tumors: evidence for a stepwise mechanism for malignant conversion in VHL tumorigenesis. *Cancer Res.* **62**, 1952–1955 (2002).
237. Glasker, S. *et al.* Reconsideration of biallelic inactivation of the VHL tumour suppressor gene in hemangioblastomas of the central nervous system. *J. Neurol. Neurosurg. Psychiatry* **70**, 644–648 (2001).
238. Prowse, A. H. *et al.* Somatic inactivation of the VHL gene in Von Hippel-Lindau disease tumors. *Am. J. Hum. Genet.* **60**, 765–771 (1997).
239. Steichen, C., Giraud, S. & Hauet, T. Les organoïdes rénaux. *médecine/sciences* **35**, 470–477 (2019).
240. Rochette, C. *et al.* Psychological impact of von Hippel-Lindau genetic screening in patients with a previous history of hemangioblastoma of the central nervous system. *J. Psychosoc. Oncol.* **36**, 624–634 (2018).





## RÉSUMÉ

---

La maladie de von Hippel-Lindau (VHL) est une maladie héréditaire prédisposant les patients au développement de tumeurs du système nerveux central, de la rétine, du rein et des surrénales tout au long de leur vie. La polyglobulie de Chuvash est une maladie héréditaire caractérisée par un taux de globules rouges élevé rendant le sang visqueux et augmentant le risque d'accident vasculaire cérébral.

Ces pathologies sont dues à des mutations d'un gène nommé *VHL*. Durant ma thèse, j'ai trouvé de nouvelles mutations dans des régions non étudiées de *VHL* chez des patients atteints de l'une de ces pathologies qui n'avaient pas eu de diagnostic génétique. Cette découverte permettra aux patients et à leurs familles de bénéficier d'un meilleur suivi médical. J'ai aussi identifié les mécanismes mis en jeu par ces nouvelles anomalies moléculaires à l'origine des pathologies. Ma thèse a permis des avancées majeures dans le diagnostic et la compréhension des pathologies héréditaires liées au gène *VHL*.

## MOTS CLÉS

---

VHL, épissage, hypoxie

## ABSTRACT

---

The von Hippel-Lindau disease is a hereditary syndrome which predispose patients to the development of nervous central, retina, kidney or adrenals tumors all along their lives. Chuvash polycythemia is a hereditary disease which is characterized by a high level of red blood cells which increases the blood viscosity and the risk of thrombosis or cerebrovascular accident.

These pathologies are due to mutations of one gene named *VHL*. During my thesis, I identified new mutations in non-investigated regions of this gene in patients who developed one of the pathologies and who don't have genetic diagnosis yet. This discovery will allow patients and their families to benefit from a better medical surveillance and therapy. I have also identified the mechanisms involved in these new molecular abnormalities at the origin of the pathologies. My thesis allowed major advances into the diagnosis and the comprehension of hereditary pathologies linked to the *VHL* gene.

## KEYWORDS

---

VHL, splicing, hypoxia

