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

Marion Lenglet, Florence Robriquet, Klaus Schwarz, Carme Camps, Anne Couturier, David Hoogewijs, Alexandre Buffet, Samantha Jl Knight, Sophie Gad, Sophie Couvé, et al.

TITLE:

New lessons from an old gene: complex splicing and a novel cryptic exon in *VHL* gene cause erythrocytosis and VHL disease

RUNNING TITLE:

Complex genetics in VHL disease and erythrocytosis

AUTHORS:

Marion Lenglet,[1,2,3§] Florence Robriquet,[2,3§] Klaus Schwarz,[4] Carme Camps,[5,6] Anne Couturier,[7] David Hoogewijs,[8] Alexandre Buffet,[9,10] Samantha JL. Knight,[5,6] Sophie Gad,[1,11] Sophie Couvé,[1,11] Franck Chesnel,[7] Mathilde Pacault,[2,12] Pierre Lindenbaum,[3] Sylvie Job,[13] Solenne Dumont,[2] Thomas Besnard,[3,12] Marine Cornec,[3] Helene Dreau,[14] Melissa Pentony,[5,6] Erika Kvikstad,[5,6] Sophie Deveaux,[15,16] Nelly Burnichon,[9,10,16,17] Sophie Ferlicot,[18,19] Mathias Vilaine,[2] Jean-Michaël Mazzella,[9,10,16,17] Fabrice Airaud,[12] Céline Garrec,[12] Laurence Heidet,[20] Sabine Irtan,[21] Elpis Mantadakis,[22] Karim Bouchireb,[20] Klaus-Michael Debatin,[23] Richard Redon,[3] Stéphane Bezieau,[3,12] Brigitte Bressac-de Paillerets,[24] Bin Tean Teh,[25] François Girodon,[26-27,36] Maria-Luigia Randi,[28] Maria Caterina Putti,[29] Vincent Bours,[30] Richard Van Wijk,[31] Joachim R. Göthert,[32] Antonis Kattamis,[33] Nicolas Janin,[34] Celeste Bento,[35] Jenny C. Taylor,[5,6] Yannick Arlot-Bonnemains,[7] Stéphane Richard,[1,11,15,16ß] Anne-Paule Gimenez-Roqueplo,[9,10,16,17ß], Holger Cario,[23]* Betty Gardie,[1,2,3, 36]*.

CORRESPONDENCE

Correspondence should be addressed to Betty Gardie: betty.gardie@inserm.fr

AFFILIATIONS

§ ß * These authors contributed equally to this work.

1- Ecole Pratique des Hautes, EPHE, PSL research University, France.
2- CRCINA, INSERM, Université de Nantes, Université d'Angers, Nantes, France.
3- L'institut du thorax, INSERM, CNRS, UNIV Nantes, Nantes, France.
4- Institute for Transfusion Medicine, University of Ulm and Institute for Clinical Transfusion Medicine and Immunogenetics Ulm, German Red Cross Blood Service Baden-Württemberrg-Hessen, Ulm, Germany.
5- Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.
6- Oxford NIHR Biomedical Research Centre, Oxford, UK.

7- Univ Rennes, CNRS, IGDR (Institut de génétique et développement de Rennes) - UMR 6290, F- 35000 Rennes, France.
8- Department of Medicine/Physiology, University of Fribourg, 1700 Fribourg, Switzerland.
9- INSERM UMR970, Paris-Cardiovascular Research Center at HEGP, Paris, France.
10- Université Paris Descartes, Faculté de Médecine, Paris, France, Equipe labellisée Ligue contre le Cancer.
11- INSERM UMR 1186, Institut Gustave Roussy, Université Paris-Saclay, Villejuif, France.
12- Service de Génétique Médicale, CHU de Nantes, Nantes, France.
13- Programme Cartes d'Identité des Tumeurs, Ligue Nationale Contre le Cancer, F-75013 Paris, France.
14- Molecular Diagnostics Laboratories, Molecular Haematology Dept, Oxford University Hospitals Trust, Oxford, UK.
15- Faculté de Médecine Paris-Sud, Le Kremlin-Bicêtre, France.
16- Réseau Expert National pour Cancers Rares de l'Adulte INCa "PREDIR" and Réseau d'Oncogénétique National INCa "Maladie de VHL et prédispositions au cancer du rein," Service d'Urologie, Assistance publique, Hôpitaux de Paris, Hôpital Bicêtre, Le Kremlin-Bicêtre, France.
17- Assistance Publique Hôpitaux de Paris, Hôpital européen Georges Pompidou, Service de Génétique, Paris, France.
18- Pathology Department, Hôpitaux Universitaires Paris Sud, Assistance Publique Hôpitaux de Paris, Le Kremlin Bicêtre, France.
19- Université Paris 11, Faculté de Médecine Paris Sud, Le Kremlin Bicêtre, France.
20- Assistance Publique Hôpitaux de Paris Centre de Référence des Maladies Rénales Héréditaires de l'Enfant et de l'Adulte (MARHEA), Service de Néphrologie Pédiatrique, Hôpital Universitaire Necker-Enfants malades, Paris, France
21- Assistance Publique Hôpitaux de Paris, Département de Chirurgie Pédiatrique, Hôpital Universitaire Necker-Enfants malades, Université Paris Descartes-Sorbonne Paris Cité, Paris, France.
22- Democritus University of Thrace Faculty of Medicine Alexandroupolis, Thrace, Greece.
23- Department of Pediatrics and Adolescent Medicine, University Medical Center Ulm, Ulm, Germany.
24- Gustave Roussy and INSERM U1186, Université Paris Saclay, Département de Biologie et Pathologies Médicales, Villejuif, F-94805, France.
25- SingHealth/Duke-NUS Institute of Precision Medicine, National Heart Centre Singapore, Singapore.
26- Service d'hématologie Biologique, Pôle Biologie, CHU Dijon, Dijon, France.
27- Inserm UMR1231 "Lipides Nutrition Cancer" équipe "Protéines de Stress et Cancer", FCS Bourgogne Franche Comté, LipSTIC Labex, F-21000 Dijon, France.
28- First Medical Clinic, Department of Medicine- DIMED, University of Padua, Padua, Italy.
29- Clinic of Pediatric Hemato-Oncology, Department of Woman's and Child's Health, University of Padua, Padua, Italy.
30- Service de génétique humaine du CHU Sart Tilman, B-4000 Liège, Belgium.
31- Department of Clinical Chemistry and Haematology, University Medical Center Utrecht, Utrecht, The Netherlands.
32- Department of Hematology, West German Cancer Center, University Hospital Essen, Essen, Germany.
33- First Department of Pediatrics, National and Kapodistrian University of Athens, Greece
34- Centre de Génétique Humaine, Cliniques universitaires Saint-Luc, B-1200 Bruxelles,

Belgium.
35- Department of Hematology, Centro Hospitalar e Universitario de Coimbra, Coimbra, Portugal.
36- Laboratory of Excellence GR-Ex.

Abstract:

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 von Hippel–Lindau disease, which is characterized by the development of highly vascularized tumors. Here, we identified a new *VHL* cryptic-exon (termed E1') deep in intron 1 that is naturally expressed in many tissues. More importantly, we identified mutations in E1' in seven families with erythrocytosis (one homozygous case and six compound-heterozygous cases with a mutation in E1' in addition to a mutation in *VHL* coding sequences) and in one large family with typical VHL disease but without any alteration in the other *VHL* exons. In this study we have shown that the mutations induced a dysregulation of the *VHL* splicing with excessive retention of E1' and are associated with a downregulation of VHL protein expression. In addition, we have demonstrated a pathogenic role for synonymous mutations in *VHL*-Exon 2 that alter splicing through E2-skipping in five families with erythrocytosis or VHL disease. In all the studied cases, the mutations differentially impact splicing, correlating with phenotype severity. This study demonstrates that cryptic-exon-retention or exon-skipping are new *VHL* alterations and reveals a novel complex splicing regulation of the *VHL* gene. These findings open new avenues for diagnosis and research into the VHL-related-hypoxia-signaling pathway.

Key points
• Mutations in a *VHL* cryptic exon may be found in patients with familial erythrocytosis or von Hippel-Lindau disease
• Synonymous mutations in *VHL* exon 2 may induce exon-skipping and cause familial erythrocytosis or von Hippel-Lindau disease

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, or 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 an adequate clinical management. However, even with the use of NGS panel diagnostics, the underlying genetic cause of presumed congenital erythrocytosis has been identified in less than one third of the patients in most published cohorts.

The molecular basis of VHL-related congenital erythrocytosis was first described in the autonomous Russian Republic of Chuvashia where this condition is an endemic disorder.[1] 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.[2]

Chuvash polycythemia arose 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) in Chuvash polycythemia. Subsequently, other missense *VHL* mutations in both alleles have been described in patients with congenital erythrocytosis.[3,4] Interestingly, it has been described some unexplained cases of patients with erythrocytosis in which only one heterozygous *VHL* mutation has been identified to date.[4-6]

VHL is located on 3p25-26 and has been reported to contain three exons (E1, E2, E3). The commonly described *VHL* transcript contains the three 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.[7] 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.[8-12] 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 α–subunits of HIF (HIF-

1α, 2α and 3α) 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-α ubiquitination and subsequent proteasomal degradation.[13,14] Under hypoxic conditions or when *VHL* is mutated, HIF-α remains stable and heterodimerizes with HIF-β, constituting a functional HIF factor. HIF transcriptionally activates a variety of genes involved in adaptation to reduced oxygen supply (e.g. erythropoiesis, angiogenesis, metabolism and cell survival). Dysregulation of the hypoxia pathway[13,14] is central to the development of erythrocytosis[1,4] (via upregulation of erythropoietin (EPO), a HIF2α target gene), but also in the development of tumors.[15] Indeed, *VHL* is a tumor suppressor gene, heterozygous mutations of which are associated with von Hippel-Lindau disease (Figure 1A).[16,17] The 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 haemangioblastomas, pancreatic neuroendocrine tumors, pheochromocytomas and clear-cell renal cell carcinomas (ccRCC).[18-20]

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[4-6] 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 twelve families linked to unexplained disease, including nine families with erythrocytosis and three families with VHL disease (Figure 1A). This study has led to the discovery of a novel cryptic-exon in the *VHL* gene and a complex regulation of *VHL* splicing.

Methods:

Complete materials and methods are detailed in supplemental data.

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.

Sanger sequencing

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

Whole Genome Sequencing

Whole genome sequencing (for Families F2, F3, F7) 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.[22,23] 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 PCR was performed using primers localized to flanking E1 and E3 exons. Taqman real-time PCR were performed on 20 ng of cDNA with the qPCR 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 library were pooled and prepared according to the denaturing and diluting libraries protocol for the Hiseq and GAIIx (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 Supplemental methods.

Reporter assays

The plasmid encoding the wild-type HA-VHL was kindly provided by Prof. William G. Kaelin Jr. An expression plasmid for the hypothetical X1-protein was constructed by subcloning the coding sequence following its synthesis (LifeTechnologies). Mutations of interest were introduced by site-directed mutagenesis (New-England Biolabs). 786-O cells were transfected (Polyplus) with constructs encoding wild-type or mutant proteins, Firefly-luciferase under the control of Hypoxia-Response-Elements (HRE-luciferase), and Renilla-luciferase for normalization.[24] After 24h, luciferase assays were performed using a Dual-Luciferase®Reporter Assay System (Promega).

Western Blotting

Cell lysates from the luciferase reporter assay were loaded into a Bis-Tris Mini Gel (4-12%) (Invitrogen). After transfert, the membrane (GE-Healthcare) was subsequently incubated with a mouse anti-HA antibody (BioLegend), and then a goat anti-mouse HRP-conjugated antibody (Jackson-Immuno-Research). 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.[8]

Minigene experiments

Minigene constructs were prepared in pCas2 plasmid, containing two artificial exons A and B (as described[25]), between which *VHL*-Exon1' or *VHL*-Exon2 with intronic flanking sequences were cloned. Cells were transfected (Polyplus) or nucleofected (Lonza). RNA were extracted 24h 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 (Family 1, Table 1). No other mutations in the three *VHL* canonical exons had been identified in genomic DNA of this patient. A RT-PCR using primers in E1 and E3 was performed using mRNA samples extracted from lymphoblastoid cell lines (LCL) 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 wild-type LCL. 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, is spliced to E1 at its 5' end, and to either E2 or E3 at its 3' terminus (Figure 1C). We showed that these isoforms are expressed in a variety of tissues and cell lines (Supplemental Fig.1). Their translation may theoretically lead to the production of a protein of 193 aa that contains the first 114 aa encoded by E1[26], and 79 additional 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), was deposited in NCBI. This isoform was predicted to encode a protein of 193 aa, named X1 (XP_011532380.1) with the sequence described above. The analysis of this region revealed strong conservation in primates, but a moderate to low conservation in more distant species (Supplemental Fig.2A-C); notably, the splice sites are identical to canonical sites (ttc**ag**/TC, AG/**gt**aag), and are highly conserved. *In silico* analysis of the donor (SD) and acceptor (SA) splice sites of E1' showed similar consensus values compared to other *VHL* exons (Figure 1C, upper panel). The capacity to translate a potential X1 protein is only conserved in higher primates (Supplemental Fig.2D). The deposited sequence has now been removed and replaced by a non-coding 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 Fig.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 Fig.1).

The new E1' cryptic-exon is mutated in patients with erythrocytosis or von Hippel-Lindau disease.

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

We investigated Chuvash polycythemia patients for whom the *VHL*-R200W had been found in a heterozygous state. Sanger sequencing identified the identical intronic variant c.340+770T>C in a singleton (F2)[5] and two affected brothers (F3) (Figure 2B). In addition, we identified a duplication c.340+694_711dup in the proband of F4 and F5[6] 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 splice-acceptor (SA) site **ag**/TC of E1'. This variant is described as a rare polymorphism (rs98274567) in the NCBI database (Table 1).

As 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 SNP (single nucleotide polymorphism) rs779808 associated with the *VHL*-R200W mutation (described[27] to be located downstream of E1', in position c.340+1150). Segregation analysis (F4, F5, F6) or Chuvash core-haplotype analysis (F2, F3) demonstrated that all patients are compound heterozygous, with one mutation being inherited from each parent (Supplemental Fig.5).

In a parallel independent study, whole genome sequencing (WGS) was being 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 further significant mutations in biologically relevant genes (Supplemental Fig.6). This study also conducted WGS for a trio with congenital erythrocytosis for which prior whole exome sequencing (WES) 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 von Hippel-Lindau disease may also fail to detect mutations in canonical *VHL* exons. An example of such family, F8, with hereditary hemangioblastoma, clear-cell renal-cell carcinoma and pheochromocytoma has been studied. Microsatellite analysis demonstrated a co-segregation of markers surrounding the *VHL* region with the disease (Supplemental Fig.7). Sequencing of the new E1' cryptic-exon in this family identified two 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 segregate in six patients who developed the disease, and were absent in four healthy tested descendants, indicating their presence on a single disease-associated allele (Figure 2B). Sequencing of tumor DNAs did not display loss of heterozygosity (LOH), suggesting that a wild-type *VHL* deletion (as observed in classic VHL disease) may not be prerequisite in cells of patients with this specific *VHL* genotype (Supplemental Fig.4).

Expression study of the new VHL isoforms in patients' cells.

RNA-sequencing (RNA-Seq) analysis of patients' LCL and tumors demonstrated that genomic variants in E1' are associated with an up-regulation of transcripts containing the E1' cryptic-exon compared with controls (Figure 3A, Supplemental Fig.8). In addition, the mutated allele is preferentially expressed (more than 70%), suggesting a causal relationship between the genetic variants and the E1' retention (Supplemental Fig.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 Figure 1C), which showed a strong upregulation of a unique isoform containing E1' spliced with E1 in LCL and tumors from patients of F8 compared with controls (Figure 3B). The study therefore focused on the isoform E1-E1'. Transcripts that retain E1' spliced with E1 contain (in-frame with E1) a premature termination codon (Figure 1C) and are likely targeted for degradation according to nonsense-mediated mRNA Decay (NMD) mechanisms. We therefore investigated the expression of transcripts that

contain 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' *versus* other isoforms (Figure 3C, Supplemental Fig.9). These results indicated that isoforms with E1-E1' junction are degraded by NMD and may fail to produce proteins. We performed RT-PCR on LCL sample 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 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 Fig.10).

In order to measure the impact of the splicing dysregulation on pVHL expression, we performed an immunoblot with an antibody able to recognize the aa encoded by E1 and, therefore, able to detect the different pVHL isoforms.[8] The antibody detected overexpressed exogenous X1 protein following transfection of plasmid encoding X1 (Figure 4A). However, it failed to detect the endogenous protein, even in patients' samples that overexpress 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 and 4B).

Functional characterization of the mutated VHL-E1’.

As the hypoxia pathway represents the major pathway involved in the genesis of the secondary erythrocytosis and VHL disease[13-15,28], we performed functional studies of the hypothetical wild-type or mutated X1 proteins using Hypoxia Response Element (HRE)-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 Fig.11). We next focused our study on a potential impact of the E1' variants on splicing by performing splicing reporter minigene assays in various cell lines. The experiments showed that splicing of the wild-type E1' is 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 the 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 has a more pronounced effect than each variant individually (the SNP showing very low effect), suggesting a synergistic effect of the variants on splicing (Supplemental Fig.12). *In silico* analysis of the c.340+617C>G consistently predicted a severe alteration of splicing by the creation of an Exonic Splicing Enhancer (ESE) site (Supplemental Fig.13).

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

Synonymous mutations in VHL-Exon 2 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 two patients with erythrocytosis (F9, F10, Figure 5A, Table 1). The mRNAs extracted from LCLs established from the different members of these two families were reverse transcribed and sequenced. Comparison of chromatograms obtained by sequencing of DNA *versus* 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), has been identified in two families (F11, F12) with von Hippel-Lindau disease (Figure 5B). This heterozygous mutation segregates with the disease in three generations. Sequencing of DNA extracted from the pheochromocytoma of F11-III.1 showed a loss of the wild-type allele in the tumor, demonstrating LOH 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 showed a significant change in the ratio of expressed *VHL* isoforms, with a higher expression of the E1-E3 transcripts in LCL of patients homozygous for D143D and in the pheochromocytomas with P138P mutation. These results suggested 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 (ESE) motifs (Supplemental Fig.14).

In order to specifically quantify the different *VHL* isoforms, we performed quantitative RT-PCR on LCL using Taqman probes complementary to the *VHL* E1-E2 or E1-E3 junctions (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 wild-type isoform expression level. These results demonstrated that synonymous mutations in E2 affect mRNA isoform production.

In order 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 to disease severity in minigene experiments.

In order 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).[4,6,29] Indeed, these mutations may also impact splicing rather than VHL protein function. We first evaluated the loss-of-function of these mutants by reporter assays. We showed that these could downregulate HIF in a manner similar to the wild-type protein, in contrast to VHL proteins lacking E2 (pVHLΔE2) which are described to be non-functional in terms of their regulation of HIF activity[8-10,12] (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 cause 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 EPO).

Splicing dysregulation of VHL is causal in the developement of disease.

RNA-seq confirmed that the synonymous mutations were not silent but instead induce potent E2-skipping (Figure 6C, Supplemental Fig.15A). Further 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 Fig.15B)[30]. We then re-analyzed our independent cohort of pheochromocytomas[30] 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, the functional studies of *VHL* mutants have been performed on missense mutations. We describe here, for the first time, functional studies of *VHL* mutations which do not have an impact on the coding sequence, but which influence the *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) can have an impact on *VHL* splicing and should be considered as pathogenic mutations. Our study points to a particular region in the 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[17] on splicing, in the same way as we described for two missense mutations (P138L, G144R). We observed that, depending on the mutation in this region, the impact on splicing can be moderate (D143D, G144R, P138L) or severe (P138P), which correlates with the severity of the disease seen in individuals carrying these *VHL* mutations (erythrocytosis *versus* cancers). This observation confirms the hypothesis of a continuum model of tumor suppression by *VHL*.[21,31]
Regarding the erythrocytosis developed in patients homozygous for D143D, both probands (F9 II.1 and F10 II.1) present mutations in the beta-globin gene (*HBB*) that induce hemoglobin

instability or thalassemia (Table 1). This may compensate 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 six families with an erythrocytosis previously associated with a heterozygous mutation in *VHL* rather than a homozygous mutation. Our investigations further confirmed that polycythemia associated with *VHL* mutation is definitely an autosomal recessive disease. In addition, we identified an E1' homozygous mutation in a patient with an erythrocytosis of unknown origin. This result demonstrates the causal role of the alteration 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 WES data. Our data showed that these newly described E1'-containing transcripts may be polyadenylated (as they are captured by polydT in RNAseq) but are also likely to be subjected to NMD and may therefore fail to produce a protein. However, we cannot exclude the possibility that they are 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.[26] Nonetheless, the hypoxia-dependent reporter assays failed to identify a potential direct role of this isoform on the HIF pathway. Instead, our results provide compelling evidence that mutations in E1' induce a severe retention of this E1'cryptic-exon which correlates 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 a reduced HIF binding by the mutant pVHL (as seen in the Chuvash polycythemia mutation *VHL*-R200W[1]) may lead 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 is 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 about one third of the 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 in the list of regions routinely sequenced in patients with congenital erythrocytosis. Second, the detection of molecular changes has implications for the clinical management of patients. For example, phlebotomy in patients with VHL- or HIF2α-related erythrocytosis may even worsen the clinical situation by increasing the risk and severity of pulmonary hypertension in these patients.[32] Third, the confirmation of the continuum model of tumor suppression by VHL helps to understand the very low frequency of secondary tumors in patients with VHL-related erythrocytosis.[33] On the other hand, 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 due to E1' mutations will also benefit from regular screening for tumors. Finally, the detection of these new genetic changes will also allow appropriate genetic counselling 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 (e.g. angiogenesis, erythropoiesis) and pathological (e.g. 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 in 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, hemangioblastomas, in addition to all types of sporadic tumors with altered hypoxia signaling.

Acknowledgments:
This study was supported by grants from the Région Pays de la Loire, project "EryCan"; the ANR (PRTS 2015 "GenRED"); the labex 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 NIHR 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 Wellcome Trust and Department of Health. We acknowledge also funding contribution from the Wellcome Trust Core Award Grant Number 203141/Z/16/Z. AB 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. DH is supported by the National Center of Competence in Research Kidney.CH. The authors thank Richard Breathnach, Jean Feunteun, Judith Favier, Sylvie Hermouet for scientific discussions; Helena M Silva, Caroline Abadie, Sophie Giraud, Florence Fellmann, Pascal Pigny, Vinciane Dideberg, Segers Karine for patient recruitment and medical advices; 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**.** We are grateful to Mario Tosi, Pascaline Gaildrat, and Alexandra Martins (Inserm U1079-IRIB, University of Rouen, Rouen, France) for kindly providing the pCAS2 minigene plasmid.
We would also like to thank the European COST Networks MPN&MPNr-EuroNet (COST Action BM0902 "Molecular Diagnosis of MyeloProliferative Neoplasms Euronet") and Hypoxianet (COST Action action TD0901 "Hypoxia sensing, signaling and adaptation").

Authorship contributions:

ML, FR, KS, AC, DH, AB, SG, SC, FC, MP, TB, SF, MV, YAB, BG performed experiments; CC, SJLK, JCT, MP, EK and HD conducted genome sequencing or data analysis & interpretation; PL, MC, SDu, SJ, ML performed bioinformatics analyses; HC, KS, SDe, NB, JMM, FA, CG, LH, SI, EM, KB, KMD, BBdP, FG, MLR, MCP, VB, RVW, JG, AK, NJ, CB, APGR, SR conducted the medical study; BG, DH, ML, FR wrote the manuscripts; BG, HC, designed the study; BG directed the study; all authors contributed to the research and approved the final manuscript.

Disclosure of Conflicts of Interests:

The authors declare no competing financial interests.

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. *Book MPN & MPNrEuroNet COST final publication, 2015*. 2015.
4. Bento C, Percy MJ, Gardie B, et al. 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 U S A*. 1998;95(15):8817-8822.
8. 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.
9. Clifford SC, Cockman ME, Smallwood AC, et al. Contrasting effects on HIF-1alpha regulation by disease-causing pVHL mutations correlate with patterns of tumourigenesis in von Hippel-Lindau disease. *Hum Mol Genet*. 2001;10(10):1029-1038.
10. Gnarra 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 pVHL172 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, Gnarra 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, Couve S, Gad S. Von Hippel-Lindau: how a rare disease illuminates cancer biology. *Semin Cancer Biol*. 2013;23(1):26-37.
21. 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.
22. 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.
23. 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.
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, 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.
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.

Family n°/patient	nucleotide variant on allele 1 / allele 2	Impact on protein	Year of birth/Age at diag	Sex	Hb	Ht	Red cells	EPO	Phenotye	Other mutation
F1, II.1	c.429C>T/ c.340+770T>C	VHL p.Asp143Asp/ X1 p.Ser179Pro?	1991/ 15 years	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 years	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 months	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/ 6months	M	14.6	45			Erythrocytosis	-
F4, II.1	c.598C>T/ c.340+694_711dup	VHL p.Arg200Trp/ X1 p.Trp159X?	1990/ 7 years	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 years	M	16.2	60	5.9		Erythrocytosis	-
F6 , II.1	c.492G>T/ c.340+574A>T♦	VHL p.Gln164His/ SA	2014/ 6 months	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 years	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 months	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 years	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	-

Table Legend:

Table 1. Description of variants identified in *VHL*-E1' and *VHL*-E2 with associated clinical manifestations. 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). WT: wild-type. ♦: the nucleotide change has been reported as rs982745672 with a global MAF that corresponds to T=0.00007/2 (TOPMED). ♦♦: the nucleotide change has been reported as rs73024533 with a global MAF that corresponds to C=0.0026/13 (1000 Genomes) and C=0.0050/147 (TOPMED). 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'). SA: mutation in splice acceptor site; Diag: diagnosis, M: Male, F: Female, N: normal values. The normal values corresponds to: Hb (Hemoglobin) (g/dL) = M: 13-18 and F: 12-15; Ht (hematocrit) (%)= M: 40-52, F: 37-47; Red cells (million/mm^3)= M: 4.2-5.7, F: 4.2-5.2; EPO (mU/ml)= 5-25.

Figures Legend:

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. Von Hippel–Lindau 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 an elevated red blood cell number. This study describes families with typical VHL-related phenotypes associated with an unexpected *VHL* status (i.e. either synonymous mutations or no alterations in *VHL*). (B) RT-PCR using primers specific for E1 and E3 was performed on mRNA extracted from lymphoblastoid cell line (LCL) established from Controls and patients of the Family 1. WT: wild-type. *: denotes larger fragments that were subsequently cloned and sequenced. (C) Schematic representation of the *VHL* gene and its products. The different *VHL* exons are represented on a scale: E1: 340bp from the ATG initiation codon, E1': 259bp, E2: 123 bp, E3: 179bp 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 isoforms containing E1' spliced with *VHL* exons have been identified by cloning and sequencing in the laboratory but have been described later on by NCBI as transcripts able to produce a protein termed X1. Consensus values of donor (SD) and acceptor (SA) splice sites 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.

Figure 2: Identification of mutations in the new *VHL* cryptic-exon in seven patients with erythrocytosis and a large family with the 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 von Hippel Lindau disease. The genotypes have been elucidated by sequencing both parents and proband (F1, F5, F6), deduced by sequencing of one parent and proband (F4, the mutation deduced being under brackets), or deduced from allele cloning of proband carrying the conserved Chuvash mutation and core haplotype (F2, F3) confirming the transmission of the mutations by one of each parents (for F2 and F3, the identity of the transmitting parent being unknown, the mutation is represented by a white circle under brackets). The genotype of parents from F7 has been elucidated from WGS data. The numbers in italics (F8) indicate the age of the patient at tumor diagnosis.

Figure 3: Expression study of the new *VHL* transcripts isoforms in patient cells.

(A) Sashimi plots from RNA-Seq data obtained from samples (LCLs or pheochromocytoma, Pheo), of three 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 Fig. 8. (B) TaqMan quantification of the different *VHL* isoforms from samples of F8 performed using probes specific to the *VHL* E1-E2, the translated sequence upstream E1' or E1-E1' junctions. Relative gene expression has been normalized to LCL control (C1) fixed at 1 (mean results of technical duplicates). C: healthy control. (C) TaqMan quantification of the different *VHL* isoforms in LCLs (established from two independent controls and from patients of F1, F2 and F8) cultured in the absence or presence of puromycin, an inhibitor of Nonsense-Mediated mRNA Decay. TaqMan probes are specific to the

VHL E1-E1' or E1-E2 junctions. The graph resumes experiments performed on LCL cultured independently 3 times and quantified in duplicates. Data are represented as the mean +/- SEM. Statistical p value: * $p<0.05$, ** $p<0.01$ based on a *t*-test. (D) RT-PCR using primers specific for E1 and E3 was performed on mRNA extracted from LCLs established from controls and patients of the Family 6 (carrying the mutation c.340+574A>T that targets the Splice Acceptor (SA) site of E1'). Patients with erythrocytosis are indicated in red. On the right, the spliced isoforms are schematically represented. *: denotes larger fragments that contains 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 Fig.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 are presented (B). Relative gene expression has been normalized to GAPDH expression and results obtained with LCL control (C1) have been fixed at 100%. Data are represented as the mean +/- SEM. ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$ based on a *t*-test. A VHL-antibody which recognized VHL-E1 was able to detect the hypothetical X1 protein. In the most left lane, control LCLs were transfected with an expression vector containing the coding sequence for X1. Five µg of proteins were blotted *vs.* 45µg for other samples to avoid signal saturation with overexpressed X1. Patients with erythrocytosis are indicated in red. (C) Functional HRE-dependent reporter assays were performed in 786-O cells (i.e. *VHL* negative 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 wild-type and mutated X1 to downregulate Firefly luciferase activity (related to HIF activity) was compared to pVHL and in competition with pVHL. An immunoblot using an antibody specific to the hemagglutinin tag was used to detect HA-VHL and HA-X1. The 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. ** $p<0.01$, based on *t*-test. (D) Characterization of *VHL*-E1' retention by the Minigene experiment (representative picture of agarose gel from n=3). RT-PCR was performed on mRNA obtained from cell lines transfected with a minigene construct containing *VHL*-E1' (wild-type or mutated) flanked by large intronic

sequences cloned between the *SERPING1* 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 with or without E1') was analyzed. Two wild-type constructs containing E1' were used; one contains SNP rs779808, 340+1150T>C. Bands corresponding to EA and EB spliced together or with *VHL*-E1' are indicated on the right. * corresponds to unspecific bands verified by sequencing. The minigene experiment performed with the construction carrying the mutation c.340+574A>T (that targets the Splice Acceptor (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).

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 two families (F9, F10) with erythrocytosis (A), and two families (F11, F12) with von Hippel-Lindau disease. (C) Results of RT-PCR using mRNA extracted from LCLs (F9, F10) and leukocytes and tumor material (pheochromocytoma) (F11, 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 junctions. Relative gene expression has been normalized to LCL control (C1). C: healthy control, F: Families. ** $P < 0.01$, *** $P < 0.001$ (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 has been normalized to GAPDH expression and results obtained with LCL control (C1) have been fixed at 100%. Data are represented as the mean +/- SEM. ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$ based on a *t*-test.

Figure 6: Functional study of synonymous mutations in *VHL*-E2.

(A) Functional 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/VHLΔE2) is used as a negative controle. 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-2α using the HRE-luciferase reporter plasmid. Immunoblots using an antibody specific for the hemagglutinin tag 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, HK2), pheochromocytoma (PC12), erythroid cell line (UT7 cultured with EPO); 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 *SERPING1* 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 Fig. 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 paraganglioma/pheochromocytomas (recruited by the French COMETE network) that identified homogeneous molecular subgroups associated with susceptibility genes (Burnichon et al., 2011). exp: relative expression compared to the median.

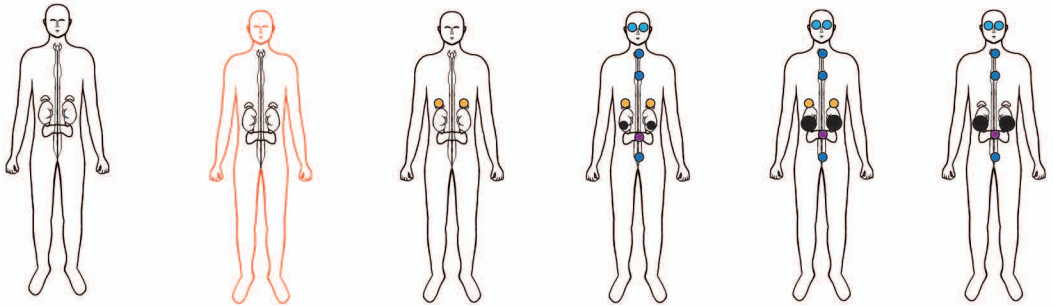
Figure 1

A

Phenotype :

Healthy | Erythrocytosis | von Hippel-Lindau Disease (2C | 2A | 2B | 1)

- Pheocromocytoma
- Neuroendocrine Pancreatic tumor
- CNS Hemangioblastoma
- Retinal Hemangioblastoma
- Renal cell carcinoma

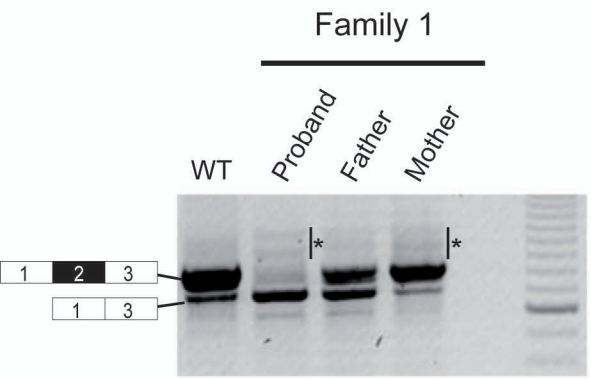


Representative Genotype : R200W/WT | R200W/R200W | L188V/WT | Y98H/WT | R167Q/WT | C162F/WT

Genotype of unsolved cases :

F1 : D143D/WT?
F2,F3, F4 : R200W/WT?
F5 : G144R/WT?
F6 : Q164H/WT
F7 : WT/WT?
F9,F10 : D143D/D143D
F8 : WT/WT?
F11, F12 : P138P/WT

B



C

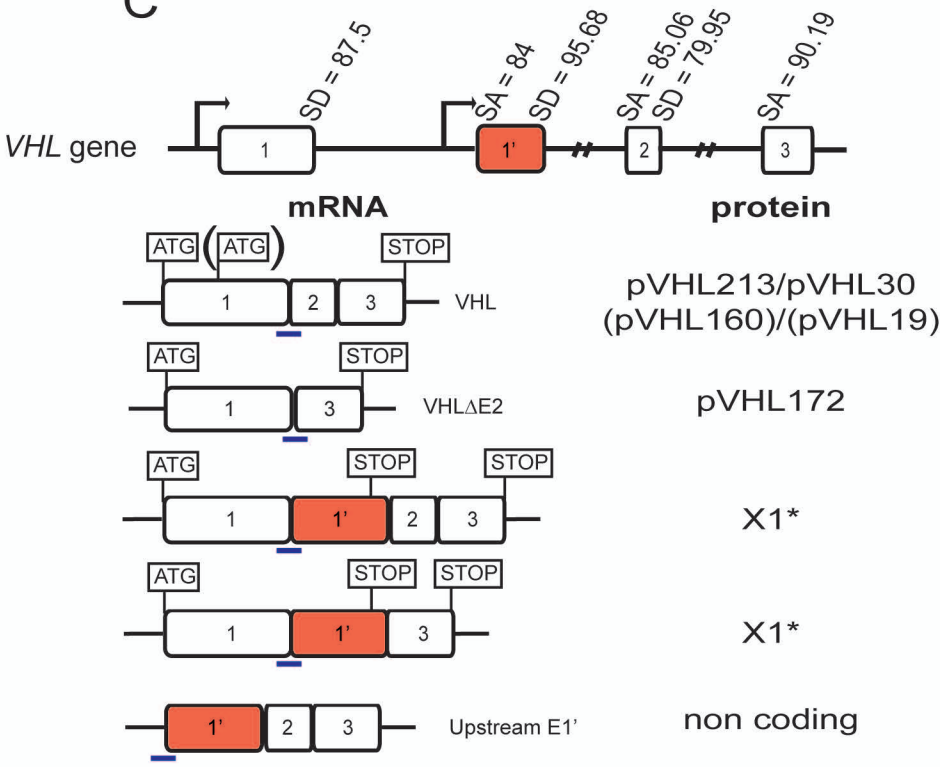


Figure 2

A

- c.340+574A>T
- c.340+694_711dup
- c.340+770T>C
- c.340+816A>C

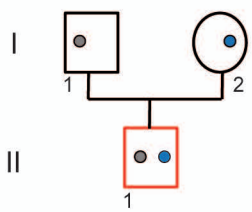
Erythrocytosis

- c.340+617C>G + c.340+648T>C

VHL disease

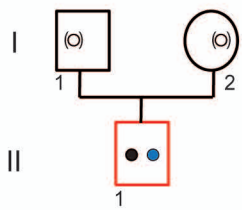
B

Family 1



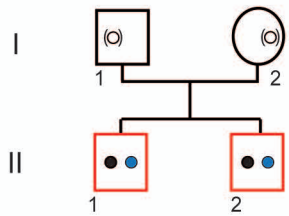
- c.429C>T, p.Asp143Asp
- c.340+770T>C

Family 2



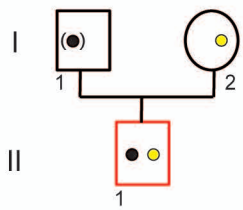
- c.598C>T, p.Arg200Trp
- c.340+770T>C

Family 3



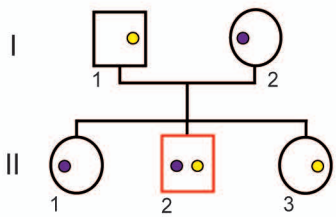
- c.598C>T, p.Arg200Trp
- c.340+770T>C

Family 4



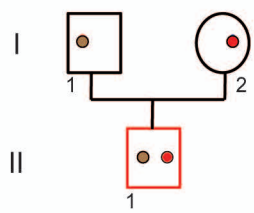
- c.598C>T, p.Arg200Trp
- c.340+694_711dup

Family 5



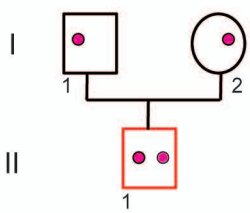
- c.430G>A, p.Gly144Arg
- c.340+694_711dup

Family 6



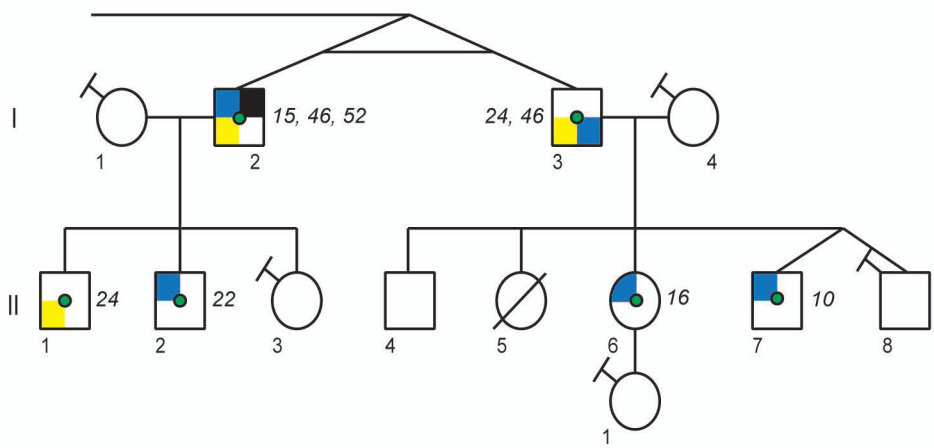
- c.492G>T, p.Gln164His
- c.340+574A>T

Family 7



- c.340+816A>C

Family 8

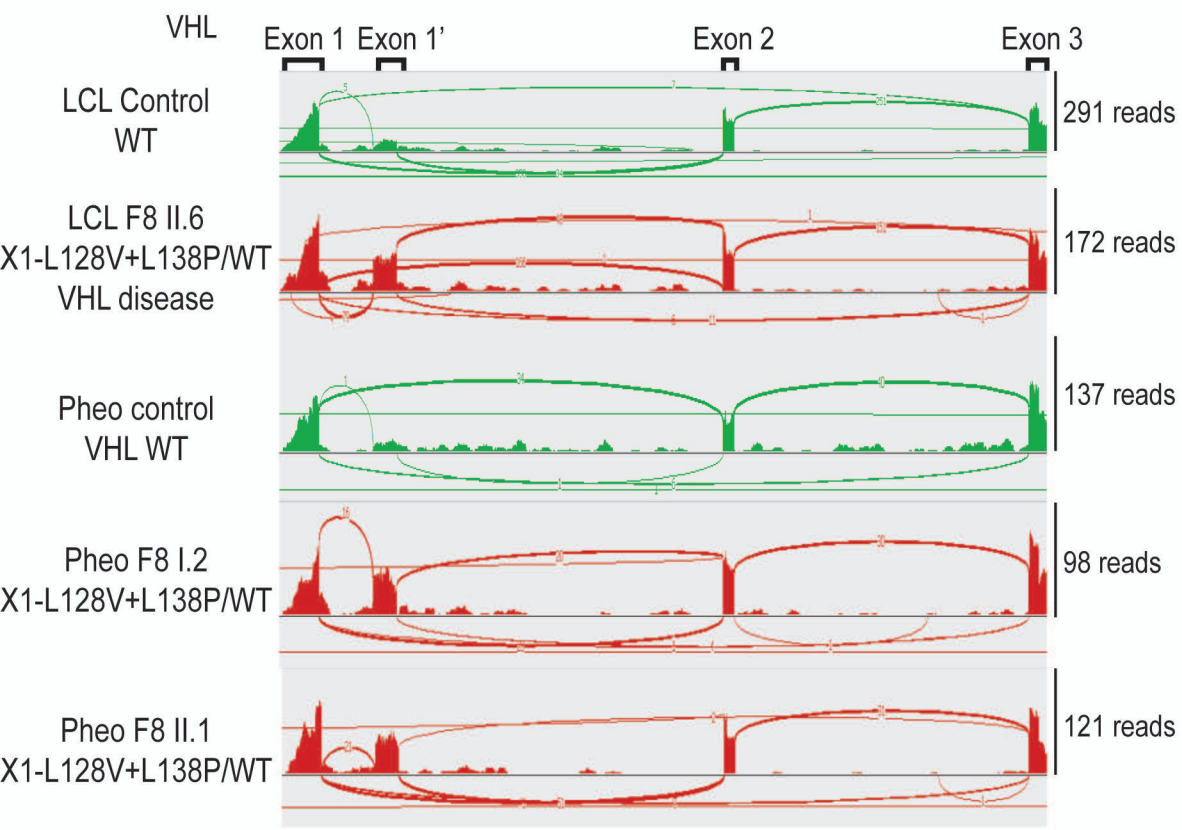


Legend:

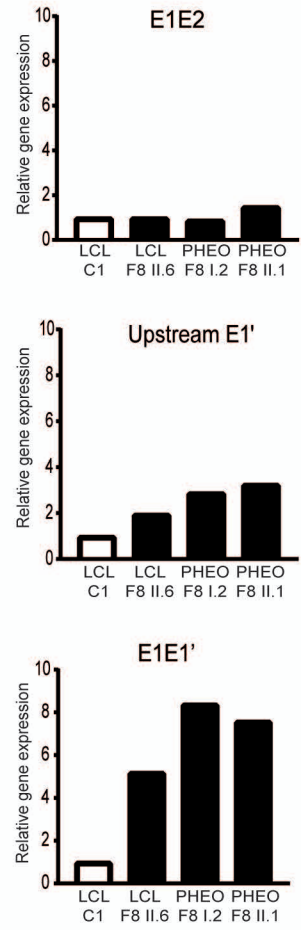
- Erythrocytosis
- Retinal Hemangioblastoma
- Pheochromocytoma
- CNS Hemangioblastoma
- Renal cell carcinoma
- c.340+617C>G c.340+648T>C
- Genetically tested, *VHL* WT

Figure 3

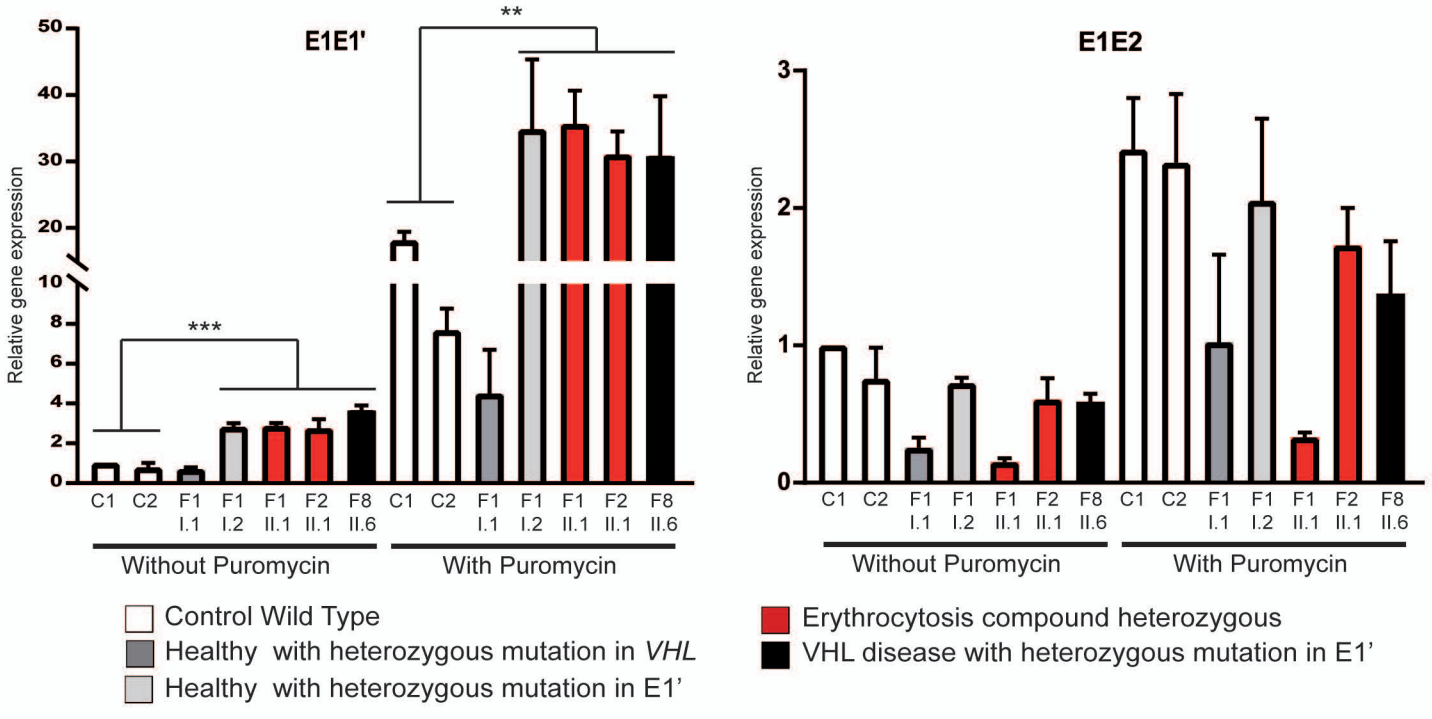
A



B



C



D

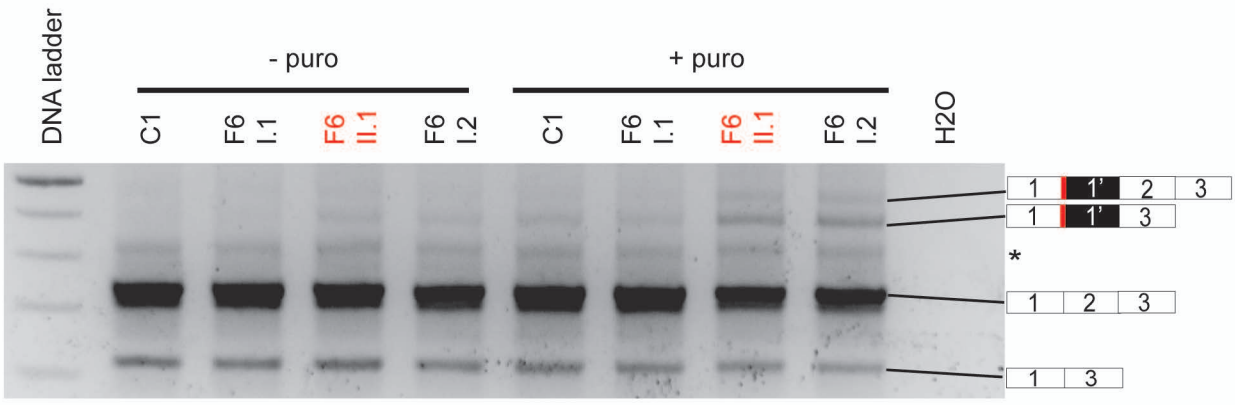
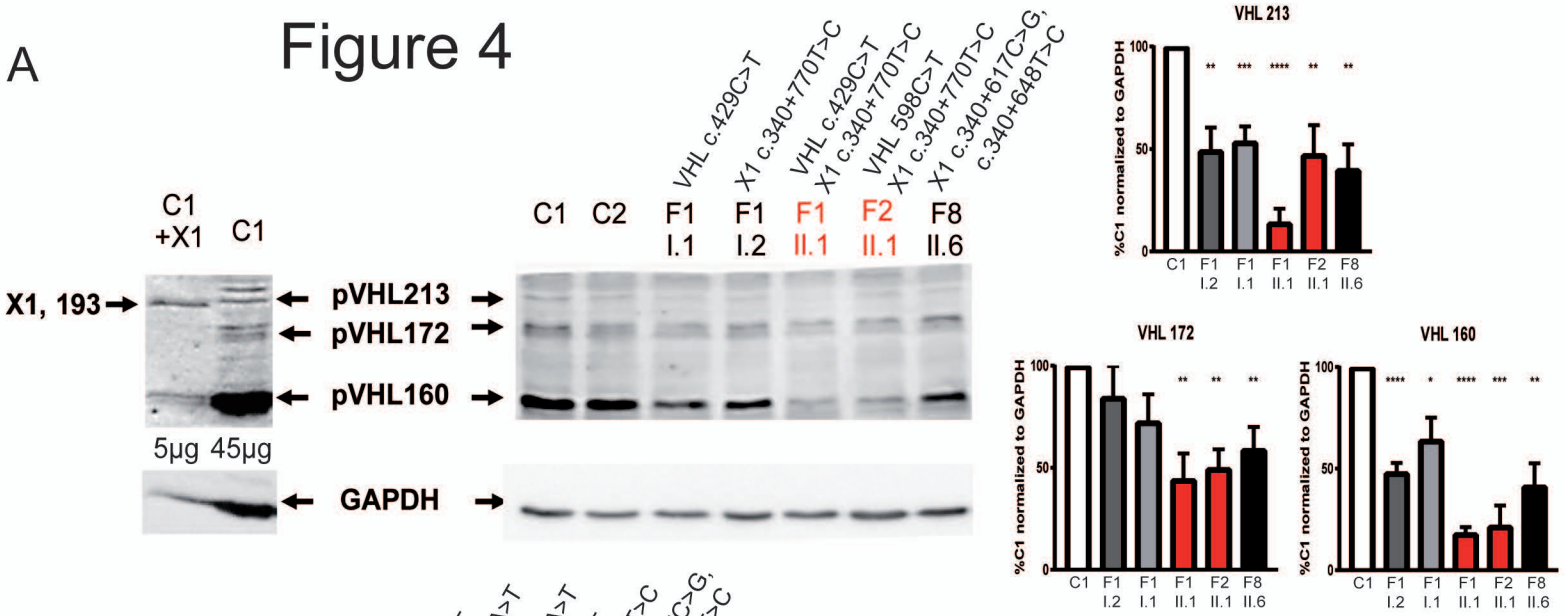
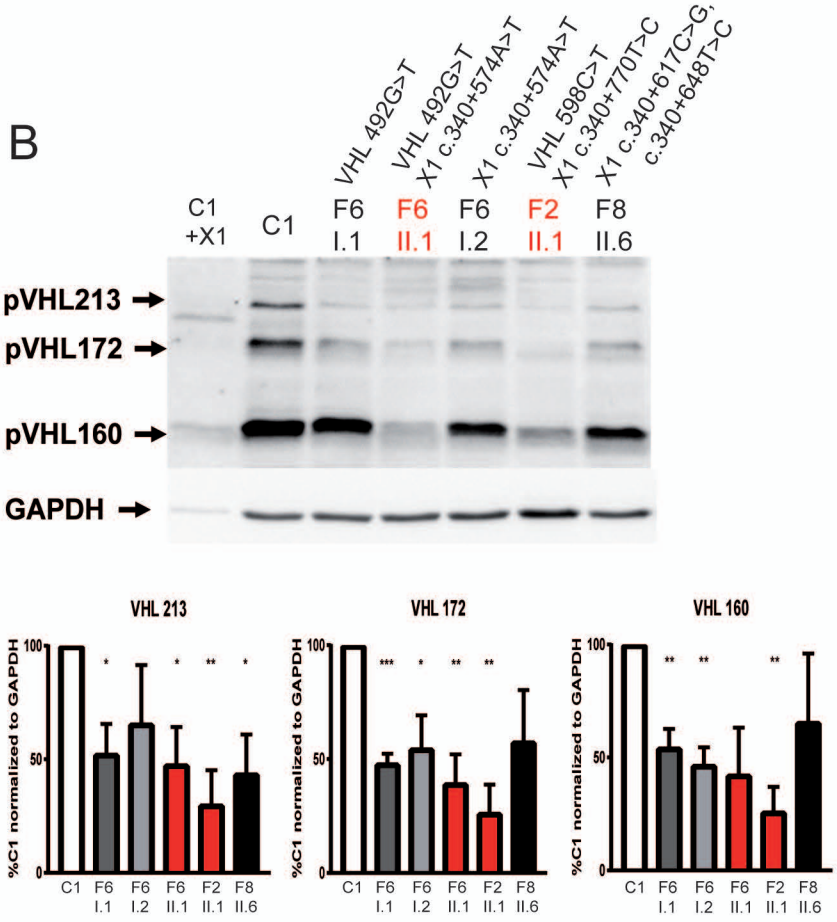


Figure 4

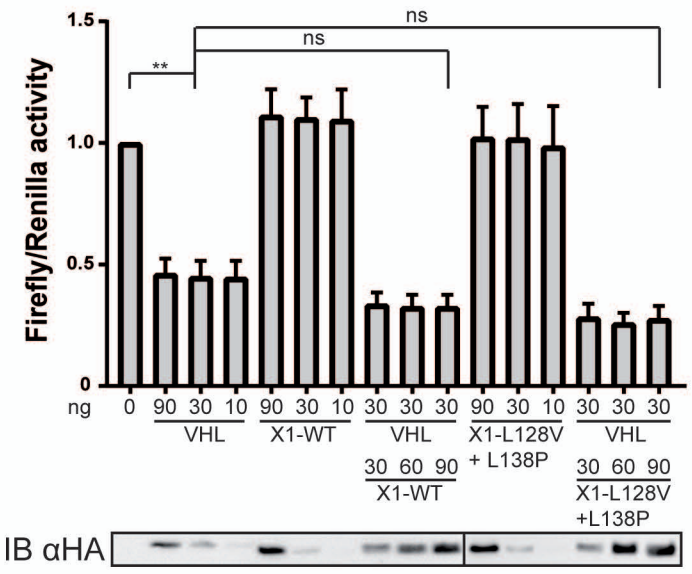
A



B



C



D

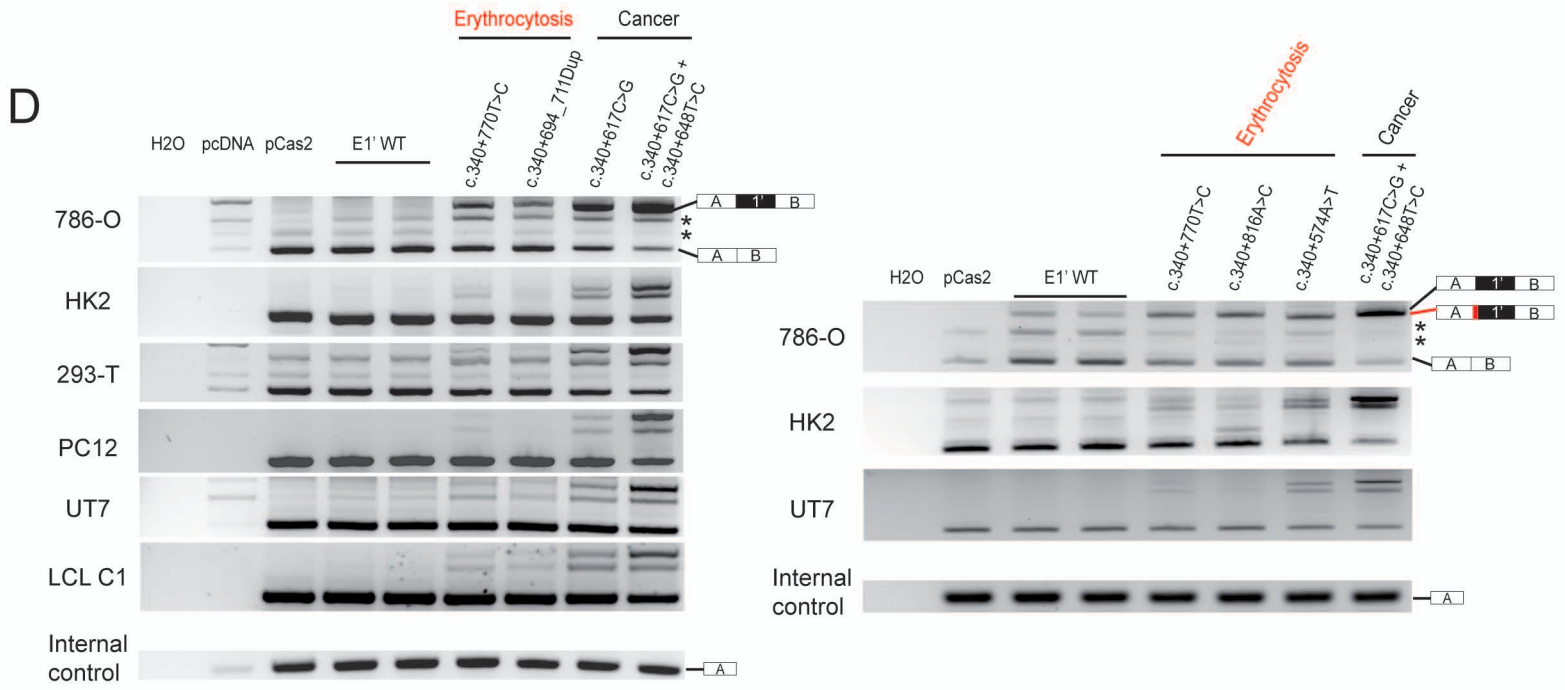


Figure 5

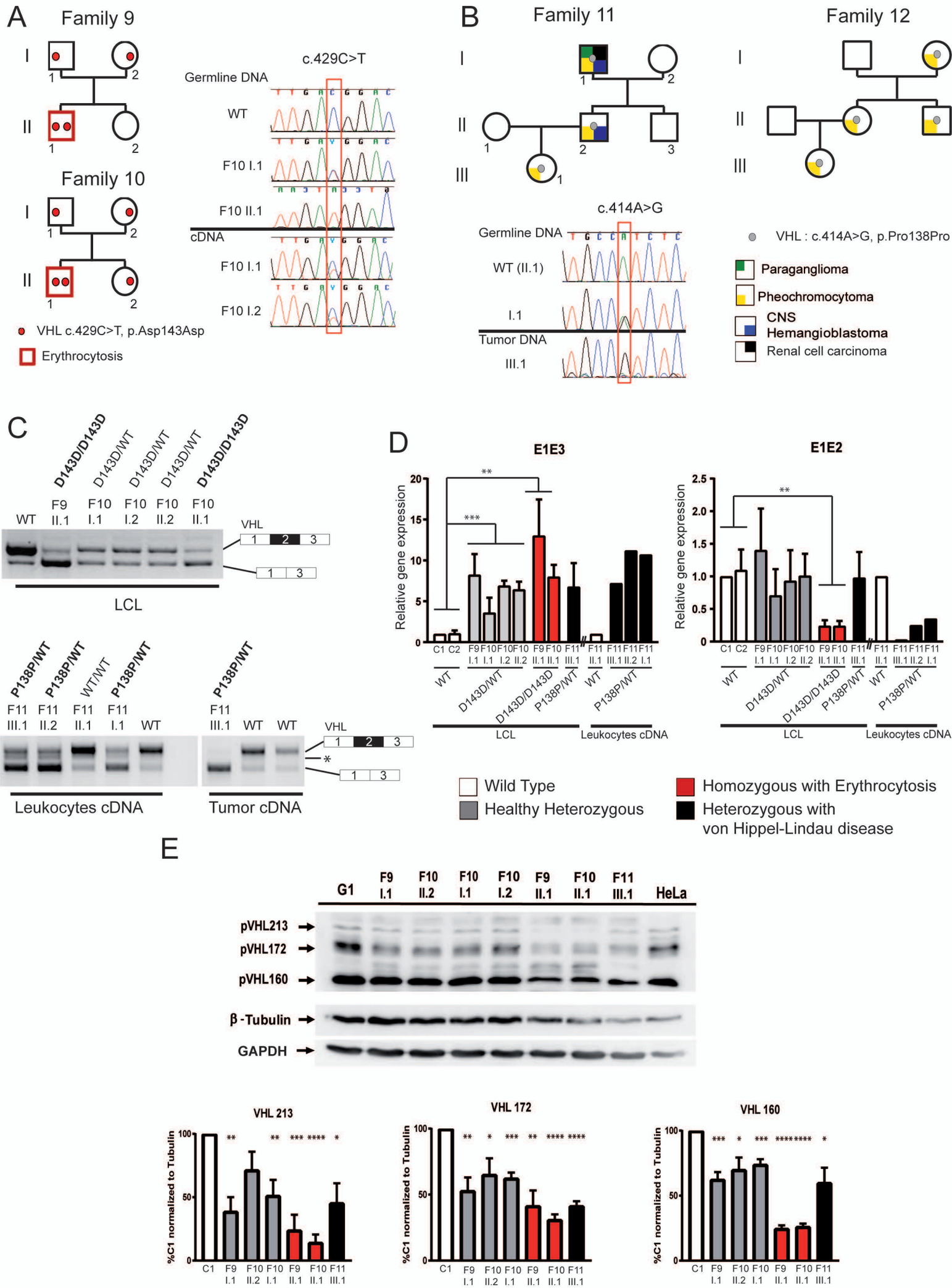
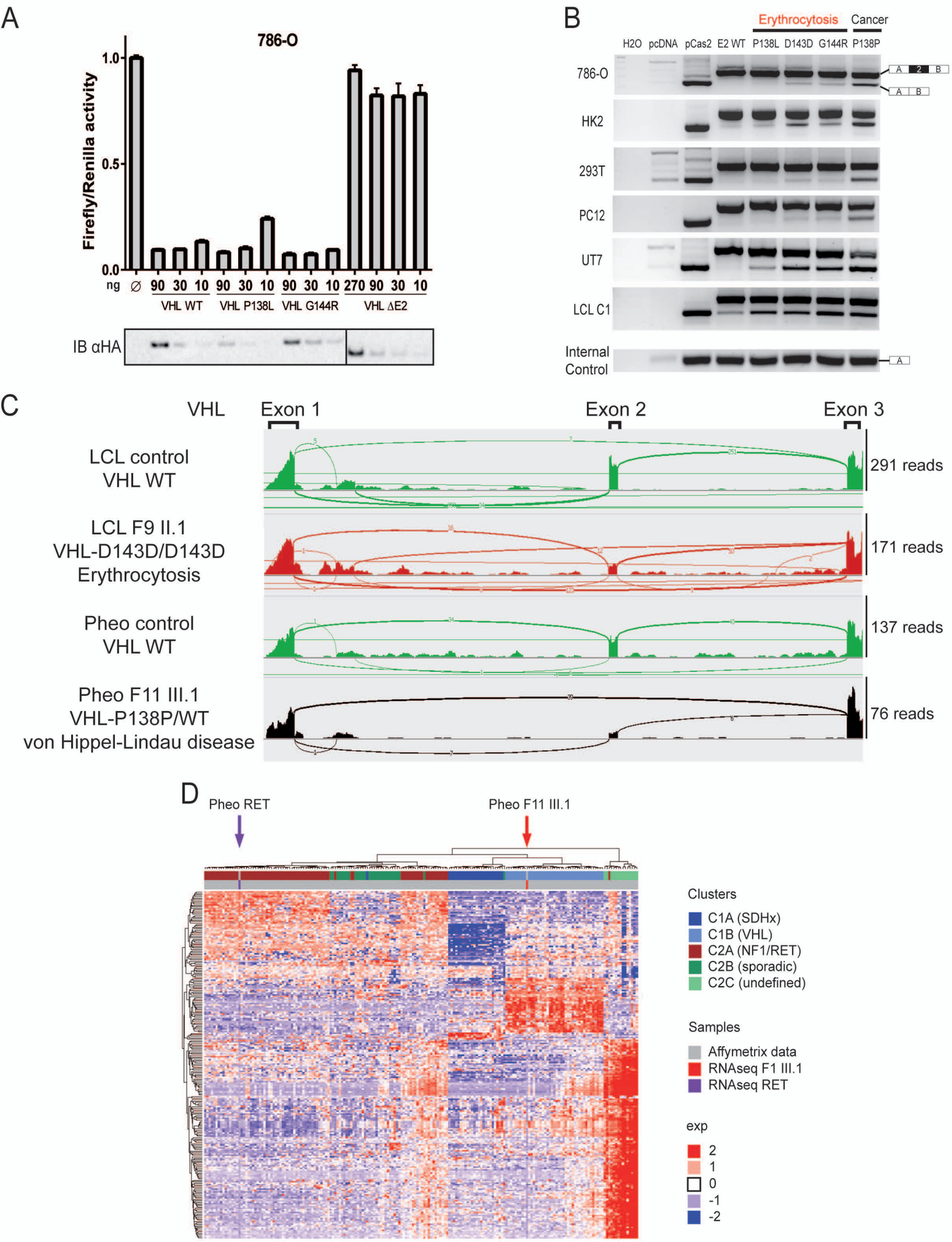

A
Family 9
I
1
2
II
1
2
Family 10
I
1
2
II
1
2
VHL c.429C>T, p.Asp143Asp
Erythrocytosis
c.429C>T
Germline DNA
WT
F10 I.1
F10 II.1
cDNA
F10 I.1
F10 I.2
B
Family 11
I
1
2
II
1
2
3
III
1
c.414A>G
Germline DNA
WT (II.1)
I.1
Tumor DNA
III.1
Family 12
I
II
III
VHL : c.414A>G, p.Pro138Pro
Paraganglioma
Pheochromocytoma
CNS Hemangioblastoma
Renal cell carcinoma
C
D143D/D143D
D143D/WT
D143D/WT
D143D/WT
D143D/D143D
WT
F9 II.1
F10 I.1
F10 I.2
F10 II.2
F10 II.1
VHL
1 2 3
1 3
LCL
P138P/WT
P138P/WT
WT/WT
P138P/WT
P138P/WT
F11 III.1
F11 II.2
F11 II.1
F11 I.1
WT
F11 III.1
WT
WT
VHL
1 2 3
*
1 3
Leukocytes cDNA
Tumor cDNA
D
E1E3
Relative gene expression
20
15
10
5
0
**
***
C1 C2 F9 I.1 F10 I.1 F10 I.2 F10 II.2 F9 II.1 F10 II.1 F11 III.1 F11 II.1 F11 III.1 F11 II.2 F11 I.1
WT
D143D/WT
D143D/D143D
P138P/WT
WT
P138P/WT
LCL
Leukocytes cDNA
E1E2
Relative gene expression
2.5
2.0
1.5
1.0
0.5
0
**
C1 C2 F9 I.1 F10 I.1 F10 I.2 F10 II.2 F9 II.1 F10 II.1 F11 III.1 F11 II.1 F11 III.1 F11 II.2 F11 I.1
WT
D143D/WT
D143D/D143D
P138P/WT
WT
P138P/WT
LCL
Leukocytes cDNA
Wild Type
Healthy Heterozygous
Homozygous with Erythrocytosis
Heterozygous with von Hippel-Lindau disease
E
G1
F9 I.1
F10 II.2
F10 I.1
F10 I.2
F9 II.1
F10 II.1
F11 III.1
HeLa
pVHL213
pVHL172
pVHL160
β-Tubulin
GAPDH
VHL 213
VHL 172
VHL 160
%C1 normalized to Tubulin
100
50
0
C1 F9 I.1 F10 II.2 F10 I.1 F9 II.1 F10 II.1 F11 III.1

Figure 6


A
786-O
Firefly/Renilla activity
1.0
0.5
0.0
ng ∅ 90 30 10 90 30 10 90 30 10 270 90 30 10
VHL WT
VHL P138L
VHL G144R
VHL ΔE2
IB αHA
B
Erythrocytosis
Cancer
H2O pcDNA pCas2 E2 WT P138L D143D G144R P138P
786-O
HK2
293T
PC12
UT7
LCL C1
Internal Control
A 2 B
A B
A
C
VHL
Exon 1
Exon 2
Exon 3
LCL control
VHL WT
291 reads
LCL F9 II.1
VHL-D143D/D143D
Erythrocytosis
171 reads
Pheo control
VHL WT
137 reads
Pheo F11 III.1
VHL-P138P/WT
von Hippel-Lindau disease
76 reads
D
Pheo RET
Pheo F11 III.1
Clusters
C1A (SDHx)
C1B (VHL)
C2A (NF1/RET)
C2B (sporadic)
C2C (undefined)
Samples
Affymetrix data
RNAseq F1 III.1
RNAseq RET
exp
2
1
0
-1
-2