Elongin C (*ELOC/TCEB1*) associated von Hippel-Lindau disease

Avgi Andreou,[1] Bryndis Yngvadottir,[1] Laia Bassaganyas,[1] Graeme Clarke,[1,2] Ezequiel Martin,[1,2] James Whitworth,[1] Alex J Cornish,[3] Genomics England Research Consortium, Richard S Houlston,[3] Philip Rich,[4] Catherine Egan,[5] Shirley V Hodgson,[6] Anne Y Warren,[7] Katie Snape,[8] Eamonn R Maher[1,*]

[*]Corresponding author

1. Department of Medical Genetics, University of Cambridge, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK
2. Stratified Medicine Core Laboratory NGS Hub, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK
3. Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK.
4. Department of Neuroradiology, St George's University Hospitals NHS Foundation Trust, London, SW17 0QT, UK
5. Moorfields Eye Hospital NHS Foundation Trust, London, EC1V 2PD, UK
6. Department of Medical Genetics, St. George's University of London, London, SW17 0QT, UK
7. Department of Histopathology, Cambridge University NHS Foundation Trust, Cambridge CB2 OQQ, UK

8. Department of Clinical Genetics, St George's University Hospitals NHS Foundation Trust and St George's University of London, London, SW17 0QT, UK

Corresponding author:

Professor Eamonn R Maher, Department of Medical Genetics, University of Cambridge, Box 238, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK, Tel +44 01223 746715, Fax +44 01223 746777

erm1000@medschl.cam.ac.uk.

Abstract

Background: Around 95% of patients with clinical features diagnostic of Von Hippel-Lindau disease (VHL) have a detectable inactivating germline variant in *VHL*. The VHL protein (pVHL) functions as part of the VCB-CR complex which plays a key role in oxygen sensing and degradation of hypoxia inducible factors. To date, only variants in *VHL* has been shown to cause VHL disease.

Materials and Methods: We undertook trio analysis by Whole-exome sequencing (WES) in a proband with VHL disease but without a detectable *VHL* mutation. Molecular studies were also performed on paired DNA extracted from the proband's kidney tumour and blood and bioinformatics analysis of sporadic renal cell carcinoma data set was undertaken.

Results: A *de novo* pathogenic variant in *ELOC* (NM_005648.4:c.236A>G [p.Tyr79Cys]) gene was identified in the proband. *ELOC* encodes elongin C, a key component [C] of the VCB-CR complex. The p.Tyr79Cys substitution is a mutational hotspot in sporadic VHL-competent renal

cell carcinoma (RCC) and has previously been shown to mimic the effects of pVHL deficiency on hypoxic signalling. Analysis of a RCC from the proband showed similar findings to that in somatically *ELOC* mutated RCC (expression of hypoxia responsive proteins, no somatic *VHL* variants and chromosome 8 loss).

Conclusions: These findings are consistent with pathogenic *ELOC* variants being a novel cause for VHL disease and suggest that genetic testing for *ELOC* variants should be performed in individuals with suspected VHL disease with no detectable *VHL* variant.

Introduction

Genetic studies of rare familial cancer syndromes have provided important insights into cancer biology and mechanisms of human disease. This is exemplified by von Hippel-Lindau (VHL) disease/syndrome (VHL) (MIM:193300), an autosomal dominant multisystem cancer predisposition disorder characterised by predisposition to retinal and central nervous system haemangioblastomas, clear cell renal cell carcinoma (ccRCC), pheochromocytoma/paraganglioma (PPGL), non-secretory pancreatic neuroendocrine tumours and endolymphatic sac tumours (ELSTs) (1, 2). The cardinal features diagnostic of VHL disease were defined in the early 1960s: two or more retinal or central nervous system haemangioblastomas or a hemangioblastoma and ccRCC or phaeochromocytoma or a positive family history of VHL disease and a single tumour (hemangioblastoma, ccRCC or phaeochromocytoma) (3).

The incidence of VHL disease is ~1 in 36,000 live births (4) and following clinical descriptions of large affected families, genetic linkage studies mapped a gene to chromosome 3p25-26 with

no evidence of locus heterogeneity (5). The von Hippel-Lindau tumour suppressor gene (*VHL* [MIM: 608537]) was identified in 1993 (6) and over 1,000 pathogenic germline and somatic *VHL* variants have now been described (7). Around 95% of individuals with clinical features diagnostic of VHL disease have an inactivating germline *VHL* variant detectable by standard molecular genetic testing. Recently some "*VHL* mutation-negative" cases have been demonstrated to have mosaicism, promoter region variants or an intronic *VHL* mutation but no other genes have been reported to cause VHL disease (8, 9). Germline *VHL* pathogenic variants may also be detected in individuals with a clinical diagnosis of VHL disease (e.g. apparently sporadic haemangioblastoma or with familial phaeochromocytoma/paraganglioma) and rare biallelic missense variants have been shown to cause autosomal recessive polycythaemia (10, 11).

Tumours from individuals with VHL disease show somatic inactivation of the wild-type allele consistent with the Knudsen two-hit model of tumourigenesis (12). Furthermore, in sporadic ccRCC and haemangioblastomas somatic biallelic inactivation of the *VHL* tumour suppressor gene (TSG) occurs as a critical and early event in tumourigenesis (13, 14). The identification of the *VHL* TSG led to the discovery of its role in the pathogenesis of sporadic ccRCC and the fundamental role of the gene product in cellular oxygen sensing (1, 15). Tumours with *VHL* TSG inactivation are highly vascular and demonstrate hypoxia-independent activation of the hypoxic gene response pathway targets, with overexpression of angiogenic (e.g. VEGF and PDGFB) and oncogenic (Cyclin D1) factors (16, 17). The VHL protein (pVHL) has a critical role in regulating the expression of the α-subunits of the hypoxia inducible transcription factors, HIF-1 and HIF-2, that regulate the cellular response to hypoxia such that pVHL functions as the target binding component of an E3 ubiquitin ligase complex comprising pVHL, elongin C, elongin B, cullin 2

and RBX1 (VCB-CR complex) (15, 18, 19). To date, germline mutations in non-VHL components of the VCB-CR complex have not been reported. Herein, we describe the association of VHL disease-like phenotype with a pathogenic variant in the *ELOC* gene encoding the elongin C protein, which binds to pVHL.

Results

Case report

A 37-year-old female of Northern European origin presented with two left retinal haemangioblastomas that were treated by laser treatment (**Figure 1A-C**). Two years later she developed an RCC and cyst of the right kidney that were treated by partial right nephrectomy (**Figure 1D-E**). At age 47 years a further RCC was detected in the left kidney and treated by cryoablation. A spinal haemangioblastoma was removed at age 52 years (**Figure 1F)** and a haemangioblastoma at the cervicomedullary junction remains under surveillance (**Figure 1G**). Before developing features indicative of VHL disease, she had presented with Henoch-Schonlein purpura aged 23 years and underwent unilateral parathyroidectomy for two parathyroid adenomas at age 28 years. Family history was unremarkable with both parents, three siblings and three children not reporting any features of VHL disease (**Figure 1H**).

Pathological examination of the right partial nephrectomy sample was consistent with ccRCC, staged as Fuhrman grade 2 (g2) pT1 NX. Sections from the right renal cyst showed fibrous

walled cysts lined by regular clear epithelial cells, with small nuclei (**Figure 2A-C**). The tumour was diffusely positive for AE1/3, CA-IX (**Figure 2D**) and Vimentin and showed focal positivity for CK7 and RCC (weak) on immunohistochemistry. Fumarate hydratase staining was retained and 2SC staining was negative. The appearances were considered typical of those seen in von Hippel-Lindau disease. In addition, the presence of a leiomyomatous stroma and occasional branched tubular structures lined by cells with voluminous cytoplasm, features of RCC with somatic *ELOC* variants, were noted focally in the RCC (**Figure 2B)** (20, 21). The spinal tumour showed features of a haemangioblastoma and was positive for inhibin, Vimentin, S100 and CA-IX expression but negative for AE1/3, Pax-8 and CK7. CD31 and CD34 stains highlighted a network of vascular structures (**Figure 2E-F)**.

Routine diagnostic testing by Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA) for a germline *VHL* variant showed no abnormality and after informed written consent, the proband and her parents underwent research testing. Whole exome sequencing (WES) and whole genome sequencing (WGS) was performed. No candidate *VHL* variants were detected in the proband but trio analysis identified 16 rare variants (gnomAD maximum allele frequency $\leq$ 0.5%) (**Supplementary Table 1**) that were not detectable in either parent. A *de novo* missense variant in *ELOC* (NC_000008.11:g.73946733A>G; NM_005648.4:c.236A>G [p.Tyr79Cys]) was identified. Direct (Sanger) sequencing validated the presence of the *de novo ELOC* variant in the proband (**Figure 3A**). Tyrosine at codon 79 (Y79) is evolutionary conserved across vertebrates and invertebrates (**Figure 3B**) (22) and is located in the tetramerisation domain of the *ELOC* gene (**Figure 3C**) (23). Elongin C Tyr79 residue is known to form a critical hydrogen bond with the Pro154 residue within the pVHL

alpha domain (24-26) (**Figure 3D**). NM_005648.4(ELOC):c.236A>G(p.Tyr79Cys) was not seen in 76,156 genomes catalogued by gnomAD (v3.1). Deep intronic and promoter variants, described previously in VHL disease or erythrocytosis, were excluded from the proband (**Supplementary Table 2**).

Microarray-based comparative genomic hybridization (aCGH) performed on the DNA pair extracted from the proband's right RCC and blood showed evidence of monosomy for chromosomes 8, 21 and 22 and no somatic alterations commonly seen in ccRCC (i.e. deletion of 3p, 9p, 14q or 5q gain) (27) (**Figure 4A**). Paired WES for tumour/blood DNA was analysed for copy number variants and SNVs/indels and was consistent with loss of chromosomes 8, 21 and 22 and no evidence of a somatic *VHL* mutation (**Figure 4B**). The c.236A>G *ELOC* variant was present in 35% and 46% (46/130 and 33/72) of reads in blood and tumour DNAs respectively. The allele read counts for variant and wild-type alleles in blood being biased towards the wild-type suggesting mosaicism for the c.236A>G variant (**Supplementary Figure 1**).

In view of the parathyroid adenomas diagnosed in the proband at an early age, we analysed for the presence of any variants in genes predisposing to any endocrine neoplasia syndromes. No pathogenic/likely pathogenic SNVs, CNVs or SVs were found in *AIP*, *CDC73*, *CDKN1B*, *MEN1* and *RET*.

ELOC c.236A>G (p.Tyr79Cys) variation in human disease

ELOC c.236A>G (p.Tyr79Cys) was originally described as a somatic variant in six RCCs without VHL inactivation (24), in three cases within The Cancer Genome Atlas (TCGA) (28) and subsequently in 5 cases from the Memorial Sloan Kettering Cancer Centre (MSKCC) cohort

(details of specific amino acid substitution at residue 79 not available) (21) (**Supplementary Table 4**). To further explore the role of the *ELOC* variants in *VHL*-independent renal tumorigenesis we searched for additional examples of germline and somatic *ELOC* variants. Germline *ELOC* variants were sought (by Sanger sequencing or WES) in 91 individuals recruited inhouse with either a VHL-like phenotype (n=91) which was defined as multiple VHL-related tumours or a single VHL-related tumour plus a family history of a VHL-related tumour. None of the 91 individuals had evidence of the *ELOC* c.236A>G (p.Tyr79Cys) variant, or other *ELOC* pathogenic variants. The *ELOC* c.236A>G (p.Tyr79Cys) variant was also absent from the germline of 78,195 participants in the 100,000 Genomes Project (29), including 1,336 individuals with RCC. To further investigate the role of somatic *ELOC* mutations and in particular c.236A>G in *VHL*-independent renal tumourigenesis we interrogated the 100,000 Genomes Project dataset (29) and identified 8 of 1,336 RCC with a candidate pathogenic *ELOC* somatic variant. Four of the 8 RCC tumours had a somatic NM_005648.4(*ELOC*):c.236A>G (p.Tyr79Cys) variant (Cases 1-4). Three of the 8 RCC had a non-codon 79 missense *ELOC* variant (Case 5: NM_005648.4:c.274G>A (p.Glu92Lys), Case 6: NM_005648.4:c.74A>T (p.Asp25Val), Case 7: NM_005648.4:c.311T>A (p.Leu104Gln)) and the remaining RCC (Case 8) harboured a somatic *ELOC* in-frame deletion (NM_005648.4:c.261_272del (p.Thr88_Pro91del)) (**Supplementary Table 5**). All eight of the RCCs with a candidate somatic pathogenic *ELOC* variant demonstrated chromosome 8 loss and were *VHL* mutation negative (**Figure 5** and **Supplementary Table 5)**.

Discussion

We report a germline *de novo* missense substitution c.236A>G (p.Tyr79Cys) in the Elongin C gene (*ELOC*), previously known as *TCEB1*, in a female who satisfied clinical diagnostic criteria for VHL disease but who did not have a detectable *VHL* mutation. In particular, there was no evidence for intronic or promoter region VHL mutation (**Supplementary Table 2)** and no evidence for a mosaic *VHL* mutation after analysis of blood and tumour DNA. To our knowledge the c.236A>G (p.Tyr79Cys) *ELOC* missense substitution has not been detected as a germline variant previously (20, 21, 24, 28). To date, 20 somatic *ELOC*-mutated RCCs have been reported (21, 28) (**Supplementary Table 4**). Our analysis has confirmed the finding of recurrent somatic p.Tyr79Cys substitutions as a hotspot mutational event in sporadic RCC but also identified additional candidate pathogenic *ELOC* somatic variants that were mostly also missense substitutions. Consistent with DiNatale (21) we found evidence of chromosome 8 deletions in the *ELOC*-mutated sporadic RCCs and also in the RCC associated with a germline *ELOC* mutation.

pVHL has two critical functional domains. Under normoxic conditions the β-domain binds to two conserved proline residues within the oxygen-dependent degradation domains of the α-subunits of the HIF-1 and HIF-2 transcription factors and targets them for ubiquitin-mediated proteolysis (15, 18, 19). pVHL deficiency or hypoxia results in HIF-1 and HIF-2 being stably expressed and activating hypoxic-gene response pathways (15-17). The second critical pVHL domain, the alpha-domain (residues 155-192) (18), interacts with other components of the VCB-CR complex by binding directly to elongin C (18). Germline or somatic *VHL* mutations that disrupt pVHL binding to elongin C result in HIF-stabilisation and activation of hypoxic-gene

response pathways. Within the pVHL alpha domain, the Pro154 residue forms a critical hydrogen bond with the Elongin C Tyr79 residue (24) (**Figure 3D**). Previously experiments in human HEK 293T cells have shown that whilst ELOC-wild type co-precipitates with pVHL and CUL2, this is greatly reduced for mutant ELOC-Tyr79 (24). Furthermore, ELOC-Tyr79Cys leads to accumulation of HIF-1α and HIF-2α as compared to tumours without *ELOC* or *VHL* mutations (24). These studies are compatible with our observation of a VHL disease phenotype in an individual with a germline *ELOC* p.Tyr79Cys variant. The previously reported *in vitro* studies are consistent with p.Tyr79Cys functioning as a loss of function variant and we and others have found that chromosome 8 loss is a feature of *ELOC* mutated RCC (21). We confirmed this finding in p.Tyr79Cys mutated RCC and also identified other candidate somatic *ELOC* mutations in sporadic RCC that were also associated with chromosome 8 loss. It is clear that *ELOC* p.Tyr79Cys is a mutation hotspot but the explanation for this is currently unclear. One possibility is that *ELOC* p.Tyr79 substitutions might disrupt pVHL related functions of the VBC-CR complex whilst leaving other functions (e.g.RNA polymerase II elongation) intact and/or there is a requirement for a specific level of ELOC function to promote tumourigenesis, similar to the "just-right" signalling model proposed for APC tumour suppressor function (30). Though inactivation of the *VHL* and *ELOC* tumour suppressor genes will both result in dysregulation of hypoxic gene response pathways and other HIF-independent pVHL functions, there will be differences in the effects on other cellular pathways and this might result in additional or varied presentation of clinical features within patients with a germline *ELOC* mutation. For example, Elongin C is known to link SOCS proteins, which are negative feedback inhibitors of cytokine and growth factor-induced signal transduction, to the proteasome and target them for degradation (31). SOCS1 was shown to interact with Elongin B, Elongin C and

Cul2 and to target JAK2, Vav, IRS1, and IRS2 for ubiquitylation and proteasomal degradation (32-34). SOCS2 forms a complex with elongin B and elongin C (SOCS2–elongin C–elongin B complex) which acts as an E3 ubiquitin ligase similarly to the VCB-CR complex showing a shared mechanism of ubiquitination between these cullin-dependent E3 ligases (35). Phosphopeptide substrates derived from the Growth hormone receptor (GHR) and the erythropoietin receptor are recognised targets of SOCS2 (31, 35, 36). Therefore, the reason for specific mutation hotspot (and absence of truncating mutations) in *ELOC* might relate to the fact that these alterations affect the interaction with VHL but not with other proteins such as SOCS1/2 (or other pVHL-unrelated ELOC functions). The overlapping but distinct functional effects of pVHL and ELOC inactivation appear to be reflected in differing patterns of somatic copy number events and mutations in *VHL* and *ELOC*-mutated RCC. RCC with germline and somatic *VHL* TSGs mutations have a high frequency of somatic chromosome 3p deletions affecting both *VHL* and other chromosome 3p TSGs such as *BAP1*, *PBRM1* and *SETD2* (37*)*. In contrast, *ELOC*-mutated RCC have a high frequency of chromosome 8 deletions but chromosome 3p deletions are infrequent. Whilst these patterns of chromosomal loss reflect the occurrence of "second hit" deletion events in the two categories of RCC, it is interesting that there are not more similarities in the somatic mutation patterns outside of *VHL/ELOC*. These differences in tumour evolution may lead to differences in tumour growth patterns (e.g. *VHL*-related RCC may show gain of chromomere 8q, including amplification of MYC, which has been associated with a more aggressive tumour phenotype (37-39) and a more indolent course of ELOC-mutated RCC has been suggested previously (20). In addition, differences in the copy number profiles and pathological appearances of *VHL* and *ELOC*-mutated RCCs could be

utilised to differentiate between *ELOC*-associated VHL disease and classical *VHL*-related VHL disease.

Given the effect of inactivation of ELOC on the function of the VBC-CR, a VHL phenotype being associated with germline p.Tyr79Cys is perhaps not unexpected. However, at this stage, it is unclear to whether germline *ELOC* mutations will solely mimic VHL disease or be associated with other clinical phenotypes. The presence of parathyroid adenomas at a young age is not a known feature of VHL disease and this may be coincidental in our case. The haemangioblastoma and two RCCs from the proband showed typical features of those associated with germline *VHL* mutations. In addition, on pathology review, the presence of a leiomyomatous stroma and occasional branched tubular structures lined by cells with voluminous cytoplasm, features of RCC with somatic *ELOC* mutations, were noted focally in the RCC (**Figure 2A-C**) (20, 21). Currently we would suggest that testing of *ELOC* should be performed in patients with suspected VHL disease but without an identifiable *VHL* mutation. The clinical course of *ELOC*-mutated RCC is variable (21), however, based on existing data we would propose that individuals with a pathogenic germline variant should be managed as per VHL disease (40). Whilst the emphasis of VHL management is primarily early diagnosis and treatment, the mechanistic similarities between VHL and ELOC p.Tyr79Cys-associated tumours suggests that treatment with HIF-2α antagonists such as bezultifan may be a therapeutic option for *ELOC* mutated tumours (41).

Materials and Methods

Patient ascertainment

All subjects gave written informed consent for genetic studies; the investigations were approved by the South Birmingham Research Ethics committee and were conducted in accordance with the Declaration of Helsinki. Participants from the participants from 100,000 Genomes Project were consented as per the 100,000 Genomes Project protocol (29).

Germline sequencing

DNA was extracted from peripheral blood samples of patients according to standard protocols. Whole exome sequencing (WES) was performed in-house using Illumina DNA Prep with Enrichment (formerly named Nextera Flex for Enrichment) (42) on Illumina's HiSeq 4000 platform with 150bp paired end reads. Raw Illumina BCL files were demultiplexed and converted to FASTQ format using Illumina's bcl2fastq version 2.19. All sample pairs were aligned to the hg38 version of the reference human genome using BWA-0.7.15 as previously described (43). The generated SAM file was compressed into a BAM file and sorted by genomic position using SAMtools version 1.3.1 (44). The sorted BAM files were subject to Base Quality Score Recalibration and Indel Realignment followed by variant calling using the Haplotype Caller algorithm as specified in the Genome-Analysis Toolkit (GATK) version 3.8 best practices (45-47). VCF files were filtered for a minimum depth of 20 reads and a Genotype Quality (GQ) of 30 using VCFtools version 0.1.15 (48). VCF files were annotated with ANNOVAR (49).

Trio analysis in the proband and parents identified 126 exonic *de novo* variants in the proband. After filtering for rare exonic *de novo* variants (gnomAD MAX AF ≤ 0.5%), 16 exonic *de novo* variants were further analysed (**Supplementary Table 1**).

Deep intronic and promoter *VHL* variants, described previously in VHL disease or erythrocytosis, were excluded from WGS data available for the proband (**Supplementary Table 2**). WGS for the 100,000 Genomes Project participants was performed according to the 100,000 Genomes Project protocol (29). Whole genome 150bp paired-end TruSeq PCR–free libraries were sequenced on a single lane using Illumina (San Diego, USA) HiSeq X technology and uniformly processed on the Illumina North Star Version 4 Whole Genome Sequencing Workflow (NSV4, version 2.6.53.23). Raw sequencing data was aligned to the NCBI GRCh38 assembly (with decoys) using iSAAC Aligner (version 03.16.02.19) and small germline variants were called using Starling (version 2.4.7). VCF files from WGS were annotated using VEP version 99 (50).

SNVs, CNVs and SVs in *AIP*, *CDC73*, *CDKN1B*, *MEN1* and *RET* were also excluded from WES data available for the proband. SVs and CNVs were annotated with bedtools (51). SVs were called with Delly v0.8.1 (52) and CNVs were called with GATK4 version 4.1.4.0 best practices (45, 53).

Targeted Sanger sequencing (n=25) and exome sequencing (n=66) on DNA extracted from blood in cohorts of patients previously examined for germline mutations in *VHL* without a mutation were performed to determine the likely frequency of germline variants in *ELOC.* Previous clinical testing using Sanger sequencing analysis, multiplex-ligation dependent probe

amplification (MLPA) and methylation analysis of *VHL* had not identified a pathogenic single-nucleotide variant (SNV) or copy number variant (CNV) in any of the samples.

Tumour studies

Targeted tumour sequencing was performed on the DNA pair extracted from the proband's macro-dissected formalin-fixed paraffin-embedded (FFPE) right kidney tumour specimen and DNA extracted from blood (germline). Library preparation was performed using Illumina DNA Prep with Enrichment (42) on Illumina's HiSeq 4000 platform. Paired WES for tumour/germline DNA was analysed for copy number variants and SNVs/indels. SNV and SV analysis was performed as described above. Microarray-based comparative genomic hybridization (aCGH) was performed on the paired tumour/germline DNA samples using Illumina's 750K SNP genotyping array (54).

Sanger Sequencing of germline samples

Sanger sequencing was performed using standard techniques, as per the Eurofins protocol (55). The following primer pairs were used:

ELOC exon 1: 5′-ccacccctagatggcttgaa-3′, 3'-tgcaaacgacgctttatagtc-5',
ELOC exon 2: 5′-gtgggtggatcatgaggtca-3′, 3'-cagtttcttctgcaaaagctgt-5',
ELOC exon 3: 5′-tttgagaccagcctgaccaa-3′, 3'-agctgtacctagtaaccttcca-5',
ELOC exon 4: 5′-aaaattagccggtcgtggtg-3′, 3'-cttctgcaaaagctgtacctagt-5'.

The following conditions were used: (i) 95°C for 30 seconds, (ii) 60°C for 30 seconds, (iii) 72°C for 45 seconds, (iv) repeat (i) to (iii) 30 times, (v) incubate at 72°C for 10 minutes.

Acknowledgments

This work was supported by the European Research Council (Advanced Researcher Award (ERM)), NIHR (Senior Investigator Award (ERM) and Cambridge NIHR Biomedical Research Centre (ERM, AYW)), Cancer Research UK Cambridge Cancer Centre (SSAG/085 to AA, ERM, AYW) and VHL UK/Ireland. The University of Cambridge has received salary support in respect of ERM from the NHS in the East of England through the Clinical Academic Reserve. RSH. and AC are supported by a grant from Cancer Research UK (C1298/A8362). The views expressed are those of the authors and not necessarily those of the NHS or Department of Health.

Part of this research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support. We acknowledge support from the NIHR UK Rare Genetic Disease Research Consortium.

Genomics England Research Consortium authors: Ambrose, J. C. ; Arumugam, P. ; Bevers, R. ; Bleda, M. ; Boardman-Pretty, F. ; Boustred, C. R. ; Brittain, H. ; Caulfield, M. J. ; Chan, G. C. ; Fowler, T. ; Giess A. ; Hamblin, A. ; Henderson, S. ; Hubbard, T. J. P. ; Jackson, R. ; Jones, L. J. ; Kasperaviciute, D. ; Kayikci, M. ; Kousathanas, A. ; Lahnstein, L. ; Leigh, S. E. A. ; Leong, I. U. S. ; Lopez, F. J. ; Maleady-Crowe, F. ; McEntagart, M. ; Minneci F. ; Moutsianas, L. ; Mueller, M. ; Murugaesu, N. ; Need, A. C. ; O'Donovan P. ; Odhams, C. A. ; Patch, C. ; Perez-

Gil, D. ; Pereira, M. B. ; Pullinger, J. ; Rahim, T. ; Rendon, A. ; Rogers, T. ; Savage, K. ; Sawant, K. ; Scott, R. H. ; Siddiq, A. ; Sieghart, A. ; Smith, S. C. ; Sosinsky, A. ; Stuckey, A. ; Tanguy M. ; Taylor Tavares, A. L. ; Thomas, E. R. A. ; Thompson, S. R. ; Tucci, A. ; Welland, M. J. ; Williams, E. ; Witkowska, K. ; Wood, S. M.

Conflict of interest statement

ERM declares invited speaker fees from MSD. The other authors declare no conflict of interest.

References

1 Gossage, L., Eisen, T. and Maher, E.R. (2015) VHL, the story of a tumour suppressor gene. *Nat. Rev. Cancer*, **15**, 55-64.

2 Maher, E.R., Yates, J.R., Harries, R., Benjamin, C., Harris, R., Moore, A.T. and Ferguson-Smith, M.A. (1990) Clinical features and natural history of von Hippel-Lindau disease. *Q. J. Med*, **77**, 1151-1163.

3 Melmon, K.L. and Rosen, S.W. (1964) Lindau's Disease. Review of the Literature and Study of a Large Kindred. *Am. J. Med.*, **36**, 595-617.

4 Maher, E.R., Iselius, L., Yates, J.R., Littler, M., Benjamin, C., Harris, R., Sampson, J., Williams, A., Ferguson-Smith, M.A. and Morton, N. (1991) Von Hippel-Lindau disease: a genetic study. *J. Med. Genet.*, **28**, 443-447.

5 Seizinger, B.R., Rouleau, G.A., Ozelius, L.J., Lane, A.H., Farmer, G.E., Lamiell, J.M., Haines, J., Yuen, J.W., Collins, D., Majoor-Krakauer, D. *et al.* (1988) Von Hippel-Lindau disease maps to the region of chromosome 3 associated with renal cell carcinoma. *Nature*, **332**, 268-269.

6 Latif, F., Tory, K., Gnarra, J., Yao, M., Duh, F.M., Orcutt, M.L., Stackhouse, T., Kuzmin, I., Modi, W., Geil, L. *et al.* (1993) Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science*, **260**, 1317-1320.

7 Tabaro, F., Minervini, G., Sundus, F., Quaglia, F., Leonardi, E., Piovesan, D. and Tosatto, S.C. (2016) VHLdb: A database of von Hippel-Lindau protein interactors and mutations. *Sci. Rep.*, **6**, 31128.

8 Coppin, L., Grutzmacher, C., Crepin, M., Destailleur, E., Giraud, S., Cardot-Bauters, C., Porchet, N. and Pigny, P. (2014) VHL mosaicism can be detected by clinical next-generation

sequencing and is not restricted to patients with a mild phenotype. *Eur. J. Hum. Genet.*, **22**, 1149-1152.

9 Lenglet, M., Robriquet, F., Schwarz, K., Camps, C., Couturier, A., Hoogewijs, D., Buffet, A., Knight, S.J.L., Gad, S., Couve, S. *et al.* (2018) Identification of a new VHL exon and complex splicing alterations in familial erythrocytosis or von Hippel-Lindau disease. *Blood*, **132**, 469-483.

10 Woodward, E.R., Eng, C., McMahon, R., Voutilainen, R., Affara, N.A., Ponder, B.A. and Maher, E.R. (1997) Genetic predisposition to phaeochromocytoma: analysis of candidate genes GDNF, RET and VHL. *Hum. Mol. Genet.*, **6**, 1051-1056.

11 Pastore, Y., Jedlickova, K., Guan, Y., Liu, E., Fahner, J., Hasle, H., Prchal, J.F. and Prchal, J.T. (2003) Mutations of von Hippel-Lindau tumor-suppressor gene and congenital polycythemia. *Am. J. Hum. Genet.*, **73**, 412-419.

12 Prowse, A.H., Webster, A.R., Richards, F.M., Richard, S., Olschwang, S., Resche, F., Affara, N.A. and Maher, E.R. (1997) Somatic inactivation of the VHL gene in Von Hippel-Lindau disease tumors. *Am. J. Hum. Genet.*, **60**, 765-771.

13 Gnarra, J.R., Tory, K., Weng, Y., Schmidt, L., Wei, M.H., Li, H., Latif, F., Liu, S., Chen, F., Duh, F.M. *et al.* (1994) Mutations of the VHL tumour suppressor gene in renal carcinoma. *Na.t Genet.*, **7**, 85-90.

14 Foster, K., Prowse, A., van den Berg, A., Fleming, S., Hulsbeek, M.M., Crossey, P.A., Richards, F.M., Cairns, P., Affara, N.A., Ferguson-Smith, M.A. *et al.* (1994) Somatic mutations of the von Hippel-Lindau disease tumour suppressor gene in non-familial clear cell renal carcinoma. *Hum. Mol. Genet.*, **3**, 2169-2173.

15 Maxwell, P.H., Wiesener, M.S., Chang, G.W., Clifford, S.C., Vaux, E.C., Cockman, M.E., Wykoff, C.C., Pugh, C.W., Maher, E.R. and Ratcliffe, P.J. (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature*, **399**, 271-275.

16 Iliopoulos, O., Levy, A.P., Jiang, C., Kaelin, W.G., Jr. and Goldberg, M.A. (1996) Negative regulation of hypoxia-inducible genes by the von Hippel-Lindau protein. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 10595-10599.

17 Bindra, R.S., Vasselli, J.R., Stearman, R., Linehan, W.M. and Klausner, R.D. (2002) VHL-mediated hypoxia regulation of cyclin D1 in renal carcinoma cells. *Cancer Res.*, **62**, 3014-3019.

18 Stebbins, C.E., Kaelin, W.G., Jr. and Pavletich, N.P. (1999) Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science*, **284**, 455-461.

19 Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J.M., Lane, W.S. and Kaelin, W.G., Jr. (2001) HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science*, **292**, 464-468.

20 Hakimi, A.A., Tickoo, S.K., Jacobsen, A., Sarungbam, J., Sfakianos, J.P., Sato, Y., Morikawa, T., Kume, H., Fukayama, M., Homma, Y. *et al.* (2015) TCEB1-mutated renal cell carcinoma: a distinct genomic and morphological subtype. *Mod. Pathol.*, **28**, 845-853.

21 DiNatale, R.G., Gorelick, A.N., Makarov, V., Blum, K.A., Silagy, A.W., Freeman, B., Chowell, D., Marcon, J., Mano, R., Sanchez, A. *et al.* (2021) Putative Drivers of Aggressiveness in TCEB1-mutant Renal Cell Carcinoma: An Emerging Entity with Variable Clinical Course. *Eur. Urol. Focus*, **7**, 381-389.

22 Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996-1006.

23 Jay, J.J. and Brouwer, C. (2016) Lollipops in the Clinic: Information Dense Mutation Plots for Precision Medicine. *PLoS One*, **11**, e0160519.

24 Sato, Y., Yoshizato, T., Shiraishi, Y., Maekawa, S., Okuno, Y., Kamura, T., Shimamura, T., Sato-Otsubo, A., Nagae, G., Suzuki, H. *et al.* (2013) Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.*, **45**, 860-867.

25 Schrodinger, L. (2010) The PyMOL Molecular Graphics System, Version 2.0. in press.

26 Nguyen, H.C., Yang, H., Fribourgh, J.L., Wolfe, L.S. and Xiong, Y. (2015) Insights into Cullin-RING E3 ubiquitin ligase recruitment: structure of the VHL-EloBC-Cul2 complex. *Structure*, **23**, 441-449.

27 Beroukhim, R., Brunet, J.P., Di Napoli, A., Mertz, K.D., Seeley, A., Pires, M.M., Linhart, D., Worrell, R.A., Moch, H., Rubin, M.A. *et al.* (2009) Patterns of gene expression and copy-number alterations in von-hippel lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer Res.*, **69**, 4674-4681.

28 Ricketts, C.J., De Cubas, A.A., Fan, H., Smith, C.C., Lang, M., Reznik, E., Bowlby, R., Gibb, E.A., Akbani, R., Beroukhim, R. *et al.* (2018) The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep.*, **23**, 313-326 e315.

29 Caulfield,M., Davies,J., Dennys,M., Elbahy,L., Fowler,T., Hill,S., Hubbard,T., Jostins,L., Maltby,N. and Mahon-Pearson,J. (2017) The 100,000 Genomes Project Protocol. 10.6084/m9.figshare.4530893.v3.

30 Albuquerque, C., Breukel, C., van der Luijt, R., Fidalgo, P., Lage, P., Slors, F.J., Leitao, C.N., Fodde, R. and Smits, R. (2002) The 'just-right' signaling model: APC somatic mutations

are selected based on a specific level of activation of the beta-catenin signaling cascade. *Hum Mol. Genet.*, **11**, 1549-1560.

31 Haan, S., Ferguson, P., Sommer, U., Hiremath, M., McVicar, D.W., Heinrich, P.C., Johnston, J.A. and Cacalano, N.A. (2003) Tyrosine phosphorylation disrupts elongin interaction and accelerates SOCS3 degradation. *J. Biol. Chem.*, **278**, 31972-31979.

32 De Sepulveda, P., Ilangumaran, S. and Rottapel, R. (2000) Suppressor of cytokine signaling-1 inhibits VAV function through protein degradation. *J. Biol. Chem.*, **275**, 14005-14008.

33 Kamizono, S., Hanada, T., Yasukawa, H., Minoguchi, S., Kato, R., Minoguchi, M., Hattori, K., Hatakeyama, S., Yada, M., Morita, S. *et al.* (2001) The SOCS box of SOCS-1 accelerates ubiquitin-dependent proteolysis of TEL-JAK2. *J. Biol. Chem.*, **276**, 12530-12538.

34 Rui, L., Yuan, M., Frantz, D., Shoelson, S. and White, M.F. (2002) SOCS-1 and SOCS-3 block insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2. *J. Biol. Chem.*, **277**, 42394-42398.

35 Bullock, A.N., Debreczeni, J.E., Edwards, A.M., Sundstrom, M. and Knapp, S. (2006) Crystal structure of the SOCS2-elongin C-elongin B complex defines a prototypical SOCS box ubiquitin ligase. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 7637-7642.

36 Eyckerman, S., Verhee, A., der Heyden, J.V., Lemmens, I., Ostade, X.V., Vandekerckhove, J. and Tavernier, J. (2001) Design and application of a cytokine-receptor-based interaction trap. *Nat. Cell. Biol.*, **3**, 1114-1119.

37 Cancer Genome Atlas Research, N. (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*, **499**, 43-49.

38 Klatte, T., Kroeger, N., Rampersaud, E.N., Birkhauser, F.D., Logan, J.E., Sonn, G., Riss, J., Rao, P.N., Kabbinavar, F.F., Belldegrun, A.S. *et al.* (2012) Gain of chromosome 8q is associated with metastases and poor survival of patients with clear cell renal cell carcinoma. *Cancer*, **118**, 5777-5782.

39 Mehrazin, R., Dulaimi, E., Uzzo, R.G., Devarjan, K., Pei, J., Smaldone, M.C., Kutikov, A., Testa, J.R. and Al-Saleem, T. (2018) The correlation between gain of chromosome 8q and survival in patients with clear and papillary renal cell carcinoma. *Ther. Adv. Urol.*, **10**, 3-10.

40 Maher, E.R., Neumann, H.P. and Richard, S. (2011) von Hippel-Lindau disease: a clinical and scientific review. *Eur. J. Hum. Genet.*, **19**, 617-623.

41 Chen, W., Hill, H., Christie, A., Kim, M.S., Holloman, E., Pavia-Jimenez, A., Homayoun, F., Ma, Y., Patel, N., Yell, P. *et al.* (2016) Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature*, **539**, 112-117.

42 Illumina. Illumina DNA Prep with Enrichment [cited 2022 Jan21] Available from: https://emea.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-flex-enrichment.html.

43 Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.

44 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

45 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a

MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297-1303.

46 DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.*, **43**, 491-498.

47 Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J. *et al.* (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, **43**, 11 10 11-11 10 33.

48 Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

49 Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.

50 McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P. and Cunningham, F. (2016) The Ensembl Variant Effect Predictor. *Genome Biol.*, **17**, 122.

51 Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841-842.

52 Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V. and Korbel, J.O. (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, **28**, i333-i339.

53 GATK team. Somatic copy number variant discovery (CNVs) [cited 2022 March10]. Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360035535892-Somatic-copy-number-variant-discovery-CNVs-

54 Illumina. Illumina Microarray Solutions [cited 2022 Jan21] Available from: https://www.illumina.com/content/dam/illumina-marketing/documents/applications/genotyping/Microarray_Solutions.pdf.

55 Eurofins. Sanger Sequencing at Eurofins Genomics [cited 2022 Jan21] Available from: https://eurofinsgenomics.eu.

56 Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M. *et al.* (2012) The life history of 21 breast cancers. *Cell*, **149**, 994-1007.

Legends to Figures

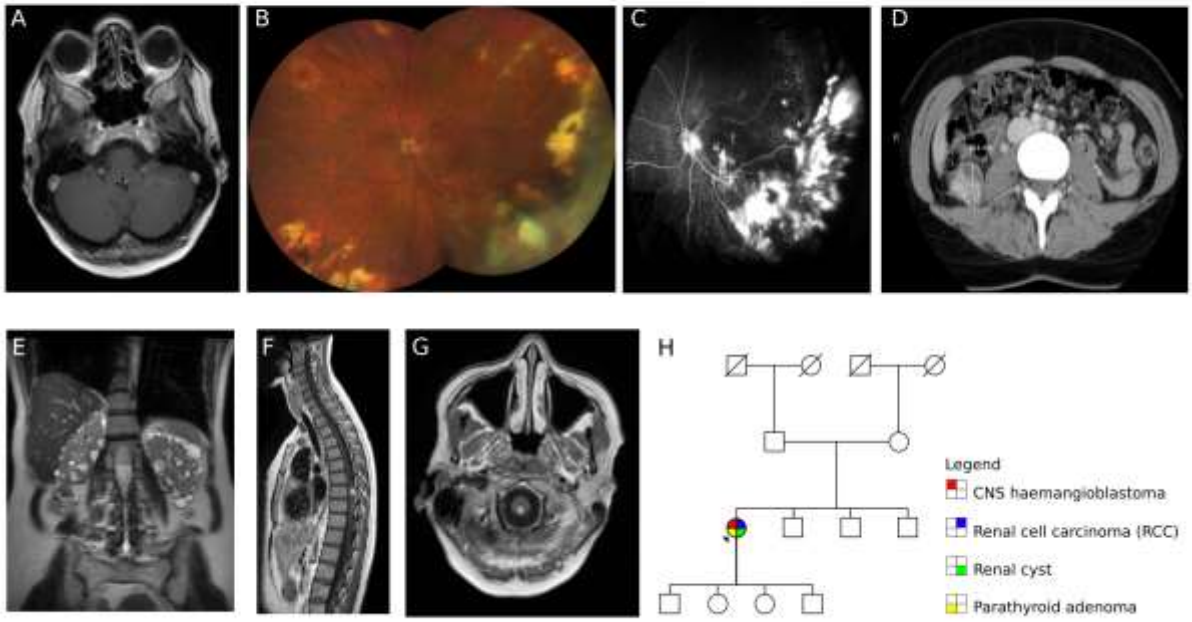

Figure 1:

A) Axial T1-weighted post contrast image through the orbits shows a retinal angioma in the left globe.

B) Colour fundus photograph of left eye at most recent clinic visit showing areas of previous laser and cryotherapy treatment with dragging of optic nerve vessels towards inferotemporal quadrant. Haemangiomas present in macula and nasal quadrant. Multiple peripheral chorioretinal scars related to previously treated haemangiomas. Superotemporal vessels with perivascular exudate. Right eye normal (not shown). Visual acuity Right 6/5 and Left 6/12.

C) Fluorescein angiogram performed 4 years previously showing areas of scarring and retinal detachment related to exudation and effect of treatment. Optic nerve leak related to the effect of traction and glial proliferation with areas of hyperfluorescence in the macula related to new small haemangiomas, visible anterior to the internal limiting membrane on OCT scans (not shown). Right eye normal (not shown).

D) MRI scan showing 29 mm diameter renal cell carcinoma in right kidney.

E) Coronal T2 weighted image through the abdomen shows numerous small cysts in both kidneys.

F) Sagittal T1-weighted post contrast image of the spinal cord shows a solid enhancing haemangioblastoma with associated hypertrophied vessels on the dorsal surface of the spinal cord.

G) Axial T1-weighted post contrast image through the cervicomedullary junction shows a small solid haemangioblastoma.

H) Family pedigree of sporadic case of von Hippel-Lindau disease.

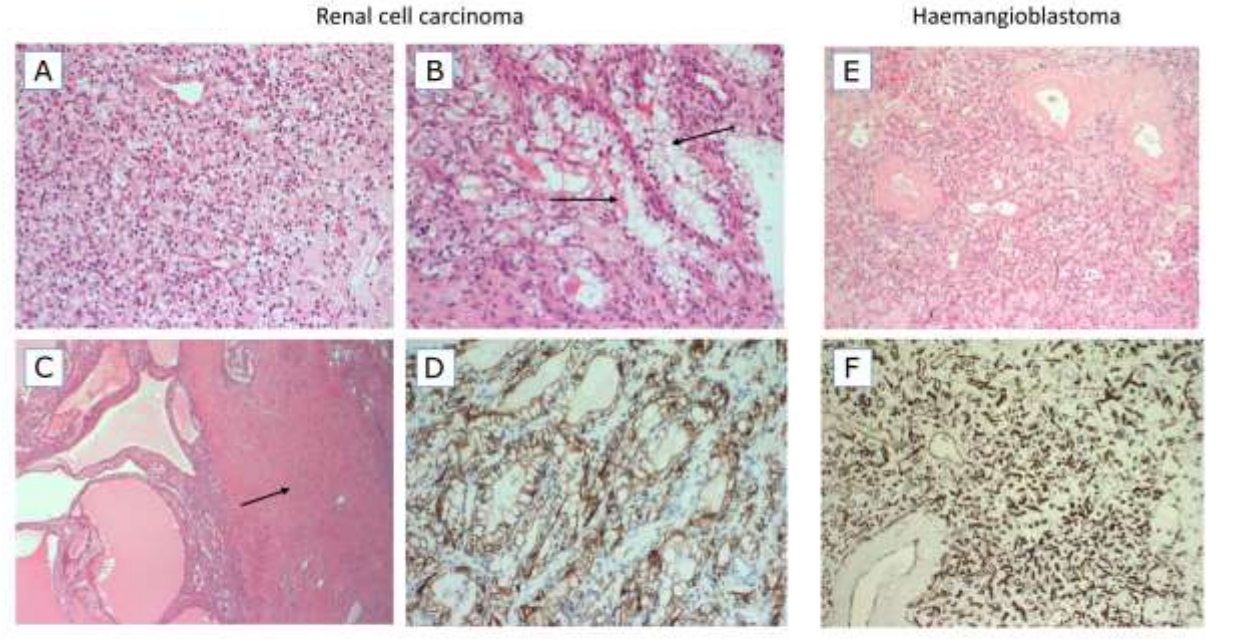

Figure 2:

Hematoxylin and eosin (H&E) stained images (A-C) and CA-IX staining image from the RCC. Hematoxylin and eosin (H&E) stained images from the haemangioblastoma (E, F).

A) An area with typical features of a clear cell renal cell carcinoma (ccRCC), composed of a sheet of small cells with clear cytoplasm and a delicate background vascular network.

B) Focus on branching tubules in which the tumour cells have more voluminous clear cytoplasm (arrows).

C) A cystic area of the RCC tumour (left) and a dense band of leiomyomatous (muscular) stroma (right, arrow).

D) The tumour was diffusely positive for CA-IX, a classic marker of HIF up-regulation.

E) The haemangioblastoma tumour is composed of very small cells with clear cytoplasm and a background vascular network. Larger blood vessels have thickened hyalinised walls.

F) The vascular network of haemangioblastoma is highlighted by CD34 immunohistochemistry.

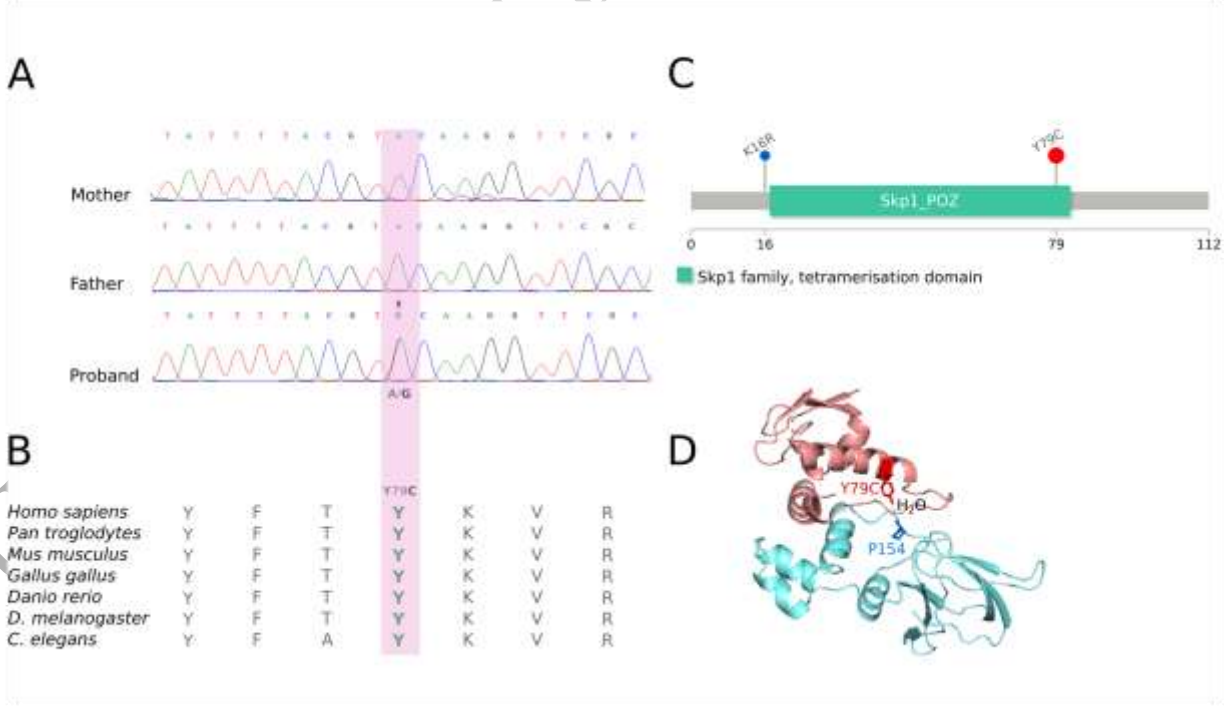

Figure 3:

A) Direct (Sanger) sequencing in trio shows the presence of the *ELOC* c.236A>G (p.Tyr79Cys) variant in the proband and absence from the parents.

B) Evolutionary conservation. Tyrosine at codon 79 (Y79) is evolutionary conserved across vertebrates and invertebrates (22).

C) ELOC domains. The Y79C variant (p.Tyr79Cys) is in the tetramerisation domain of the *ELOC* gene (23).

D) ELOC Y79C-VHL interaction. Tyr79 mediates a hydrogen bond with Pro154 of VHL via a water molecule. (adapted from (25)). The X-ray crystallographic structure of the ELOC/VHL complex was downloaded from the Protein Data Bank (PDB:4WQO) (26). Molecules other than ELOC and VHL were removed from the structure for clarity.

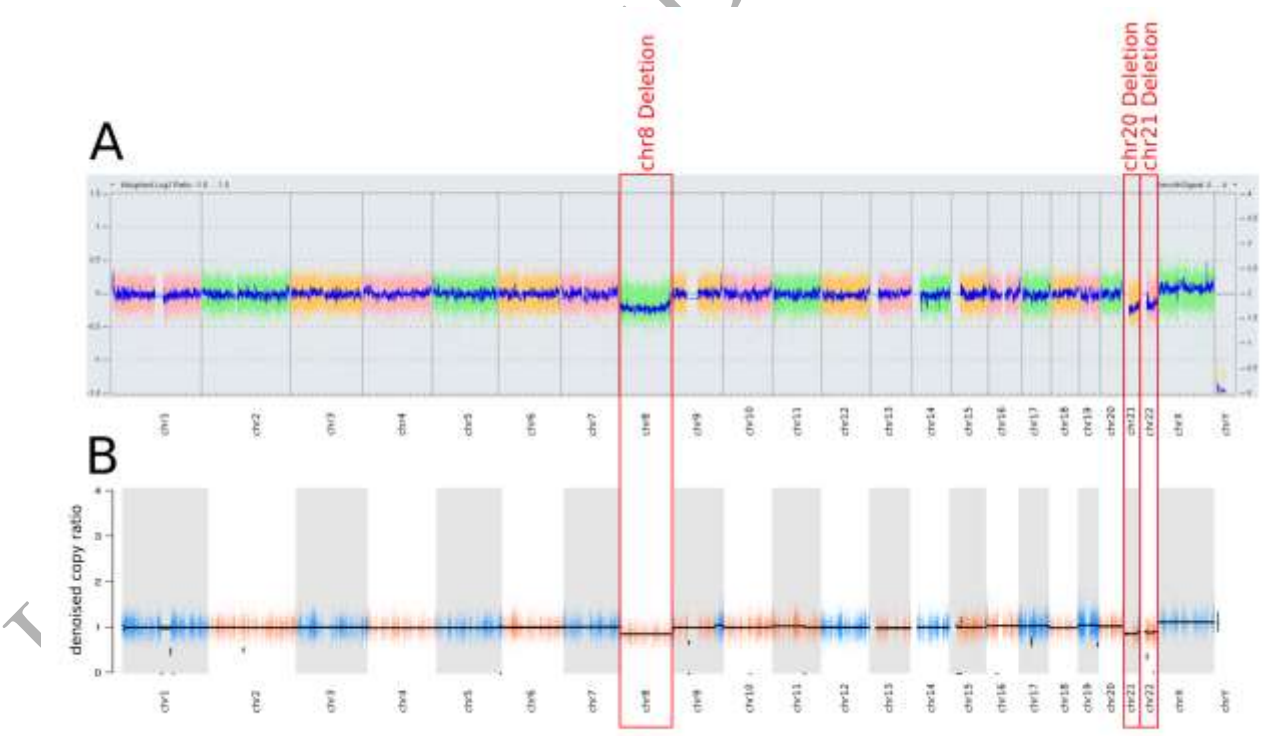

Figure 4:

A) Array CGH (750k array*)* of germline/tumour pair showed monosomy for chromosomes 8, 21, 22 (in ~20% of cells).

B) Targeted WES of germline/tumour pair identified in terms of copy number profile. This shows broad losses involving the full chromosome 8 and the long arms of chromosomes 21 and 22 (in ~40% of cells, tumour fraction ~ 47.8%). Any of the known recurrent RCC-related copy number aberrations (CNAs) (i.e. 3p, 9p or 14p losses, and 5q or chr7 amplification) were not found.

Figure 5:

Copy number analysis (CNA) profiles for the 8 RCC tumours with somatic *ELOC* variants from the 100,000 Genomes Project using Battenberg caller (subclonal copy number caller) (56). The *ELOC c.236A>G* (p.Tyr79Cys) missense variant was identified in Cases 1-4, Cases 5-7 had a non-codon 79 missense *ELOC* variant (NM_005648.4:c.274G>A (p.Glu92Lys), NM_005648.4:c.311T>A (p.Leu104Gln), NM_005648.4:c.74A>T (p.Asp25Val)) and Case 8 harboured an in-frame deletion (NM_005648.4:c.261_272del (p.Thr88_Pro91del)).

Abbreviations

2SC = 2-succinocysteine; aCGH = microarray-based comparative genomic hybridization; AE1/3 = cytokeratin AE1/3; CA-IX = carbonic anhydrase 9; ccRCC = clear cell renal cell carcinoma; CD31 = cluster of differentiation 31 protein; CD31 = cluster of differentiation 31 protein; CD31 = cluster of differentiation 31 protein; CK7 = cytokeratin 7; CNV = copy number variant; CUL2 = cullin 2; ELOC = elongin C; ELSTs = endolymphatic sac tumours; FFPE = formalin-fixed paraffin-embedded; gnomAD MAX AF = The Genome Aggregation Database maximum allele frequency; HEK 293T cells = Human embryonic kidney 293 cells; HIF-2a = hypoxia-inducible factor 2a; MLPA = multiplex-ligation dependent probe amplification; MSKCC = Memorial Sloan Kettering Cancer Centre; Pax-8 = paired box gene 8 protein; PDGFB = platelet-derived growth factor beta polypeptide; PPGL = pheochromocytoma/paraganglioma; pVHL = VHL protein; RBX1 = ring box 1; S100 = S 100 protein; SNP = single nucleotide polymorphism; SNV = single-nucleotide variant; SV= structural variant; TCGA = The Cancer Genome Atlas; VCB-CR complex = E3 ubiquitin ligase complex comprising pVHL, elongin C, elongin B, cullin 2 and RBX1; TGS= tumour suppressor gene; VEGF = vascular endothelial growth factor; VEGF =

vascular endothelial growth factor; WES = whole exome sequencing; WGS = whole genome sequencing.