# Gene panel sequencing improves the diagnostic work-up of patients with idiopathic erythrocytosis and identifies new mutations

Carme Camps,[1,2] Nayia Petousi,[3] Celeste Bento,[4] Holger Cario,[5] Richard R. Copley,[1,2] Mary Frances McMullin,[6] Richard vanWijk,[7] WGS500 consortium,[8] Peter J. Ratcliffe,[3] Peter A. Robbins,[9] and Jenny C. Taylor[1,2]

[1]National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre, Oxford, UK;
[2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK;
[3]Nuffield Department of Medicine, University of Oxford, Oxford, UK;
[4]Hematology Department, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal;
[5]Department of Pediatrics and Adolescent Medicine, University Medical Center, Ulm, Germany;
[6]Centre for Cancer Research and Cell Biology, Queen's University, Belfast, UK.
[7]Department of Clinical Chemistry and Hematology, University Medical Center Utrecht, Utrecht, Netherlands;
[8]A list of members and affiliations is provided in the supplementary information; and
[9]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK

CC and NP contributed equally to this work
PJR, PAR and JCT jointly supervised this work
Corresponding author: Nayia Petousi, Henry Wellcome Building, Roosevelt Drive, OX3 7BN, nayia.petousi@ndm.ox.ac.uk

Running head: Gene panel for idiopathic erythrocytosis

Word count for abstract: 248; Word count for text: 3996;
Figure/table count: 7; Supplemental files: 1

## *Acknowledgements*

The authors would like to thank the patients and their families who consented to this study, Melissa M. Pentony for the support provided with the management of Ion Torrent data and the Core and administration services at the Wellcome Trust Centre for Human Genetics, which are funded by the Wellcome Trust Core Award [090532/Z/09/Z]. This work was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre Oxford with funding from the Department of Health's NIHR Biomedical Research Centre's funding scheme. The WGS500 study was funded by the Wellcome Trust Core Award (090532/Z/09/Z) and a Medical Research Council Hub grant (G0900747 91070) to Peter Donnelly (director of the Wellcome Trust Centre of Human Genetics), the NIHR Biomedical Research Centre Oxford, the UK Department of Health's NIHR Biomedical Research Centres funding scheme and Illumina. NP is funded via a NIHR Clinical Lectureship. PJR is a member of the Ludwig Institute for Cancer Research.

## *Abstract*

Erythrocytosis is a rare disorder characterized by increased red cell mass and elevated hemoglobin concentration and hematocrit. Several genetic variants have been identified as causes for erythrocytosis in genes belonging to different pathways including oxygen sensing, erythropoiesis and oxygen transport. However, despite clinical investigation and screening for these mutations, the cause of disease cannot be found in a considerable number of patients, who are classified as having idiopathic erythrocytosis.

In this study, we developed a targeted next generation sequencing panel encompassing the exonic regions of 21 genes from relevant pathways (~79Kb) and sequenced 125 patients with idiopathic erythrocytosis. The panel effectively screened 97% of coding regions of these genes, with an average coverage of 450X. It identified 51 different rare variants, all leading to alteration of protein sequence, with 57 out of 125 cases (45.6%) having at least 1 of these variants. Ten of these were known erythrocytosis-causing variants, which had been missed following existing diagnostic algorithms. Twenty-two were novel variants in erythrocytosis-associated genes (*EGLN1*, *EPAS1*, *VHL*, *BPGM*, *JAK2*, *SH2B3*) and in novel genes included in the panel (*EPO*, *BHLHE41*, *OS9*, *EGLN2*, *HIF3A*), some with high likelihood of functionality, for which future segregation, functional and replication studies will be useful in providing further evidence for causality. The rest were classified as polymorphisms. Overall, these results demonstrate the benefits of using a gene panel versus existing methods where focused genetic screening is performed depending on biochemical measurements: it improves diagnostic accuracy and provides the opportunity for novel variant discovery.

## *Introduction*

Erythrocytosis is a clinical condition characterized by increased red cell mass and typically elevated hemoglobin (Hb) concentration and hematocrit (Hct)[(1)]. It can be *congenital* (e.g. genetic) or *acquired* and can be classified as *primary* or *secondary*[(1)] (Figure 1A). Several causal genetic mutations have been identified: heterozygous mutations in the Epo receptor (*EPOR*) gene cause primary congenital erythrocytosis[(3, 4)], while mutations in *JAK2* are predominantly associated with primary acquired erythrocytosis i.e. polycythemia vera[(5-7)]. Homozygous germline mutations in *VHL* e.g. Chuvash Polycythaemia and heterozygous germline mutations in *EGLN1* (*PHD2*) and *EPAS1* (*HIF2A*) have been found in patients with secondary congenital erythrocytosis[(3, 8)]. In the case of EPAS1, somatic gain-of-function mutations have also been detected in phaeochromocytomas and paragangliomas of some patients with congenital erythroctysosis, attributed to tissue mosaicism [ref]. Some patients, particularly those with polycythemia vera and some forms of genetic erythrocytosis, present with increased incidence of both arterial and venous thromboembolic events[(2).] Other congenital lesions include high oxygen-affinity hemoglobinopathies or 2,3-bisphosphoglycerate (2,3-BPG) deficiency[(9-11)], through mutations in globin genes (*HBA1, HBA2* and *HBB*) or the *BPGM* gene, respectively. These genes belong to key pathways involved in the pathogenesis of erythrocytosis e.g. oxygen-sensing (Hypoxia-Inducible

Factor, HIF) pathway, bone marrow erythropoiesis and oxygen transport (Figure 1B). Briefly, HIFs are transcription factors composed of 2 subunits: HIFα, which is oxygen-sensitive, and HIFβ. There are 3 HIFα isoforms, but HIF2α (*EPAS1*) is *EPO*'s main transcriptional regulator[12],[13]. In normoxia, HIFα is hydroxylated by oxygen-dependent prolyl hydroxylases (PHDs, encoded by *EGLN1*, *EGLN2* and *EGLN3*), binds to VHL and becomes ubiquitinated and degraded. In hypoxia, hydroxylation diminishes, HIFα stabilizes and initializes the transcription of target genes, including *EPO*[14]. Epo binds to the EPOR of erythroid progenitor cells in the bone marrow, stimulating proliferation and differentiation into red blood cells (RBC), through a JAK2-mediated signaling cascade. In RBC, BPGM promotes the release of oxygen to local tissues by producing 2,3-bisphosphoglycerate (2,3-BPG), which decreases the affinity of hemoglobin to oxygen.

Even if fully investigated (including screening for known mutations), there remains a considerable proportion of patients (~70%) in whom no cause is found, currently described as having idiopathic erythrocytosis[2, 4]. About two thirds of these patients have inappropriately normal or elevated Epo levels suggesting a defect in oxygen-sensing or oxygen delivery pathways. Most patients have early-onset disease and/or often a family history, suggesting a high probability for a genetic etiology. Logically, further investigation of these patients should begin by fully sequencing genes in which genetic variants are already known to cause erythrocytosis as opposed to simply screening for particular known variants. As many of these are in the HIF pathway, sequencing other key genes in this pathway (in which variants have not yet been observed) and also other erythropoiesis-related genes, is likely to be fruitful in the effort to resolve functional variants.

Using traditional methods of DNA sequencing, such as Sanger, to comprehensively sequence a large number of genes in a substantial number of patients with a relatively rare disease would be time-consuming, labor-intensive and impractical. On the other hand, using high-throughput technology e.g. whole genome sequencing (WGS) is not without its caveats: it generates huge volumes of data, bioinformatic analysis is complex with caution required to avoid false-positive assignment of causality, and it is currently too expensive for the number of patients involved. A way forward is the development of targeted next generation sequencing (NGS) panels of genes known to be associated with a condition.

We developed a NGS erythrocytosis gene panel to investigate the genetic origins of disease in idiopathic erythrocytosis patients. We used an ultra-high multiplex PCR method (*AmpliSeq, Thermo Fisher*), which allows rapid high-throughput sequencing of the full length of multiple genes in multiple samples. We defined a custom-made panel of 21 candidate genes, chosen from key pathways known to be involved in the pathogenesis of erythrocytosis, and used it to sequence 125 patients with idiopathic erythrocytosis. We also included novel candidate genes suggested by an initial WGS study, the WGS500 project[15], in which 500 clinical samples across a diverse spectrum of clinical disorders were sequenced, which included a few idiopathic erythrocytosis cases with high suspicion of a genetic cause.

The aims of the study were to: create a targeted sequencing panel, as a research tool, for the genetic investigation of erythrocytosis; evaluate its diagnostic utility in a cohort of idiopathic

erythrocytosis patients; search for novel variants in erythrocytosis-associated genes; and include new candidate genes identified in WGS500 to determine if they are mutated in additional patients.

## *Patients and Methods*

### Patient description

Patient DNA samples were obtained from whole blood and acquired from 4 separate databases of idiopathic erythrocytosis (UK, Portugal, Germany and the Netherlands). Participants gave informed consent and appropriate ethical approval was gained. The inclusion criteria were:

1. Confirmed absolute erythrocytosis with a red cell mass >125% predicted, and Hb >180 g/L and Hct >0.52 L/L in adult males or Hb >160 g/L and Hct >0.48 L/L in adult females, or Hb and Hct levels above the 99$^{th}$ centile of age-appropriate reference values in children.
2. Registered as idiopathic (cause of illness not identified), following appropriate investigation at each Center (Figure S1).
3. Early-onset disease, or cases with long-standing idiopathic erythrocytosis.

Details are found in Supplementary Information (SI).

Ten samples were whole-genome sequenced as part of the WGS500 project. 125 patient samples were sequenced using our erythrocytosis gene panel alongside 10 positive controls.

### Whole genome sequencing

Samples were sequenced at a 30X depth with Illumina HiSeq2000. Details are found in the SI.

### Design of a custom-made gene panel for idiopathic erythrocytosis

A customized panel, designed to encompass the coding and untranslated (UTR) regions of the candidate genes (Table 1), was created using the Ion AmpliSeq Designer (*Thermo Fisher*), whereby 635 primer pairs generating amplicons of ~200bp were designed. This panel covered 90.3% of the target region (78.96Kb), with a 97.4% of average coverage of the coding regions. The primers were synthesized in two multiplex pools.

### Ion Torrent sequencing and analysis

The primer pools were used with the Ion Ampliseq Library kit 2.0 and Ion Xpress barcode adapters (*Thermo Fisher*) to create libraries. Library quality and concentration were assessed using a High sensitivity DNA kit and Agilent 2100 Bioanalyzer (*Agilent Technologies*). Pools of 8 libraries were used for template preparation, loaded into an Ion 316 chip and sequenced on an Ion PGM instrument (500 flows). Details are found in the SI.

The Torrent Suite Software (*Thermo Fisher*) was used for quality control and alignment of the sequencing data to the human genome (Hg19). Variants were called with the Ion Reporter Software v4.2 (*Thermo Fisher*), using the germline workflow for single samples and the default parameters, and annotated with ANNOVAR[(16)]. Only variants fulfilling all of the following

conditions were selected for further analysis: confidence ≥40, read depth ≥20, frequency in 1000 Genomes (1000G) ≤3% and frequency in NHLBI ESP exomes (6500si) ≤3%.

The SIFT and PolyPhen2 HDIV scores and cut-offs from the LJB23 database in ANNOVAR were used to inform assessment of causality of non-synonymous variants. A variant was considered deleterious (D) by SIFT when sift score ≤0.05 and tolerated (T) when sift score >0.05. For PolyPhen2 HDIV, a variant was classified as probably damaging (D) when pp2_hdiv score ≥0.957, possibly damaging (P) when 0.453≤ pp2_hdiv score ≤0.956, or benign (B) when pp2_hdiv score ≤0.446. Synonymous variants were investigated for possible splicing effects using Human Splicing Finder, NetGene2 and FSPLICE.

**Sanger validation**

All relevant variants identified by Ion Torrent were confirmed by Sanger sequencing. For protocol and primer details see the SI and Table S1.

## *Results*

**Novel candidate genes and variants were identified by WGS**

The whole genomes of a small number of idiopathic erythrocytosis cases with high suspicion of a genetic cause were sequenced as part of the WGS500 project. Candidate variants were found in novel genes, not previously associated with erythrocytosis: *EPO*, *GFIb*, *KDM6A* and *BHLHE41*. Further details of the rationale and criteria used to select these genes as candidates are given in the SI and Table S2. On this basis, these genes were included in the NGS gene panel along with other erythrocytosis candidate genes (Table 1).

**The erythrocytosis gene panel has high performance in sequencing and variant detection**

Overall, 135 samples were sequenced on the Ion Torrent platform using the erythrocytosis gene panel (125 undiagnosed patients and 10 positive controls). On average, 89% of mapped reads were on target regions, which indicates a successful custom panel according to the manufacturer's guidance. The average coverage depth of the amplicons generated was 450X (Figure 2A). Most samples (133 out of 135) had over 92% of amplicons with coverage above 20X (Figure 2B). Only two samples presented substantial failure across the panel (Figure 2B), which was related to DNA quality. Only 17 amplicons (2.6%) had an average coverage lower than 20X across samples, indicating a general poor amplification of these regions within the highly-multiplexed reactions (Table S3). Ten of these (1.6% of all amplicons) had completely failed (coverage <20X in all samples), probably due to sequence context issues, meaning that any variants within these regions would not be called. Therefore, the sequencing was generally successful across samples, with a high percentage of the target sequence included at a good depth for germline variant calling.

We then compiled a list of all known erythrocytosis-associated variants from the literature[3, 4], including the variants identified in the WGS study, and cross-referenced their genomic coordinates with those of the generated amplicons. With the exception of two missense variants in *VHL*, all the other variants were within amplicons that showed good performance. The two *VHL* missense

variants – c.235C>T and c.311G>T – fall within an amplicon in exon 1 that showed complete failure and therefore would not be detected.

Importantly, our panel reliably detected 10 known variants – in different genes and hence in different amplicons – in the positive control samples, in which mutations had previously been identified either through WGS or Sanger sequencing (Table S4).

**51 exonic variants were identified across 57 patients by the erythrocytosis gene panel and validated by Sanger sequencing**

We identified 98 different variants across the coding regions of the genes examined, of which 19 were insertions or deletions (INDELS), 49 non-synonymous single nucleotide variations (SNVs) and 30 synonymous SNVs (Figure 3). None of the synonymous SNVs is predicted to alter splicing according to Human Splicing Finder, NetGene2 and FSPLICE. Therefore, we focused only on variants resulting in protein sequence alterations: following Sanger sequencing, 17 out of the 19 INDELS appeared to be false positives but 2 were confirmed. All 49 non-synonymous SNVs were confirmed, although for one SNV there was a single base discrepancy: Ion Torrent detected a triple base change from CAA to ATT in exon 12 of *JAK2* (chr9:5070025-5070027) but only a change from AA to TT (chr9:5070026-5070027) was confirmed by Sanger. As a result, a total of 51 variants (49 SNVs and 2 INDELS) were detected (Table S5). Therefore, 57 out of 125 cases had at least 1 exonic variant (45.6%); of those, 38 patients had only 1 exonic variant detected (30.4%), while 19 had more than 1 exonic variant detected (15.2%).

To investigate whether the variants discovered are unique to erythrocytosis patients (and therefore more likely to be disease-causing), we used *in silico* data from the 1000G project as control. For this, we examined the variant calls released by the 1000G project after integrating both exome and low coverage data across 1041 individuals and extracted the SNVs identified within the coordinates of the amplicons generated by our gene panel. We found that of the 49 non-synonymous SNVs discovered, 30 were uniquely found in our erythrocytosis cohort and not in the 1000G *in silico* control cohort, whereas the other 19 were also found in the control cohort (Figure 3). Those 19 SNVs (Table S6) are thus unlikely to be disease-causing mutations and most likely represent polymorphisms, found at similar or lower frequencies in the erythrocytosis cohort as in the control cohort (Fisher's exact test and Benjamini and Hochberg false discovery correction[17]).

Out of the 30 uniquely identified variants in our patient cohort, 10 were previously reported in the literature as causing erythrocytosis and hence are classified here as disease-causing variants (Table 2). The remaining 20 had no previous clinical associations. No exonic variants were identified in *EGLN3, HIF1AN (FIH), HBA1, HBA2, GFI1B* and *ZNF197*.

**Novel genes and variants identified by the erythrocytosis gene panel**

Of the 22 novel variants (20 SNVs and 2 INDELs) identified (Table 3), 14 were found in known erythrocytosis-associated genes, such as *VHL*, *EPAS1*, *JAK2, SH2B3* (*LNK*), *EGLN1* and *BPGM.* These variants are extremely rare: 5 were reported in dbSNP142 and ExAC (Exome Agregation

Consortium) databases at very low allele frequency (≤0.005) and 8 were only reported in ExAC at even lower allele frequency (≤0.0007). The remaining 7 SNVs and the 2 validated INDELs were not reported neither in dbSNP142 nor in ExAC, the latter containing exome data from 60,706 unrelated individuals. Most of these were classified as deleterious or damaging by either SIFT or PolyPhen2. There was consensus between both *in silico* tools for 12 of the variants, 8 of them being predicted as deleterious by both (Table 3).

Some of these novel or very rare variants have a high likelihood of causality based on the location and predicted effect of the protein coding change as well as on genetic evidence for causality, and are of particular physiological interest. For example, EPAS1 p.Y532H, a novel exon 12 mutation, is located one position downstream of residue 531, which is the prolyl hydroxylation site on HIF2α on the C-terminal ODD (oxygen-dependent degradation domain). Furthermore, it is part of a 6-residue domain which is highly conserved both across all HIFα isoforms and across species and which interacts with the VHL complex[18]. Thus, this mutation likely interferes with hydroxylation of HIF2α by PHDs and binding to the VHL complex, leading to upregulation of Epo. It was found in two related patients, father and son, both of whom had idiopathic erythrocytosis with raised Epo levels, and thus inherited in an autosomal dominant manner. Furthermore, EGLN1 p.L279P is affecting a conserved residue, previously reported as altered (p.L279Tfs43, a frameshift variant) in a patient with erythrocytosis[19]. Structurally, this residue is located on helix 3, which interacts with both N-terminal ODD and C-terminal ODD hydroxylation domains on HIFα[20]; a proline substitution may affect protein stability and diminish ODD binding, reducing HIFα hydroxylation. The VHL p.E52X variant introduces a stop codon, predicting termination of translation of the long VHL isoform (p30) while allowing only the translation of the alternative form of VHL (p19) from a translation site at M54. To date, only a few variants upstream of the *VHL* internal start codon 54 have been described, associated with either pheochromocytomas (codon 25 and 38) or with VHL disease (p.E46X and p.E52K)[21-23]. The role of the heterozygous VHL p.E52X in producing erythrocytosis in the patient in our study is not clear and the patient will be advised screening for the presence of VHL disease; there is evidence that erythrocytosis is seen in about 5-20% of patients with VHL disease[24].

Eight variants were identified in novel genes included in the panel because of their association with the oxygen-sensing pathway but in which no previous erythrocytosis-associated mutation has been reported, such as *EGLN2, HIF3A* and *OS9*. Interestingly, no candidate variants were found in key HIF pathway genes such as *EGLN3, HIF1AN* and importantly, *HIF1A*. This is consistent with existing literature in which variation in *EPAS1,* but not *HIF1A,* is associated with erythrocytosis. In addition, novel variants were also found in *EPO* and *BHLHE41*, two genes with no previous genetic association with erythrocytosis that were revealed by WGS500. For *EPO*, the most striking variant found is a frameshift, p.P7fs, detected in heterozygous status in one patient. Although at present it is difficult to link an apparently inactivating mutation to the generation of erythrocytosis, the variant has since been confirmed in heterozygous status in the patient's father who also presents with high hematocrit and hemoglobin levels. The patient had a normal Epo level (4.1 mU/ml). Epo levels are not known for the father. The variant is very close to an exon-intron junction, so one potential mechanism is that the reading frame is restored through alternative splicing leading to

the translation of an Epo analogue that is more potent. Two other EPO SNVs were detected in other patients but these are most likely very rare polymorphisms (see Table 3). Regarding *BHLHE41*, the novel missense variant (p.F149L) we identified is classified as benign by both PolyPhen2 and SIFT and is thus unlikely to be pathogenic, a notion supported by segregation analysis in the patient's family (see Table 3). Familial segregation studies, functional molecular studies as well as screening of larger erythrocytosis cohorts for replication of findings, will be needed to provide further evidence for causality of novel variants and novel genes.

## *Discussion*

The technical advances in next generation sequencing, together with the increasing understanding of the biological pathways underlying the pathogenesis of erythrocytosis, provide new opportunity for a refinement of the diagnostic pathway for the genetic diagnosis of patients with erythrocytosis.

Our approach allowed the creation of a NGS-based targeted gene panel with the capacity to process a large cohort of patient samples and simultaneously examine a large number of genes across several biological pathways in a systematic and efficient manner.

Our panel exhibited high performance and reliability. It produced high quality sequencing data with good target coverage. It accurately detected variants on 10 positive controls. It was excellent at reliably calling SNVs, with all SNVs identified subsequently validated in all samples by Sanger sequencing. A few limitations have, nevertheless, been recognized and should be taken into account when considering its future applications and before it can be considered optimal for immediate clinical diagnostic use. For example, a few amplicons – including a region on VHL exon 1 – showed complete failure across samples and thus potential variants within them would not be detected. Furthermore, there were some false positive INDEL detections, as has previously been reported by other users of Ion Torrent sequencing[25-27]. These could be addressed by re-designing primers covering that particular VHL genomic region, optimizing the variant calling bioinformatics workflow and employing recently proposed strategies to increase the accuracy of INDEL detection[25, 27]. Also, variant detection in genes with high sequence similarity such as HBA1 and HBA2 can be challenging and need extra analysis. Furthermore, another limitation of the panel – related more to the nature of the technology involved – is that it can only identify SNVs and short INDELs but not other structural variants such as large INDELs or copy number variations.

Currently, the clinical consensus for investigating erythrocytosis involves: establishing the diagnosis of absolute erythrocytosis, excluding systemic causes (e.g. hypoxic lung diseases or tumors) and then proceeding to focused genetic testing based on algorithms that attempt to predict the type of mutation that might be present. There is variability in procedures employed at different centers (Figure S1), but as a general rule: if Epo is low, variants in genes involved in bone marrow erythropoiesis (*EPOR*, *JAK2*) are screened for. If Epo is high or normal, P50 (partial pressure of oxygen at which 50% of Hb is saturated with oxygen) is calculated and if low, Hb electrophoresis is performed and/or variants in oxygen delivery pathways (globin genes, *BPGM*)

are screened for; if P50 is normal or not available, variants in the oxygen-sensing HIF pathway (*VHL*, *EPAS1*, *EGLN1*) are screened for[3, 28, 29].

Using our gene panel we were able to provide definitive genetic diagnoses in 9 patients that were previously missed. For example, a variant in *EPAS1*, p.G537R – a well-described gain-of-function mutation found in erythrocytosis patients[30, 31] – was detected. This was previously missed because the patient was not screened for *EPAS1* variants, owing to the fact that the Epo level was not high enough (and was thus directed to a different branch of the diagnostic algorithm). Similarly, we identified a homozygous *VHL* variant (p.H191D) known to cause erythrocytosis[32]. Interestingly, we found four variants in the *HBB* gene, all relating to high-affinity hemoglobinopathies associated with erythrocytosis: HBB p.H147P (Hb York), HBB p.H144Q (Hb Little Rock), HBB p.V110M (Hb San Diego) and HBB p.E102D (Hb Potomac)[33-37]. These were missed previously in the patients in our study, either because conventional screening with Hb electrophoresis can miss hemoglobinopathies[37] or because of difficulties obtaining optimal fresh venous blood samples for P50 measurements in all patients. In addition, we identified a heterozygous variant in *JAK2* (p.K539L) and two in *SH2B3* (p.E208Q and p.E400K), all known to associate with erythrocytosis[6, 38, 39]. The patient with variant JAK2 p.K439L, originally classified as idiopathic erythrocytosis as the conventional criteria for Polycythemia Vera (PV) including JAK2 p.V617F screening were not met, should now be considered as PV with a *JAK2* exon 12 mutation. As highlighted in previous studies[6, 7], the clinical picture of this subtype of PV is indistinguishable from that of idiopathic erythrocytosis. This emphasizes that *JAK2* exon 12 mutations should actively be screened for in idiopathic erythrocytosis patients. Furthermore, the findings of *SH2B3* variants highlight that this gene should also be surveyed, which now is not routinely done. The erythrocytosis gene panel can successfully do both. Of interest, the same *SH2B3* (p.E208Q) variant has since been found in the patient's father, who also has congenital erythrocytosis, confirming its germ-line origin and providing a definitive diagnosis in an additional (tenth) patient. Thus, we demonstrated that the panel allows reliable detection of known erythrocytosis-causing mutations, avoiding pitfalls that may occur when following existing algorithms.

In this study, 4 out of the 125 patients carried the VHL p.R200W variant in heterozygous state. VHL p.R200W causes Chuvash polycythemia in the homozygous state[40, 41]. Congenital erythrocytosis also occurs in patients who are compound heterozygotes[42-44], while heterozygous carriers of this variant are usually not affected. Nevertheless, VHL p.R200W heterozygous mutations feature significantly more frequently in erythrocytosis databases[4] than in general populations[45], suggesting a causal role. For one of the 4 patients here, the variant was newly identified. For the other 3, previous genetic tests had also identified it. Thus, within this study we aimed to detect additional genetic changes that might explain the clinical phenotype in these patients. We did not detect any other variants within *VHL*, except for 2 SNPs in the 3'UTR region which have high minor allelic frequencies (≥0.35 in dbSNP142) and are thus not considered significant. An alternative hypothesis is that the co-occurrence of this heterozygous variant with another heterozygous variant in a separate gene of the same biological pathway acts in synergy to produce disease. We did not obtain conclusive evidence for this in the four patients we studied: 2 did not have an additional variant; in the other 2, the VHL p.R200W co-occurred with additional heterozygous

missense variants that we classified as polymorphisms (Table S6), i.e. with EGLN1 p.A157Q and EGLN2 p.T405M in one patient and with EPOR p.G46E and EGLN2 p.S58L in the other.

Moreover, because the panel allows full sequencing of multiple genes instead of specific mutation screening, it gave the opportunity to detect 22 novel variants in disease-relevant genes. Some of these have strong likelihood of causality, based on the location of the mutated residues on functional or regulatory domains and the expected disturbance they would cause on protein structure and function (as explained in Results for EGLN1 p.L279P, EPAS1 p.Y532H and VHL p.E52X), and based on genetic evidence of familial segregation (e.g. EPAS1 p.Y532H which is dominantly inherited). For other variants, there is concordance between SIFT and Polyphen in their deleterious predictions and they are mostly found in known erythrocytosis-associated genes, whereas the rest have lesser evidence of functional candidacy (Table 3). While the functional significance of newly identified variants cannot at present be confirmed – and indeed clinical causation cannot be concluded –, by screening further larger erythrocytosis patient cohorts, the identification of these variants in other patients should provide genetic evidence of causality.

One limitation of our study is the lack of DNA from a source other than blood to determine germline or somatic status. This would predominantly be relevant for mutations in *JAK2* and *SH2B3,* in which somatic mutations are known to be associated with polycythaemia vera and myeloproliferative diseases. Indeed, for patients where JAK2 and SH2B3 variants are found following gene panel sequencing, further studies in skin/nail DNA should be advised.While somatic mutations in EPAS1 can be found in tumours of patients with erythrocytosis, mutations in EPAS1 and other genes of the HIF pathway and Hb transport detected in blood with this panel are likely to be germline.

Another motivation in this study was to assess whether variants in novel candidate genes identified in WGS500 – in *EPO*, *GFI1b*, *KDM6A* and *BHLHE41* – could be replicated. The precise WGS-identified variants were not found in this cohort of 125 cases, suggesting that larger patient cohorts need to be sequenced before the significance of variation in these genes can be properly interpreted. However, in the case of *EPO,* other variants were identified suggesting that *EPO* should be actively surveyed as an erythrocytosis-associated candidate gene.

Thus, despite the few technical limitations described, this gene panel has the potential to be useful in the diagnostic work-up of patients with erythrocytosis in clinical practice, after appropriate optimization. A point to note is that the gene panel in our study was applied to a highly-selected group that had undergone significant “filtering”, clinical and genetic (Figure S1) before inclusion in the study. Despite this, candidate variants – known causal and novel – were detected in 29% of patients. Thus, we propose that gene panel sequencing should be applied directly on “erythrocytosis cases where a genetic cause is suspected”, i.e. after clinical exclusion of acquired systemic causes and at the point where genetic testing is considered (Figure 4). This would undoubtedly increase the diagnostic yield and, because genetic testing would be conducted in an unbiased manner, it would improve diagnostic accuracy by decreasing the number of missed diagnoses. Furthermore, with accrued use in future patients, the erythrocytosis gene panel is likely

to generate an increasing number of “diagnostic” mutations through replication. Taking together all these results and observations, we hope to demonstrate the immediate utility that a targeted gene panel would have in the investigation of erythrocytosis at a time where next-generation sequencing is revolutionizing clinical medicine.

## *References*


1. McMullin MF. The classification and diagnosis of erythrocytosis. International journal of laboratory hematology. 2008 Dec;30(6):447-59.
2. McMullin MF. Idiopathic erythrocytosis: a disappearing entity. Hematology Am Soc Hematol Educ Program. 2009:629-35.
3. Hussein K, Percy M, McMullin MF. Clinical utility gene card for: familial erythrocytosis. European journal of human genetics : EJHG. 2012 May;20(5).
4. Bento C, Percy MJ, Gardie B, Maia TM, van Wijk R, Perrotta S, et al. Genetic basis of congenital erythrocytosis: mutation update and online databases. Human mutation. 2014 Jan;35(1):15-26.
5. Percy MJ, Jones FG, Green AR, Reilly JT, McMullin MF. The incidence of the JAK2 V617F mutation in patients with idiopathic erythrocytosis. Haematologica. 2006 Mar;91(3):413-4.
6. Scott LM, Tong W, Levine RL, Scott MA, Beer PA, Stratton MR, et al. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. The New England journal of medicine. 2007 Feb 1;356(5):459-68.
7. Percy MJ, Scott LM, Erber WN, Harrison CN, Reilly JT, Jones FG, et al. The frequency of JAK2 exon 12 mutations in idiopathic erythrocytosis patients with low serum erythropoietin levels. Haematologica. 2007 Dec;92(12):1607-14.
8. Lee FS, Percy MJ, McMullin MF. Oxygen sensing: recent insights from idiopathic erythrocytosis. Cell Cycle. 2006 May;5(9):941-5.
9. Percy MJ, Butt NN, Crotty GM, Drummond MW, Harrison C, Jones GL, et al. Identification of high oxygen affinity hemoglobin variants in the investigation of patients with erythrocytosis. Haematologica. 2009 Sep;94(9):1321-2.
10. Hoyer JD, Allen SL, Beutler E, Kubik K, West C, Fairbanks VF. Erythrocytosis due to bisphosphoglycerate mutase deficiency with concurrent glucose-6-phosphate dehydrogenase (G-6-PD) deficiency. American journal of hematology. 2004 Apr;75(4):205-8.
11. Petousi N, Copley RR, Lappin TR, Haggan SE, Bento CM, Cario H, et al. Erythrocytosis associated with a novel missense mutation in the BPGM gene. Haematologica. 2014 Oct;99(10):e201-4.
12. Gruber M, Hu CJ, Johnson RS, Brown EJ, Keith B, Simon MC. Acute postnatal ablation of Hif-2alpha results in anemia. Proceedings of the National Academy of Sciences of the United States of America. 2007 Feb 13;104(7):2301-6.
13. Rankin EB, Biju MP, Liu Q, Unger TL, Rha J, Johnson RS, et al. Hypoxia-inducible factor-2 (HIF-2) regulates hepatic erythropoietin in vivo. J Clin Invest. 2007 Apr;117(4):1068-77.
14. Webb JD, Coleman ML, Pugh CW. Hypoxia, hypoxia-inducible factors (HIF), HIF hydroxylases and oxygen sensing. Cell Mol Life Sci. 2009 Nov;66(22):3539-54.
15. Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nature genetics. 2015 Jul;47(7):717-26.
16. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010 Sep;38(16):e164.
17. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995;57(1):289-300.
18. Min JH, Yang H, Ivan M, Gertler F, Kaelin WG, Jr., Pavletich NP. Structure of an HIF-1alpha -pVHL complex: hydroxyproline recognition in signaling. Science. 2002 Jun 7;296(5574):1886-9.
19. Jang JH, Seo JY, Jang J, Jung CW, Lee KO, Kim SH, et al. Hereditary gene mutations in Korean patients with isolated erythrocytosis. Annals of hematology. 2014 Jun;93(6):931-5.
20. Chowdhury R, McDonough MA, Mecinovic J, Loenarz C, Flashman E, Hewitson KS, et al. Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. Structure. 2009 Jul 15;17(7):981-9.
21. van der Harst E, de Krijger RR, Dinjens WN, Weeks LE, Bonjer HJ, Bruining HA, et al. Germline mutations in the vhl gene in patients presenting with phaeochromocytomas. International journal of cancer Journal international du cancer. 1998 Jul 29;77(3):337-40.

22. Olschwang S, Richard S, Boisson C, Giraud S, Laurent-Puig P, Resche F, et al. Germline mutation profile of the VHL gene in von Hippel-Lindau disease and in sporadic hemangioblastoma. Human mutation. 1998;12(6):424-30.
23. Dollfus H, Massin P, Taupin P, Nemeth C, Amara S, Giraud S, et al. Retinal hemangioblastoma in von Hippel-Lindau disease: a clinical and molecular study. Investigative ophthalmology & visual science. 2002 Sep;43(9):3067-74.
24. Friedrich CA. Genotype-phenotype correlation in von Hippel-Lindau syndrome. Human molecular genetics. 2001 Apr;10(7):763-7.
25. Costa JL, Sousa S, Justino A, Kay T, Fernandes S, Cirnes L, et al. Nonoptical massive parallel DNA sequencing of BRCA1 and BRCA2 genes in a diagnostic setting. Human mutation. 2013 Apr;34(4):629-35.
26. Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, et al. Updating benchtop sequencing performance comparison. Nature biotechnology. 2013 Apr;31(4):294-6.
27. Yeo ZX, Chan M, Yap YS, Ang P, Rozen S, Lee AS. Improving indel detection specificity of the Ion Torrent PGM benchtop sequencer. PloS one. 2012;7(9):e45798.
28. Cario H, McMullin MF, Bento C, Pospisilova D, Percy MJ, Hussein K, et al. Erythrocytosis in children and adolescents-classification, characterization, and consensus recommendations for the diagnostic approach. Pediatric blood & cancer. 2013 Nov;60(11):1734-8.
29. Bento C, Almeida H, Maia TM, Relvas L, Oliveira AC, Rossi C, et al. Molecular study of congenital erythrocytosis in 70 unrelated patients revealed a potential causal mutation in less than half of the cases (Where is/are the missing gene(s)?). European journal of haematology. 2013 Oct;91(4):361-8.
30. Percy MJ, Furlow PW, Lucas GS, Li X, Lappin TR, McMullin MF, et al. A gain-of-function mutation in the HIF2A gene in familial erythrocytosis. The New England journal of medicine. 2008 Jan 10;358(2):162-8.
31. Gale DP, Harten SK, Reid CD, Tuddenham EG, Maxwell PH. Autosomal dominant erythrocytosis and pulmonary arterial hypertension associated with an activating HIF2 alpha mutation. Blood. 2008 Aug 1;112(3):919-21.
32. Tomasic NL, Piterkova L, Huff C, Bilic E, Yoon D, Miasnikova GY, et al. The phenotype of polycythemia due to Croatian homozygous VHL (571C>G:H191D) mutation is different from that of Chuvash polycythemia (VHL 598C>T:R200W). Haematologica. 2013 Apr;98(4):560-7.
33. Charache S, Jacobson R, Brimhall B, Murphy EA, Hathaway P, Winslow R, et al. Hb Potomac (101 Glu replaced by Asp): speculations on placental oxygen transport in carriers of high-affinity hemoglobins. Blood. 1978 Feb;51(2):331-8.
34. Gonzalez Fernandez FA, Villegas A, Ropero P, Carreno MD, Anguita E, Polo M, et al. Haemoglobinopathies with high oxygen affinity. Experience of Erythropathology Cooperative Spanish Group. Annals of hematology. 2009 Mar;88(3):235-8.
35. Bromberg PA, Alben JO, Bare GH, Balcerzak SP, Jones RT, Brimhall B, et al. High oxygen affinity variant of haemoglobin Little Rock with unique properties. Nature: New biology. 1973 Jun 6;243(127):177-9.
36. Misgeld E, Gattermann N, Wehmeier A, Weiland C, Peters U, Kohne E. Hemoglobinopathy York [beta146 (HC3) His==>Pro]: first report of a family history. Annals of hematology. 2001 Jun;80(6):365-7.
37. Wajcman H, Galacteros F. Abnormal hemoglobins with high oxygen affinity and erythrocytosis. Hematology and cell therapy. 1996 Aug;38(4):305-12.
38. Spolverini A, Pieri L, Guglielmelli P, Pancrazzi A, Fanelli T, Paoli C, et al. Infrequent occurrence of mutations in the PH domain of LNK in patients with JAK2 mutation-negative 'idiopathic' erythrocytosis. Haematologica. 2013 Sep;98(9):e101-2.
39. McMullin MF, Wu C, Percy MJ, Tong W. A nonsynonymous LNK polymorphism associated with idiopathic erythrocytosis. American journal of hematology. 2011 Nov;86(11):962-4.
40. Ang SO, Chen H, Hirota K, Gordeuk VR, Jelinek J, Guan Y, et al. Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia. Nature genetics. 2002 Dec;32(4):614-21.
41. Smith TG, Brooks JT, Balanos GM, Lappin TR, Layton DM, Leedham DL, et al. Mutation of von Hippel-Lindau tumour suppressor and human cardiopulmonary physiology. PLoS Med. 2006 Jul;3(7):e290.
42. Bento MC, Chang KT, Guan Y, Liu E, Caldas G, Gatti RA, et al. Congenital polycythemia with homozygous and heterozygous mutations of von Hippel-Lindau gene: five new Caucasian patients. Haematologica. 2005 Jan;90(1):128-9.
43. Cario H, Schwarz K, Jorch N, Kyank U, Petrides PE, Schneider DT, et al. Mutations in the von Hippel-Lindau (VHL) tumor suppressor gene and VHL-haplotype analysis in patients with presumable congenital erythrocytosis. Haematologica. 2005 Jan;90(1):19-24.
44. Percy MJ, McMullin MF, Jowitt SN, Potter M, Treacy M, Watson WH, et al. Chuvash-type congenital polycythemia in 4 families of Asian and Western European ancestry. Blood. 2003 Aug 1;102(3):1097-9.
45. Liu E, Percy MJ, Amos CI, Guan Y, Shete S, Stockton DW, et al. The worldwide distribution of the VHL 598C>T mutation indicates a single founding event. Blood. 2004 Mar 1;103(5):1937-40.
46. Panovska-Stavridis Iea. Familiar JAK2 G571S Variant Not Linked with Essential Trombocythemia. Blood, 2014:21.

## *Figure Legends*

**Figure 1. (A) Causes of erythrocytosis.** Erythrocytosis can be congenital or acquired. It is classified as primary, where there is an intrinsic defect in erythropoietic cells and erythropoietin (Epo) levels are low, or secondary, where the increased red cell production is externally driven through increased Epo production and Epo levels are high or inappropriately normal. **(B) Pathways involved in the pathogenesis of erythrocytosis.** (i) Hypoxia inducible factor (HIF) oxygen sensing pathway in renal EPO producing cells. HIFs are dimeric transcription factors composed of one α- and one β- subunit. In normoxia, HIFα subunits are hydroxylated by oxygen-dependent prolyl-hydroxylases (PHDs) and asparaginyl hydroxylase (HIF1AN). The hydroxylated prolines (P) are recognised by VHL, which mediates the ubiquitination and proteasomal degradation of HIFα. The hydroxylated asparagine (N) compromises the interaction of HIFα with cofactors necessary for transcriptional activity (p300/CBP). In hypoxia, PHDs and HIF1AN are less active, HIFα subunits stabilize and translocate into the nucleus where they interact with HIFβ subunit and cofactors and initiate transcription of target genes, including *EPO* (ii) Erythropoiesis in the bone marrow. It is triggered by the binding of EPO to the EPO receptor (EPOR) located in the surface of erythroid progenitor cells and the subsequent activation of JAK2-signalling cascade. The process is inhibited by the interaction of SH2B3 and JAK2. (iii) Hemoglobin (Hb) synthesis and oxygen transport. BPGM produces 2,3-BPG, which promotes the release of oxygen to local tissues by decreasing the affinity of deoxygenated Hb to oxygen. Alterations in the Hb chains (Hb-α and Hb-β) or BPGM could shift the Hb-oxygen dissociation curve and alter oxygen levels, which directly influences Epo production. (PV: Polycythemia Vera; ECYT 1-4: erythrocytosis type 1-4**;** Hb: hemoglobin; $O_2$: oxygen; 2,3 BPG: 2,3-bisphosphoglycerate; RBC: red blood cells, EPO: erythropoietin; PHDs: prolyl hydroxylases). PHDs: PHD1 (*EGLN2*), PHD2 (*EGLN1*) and PHD3 (*EGLN3*).

**Figure 2. Coverage of the amplicons generated by the erythrocytosis gene panel across 135 samples.** (A) Each boxplot represents the distribution of the number of reads obtained for all the amplicons generated by the panel within each sample. The horizontal line across the plot shows the average coverage (450X). (B) Each dot represents the percentage of amplicons with coverage over 20X within each sample.

**Figure 3. Overview of the exonic variants detected with Ion Torrent sequencing among 125 erythrocytosis patients, validation and further classification.**

**Figure 4. Proposed use of the erythrocytosis gene panel in the investigation of erythrocytosis.** The gene panel would make genetic testing more efficient and stream-lined. It enables the simultaneous survey of the full length of 21 candidate genes, in a systematic and unbiased manner, allowing the detection of known causal variants as well as novel variants in known and novel genes.

## *Tables*

**Table 1.** Genes included in the custom-made erythrocytosis gene panel

| Candidate Gene | Position | No of exons | Pathway | Candidacy |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| ***VHL*** | Chr3:10183319-10195354 | 3 | Oxygen-sensing | Known erythrocytosis-causing variants |
| ***EPAS1*** | Chr2:46524541-46613842 | 16 | Oxygen-sensing | Known erythrocytosis-causing variants |
| ***EGLN1*** | Chr1:231499497-231560790 | 4 | Oxygen-sensing | Known erythrocytosis-causing variants |
| ***HIF1A*** | Chr14:62162119-62214977 | 15 | Oxygen-sensing | Key gene of the HIF pathway |
| ***HIF3A*** | Chr19:46800303-46846690 | 13 | Oxygen-sensing | Key gene of the HIF pathway |
| ***EGLN2*** | Chr19:41305048-41314346 | 5 | Oxygen-sensing | Key gene of the HIF pathway |
| ***EGLN3*** | Chr14:34393421-34420284 | 5 | Oxygen-sensing | Key gene of the HIF pathway |
| ***HIF1AN*** | Chr10:102295641-102313681 | 6 | Oxygen-sensing | Key gene of the HIF pathway |
| ***EPO*** | Chr7:100318423-100321323 | 5 | Erythropoiesis/ Oxygen-sensing | 1. Key gene in erythropoiesis<br>2. Identified in WGS500 |
| ***EPOR*** | Chr19:11487881-11495018 | 8 | Erythropoiesis | Known erythrocytosis-causing variants |
| ***JAK2*** | Chr9:4985245-5128183 | 25 | Erythropoiesis | Known erythrocytosis-causing variants |
| ***SH2B3*** | Chr12:111843752-111889427 | 8 | Erythropoiesis | Known erythrocytosis-causing variants |
| ***BPGM*** | Chr7:134331531-134364568 | 3 | Oxygen transport | 1. Known erythrocytosis-causing variants<br>2. Identified in WGS |
| ***HBB*** | Chr11: 5246696-5248301 | 3 | Oxygen transport/ Hemoglobin synthesis | Known erythrocytosis-causing variants |
| ***HBA1*** | Chr16:226679-227520 | 3 | Oxygen transport/ Hemoglobin synthesis | Key gene in oxygen transport |
| ***HBA2*** | Chr16:222846-223709 | 3 | Oxygen transport/ Hemoglobin synthesis | Key gene in oxygen transport |
| ***KDM6A*** | ChrX:44732423-44971747 | 29 | Oxygen-regulated demethylase | Identified in WGS500 |
| ***GFI1b*** | Chr9:135854098-135867084 | 11 | Erythropoiesis | Identified in WGS500 |

| | | | | |
|---|---|---|---|---|
| ***BHLHE41*** | Chr12:26272959-26278003 | 9 | Factor associated with HIF | Identified in WGS500 |
| ***OS9*** | Chr12:58087738-58115340 | 15 | Factor associating with HIF | Related to the HIF pathway |
| ***ZNF197*** | Chr3: 44666511-44689963 | 5 | Factor associating with HIF | Related to the HIF pathway |

Official gene symbols according to HUGO Gene Nomenclature Committee are given here. Other gene symbols used frequently in the literature are: *HIF2A* (*EPAS1*), *PHD2* (*EGLN1*), *PHD1* (*EGLN2*), *PHD3* (*EGLN3*), *FIH* (*HIF1AN*), *LNK* (*SH2B3*), *DEC2* (BHLHE41).

**Table 2.** Variants detected by the erythrocytosis gene panel, known to cause erythrocytosis

| Chr | Position | Ref | Alt | Gene | Variant type | cDNA Change | Protein Change | Geno type | No of cases | Mechanism of action | Previous Publication |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 46607420 | G | A | *EPAS1* | SNV | c.1609G>A | p.G537R | Het | 1 | Gain of function of HIF2A | Percy et al 2008(30) Gale et al 2008(31) |
| 3 | 10191578 | C | G | *VHL* | SNV | c.571C>G | p.H191D | Hom | 1 | Loss of function (enhances HIF regulated gene expression) | Tomasic et al 2013(32) |
| 3 | 10191605 | C | T | *VHL* | SNV | c.598C>T | p.R200W | Het* | 4 | Loss of function (decreased HIF binding & hydroxylation, enhances HIF-regulated gene expression) | Ang et al 2002(40) |
| 9 | 5070025 | CAA | ATT | *JAK2* | SNV | c.1615_ 1616invAA | p.K539L | Het | 1 | Gain of function of JAK2 (K539L) | Scott et al. 2007(6) |
| 11 | 5246832 | T | G | *HBB* | SNV | c.440A>C | p.H147P | Het | 1 | High oxygen affinity Hb (Hb York) | Misgeld et al 2001(36) |
| 11 | 5246840 | G | C | *HBB* | SNV | c.432C>G | p.H144Q | Het | 1 | High oxygen affinity Hb (Hb Little Rock) | Bromberg et al 1973(35) Wajcman et al 1996(37) |
| 11 | 5246944 | C | T | *HBB* | SNV | c.328G>A | p.V110M | Het | 1 | High oxygen affinity Hb (Hb San Diego) | Wajcman et al 1996(37) Gonzalez et al 2009(34) |
| 11 | 5247816 | C | G | *HBB* | SNV | c.306G>C | P.E102D | Het | 1 | High oxygen affinity Hb (Hb Potomac) | Charache et al 1978(33) |
| 12 | 111856571 | G | C | *SH2B3* | SNV | c.622G>C | p.E208Q | Het | 1 | Enhances JAK2 | Spolverini |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | signalling | et al(38) |
| 12 | 111885310 | G | A | *SH2B3* | SNV | c.1198G>A | p.E400K | Het | 1 | Interacts with JAK2 signalling | McMullin et al 2011(39) Spolverini et al(38) |

Official gene symbols according to HUGO Gene Nomenclature Committee are given here. Other gene symbols used frequently in the literature are: *HIF2A* (*EPAS1*), *LNK* (*SH2B3*). Chr indicates chromosome; Ref, reference allele; Alt, alternate allele; SNV, single nucleotide variant; Het, heterozygous; Hom, homozygous; and Hb, hemoglobin. * This variant causes Chuvash polycythemia in the homozygous state. In one of the patients, this variant was discovered in this study, whereas the other 3 had been detected in previous genetic tests.

**Table 3.** Novel variants detected by the erythrocytosis gene panel

| Chr | Position | Ref | Alt | Gene | Variant type | cDNA Change | Protein Change | Geno type | No of cases | SIFT/ Polyphen Score | Allele freq.: dbSNP142 ExAc | Evidence of causality |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 231556799 | A | G | *EGLN1* | SNV | c.T836C | p.L279P | Het | 1 | D/D | Not found Not found | Predicted structural/ functional effects |
| 2 | 46607405 | T | C | *EPAS1* | SNV | c.T1594C | p.Y532H | Het | 2 | D/D | Not found Not found | Predicted structural/ functional effects Segregation |
| 3 | 10183685 | G | T | *VHL* | SNV | c.G154T | p.E52X | Het | 1 | T/NA | Not found 4.16E-05 | Predicted structural/ functional effects |
| 7 | 100319185 | TC | T | *EPO* | deletion | c.19delC | p.P7fs | Het | 1 | NA/NA | Not found Not found | Segregation |
| 2 | 46574031 | AAGG | A | *EPAS1* | deletion | c.47delAGG | p.del17E | Het | 1 | NA/NA | Not found Not found | Deleterious (by both SIFT and Poliphen) Most not found in large population databases Most in known erythrocytosis-associated genes |
| 7 | 100320336 | A | G | *EPO* | SNV | c.A296G | p.E99G | Het | 1 | D/D | Not found Not found | |
| 9 | 5050747 | A | T | *JAK2* | SNV | c.A530T | p.E177V | Het | 1 | D/D | Not found Not found | |
| 12 | 111856181 | G | A | *SH2B3* | SNV | c.G232A | p.E78K | Het | 1 | D/P | Not found 7.40E-04 | |
| 12 | 111884812 | G | A | *SH2B3* | SNV | c.G901A | p.E301K | Het | 1 | D/D | 3.20E-05 3.30E-05 | |
| 12 | 111885466 | C | T | *SH2B3* | SNV | c.C1243T | p.R415C | Hom | 1 | D/D | 1.00E-03 4.19E-05 | |
| 19 | 41313427 | G | T | *EGLN2* | SNV | c.G1139T | p.R380L | Het | 1 | D/D | Not found Not found | |
| 2 | 46611651 | T | C | *EPAS1* | SNV | c.T2465C | p.M822T | Het | 1 | D/B | Not found 8.24E-06 | Extremely rare variants |
| 3 | 10183605 | C | T | *VHL* | SNV | c.C74T | p.P25L | Het | 2 | D/B | 4.00E-04 5.17E-03 | |
| 7 | 100320290 | G | C | *EPO* | SNV | c.G250C | p.G84R | Het | 2 | T/D | Not found 8.04E-05 | |
| 7 | 134346563 | C | A | *BPGM* | SNV | c.C304A | p.Q102K | Het | 1 | D/B | Not found Not found | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 5022168 | G | A | *JAK2* | SNV | c.G181A | p.E61K | Het | 1 | T/B | Not found<br>Not found | |
| 9 | 5054775 | G | C | *JAK2* | SNV | c.G827C | p.G276A | Het | 1 | T/B | Not found<br>8.29E-06 | |
| 9 | 5072561 | G | A | *JAK2* | SNV | c.G1711A | p.G571S* | Het | 1 | T/D | 7.4 E-04<br>4.81E-04 | |
| 12 | 26276001 | A | C | *BHLHE41* | SNV | c.T447G | p.F149L | Het | 1 | T/B | Not found<br>1.54E-04 | |
| 12 | 58109559 | G | A | *OS9* | SNV | c.G497A | p.G166D | Het | 1 | T/D | Not found<br>3.42E-05 | |
| 19 | 46811511 | A | C | *HIF3A* | SNV | c.A190C | p.I64L | Het | 1 | D/B | Not found<br>1.65E-05 | |
| 19 | 46823777 | C | A | *HIF3A* | SNV | c.C896A | p.A299D | Het | 1 | T/B | 0.00066<br>7.93E-04 | |

Official gene symbols according to HUGO Gene Nomenclature Committee are given here. Other gene symbols used frequently in the literature are: *HIF2A* (*EPAS1*), *PHD2* (*EGLN1*), *PHD1* (*EGLN2*), *LNK* (*SH2B3*), *DEC2* (BHLHE41). Chr indicates chromosome; Ref, reference allele; Alt, alternate allele; SNV, single nucleotide variant; Het, heterozygous; Hom, homozygous; D/, deleterious by SIFT; T/, tolerated by SIFT; /D, probably damaging by Polyphen2 HDIV; /P, possibly damaging by Polyphen2 HDIV; /B, benign by Polyphen2 HDIV; NA, non-applicable. ExAc: Exome Aggregation Consortium, Cambridge, MA (http://exac.broadinstitute.org). * JAK2 G751S has been reported previously in myeloproliferative disorders[46] but it is thought to be a silent non-functioning polymorphism; indeed, it was subsequently detected in the unaffected mother of the patient in our study by sanger sequencing, supporting this possibility.