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 van Wijk,[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, UK; [3]Nuffield Department of Medicine, University of Oxford, UK; [4]Hematology Department, Centro Hospitalar e Universitário de 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, the Netherlands; [8]A list of members and affiliations is provided in the Online Supplementary Information; and [9]Department of Physiology, Anatomy and Genetics, University of Oxford, UK

**CC and NP contributed equally to this work*

***PJR, PAR and JCT jointly supervised this work*

Haematologica 2016
Volume 101(11):1306-1318

Correspondence:

nayia.petousi@ndm.ox.ac.uk

Received: February 9, 2016.

Accepted: July 26, 2016.

Pre-published: September 20, 2016.

doi:10.3324/haematol.2016.144063

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/101/11/1306

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 (~79 Kb) 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 alterations of protein sequence, with 57 out of 125 cases (45.6%) having at least one 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 (e.g. *EPO*, *EGLN2*, *HIF3A*, *OS9*), some with a high likelihood of functionality, for which future segregation, functional and replication studies will be useful to provide further evidence for causality. The rest were classified as polymorphisms. Overall, these results demonstrate the benefits of using a gene panel rather than existing methods in which focused genetic screening is performed depending on biochemical measurements: the gene panel improves diagnostic accuracy and provides the opportunity for discovery of novel variants.

Introduction

Erythrocytosis is a clinical condition characterized by increased red cell mass and typically elevated hemoglobin concentration and hematocrit.[1] It can *be congenital* (e.g. genetic) or *acquired* and classified as *primary* or *secondary*[1] (Figure 1A). Several causal genetic mutations have been identified. Heterozygous mutations in the erythropoietin receptor (EPOR) gene cause primary congenital

erythrocytosis,[2,3] while *JAK2* mutations are predominantly associated with primary acquired erythrocytosis i.e. polycythemia vera.[4-6] Homozygous germline mutations in *VHL* e.g. Chuvash polycythemia and heterozygous germline mutations in *EGLN1* (*PHD2*) and *EPAS1* (*HIF2A*) have been found in patients with secondary congenital erythrocytosis.[2,7] Regarding *EPAS1*, somatic gain-of-function mutations have been detected in pheochromocytomas and paragangliomas in patients with congenital erythrocytosis, attributed to tissue mosaicism.[8] Some patients, particularly those with polycythemia vera and some forms of genetic erythrocytosis, have increased incidences of both arterial and venous thromboembolic events.[9] Other congenital lesions include high oxygen-affinity hemoglobinopathies or 2,3-bisphosphoglycerate deficiency,[10-12] caused by mutations in globin genes (*HBA1*, *HBA2*, *HBB*) or the *BPGM* gene, respectively. These genes belong to key pathways involved in the pathogenesis of erythrocytosis e.g. the oxygen-sensing (hypoxia-inducible factor, HIF) pathway, erythropoiesis and oxygen transport (Figure 1B). Briefly, HIF are transcription factors composed of two subunits: HIFα, which is oxygen-sensitive, and HIFβ. There are three HIFα isoforms, but HIF2α (*EPAS1*) is erythropoietin's (EPO) main transcriptional regulator.[13,14] In normoxia, HIFα is hydroxylated by oxygen-dependent prolyl hydroxylases (encoded by *EGLN1*, *EGLN2* and *EGLN3*), binds to VHL and becomes ubiquitinated and degraded. In hypoxia, hydroxylation diminishes and HIFα stabilizes and initiates the transcription of target genes, including *EPO*.[15] Erythropoietin binds to the EPOR of erythroid progenitor cells in the bone marrow, stimulating proliferation and differentiation into red blood cells, through a JAK2-mediated signaling cascade. In red blood cells, BPGM promotes the release of oxygen to local tissues by producing 2,3-bisphosphoglycerate, which decreases the affinity of hemoglobin to oxygen.

Even if fully investigated (including screening for known mutations), a considerable proportion of patients (~70%) remain without an identified cause of their erythrocytosis and are described as having idiopathic erythrocytosis.[3,9] About two thirds of these patients have inappropriately normal or elevated erythropoietin 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 of 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 DNA sequencing methods, e.g. Sanger sequencing, to comprehensively sequence a large number of genes in a substantial number of patients with a relatively rare disease is time-consuming, labor-intensive and impractical. Conversely, high-throughput technology e.g. whole-genome sequencing (WGS), has its own drawbacks with generation of huge volumes of data, high cost and complex bioinformatic analysis. A way forward is the development of disease-relevant, targeted, next-generation sequencing gene panels.

We developed a next-generation sequencing erythrocytosis gene panel, using an ultra-high multiplex polymerase chain reaction 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 from key pathways 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,[16] in which 500 samples across a diverse spectrum of clinical disorders were sequenced, including some cases of idiopathic erythrocytosis strongly suspected of having a genetic cause.

The aims of the study were: (i) to create a targeted sequencing panel, as a research tool, for the genetic investigation of erythrocytosis; (ii) to evaluate the panel's diagnostic utility in a cohort of patients with idiopathic erythrocytosis; (iii) to search for novel variants in erythrocytosis-associated genes; and (iv) to include new candidate genes identified in WGS500 to determine whether they are mutated in additional patients.

Methods

Patients

DNA samples extracted from the blood of patients with idiopathic erythrocytosis were acquired from four separate idiopathic erythrocytosis databases (UK, Portugal, Germany and The Netherlands). Participants gave informed consent and appropriate ethical approval was gained. The inclusion criteria were: (i) confirmed absolute erythrocytosis with a red cell mass >125% predicted, and hemoglobin >180 g/L and hematocrit >0.52% in adult males or hemoglobin >160 g/L and hematocrit >0.48% in adult females, or hemoglobin and hematocrit levels above the 99th centile of age-appropriate reference values in children; (ii) registered as idiopathic (unidentified cause of illness), following appropriate investigation at each Center (*Online Supplementary Figure S1*); and (iii) early-onset disease, or cases with long-standing idiopathic erythrocytosis. Details are given in the *Online Supplementary Information*.

Ten samples were whole-genome sequenced as part of the WGS500 project, whereas we used our erythrocytosis gene panel to sequence 125 samples from patients with idiopathic erythrocytosis as well as ten positive controls.

Whole-genome sequencing

Samples were sequenced at a 30X depth with Illumina HiSeq2000. Details are provided in the *Online Supplementary Information*.

Ion Torrent sequencing and analysis

A customized panel, encompassing the coding and untranslated regions of the candidate genes (Table 1), was created using the Ion AmpliSeq Designer (Thermo Fisher), whereby 635 primer pairs generating amplicons of ~200 bp were designed. This panel covered 90.3% of the target region (78.96 Kb), with 97.4% average coverage of the coding regions. The primers, synthesized in two multiplex 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 2100 Bioanalyzer (Agilent Technologies). Pools of eight libraries were used for template preparation, loaded into an Ion 316 chip and sequenced on an Ion PGM instrument (500 flows).

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.[17] 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%. Provean and the SIFT and PolyPhen2-HDIV scores and cut-offs from the ANNOVAR LJB23 database were used to assess causality of non-synonymous variants. Synonymous variants were investigated for possible splicing effects using Human Splicing Finder, NetGene2 and FSPLICE. Further details are given in the *Online Supplementary Information*.

Sanger validation

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

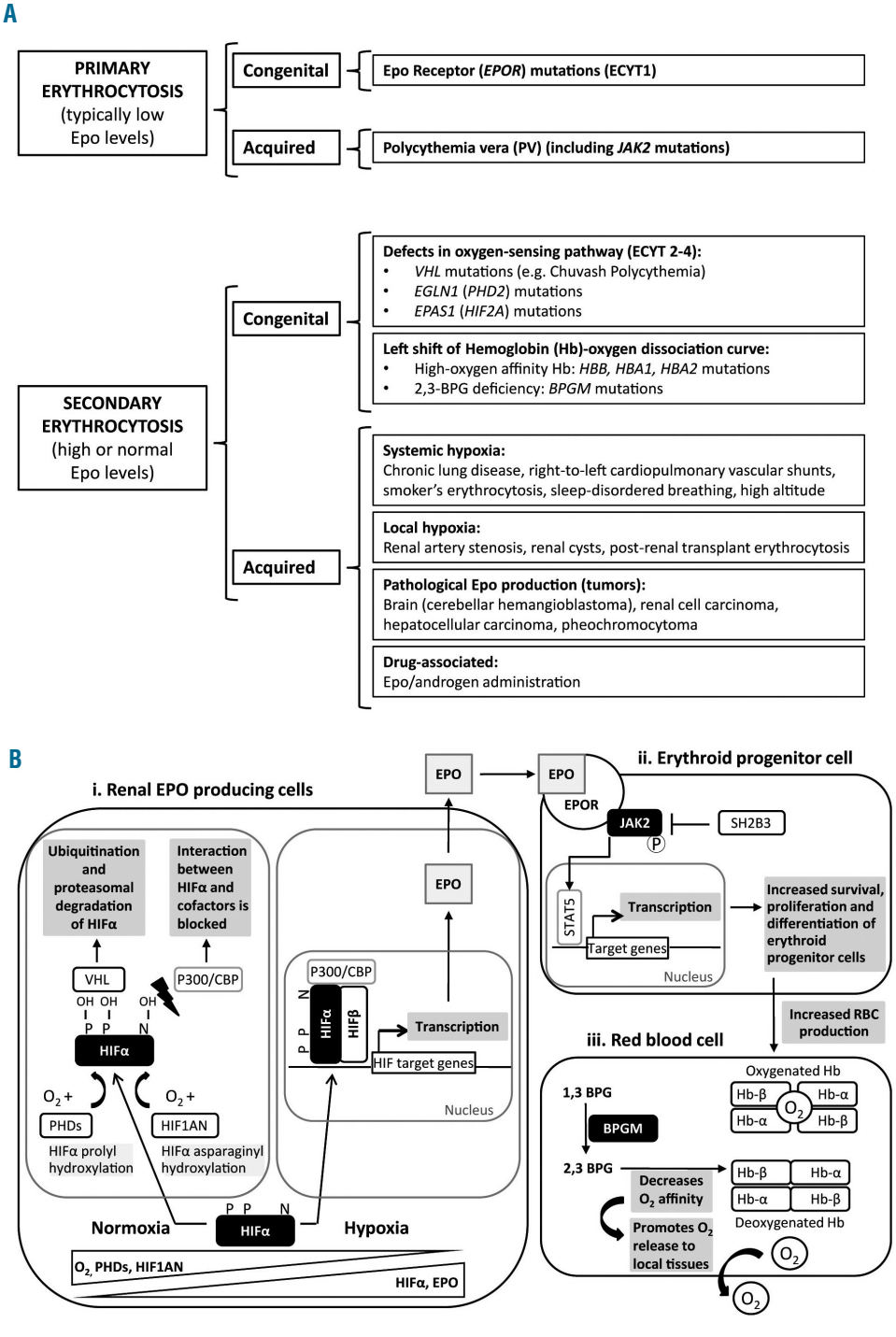

Figure 1. Classification and pathogenesis of erythrocytosis. (A) Causes of erythrocytosis. Erythrocytosis can be congenital or acquired. It is classified as primary, when there is an intrinsic defect in erythropoietic cells and erythropoietin (Epo) levels are low, or secondary, when the increased red cell production is externally driven through increased EPO production and EPO levels are high or inappropriately normal. Note: in this article, the term erythrocytosis rather than polycythemia is used consistently throughout (B) Pathways involved in the pathogenesis of erythrocytosis. (i) Hypoxia inducible factor (HIF) oxygen sensing pathway in renal EPO-producing cells. HIF are dimeric transcription factors composed of one α- and one β- subunit. In normoxia, HIFα subunits are hydroxylated by oxygen-dependent prolyl-hydroxylases (PHD) and asparaginyl hydroxylase (HIF1AN). The hydroxylated prolines (P) are recognized 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, PHD and HIF1AN are less active, HIFα subunits stabilize and translocate into the nucleus where they interact with the HIFβ subunit and cofactors and initiate transcription of target genes, including *EPO* (ii) Erythropoiesis in the bone marrow. This is triggered by the binding of EPO to the EPO receptor (EPOR) located on the surface of erythroid progenitor cells and subsequent activation of the JAK2-signaling 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 influence 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*).

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

Candidate gene	Position	N. of exons	Transcript ID	Pathway	Candidacy	Inheritance
VHL	Chr3:10183319-10195354	3	NM_000551	Oxygen-sensing	Known erythrocytosis-causing variants	Recessive / compound heterozygous (based on reported cases)
EPAS1	Chr2:46524541-46613842	16	NM_001430	Oxygen-sensing	Known erythrocytosis-causing variants	Dominant (based on reported cases)
EGLN1	Chr1:231499497-231560790	4	NM_022051	Oxygen-sensing	Known erythrocytosis-causing variants	Dominant (due to haploinsufficiency, based on reported cases)
HIF1A	Chr14:62162119-62214977	15	NM_001530	Oxygen-sensing	Key gene of the HIF pathway	Unknown (no cases reported). Likely dominant by function similarity to *EPAS1*
HIF3A	Chr19:46800303-46846690	13	NM_022462	Oxygen-sensing	Key gene of the HIF pathway	Unknown (no cases reported). Likely dominant by function similarity to *EPAS1*
EGLN2	Chr19:41305048-41314346	5	NM_053046	Oxygen-sensing	Key gene of the HIF pathway	Unknown (no cases reported). Likely dominant by function similarity to *EGLN1*
EGLN3	Chr14:34393421- 34420284	5	NM_022073	Oxygen-sensing	Key gene of the HIF pathway	Unknown (no cases reported). Likely dominant by function similarity to *EGLN1*
HIF1AN	Chr10:102295641-102313681	6	NM_017902	Oxygen-sensing	Key gene of the HIF pathway	Unknown (no cases reported). No function similarity to any known associated gene
EPO	Chr7:100318423-100321323	5	NM_000799	Erythropoiesis/ oxygen-sensing	1. Key gene in erythropoiesis 2. Identified in WGS500	Dominant (based on WGS500 variant), but cannot discard other patterns of inheritance
EPOR	Chr19:11487881-11495018	8	NM_000121	Erythropoiesis	Known erythrocytosis-causing variants	Dominant (based on reported cases)
JAK2	Chr9:4985245-5128183	25	NM_004972	Erythropoiesis	Known erythrocytosis-causing variants	Somatic (based on reported cases)
SH2B3	Chr12:111843752-111889427	8	NM_005475	Erythropoiesis	Known erythrocytosis-causing variants	Dominant (somatic or germline, based on reported cases)
BPGM	Chr7:134331531-134364567	3	NM_001724	Oxygen transport	1. Known erythrocytosis-causing variants 2. Identified in WGS	Dominant / compound heterozygous / recessive (based on reported cases)
HBB	Chr11: 5246696-5248301	3	NM_000518	Oxygen transport/ hemoglobin synthesis	Known erythrocytosis-causing variants	Dominant (based on reported cases)
HBA1	Chr16:226650-227521	3	NM_000558	Oxygen transport/ hemoglobin synthesis	Key gene in oxygen transport	Dominant (based on reported cases)
HBA2	Chr16:222846-223709	3	NM_000517	Oxygen transport/ hemoglobin synthesis	Key gene in oxygen transport	Dominant (based on reported cases)
KDM6A	ChrX:44732421-44971857	29	NM_021140	Oxygen-regulated demethylase	Identified in WGS500	Recessive X-linked inheritance (based on WGS500 variant)
GFI1B	Chr9:135854098-135867084	11	NM_004188	Erythropoiesis	Identified in WGS500	Recessive (based on WGS500 variant), but cannot discard other patterns of inheritance
BHLHE41	Chr12:26272959-26278003	9	NM_030762	Factor associated with HIF	Identified in WGS500	Recessive (based on WGS500 variant), but cannot discard other patterns of inheritance
OS9	Chr12:58087738-58115340	15	NM_001261421	Factor associating with HIF	Related to the HIF pathway	Unknown (no cases reported). No function similarity to any known associated gene
ZNF197	Chr3: 44666511-44689963	5	NM_006991	Factor associating with HIF	Related to the HIF pathway	Unknown (no cases reported). No function similarity to any known associated gene

Official gene symbols according to the 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).

Results

Novel candidate genes and variants were identified by whole-genome sequencing

The whole genomes of a small number of idiopathic erythrocytosis cases strongly suspected of having a genetic cause were sequenced as part of the WGS500 project. Candidate variants were found in novel genes, not previously associated with erythrocytosis: *EPO*, *GFI1B*, *KDM6A* and *BHLHE41*. Details of the rationale and criteria used to select these genes as candidates are given in the *Online Supplementary Information* and *Online Supplementary Table S2*. On this basis, these genes were included in the next-generation sequencing 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 using the gene panel (125 undiagnosed patients, 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 below 20X across samples, indicating a general poor amplification of these regions within the highly-multiplexed reactions (*Online Supplementary Table S3*). Ten of these (1.6% of all amplicons) had complete failure (coverage <20X in all samples), probably due to sequence context issues. The sequencing was, therefore, generally successful across samples, with a high percentage of the target sequence included at a good depth for germline variant calling.

We compiled a list of all known erythrocytosis-associated variants from the literature,[2,3] 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 performed well. The two *VHL* missense variants – c.235C>T and c.311G>T – fall within an amplicon in exon 1 that showed complete failure and would not, therefore, be detected.

Importantly, our panel reliably detected ten 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 (*Online Supplementary Table S4*).

Fifty-one 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

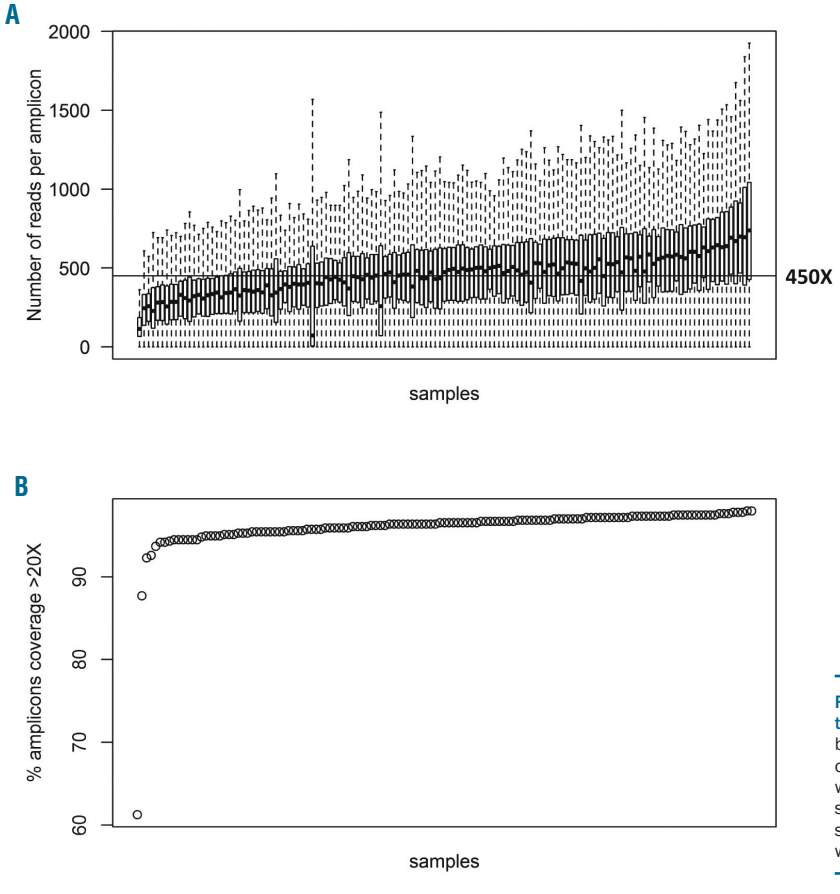

Figure 2. Coverage of the amplicons generated by the erythrocytosis gene panel across 135 samples. (A) Each boxplot represents the distribution of the number of reads obtained for all the amplicons generated by the panel within each sample. The horizontal line across the plot shows the average coverage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

regions of the genes examined, of which 19 were insertions or deletions (INDEL), 49 non-synonymous single nucleotide variations (SNV) and 30 synonymous SNV (Figure 3). None of the synonymous SNV is predicted to alter splicing according to Human Splicing Finder, NetGene2 and FSPLICE. We, therefore, focused on variants resulting in protein sequence alterations: following Sanger sequencing, 17 out of the 19 INDEL appeared to be false positives but two were confirmed. All 49 non-synonymous SNV were confirmed, although for one SNV there was a single base discrepancy: Ion Torrent detected a triple base change (CAA>ATT) in exon 12 of *JAK2* (chr9:5070025-5070027) but only a double change (AA>TT, chr9:5070026-5070027) was confirmed by Sanger sequencing. As a result, a total of 51 variants (49 SNV, 2 INDEL) were detected (*Online Supplementary Table S5*). Therefore, 57 out of 125 cases had at least one exonic variant (45.6%); of those, 38 patients had only one exonic variant detected (30.4%), while 19 had more than one (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 a control. For this, we examined the variant calls from the 1000G project after integrating both exome and low coverage data across 1041 individuals and extracted the SNV identified within the coordinates of the amplicons generated by our gene panel. We found that of the 49 non-synonymous SNV 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 at similar or higher frequencies (Fisher exact test and Benjamini and Hochberg false discovery correction[18]) (Figure 3). Those 19 SNV (*Online Supplementary Table S6*) are thus unlikely to be disease-causing mutations and most likely represent polymorphisms.

Out of the 30 uniquely identified variants in our cohort of patients, ten had been 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* or *ZNF197*.

Novel genes and variants identified by the erythrocytosis gene panel

The 22 novel variants (20 SNV and 2 INDEL) identified (Table 3) are extremely rare: nine were absent from both the dbSNP142 and Exome Aggregation Consortium (ExAC) databases, the latter containing data from 60,706 unrelated individuals; eight were reported only in ExAC at extremely low allele frequencies (≤0.0007), and only five were reported in both databases at very low allele frequencies (≤0.005).

Fourteen of these novel or very rare variants were found in known erythrocytosis-associated genes, such as *VHL*, *EPAS1*, *JAK2*, *SH2B3* (*LNK*), *EGLN1* and *BPGM*. Some of these 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 oxygen-dependent degradation domain (ODD). Furthermore, it is

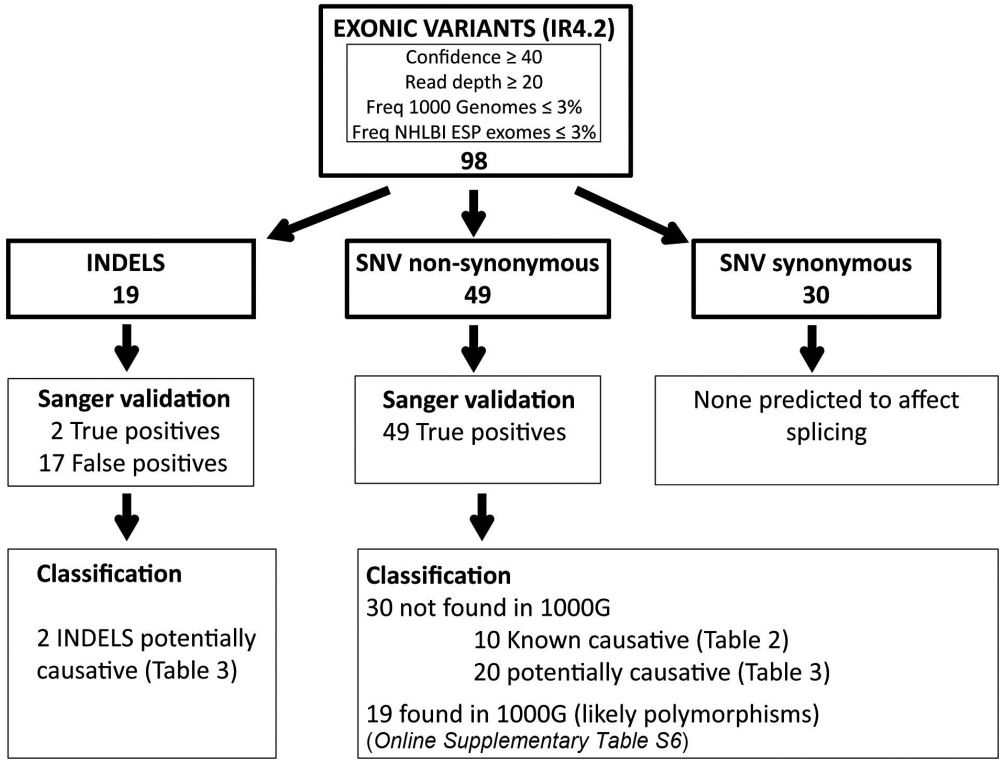

Figure 3. Overview of the exonic variants detected with Ion Torrent sequencing among 125 patients with erythrocytosis, their validation and further classification.

part of a six-residue domain which is highly conserved both across all HIFα isoforms and across species and which interacts with the VHL complex.[19] Thus, this mutation likely interferes with hydroxylation of HIF2α by prolyl hydroxylases and binding to the VHL complex, leading to upregulation of erythropoietin. It was found in two related patients, father and son, both of whom had idiopathic erythrocytosis with raised erythropoietin levels, and was, therefore, inherited in an autosomal dominant manner. *EGLN1* p.L279P affects a conserved residue, previously reported as altered (p.L279Tfs43, a frameshift variant) in a patient with erythrocytosis.[20] Structurally, this

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

Genomic location	Gene	cDNA/protein change	Genotype	N. of cases	Patient information	Type of erythrocytosis	Mechanism of action	Previous publication
chr2:46607420 G>A	*EPAS1*	c.G1609A p.G537R	Het	1	Female; age at diagnosis, 13 y; Hb, 196 g/L; Hct, 59.1%; Epo, 7.5 mIU/mL; chronic headache; pulmonary hypertension	Secondary	Gain of function of HIF2A	Percy *et al.* 2008[31], Gale *et al.* 2008[32]
chr3:10191578 C>G	*VHL*	c.C571G p.H191D	Hom	1	Male; age at diagnosis, 12 y; Hb, 154 g/L; Hct, 59%; Epo, 23 mIU/mL	Secondary	Loss of function (enhances HIF regulated gene expression)	Tomasic *et al.* 2013[33]
chr3:10191605 C>T	*VHL*	c.C598T p.R200W	Het*	4	Patient 1: Male; age at diagnosis, 47 y; Hb, 182 g/L; Hct, 54%; Epo, 8 mIU/mL; no family history Patient 2: Female; age at diagnosis, 48 y; Hb, 199 g/L; Hct, 67%; Epo, 22 mIU/mL; no family history Patient 3: Male; age at diagnosis, 1 y; Hb, 179 g/L; Hct, 54%; Epo, 60 mIU/mL; brother of patient 4 Patient 4: Male; age at diagnosis, 2 y; Hb, 146 g/L; Hct, 44.4%; Epo, high; brother of patient 3	Secondary	Loss of function (decreased HIF binding & hydroxylation, enhances HIF-regulated gene expression)	Ang *et al.* 2002[41]
chr9:5070026 AA>TT	*JAK2*	c.1615_1616invAA p.K539L	Het	1	Male, age at diagnosis, 35 y; Hb, 155 g/L; Hct, 52%; RBC, 6.37x10^{12}/L; WBC and platelets, normal range; Epo, 5-46 mIU/mL (after venesection); first presentation with large stroke; splenomegaly; BM biopsy, erythropoietic hyperplasia; JAK2 V617F negative	Primary	Gain of function of JAK2 (K539L)	Scott *et al.* 2007[5]
chr11:5246832 T>G	*HBB*	c.A440C p.H147P	Het	1	Female; age at diagnosis, 27 y; Hb, 173 g/L; Hct, 52.4%; Epo, 24 mIU/mL; family history, two brothers, mother, grandfather and great grand-mother affected by erythrocytosis (maternal line)	Secondary	High oxygen affinity Hb (Hb York)	Misgeld *et al.* 2001[37]
chr11:5246840 G>C	*HBB*	c.C432G p.H144Q	Het	1	Male; age at diagnosis, 34 y; Hb, 200 g/L; Hct, 58%; Epo, 7.5-27 mIU/mL; asymptomatic	Secondary	High oxygen affinity Hb (Hb Little Rock)	Bromberg *et al.* 1973[36], Wajcman *et al.* 1996[38]
chr11:5246944 C>T	*HBB*	c.G328A p.V110M	Het	1	Male; age at diagnosis, 49 y; Hb, 181 g/L; Hct, 53%; Epo, 25 mIU/mL	Secondary	High oxygen affinity Hb (Hb San Diego)	Wajcman *et al.* 1996[38], Gonzalez *et al.* 2009[35]
chr11:5247816 C>G	*HBB*	c.G306C p.E102D	Het	1	Male; age at diagnosis, 25 y; Hb, 204 g/L; Hct, 58%; Epo, 5.5 mIU/mL	Secondary	High oxygen affinity Hb (Hb Potomac)	Charache *et al.* 1978[34]
chr12:111856571 G>C	*SH2B3*	c.G622C p.E208Q	Het	1	Male; age at diagnosis, 14 y; Hb, 210 g/L; Hct, not available; Epo, 20.2 mIU/mL (after venesection)	Primary	Enhances JAK2 signaling	Spolverini *et al.* 2013[39]
chr12:111885310 G>A	*SH2B3*	c.G1198A p.E400K	Het	1	Male; age at diagnosis, 40 y; Hb, 189 g/L; Hct, 52.7%; RBC, 5.78x10^{12}/L; Epo, 2.5 mIU/mL	Primary	Interacts with JAK2 signaling	McMullin *et al.* 2011[40], Spolverini *et al.* 2013[39]

*Official gene symbols according to the HUGO Gene Nomenclature Committee are given here. Other gene symbols used frequently in the literature are: HIF2A (EPAS1), LNK (SH2B3). Chr : chromosome; Het: heterozygous; Hom: homozygous; y: years; Hb: hemoglobin; Hct: hematocrit; WBC: white blood cell count; RBC: red blood cell count; BM: bone marrow. Typical normal ranges: Hb, 130-180 g/L (adult males) and 115-155 g/L (adult females); Hct, 45-52% (adult males) and 37-48% (adult females); RBC, 4.7-6.1 x10^{12}/L (adult males) and 4.2-5.4 x10^{12}/L (adult females); Epo, 3.3-15.8 mIU/mL (adult males and females, although this range can vary between laboratories). *This variant causes Chuvash polycythemia in the homozygous state. In one of the patients, this variant was discovered in this study, whereas in the other three it had been detected in previous genetic tests.*

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

Genomic location	Gene	cDNA/ protein change	Genotype	N. of cases	Patient information	Type of erythrocytosis	SIFT/ Polyphen/ Provean	Allele frequency dbSNP142 ExAc	DNA studies in family members	Evidence of causality
chr1:231556799 A>G	*EGLN1*	c.T836C p.L279P	Het	1	Male; age at diagnosis, 52 y; Hb, 198 g/L; Hct, 61.2%; RBC, 6.42x10^{12}/L; Epo, 12.2 mIU/mL; headaches and dizziness; family history	Putative secondary	D/D/D	Not found Not found	Not available	Predicted structural/ functional effects
chr2:46607405 T>C	*EPAS1*	c.T1594C p.Y532H	Het	2	Patient 1: Male; age at diagnosis, 12 y; Hb, 190 g/L; Hct, 54%; Epo, not available; clinically well; no pulmonary hypertension; family history, father with congenital erythrocytosis (Patient 2) Patient 2: Male; age at diagnosis, 42 y	Putative secondary	D/D/D	Not found Not found	Variant present in both affected father and son	Predicted structural/ functional effects & segregation
chr3:10183685 G>T	*VHL*	c.G154T p.E52X	Het	1	Male; age at diagnosis, 28 y; Hb, 184 g/L; Hct, 54.6%; Epo, not available; headaches and dizziness; family history, affected brother	Putative secondary	T/NA/NA	Not found 4.16E-05	Not available	Predicted structural/ functional effects
chr7:100319185 TC>T	*EPO*	c.19delC p.P7fs	Het	1	Female; age at diagnosis, 3 y; Hb, 194 g/L; Hct, 58%; Epo, 4.1 mIU/mL; asymptomatic; family history, affected father and paternal grandmother with high Hb and Hct	Putative secondary	NA/NA/NA	Not found Not found	Variant present in father, who has high Hb and Hct	Segregation
chr2:46574031 AAGG>A	*EPAS1*	c.47delAGG p.del17E	Het	1	Male; age at diagnosis, 47 y; Hb, 186 g/L; Hct, 58.6%; Epo, normal range; headaches and dizziness	Putative secondary	NA/NA/D	Not found Not found	Not available	1. Deleterious by at least two prediction tools 2. Most in known erythrocytosis-causing genes 3. Most not found in large population databases
chr9:5050747 A>T	*JAK2*	c.A530T p.E177V	Het	1	Male; age at diagnosis, 14 y; Hb, 158 g/L; Hct, 52%; Epo, 7.6 mIU/mL; normal liver and spleen; normal cardiopulmonary function; no family history	Putative primary	D/D/D	Not found Not found	Not available	
chr12:111856181 G>A	*SH2B3*	c.G232A p.E78K	Het	1	Female; age at diagnosis, not reported (current age 85 y); Hb, 196 g/L; Hct, 57%; Epo, 12.2 mIU/mL; normal liver and spleen size; no family history	Putative primary	D/P/D	Not found 7.40E-04	Not available	
chr12:111884812 G>A	*SH2B3*	c.G901A p.E301K	Het	1	Male; age at diagnosis, 46 y; Hb, 187 g/L; Hct, 52%; RBC, 6.16x10^{12}/L; WBC and platelets, normal range; Epo, 10 mIU/mL; fatigue; mild splenomegaly; BM biopsy, increased erythropoiesis and slight dyserythropoiesis, normal megakaryopoiesis, no myeloproliferative neoplasia; no family history	Putative primary	D/D/Neutral	3.20E-05 3.30E-05	Not available	

Continued on the next page

continued from the previous page

Genomic location	Gene	cDNA/ protein change	Genotype	N. of cases	Patient information	Type of erythrocytosis	SIFT/ Polyphen/ Provean	Allele frequency dbSNP142 ExAc	DNA studies in family members	Evidence of causality
chr12:111885466 C>T	*SH2B3*	c.C1243T p.R415C	Hom	1	Male; age at diagnosis, 18 y; Hb, 188 g/L; Hct, 57%; Epo, not available; splenomegaly; no family history	Putative primary	D/D/D	1.00E-03 4.19E-05	Not available	
chr19:41313427 G>T	*EGLN2*	c.G1139T p.R380L	Het	1	Male; age at diagnosis, 16 y; Hb, 183 g/L; Hct, 53.7%; Epo, 2.8 mIU/mL; family history	Putative secondary	D/D/D	Not found Not found	Not available	
chr2:46611651 T>C	*EPAS1*	c.T2465C p.M822T	Het	1	Female; age at diagnosis, 9 y; Hb, 162 g/L; Hct, 48%; Epo, 5.8 mIU/mL; no family history	Putative secondary	D/B/Neutral	Not found 8.24E-06	Not available	Extremely rare variants
chr3:10183605 C>T	*VHL*	c.C74T p.P25L	Het	2	Patient 1: Male; age at diagnosis, 21 y; Gitelman syndrome (with positive *SLC12A3* mutation); Patient 2: Male; age at diagnosis, 15 y; Hb, 190 g/L; Hct, 55%; Epo, not available; mild headaches; family history, father also affected	Putative secondary	D/B/Neutral	4.00E-04 5.17E-03	Not available	
chr7:134346563 C>A	*BPGM*	c.C304A p.Q102K	Het	1	Male; age at diagnosis, 52 y; Hb, 186 g/L; Hct, 52.5%; Epo, normal range; myocardial infarction; no family history	Putative secondary	D/B/Neutral	Not found Not found	Not available	
chr9:5022168 G>A	*JAK2*	c.G181A p.E61K	Het	1	Male; age at diagnosis, 41 y; Hb, 172 g/L; Hct, 53%; WBC and platelets, normal range; Epo, 10-27 mIU/mL (while venesected); no family history	Putative primary	T/B/Neutral	Not found Not found	Not available	
chr9:5054775 G>C	*JAK2*	c.G827C p.G276A	Het	1	Female; age at diagnosis, 33 y; Hb, 172 g/L; Hct, 53.3%; Epo, 6 mIU/mL; headaches; dizziness; family history, affected brother	Putative primary	T/B/Neutral	Not found 8.29E-06	Not available	
chr12:58109559 G>A	*OS9*	c.G497A p.G166D	Het	1	Male; age at diagnosis, 37 y; Hb, 188 g/L; Hct, 51.8%; Epo, 8.24 mIU/mL	Unknown	T/D/D	Not found 3.42E-05	Not available	
chr19:46811511 A>C	*HIF3A*	c.A190C p.I64L	Het	1	Female; age at diagnosis, 19 y; Hb, 183 g/L; Hct, 58%; Epo, not available	Putative secondary	D/B/Neutral	Not found 1.65E-05	Not available	
chr19:46823777 C>A	*HIF3A*	c.C896A p.A299D	Het	1	Male; age at diagnosis, 53 y; Hb, 188 g/L; Hct, 58%; Epo, 23 mIU/mL; coronary heart disease	Putative secondary	T/B/Neutral	0.00066 7.93E-04	Not available	
chr9:5072561 G>A	*JAK2*	c.G1711A p.G571S*	Het	1	Male; age at diagnosis, 4 y; Hb, 180 g/L; Hct, not available; Epo, 12.4 mIU/mL	Unknown	T/D/D	7.4 E-04 4.81E-04	Variant present in non-affected parent	Unlikely disease-causing in erythrocytosis
chr12:26276001 A>C	*BHLHE41*	c.T447G p.F149L	Het	1	Male; age at diagnosis, 23 y; Hb, 183 g/L; Hct, 57%; Epo, not available	Unknown	T/B/Neutral	Not found 1.54E-04	Variant present in non-affected paternal aunt; absent in non-affected mother and sibling	

Continued on the next page

continued from the previous page

Genomic location	Gene	cDNA/ protein change	Genotype	N. of cases	Patient information	Type of erythrocytosis	SIFT/ Polyphen/ Provean	Allele frequency dbSNP142 ExAc	DNA studies in family members	Evidence of causality
chr7:100320336 A>G	*EPO*	c.A296G p.E99G**	Het	1	Male; age at diagnosis, 13 y; Hb, 149 g/L; Hct, 49.6%; Epo, 7.5 mIU/mL; subsequent red cell mass measurement negative for absolute erythrocytosis despite high Hb	Not absolute erythrocytosis	D/D/D	Not found	Not available	
chr7:100320290 G>C	*EPO*	c.G250C p.G84R***	Het	2	Patient 1: Male; age at diagnosis, 12 y; Hb, 190 g/L; Hct, 54%; Epo, not available; clinically well; no pulmonary hypertension; family history, father with congenital erythrocytosis (Patient 2) Patient 2: Male; age at diagnosis, 42 y	Putative secondary	T/D/Neutral	Not found 8.04E-05	Variant present in both affected father and son	

*Official gene symbols according to the 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: chromosome; Het: heterozygous; Hom: homozygous; y: years; Hb: hemoglobin; Hct: hematocrit; WBC: white blood cell count; RBC: red blood cell count; BM: bone marrow; D: deleterious (applicable to SIFT and Provean predictions) and probably damaging (applicable to Polyphen2 HDIV predictions); T: tolerated by SIFT; P: possibly damaging by Polyphen2 HDIV; B: benign by Polyphen2 HDIV and NA, non-applicable. Typical normal ranges: Hb, 130-180 g/L (adult males) and 115-155 g/L (adult females); Hct, 45-52% (adult males) and 37-48% (adult females); RBC, 4.7-6.1x10^{12}/L (adult males) and 4.2-5.4 x10^{12}/L (adult females); Epo: 3.3-15. 8 mIU/mL (adult males and females, although this range can vary between laboratories). ExAc: Exome Aggregation Consortium, Cambridge, MA, USA (http://exac.broadinstitute.org). All variants except the ones included in the "unlikely disease-causing in erythrocytosis" section have been submitted to a dedicated database (www.erythrocytosis.org). * JAK2 p.G571S has been reported previously in myeloproliferative disorders[47] but it is thought to be a silent non-functioning polymorphism; ** predicted deleterious but although patient has high Hb and Hct was subsequently found to have normal red cell mass. *** Likely to be non-pathogenic: predicted benign/tolerated and present in patients with an identified variant in EPAS1 with very high likelihood of causality.*

residue is located on helix 3, which interacts with both N-terminal and C-terminal ODD hydroxylation domains on HIFα;[21] a proline substitution may affect protein stability and diminish ODD binding, reducing HIFα hydroxylation. The *VHL* p.E52X variant introduces a stop codon, predicting translation termination of the long VHL isoform (p30) while allowing translation only 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 and have been associated with either pheochromocytomas (codon 25 and 38) or with von Hippel-Landau (VHL) disease (p.E46X and p.E52K).[22-24] 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 to undergo investigations for the presence of VHL disease; there is evidence that erythrocytosis is seen in about 5-20% of patients with VHL disease.[25] For the remaining variants, most were classified as deleterious by either SIFT, PolyPhen2 or Provean (Table 3), with a high degree of agreement between tools, so further investigations are needed to elucidate their functional impact.

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* (Table 3). In addition, novel variants were also found in *EPO* and *BHLHE41*, two genes without previous genetic association with erythrocytosis which were revealed by WGS500. For *EPO*, the most striking variant found is a frameshift, p.P7fs, detected in a heterozygous state 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 a heterozygous state in the patient's father who also has high hematocrit and hemoglobin levels. Two other *EPO* SNV were detected in other patients but these are most likely very rare polymorphisms (Table 3). Regarding *BHLHE41*, the novel missense variant identified (p.F149L) is classified as benign by Provean, PolyPhen2 and SIFT and is thus unlikely to be pathogenic, a notion supported by segregation analysis in the patient's family (Table 3).

Discussion

The technical progress in next-generation sequencing, together with the increasing understanding of the biological pathways underlying the pathogenesis of erythrocytosis, provide new opportunities to advance the genetic investigation of patients with erythrocytosis.

Our approach allowed the creation of a next-generation sequencing targeted gene panel with the capacity to process a large group of 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 in ten positive controls. It was excellent at reliably calling SNV, with all SNV identified subsequently validated in all samples by Sanger sequencing. Nevertheless, a few limitations are recognized and should be taken into account when considering its future applications. 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, as previously reported by other Ion

Torrent sequencing users.[26-28] 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.[26,28] Another limitation of the panel – related to the nature of its technology – is that it can only identify SNV and short INDEL but not other structural variants such as large INDEL or copy number variations. Also, variant detection in genes with high sequence similarity such as *HBA1* and *HBA2* is challenging and caution is needed for variant calling.

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. The procedures employed at different Centers vary (*Online Supplementary Figure S1*), but as a general rule if the patient's erythropoietin level is low, variants in genes involved in erythropoiesis (*EPOR*, *JAK2*) are screened for. If the patient's erythropoietin level is high or normal, the P50 (partial pressure of oxygen at which 50% of hemoglobin is saturated with oxygen) is calculated and if low, hemoglobin 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.[2,29,30]

Using our gene panel we were able to provide definitive genetic diagnoses in nine patients whose mutations had been previously missed. For example, a variant in *EPAS1*, p.G537R – a well-described gain-of-function mutation found in erythrocytosis patients[31,32] – was detected. This was previously missed because the patient was not screened for *EPAS1* variants, owing to the fact that the erythropoietin level was not high enough (and investigations were thus directed to a different branch of the diagnostic algorithm). Similarly, we identified a homozygous *VHL* variant (p.H191D) known to cause erythrocytosis.[33] 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).[34-38] These were missed previously, either because conventional screening with hemoglobin electrophoresis can miss hemoglobinopathies[38] or because of difficulties in 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.[5,39,40] The patient with variant *JAK2* p.K439L, originally classified as having idiopathic erythrocytosis as the conventional criteria for polycythemia vera, including *JAK2* p.V617F screening, were not met, should now be considered as having polycythemia vera with a *JAK2* exon 12 mutation. As highlighted in previous studies,[5,6] the clinical picture of this subtype of polycythemia vera is indistinguishable from that of idiopathic erythrocytosis. This emphasizes that *JAK2* exon 12 mutations should be actively screened for in patients with idiopathic erythrocytosis. Furthermore, the finding of *SH2B3* variants highlights that this gene should also be surveyed, which is currently not done routinely. The erythrocytosis gene panel can successfully do both. 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, four out of the 125 patients were heterozygous for *VHL* p.R200W. In the homozygous state, this variant causes Chuvash polycythemia.[41,42] Congenital erythrocytosis also occurs in patients who are compound heterozygotes,[43-45] but heterozygous carriers are usually unaffected. Nevertheless, *VHL* p.R200W heterozygous mutations feature significantly more frequently in erythrocytosis databases[3] than in general populations,[46] suggesting a causal role for this mutation. For one of the four patients here, the variant was newly identified. For the other three, previous genetic tests had also identified it. Thus, within this study we aimed to detect additional genetic changes that might explain the patients' clinical phenotype. We did not detect any other variants within *VHL*, except for two single nucleotide polymorphisms in the 3' untranslated region with high minor allelic frequencies (≥0.35 in dbSNP142). Alternatively, the co-occurrence of this heterozygous variant with another heterozygous variant in a separate gene of the same biological pathway could act in synergy to produce disease. We did not obtain conclusive evidence for this in our four patients: two did not have an additional variant and in the other two, the *VHL* p.R200W co-occurred with heterozygous missense variants classified as polymorphisms (*Online Supplementary 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.

As this research panel provides full-gene sequencing instead of specific mutation screening, it allowed the detection of 22 novel variants. For some of these, there is a strong likelihood of causality, based on the location of the mutated residues on functional or regulatory domains and the expected disturbance they would cause to protein structure and function (as explained in the Results section for *EGLN1* p.L279P, *EPAS1* p.Y532H and *VHL* p.E52X), and based on genetic evidence of familial segregation (e.g. *EPAS1* p.Y532H and *EPO* p.P7fs, which are dominantly inherited). For other variants, mostly found in known erythrocytosis-associated genes, there is strong consensus in the *in silico* prediction of deleterious effect, whereas for some there is less evidence of functional candidacy (Table 3). While the functional significance of newly identified variants cannot currently be confirmed – and indeed clinical causation cannot be concluded – future functional studies and screening of larger cohorts of erythrocytosis patients are needed to replicate the findings and to provide further evidence of causality.

In this study we explored some genes, not previously associated with erythrocytosis, because of their involvement in the HIF pathway or their discovery through WGS500 as potential candidates. Candidate variants were found in *EGLN2* and *HIF3A* but not 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. The precise WGS-identified variants in *EPO*, *GFI1B*, *KDM6A* and *BHLHE41* were not found in this cohort of 125 cases, suggesting that larger cohorts of patients 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 identi-

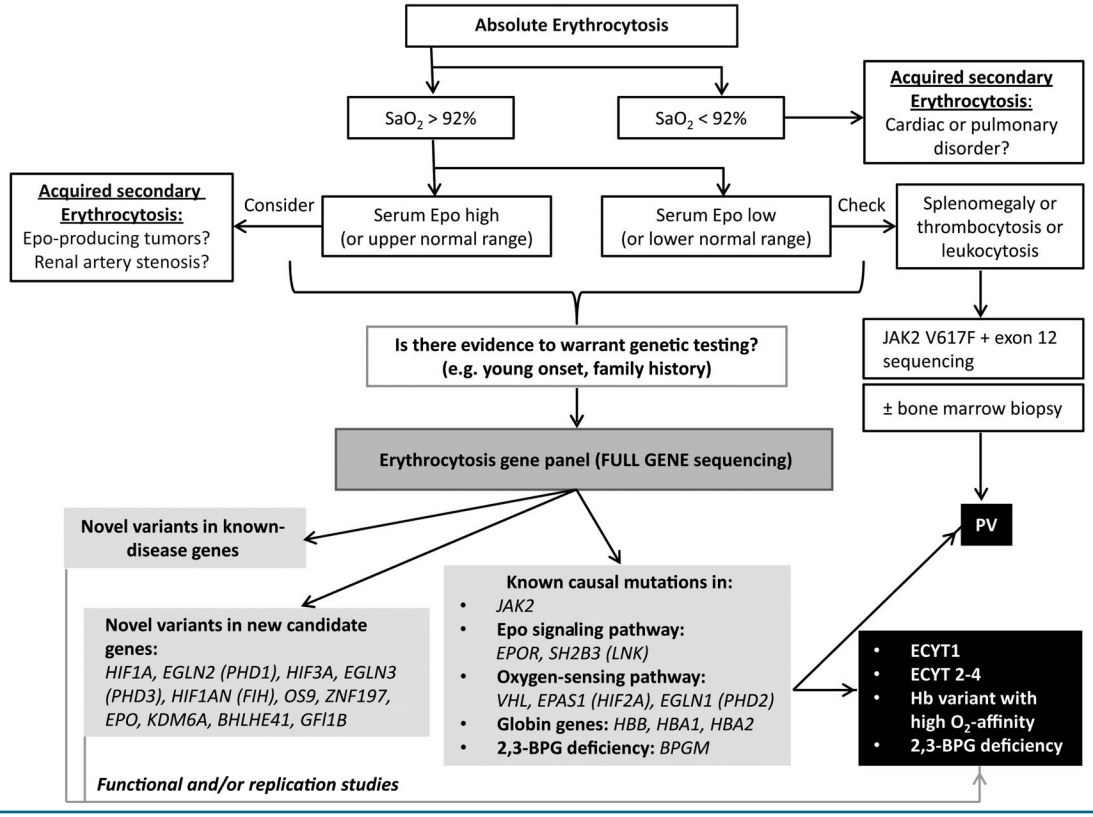

Figure 4. Proposed use of a gene panel in the investigation of erythrocytosis. A gene panel would make genetic testing more efficient and streamlined. It enables the simultaneous survey of the full length of 21 candidate genes, in a systematic and unbiased manner, allowing the detection of known causal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

fied suggesting that *EPO* should be actively surveyed as an erythrocytosis-associated candidate gene. Accrued use of the panel in further patients will provide insight into which novel genes play a role in erythrocytosis and will allow refinement of any future diagnostic panels.

One limitation of our study is the lack of DNA from a source other than blood to determine germline or somatic status. This would only be a concern for *JAK2* and *SH2B3*, in which somatic mutations are associated with polycythemia vera and myeloproliferative diseases. When variants in *JAK2* and *SH2B3* are found by the panel, further studies in skin/nail DNA are probably warranted. For all other genes, variants detected in blood with the panel are most likely germline. While somatic mutations in *EPAS1* can be found in tumors of patients with erythrocytosis,[8] these would not be detectable in blood with our methodology.

Thus, despite the few technical limitations described, the erythrocytosis gene panel is useful in the genetic investigation of patients with erythrocytosis from a research perspective. Furthermore, following appropriate optimization and refinement, gene panel sequencing has the potential to improve the diagnostic work-up of erythrocytosis patients in clinical practice. A point to note is that the gene panel in our study was applied to a highly-selected group of patients who had undergone significant clinical and genetic "filtering" (*Online Supplementary 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 to "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. In conclusion, we hope to demonstrate the immediate utility of a targeted gene panel in the investigation of erythrocytosis at a time when next-generation sequencing is revolutionizing clinical medicine.

Acknowledgments

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.

References

1. McMullin MF. The classification and diagnosis of erythrocytosis. Int J Lab Hematol. 2008;30(6):447-459.
2. Hussein K, Percy M, McMullin MF. Clinical utility gene card for: familial erythrocytosis. Eur J Hum Genet. 2012;20(5):
3. Bento C, Percy MJ, Gardie B, et al. Genetic basis of congenital erythrocytosis: mutation update and online databases. Hum Mutat. 2014;35(1):15-26.
4. Percy MJ, Jones FG, Green AR, Reilly JT, McMullin MF. The incidence of the JAK2 V617F mutation in patients with idiopathic erythrocytosis. Haematologica. 2006;91(3): 413-414.
5. Scott LM, Tong W, Levine RL, et al. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. N Engl J Med. 2007;356(5):459-468.
6. Percy MJ, Scott LM, Erber WN, et al. The frequency of JAK2 exon 12 mutations in idiopathic erythrocytosis patients with low serum erythropoietin levels. Haematologica. 2007;92(12):1607-1614.
7. Lee FS, Percy MJ, McMullin MF. Oxygen sensing: recent insights from idiopathic erythrocytosis. Cell Cycle. 2006;5(9):941-945.
8. Zhuang Z, Yang C, Lorenzo F, et al. Somatic HIF2A gain-of-function mutations in paraganglioma with polycythemia. N Engl J Med. 2012;367(10):922-930.
9. McMullin MF. Idiopathic erythrocytosis: a disappearing entity. Hematology Am Soc Hematol Educ Program. 2009;629-635.
10. Percy MJ, Butt NN, Crotty GM, et al. Identification of high oxygen affinity hemoglobin variants in the investigation of patients with erythrocytosis. Haematologica. 2009;94(9):1321-1322.
11. 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. Am J Hematol. 2004;75(4):205-208.
12. Petousi N, Copley RR, Lappin TR, et al. Erythrocytosis associated with a novel missense mutation in the BPGM gene. Haematologica. 2014;99(10):e201-204.
13. Gruber M, Hu CJ, Johnson RS, Brown EJ, Keith B, Simon MC. Acute postnatal ablation of Hif-2alpha results in anemia. Proc Natl Acad Sci USA. 2007;104(7):2301-2306.
14. Rankin EB, Biju MP, Liu Q, et al. Hypoxia-inducible factor-2 (HIF-2) regulates hepatic erythropoietin in vivo. J Clin Invest. 2007;117(4):1068-1077.
15. Webb JD, Coleman ML, Pugh CW. Hypoxia, hypoxia-inducible factors (HIF), HIF hydroxylases and oxygen sensing. Cell Mol Life Sci. 2009;66(22):3539-3554.
16. Taylor JC, Martin HC, Lise S, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet. 2015;47(7):717-726.
17. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
18. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57(1):289-300.
19. 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;296(5574):1886-1889.
20. Jang JH, Seo JY, Jang J, et al. Hereditary gene mutations in Korean patients with isolated erythrocytosis. Ann Hematol. 2014;93(6): 931-935.
21. Chowdhury R, McDonough MA, Mecinovic J, et al. Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. Structure. 2009;17(7):981-989.
22. van der Harst E, de Krijger RR, Dinjens WN, et al. Germline mutations in the vhl gene in patients presenting with phaeochromocytomas. Int J Cancer. 1998;77 (3):337-340.
23. Olschwang S, Richard S, Boisson C, et al. Germline mutation profile of the VHL gene in von Hippel-Lindau disease and in sporadic hemangioblastoma. Hum Mutat. 1998;12(6):424-430.
24. Dollfus H, Massin P, Taupin P, et al. Retinal hemangioblastoma in von Hippel-Lindau disease: a clinical and molecular study. Invest Ophthalmol Vis Sci. 2002;43(9): 3067-3074.
25. Friedrich CA. Genotype-phenotype correlation in von Hippel-Lindau syndrome. Hum Mol Genet. 2001;10(7):763-767.
26. Costa JL, Sousa S, Justino A, et al. Nonoptical massive parallel DNA sequencing of BRCA1 and BRCA2 genes in a diagnostic setting. Hum Mutat. 2013;34(4):629-635.
27. Junemann S, Sedlazeck FJ, Prior K, et al. Updating benchtop sequencing performance comparison. Nat Biotechnol. 2013;31(4):294-296.
28. 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.
29. Cario H, McMullin MF, Bento C, et al. Erythrocytosis in children and adolescents-classification, characterization, and consensus recommendations for the diagnostic approach. Pediatr Blood Cancer. 2013;60 (11):1734-1738.
30. Bento C, Almeida H, Maia TM, et al. Molecular study of congenital erythrocytosis in 70 unrelated patients revealed a potential causal mutation in less than half of the cases (Where is/are the missing gene(s)?). Eur J Haematol. 2013;91(4):361-368.
31. Percy MJ, Furlow PW, Lucas GS, et al. A gain-of-function mutation in the HIF2A gene in familial erythrocytosis. N Engl J Med. 2008;358(2):162-168.
32. 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;112 (3):919-921.
33. Tomasic NL, Piterkova L, Huff C, 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;98 (4):560-567.
34. Charache S, Jacobson R, Brimhall B, et al. Hb Potomac (101 Glu replaced by Asp): speculations on placental oxygen transport in carriers of high-affinity hemoglobins. Blood. 1978;51(2):331-338.
35. Gonzalez Fernandez FA, Villegas A, Ropero P, et al. Haemoglobinopathies with high oxygen affinity. Experience of Erythropathology Cooperative Spanish Group. Ann Hematol. 2009;88(3):235-238.
36. Bromberg PA, Alben JO, Bare GH, et al. High oxygen affinity variant of haemoglobin Little Rock with unique properties. Nat New Biol. 1973;243(127):177-179.
37. Misgeld E, Gattermann N, Wehmeier A, Weiland C, Peters U, Kohne E. Hemoglobinopathy York [beta146 (HC3) His==>Pro]: first report of a family history. Ann Hematol. 2001;80(6):365-367.
38. Wajcman H, Galacteros F. Abnormal hemoglobins with high oxygen affinity and erythrocytosis. Hematol Cell Ther. 1996;38(4): 305-312.
39. Spolverini A, Pieri L, Guglielmelli P, et al. Infrequent occurrence of mutations in the PH domain of LNK in patients with JAK2 mutation-negative 'idiopathic' erythrocytosis. Haematologica. 2013;98(9):e101-102.
40. McMullin MF, Wu C, Percy MJ, Tong W. A nonsynonymous LNK polymorphism associated with idiopathic erythrocytosis. Am J Hematol. 2011;86(11):962-964.
41. Ang SO, Chen H, Hirota K, et al. Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia. Nat Genet. 2002;32(4):614-621.
42. Smith TG, Brooks JT, Balanos GM, et al. Mutation of von Hippel-Lindau tumour suppressor and human cardiopulmonary physiology. PLoS Med. 2006;3(7):e290.
43. Bento MC, Chang KT, Guan Y, et al. Congenital polycythemia with homozygous and heterozygous mutations of von Hippel-Lindau gene: five new Caucasian patients. Haematologica. 2005;90(1):128-129.
44. Cario H, Schwarz K, Jorch N, et al. Mutations in the von Hippel-Lindau (VHL) tumor suppressor gene and VHL-haplotype analysis in patients with presumable congenital erythrocytosis. Haematologica. 2005;90(1):19-24.
45. Percy MJ, McMullin MF, Jowitt SN, et al. Chuvash-type congenital polycythemia in 4 families of Asian and Western European ancestry. Blood. 2003;102(3):1097-1099.
46. Liu E, Percy MJ, Amos CI, et al. The worldwide distribution of the VHL 598C>T mutation indicates a single founding event. Blood. 2004;103(5):1937-1940.
47. Panovska-Stavridis I, Eftimov A, Pivkova-Veljanovska A, Ivanovski M, Cevreska L, Dimovski AJ. Familiar JAK2 G571S variant not linked with essential trombocythemia. Blood. 2014:124(21);558.