bjh research paper

# Gene expression profiling in polycythaemia vera: overexpression of transcription factor NF-E2

**Philipp S. Goerttler,[1,2] Clemens Kreutz,[3] Johannes Donauer,[4] Daniel Faller,[3] Thomas Maiwald,[3] Edith März,[1] Brigitta Rumberger,[4] Titus Sparna,[4] Annette Schmitt-Gräff,[5] Jochen Wilpert,[4] Jens Timmer,[3] Gerd Walz[4] and Heike L. Pahl[1]**

*[1]Department of Experimental Anaesthesiology, [2]Department of Biology, University Hospital Freiburg, Center for Clinical Research, [3]Freiburg Center for Data Analysis and Modelling, Department of Mathematics and Physics, University of Freiburg, and [4]Renal Division, Department of Medicine, University Hospital Freiburg, [5]Department of Pathology, University Hospital Freiburg, Freiburg, Germany*

Received 25 November 2004; accepted for publication 24 January 2005

Correspondence: Heike L. Pahl, Department of Experimental Anaesthesiology, University Hospital Freiburg, Centre for Clinical Research, Breisacher Str. 66, 79106 Freiburg, Germany. E-mail: Heike.Pahl@klinikum.uni-freiburg.de

## Summary

The molecular aetiology of polycythaemia vera (PV) remains unknown and the differential diagnosis between PV and secondary erythrocytosis (SE) can be challenging. Gene expression profiling can identify candidates involved in the pathophysiology of PV and generate a molecular signature to aid in diagnosis. We thus performed cDNA microarray analysis on 40 PV and 12 SE patients. Two independent data sets were obtained: using a two-step training/validation design, a set of 64 genes (class predictors) was determined, which correctly discriminated PV from SE patients. Separately 253 genes were identified to be upregulated and 391 downregulated more than 1·5-fold in PV compared with healthy controls ($P < 0{\cdot}01$). Of the genes overexpressed in PV, 27 contained Sp1 sites: we therefore propose that altered activity of Sp1-like transcription factors may contribute to the molecular aetiology of PV. One Sp1 target, the transcription factor NF-E2 [nuclear factor (erythroid-derived 2)], is overexpressed 2- to 40-fold in PV patients. In PV bone marrow, NF-E2 is overexpressed in megakaryocytes, erythroid and granulocytic precursors. It has been shown that overexpression of NF-E2 leads to the development of erythropoietin-independent erythroid colonies and that ectopic NF-E2 expression can reprogram monocytic cells towards erythroid and megakaryocytic differentiation. Transcription factor concentration may thus control lineage commitment. We therefore propose that elevated concentrations of NF-E2 in PV patients lead to an overproduction of erythroid and, in some patients, megakaryocytic cells/platelets. In this model, the level of NF-E2 overexpression determines both the severity of erythrocytosis and the concurrent presence or absence of thrombocytosis.

**Keywords:** polycythaemia vera, secondary erythrocytosis, microarrays, NF-E2, transcription factors.

Polycythaemia vera (PV) is one of four diseases termed the myeloproliferative disorders (MPDs) (Dameshek, 1951). Besides PV this group includes essential thrombocythaemia (ET), idiopathic myelofibrosis (IMF) and chronic myeloid leukaemia (CML). The four diseases share several clinical features (Michiels *et al*, 1999; Tefferi, 1999). All arise clonally from the mutation of a single pluripotent haematopoietic stem cell (Adamson *et al*, 1976; Jacobson *et al*, 1978; Fialkow *et al*, 1980, 1981) and can transform to acute leukaemia. Each MPD leads to the increased production of one or more haematopoietic elements. In PV, the predominant clinical feature is erythrocytosis, defined by an increased red cell mass (Bilgrami & Greenberg, 1995). Thromboembolic complications constitute the major cause of morbidity and mortality in PV and reduce life expectancy to a median of 15 years after diagnosis. However, the clinical course of individual PV patients is very heterogeneous and prolonged thrombosis-free survival is possible (Samuelsson, 1998).

Today, CML is regarded as a separate entity, defined by the 9;22(q34;q11) translocation, the 'Philadelphia chromosome' ($Ph^-$), which results in production of the Bcr/Abl fusion protein. The Philadelphia chromosome is not found in any of the three remaining subtypes of MPD (Carbonell *et al*, 1983;

doi:10.1111/j.1365-2141.2005.05416.x

Diez Martin *et al*, 1991). Consequently, they have been termed the 'Ph-negative-MPDs' (MPDs).

Contrary to CML, the Ph-negative MPDs do not carry specific or diagnostic molecular markers. Hence, both the diagnosis of PV, and the differential diagnosis to reactive, secondary erythrocytosis (SE) are accomplished by clinical parameters (Pearson, 1998; Murphy, 1999; WHO, 2001). The lack of molecular diagnostic tools causes inaccurate or even mis-diagnoses in a significant proportion of patients (Johansson *et al*, 2002). While novel molecular markers for the diagnosis of PV have recently been described, they have not yet been validated in prospective studies (Moliterno *et al*, 1998; Temerinac *et al*, 2000; Klippel *et al*, 2003).

To date, our inadequate understanding of the molecular aetiology of PV has prevented the development of rationally designed, targeted therapies. Current cytoreductive treatment options, which include hydroxyurea, anagrelide and interferon, are therefore non-specific and often cause considerable side effects and/or raise the risk of transformation to acute leukaemia (Solberg, 2002; Fruchtman, 2004; Spivak, 2004).

Gene expression profiling can identify a disease-specific molecular signature, to aid in diagnosis, and reveal candidates involved in the aetiology or pathophysiology of disease development. We therefore performed cDNA microarray analysis on a cohort of 40 PV and 12 SE patients. We used a two-step training/validation design to identify a set of class predictor genes, which represent a molecular signature for PV and distinguish PV from SE patients. In a separate analysis, candidate genes for the aetiology of disease development were identified. From these data, we propose a molecular model for the pathophysiology of PV.

## Patients and methods

### *Patients*

Peripheral blood samples from therapeutic phlebotomies were obtained from a total of 57 patients with PV and 12 patients with SE (of these 57, 52 patients, 40 PV and 12 SE, were analysed by cDNA microarray, the remainder were studied to confirm gene expression data by independent methods, see Table SI). Venous blood samples were anticoagulated with EDTA and shipped by courier to Freiburg without cooling. The maximum time interval between venepuncture and arrival in the laboratory was 24 h. All PV patients fulfilled the WHO criteria for diagnosis (WHO, 2001). All patients with SE displayed a documented cause for SE: chronic obstructive pulmonary disease (COPD, $n = 6$), reactive erythrocytosis [caused by very heavy smoking ($n = 3$), severe sleep apnoe ($n = 2$) or a haemoglobinopathy type Alberta, which leads to a compensatory erythrocytosis ($n = 1$)]. Patient characteristics are shown in Table SI. Buffy coats from 50 healthy volunteer blood donors were used to isolate the control granulocytes. The study protocol was approved by the local ethics committee and informed consent was obtained from all patients. Each patient was assigned a unique patient number (UPN), which was used thereafter for the protection of privacy.

### *Separation of cells*

Anticoagulated blood from patients and healthy controls was used as a source of granulocytes. Granulocytes were purified by dextran sedimentation followed by Ficoll–Paque (Pharmacia, Freiburg, Germany) separation (Kruisbeck *et al*, 1991). Erythrocytes were eliminated by hypotonic lysis (0·2% NaCl for 30 s). This method consistently yielded granulocyte preparations with greater than 98% purity as judged by visual inspection of Wright–Giemsa stained slide preparations.

### *RNA preparation*

For microarray analysis, freshly prepared granulocytes were homogenized in 4 mol/l guanidinium isothiocyanate containing 0·5% *N*-laurylsarcosine, 25 mmol/l sodium citrate and 0·72% beta-mercaptoethanol using a 20G syringe. Total RNA was subsequently purified by caesium chloride density gradient centrifugation. After ethanol precipitation, RNA was resuspended in double distilled (dd) $H_2O$ at a concentration of 0·4–4 μg/μl and stored at −80°C.

For polycythaemia rubra vera-1 (PRV-1) quantification, RNA was harvested using an acidic phenol extraction as previously described (Klippel *et al*, 2003).

### *PRV-1 Quantification*

PRV-1 mRNA levels were determined precisely as previously described (Klippel *et al*, 2003).

### *Microarrays*

cDNA microarrays were produced and processed according to the Stanford protocol described by Eisen and Brown (1999). A total of 7497 annotated genes from the RZPD (Resource Center and Primary Database, Berlin, Germany) were obtained as bacterial stocks. Plasmids were purified using the Qiagen 96-well Turbo Kit (Qiagen, Hilden, Germany), and inserts were purified by polymerase chain reaction (PCR) using vector primers flanking the individual inserts (5′-CTG CAA GGC GAT TAA GTT GGG TAA C-3′ and 5′-GTG AGC GGA TAA CAA TTT CAC ACA GGA AAC AGC-3′). PCR products were purified by ethanol precipitation and resuspended in $ddH_2O$. Aliquots were transferred into 384-well plates, dried, and resuspended in 3× standard saline citrate (SSC) or 10% dimethyl sulphoxide (DMSO) to a final concentration of approximately 40 ng/μl. Printing was performed on aminosilane-coated slides (CMT-GAP II Slides, Corning, NY, USA), using an arrayer that was assembled according to specifications by the Stanford group with software provided by J. de Risi (http://cmgm.stanford.edu/pbrown).

### *Hybridization*

A pool of RNA extracted from isolated granulocytes of 50 healthy controls was produced and used as a reference RNA in each hybridization. Each patient RNA was hybridized together with this control pool to two arrays. Duplicates were performed with dye-swap to control for possible differences in the incorporation rate of the two fluorochromes (first slide: patient cDNA labelled with Cy3; control cDNA labelled with Cy5; second slide: patient cDNA labelled with Cy5; control cDNA labelled with Cy3). For each slide, 12 μg of patient and control pool RNA, respectively, were reverse transcribed into cDNA in the presence of Cy3- or Cy5-labelled dUTP, using Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA). A PCR-purification kit (Qiagen) was used for cDNA purification after dye labelling.

The subsequent procedures were performed according to the published protocol (http://cmgm.stanford.edu/pbrown/mguide/index.html). Briefly, the microarray slides were prehybridized for 20 min at 65°C in prehybridization buffer [3·5X SSC; 0,1% sodium dodecyl sulphate (SDS); 10 mg/ml bovine serum albumin]. After rinsing with water, the slides were hybridized to 24 μg of labelled cDNA resuspended in hybridization solution (3·4× SSC, 0·3% SDS, 18 μg poly adenylic acid potassium salt, 18 μg tRNA, 18 μg Cot-1 DNA; Sigma, Steinheim, Germany) for 16–20 h at 62–65°C in chambers (Monterey Industries, Richmond, CA, USA). After hybridization the slides were washed for 2 min in 1X SSC and 0·03% SDS, followed by 5 min in 0·2X SSC. After a final washing step for 5 min in 0·05X SSC the slides were dried by centrifugation and stored in the dark until analysis.

### *Data analysis*

Signal intensities were measured by an Axon 4000A scanner using GenePix 3.0 software (Axon Instruments Inc., Union City, CA, USA). Image and data files, array layout, as well as all relevant information according to the MIAME guidelines [Minimum Information About a Microarray Experiment (Brazma *et al*, 2001)] were transferred into the GeneTraffic-Duo database (Microarray Data Management and Analysis Software, Iobion Informatics, LLC, USA). To exclude artefacts near the background range, all spots were eliminated when sample or reference intensity was less than 50 or less than the local background. Local background was subtracted from all spot intensities. Normalization was performed with the Lowess (Locally weighted scatter plot smoother) sub-grid normalization method (Workman *et al*, 2002). In contrast to global normalization, sub-grid normalization calculates the normalization factor for each of the 16 subgrids independently and is therefore relatively insensitive to local variations on the array (Workman *et al*, 2002).

For each spot, the logarithm to the base of two of the Cy3 to Cy5 ratio was computed (LogRatio).

For analysis of the genes discriminating between PV and SE, a two-sample *t*-test was used for a statistical analysis of differentially expressed genes after application of the above mentioned normalization and filtering criteria. To control for multiple testing the obtained *P*-values were adjusted by calculating the false discovery rate (fdr) using the method by Benjamini and Hochberg (1995).

Differential expression was defined by *P*-values below 0·01 (fdr). Agglomerative hierarchical clustering, introduced by Kaufman and Rousseeuw (1990) was performed using the R-statistical software package (http://www.r-project.org).

In order to confirm that the 10 PV patients randomly selected for the training set constituted a representative sample, expression of the resulting 64 class predictor genes was independently analysed in the remaining 30 PV patients. Fifty-three of the 64 genes (82·8%) were also differentially expressed in the 30 patients in the validation set ($P < 0{\cdot}05$, fdr corrected). This demonstrated that the choice of PV patients in the training set had little influence on our gene expression signature.

### *Northern blot analysis*

Ten μg of total RNA were used for Northern blots. After transfer to Hybond-N-Membranes and immobilization, the blots were hybridized in ExpressHyb Hybridization Solution (Clontech, Heidelberg, Germany) at 68°C. The probes were labelled using the Prime-It-II labelling kit (Stratagene, Amsterdam, The Netherlands) and alpha-$^{32}$P-dCTP (Amersham, Freiburg, Germany). The blots were washed three times for 10 min in 2X SSC, 0·05% SDS at room temperature and twice at 50°C for 20 min in 0·1X SSC, 0·1% SDS. Hybridization was detected by autoradiography.

### *Quantitative reverse transcription (RT)-PCR assays*

Quantitative RT-PCR experiments were performed using assay on demand (Applied Biosystems, Foster City, CA, USA) products for gene expression analysis:

- Human NF-E2 Assay on demand (no. Hs00232351_m1).
- Human CAECAM-1 Assay on demand (no. Hs00236077_m1).
- Human Thrombomodulin Assay on demand (no. Hs00264920_s1).
- Human Pim-1 Assay on demand (no. Hs00171473_m1).
- Human 18S Pre-Developed TaqMan Assay Reagents (4310893E).

Reverse Transcription of 50 ng of total granulocyte RNA was performed using the TaqMan Reverse Transcription Kit (Applied Biosystems). 5 μl of a 1:4 dilution of these cDNAs were included into each Quantitative RT-PCR measurement.

All measurements were performed in triplicate according to the manufacturers instructions in an ABi Prism® 7000 Cycler (Applied Biosystems, Foster City, CA, USA). A cDNA standard curve, in which copy numbers had previously been determined by quantification with a plasmid standard, was included in each experiment. Using this reference, gene expression was

determined and is reported in copy number per 1 000 000 copies of 18S rRNA.

### *Western Blot*

Total granulocyte cell extracts were prepared using a high-salt detergent buffer (Totex) as previously described (Pahl *et al*, 1996). Cell extracts (100 μg) were boiled in Laemmli sample buffer and subjected to SDS-polyacrylamide gel electrophoresis and Western blotting as described (Pahl *et al*, 1996). Primary polyclonal antibodies against NF-E2 [nuclear factor (erythroid-derived 2)], were a kind and generous gift of Dr E. Bresnick. The rabbit serum was used at a dilution of 1:2000. Bound antibody was decorated with peroxidase-conjugated secondary antibody (donkey anti-rabbit IgG; Amersham, Freiburg, Germany). The immunocomplexes were detected using enhanced chemiluminescence Western blotting reagents (Amersham). Exposure to Kodak XAR-5 films was performed for 5–10 s.

### *Immunohistochemistry*

Representative pretreatment trephine biopsies were obtained from six PV patients, diagnosed according to the WHO criteria, and three healthy donors. Samples were fixed in 4% buffered formaldehyde, decalcified in 10% buffered-EDTA, pH 7·2 and paraffin embedded. Heat retrieval of deparaffinized tissue sections cut at 4 μm was performed in a pressure cooker using a target retrieval solution at pH 6 (S1699; Dako, Hamburg, Germany). Affinity purified rabbit anti-NF-E2 antibodies E19X at a 1:1000 dilution and H-230X at a 1:500 dilution (Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA) were used as primary antibodies and detected by the alkaline-phosphatase-anti-alkaline phosphatase method according to the manufacturers' instructions (Dako). To control for non-specific binding the first antibodies were omitted.

## Results

In order to identify candidate genes for the molecular aetiology of PV and to define a PV specific gene expression signature, cDNA microarray analysis was conducted. A control pool was generated consisting of RNA from isolated granulocytes of 50 healthy volunteers. RNA isolated from purified granulocytes of individual PV ($n = 40$) or SE ($n = 12$) patients, labelled with one fluorochrome, together with this control pool, labelled with a second fluorochrome, was hybridized to cDNA arrays (Fig S1). Each patient was analysed in duplicate, using a dye-swap to control for differences in fluorochrome incorporation. By separate analyses, two different data sets were obtained (Fig S1). By comparing data in PV and SE patients, genes whose expression can differentiate the MPD from reactive erythrocytosis were determined. These genes may be considered a molecular signature for PV. Secondly, comparing gene expression in PV patients to healthy controls identified possible candidates contributing to PV pathophysiology.

### *A molecular signature for PV*

A two-step design was implemented to obtain a gene expression signature for PV. A training set of 10 PV and 10 SE patients was randomly chosen from the cohort. The following criteria were used to define class predictor genes, which discriminate the two forms of erythrocytoses:

(1) Candidate genes had to have yielded evaluable signals in 80% of the samples analysed (see *Patients and methods* for details on analyses of hybridization intensity and background measurements).
(2) Candidate genes must discriminate between the two entities with $P < 0{\cdot}01$ (corrected for the fdr, see *Patients and methods*).

Sixty-four genes fulfilling both criteria were identified and used to conduct a two-dimensional hierarchical clustering analysis of the training set samples (Fig 1A). The class predictor was able to discriminate PV from SE in 100% of the samples in the training set. This, however, is not sufficient to establish the specificity and sensitivity of class prediction. Validation of the class predictor for diagnostic purposes can only be achieved by testing it on an independent cohort of patients. We therefore performed hierarchical clustering analysis on an additional 32 patients (30 PV and two SE) as well as the 20 samples in the training cohort. The resulting dendrogram (Fig 1B) displays a 100% discrimination between PV and SE in this cohort of 52 patients. Hence the 64 genes identified here constitute a molecular signature for PV, one that is able to discriminate this MPD from secondary, reactive erythrocytosis.

### *Differential gene expression between PV and healthy controls*

Those genes identified as class predictors for PV *versus* SE, however, are not necessarily those that contribute to the pathophysiology of PV. On the contrary, it is quite likely that factors, which are physiologically upregulated in response to exogenous stimuli causing reactive erythrocytosis, are endogenously and malignantly activated in PV patients (Fig S1A). In this case, there would be no difference in gene expression between PV and SE patients, but the aberrant transcription in PV nonetheless would contribute to PV pathophysiology.

Hence, candidates whose altered expression may contribute to PV disease development, are most likely to be found among the genes that are differentially expressed between healthy controls and PV patients.

To identify such genes the following criteria were applied:

(1) Candidate genes must show a difference in expression between PV and healthy controls with $P < 0{\cdot}01$ (fdr corrected, see Patients and methods).
(2) Candidate genes had to have yielded evaluable signals in 80% of the samples analysed (see Patients and methods).
(3) Candidate genes must be up- or downregulated at least 1·5-fold.

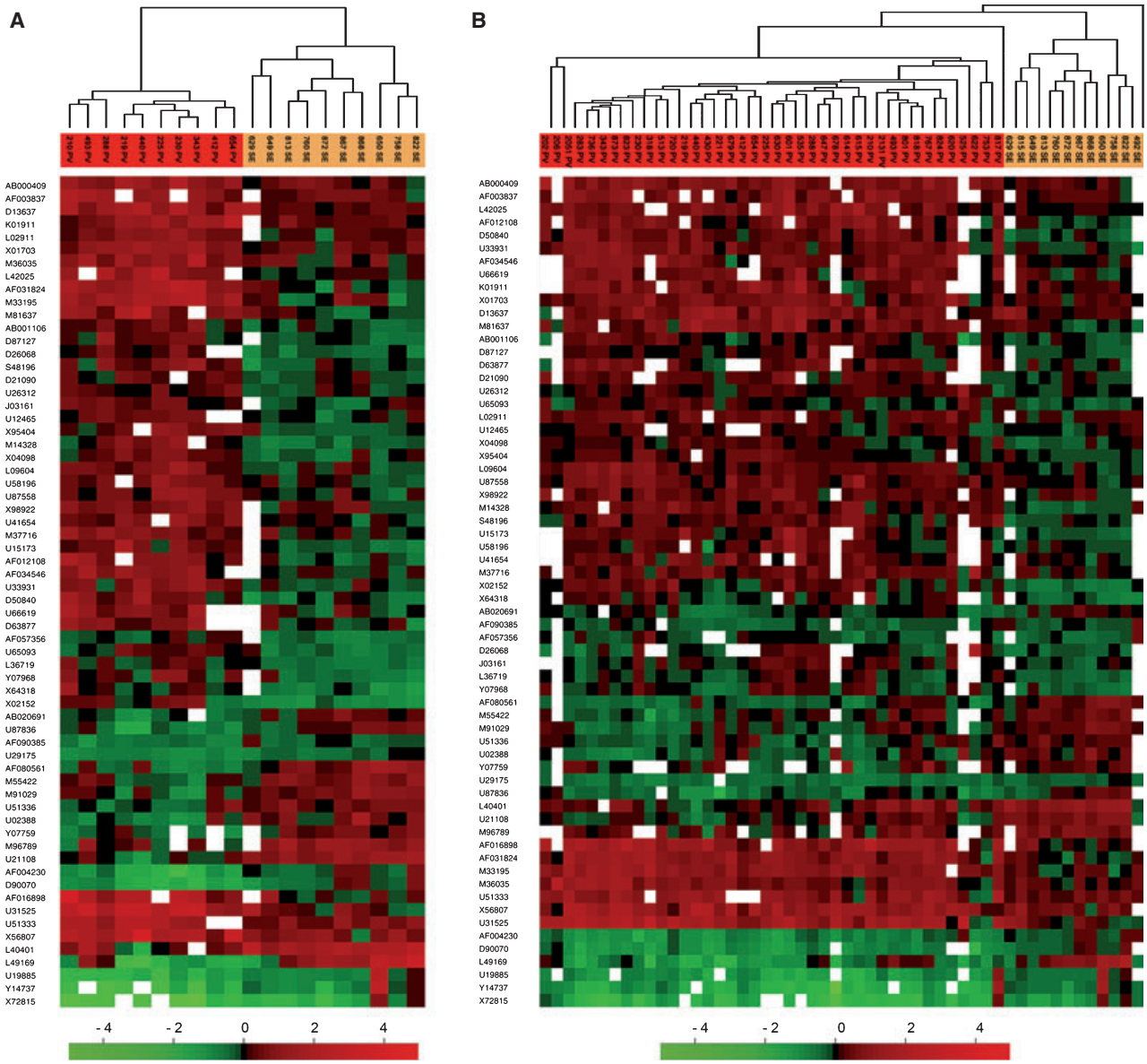


**Fig 1.** Gene expression profiling in patients with PV and SE. (A) Training Set. A training set of 10 PV and 10 SE patients was defined, randomly chosen from the cohort of available patients. A two-sample *t*-test was used to identify class predictor genes, which are differentially expressed between PV and SE. To control for multiple testing the obtained *P*-values were adjusted for the false discovery rate (fdr) (Benjamini & Hochberg, 1995). Differential expression was defined by $P < 0{\cdot}01$ (fdr corrected). Agglomerative hierarchical clustering was used to generate the dendrogram (Kaufman & Rousseeuw, 1990). The columns represent individual patients and the rows represent specific cDNAs, identified by GenBank accession number. Gene expression is depicted according to the colour scale shown below the figure. White squares depict missing data points. (B) Validation Set. The 64 class predictor genes defined in the training set were applied to the entire cohort of 40 PV and 12 SE patients. Agglomerative hierarchical clustering was used to generate the dendrogram (Kaufman & Rousseeuw, 1990). For detailed description of the dendrogram see (A) above.

Of the 7497 cDNA spots represented on the array, 2515 displayed a significantly different expression between PV and healthy controls ($P < 0{\cdot}01$ fdr corrected). Of these, 1191 were evaluable in at least 80% of the samples (32 of the 40 PV patients). In this group, 644 genes displayed a difference of at least 1·5-fold between PV and healthy controls. We felt that this latter criterion, which excluded 547 genes that display a significant and consistently observed difference between PV and healthy controls, albeit a very small one, was necessary in order to be able to confirm the results using independent methods such as quantitative RT-PCR or Western blot. However, we recognize that in some cases, a consistently present overexpression of 1·3-fold, for example, may carry physiological consequences. Such differences, however, are not analysable with the methods used in this study.

### *644 cDNAs are differentially expressed in PV versus healthy controls*

The 644 cDNAs identified as differentially expressed were annotated using the Gene Ontology Consortium database (http://www.geneontology.org). The results are summarized in Table SII, which depicts genes up- or downregulated in PV according to their biological function, as well as those genes, whose function is unknown. Of the 644 differentially expressed genes, 253 were upregulated and 391 were downregulated in PV patients.

### *Many genes upregulated in PV patients are controlled by Sp1-like transcription factors*

Seventeen of the 253 genes upregulated in PV patients contained Sp1 binding sites in their promoter. An additional 10 genes, consistently upregulated in PV patients ($P < 0{\cdot}01$), but not analysed in at least 80% of the samples were also regulated by Sp1. In 18 of the 27 genes (66%), the Sp1 site was demonstrated to be functional and critical for promoter activity, the rest were putative sites based on sequence homology (Fig 2).

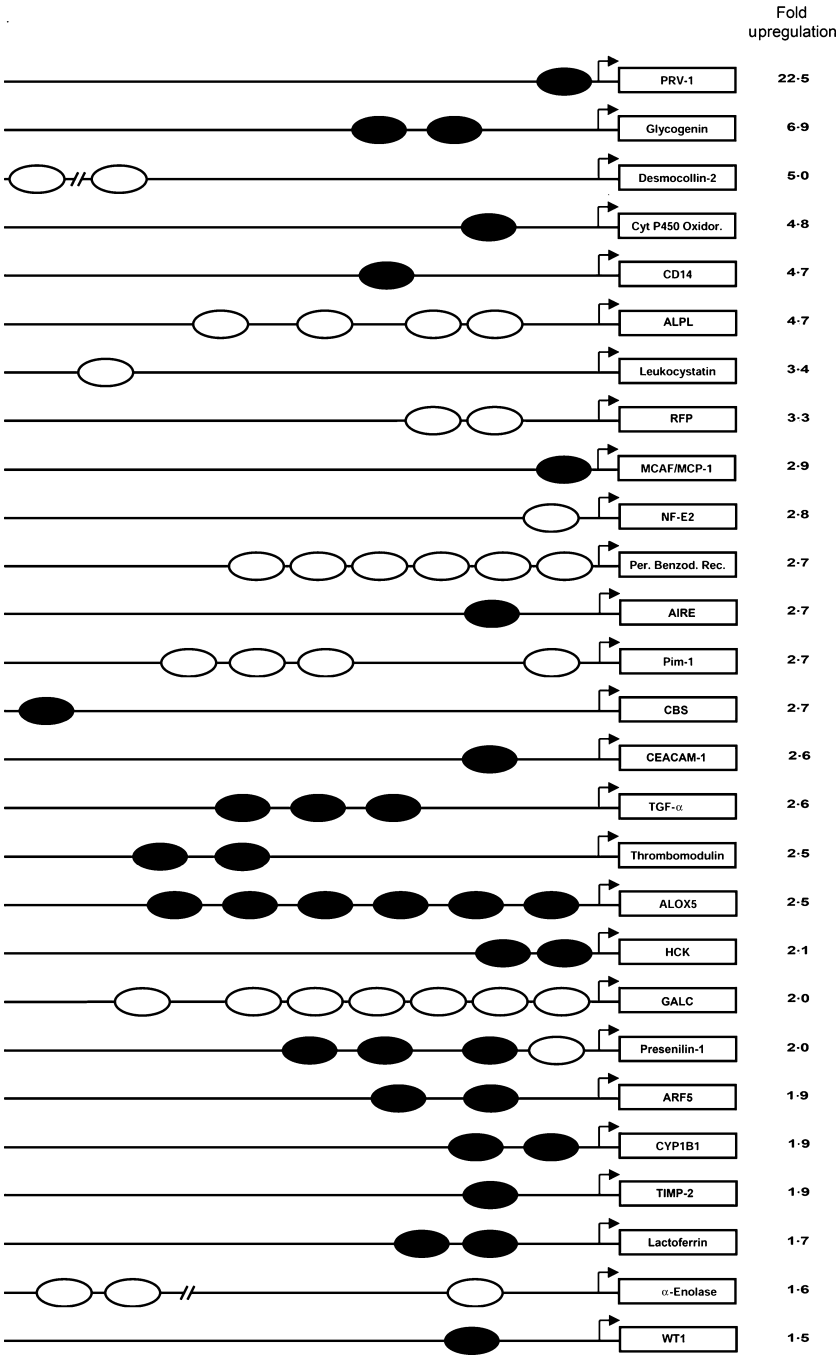


**Fig 2.** The 27 Sp1 target genes overexpressed in PV. Schematic representation of the Sp1 recognition sequences found in promoters of genes overexpressed in PV patients. Filled symbols denote sites, whose integrity is critical for promoter activity in functional assays (Hoshiko *et al*, 1990; Tazawa *et al*, 1993; Zhang *et al*, 1994; Chen *et al*, 1996; Cohen *et al*, 1997; Marsden *et al*, 1997; Wo *et al*, 1997; Hauses *et al*, 1998; Lebeda *et al*, 1999; van Maanen *et al*, 1999; Pastorcic & Das, 1999; Khanna-Gupta *et al*, 2000; O'Leary & Kasper, 2000; Strunck *et al*, 2000; Vyhlidal *et al*, 2000; Zhong *et al*, 2000; Nathanson *et al*, 2002; Maclean *et al*, 2004). Open symbols denote putative sites identified by sequence homology (Weiss *et al*, 1988; Giallongo *et al*, 1990; Meeker *et al*, 1990; Pischedda *et al*, 1995; Luzi *et al*, 1997; Iwata *et al*, 1999; Nathanson *et al*, 2002). Fold upregulation in PV *versus* healthy controls is depicted on the right.

### *Verification of differential expression by Northern blot and quantitative RT-PCR*

The overexpression of Sp1-regulated genes observed by microarray analysis was verified with several independent methods (Figs 3 and 4). Northern blots were used to confirm overexpression of the protease inhibitor leucocystatin (Fig 3A), the Ser/Thr kinase Pim-1 (Fig 3B) and the transcription factor NF-E2 (Fig 3C). More precise quantitation of the degree of overexpression was obtained through quantitative RT-PCR, conducted for NF-E2 (Fig 4A), Pim-1 (Fig 4B), the cell adhesion molecule CEACAM-1 (Fig 4C) and the thrombin cofactor thrombomodulin (Fig 4D). These four Sp1 target genes were overexpressed a median of 7-fold (NF-E2, range 2·3- to 40-fold), threefold (Pim-1, range 1·2- to 4·7-fold), 6·6-fold (CEACAM-1, range 4·1- to 11·1-fold) and fivefold (thrombomodulin, range 2- to 11·7-fold).

### *The transcription factor NF-E2 is overexpressed in granulocytes, erythroid and myeloid precursors as well as megakaryocytes in PV patients*

Of the Sp1 target genes overexpressed in PV patients, the transcription factor NF-E2 represented an especially promising candidate in the molecular aetiology of PV. NF-E2 is expressed in haematopoietic precursor cells as well as in those lineages affected by PV. In addition, ectopic expression of this factor can reprogram cells towards erythroid differentiation and can induce erythroid maturation in the absence of erythropoietin (Sayer *et al*, 2000).

We therefore sought to confirm NF-E2 overexpression in PV patients at the protein level. This is especially important in light of the observation that the *prv-1* mRNA is overexpressed 8- to 64-fold in PV patients compared with healthy controls, but PRV-1 protein levels are not different between PV patients and controls (Klippel *et al*, 2002, 2003). Western blot analysis of cellular extracts from purified granulocytes of 13 PV patients and four healthy controls showed that the NF-E2 protein was readily detectable in PV cells, but not in healthy control granulocytes (Fig 5A and data not shown). Hence, NF-E2 overexpression was manifested at the protein level.

Bone marrow immunohistochemistry was used to confirm NF-E2 protein overexpression and to determine whether other cell types, in addition to mature peripheral granulocytes, overexpress NF-E2. Two different antibodies, recognizing distinct epitopes at the N-terminus and the C-terminus, respectively, were used. Tissue sections of normal marrow showed no or very weak levels of immunoreactivity near the threshold of detection using both antibodies (Fig 5B, panels 1 and 3). By contrast, numerous clearly immunoreactive bone marrow cells were identified in PV patients (Fig 5B, panels 2 and 4). The pleiomorphic megakaryocytes of PV yielded a distinct nuclear NF-E2 protein accumulation in a pattern typical for a transcription factor. In addition, PV cells with morphological features of erythro- and myelopoiesis were immunolabelled by both NF-E2 antibodies. No specific nuclear labelling was seen in negative control sections. Immunohistochemistry hence depicts NF-E2 overexpression in megakaryocytes as well as in erythroid and myeloid precursors in PV bone marrow.

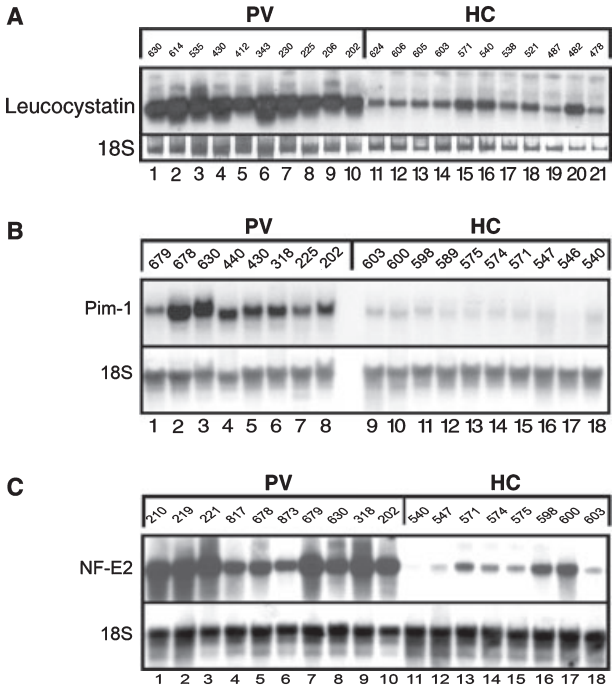


**Fig 3.** Validation of Sp1 target genes overexpressed in PV patients by Northern blot or semi-quantitative RT-PCR. RNA was isolated from purified granulocytes from the indicated PV patients and healthy controls (HC) and analysed by Northern blot. To control for equal amounts of RNA the housekeeping gene 18S rRNA was analysed in parallel, as indicated.

## Discussion

Gene expression profiling has successfully been used to achieve various goals. Firstly, the identification of molecular signatures may aid in disease diagnosis (Shaffer *et al*, 2001). Secondly, the recognition of distinct subgroups within a disease entity can facilitate classification (Yeoh *et al*, 2002). Thirdly, expression profiling can depict the molecular aetiology underlying disease development and thereby reveal potential drug targets (Savage *et al*, 2003).

Applying this technique to the diffuse large B-cell lymphomas (DLBCL) has led to the recognition of three distinct subgroups of disease, provided molecular signatures to diagnose these subgroups and culminated in the initiation of a phase II clinical trial using Velcade (bortezomib), an NF-κB inhibitor, whose utility was suggested by the identification of aberrant NF-κB activity in the ABC DLBCL subtype (Alizadeh *et al*, 2000; Rosenwald *et al*, 2003; Savage *et al*, 2003).

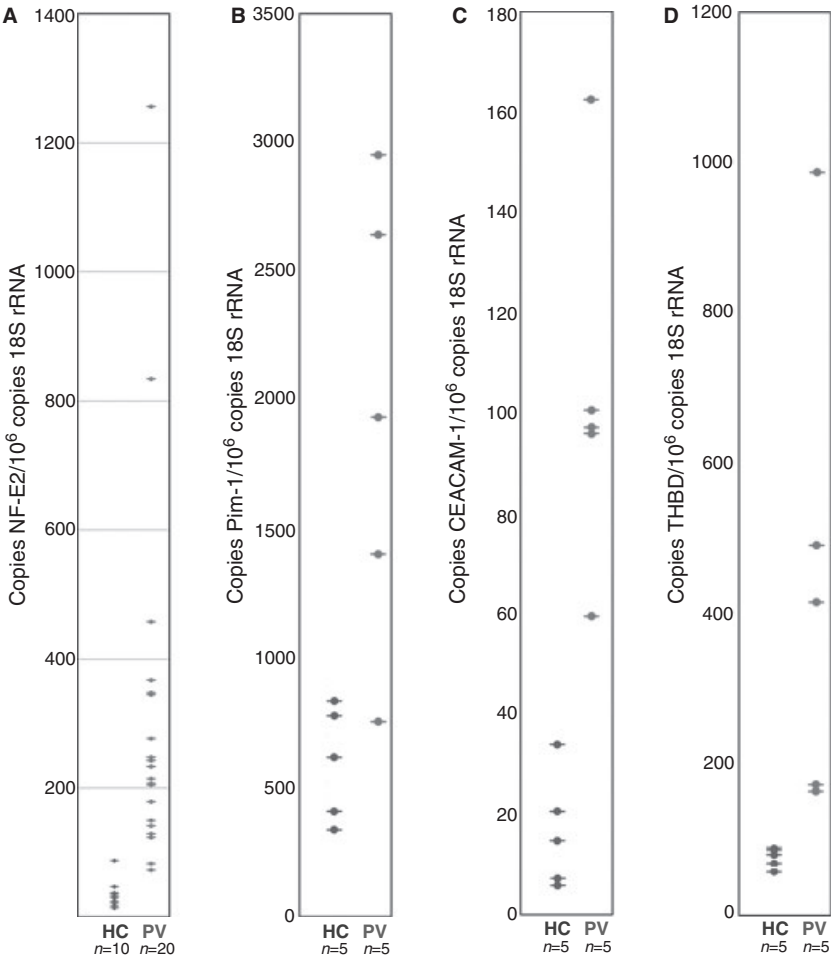


**Fig 4.** Validation of Sp1 target genes overexpressed in PV patients by quantitative RT-PCR. RNA was isolated from purified granulocytes of PV patients and healthy controls (HC) as indicated and subjected to quantitative RT-PCR analysis. A standard curve with known copy numbers of NF-E2, Pim-1, CEACAM-1 or THBD, respectively, and 18S rRNA was included on each plate. Sample copy numbers of target genes and 18S rRNA were determined from the standard curve and are expressed as relative ratios (molecules target gene per $10^6$ 18S molecules).

### *A molecular signature for PV*

Gaining similar knowledge about the transcriptional profile of PV could therefore greatly contribute to our comprehension of this disease entity. Here, we defined a molecular signature for PV composed of 64 genes, which correctly identified all PV patients in a validation cohort of 52 individuals (Fig 1B). The fact that SE patients are divided into two groups, patient 492 being the sole member of a separate subclassification, is most probably due to the fact that a large proportion of the 64 class predictors genes were not analysed in this individual (missing data, indicated by white squares, Fig 1B). In addition, SE can be caused by diverse physiological and pathophysiological conditions, which are unlikely to generate a homogenous pattern of gene expression.

Our gene expression signature differs from one recently published (Pellagatti *et al*, 2003) in that it can discriminate between PV and SE. Seven of the 11 genes proposed by Pellagatti *et al* (2003) do not discriminate between PV and SE as the authors did not include SE patients in their analysis. For the differential diagnosis of a patient presenting with erythrocytosis, a molecular signature that correctly discriminates primary from secondary polycythaemia is crucial.

Until the currently unknown disease-inducing alteration(s) can be detected directly, employing a molecular signature composed of 64 genes has distinct advantages over using a single marker for diagnosis. Epigenetic markers, while closely associated with disease development, will not invariably be present in all patients (Spivak, 2004). Hence, a larger collection of genes is more likely to accurately diagnose patients with unusual presentation than single markers. However, limited availability and high costs of transcriptional profiling currently prohibit the widespread use of this technique for routine diagnosis.

### *PV is molecularly homogeneous*

Because the clinical course of PV patients is heterogeneous and poorly predictable by clinical parameters, we used unsupervised hierarchical clustering to seek possible subtypes within the disease entity. The resulting dendrogram did not reveal any distinct subtypes, rather, gene expression appeared relatively homogeneous across all PV patients (data not shown). A gene expression profile common to all cases of PV was reminiscent of CLL, where all patients, although even

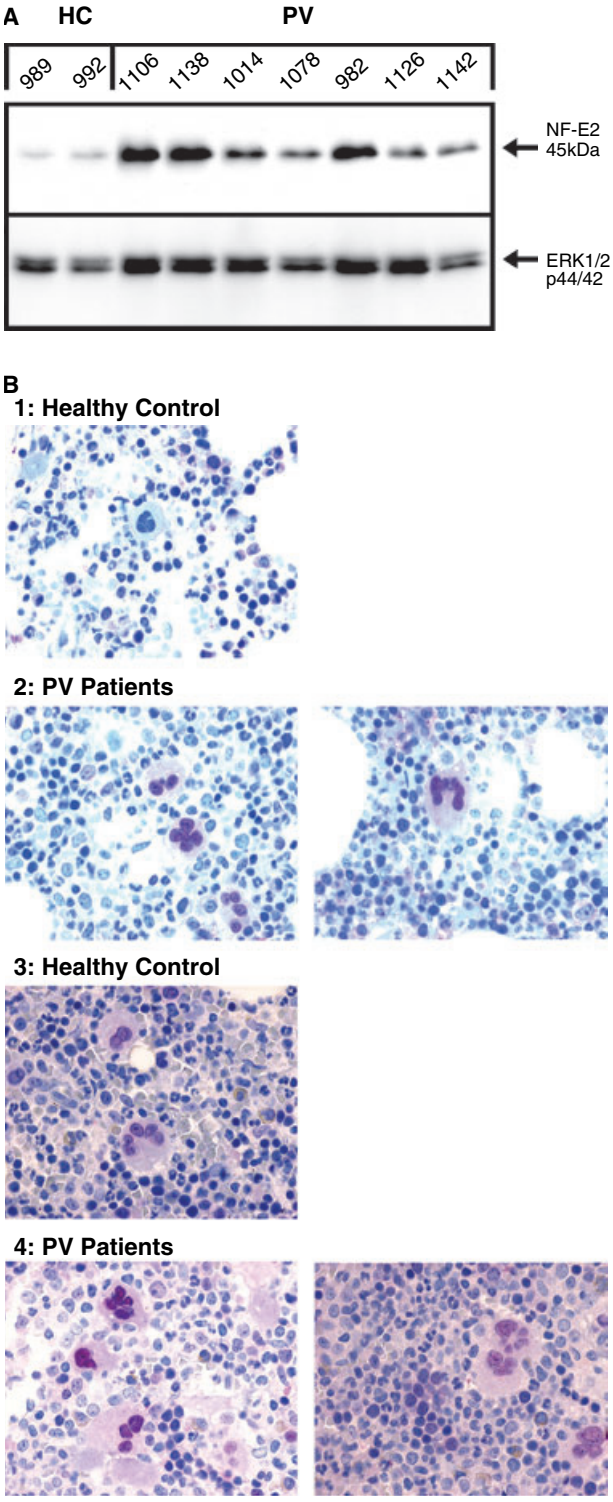


**Fig 5.** (A) NF-E2 protein expression in PV patients and healthy controls. Total cellular protein was isolated from purified granulocytes of seven PV patients and two healthy controls (HC). Protein (100 μg) was subjected to Western blotting and hybridized with an antibody against p45/NF-E2 (top). For verification that similar amounts of protein were loaded in each lane, the blot was stripped and reprobed with an antibody against ERK1/2 p44/p42 MAP kinases (bottom). Similar results were obtained in six additional PV patients and two additional healthy controls (data not shown). (B) Immunohistochemical detection of NF-E2 in PV patients and healthy controls. Bone marrow trephines from healthy adults (panels 1 and 3) or PV patients (panels 2 and 4) were stained for NF-E2 using two different antibodies (C-19X, panels 1 and 2, H-230X, panels 3 and 4; original magnifications ×400). A total of 6 PV patients and three healthy controls were stained with similar results, typical results are shown.

more heterogeneous in clinical course, appeared to share a common mechanism of malignant transformation (Klein *et al*, 2001; Rosenwald *et al*, 2001). The consistently observed overexpression of *prv-1* mRNA in most patients with PV (Klippel *et al*, 2003; Kralovics *et al*, 2003) likewise suggests that similar or closely related alterations lead to disease development in individual PV patients. However, it is possible that the cross-sectional cohort studied here, which included only eight individuals analysed at diagnosis, but 13 patients treated with cytoreductive medication, may obscure the identification of PV subtypes.

### *A model for the molecular aetiology of PV*

We demonstrated overexpression of the transcription factor NF-E2 in 53 of 57 (93%) of PV patients. Because NF-E2, as well as many additional overexpressed genes, are regulated by members of the Sp1 family of transactivators, we hypothesize that altered activity of an Sp1-like transcription factor contributes to the molecular pathology of PV.

Several published observations make NF-E2 an exceptionally promising candidate for the molecular aetiology of PV. It is expressed in haematopoietic precursors as well as in erythroid, megakaryocytic and granulocytic cells (Andrews *et al*, 1993; Toki *et al*, 1996). As a stem cell disorder in which trilineage hyperplasia is frequently observed, PV affects precisely these cell types.

Sayer *et al* (2000) have previously shown that overexpression of NF-E2 in fetal liver cells leads to the development of erythropoietin (Epo)-independent erythroid colonies. These colonies were 'pale and appeared to contain less haemoglobin than controls' (Sayer *et al*, 2000). Growth of Epo-independent erythroid colonies, termed 'endogenous erythroid colonies (EECs)', is characteristically and specifically observed in PV patients (Prchal & Axelrad, 1974) (reviewed in Westwood & Pearson, 1996). Strikingly, EECs are also often poorly haemoglobinized (Eridani *et al*, 1987; Dobo *et al*, 2001).

Sayer *et al* (2000) reported that NF-E2 overexpression in the erythroid cell line J2E promoted spontaneous morphological erythroid maturation in the absence of Epo. Parental J2E cells require Epo stimulation to mature past the proerythroblast stage. In contrast, 5–8% of NF-E2 overexpressing cells formed mature, enucleated reticulocytes without Epo (Sayer *et al*, 2000). Interestingly, the level of NF-E2 overexpression correlated with the amount of Epo-independent maturation, with the subclone expressing higher levels of NF-E2 displaying more Epo-independent maturation, a higher proliferation rate and a higher viability in the absence of Epo. Again, Epo-independent

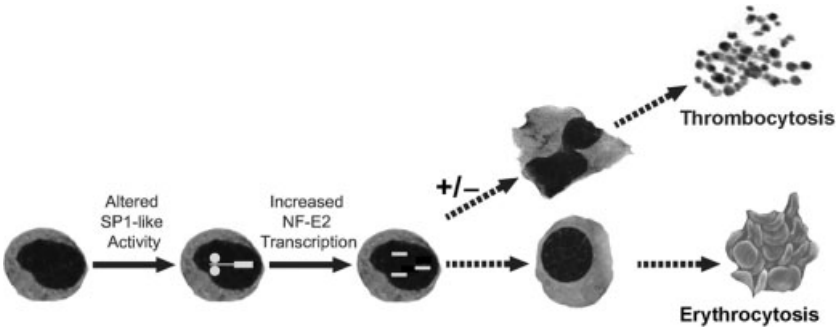


**Fig 6.** Proposed model for the molecular aetiology of PV. In PV cells, aberrant activity of Sp1-like transcription factors leads to increased expression of NF-E2 (Figs 4–6). Elevated concentrations of NF-E2 lead to an overproduction of erythroid and, in select patients, megakaryocytic cells and platelets. In this model the level of NF-E2 overexpression determines both the severity of erythrocytosis and the concurrent presence or absence of thrombocytosis.

maturation of NF-E2 overexpressing J2E cells resulted in very low levels of haemoglobin synthesis, which could not be increased by Epo stimulation.

More unexpected was the observation that ectopic expression of NF-E2 in the myeloid cell line M1 lead to the spontaneous emergence of erythroid and, in one subline, immature megakaryocytic cells (Sayer *et al*, 2000). Again, the level of ectopic NF-E2 expressed correlated with phenotype, the highest levels of NF-E2 inducing megakaryocytic maturation. NF-E2-induced erythroid cells were morphologically mature, including orthochromatic erythroblasts and reticulocytes, and expressed the Epo receptor. Taken together, these data demonstrate that augmenting the level of NF-E2 can promote erythroid maturation in the absence of Epo and can reprogramme precursor cells towards erythroid maturation.

Using antisense-oligonucleotides, Labbaye *et al* (1995) demonstrated that reduction of NF-E2 mRNA levels selectively abrogated erythroid colony formation (BFU-E), whereas granulocytic/monocytic colonies (CFU-GM) were unaffected. NF-E2 thus appears to be required for erythroid differentiation. In addition, overexpression of NF-E2 in J2E cells increased proliferation three- to fivefold and significantly augmented viability in the absence of serum (Sayer *et al*, 2000).

In light of these data, the NF-E2 overexpression in PV patients reported here, suggests a compelling model for the molecular disease aetiology (Fig 6). We propose that in PV patients, elevated concentrations of NF-E2 in bone marrow precursors (Fig 5B) lead to an overproduction of erythroid and, in some patients, megakaryocytic cells/platelets. The level of NF-E2 overexpression in individual PV patients varied from 2·3- to 40-fold (Fig 4A). We hypothesize that the level of NF-E2 overexpression determines both the severity of erythrocytosis and the concurrent presence or absence of thrombocytosis. The remaining 7% of PV patients with normal NF-E2 expression may have incurred alterations in downstream targets that are part of the same signal transduction pathway. The heterogeneity of disease manifestation could thus be explained by a similar but variably pronounced molecular alteration.

This hypothesis is readily testable by correlating NF-E2 levels with laboratory parameters, such as red cell mass (as a percentage of the expected) at diagnosis, the required frequency of phlebotomy and platelet numbers. Unfortunately, for the cross-sectional cohort retrospectively analysed here, the necessary detailed clinical data was not available, prohibiting such analysis. However, the large, prospective trial of PV patients currently planned by the MPD Research Consortium will soon enable us to test this hypothesis.

## Acknowledgements

Our first and foremost thanks go to the patients who consented to be included in this study. Our gratitude extends to their physicians and nurses, whose contribution was pivotal to this endeavour. In this respect, Prof. Heimpel, Prof. Peter, PD Dr Waller, Dr Loop and Dr Harris deserve special mention. We sincerely thank Marinella Klein for excellent technical assistance, Dr Emery Bresnick, for the generous gift of NF-E2 antibody, Cordula Steimle for critical review of the manuscript and Prof. Dr K. K. Geiger for his continued support. This work was supported by a grant from the Else-Kröner-Fresenius-Stiftung to H.L.P.

## Supplementary material

The following materials are available from http://www.blackwellpublishing.com/products/journals/suppmat/BJH/BJH5416/BJH5416sm.htm

**Table SI**. Characteristics of PV and SE patients analysed in this study.

**Table SII**. A complete listing of the 644 cDNAs differentially regulated between PV patients and healthy controls, including those whose function or protein product is unknown.

**Fig S1**. Logical approach and experimental design of the microarray analyses.

## References


Adamson, J.W., Fialkow, P.J., Murphy, S., Prchal, J.F. & Steinmann, L. (1976) Polycythemia vera: stem-cell and probable clonal origin of the disease. *New England Journal of Medicine*, **295,** 913–916.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Jr, Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O. & Staudt, L.M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403,** 503–511.

Andrews, N., Erdjument-Bromage, H., Davidson, M.B., Tempst, P. & Orkin, S.H. (1993) Erythroid transcription factor NF-E2 is a hematopoietic-specific basic-leucine zipper protein. *Nature*, **362,** 722–728.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of Royal Statistical Society*, **57,** 289–295.

Bilgrami, S. & Greenberg, B.R. (1995) Polycythemia rubra vera. *Seminars in Oncology*, **22,** 307–326.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J. & Vingron, M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genetics*, **29,** 365–371.

Carbonell, F., Ganser, A. & Heimpel, H. (1983) Cytogenetic studies in chronic myeloproliferative diseases. *Acta Haematologica*, **69,** 145–151.

Chen, C.J., Lin, T.T. & Shively, J.E. (1996) Role of interferon regulatory factor-1 in the induction of biliary glycoprotein (cell CAM-1) by interferon-gamma. *Journal of Biological Chemistry*, **271,** 28181–28188.

Cohen, H.T., Bossone, S.A., Zhu, G., McDonald, G.A. & Sukhatme, V.P. (1997) Sp1 is a critical regulator of the Wilms' tumor-1 gene. *Journal of Biological Chemistry*, **272,** 2901–2913.

Dameshek, W. (1951) Some speculations on the myeloproliferative syndromes. *Blood*, **6,** 372–375.

Diez Martin, J.L., Graham, D.L., Petitt, R.M. & DeWald, G.W. (1991) Chromosome studies in 104 patients with polycythemia vera. *Mayo Clinic Proceedings*, **66,** 287–299.

Dobo, I., Mossuz, P., Campos, L., Girodon, F., Allegraud, A., Latger-Cannard, V., Boiret, N., Pineau, D., Wunder, E., Zandecki, M., Praloran, V. & Hermouet, S. (2001) Comparison of four serum-free, cytokine-free media for analysis of endogenous erythroid colony growth in polycythemia vera and essential thrombocythemia. *Hematol J*, **2,** 396–403.

Eisen, M.B. & Brown, P.O. (1999) DNA arrays for analysis of gene expression. *Methods in Enzymology*, **303,** 179–205.

Eridani, S., Sawyer, B. & Pearson, T.C. (1987) Patterns of in vitro BFU-E proliferation in different forms of polycythaemia and in thrombocythaemia. *European Journal of Haematology*, **38,** 363–369.

Fialkow, P.J., Jacobson, R.J., Singer, J.W., Sacher, R.A., W, M.R. & Neefe, J.R. (1980) Philadelphia chromosome (Ph1)-negative chronic myelogenous leukemia (CML): a clonal disease with origin in a multipotent stem cell. *Blood*, **56,** 70–73.

Fialkow, P.J., Faguet, G.B., Jacobson, R.J., Vaidya, K. & Murphy, S. (1981) Evidence that essential thrombocythemia is a clonal disorder with origin in a multipotent stem cell. *Blood*, **58,** 916–919.

Fruchtman, S.M. (2004) Treatment paradigms in the management of myeloproliferative disorders. *Seminars in Hematology*, **41,** 18–22.

Giallongo, A., Oliva, D., Cali, L., Barba, G., Barbieri, G. & Feo, S. (1990) Structure of the human gene for alpha-enolase. *European Journal of Biochemistry*, **190,** 567–573.

Hauses, M., Tonjes, R.R. & Grez, M. (1998) The transcription factor Sp1 regulates the myeloid-specific expression of the human hematopoietic cell kinase (HCK) gene through binding to two adjacent GC boxes within the HCK promoter-proximal region. *Journal of Biological Chemistry*, **273,** 31844–31852.

Hoshiko, S., Radmark, O. & Samuelsson, B. (1990) Characterization of the human 5-lipoxygenase gene promoter. *Proceedings of the National Academy of Sciences, USA*, **87,** 9073–9077.

Iwata, Y., Nakayama, A., Murakami, H., Iida, K., Iwashita, T., Asai, N. & Takahashi, M. (1999) Characterization of the promoter region of the human RFP gene. *Biochemical and Biophysical Research Communications*, **261,** 381–384.

Jacobson, R.J., Salo, A. & Fialkow, P.J. (1978) Agnogenic myeloid metaplasia: a clonal proliferation of hematopoietic stem cells with secondary myelofibrosis. *Blood*, **51,** 189–194.

Johansson, P., Andreasson, B., Safai-Kutti, S., Rhedin, C., Vilen, L., Vaart, J. & Kutti, J. (2002) On the diagnosis of polycythaemia vera as assessed in the health and medical care in the Vastra Gotaland region, Sweden. *Journal of Internal Medicine*, **251,** 348–354.

Kaufman, L. & Rousseeuw, P. (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, New York.

Khanna-Gupta, A., Zibello, T., Simkevich, C., Rosmarin, A.G. & Berliner, N. (2000) Sp1 and C/EBP are necessary to activate the lactoferrin gene promoter during myeloid differentiation. *Blood*, **95,** 3734–3741.

Klein, U., Tu, Y., Stolovitzky, G.A., Mattioli, M., Cattoretti, G., Husson, H., Freedman, A., Inghirami, G., Cro, L., Baldini, L., Neri, A., Califano, A. & Dalla-Favera, R. (2001) Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *Journal of Experimental Medicine*, **194,** 1625–1638.

Klippel, S., Strunck, E., Busse, C.E., Behringer, D. & Pahl, H.L. (2002) Biochemical characterization of PRV-1, a novel hematopoietic cell surface receptor, which is overexpressed in polycythemia rubra vera. *Blood*, **100,** 2441–2448.

Klippel, S., Strunck, E., Temerinac, S., Bench, A.J., Meinhardt, G., Mohr, U., Leichtle, R., Green, A.R., Griesshammer, M., Heimpel, H. & Pahl, H.L. (2003) Quantification of PRV-1 mRNA distinguishes polycythemia vera from secondary erythrocytosis. *Blood*, **102,** 3569–3574.

Kralovics, R., Buser, A.S., Teo, S.S., Coers, J., Tichelli, A., Van Der Maas, A.P. & Skoda, R.C. (2003) Comparison of molecular markers in a cohort of patients with chronic myeloproliferative disorders. *Blood*, **102,** 1869–1871.

Kruisbeck, A.M., Margulies, D.H., Shevach, E.M. & Strober, W. (1991) *Current Protocols in Immunology*. John Wiley & Sons, New York.

Labbaye, C., Valtieri, M., Barberi, T., Meccia, E., Masella, B., Pelosi, E., Condorelli, G.L., Testa, U. & Peschle, C. (1995) Differential expression and functional role of GATA-2, NF-E2, and GATA-1 in normal adult hematopoiesis. *Journal of Clinical Investigation*, **95,** 2346–2358.

Lebeda, R.A., Johnson, S.K. & Haun, R.S. (1999) Transcriptional regulation of the human ADP-ribosylation factor 5 (ARF5) gene. *Biochimica et Biophysica Acta*, **1445,** 314–320.

Luzi, P., Victoria, T., Rafi, M.A. & Wenger, D.A. (1997) Analysis of the 5′ flanking region of the human galactocerebrosidase (GALC) gene. *Biochemical and Molecular Medicine*, **62,** 159–164.

van Maanen, M.H., Fournier, P.A., Palmer, T.N. & Abraham, L.J. (1999) Characterization of the human glycogenin-1 gene: identification of a muscle-specific regulatory domain. *Gene*, **234,** 217–226.

Maclean, K.N., Kraus, E. & Kraus, J.P. (2004) The dominant role of Sp1 in regulating the cystathionine beta-synthase -1a and -1b promoters facilitates potential tissue-specific regulation by Kruppel-like factors. *Journal of Biological Chemistry*, **279,** 8558–8566.

Marsden, M.D., Collins, J.E., Greenwood, M.D., Adams, M.J., Fleming, T.P., Magee, A.I. & Buxton, R.S. (1997) Cloning and transcriptional analysis of the promoter of the human type 2 desmocollin gene (DSC2). *Gene*, **186,** 237–247.

Meeker, T.C., Loeb, J., Ayres, M. & Sellers, W. (1990) The human Pim-1 gene is selectively transcribed in different hemato-lymphoid cell lines in spite of a G + C-rich housekeeping promoter. *Molecular Cellular Biology*, **10,** 1680–1688.

Michiels, J.J., Kutti, J., Stark, P., Bazzan, M., Gugliotta, L., Marchioli, R., Griesshammer, M., van Genderen, P.J., Briere, J., Kiladjian, J.J., Barbui, T., Finazzi, G., Berlin, N.I., Pearson, T.C., Green, A.C., Fruchtmann, S.M., Silver, R.T., Hansmann, E., Wehmeier, A., Lengfelder, E., Landolfi, R., Kvasnicka, H.M., Hasselbalch, H., Cervantes, F., Thiele, J. et al. (1999) Diagnosis, pathogenesis and treatment of the myeloproliferative disorders essential thrombocythemia, polycythemia vera and essential megakaryocytic granulocytic metaplasia and myelofibrosis. *Netherlands Journal of Medicine*, **54,** 46–62.

Moliterno, A.R., Hankins, D. & Spivak, J.L. (1998) Impaired expression of the thrombopoietin receptor by platelets from patients with polycythemia vera. *New England Journal of Medicine*, **338,** 572–580.

Murphy, S. (1999) Diagnostic criteria and prognosis in polycythemia vera and essential thrombocythemia. *Seminars in Hematology*, **36,** 9–13.

Nathanson, C.M., Wasselius, J., Wallin, H. & Abrahamson, M. (2002) Regulated expression and intracellular localization of cystatin F in human U937 cells. *European Journal of Biochemistry*, **269,** 5502–5511.

O'Leary, K.A. & Kasper, C.B. (2000) Molecular basis for cell-specific regulation of the NADPH-cytochrome P450 oxidoreductase gene. *Archives of Biochemistry and Biophysics*, **379,** 97–108.

Pahl, H.L., Krauß, B., Schulze-Osthoff, K., Decker, T., Traenckner, E.B.-M., Vogt, M., Myers, C., Parks, T., Warring, P., Mühlbacher, A., Czernilofsky, A.-P. & Baeuerle, P.A. (1996) The immunosuppressive fungal metabolite gliotoxin specifically inhibits transcription factor NF-κB. *Journal of Experimental Medicine*, **183,** 1829–1840.

Pastorcic, M. & Das, H.K. (1999) An upstream element containing an ETS binding site is crucial for transcription of the human presenilin-1 gene. *Journal of Biological Chemistry*, **274,** 24297–24307.

Pearson, T.C. (1998) Diagnosis and classification of erythrocytoses and thrombocytoses. *Baillieres Clinical Haematology*, **11,** 695–720.

Pellagatti, A., Vetrie, D., Langford, C.F., Gama, S., Eagleton, H., Wainscoat, J.S. & Boultwood, J. (2003) Gene expression profiling in Polycythemia Vera using cDNA microarray technology. *Cancer Research*, **63,** 3940–3944.

Pischedda, C., Cocco, S., Melis, A., Marini, M.G., Kan, Y.W., Cao, A. & Moi, P. (1995) Isolation of a differentially regulated splicing isoform of human NF-E2. *Proceedings of the National Academy of Sciences, USA*, **92,** 3511–3515.

Prchal, J.F. & Axelrad, A.A. (1974) Bone-marrow responses in polycythemia vera. *New England Journal of Medicine*, **290,** 1382.

Rosenwald, A., Alizadeh, A.A., Widhopf, G., Simon, R., Davis, R.E., Yu, X., Yang, L., Pickeral, O.K., Rassenti, L.Z., Powell, J., Botstein, D., Byrd, J.C., Grever, M.R., Cheson, B.D., Chiorazzi, N., Wilson, W.H., Kipps, T.J., Brown, P.O. & Staudt, L.M. (2001) Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *Journal of Experimental Medicine*, **194,** 1639–1647.

Rosenwald, A., Wright, G., Leroy, K., Yu, X., Gaulard, P., Gascoyne, R.D., Chan, W.C., Zhao, T., Haioun, C., Greiner, T.C., Weisenburger, D.D., Lynch, J.C., Vose, J., Armitage, J.O., Smeland, E.B., Kvaloy, S., Holte, H., Delabie, J., Campo, E., Montserrat, E., Lopez-Guillermo, A., Ott, G., Muller-Hermelink, H.K., Connors, J.M., Braziel, R., Grogan, T.M., Fisher, R.I., Miller, T.P., LeBlanc, M., Chiorazzi, M., Zhao, H., Yang, L., Powell, J., Wilson, W.H., Jaffe, E.S., Simon, R., Klausner, R.D. & Staudt, L.M. (2003) Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *Journal of Experimental Medicine*, **198,** 851–862.

Samuelsson, J. (1998) Survival in a patient with polycythaemia vera for over thirty years: implications for treatment decisions in younger patients. *Leukemia and Lymphoma*, **32,** 195–198.

Savage, K.J., Monti, S., Kutok, J.L., Cattoretti, G., Neuberg, D., De Leval, L., Kurtin, P., Dal Cin, P., Ladd, C., Feuerhake, F., Aguiar, R.C., Li, S., Salles, G., Berger, F., Jing, W., Pinkus, G.S., Habermann, T., Dalla-Favera, R., Harris, N.L., Aster, J.C., Golub, T.R. & Shipp, M.A. (2003) The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. *Blood*, **102,** 3871–3879. E-pub August 2003, 3821.

Sayer, M.S., Tilbrook, P.A., Spadaccini, A., Ingley, E., Sarna, M.K., Williams, J.H., Andrews, N.C. & Klinken, S.P. (2000) Ectopic expression of transcription factor NF-E2 alters the phenotype of erythroid and monoblastoid cells. *Journal of Biological Chemistry*, **275,** 25292–25298.

Shaffer, A.L., Rosenwald, A., Hurt, E.M., Giltnane, J.M., Lam, L.T., Pickeral, O.K. & Staudt, L.M. (2001) Signatures of the immune response. *Immunity*, **15,** 375–385.

Solberg, L.A., Jr (2002) Therapeutic options for essential thrombocythemia and polycythemia vera. *Seminars in Oncology*, **29,** 10–15.

Spivak, J.L. (2004) The chronic myeloproliferative disorders: clonality and clinical heterogeneity. *Seminars in Hematology*, **41,** 1–5.

Strunck, E., Bliestle, H., Temerinac, S. & Pahl, H.L. (2000) Characterization of the promoter for PRV-1, a gene overexpressed in *Polycythemia vera. Blood*, **96,** 511a.

Tazawa, R., Hirosawa, S., Suzuki, K., Hirokawa, K. & Aoki, N. (1993) Functional characterization of the 5′-regulatory region of the human thrombomodulin gene. *Journal of Biochemistry*, **113,** 600–606.

Tefferi, A. (1999) Pathogenetic mechanisms in chronic myeloproliferative disorders: polycythemia vera, essential thrombocythemia, agnogenic myeloid metaplasia, and chronic myelogenous leukemia. *Seminars in Hematology*, **36,** 3–8.

Temerinac, S., Klippel, S., Strunck, E., Röder, S., Lübbert, M., Lange, W., Azemar, M., Meinhardt, G., Schaefer, H.E. & Pahl, H.L. (2000) Cloning of PRV-1, a novel member of the uPAR receptor superfamily, which is overexpressed in *Polycythemia rubra vera*. *Blood*, **95,** 2569–2576.

Toki, T., Itoh, J., Arai, K., Kitazawa, J., Yokoyama, M., Igarashi, K., Yamamoto, M. & Ito, E. (1996) Abundant expression of erythroid transcription factor P45 NF-E2 mRNA in human peripheral granurocytes. *Biochemical and Biophysical Research Communications*, **219,** 760–765.

Vyhlidal, C., Samudio, I., Kladde, M.P. & Safe, S. (2000) Transcriptional activation of transforming growth factor alpha by estradiol: requirement for both a GC-rich site and an estrogen response element half-site. *Journal of Molecular Endocrinology*, **24,** 329–338.

Weiss, M.J., Ray, K., Henthorn, P.S., Lamb, B., Kadesch, T. & Harris, H. (1988) Structure of the human liver/bone/kidney alkaline phosphatase gene. *Journal of Biological Chemistry*, **263,** 12002–12010.

Westwood, N.B. & Pearson, T.C. (1996) Diagnostic applications of haemopoietic progenitor culture techniques in polycythaemias and thrombocythaemias. *Leukemia & Lymphoma*, **22,** 95–103.

WHO (2001) *Tumours of the Hemopoietic and Lymphoid Tissues*. IARC Press, Lyon.

Wo, Y.Y., Stewart, J. & Greenlee, W.F. (1997) Functional analysis of the promoter for the human CYP1B1 gene. *Journal of Biological Chemistry*, **272,** 26702–26707.

Workman, C., Jensen, L.J., Jarmer, H., Berka, R., Gautier, L., Nielser, H.B., Saxild, H.H., Nielsen, C., Brunak, S. & Knudsen, S. (2002) A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biology*, **3,** 0048.0041–0048.0016.

Yeoh, E.J., Ross, M.E., Shurtleff, S.A., Williams, W.K., Patel, D., Mahfouz, R., Behm, F.G., Raimondi, S.C., Relling, M.V., Patel, A., Cheng, C., Campana, D., Wilkins, D., Zhou, X., Li, J., Liu, H., Pui, C.H., Evans, W.E., Naeve, C., Wong, L. & Downing, J.R. (2002) Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*, **1,** 133–143.

Zhang, D.E., Hetherington, C.J., Tan, S., Dziennis, S.E., Gonzalez, D.A., Chen, H.M. & Tenen, D.G. (1994) Sp1 is a critical factor for the monocytic specific expression of human CD14. *Journal of Biological Chemistry*, **269,** 11425–11434.

Zhong, Z.D., Hammani, K., Bae, W.S. & DeClerck, Y.A. (2000) NF-Y and Sp1 cooperate for the transcriptional activation and cAMP response of human tissue inhibitor of metalloproteinases-2. *Journal of Biological Chemistry*, **275,** 18602–18610.