Leukemia (2008), 1–9

www.nature.com/leu

ORIGINAL ARTICLE

Gene expression profiling distinguishes *JAK2*V617F-negative from *JAK2*V617F-positive patients in essential thrombocythemia

E Puigdecanet[1,2,3,4], B Espinet[1,3,5], J Lozano[6,7], L Sumoy[6], B Bellosillo[1], L Arenillas[2,3,5], A Álvarez-Larrán[8], F Solé[1,3,5], S Serrano[1,2,3], C Besses[8] and L Florensa[2,3,5]

[1]Laboratori de Citogenètica i Biologia Molecular, Servei de Patologia. Hospital del Mar, IMAS, Barcelona, Spain; [2]Laboratori de Citologia Hematològica, Servei de Patologia, Hospital del Mar, IMAS, Barcelona, Spain; [3]GRETNHE, IMIM-Hospital del Mar, Barcelona, Spain; [4]Departament de Biologia Cellular, Fisiologia i Immunologia. Facultat de Biociències. Universitat Autònoma de Barcelona, Bellaterra, Spain; [5]Escola de Citologia Hematològica Soledad Woessner-IMAS, Barcelona, Spain; [6]Laboratori de Microarrays, Programa de Bioinformàtica i Genòmica, Centre de Regulació Genòmica-UPF, Barcelona, Spain; [7]Plataforma de Bioinformática, CIBERehd, Hospital Clinic, Barcelona, Spain and [8]Servei d'Hematologia Clínica, Hospital del Mar, IMAS, Barcelona, Spain

To explore the gene expression signature in essential thrombocythemia (ET) patients in relation to *JAK2*V617F mutational status, expression profiling in circulating granulocytes was performed. Twenty ET were studied by microarray analysis and the results were confirmed by real-time quantitative RT-PCR in 40 ET patients, not receiving cytoreductive treatment. A heterogeneous molecular signature characterized by two main gene expression patterns was found: one with an upregulation of inflammatory genes related to neutrophil activation and thrombosis, and the other with significantly lower expression of these genes. Supervised clustering analysis showed 30 genes differentially expressed between *JAK2*V617F-negative and *JAK2*V617F-positive ET patients. Among the *JAK2*V617F-negative, a set of 14 genes (*CISH*, *C13orf18*, *CCL3*, *PIM1*, *MAFF*, *SOCS3*, *ID2*, *GADD45B*, *KLF5*, *TNF*, *LAMB3*, *HRH4*, *TAGAP* and *TRIB1*) showed an abnormal expression pattern. In this group of patients, *CISH*, *SOCS2*, *SOCS3* and *PIM1* genes, all involved in JAK-STAT signalling pathway, presented a lower expression. A two-gene predictor model was built comprising *FOSB* and *CISH* genes, which were the best discriminators of *JAK2*V617F status. In conclusion, *JAK2*V617F-negative ET patients present a characteristic gene expression profile, different from *JAK2*V617F-positive patients. Other pathways, besides JAK-STAT, might be implicated in the pathophysiology of *JAK2*V617F-negative ET patients.

Leukemia advance online publication, 15 May 2008; doi:10.1038/leu.2008.112

Keywords: essential thrombocythemia; *JAK2*V617F; gene expression profiling; real-time quantitative RT-PCR

Introduction

BCR-ABL-negative myeloproliferative disorders (MPDs) including polycythemia vera (PV), essential thrombocythemia (ET) and primary myelofibrosis (PMF) are characterized by a clonal expansion of a multipotent haematopoietic progenitor cell. ET is a heterogeneous entity characterized by increased bone marrow megakaryocytes, persistent thrombocytosis and an increased risk of thrombohaemorrhagic complications.[1] Recently, the existence of the *JAK2*V617F mutation has been reported in a high proportion of MPD BCR-ABL-negative patients (90% of PV, 50–60% of ET and PMF patients).[2–6] Although the description of the *JAK2*V617F mutation has provided important insight into the pathogenesis of PV, ET and PMF, the molecular abnormalities associated to *JAK2*V617F-negative BCR-ABL-negative MPD remain unknown. Regarding ET, *JAK2*V617F-positive patients present clinical features resembling the PV phenotype, in contrast to *JAK2*V617F-negative ET patients, and according to this observation, two molecular distinct subtypes of ET have been proposed.[7] The constitutive kinase activity of V617F mutation causes cytokine-independent activation of the JAK-STAT, PI3K-AKT and Ras-MAPK pathways. This multiple pathway activation results in the independence or hypersensitivity of haematopoietic progenitors to numerous cytokines, a hallmark of MPD.[2–5] Recently, mutations involving exon 12 of the *JAK2* gene in PV[8,9] and *MPL*W515L mutation in approximately 10% of *JAK2*V617F-negative PMF and in a smaller proportion of ET have been reported.[10] However, in more than 40% of ET patients a molecular marker is still lacking.

Global gene expression analysis has been used as a powerful tool to provide insight into the molecular aetiology of haematologic malignancies. Granulocytes with altered gene expression have been detected in PV[11–13] and new markers have been defined in relation to *JAK2*V617F.[14] Recently, distinct gene expression subclasses in ET in relation to *JAK2*V617F status have been described by using cDNA microarrays.[15]

Our aim was to characterize the gene expression profile of peripheral blood granulocytes in ET using whole genome oligonucleotide microarray technology and real-time quantitative RT-PCR. The objective was to extend the set of markers for ET and analyse the relation between gene expression data, *JAK2*V617F status and clinical manifestations of the disease.

Patients and methods

Patients

A total of 40 cases (28 women/12 men) with a median age of 54 years (ranging from 31 to 85 years), diagnosed with ET according to the WHO criteria (2001), who had never received cytoreductive treatment were studied. At the time of diagnosis four patients had presented thrombotic events and three patients had shown haemorrhages. Among the 40 patients, 32 presented at least one of the following clonality markers: *JAK2*V617F, positivity for *HUMARA* assay and myeloid endogenous growth. *JAK2*V617F was detected in 52.5% (21/40) patients, always in

Correspondence: Dr L Florensa, Laboratori de Citologia Hematològica, Servei de Patologia-GRETNHE, Hospital del Mar, Pg. Marítim, 25-29, Barcelona 08003, Spain.
E-mail: epuigdecanet@imim.es
Received 4 July 2007; revised 2 April 2008; accepted 4 April 2008

heterozygosity, with a percentage of mutant alleles ranging from 7.8 to 40.7% (mean of 28.3%). Among the 17 out of 28 informative women, seven showed clonality by the *HUMARA* assay. Out of 40 patients, 26 showed endogenous myeloid growth (eBFU-E and/or eCFU-MK). Out of 19 *JAK2*V617F-negative patients, 11 presented *HUMARA* clonality and/or myeloid endogenous growth. Therefore, only eight patients did not present any clonal marker. We analysed the association between presence or absence of *JAK2*V617F and clinico-biological data from all patients (Table 1). Ten healthy individuals (5 women/5 men) with age ranging from 25 to 58 years and with a normal blood cell count were used as assay validation controls. The study was approved by the local ethics committee and written informed consent was obtained from all patients, according to the Declaration of Helsinki.

RNA extraction

Venous blood (20 ml) was collected in ethylendiaminetetraacetic acid (EDTA) and immediately processed. Granulocytes were isolated as previously described[16] and total cellular RNA from each sample was isolated from 10×10^6 granulocytes using a guanidinium thiocyanate-based method (Ultraspec; Biotecx Laboratories, Houston, TX, USA) following the protocol supplied by the manufacturer.

RNA quality was assessed by nanoelectrophoresis using the Nano lab-on-a-chip assay for total eukaryotic RNA (Bioanalyzer, Agilent Technologies, Palo Alto, CA, USA). Only samples with RNA integrity number >7 were subsequently used in microarray experiments.

Microarray analysis

RNA labelling and hybridization. In each experiment, RNA obtained from granulocytes from a single ET patient was compared with a pool of granulocyte RNAs from 10 healthy individuals. Total RNA (500 ng) were reverse transcribed into cDNA and amplified by *in vitro* transcription in the presence of fluorescent-labelled CTP using the Low Input RNA Labelling and Amplification Kit from Agilent following the manufacturer's instructions.

Each microarray was hybridized with 750 ng of each amplified cRNA labelled with Cy5 or Cy3 at a specific activity between 7 and 15 pmol/μg. Duplicate hybridizations were performed for each comparison with dye swapping to control for possible differences in the incorporation rate of the Cy5 and Cy3 fluorochromes. Microarray expression profiles were obtained using Whole Human Genome oligonucleotide microarrays (G4112A, Agilent), comprising 41 000 60-nt oligonucleotide probes, most represented as single spots, using the protocol recommended by Agilent with SSC-based washes.

Scanning and analysis. Fluorescent images were obtained using an Agilent G2565BA scanner. Fluorescence intensity data were extracted and quantified using Genepix 6.0 (Molecular Dynamics).

Only spots with signal intensities twice above the local background, not saturated and not flagged as absent by GenePix, were considered reliable and used for subsequent analysis. Normalized log2ratios were scaled between arrays to make all data comparable. Statistically significant differences in gene expression were determined by computing a Bayesian statistic using all log2ratios from replicate hybridizations.

Target genes were considered as differentially expressed when their empirical Bayes statistic B was higher than 15 (B_rank $>$ 98%). All quantitative and statistical analyses were performed using MMARGE tool, a web implementation of the Limma package in the R environment.[17,18]

Supervised classification of ET samples into categories according to clinical and biological parameters based on gene expression profiles was performed using standard two-tailed Student's test using the *t*-test function from R (http://www.cran.r-project.org) and was used to determine statistical significance of differential gene expression. Gene Ontology terms for annotated genes and common pathways were analysed using the Ingenuity Pathways Analysis software (Ingenuity Systems Inc.) and the FATIGO+ gene ontology-based application (http://fatigo.bioinfo.cnio.es/bioinfo/).[19] The differentially expressed gene lists were linked to the Internet genome databases (that is, GeneCards, GenBank, Swiss Prot, Online Mendelian Inheritance in Man).

Microarray data have been deposited in NCBIs Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/projects/geo/)

Table 1 Main clinico-biological data at diagnosis of the 40 ET patients in relation to JAK2 status

	JAK2V617F-positive (n = 21)	*JAK2V617F-negative (n = 19)*	P-*value*
Age[a]	52 (31–76)	55 (35–85)	NS
Sex	15F/6M	13F/6M	NS
Platelets ($\times 10^9$ per l)[b]	719.29 (± 235.69)	832.61 (± 243.08)	NS
WBC ($\times 10^9$ per l)[b]	9.12 (± 2.57)	7.97 (± 1.33)	0.097
Haemoglobin (g per 100 ml)[b]	14.6 (± 1.4)	13.5 (± 1.6)	NS
Increased LDH	3/20	5/19	NS
LAP index[b]	86.17 (± 30.99)	55.64 (± 16.28)	0.002
Splenomegaly	1/21	1/19	NS
Thrombotic events	2/21	2/19	NS
Haemorrhagic events	0/21	3/19	NS
Cardiovascular risk factors	10/21	8/19	NS
% *JAK2*V617F alleles[a]	28.3% (7.8–40.7%)	—	NA
HUMARA clonality	2/7	5/10	NS
Myeloid endogenous growth (eBFU-E and/or eCFU-MK)	18/21	8/18	0.015
eBFU-E	17/21	1/18	<0.001
eCFU-MK	15/21	8/18	0.112
Increased PRV-1	10/20	1/19	0.003

Abbreviations: eBFU-E, erythroid endogenous growth; eCFU-MK, megakaryocytic endogenous growth; ET, essential thrombocythemia; F, female; LAP, leucocyte alkaline phosphatase; LDH, lactate dehidrogenase; M, male; NA: not applicable; NS: not statistically significant.
[a]Median value (range) is reported.
[b]Mean value (± s.d.) is reported.

and are accessible through GEO Series accession number GSE11003.

Real-time quantitative RT-PCR analysis by TaqMan low density arrays

On the basis of array results, we evaluated expression levels of selected genes by real-time quantitative RT-PCR analyses using the TaqMan Low Density Arrays (LDA platform, Applied Biosystems, Foster City, CA, USA) in all patients. cDNA was reverse transcribed from total RNA with random primers using High Capacity cDNA Archive Kit (Applied Biosystems), as described by the manufacturer. TaqMan PCR reactions were performed on cDNA samples using the TaqMan Universal PCR Master Mix (Applied Biosystems) according to the manufacturer's instructions, and fluorescence was detected in an ABI PRISM 7900 HT (Applied Biosystems). Cycle threshold (C_t) values were calculated by the ABI PRISM software, and relative gene expression levels were expressed as the difference in C_t values (ΔC_t) of the target gene and the geometric mean of the three housekeeping genes: *18S rRNA* (eukaryotic 18S ribosomal RNA), *GUSB* (glucuronidase, beta) and *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase).[20] $\Delta\Delta C_t$ values were calculated for each sample using the mean of its ΔC_t subtracted from the mean ΔC_t value measured in the pool of the 10 healthy subjects, considered as a calibrator. Gene expression quantification was achieved using the comparative C_t method for relative quantification, in which the amount of target is expressed as $2^{-\Delta\Delta Ct}$. We designed two different LDA platforms, containing 96 and 48 assays in duplicate, which were selected according to the microarray results (see Supplementary Table 1 and 2). Supervised classification of ET samples based on real-time quantitative RT-PCR results into categories according to clinical and biological parameters was performed as described above (Microarray data analysis).

Statistical analyses of the clinical and biological data

Statistical analyses of the clinical and biological data and their relationship with the gene expression data were performed using the SPSS12 (SPSS Inc., Chicago, Illinois, USA). Differences between normally distributed group data were analysed by the unpaired Student's *t*-test, as well as one-factor analysis of variance (ANOVA) and nonparametric tests when appropriate. The χ^2-test was used to compare categorical variables among groups. Statistical significance was considered when $P<0.05$.

Predictor modelling

Predictor model was constructed on a training data set and evaluated on an independent test data set. We used the misclassification-penalized posterior algorithm (MiPP).[21] MiPP is based on stepwise incremental classification modelling for discovery of the most parsimonious prediction models. Model training was performed using different classification algorithms including linear discriminant analysis, quadratic discriminant analysis, support vector machine learning and logistic regression.

Results and discussion

Global gene expression patterns in ET

A flow-chart describing the design of the entire study is provided in Figure 1. Analysing the microarray data from the first 20 patients studied (patients 1–20), a common gene expression signature was found in 17 out of 20 ET patients (patients 1–17) compared to the healthy control pool. The other three patients (patients 18–20) showed a different gene expression pattern, being independent among them, and were considered outliers. The signature of the first 17 patients was composed of 124 upregulated and 14 downregulated genes, with >2-fold changes over control granulocytes.

Applying Ingenuity Pathways Analysis (Ingenuity Systems, www.ingenuity.com) of the 124 upregulated genes, 42 were implicated in the immune response ($P<0.0003$), 39 of them were involved in the development and function of the haematological system ($P<0.0005$) and 34 were related to cancer ($P<0.0004$). We studied common pathways for these genes, detecting two main networks related to the immune response ($P<0.05$). One important network comprised 24 genes, among which *TNF* and *IL-1* were overexpressed. These genes are the main initiators of the immune and inflammatory responses. Another network more related to haematological development and function comprised 39 overexpressed genes. These genes were directly involved in neutrophil chemotaxis (*CXCL2*, *PPBP*, *CCL4*, *CCL3*, *CCL20* and *CCL23*), neutrophil adhesion (*ICAM1*), complement and coagulation mediators (*PLAU*, plasminogen activator, urokinase, and *PLAUR*, *PLAU* receptor) and other relevant molecules, such as *OSM*, *PTGS2*, *ETS2*, *CD83* and *EDN1*. These findings suggest that there is an activation of the immune response in ET patients over healthy controls.

We selected a set of 96 genes, including 81/124 of the most upregulated and 9/14 of the most downregulated with functional annotation (comprising all the above referred genes), 3 other genes previously described as associated with MPDs (*CD44*, *NFE2* and *ANKRD15*),[11,12,14] and *18S rRNA*, *GAPDH* and *GUSB* genes as endogenous controls. This set of 96 genes was validated by real-time quantitative RT-PCR in the same cohort of 20 patients previously studied by microarrays. The concordance between the two assays, calculated using the Spearman correlation test, was high for most of the genes validated, with a mean correlation coefficient across 17 samples of 0.76 (s.d. = 0.06). Subsequently, 20 additional ET patients (patients 21–40) were studied using the 96 gene custom designed LDA platform, expanding the series to 40 patients. A hierarchical clustering of the gene expression data of the 96 genes in the 40 patients studied showed a classification into two main groups. The first one (group A) comprised 18 patients, showing the common expression pattern found in the microarray data (which was characterized by 81 genes upregulated and 9 downregulated, in relation to control granulocytes). The other cluster (group B) included a set of 22 patients. In this cluster two subgroups were observed: one of them (group B.1) comprised 9 patients and the other (group B.2) included 13 patients. Concerning expression data of the different patient clusters (groups A, B.1 and B.2), one-factor ANOVA test confirmed a progressive decrease in expression levels of certain genes in the three respective clusters (genes with F ratio>33 were *ARL8*, *CCL3*, *CCL4*, *CD83*, *CXCL2*, *DUSP2*, *EDN1*, *ICAM1*, *IL1A*, *IL1RN*, *KLF10*, *MAFF*, *NFKBIZ*, *PLAU*, *PLEK*, *PLEKHG2*, *PLK2*, *PTX3*, *TNF*, *TNFAIP3*, *XBP1* and *ZFP36*). In addition, the same test confirmed that the nine downregulated genes in microarray analysis (*AP3M1*, *ARRDC3*, *C4BPA*, *C1QC*, *DDX28*, *SOLH*, *SOX4*, *TBCC*, *ZNF217*) had similar expression levels in the patients studied, as some other genes such as *B2M*, *CCL23*, *CD44*, *NBN*, *NFE2*, *RABGEF1* and *RBF600*. Cluster data are shown in Figure 2.

We did not find any statistically significant association between the two cluster-derived patient groups and the clinical and biological parameters described in Table 1. However, three

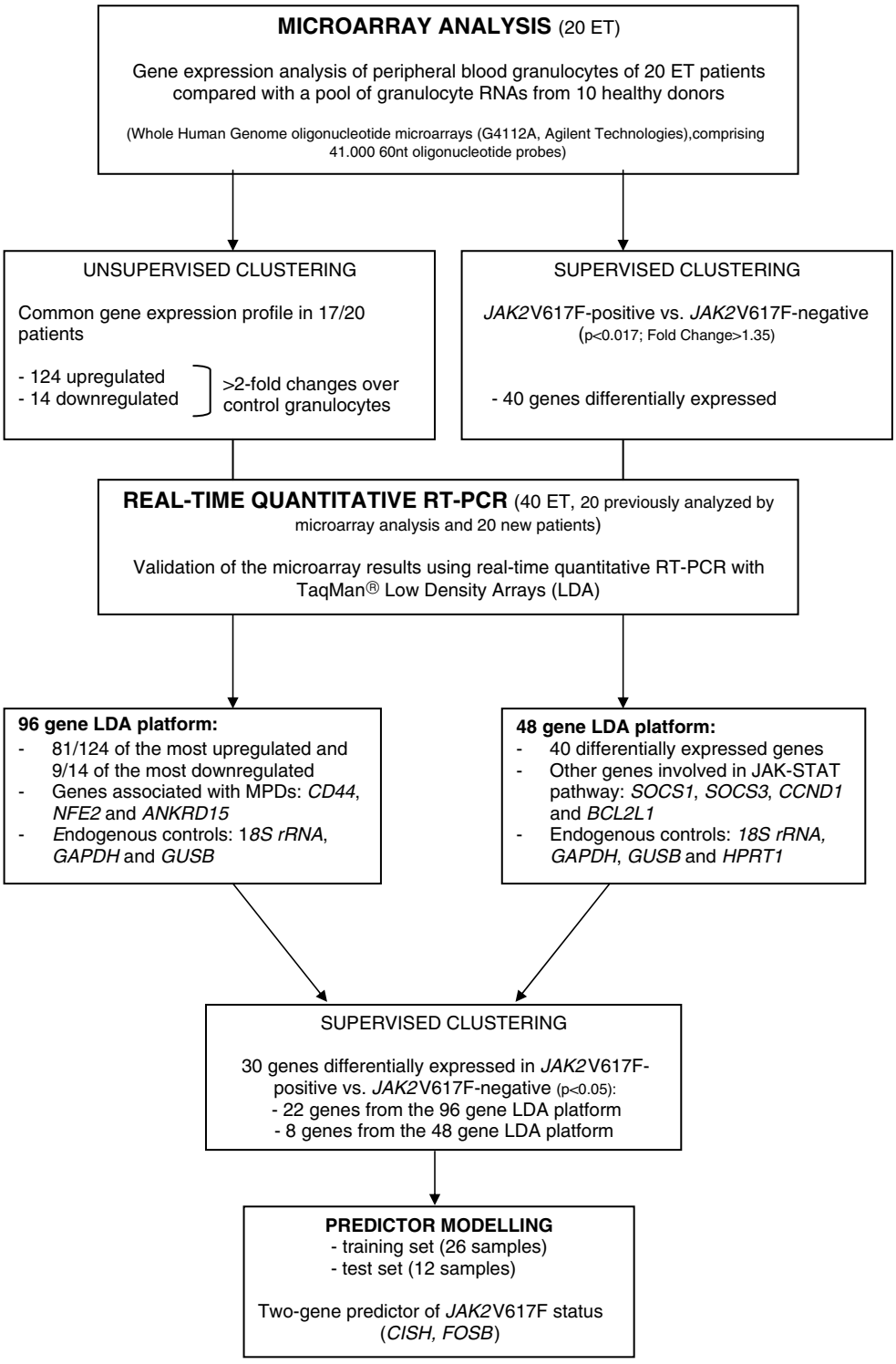

Figure 1 Flow-chart describing the design of the study.

out of four patients that presented thrombotic complications were included in group A. Moreover, 76.9% (10/13) patients in group B.2 were *JAK2*V617F-positive.

Besides, an overexpression of *NFE2* and downregulation of *ANKRD15* has been described in PV.[12,14] In our results, the expression of *NFE2* and *ANKRD15* genes is not altered in ET patients and is similar to healthy controls. We detected an upregulation of *CD44* in ET patients, also previously reported in PV patients.[11] CD44 is an adhesion molecule involved in cell–cell contact between progenitor cells and stromal cells and might therefore be important for haemopoiesis, in proliferation of progenitor cells and myeloid differentiation,[22] as well as in the pathophysiology of bone marrow fibrosis.[23]

It is remarkable that some overexpressed genes in group A in relation to group B are involved in the immune response. An activated neutrophil phenotype has been described by Falanga

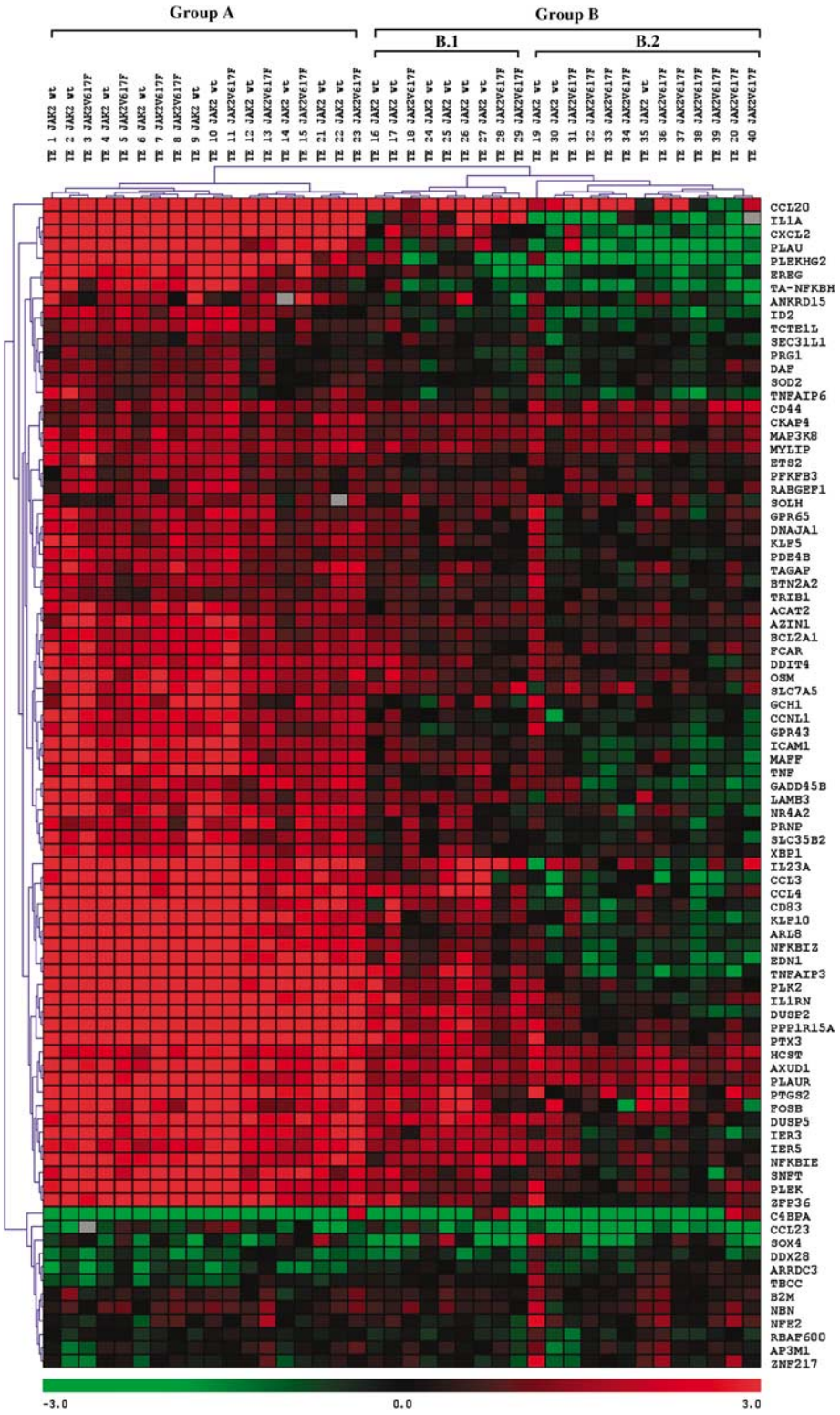

Figure 2 Gene expression profiling in 40 essential thrombocythemia (ET) patients using the 96 TaqMan Low Density Arrays platform (LDA). Among the 96 genes assayed, gene expression data of 87 genes is shown. The rest of assays data are not shown because in six genes (*C1QC, CNGB1, HLA-G, LIMS1, MT1M* and *PHACTR1*) no amplification was detected in any of the samples studied, and three genes (*18S rRNA, GAPDH* and *GUSB*) were used as endogenous controls. A hierarchical clustering (Euclidean distance Complete Linkage) of the gene expression data of the 96 genes in 40 patients studied showed a classification into two main groups: A and B. In this cluster B two subgroups were observed: B.1 and B.2. The columns represent individual patients and the rows represent specific genes. The expression values are the log2ratios between every patient and the control pool. Grey squares depict missing data points.

et al.[24] in ET and PV patients. These authors have proposed a probable role for leucocytes in the thrombosis of ET. On the other hand, a high monocyte and platelet activation has been reported in ET patients with thrombosis.[25] In this sense, activated leucocytes and platelets could contribute to the local injury of the endothelium by engagement of adhesion receptors and release of proteases or cytokines resulting in coagulation activation. Taking these results into account, we could hypothesize that the ET expression pattern found in group A, with overexpression of genes involved in the immune response, could primarily affect thrombosis mechanisms.

Gene expression patterns in relation to JAK2V617F

Supervised clustering of the microarray data of the 20 ET patients, showed some genes differentially expressed in *JAK2*V617F-positive versus *JAK2*V617F-negative ET. Most of these genes showed little expression differences in relation to controls.

We selected 40 of these genes, applying a cutoff based on the Bayesian moderated *t*-test (using Limma package,[26] $P<0.017$; fold change >1.35 between the *JAK2*V617F-postive and *JAK2*V617F-negative patients) and we designed a new LDA platform with 48 genes that included the 40 differentially expressed genes and the *SOCS1*, *SOCS3*, *CCND1* and *BCL2L1* genes which are directly or indirectly related to the JAK-STAT pathway. The *18S rRNA*, *GAPDH*, *GUSB* and *HPRT1* (hypoxanthine phosphoribosyl-transferase) genes were used as endogenous controls. Using real-time quantitative RT-PCR, this set of 48 genes was validated in 19/20 patients previously studied by microarrays, and tested in 19 additional ET patients. A *t*-test of the gene expression data of the 38 ET patients confirmed the differential expression between *JAK2*V617F-positive and *JAK2*V617F-negative patients in 8/48 genes studied ($P<0.052$): *CISH*, *C13orf18*, *CD44*, *PIM1*, *SOCS2*, *SOCS3*, *HRH4*, *BATF*. Among these eight genes, four were induced by the JAK-STAT pathway in response to cytokines: *CISH* (cytokine inducible SH2-containing protein), *SOCS2* (suppressor of cytokine signalling 2), *SOCS3* (suppressor of cytokine signalling 3) and *PIM1* (*PIM1* oncogene).[27] SOCS2 and SOCS3 were overexpressed in *JAK2*V617F-positive ET and no differences were found between *JAK2*V617F-negative ET and controls. *CISH* and *PIM1* expression levels in mutated patients were similar to controls, whereas a downregulation was found in *JAK2*V617F-negative ET patients (Table 2). These genes have been described as targets of the STATs (signal transducers and activators of transcription). *CISH*, *SOCS2* and *SOCS3* are *SOCS* family members that are activated by STATs. These three genes as well as *PIM1* act as negative regulators of the JAK-STAT signalling pathway.[28,29] As *JAK2*V617F causes constitutive activation of the JAK-STAT signalling in haematopoietic cells,[2–5] it consequently activates several negative feedback systems, mainly including members of the SOCS family, which initially might act to suppress the tumour promoting effects of the *JAK2*V617F oncogene.[30] The activation of the negative feedback of JAK-STAT pathway could explain the higher expression of *SOCS2* and *SOCS3* detected in *JAK2*V617F-positive ET. The lower expression of these two genes in *JAK2*V617F-negative ET indicates that there is no activation of the JAK-STAT pathway in these patients. The finding of similar expression levels of *CISH* and *PIM1* in mutated patients and controls could be explained by a perturbed expression or imbalances in the negative feedback system of the JAK-STAT pathway. Similarly, a recent study in MPD Philadelphia positive patients has described that BCR/ABL signalling activates several negative feedback regulatory systems (such as JAK-STAT inhibitors).[28] This effect might be overcome by increased BCR/ABL signalling or by signalling via additional or alternative pathways, which subsequently causes an increased signalling in the JAK-STAT pathway. Moreover, Verdier *et al*.[31] described that the overexpression of STAT5 overcomes the inhibitory effects of *CISH* on erythropoietin signalling. To our knowledge, this is the first time that *CISH* and *SOCS3* have been described to present different expression between *JAK2*V617F-negative and *JAK2*V617F-positive ET patients. Concerning to the relative differential expression of *SOCS2* and *PIM1*, our results agree with those recently reported by Schwemmers *et al*.[15] These authors revealed two distinct gene expression subclasses in a series of 16 ET patients in relation to *JAK2*V617F status, by using cDNA microarrays (home-made platform) analysing a reduced set of genes (7497 genes).

Regarding the other four differentially expressed genes, a higher expression of *BATF* (basic leucine zipper transcription factor, ATF-like) and *CD44* (CD44 molecule) was detected in *JAK2*V617F-positive versus *JAK2*V617F-negative ET patients. *C13orf18* (chromosome 13 open reading frame 18) and *HRH4* (histamine H4 receptor) were downregulated in *JAK2*V617F-negative patients. Interestingly, *C13orf18* expression was the best marker to discriminate *JAK2*V617F-negative from *JAK2*V617F-positive and controls (Table 2).

On the other hand, when we applied a *t*-test to the data obtained on the 96 genes LDA platform, we found 22 genes differentially expressed between *JAK2*V617F-positive and *JAK2*V617F-negative patients ($P<0.05$). All these 22 genes presented a lower expression in *JAK2*V617F-positive patients. Interestingly, the same set of genes was downregulated in the B.2 group of the unsupervised clustering obtained from the 96 genes LDA data that had a majority of cases carrying the *JAK2*V617F mutation. These findings support the hypothesis that there is an effect of mutational status of *JAK2*V617F on the granulocytic expression profile of ET patients, in agreement with other studies.[15]

This set of 22 genes, together with the 8 previously defined in 48 genes LDA, established a group of 30 genes that showed different expression in relation to *JAK2*V617F status (Table 2).

A set of genes predicts the JAK2V617F mutational status

Grouping real-time PCR results from the two LDA custom designs (137 genes plus 4 endogenous controls) we sought to obtain a gene expression signature that differentiated *JAK2*V617F-positive and *JAK2*V617F-negative ET. The predictor model was constructed on the training data set (26 samples) and evaluated on an independent test data set (12 samples). We performed an exhaustive sequential search to select the genes with the lowest error classification in the training set and optimal prediction power in the independent set. Applying successive runs of MiPP R package, using linear discriminant analysis and the settings (n.seq = 20, n.split = 100, n.fold = 10, n.split.eval = 100) and changing the number of preselected genes (using percent.cut parameter), two models with an error rate lower than 20% in test set were selected: the first model was a three gene predictor, including *CISH*, *FOSB* (FBJ murine osteosarcoma viral oncogene homologue B) and *SOCS2*, with validation error rates of 15.38% in the training set and 16.67% in the test set (percent.cut = 0.3). The second model, comprising only the genes *CISH* and *FOSB*, yielded an error rate of 11.54% in the training set and of 16.67% in the test set (percent.cut = 0.02). It is noteworthy that the two genes with highest predictive value were consistently recovered in all predictor models (Figure 3). These results highlight the power of

Table 2 Genes differentially expressed between *JAK2*V617F-positive and *JAK2*V617F-negative ET patients ($P<0.052$)

Symbol	*Gene description*	*GenBank accesion no.*	*Taqman assay*	*Fold change vs control pool*		*Differential FC JAK2V617F-positive/ JAK2V617F-negative*	P-*value*
				*JAK2*V617F-*positive*	*JAK2*V617F-*negative*		
CISH	Cytokine inducible SH2-containing protein	NM_013324	Hs00367082_g1	−1.33	−2.79	2.08	0.00062
FOSB	FBJ murine osteosarcoma viral oncogene homologue B	NM_006732	Hs00171851_m1	1.87	5.32	0.35	0.00204
C13orf18	Chromosome 13 open reading frame 18	NM_025113	Hs00228336_m1	−1.17	−1.63	1.39	0.00447
CD44	CD44 molecule (Indian blood group)	NM_000610	Hs00153304_m1	1.81	1.49	1.22	0.00500
IER3	Immediate early response 3	NM_003897	Hs00174674_m1	1.64	3.80	0.43	0.00749
HCST	Haematopoietic cell signal transducer	NM_014266	Hs00367159_m1	2.58	3.82	0.68	0.00803
CCL3	Chemokine (C-C motif) ligand 3	NM_002983	Hs00234142_m1	1.27	3.94	0.32	0.01113
PIM1	Pim-1 oncogene	NM_002648	Hs00171473_m1	−1.28	−1.77	1.37	0.01263
MAFF	v-Maf musculoaponeurotic fibrosarcoma oncogene homologue F (avian)	NM_012323	Hs00544822_m1	1.37	2.73	0.50	0.01352
PPP1R15A	Protein phosphatase 1, regulatory (inhibitor) subunit 15A	NM_014330	Hs00169585_m1	3.09	7.06	0.44	0.01379
KLF10	Kruppel-like factor 10	NM_005655	Hs00194622_m1	1.55	3.68	0.42	0.01403
SOCS3	Suppressor of cytokine signalling 3	NM_003955	Hs00269575_s1	1.46	−1.04	1.52	0.01823
ID2	Inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	NM_002166	Hs00747379_m1	−1.36	1.21	0.61	0.01891
TNFAIP3	Tumor necrosis factor, alpha-induced protein 3	NM_006290	Hs00234713_m1	1.81	5.57	0.32	0.01914
CCL4	Chemokine (C-C motif) ligand 4	NM_002984	Hs99999148_m1	2.08	6.29	0.33	0.02024
GADD45B	Growth arrest and DNA-damage-inducible, beta	NM_015675	Hs00169587_m1	1.10	2.15	0.51	0.02560
TNFAIP6	Tumor necrosis factor, alpha-induced protein 6	NM_007115	Hs00200180_m1	−1.42	1.22	0.57	0.02565
TA-NFKBH	T-cell activation NFKB-like protein	NM_139239	Hs00262018_m1	−1.40	1.57	0.45	0.02767
CXCL2	Chemokine (C-X-C motif) ligand 2	NM_002089	Hs00601975_m1	1.60	6.98	0.23	0.02775
KLF5	Kruppel-like factor 5 (intestinal)	NM_001730	Hs00156145_m1	1.38	1.98	0.70	0.03059
TNF	Tumor necrosis factor (TNF superfamily, member 2)	NM_000594	Hs00174128_m1	1.40	2.69	0.52	0.03283
PLEKHG2	Pleckstrin homology domain containing, family G (with RhoGef domain) member 2	NM_022835	Hs00293943_m1	−1.51	2.79	0.24	0.03572
DDIT4	DNA-damage-inducible transcript 4	NM_019058	Hs00430304_g1	1.56	2.49	0.63	0.04051
LAMB3	Laminin, beta 3	NM_000228	Hs00165078_m1	1.31	2.29	0.57	0.04114
HRH4	Histamine receptor H4	NM_021624	Hs00222094_m1	−1.32	−2.28	1.73	0.04516
TAGAP	T-cell activation GTPase activating protein	NM_138810	Hs00299284_m1	1.43	2.22	0.65	0.04650
BATF	Basic leucine zipper transcription factor, ATF-like	NM_006399	Hs00232390_m1	1.91	1.53	1.25	0.04701
TRIB1	Tribbles homologue 1 (*Drosophila*)	NM_025195	Hs00179769_m1	1.18	1.48	0.80	0.04993
DUSP5	Dual specificity phosphatase 5	NM_004419	Hs00244839_m1	2.71	5.05	0.54	0.05023
SOCS2	Suppressor of cytokine signalling 2	NM_003877	Hs00153304_m1	1.96	−1.06	2.09	0.05126

The fold change (FC) were subtracted from de $2^{\Delta\Delta Ct}$ real-time quantitative RT-PCR values.
Genes involved in JAK-STAT pathway are in bold.

CISH and *FOSB* for the discrimination of ET patients according to *JAK2*V617F status. *CISH* is a negative regulator of the JAK-STAT signalling pathway and its expression could be altered by *JAK2*V617F. *FOSB* is an oncogene involved in the regulation of progression through cell cycle. *FOSB* upregulation in ET *JAK2*V617F-negative patients may indicate its implication in the molecular pathology of this group of ET patients. Interestingly, the differential expression of this set of genes in *JAK2*V617F-negative patients could be useful to identify other pathways besides JAK-STAT in the pathogenesis of ET.

Gene expression signature in JAK2*V617F-negative patients*

Among *JAK2*V617F-negative ET patients (that represent approximately 50% of ET), a proportion of them do not present any clonal marker, such as HUMARA clonality or endogenous myeloid growth. It would be interesting to find new biological markers that could characterize this group of patients. In our study, we defined a set of genes differentially expressed in *JAK2*V617F-negative ET patients versus *JAK2*V617F-positive and controls. Among them, the following genes showed an abnormal expression pattern (with >1.5-fold changes over control granulocytes) in *JAK2*V617F-negative patients: *CISH*, *C13orf18*, *CCL3*, *PIM1*, *MAFF*, *SOCS3*, *ID2*, *GADD45B*, *KLF5*, *TNF*, *LAMB3*, *HRH4*, *TAGAP* and *TRIB1* (Table 2). Noteworthy, the expression of these genes was not altered either in *JAK2*V617F-positive ET patients or in controls. The identification of a characteristic gene expression signature in this group of patients is worthy of consideration to identify other signal transduction pathways besides JAK-STAT involved in the pathogenesis of ET.

In conclusion, our results showed a heterogeneous molecular signature in ET patients characterized by two main gene expression patterns: one (group A) with an upregulation of inflammatory markers that could participate in neutrophil activation and thrombosis, and the other (group B) with significantly lower expression of these genes. Interestingly, we found a group of 30 genes that showed a different expression in relation to *JAK2*V617F status. Among them, a set of 14 genes (*CISH*, *C13orf18*, *CCL3*, *PIM1*, *MAFF*, *SOCS3*, *ID2*, *GADD45B*, *KLF5*, *TNF*, *LAMB3*, *HRH4*, *TAGAP* and *TRIB1*) showed an abnormal expression pattern only in *JAK2*V617F-negative. Remarkably, in this group of patients *CISH*, *SOCS2*, *SOCS3* and *PIM1* presented a significant lower expression, all of them involved in the JAK-STAT signalling pathway. A two-gene predictor model comprising *FOSB* and *CISH* genes, which were the best discriminators of *JAK2*V617F status, was built. The identification of a gene expression signature characteristic in this group of patients is interesting to identify other signal transduction pathways besides JAK-STAT involved in the pathogenesis of ET. These data would provide new biological markers that characterize *JAK2*V617F-negative ET patients. Additional biological investigations are required to confirm these results.

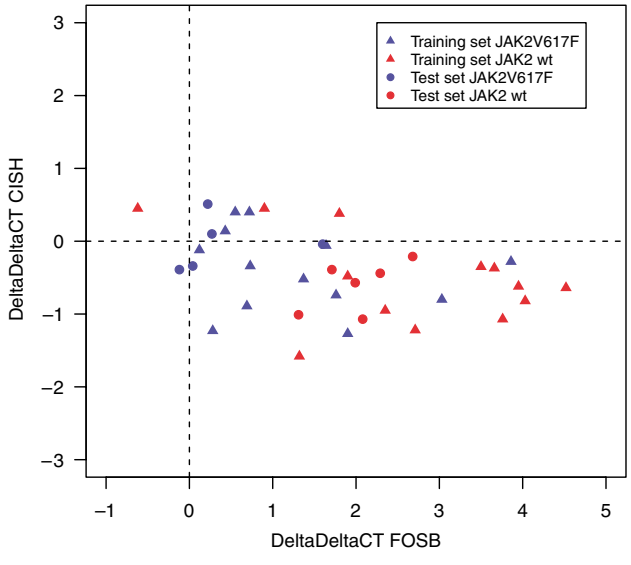

Figure 3 Diagram showing the relative expression of CISH and FOSB versus the control pool in the training data set (26 samples; 13 JAK2V617F-positive and 13 JAK2V617F-negative) and the independent test data set (12 samples; 6 JAK2V617F-positive and 6 JAK2V617F-negative). These genes show a good separation between the two JAK2 status groups.

Acknowledgements

This work was supported by grants FIS PI030345 and PI071009 from the Spanish Ministry of Health and Beca de Investigación FEHH. We thank Manuel Delgado, Norma Carmen Gutiérrez and Jesús Mª Hernández-Rivas from Servicio de Hematología, Hospital Universitario de Salamanca and Centro de Investigación del Cáncer (CIC), for their help in microarray data analyses and Josep M Manresa from Assessorament Metodològic en Investigació Biomèdica (AMIB) Department, IMIM-Hospital del Mar, for the statistical analysis of the clinical data.

Authorship contribution: EP processed the samples, carried out the microarray study and real-time quantitative RT-PCR analysis, and wrote the manuscript; BE designed the study and wrote the paper; JJL, LS performed the statistical analysis of the microarray and the real-time quantitative RT-PCR data; BB carried out the *JAK2* analysis and provided molecular biology laboratory support; LA provided basic laboratory support; AA and CB collected the patients data; FS performed cytogenetics studies; SS was head of the laboratory that was involved in the biology studies; LF designed the study and wrote the paper. All authors reviewed and accepted the manuscript.

References

1 Tefferi A, Vardiman JW. Classification and diagnosis of myeloproliferative neoplasms: the 2008 World Health Organization criteria and point-of-care diagnostic algorithms. *Leukemia* 2008; **22**: 14–22.
2 Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S *et al.* Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 2005; **365**: 1054–1061.
3 Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ *et al.* Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 2005; **7**: 387–397.
4 James C, Ugo V, Le Couedic JP, Staerk J, Delhommeau F, Lacout C *et al.* A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* 2005; **434**: 1144–1148.
5 Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR *et al.* A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 2005; **352**: 1779–1790.
6 Campbell PJ, Baxter EJ, Beer PA, Scott LM, Bench AJ, Huntly BJ *et al.* Mutation of JAK2 in the myeloproliferative disorders: timing, clonality studies, cytogenetic associations, and role in leukemic transformation. *Blood* 2006; **108**: 3548–3555.
7 Campbell PJ, Scott LM, Buck G, Wheatley K, East CL, Marsden JT *et al.* Definition of subtypes of essential thrombocythaemia and

relation to polycythaemia vera based on JAK2 V617F mutation status: a prospective study. *Lancet* 2005; **366**: 1945–1953.
8 Scott LM, Tong W, Levine RL, Scott MA, Beer PA, Stratton MR *et al*. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *N Engl J Med* 2007; **356**: 459–468.
9 Martínez-Avilés L, Besses C, Alvarez-Larrán A, Cervantes F, Hernández-Boluda JC, Bellosillo B. JAK2 exon 12 mutations in polycythemia vera or idiopathic erythrocytosis. *Haematologica* 2007; **92**: 1717–1718.
10 Pikman Y, Lee BH, Mercher T, McDowell E, Ebert BL, Gozo M *et al*. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med* 2006; **3**: e270.
11 Pellagatti A, Vetrie D, Langford CF, Gama S, Eagleton H, Wainscoat JS *et al*. Gene expression profiling in polycythemia vera using cDNA microarray technology. *Cancer Res* 2003; **63**: 3940–3944.
12 Goerttler PS, Kreutz C, Donauer J, Faller D, Maiwald T, Marz E *et al*. Gene expression profiling in polycythaemia vera: overexpression of transcription factor NF-E2. *Br J Haematol* 2005; **129**: 138–150.
13 Steidl U, Schroeder T, Steidl C, Kobbe G, Graef T, Bork S *et al*. Distinct gene expression pattern of malignant hematopoietic stem and progenitor cells in polycythemia vera. *Ann NY Acad Sci* 2005; **1044**: 94–108.
14 Kralovics R, Teo SS, Buser AS, Brutsche M, Tiedt R, Tichelli A *et al*. Altered gene expression in myeloproliferative disorders correlates with activation of signaling by the V617F mutation of Jak2. *Blood* 2005; **106**: 3374–3376.
15 Schwemmers S, Will B, Waller CF, Abdulkarim K, Johansson P, Andreasson B *et al*. JAK2V617F-negative ET patients do not display constitutively active JAK/STAT signaling. *Exp Hematol* 2007; **35**: 1695–1703.
16 Florensa L, Bellosillo B, Besses C, Puigdecanet E, Espinet B, Perez-Vila E *et al*. JAK2 V617F mutation analysis in different myeloid lineages (granulocytes, platelets, CFU-MK, BFU-E and CFU-GM) in essential thrombocythemia patients. *Leukemia* 2006; **20**: 1903–1905.
17 Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; **3**: Article3.
18 Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S *et al*. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; **5**: R80.
19 Al-Shahrour F, Minguez P, Tárraga J, Montaner D, Alloza E, Vaquerizas JMM *et al*. BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Res* 2006; **34**: W472–W476.
20 Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A *et al*. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002; **3**: RESEARCH0034.
21 Soukup M, Cho H, Lee JK. Robust classification modeling on microarray data using misclassification penalized posterior. *Bioinformatics* 2005; **21** (Suppl): i423–i430.
22 Moll J, Khaldoyanidi S, Sleeman JP, Achtnich M, Preuss I, Ponta H *et al*. Two different functions for CD44 proteins in human myelopoiesis. *J Clin Invest* 1998; **102**: 1024–1034.
23 Rameshwar P, Chang VT, Gascón P. Implication of CD44 in adhesion-mediated overproduction of TGF-beta and IL-1 in monocytes from patients with bone marrow fibrosis. *Br J Haematol* 1996; **93**: 22–29.
24 Falanga A, Marchetti M, Vignoli A, Balducci D, Russo L, Guerini V *et al*. V617F JAK-2 mutation in patients with essential thrombocythemia: relation to platelet, granulocyte, and plasma hemostatic and inflammatory molecules. *Exp Hematol* 2007; **35**: 702–711.
25 Arellano-Rodrigo E, Alvarez-Larran A, Reverter JC, Villamor N, Colomer D, Cervantes F. Increased platelet and leukocyte activation as contributing mechanisms for thrombosis in essential thrombocythemia and correlation with the JAK2 mutational status. *Haematologica* 2006; **91**: 169–175.
26 Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds). *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Springer: New York, 2005, pp 397–420.
27 Ward AC, Touw I, Yoshimura A. The Jak-Stat pathway in normal and perturbed hematopoiesis. *Blood* 2000; **95**: 19–29.
28 Håkansson P, Nilsson B, Andersson A, Lassen C, Gullberg U, Fioretos T. Gene expression analysis of BCR/ABL1-dependent transcriptional response reveals enrichment for genes involved in negative feedback regulation. *Genes Chromosomes Cancer* 2008; **47**: 267–275.
29 Peltola KJ, Paukku K, Aho TL, Ruuska M, Silvennoinen O, Koskinen PJ. Pim-1 kinase inhibits STAT5-dependent transcription via its interactions with SOCS1 and SOCS3. *Blood* 2004; **103**: 3744–3750.
30 Cooney RN. Suppressors of cytokine signaling (SOCS): inhibitors of the JAK/STAT pathway. *Shock* 2002; **17**: 83–90.
31 Verdier F, Chrétien S, Muller O, Varlet P, Yoshimura A, Gisselbrecht S *et al*. Proteasomes regulate erythropoietin receptor and signal transducer and activator of transcription 5 (STAT5) activation. Possible involvement of the ubiquitinated Cis protein. *J Biol Chem* 1998; **273**: 28185–28190.

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)