Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Dissociation of the JAK2 V617F mutation from STAT transcriptional activation in polycythemia vera: erythroid expansion as the dominant transcriptomic signature

yu wang
Chinese PLA General Hospital
yu wang
Chinese PLA General Hospital
wei fei
Tibet
yukun wang
Chinese PLA General Hospital
tianhao wang
Chinese PLA General Hospital
huiru an
anhuiru74@163.com

Chinese PLA General Hospital

Research Article

Keywords: polycythemia vera, JAK2 V617F, JAK-STAT, EPAS1, transcriptome

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11766858/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The *JAK2* V617F mutation drives polycythemia vera (PV) and is canonically thought to act through constitutive signal transducer and activator of transcription (STAT) signaling, yet whether the STAT transcriptional axis is actually engaged in primary patient samples remains unclear. We reanalyzed five public microarray datasets spanning whole blood and bone marrow CD34 + progenitor cells, together with a single-cell 3'-TARGET-seq dataset, scoring four curated gene sets (erythroid, *JAK2*-STAT, HIF-*EPO*, inflammatory) and performing ranked gene-set enrichment analysis with multiple independent STAT gene sets, followed by cell-composition adjustment. The erythroid program was strongly activated in PV (NES + 2.52; meta-analytic d + 1.30), whereas the STAT program was not: multiple STAT gene sets showed no enrichment, and direct STAT5 targets, including canonical pSTAT5 targets *BCL2L1*, *PIM1*, and *CISH*, showed no consistent upregulation (d + 0.07). *EPAS1* (*HIF2A*) was upregulated across datasets (d + 0.98) but was largely attenuated after adjusting for the increased erythroid cell fraction (adjusted p = 0.613), indicating compositional confounding; within-patient single-cell analysis from a myelofibrosis cohort (used as indirect supporting evidence) corroborated the STAT dissociation while erythroid markers tracked the mutant clone. At the transcriptomic level, the *JAK2* V617F mutation is dissociated from STAT transcriptional activation in PV, and bulk *EPAS1*/*HIF2A* upregulation largely reflects erythroid expansion, cautioning against cell-intrinsic interpretations without composition-aware analysis.

1. Introduction

Philadelphia-negative myeloproliferative neoplasms (MPNs) comprise polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF), which share a recurrent driver mutation in the tyrosine kinase *JAK2*, most commonly the V617F substitution, which is present in approximately 95% of PV and 50–60% of ET and PMF [1–4]. Despite this shared genetic lesion, the three diseases display divergent phenotypes: erythrocytosis in PV, thrombocytosis in ET, and bone marrow fibrosis in PMF. This divergence cannot be explained by the mutation alone and implies that distinct downstream transcriptional programs are engaged in each subtype [25, 26].

The canonical model of *JAK2* V617F pathogenesis posits constitutive activation of the JAK/STAT signaling cascade, leading to STAT-mediated transcription of proliferation genes [5–7, 24]. However, a growing body of evidence suggests that this model is incomplete. Transcriptomic studies of MPN have produced heterogeneous and sometimes contradictory results, and the extent to which the STAT transcriptional program is genuinely activated in primary patient samples, as opposed to being inferred from phospho-STAT measurements, has not been systematically established [8–11].

Erythropoiesis is governed by lineage-determining programs in which hypoxia-inducible factor 2 (*HIF2A*, encoded by *EPAS1*) is the principal transcriptional regulator of erythropoietin and erythroid differentiation [12–14, 28], and gain-of-function *EPAS1* mutations cause familial erythrocytosis [15]. *EPAS1* therefore marks the erythroid program, but whether its reported upregulation in PV reflects a cell-intrinsic molecular mechanism or simply the expanded erythroid compartment has not been resolved.

A second, more general issue is that bulk transcriptomic studies of PV have rarely adjusted for cell composition. Because PV expands the erythroid compartment, any bulk signal is a weighted average over cell types whose proportions shift with disease, so an erythroid-lineage gene such as *EPAS1* will appear "upregulated" even if its expression per cell is unchanged. Distinguishing such compositional confounding from a cell-intrinsic transcriptional response is essential, yet has been under-appreciated in the PV literature [18].

Here we test the STAT hypothesis at the transcriptomic level using a cross-platform reanalysis of five PV datasets spanning whole blood and CD34 + progenitor cells, asking whether the STAT axis is activated in PV, whether the PV transcriptome is instead defined by erythroid expansion, and whether *EPAS1* upregulation survives adjustment for cell composition.

2. Materials and Methods

2.1 Datasets and preprocessing

We curated six public bulk transcriptomic datasets (Table 1): whole blood (GSE26049 discovery; GSE61629 and GSE57793 same-platform extension), CD34 + progenitor cells (GSE47018 peripheral; GSE103237 bone marrow), and an siRNA cell-line model (GSE27523). For single-cell validation we additionally analyzed GSE122198, a 3'-TARGET-seq dataset of 2,798 CD34 + progenitors with per-cell *JAK2* V617F genotyping (see 2.5). Data were retrieved from the Gene Expression Omnibus through refine.bio and summarized at the gene level.

Table 1
Datasets used in this study.

Dataset	Platform	Cell type	Groups (n)
GSE26049	GPL570	Whole blood	PV 41, ET 19, PMF 9, Ctrl 21
GSE61629	GPL570	Whole blood	PV 17, ET 7, PMF 3, Ctrl 18
GSE57793	GPL570	Whole blood	PV 36, ET 15, PMF 7
GSE47018	hgu133a	CD34+	PV 20, Ctrl 7
GSE103237	U219	Bone marrow CD34+	PV 26, ET 24, Ctrl 15
GSE27523	HuEx-1_0-st	LN229 cell line	HIF1A/HIF2A siRNA

PV, polycythemia vera; ET, essential thrombocythemia; PMF, primary myelofibrosis; Ctrl, healthy controls; CD34+, CD34-positive progenitor cells; siRNA, small interfering RNA.

2.2 Gene-set scoring and ranked gene-set enrichment analysis

We scored four curated gene sets (erythroid, *JAK2*-STAT, HIF-*EPO*, and inflammatory) as the mean per-sample z-score of member genes (Supplementary Methods). Differential expression was assessed by Welch's t-test, and genes ranked by the t statistic were subjected to Kolmogorov–Smirnov-style gene-set enrichment analysis (GSEA) with 1,000 permutations [16]. To exclude gene-set selection bias, five independent STAT-related gene sets from MSigDB were tested alongside the four axes as positive controls.

2.3 Cross-platform replication and meta-analysis

For each signature gene, Cohen's d between PV and control was computed after per-sample z-scoring, with replication requiring concordant direction across datasets. Effect estimates were pooled across the four datasets with PV and control groups using a DerSimonian-Laird random-effects model [32], with between-dataset heterogeneity quantified by I^2 and τ^2.

2.4 Cell-composition deconvolution

Cell composition was estimated by lineage marker-gene scoring and removed by linear-regression residualization (Supplementary Methods), distinguishing compositional confounding from cell-intrinsic transcriptional change. The procedure was benchmarked against a mitochondrial oxidative-phosphorylation (OXPHOS) gene set as a cell-intrinsic positive control, and two permutation tests assessed the specificity of adjusted effects.

2.5 *JAK2* V617F genotype, allele burden, and single-cell analyses

Within GSE103237 CD34 + cells, *JAK2* V617F-positive and V617F-negative (*CALR*-mutated) MPN were compared by ranked GSEA; in GSE47018, genome-wide Spearman correlations between expression and V617F burden were ranked and tested by GSEA. For single-cell validation, GSE122198 (3'-TARGET-seq, 2,798 CD34 + cells with per-cell *JAK2* genotyping) was analyzed by within-patient paired comparisons of mutant versus wild-type cells (Wilcoxon signed-rank; Supplementary Methods).

3. Results

3.1 Erythroid and hypoxia–erythropoietin programs are elevated in polycythemia vera, whereas the *JAK2*–STAT program is unchanged

Following the analysis pipeline summarized in Supplementary Figure S1, four-axis scoring of the discovery cohort (GSE26049) revealed a clear pattern (Fig. 1). The erythroid axis was markedly elevated in all three MPN subtypes relative to controls (PV 2.38, ET 2.30, PMF 3.12 vs control 1.60; $p = 2.3\times10^{-5}$), consistent with the shared erythroid lineage expansion. The HIF-*EPO* axis was specifically highest in PV (PV 0.27 vs ET 0.16, PMF 0.15, control 0.13; $p = 3.5\times10^{-7}$). In contrast, the *JAK2*-STAT axis was identical

between PV and controls (1.24 vs 1.24) and lowest in PMF (1.02); the four-group Kruskal–Wallis test was significant ($p = 4.9\times10^{-4}$), but this reflected the low PMF score rather than any PV–control difference. These data indicate that, at the transcriptomic level, the STAT program is not activated in PV despite the presence of the *JAK2* V617F mutation. As a secondary observation, the inflammatory axis was highest in PMF (3.36), consistent with its pro-inflammatory and fibrotic nature (Supplementary Results S3).

3.2 *EPAS1* (*HIF2A*) and *ADM* replicate across four datasets and two cell types

We next examined individual genes, beginning with the key signature genes in the discovery cohort (Fig. 2A). *EPAS1* (*HIF2A*) and *ADM* were consistently upregulated in PV relative to controls across all four datasets for which controls were available, spanning whole blood (GSE26049, GSE61629), CD34 + peripheral cells (GSE47018), and bone marrow CD34 + progenitor cells (GSE103237) (Fig. 2B); the cross-dataset and cross-subtype distribution of *EPAS1* is shown in Fig. 2C, and its independent validation with *ADM* in bone marrow CD34 + cells is shown in Fig. 3A–B. Cohen's d for *EPAS1* was 1.08, 0.77, 0.77, and 0.88 ($p = 0.001$ in GSE103237), and for *ADM* was 1.09, 1.06, 0.58, and 0.97 ($p < 0.001$ in GSE103237). *MYC* was consistently downregulated in whole blood (d = − 1.76, − 1.29, − 0.41). By contrast, *CA9* and *SLC2A1*, although upregulated in whole blood (d = 0.74–1.16), showed no elevation in bone marrow CD34 + cells (d = − 0.77 and − 0.05, respectively), indicating that these genes mark a whole-blood-specific component of the HIF response rather than the CD34 + progenitor-intrinsic program. As shown below (see 3.5), the bulk upregulation of *EPAS1* is further attenuated after adjustment for cell composition, indicating a substantial contribution from the expanded erythroid compartment.

3.3 Ranked gene-set enrichment analysis confirms erythroid enrichment and absence of STAT enrichment

Whole-genome GSEA of the PV-versus-control comparison (Fig. 1) showed strong enrichment of the erythroid gene set among PV-upregulated genes (NES = + 2.52, $p = 0.001$). The inflammatory gene set was also enriched (NES = + 1.66, $p = 0.031$), consistent with the known inflammatory milieu of MPNs. The HIF-*EPO* gene set showed a positive but non-significant trend (NES = + 1.34, $p = 0.138$), reflecting the divergent behavior of its constituent genes (*EPAS1* and its co-regulated targets upregulated, while *HIF1A*-classical targets such as *EPO*, *BNIP3*, and *PDK1* were not upregulated). Importantly, the *JAK2*-STAT gene set showed no enrichment (NES = + 0.74, $p = 0.793$), independently confirming that the STAT transcriptional program is not engaged in PV.

Analyses of *HIF1A*/*HIF2A* target specificity (siRNA), diagnostic discrimination, and the PMF inflammatory gradient are reported in Supplementary Results S1, S2, and S3, respectively.

3.4 Multiple independent STAT gene sets confirm the absence of STAT transcriptional activation

To exclude gene-set selection bias, we repeated the GSEA using five independent STAT-related gene sets (Fig. 3). None was positively enriched in PV relative to controls: HALLMARK_IL6_JAK_STAT3_SIGNALING (NES = + 1.17, p = 0.25), AZARE_STAT3_TARGETS (− 1.21, p = 0.23), BAKER_STAT3_TARGETS (− 1.01, p = 0.41), BAKER_STAT5_TARGETS (− 1.10, p = 0.35), and the STAT5 direct-target set (− 1.91, p = 0.006). Because this analysis re-scored the positive-control axes with an independent set of 1,000 permutations, their NES values differ slightly from 3.3 (e.g., erythroid + 2.45 vs + 2.52), reflecting permutation variance rather than different gene-set definitions. The single-cohort GSEA suggested negative enrichment of direct STAT5 targets, but a meta-analytic assessment across four datasets (3.6) showed no consistent upregulation in PV (d = + 0.07, 95% CI − 0.95 to + 1.09, I^2 = 89%), indicating transcriptional quiescence with directionally inconsistent estimates rather than robust downregulation. Consistent with this, *EPAS1* did not differ between *JAK2* V617F-positive and V617F-negative MPN samples in CD34 + cells (d = + 0.04, p = 0.35), supporting *EPAS1* as an erythroid marker rather than a direct V617F target. None of the canonical pSTAT5 ChIP-seq direct targets *BCL2L1*, *PIM1*, or *CISH* was consistently upregulated across the four datasets (all p > 0.3) [20, 27].

3.5 Cell-composition deconvolution shows that *EPAS1* upregulation is largely attenuated by adjustment for erythroid expansion

Marker-gene deconvolution of the whole-blood discovery cohort (GSE26049) revealed the expected PV cell-composition changes: increased erythroid (d = + 0.95, p < 0.001) and neutrophil (d = + 0.60) fractions, with reduced lymphoid (T-cell d = − 1.17; B-cell d = − 1.26) and monocyte (d = − 0.59) fractions (Fig. 4). After adjusting for these fractions, *EPAS1* upregulation in whole blood was no longer significant (adjusted mean difference = + 0.012, p = 0.613), and the same held for *ADM* (p = 0.149) and *EPOR* (p = 0.206). In the bone marrow CD34 + cohort (GSE103237), erythroid-progenitor content was markedly increased in PV (d = + 1.12, p < 0.001), and *EPAS1* upregulation was again attenuated once this fraction was accounted for (adjusted p = 0.126) (Fig. 4). In the peripheral CD34 + cohort (GSE47018), *EPAS1* remained elevated after full adjustment (d = + 0.21, p < 0.001), though its control group is small (n = 7). These results indicate that the bulk *EPAS1* signal in PV is substantially confounded by the expanded erythroid compartment, consistent with *EPAS1* as an erythroid-lineage marker, while the question of cell-intrinsic *EPAS1* upregulation cannot be cleanly resolved from bulk or CD34 + data because *EPAS1* is itself an erythroid gene and our marker-gene residualization provides only a qualitative, not quantitative, adjustment. By contrast, *STAT5A*/*STAT5B* mRNA was not elevated in CD34 + progenitors (GSE103237: d = − 0.07 and − 0.02, both p > 0.1), and the STAT5 target program remained non-activated in whole blood after composition adjustment, indicating that the quiescence of the STAT axis is not a composition artifact.

To confirm that composition adjustment does not indiscriminately erase all signals, we benchmarked the procedure against cell-intrinsic biology. Mitochondrial oxidative-phosphorylation (OXPHOS) genes remained significantly upregulated after adjustment (*NDUFA3* d = + 1.00, p = 0.0002; *COX5B* d = + 1.04, p = 0.0006; *UQCR10* d = + 0.87, p = 0.001), exposing a cell-intrinsic metabolic program consistent with *JAK2*

V617F-driven rewiring (Figure S5), whereas lineage-related bulk signals (interferon-response, myeloid-innate) attenuated after adjustment. Two permutation tests established the specificity of the attenuation: the adjusted erythroid-module effect (d = + 0.33) fell within the null of 1,000 random gene sets (p = 0.63), the adjusted *EPAS1* effect was not significant by label permutation (p = 0.54), whereas the adjusted STAT5 direct-target effect (d = − 0.18) was significantly below the null (p = 0.032), indicating the STAT5 target program is relatively quiescent (Figure S6). To further rule out over-correction by the multi-component model, we adjusted *EPAS1* for the erythroid fraction alone: the effect fell from d = + 0.093 (p = 0.0002) to d = + 0.050 (p = 0.053), indicating that the erythroid compartment, not the simultaneous adjustment of myeloid and lymphoid fractions, accounts for most of the attenuation (Figure S7).

3.6 Meta-analysis confirms robust erythroid activation and absent STAT5-target activation across four datasets

To obtain pooled effect estimates and reduce single-dataset bias, we combined Cohen's d estimates across the four datasets with PV and control groups using a DerSimonian-Laird random-effects model (Fig. 5). The erythroid module was robustly upregulated with low heterogeneity (d = + 1.30, 95% CI + 0.89 to + 1.70, p < 0.001, I^2 = 22%), as were *ADM* (d = + 1.21, I^2 = 0%) and *EPAS1* (d = + 0.98, I^2 = 0%), whereas the STAT5 direct-target set showed no significant pooled effect and marked heterogeneity (d = + 0.07, 95% CI − 0.95 to + 1.09, p = 0.90, I^2 = 89%). A sensitivity analysis restricted to the two fully independent datasets (GSE47018 and GSE103237) preserved this contrast: the erythroid module remained robustly upregulated (d = + 1.55, p < 0.001), whereas STAT5 direct targets showed no consistent upregulation (d = + 0.74, 95% CI − 0.52 to + 2.01, p = 0.248, I^2 = 81%); the positive point estimate was driven by GSE103237 (d = + 1.36) while GSE47018 was near zero, so it reflects between-dataset heterogeneity rather than a consistent upregulation.

3.7 STAT5 target expression is uncorrelated with *JAK2* V617F allele burden

In GSE47018 (CD34 + PV, n = 20, allele burden 55–100%), the STAT5 direct-target score showed no correlation with *JAK2* V617F allele burden (Spearman ρ = −0.005, p = 0.98), and neither did the erythroid module (ρ = −0.055, p = 0.82) nor *EPAS1* (ρ = −0.058, p = 0.81) (Fig. 5). Although this PV-only cohort is small and lacks a low-burden comparison group, the complete absence of a dose–response relationship between mutation burden and STAT target expression is consistent with the conclusion that higher *JAK2* V617F dosage does not engage the STAT transcriptional program.

3.8 Genotype-matched and dose-response analyses confirm that *JAK2* V617F does not engage the STAT program

To test the STAT-dissociation claim in a genotype-matched design that is not confounded by cell composition, we compared *JAK2* V617F-positive and V617F-negative (*CALR*-mutated) MPN samples

within the bone marrow CD34 + cohort (GSE103237, n = 43 vs 7). Whole-transcriptome ranked GSEA showed no enrichment of the STAT5 direct-target set (NES = − 0.82, p = 0.68) or the *JAK2*-STAT axis (NES = − 0.58, p = 0.95), indicating that the STAT transcriptional program does not differ between *JAK2* V617F-mutated and *CALR*-mutated progenitors (Fig. 3). In a parallel dose-response analysis of the GSE47018 CD34 + PV cohort, genome-wide gene-by-allele-burden Spearman correlations ranked by effect size showed no enrichment of the STAT5 direct-target set (NES = − 0.74, p = 0.80), confirming that higher mutation burden does not engage the STAT program. These analyses are underpowered by the small negative-control group (n = 7) and the narrow burden range (55–100%), but their consistency with the bulk findings strengthens the transcriptome-level dissociation claim.

3.9 Single-cell analysis provides indirect supportive evidence that JAK2 V617F does not engage the STAT program while erythroid markers are elevated

To resolve the STAT dissociation at single-cell resolution and control for inter-patient heterogeneity, we analyzed a 3'-TARGET-seq dataset (GSE122198) in which the *JAK2* V617F genotype of each of 2,798 CD34 + progenitors was measured in parallel with its transcriptome [21]. This dataset derives from myelofibrosis (PMF) rather than PV, so it provides indirect, disease-related supporting evidence rather than PV-specific single-cell validation. The rationale for this cross-disease design is that bulk PV transcriptomes cannot separate a change occurring within *JAK2* V617F-mutant cells from a change driven by the expanded erythroid cell fraction; the within-patient mutant-versus-wild-type comparison available in GSE122198 isolates the cell-intrinsic transcriptional effect of the mutation itself in the progenitor compartment, a mutation-intrinsic property that is expected to transfer across MPN subtypes even though the surrounding PMF microenvironment differs from that of PV. Within-patient paired comparison of *JAK2* V617F-mutant and wild-type cells (six patients with ≥ 5 cells per class) showed no difference in STAT5 direct-target expression (Wilcoxon p = 0.109) or the *JAK2*-STAT axis (p = 0.219), and no allele-dose effect: the homozygous-minus-wild-type difference was ≈ 0 (p = 0.31), with a homozygous-minus-heterozygous difference of − 0.027 (p = 0.5). By contrast, erythroid markers were elevated in mutant cells (p = 0.047) (Figure S4; Supplementary Results S4). This cell-intrinsic comparison, free of the cell-composition and inter-patient confounding that affects bulk data, provides supportive evidence that the *JAK2* V617F mutation does not, by itself, engage the STAT transcriptional program in the progenitor compartment, while erythroid-lineage markers track the mutant clone.

4. Discussion

The central finding of this study is that, in bulk patient-derived transcriptomic profiles, the *JAK2* V617F mutation and the STAT transcriptional program are dissociated in PV. We emphasize that this conclusion is drawn at the transcriptomic level: it does not demonstrate protein-level uncoupling, and it does not exclude constitutive STAT protein phosphorylation, transient STAT signaling, or STAT activation restricted to minor cell subpopulations. The PV transcriptome is defined by erythroid expansion, manifested as erythroid gene-set enrichment and, in bulk, *EPAS1* upregulation, in the absence of STAT transcriptional activation. This pattern emerged consistently across multiple datasets, cell types, and

platforms and was confirmed with multiple independent STAT gene sets, which substantially reduces the probability that the observation reflects gene-set selection bias or technical artifact. Meta-analytic pooling of effect sizes across the four datasets reinforced this contrast: the erythroid module was robustly upregulated with low heterogeneity (random-effects d = + 1.30, I^2 = 22%), whereas the STAT5 target set was not (d = + 0.07, I^2 = 89%), and STAT5 target expression was uncorrelated with *JAK2* V617F allele burden, indicating that the quiescence of the STAT axis is a genuine transcriptomic state rather than a consequence of cell composition or mutation dosage.

The absence of STAT transcriptional activation in PV is notable because it conflicts with the canonical model of *JAK2* V617F pathogenesis. While phospho-STAT5 measurements demonstrate constitutive STAT phosphorylation in MPN cells [5, 6], our data show that this phosphorylation does not translate into a detectable STAT target-gene program in bulk patient transcriptomes. This discrepancy has been noted in prior MPN transcriptomic studies [8–11] and may reflect a relatively modest transcriptional output of STAT signaling compared with the dominant lineage-determining programs (erythroid, myeloid, inflammatory) that scale with disease burden. Consistent with this, genome-wide pSTAT5 ChIP-seq in erythroid cells identified a limited set of direct targets whose immediate transcriptional response is largely confined to proliferation and survival genes (e.g., *BCL2L1*, *PIM1*, *CISH*) [20, 27]. Importantly, STAT protein phosphorylation and STAT-mediated target-gene transcription are therefore not equivalent readouts: constitutive kinase phosphorylation can coexist with a weak, delayed, or cell-type-restricted transcriptional output, which reconciles the phospho-STAT evidence with the absence of a STAT transcriptomic signature in bulk PV samples. Several mechanisms could underlie this dissociation. First, STAT-driven transcription may be transient or burst-like rather than sustained; a short-lived transcriptional pulse would be averaged out in steady-state bulk transcriptomes, so a failure to detect a STAT signature does not rule out a transient STAT response. Second, STAT5 promoter occupancy may be insufficient to initiate transcription without cooperating lineage-specific transcription factors: in erythroid cells, the direct targets and pro-survival function of STAT5 depend on the erythroid transcriptional context (e.g., *GATA1*) [20, 27], so STAT5-bound promoters lacking these cofactors would not be efficiently transcribed. Third, the STAT axis is tightly restrained by negative feedback (*SOCS1*/*SOCS3* are themselves STAT target genes [5, 6]), so a brief transcriptional pulse could be rapidly shut off and rendered undetectable in bulk samples. Under any of these scenarios, phospho-STAT5 abundance is not equivalent to target-gene transcription, which is precisely the level at which our transcriptome-level analysis operates.

The genotype-matched and dose-response analyses reinforce this interpretation. Within the bone marrow CD34 + cohort, the STAT program did not differ between *JAK2* V617F-positive and *CALR*-mutated MPN (3.8), and STAT target expression showed no relationship to mutation burden. Importantly, our data do not identify the transcriptional program that *JAK2* V617F directly engages; they establish what it does not engage (STAT) and what dominates the bulk PV transcriptome (erythroid expansion). We therefore refrain from attributing the erythroid phenotype to any specific V617F-driven transcriptional output beyond the erythroid lineage program itself.

A further caveat is that bulk transcriptomes cannot exclude cell-type-specific STAT activation. Independent single-cell evidence, however, points in the same direction as our bulk finding. Van Egeren and colleagues performed joint single-cell RNA-seq and *JAK2* genotyping of CD34+-enriched bone marrow cells from patients with ET and PV and found that the V617F mutation increased the expression of interferon-response genes (including HLA molecules) and the leptin receptor, rather than classical STAT5 target genes [19]. A single-cell 3'-TARGET-seq analysis (from a myelofibrosis cohort, used as indirect supporting evidence) further addresses this caveat: within-patient paired comparisons of *JAK2* V617F-mutant and wild-type CD34 + progenitors showed no difference in STAT5 target expression and no allele-dose effect, whereas erythroid markers were elevated in mutant cells (Figure S4, Supplementary Results S4). Concurrent single-cell work supports this resolution: a recent preprint identified the basophilic erythroblast as the stage at which *JAK2* V617F shifts from transcriptionally silent to trajectory-altering, with erythroid-lineage divergence as the dominant readout [33], and a multi-compartment study likewise reported compositionally sensitive inflammatory/interferon enrichment rather than a canonical STAT5 target-gene response [34], aligning with the interferon/HLA program that single-cell genotyping attributes to the V617F clone [19]. Together, these single-cell and composition-aware studies corroborate our central dissociation claim from the cell-intrinsic direction: the *JAK2* V617F mutation does not engage the STAT program in the progenitor compartment, while erythroid-lineage and interferon/inflammatory programs remain dominant.

Our deconvolution analysis substantially qualifies the *EPAS1* finding. Because *EPAS1* is itself an erythroid gene, its bulk upregulation is substantially confounded by the expanded erythroid compartment, and we could not resolve cell-intrinsic *EPAS1* upregulation from bulk or CD34 + data; we therefore interpret *EPAS1* as an erythroid-lineage marker of PV rather than as evidence for a cell-intrinsic *HIF2A*-driven mechanism. Consistent with this, *EPAS1* did not differ between *JAK2* V617F-positive and V617F-negative MPN samples in the CD34 + cohort (d = + 0.04, p = 0.35), and the *HIF2A* signature's discriminative capacity overlapped with that of the erythroid cell fraction (Supplementary Results S2). This carries a broader methodological caution: bulk PV studies that did not adjust for cell composition may have overestimated the cell-intrinsic significance of *EPAS1*/*HIF2A* and other erythroid-lineage signals: early CD34 + profiling attributed part of the PV signature to *JAK2* V617F-dependent and -independent programs without resolving the contribution of the expanded erythroid compartment [8], and reports of elevated HIF transcriptional activity in V617F-positive PV [18] do not distinguish an intrinsic HIF program from an expanded erythroid fraction. Our composition-adjusted analysis suggests that, at least for *EPAS1*, a large share of the bulk signal is compositional, so the literature's cell-intrinsic interpretation may need revisiting. A cell-intrinsic role for *HIF2A* in PV remains plausible [18], but establishing it would require single-cell or functional experiments that lie beyond this study.

An siRNA analysis further cautioned against over-interpreting the downstream targets: *CA9*, *SLC2A1*, and *ADM* were predominantly *HIF1A*-dependent in a glioma cell line, indicating that *HIF1A*/*HIF2A* target partitioning depends on cell type (Supplementary Results S1). We therefore emphasize *EPAS1* itself as the core of the signature.

This study has several limitations. First, three of the five patient datasets originate from overlapping investigators, so only GSE47018 and GSE103237 are fully independent; a sensitivity analysis restricted to these two datasets confirmed the central contrast (3.6). Second, PMF samples are few (n = 9/3/7) and the independent CD34 + cohorts have small control groups (n = 7 and 15), so subtype- and CD34+-specific conclusions are underpowered. Third, the siRNA validation is from a single non-erythroid glioma cell line (n = 3 per condition), which may not transfer to PV erythroid progenitors. Fourth, all bulk analyses rely on 2010–2017 microarray platforms with limited dynamic range, and are transcriptomic and correlative without functional or protein-level validation; we cannot exclude STAT protein phosphorylation, transient signaling, or STAT activation in minor subpopulations. Fifth, the STAT5 direct-target meta-analysis showed substantial heterogeneity (I^2 = 89%), so the pooled effect should be read as "no consistent upregulation" rather than a precise null, and the allele-burden analysis is limited to 20 PV patients with uniformly high burden (≥ 55%). Finally, our marker-gene deconvolution produces relative rather than absolute cell-fraction estimates, so the residualization is a qualitative correction; we cannot quantify the proportion of the bulk *EPAS1* signal attributable to composition, and the claim that past studies "overestimated" *EPAS1*'s significance is an inference from our framework rather than a formal re-analysis. Two safeguards argue that this qualitative correction does not simply erase genuine biological signal. First, the same residualization procedure preserved a cell-intrinsic mitochondrial oxidative-phosphorylation signal (Figure S5), showing that the adjustment does not indiscriminately attenuate true differences. Second, an erythroid-only sensitivity analysis, in which *EPAS1* was adjusted for the erythroid compartment alone, without the myeloid or lymphoid components, already reduced the *EPAS1* effect from d = + 0.093 (p = 0.0002) to d = + 0.050 (p = 0.053), indicating that the erythroid cell fraction alone accounts for most of the bulk signal and that the attenuation is not an artifact of over-correction by a multi-component model (Figure S7).

In conclusion, we show that PV is transcriptionally defined by erythroid expansion in the absence of STAT transcriptional activation, a pattern that replicates across datasets, cell types, and platforms and is confirmed by multiple independent STAT gene sets. At the transcriptomic level, the *JAK2* V617F mutation is dissociated from STAT-driven transcription, which challenges the canonical model of PV pathogenesis without excluding STAT protein phosphorylation or transient signaling; *EPAS1* upregulation is largely attenuated after composition adjustment, consistent with an erythroid cell-fraction effect. These findings motivate functional and single-cell studies to resolve the cell-intrinsic consequences of *JAK2* V617F in the erythroid lineage.

Declarations

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This research received no external funding.

Author Contribution

Yu Wang and Yu Wang contributed equally to this work. H.A. and T.W. conceived and supervised the study. Yu Wang, Yu Wang, Yukun Wang and W.F. performed the bioinformatic analyses and interpreted the data. All authors read and approved the final manuscript.

Data Availability

All publicly available transcriptomic datasets analysed in this study were obtained from the Gene Expression Omnibus (GEO) repository under the following accession numbers: GSE26049, GSE61629, GSE57793, GSE47018, GSE103237, GSE27523 and GSE122198.Access to these datasets is available via the NCBI GEO website (https://www.ncbi.nlm.nih.gov/geo/). Custom R analysis scripts used for re-analysis are available from the corresponding author upon reasonable request. No new raw sequencing data were generated in this work.

References

1. James C, Ugo V, Le Couédic JP, Staerk J, Delhommeau F, Lacout C et al (2005) A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature 434(7037):1144–1148. 10.1038/nature03546
2. Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ et al (2005) Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell 7(4):387–397. 10.1016/j.ccr.2005.03.023
3. Kralovics R, Passamonti F, Buser AS, Teo S, Tiedt R, Passweg JR et al (2005) A gain-of-function mutation of JAK2 in myeloproliferative disorders. N Engl J Med 352(17):1779–1790. 10.1056/NEJMoa051113
4. Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S et al (2005) Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet 365(9464):1054–1061. 10.1016/S0140-6736(05)71142-9
5. Levine RL, Gilliland DG (2008) Myeloproliferative disorders. Blood 112(6):2190–2198. 10.1182/blood-2008-03-077966
6. Vainchenker W, Delhommeau F, Constantinescu SN, Bernard OA (2011) New mutations and pathogenesis of myeloproliferative neoplasms. Blood 118(7):1723–1735. 10.1182/blood-2011-02-292102

7. Quintás-Cardama A, Verstovsek S (2013) Molecular pathways: JAK/STAT pathway: mutations, inhibitors, and resistance. Clin Cancer Res 19(8):1933–1940. 10.1158/1078-0432.CCR-12-0284
8. Berkofsky-Fessler W, Buzzai M, Kim MK, Fruchtman S, Najfeld V, Min DJ et al (2010) Transcriptional profiling of polycythemia vera identifies gene expression patterns both dependent and independent from the action of JAK2V617F. Clin Cancer Res 16(17):4339–4352. 10.1158/1078-0432.CCR-10-1092
9. Skoda RC, Duek A, Grisouard J (2015) Pathogenesis of myeloproliferative neoplasms. Exp Hematol 43(8):599–608. 10.1016/j.exphem.2015.06.007
10. Rampal R, Al-Shahrour F, Abdel-Wahab O, Patel JP, Brunel JP, Mermel CH et al (2014) Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. Blood 123(22):e123–e133. 10.1182/blood-2014-02-554634
11. Zini R, Guglielmelli P, Pietra D, Rumi E, Rossi C, Rontauroli S et al (2017) CALR mutational status identifies different disease subtypes of essential thrombocythemia showing distinct expression profiles. Blood Cancer J 7(12):638. 10.1038/s41408-017-0010-2
12. Rankin EB, Biju MP, Liu Q, Unger TL, Rha J, Johnson RS et al (2007) Hypoxia-inducible factor-2 (HIF-2) regulates hepatic erythropoietin in vivo. J Clin Invest 117(4):1068–1077. 10.1172/JCI30117
13. Gruber M, Hu CJ, Johnson RS, Brown EJ, Keith B, Simon MC (2007) Acute postnatal ablation of Hif-2alpha results in anemia. Proc Natl Acad Sci U S A 104(7):2301–2306. 10.1073/pnas.0608382104
14. Kaelin WG Jr, Ratcliffe PJ (2008) Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Mol Cell 30(4):393–402. 10.1016/j.molcel.2008.04.009
15. Percy MJ, Furlow PW, Lucas GS, Li X, Lappin TR, McMullin MF et al (2008) A gain-of-function mutation in the HIF2A gene in familial erythrocytosis. N Engl J Med 358(2):162–168. 10.1056/NEJMoa073123
16. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102(43):15545–15550. 10.1073/pnas.0506580102
17. Semenza GL (2012) Hypoxia-inducible factors in physiology and medicine. Cell 148(3):399–408. 10.1016/j.cell.2012.01.021
18. Song J, Kim SJ, Gordeuk VR, Prchal JT (2019) Hypoxia inducible factors' transcriptional activity in polycythemia vera. Blood 134(Suppl 1):4211. 10.1182/blood-2019-131909
19. Van Egeren D, Kamaz B, Liu S, Nguyen M, Reilly CR, Kalyva M et al (2022) Transcriptional differences between JAK2-V617F and wild-type bone marrow cells in patients with myeloproliferative neoplasms. Exp Hematol 107:14–19. 10.1016/j.exphem.2021.12.364
20. Gillinder KR, Tuckey H, Bell CC, Magor GW, Huang S, Ilsley MD et al (2017) Direct targets of pSTAT5 signalling in erythropoiesis. PLoS ONE 12(7):e0180922. 10.1371/journal.pone.0180922
21. Rodriguez-Meira A, Buck G, Clark SA, Povinelli BJ, Alcolea V, Louka E et al (2019) Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. Mol Cell 73(6):1292–1305e8. 10.1016/j.molcel.2019.01.009

22. Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC et al (2013) Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. N Engl J Med 369(25):2391–2405. 10.1056/NEJMoa1312542
23. Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD et al (2013) Somatic mutations of calreticulin in myeloproliferative neoplasms. N Engl J Med 369(25):2379–2390. 10.1056/NEJMoa1311347
24. Mullally A, Lane SW, Ball B, Megerdichian C, Okabe R, Al-Shahrour F et al (2010) Physiological Jak2V617F expression causes a lethal myeloproliferative neoplasm with differential effects on hematopoietic stem and progenitor cells. Cancer Cell 17(6):584–596. 10.1016/j.ccr.2010.05.015
25. Tefferi A, Vainchenker W (2011) Myeloproliferative neoplasms: molecular pathophysiology, essential clinical understanding, and treatment strategies. J Clin Oncol 29(5):573–582. 10.1200/JCO.2010.29.8711
26. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM et al (2016) The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 127(20):2391–2405. 10.1182/blood-2016-03-643544
27. Socolovsky M, Nam H, Fleming MD, Haase VH, Brugnara C, Lodish HF (2001) Ineffective erythropoiesis in Stat5a(-/-)5b(-/-) mice due to decreased survival of early erythroblasts. Blood 98(12):3261–3273. 10.1182/blood.V98.12.3261
28. Haase VH (2013) Regulation of erythropoiesis by hypoxia-inducible factors. Blood Rev 27(1):41–53. 10.1016/j.blre.2012.12.003
29. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y et al (2015) Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12(5):453–457. 10.1038/nmeth.3337
30. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F et al (2019) Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 37(7):773–782. 10.1038/s41587-019-0114-2
31. Aran D, Hu Z, Butte AJ (2017) xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol 18(1):220. 10.1186/s13059-017-1349-1
32. Borenstein M, Hedges LV, Higgins JP, Rothstein HR (2010) A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 1(2):97–111. 10.1002/jrsm.12
33. Konak D, Pir P (2026) Basophilic erythroblast emerges as the key turning point in polycythemia vera. bioRxiv [Preprint]. 10.64898/2026.09.01.748468
34. Alruwetei AM (2026) Multi-compartment transcriptomics identifies a persistent inflammatory program and a network-derived diagnostic signature in polycythemia vera. Int J Mol Sci 27(10):4580. 10.3390/ijms27104580