Germline duplication of ATG2B and GSKIP predisposes to familial myeloid malignancies

J. Saliba, C. Saint-Martin, A. Di Stefano, G. Lenglet, C. Marty, B. Keren, F. Pasquier, V. D. Valle, L. Secardin, G. Leroy, et al.

Germline duplication of *ATG2B* and *GSKIP* predisposes to familial myeloid malignancies

Joseph Saliba[1–4,14], Cécile Saint-Martin[1–3,5,14], Antonio Di Stefano[1–4,14], Gaëlle Lenglet[1–4], Caroline Marty[1–4], Boris Keren[5], Florence Pasquier[1–4], Véronique Della Valle[1–3], Lise Secardin[1–4], Gwendoline Leroy[5], Emna Mahfoudhi[1–4,6], Sarah Grosjean[1–4], Nathalie Droin[1–3], M'boyba Diop[1–3], Philippe Dessen[1–3], Sabine Charrier[7], Alberta Palazzo[1–3], Jane Merlevede[1–3], Jean-Côme Meniane[8], Christine Delaunay-Darivon[8], Pascal Fuseau[9], Françoise Isnard[10,11], Nicole Casadevall[1,12], Eric Solary[1–3], Najet Debili[1–3], Olivier A Bernard[1–3], Hana Raslova[1–3], Albert Najman[10,11], William Vainchenker[1–4,13,15], Christine Bellanné-Chantelot[1–3,5,11,15] & Isabelle Plo[1–4,15]

No major predisposition gene for familial myeloproliferative neoplasms (MPN) has been identified. Here we demonstrate that the autosomal dominant transmission of a 700-kb duplication in four genetically related families predisposes to myeloid malignancies, including MPN, frequently progressing to leukemia. Using induced pluripotent stem cells and primary cells, we demonstrate that overexpression of ATG2B and GSKIP enhances hematopoietic progenitor differentiation, including of megakaryocytes, by increasing progenitor sensitivity to thrombopoietin (TPO). ATG2B and GSKIP cooperate with acquired *JAK2*, *MPL* and *CALR* mutations during MPN development. Thus, the germline duplication may change the fitness of cells harboring signaling pathway mutations and increases the probability of disease development.

Most myeloid malignancies, including acute myeloid leukemias (AML), myelodysplastic syndromes (MDS) and MPN, are sporadic diseases. Familial forms are rare but informative, as germline mutations may phenocopy initiating mutations in sporadic leukemias. Such germline mutations have been identified in *RUNX1* (in a familial platelet disorder that predisposes to AML, or FPD/AML)[1], *CEBPA*[2] and *GATA2* (refs. 3,4) coding sequences as well as in the 5′UTR of the *ANKRD26* (*THC2*) gene[5].

Familial cases of MPN are usually transmitted by autosomal inheritance with incomplete penetrance, appear in adulthood and exhibit acquired genetic abnormalities that are similar to those identified in sporadic cases, such as mutations of *JAK2* (encoding p.Val617Phe) and *TET2* (refs. 6,7). The independent acquisition of several oncogenic hits in the same patient with MPN suggested that an unidentified germline predisposition locus might be present in some sporadic cases[6–8]. For example, the *JAK2* 46/1 haplotype was shown to increase the risk of developing MPN with *JAK2* mutation encoding p.Val617Phe[9–11], whereas a germline intronic SNP in the *TERT* gene is another susceptibility factor for MPN development[12,13]. Linkage and segregation analyses indicate that genetic predisposition to MPN cannot be related to a common alteration and rather involves a number of susceptibility loci responsible for independent familial aggregations. Identification of these susceptibility loci may improve understanding of the mechanisms of predisposition, which might either result in the induction of genetic instability, favoring the acquisition of oncogenic mutations, or correspond to a fertile ground for the selection of somatic mutations.

Here we describe a newly identified germline copy number variation (CNV) that predisposes to several myeloid malignancies, particularly essential thrombocythemia, which can progress to myelofibrosis or leukemia.

RESULTS

Identification of a germline CNV on chromosome 14

We identified two large families (F1 and F2) from the French West Indies with affected individuals having clinical features distinct from those of other familial cases of MPN (**Fig. 1**). Adult-onset hematological malignancies, including MPN, chronic myelomonocytic leukemia (CMML) and AML, were transmitted in these families with autosomal dominant inheritance (**Table 1** and **Supplementary Table 1**).

[1]INSERM, Unité Mixte de Recherche (UMR) 1170, Villejuif, France. [2]Université Paris XI, UMR 1170, Gustave Roussy, Villejuif, France. [3]Gustave Roussy, Villejuif, France. [4]Laboratory of Excellence GR-Ex, Villejuif, France. [5]Assistance Publique–Hôpitaux de Paris, Département de Génétique, Hôpital Pitié-Salpêtrière, Paris, France. [6]Laboratoire d'Hématologie Moléculaire et Cellulaire (Salem Abbes), Institut Pasteur de Tunis, Belvédère, Tunisia. [7]Généthon, Evry, France. [8]Service de Médecine Interne, Centre Hospitalier Universitaire (CHU) de Fort de France, Fort de France, France. [9]Laboratoire d'Hématologie, CHU Fort de France, Fort de France, France. [10]Assistance Publique–Hôpitaux de Paris, Département d'Hématologie Clinique et de Thérapie Cellulaire, Hôpital Saint-Antoine, Paris, France. [11]Sorbonne Universités, Université Pierre et Marie Curie (UPMC), Groupe de Recherche Clinique sur les Myéloprolifératives Aiguës et Chroniques (GRC 7, MYPAC), Paris, France. [12]Assistance Publique–Hôpitaux de Paris, Laboratoire d'Hématologie, Hôpital Saint-Antoine, Paris, France. [13]Assistance Publique–Hôpitaux de Paris, Consultation d'Immuno-Hématologie, Hôpital Saint-Louis, Paris, France. [14]These authors contributed equally to this work. [15]These authors jointly supervised this work. Correspondence should be addressed to W.V. (verpre@igr.fr), C.B.-C. (christine.bellanne-chantelot@aphp.fr) or I.P. (isabelle.plo@gustaveroussy.fr).

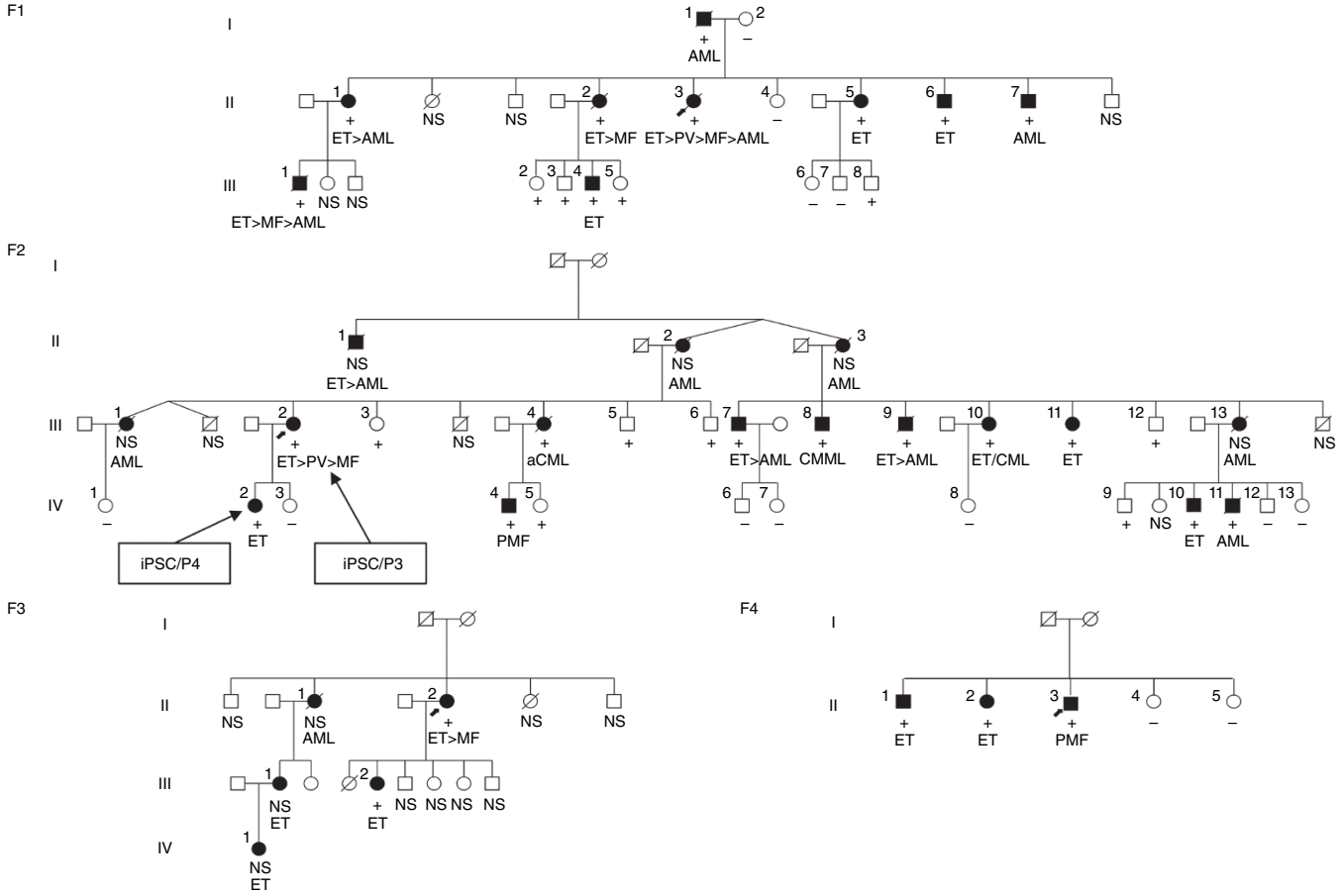


Figure 1 Pedigrees of the four families with MPN. Filled symbols represent cases. Under each symbol, the first line shows the genetic status for the germline CNV (+, the CNV was found; –, otherwise); the second line indicates the phenotype at the time of diagnosis followed by the disease evolution. ET, essential thrombocythemia; PV, polycythemia vera; MF, myelofibrosis; PMF, primary myelofibrosis; AML, acute myeloid leukemia; aCML, atypical chronic myeloid leukemia; CMML, chronic myelomonocytic leukemia; NS, not studied.

Moreover, two-thirds of the patients (22/33) initially had essential thrombocythemia, with half of them progressing to myelofibrosis or AML. Genetic linkage analysis of the two families identified positive linkage at the 14q32.13-q32.2 locus (logarithm of odds (LOD) score z_{max} = 3.7) (**Fig. 2a**). Using additional microsatellite markers and taking advantage of a crossing-over event in one patient (F1:II-7) and an ancestral patient from family F2, the susceptibility region was further narrowed down to a 1.86-Mb interval (95.76–97.62 Mb, hg19) (**Supplementary Fig. 1a**). Complete sequencing of the 1.86-Mb candidate locus did not identify any germline mutation in coding regions that segregated with the disease (data not shown). SNP array analysis detected a 700-kb duplication located within the linkage region and absent from the Database of Genomic Variants (DGVS, v10) (**Fig. 2b**). We further confirmed the presence of the CNV in all affected cases in both families by quantitative RT-PCR (qRT-PCR; **Fig. 1** and **Supplementary Table 2**). PCR analyses mapped the proximal and distal breakpoints of the CNV and established this duplication as a 700-kb head-to-tail tandem duplication (**Fig. 2c**,**d** and **Supplementary Table 2**). This region includes the genes *TCL1A*, *GSKIP*, *ATG2B*, *BDKRB1* and *BDKRB2*, together with the first exon of *AK7* (**Fig. 2e**). Using quantitative PCR, we identified in two other families the same CNV with identical breakpoints (families F3 and F4; **Fig. 1** and **Supplementary Fig. 1b**). These families shared with the two initial families their geographical origin and the clinical features of affected individuals. Analysis of the 4 families demonstrated the high penetrance of the phenotype, as 24 of 34 carriers of the germline CNV developed disease (**Fig. 1**). Asymptomatic carriers of the CNV were matched by age to patients with MPN and presented no major blood abnormalities (**Supplementary Table 3**). This CNV was not identified in 199 control DNA samples from individuals with the same geographical origin and 98 unrelated Eurocaucasian familial MPN cases.

Molecular characterization of the familial MPN cases

Analysis of the genetic abnormalities associated with disease development collected from 24 members of these families identified a signaling mutation profile in MPN (essential thrombocythemia or primary myelofibrosis (PMF)) similar to that observed in sporadic cases: cases had *JAK2* mutation encoding p.Val617Phe (15/22, 68%), *MPL* mutation (2/22, 9%) or *CALR* mutation (4/22, 18%) or were triple negative for these mutations (1/22, 5%) (**Fig. 3** and **Table 1**). Sequencing performed on the most recent samples showed the acquisition of secondary events in *TET2* (7/21, 38%), *IDH1* (2/21, 10%), *IDH2* (4/21, 19%) and *ASXL1* (1/21, 5%) with disease evolution to myelofibrosis and leukemia (**Fig. 3** and **Table 1**). Several patients exhibited biallelic mutation of *TET2* or a combination of epigenetic mutations (in *TET2* and *IDH2* or *IDH1* and *IDH2*; *n* = 3/21, 14%). No *TP53* mutation was detected. Complex karyotypes were observed only when the leukemia occurred.

Table 1 Clinical and molecular data for the four families with MPN

Family	ID	Sex	Initial diagnosis	Age at diagnosis (years)	Follow-up duration (years)	Evolution	Outcome	Signaling mutation (allele burden)	Epigenetic mutation	Karyotype
F1	I-1	Male	AML4	75.4	0.5		Died	NS	NS	46,XY,del(13)(q12q22)[1]/46,XY,del(13)(q12q22),t(1;1)(p34,3p36,3),t(2;20)(p13;q11,2),t(4;12)(q12;q15),add(5)(q35)[16]
F1	II-1	Female	ET	63.5	7.5	AML	Alive	*JAK2*: p.Val617Phe (45%)	*IDH1*: c.394C>T, p.Arg132Cys	46,XX,–7[10]/47,–7,+11/46,XX
F1	II-2	Female	ET	39.5	20.2	MF	Died	*JAK2*: p.Val617Phe (84%)	*IDH2*: c.419G>A, p.Arg140Gln	46,XX,der(6)t(6;?)(p25)[7]/46,XX,dup(12)(q13q22)[2]/46,XX[8]
F1	II-3	Female	ET	38	17	PV> MF>AML	Died	*JAK2*: p.Val617Phe (48%)	*IDH2*: c.563G>A, p.Arg188Gln *TET2*: c.1648C>T, p.Arg550*/ c.2567delG, p.Gly856fs	46,XX,+1,der(1;7)(q10;p10) [11]/46,XX[5]
F1	II-5	Female	ET	41.2	21.8		Alive	*JAK2*: p.Val617Phe (62%)	*IDH1*: c.394C>A, p.Arg132Ser *IDH2*: c.563G>A, p.Arg188Gln	46,XX
F1	II-6	Male	ET	36.8	17		Alive	*JAK2*: p.Val617Phe (2%)	*IDH2*: c.563G>A, p.Arg188Gln *TET2*: c.3321dupA, p.Ser1107fs	46,XY[20]
F1	II-7	Male	AML2	43.6	7.9		HSCT, alive	Triple negative	*IDH1*: c.394C>T, p.Arg132Cys	46,XY,–7,+14[12]/46,XY[11]
F1	III-1	Male	ET	34	8	AML	Died	*CALR*: c.1099_1150del52 (30%)	*TET2*: c.2058A>T, p.Arg686Ser	45,XY,–5,der(6)t(6;8)(p22;q21), –17+mar(16)/46,XY[1]
F1	III-4	Male	ET	35.5	8.5		Alive	*JAK2*: p.Val617Phe (2%)	None	NS
F2	II-1	Male	ET	81	0.5	AML	Died	NS	NS	NS
F2	II-2	Female	AML	52	1		Died	NS	NS	NS
F2	II-3	Female	AML	52	ns		Died	NS	NS	NS
F2	III-1	Female	AML2	37.5	0.5		Died	NS	NS	NS
F2	III-2	Female	ET	49	11	PV>MF	Alive	*JAK2*: p.Val617Phe (49%)	*TET2*: c.3500+3A>C	46,XY,inv(2)(p24q14)[11]
F2	III-4	Female	aCML	48	0.9	AML	Died	Triple negative	*TET2*: c.1954delC, p.Gln652fs/ c.2490dupA, p.Gln831fs *ASXL1*: c.2893C>T, p.Arg965*	46,XX,–7[20]
F2	III-7	Male	ET	36	22	AML	HSCT, alive	*CALR*: c.1099_1150del52 (5%)	None	46,XY,inv(2)(p24q11)c[23]
F2	III-8	Male	CMML	45.5	0.5		Alive	*JAK2*: p.Val617Phe (38%)	*TET2*: c.4469delA, p.Glu1490fs	46,XY,inv(2) (p25q13)[25]
F2	III-9	Male	ET	41	9	AML	HSCT, died	*MPL*: p.Trp515Leu (5%)	*TET2*: c.5551G>T, p.Glu1851*	46,XY,inv(2)(p2?5q1?4)[20]
F2	III-10	Female	ET	34	16	CML, MF	Alive	*CALR*: c.1099_1150del52 (35%); m-Bcr (e1/a2)	None	46,XX,t(9,22)(q34,q11)[17]
F2	III-11	Female	ET	41.5	6.5		Alive	*MPL*: p.Trp515Leu (3%)	*TET2*: c.3804–2A>T	NS
F2	III-13	Female	AML2	35	0.8		HSCT, died	NS	NS	NS
F2	IV-2	Female	ET	33	5		Alive	*JAK2*: p.Val617Phe (1%)	None	46,XY,inv(2) (p24q14)[11]
F2	IV-4	Male	PMF[i]	36	0.5		Alive	*JAK2*: p.Val617Phe (72%)	*ASXL1*: c.1934dupA, p.Gly646fs	46,XY
F2	IV-10	Male	ET	42	0.2		Alive	*JAK2*: p.Val617Phe (14%)	NS	NS
F2	IV-11	Male	AML6	34	1.6		Died	NS	NS	45,XY,–7[18]
F3	II-1	Female	AML5	64.8	0		Died	NS	NS	48,XX,+4,+8[15]
F3	II-2	Female	ET	46.3	22	MF	Alive	*JAK2*: p.Val617Phe (5%)	*TET2*: c.3689delT, p.Ile1230fs	46,XX[15]
F3	III-1	Female	ET	42	8		Alive	*JAK2*: p.Val617Phe (1.4%)	None	NS
F3	III-2	Female	ET	34	11		Alive	Triple negative	None	NS
F3	IV-1	Female	ET	29	2		Alive	*JAK2*: p.Val617Phe (11%)	NS	NS
F4	II-1	Male	ET	43.8	0.6		Alive	*CALR*: c.1099_1150del52 (10%)	None	46,XY[20]
F4	II-2	Female	ET	42	0.2		Alive	*JAK2*: p.Val617Phe (0.6%)	None	NS
F4	II-3	Male	PMF	38.5	0.8		HSCT, alive	*JAK2*: p.Val617Phe (61%)	None	46,XY[20]

AML, acute myeloid leukemia (subtype according to French-American-British (FAB) classification); ET, essential thrombocythemia; PV, polycythemia vera; aCML, atypical chronic myeloid leukemia (*BCR-ABL* negative); CMML, chronic myelomonocytic leukemia; PMF, primary myelofibrosis; HSCT, hematopoietic stem cell transplantation; m-Bcr (e1/a2), major breakpoint cluster region rearrangement between exons 1 and 2; NS, not studied. Triple-negative individuals lacked the mutation in *JAK2* encoding p.Val617Phe (c.1849G>T), the mutation in *MPL* encoding p.Trp515Leu (c.1544G>T) and the 52-bp deletion (p.Leu367Thrfs*46) or 5-bp insertion (p.Lys385Asnfs*47) in *CALR*. ID is according to the pedigrees shown in **Figure 1**.

Expression of the genes contained in the duplication

Using gene expression arrays, we found that three of the six duplicated genes were expressed in Epstein-Barr virus (EBV)-transformed cell lines (EBVCs) from three donors, namely *TCL1A*, *ATG2B* and *GSKIP*, after normalization to the levels of a housekeeping gene (*PPIA* or *B2M*). However, expression of only *ATG2B* and *GSKIP* was detected on microarrays of $CD34^+$ purified hematopoietic progenitors from donors and $CD36^+$ erythroblasts or $CD41^+$ megakaryocytes derived from $CD34^+$ progenitors cultured *in vitro*. Moreover, *ATG2B* and *GSKIP* expression levels were similar to those of *RUNX1* and *STAT5A*, two genes well expressed during normal hematopoiesis (**Fig. 4a**). Microarrays showed significantly higher levels of *ATG2B* and *GSKIP* transcripts in EBVCs derived from five patients in comparison to EBVCs derived from three control relatives but no difference for *TCL1A* levels (**Fig. 4b**). We further analyzed the expression of these two genes during normal *in vitro* erythroid and megakaryocyte differentiation from $CD34^+$ progenitors. Different cellular fractions were sorted by flow cytometry, and the expression levels of *ATG2B* and *GSKIP* were quantified by qRT-PCR (**Fig. 4c**,**d**). *GSKIP* expression did not vary, whereas *ATG2B* expression decreased during megakaryocyte and erythroid differentiation. A strong increase in *ATG2B* expression was observed in terminal erythroid differentiation. We confirmed by qRT-PCR the overexpression (by two- to threefold) of *ATG2B* and *GSKIP* in megakaryocytes derived from patients as compared to controls (**Fig. 4e**).

Derivation of human induced pluripotent stem cells

To analyze the consequences of the duplication conferring predisposition to MPN, we generated induced pluripotent stem cells (iPSCs) from $CD34^+$ progenitor cells sorted from two patients; the patients were selected from the same family to reduce genetic heterogeneity (F2:III-2 (iPSC/P3) and F2:IV-2 (iPSC/P4))[14]. Patient F2:III-2, who developed essential thrombocythemia that progressed to myelofibrosis, had a *JAK2* mutation encoding p.Val617Phe (allele burden in granulocytes, 50%) and a heterozygous *TET2* mutation (c.3500+3A>C) associated with decreased 5-hydroxymethylcytosine (5hmC) levels (**Supplementary Fig. 2**). Patient F2:IV-2, who developed essential thrombocythemia with a platelet count slightly above the normal value ($450–550 \times 10^9$ cells/l), had a *JAK2* mutation encoding p.Val617Phe with very low allele burden (<1% in granulocytes). We obtained clones bearing only the CNV predisposition locus from patient F2:IV-2 (P4-CNV) and clones harboring the CNV together with one wild-type *JAK2* allele and one allele encoding the p.Val617Phe alteration, with (P3-CNV-VF-TET2) or without (P3-CNV-VF) *TET2* mutation, from patient F2:III-2 (**Supplementary Fig. 2**). We used as controls previously published iPSC clones obtained by reprogramming $CD34^+$ cells from healthy donors or sporadic MPN cases heterozygous for the *JAK2* mutation encoding p.Val617Phe (P2-VF)[15]. All these clones formed embryonic stem cell (ES)-like colonies. We selected two clones of each genotype (a or b), performed genomic characterization using comparative genomic hybridization (CGH) arrays, cytogenetics and whole-exome sequencing (**Supplementary Fig. 3a,b**), and validated the phenotype while checking for the silencing of transgenes, the reexpression of endogenous pluripotent transcription factors, and the ability to generate embryoid bodies *in vitro* and to form teratomas *in vivo* (**Supplementary Fig. 4a–g**).

CNV increases the generation of iPSC-derived hematopoietic cells

To explore the hematopoietic differentiation of iPSC clones, sac-like structures were dissociated at day 12 and cultured on the OP9 stromal cell line in the presence of cytokines[16]. First, when hematopoietic progenitor colonies were enumerated 10–12 d after seeding day 13 unfractionated cells in semisolid medium, we observed a tenfold higher number of colonies for samples carrying the CNV predisposition locus alone, regardless of an additional mutation in *JAK2* or *TET2*, in comparison to controls (**Fig. 5a**). Second, we performed kinetic analyses from day 10 until day 21, and the percentages of megakaryocytes ($CD41^+$), erythroblasts (GPA^+) and monocytes ($CD14^+$) were determined as previously described[15,17,18]. P2-VF iPSC clones (heterozygous for the *JAK2* mutation encoding p.Val617Phe) did not show marked differences in the expression of hematopoietic markers in comparison to control iPSC clones, whereas all the iPSC clones carrying the CNV predisposition locus had a significantly higher proportion of megakaryocytes as differentiation progressed (**Fig. 5b**). Increases in the proportions of differentiated cells for clones carrying the CNV alone were further enhanced by the *JAK2* mutation in P3-CNV-VF clones and were even more remarkable, especially for erythroblasts, in P3-CNV-VF-TET2 clones mutated for both *JAK2* and *TET2*. Finally, when iPSCs were sorted for the TRA-1-81 pluripotent surface marker and grown on OP9 cells with a cocktail of cytokines, we observed significantly higher absolute numbers of $CD41^+$, $CD14^+$ and GPA^+ cells generated by P4-CNV iPSCs as compared to control iPSCs at day 18 (**Fig. 5c**).

These results indicate that the CNV predisposition locus could be sufficient to increase the generation of hematopoietic progenitor cells and the overproduction of erythroblasts, megakaryocytes and monocytes, an effect that is reinforced by additional mutations in *JAK2* and/or *TET2*.

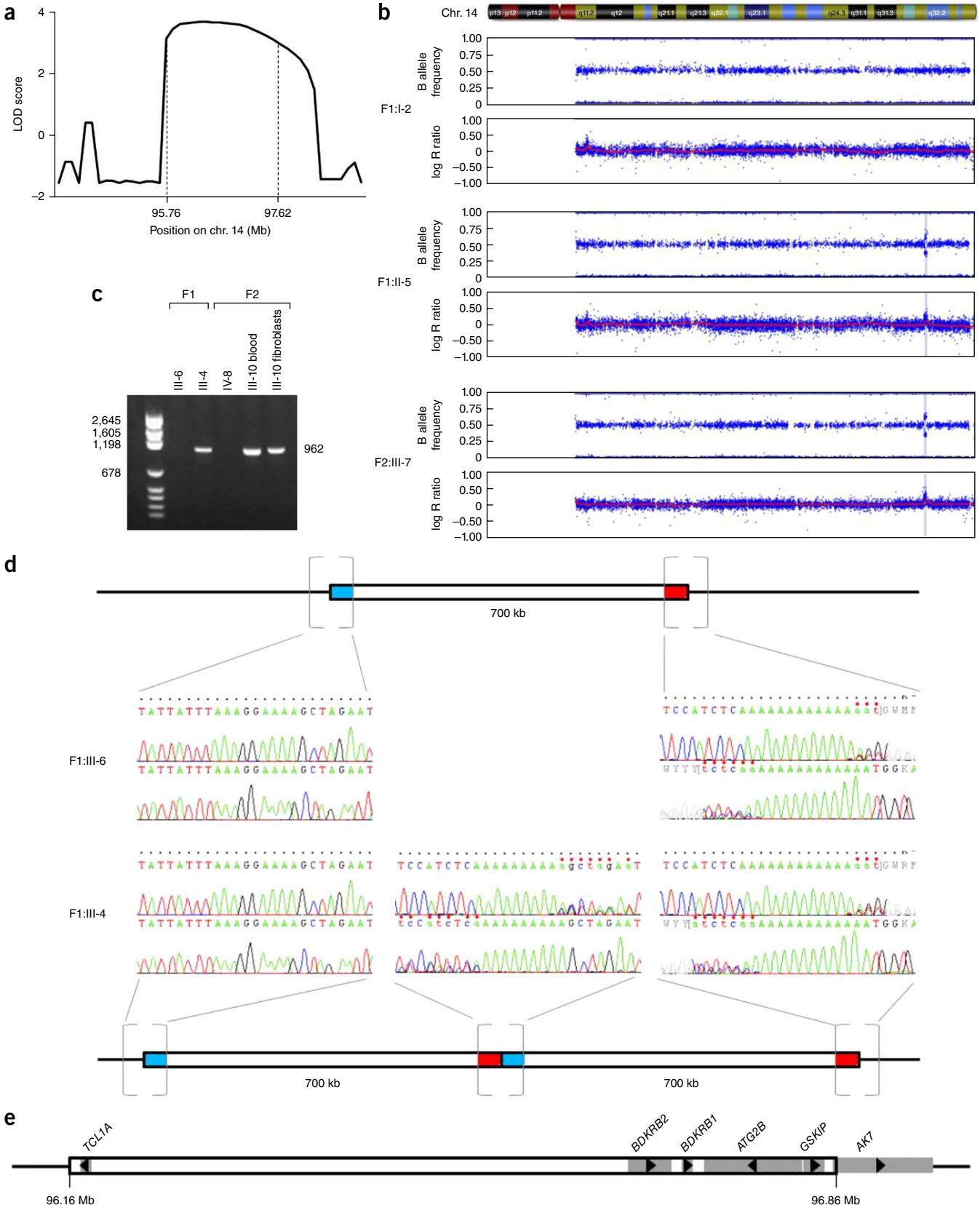
CNV and the response of progenitors to cytokines

Because hypersensitivity of hematopoietic progenitors to cytokines is a characteristic feature of MPN[19], we cultured iPSC clones on OP9 stromal cells for 12 d in the presence of vascular endothelial growth factor (VEGF) and hematopoietic cytokines and then sorted for the hematopoietic progenitor cell fraction (GPA^+CD41^+). These cells were grown for 12 d in methylcellulose in the presence of stem cell factor (SCF) and increasing concentrations of erythropoietin (EPO). The response to EPO was equivalent for control, P2-VF and P4-CNV progenitor cells. In contrast, P3-CNV-VF and P3-CNV-VF-TET2 progenitor cells demonstrated increased sensitivity to EPO, with around 30% of the colonies formed constituting endogenous erythroid colonies (EECs) (**Supplementary Fig. 5a**). Moreover, the

Figure 2 Characterization of the germline CNV. (**a**) Positive linkage signal (*z* score = 3.7) at 14q32. The dashed lines highlight the linkage signal in the two initial families (F1 and F2). (**b**) SNP array analysis showing a 700-kb duplication in both families F1 and F2. A chromosome 14 ideogram is shown at the top. For each individual, the lower plot shows the log R ratio (*y* axis) for 9,592 probes on chromosome 14 (*x* axis) and the upper plot shows the B allele frequency (*y* axis) for each probe position (*x* axis). The array profiles for F1:II-5 and F2:III-7 show 3 copies of the genomic region chr. 14: 95.76–97.62 Mb, with an increased log R ratio of ~0.4 and B allele frequencies with values at 0 for the AAA genotype, 0.33 for the AAB genotype, 0.67 for the ABB genotype and 1 for the BBB genotype. The array profile for F1:I-2 (a control) shows only 2 copies for this segment with a normal log R ratio and B allele frequencies with values at 0 for the AA genotype, 0.5 for the AB genotype and 1 for the BB genotype. (**c**) Gel electrophoresis showing the amplification of a 962-bp junction fragment in patients (F1:III-4 and F2:III-10) and its absence in non-carrier controls (F1:III-6 and F2:IV-8). Analysis of DNA extracted from blood and fibroblasts (F2:III-10) showed the germline status of the identified CNV. The primers used for PCR were Chr14_B2-C2.1F and Chr14_B1-C2.3R (**Supplementary Table 2**). (**d**) Mapping and sequencing of the proximal and distal breakpoints established the duplication as a 700-kb head-to-tail tandem duplication. (**e**) Schematic structural organization of the duplicated region including six genes. Arrowheads indicate the direction of transcription.

percentage of large primitive erythroblast colonies was higher for P3-CNV-VF-TET2 progenitor cells as compared to P3-CNV-VF and control progenitor cells (**Supplementary Fig. 5b**,**c**).

JAK2 is a key molecule in the cytokine receptor signaling cascade, and JAK2 Val617Phe induces constitutive activation of downstream signaling pathways in cell lines and primary cells[20]. iPSC-derived



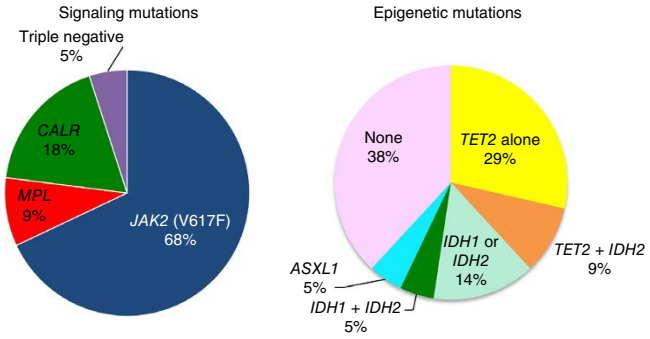


Figure 3 Pie chart representation of acquired signaling and epigenetic mutations in MPN from the four families.

erythroblasts were expanded in liquid culture in the presence of EPO and SCF from day 12 to day 18 and then deprived of cytokines. Constitutive STAT5, ERK and AKT phosphorylation was observed only in P3-CNV-VF and P3-CNV-VF-TET2 cells (**Supplementary Fig. 5d**), confirming their EPO-independent growth. This effect was not observed in either P2-VF cells, bearing only a heterozygous *JAK2* mutation encoding p.Val617Phe, or P4-CNV cells. EPO induced the phosphorylation of STAT5, ERK and AKT in all erythroblast cultures, whatever their genotype. These results argue for a synergistic effect of the CNV and the *JAK2* and *TET2* mutations on erythroid progenitor sensitivity to EPO.

We subsequently explored the TPO sensitivity of megakaryocyte progenitors. Day 12 GPA$^+$CD41$^+$ progenitors were sorted and grown for 12 d in serum-free fibrin clots in the presence or absence of increasing TPO concentrations. Rare TPO-independent megakaryocyte colonies, around 10% of the maximum number for TPO-stimulated cultures, were obtained for megakaryocyte progenitors generated from control iPSCs in contrast to the 50% of maximum TPO-stimulated colonies obtained for TPO-independent megakaryocyte colonies generated from iPSCs heterozygous for the *JAK2* mutation, as described[15] (**Supplementary Fig. 5e**). Interestingly, the predisposition locus alone (P4-CNV) caused a similar induction in the formation of TPO-independent megakaryocytes (resulting in around 40% of the maximum colony numbers for TPO-stimulated cells), an effect that was further enhanced by the *JAK2* mutation (P3-CNV-VF), whereas *TET2* mutation (P3-CNV-VF-TET2) did not further modify the response to TPO. Finally, when GPA$^+$CD41$^+$ progenitors were cultured in the presence of SCF and TPO, the megakaryocytes derived from all iPSCs harboring the predisposition locus showed enrichment for hyperploid megakaryocyte cells (**Supplementary Fig. 5f**,**g**).

To validate these observations, we went back to CD34$^+$ hematopoietic progenitors from patients, including F2:III-6 (CNV alone), F2:IV-2 (CNV and a low burden of *JAK2* mutation (p.Val617Phe)), F2:III-2 (CNV and *JAK2* (p.Val617Phe) and *TET2* mutation) and F1:II-1 (CNV and *JAK2* (p.Val617Phe) and *IDH1* mutation), and explored their ability to form EECs and endogenous CFU-MKs. We observed that the CNV alone promoted the spontaneous generation of colony forming unit–megakaryocytes (CFU-MKs) and that this effect was reinforced by *JAK2* mutation (**Fig. 6a**). We also detected a tiny amount of spontaneous EEC formation with the CNV alone that was further enhanced by the *JAK2* mutation in patients F2:IV-2 and F1:II-1 (**Fig. 6b**). The *TET2* mutation, in contrast to the *IDH1* mutation, did not have an additive effect on this response but instead affected the size of the colonies and, thus, the proliferation rate of the erythroblasts (F2:III-6 versus F2:III-2; **Fig. 6c**). Taken together, the results suggest that the CNV predisposition locus promotes megakaryopoiesis by inducing the spontaneous generation of CFU-MKs and the formation of hyperploid megakaryocytes. The CNV predisposition locus also seems to cooperate with the *JAK2* mutation to modify the response of hematopoietic progenitors to EPO.

Characterization of genes involved in the phenotype

To investigate the function of *ATG2B* and *GSKIP* in hematopoiesis, we transduced CD34$^+$ progenitor cells from controls with a lentivirus expressing either short hairpin RNA (shRNA) targeting these genes or a scrambled sequence, with GFP as a selection marker. The shRNAs to *ATG2B* and *GSKIP*, alone or in combination, resulted in about a 40–50% reduction in the transcript levels of their respective targets (**Fig. 7a**). We explored the effect of these shRNAs on megakaryocyte progenitors and found a significant decrease in the frequency of CFU-MKs with each shRNA alone or with the shRNAs used in combination. In addition, the simultaneous silencing of both *ATG2B*

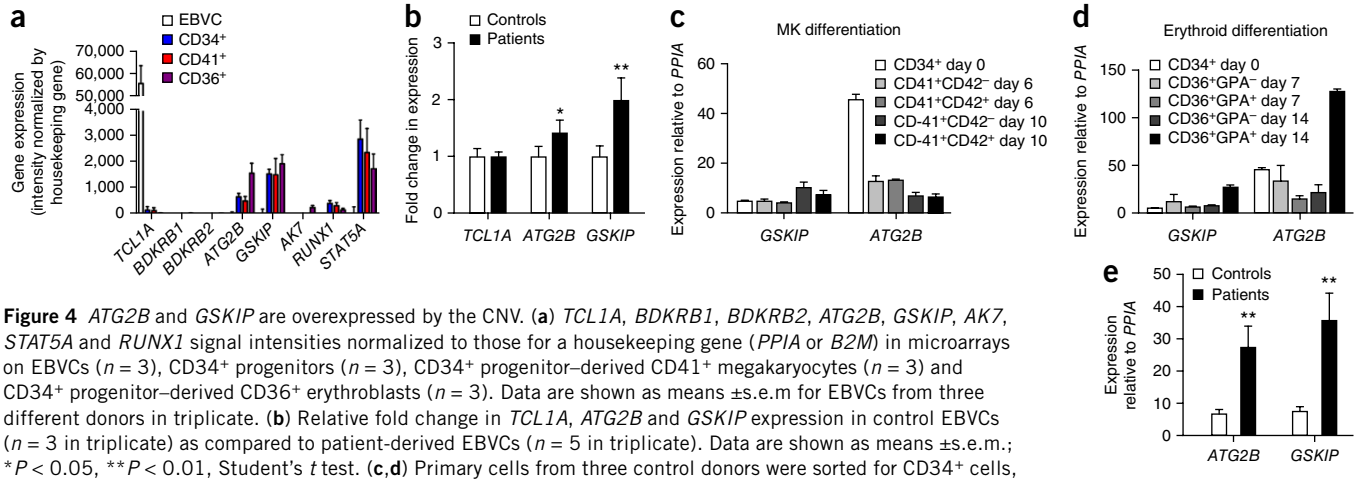


Figure 4 *ATG2B* and *GSKIP* are overexpressed by the CNV. (**a**) *TCL1A*, *BDKRB1*, *BDKRB2*, *ATG2B*, *GSKIP*, *AK7*, *STAT5A* and *RUNX1* signal intensities normalized to those for a housekeeping gene (*PPIA* or *B2M*) in microarrays on EBVCs ($n = 3$), CD34$^+$ progenitors ($n = 3$), CD34$^+$ progenitor–derived CD41$^+$ megakaryocytes ($n = 3$) and CD34$^+$ progenitor–derived CD36$^+$ erythroblasts ($n = 3$). Data are shown as means ±s.e.m for EBVCs from three different donors in triplicate. (**b**) Relative fold change in *TCL1A*, *ATG2B* and *GSKIP* expression in control EBVCs ($n = 3$ in triplicate) as compared to patient-derived EBVCs ($n = 5$ in triplicate). Data are shown as means ±s.e.m.; $^*P < 0.05$, $^{**}P < 0.01$, Student's *t* test. (**c**,**d**) Primary cells from three control donors were sorted for CD34$^+$ cells, grown with SCF (25 ng/ml) and TPO (20 ng/ml) (**c**) or with SCF (25 ng/ml), interleukin (IL)-3 (10 µg/ml) and EPO (1 U/ml) (**d**), and sorted at day 6 or 10 for megakaryocyte (MK) differentiation (**c**) or at day 7 or 14 for erythroid differentiation (**d**). *ATG2B* and *GSKIP* levels were quantified by qRT-PCR in each fraction. Expression relative to that of *PPIA* was calculated. Data are shown as means ±s.e.m. (**e**) *ATG2B* and *GSKIP* expression was quantified by qRT-PCR in CD41$^+$CD42$^+$ cells derived from three controls and three patients at day 10 of megakaryocyte differentiation. Data are shown as means ±s.e.m.; $^{**}P < 0.01$, Student's *t* test.

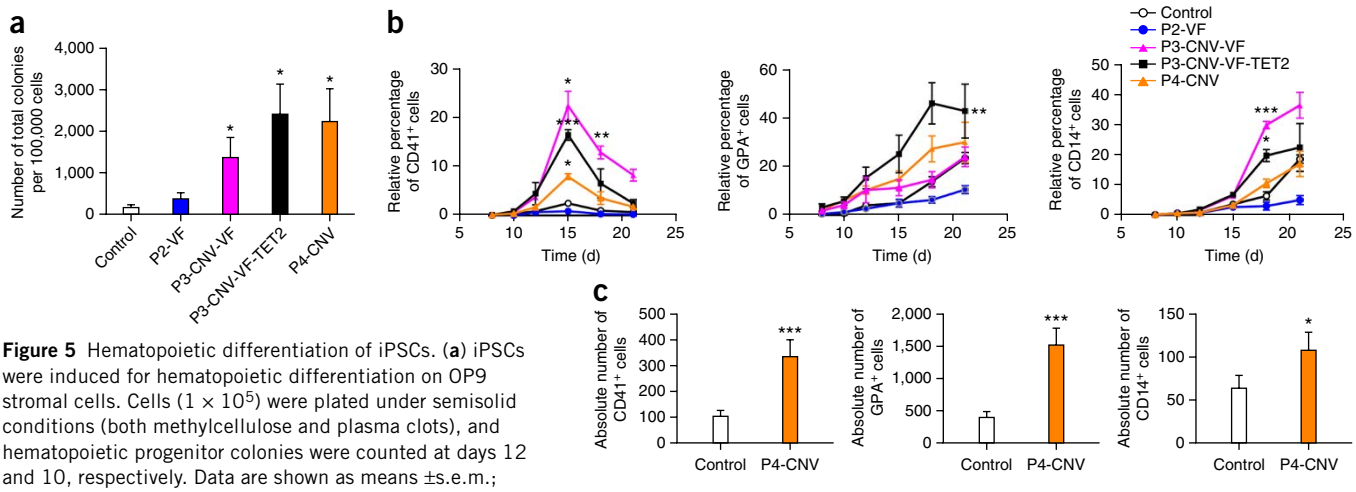


Figure 5 Hematopoietic differentiation of iPSCs. (**a**) iPSCs were induced for hematopoietic differentiation on OP9 stromal cells. Cells (1×10^5) were plated under semisolid conditions (both methylcellulose and plasma clots), and hematopoietic progenitor colonies were counted at days 12 and 10, respectively. Data are shown as means ±s.e.m.; $n = 4$; $^*P < 0.05$. (**b**) Hematopoietic differentiation of iPSCs (control, P2-VF, P3-CNV-VF, P3-CNV-VF-TET2 and P4-CNV) was induced by seeding iPSCs onto OP9 stromal cells in the presence of VEGF and hematopoietic cytokines. Bulk and hematopoietic cells collected at days 8, 10, 12, 15, 18 and 21 were analyzed by flow cytometry. The relative percentages of $CD41^+$, GPA^+ and $CD14^+$ cells were calculated. Data are shown as means ±s.e.m.; $n = 3$; $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, Student's t test for $CD41^+$ and $CD14^+$ cells and Bonferroni test for GPA^+ cells. (**c**) The hematopoietic potential of iPSCs was quantified by plating one $TRA\text{-}1\text{-}81^+$ cell per well in a 96-well plate coated with OP9 stromal cells. The absolute numbers of $CD41^+$, GPA^+ and $CD14^+$cells in each clone were measured by flow cytometry at day 18. Data are shown as means ±s.e.m.; $n > 20$ clones per genotype, 2 independent experiments; $^*P < 0.05$, $^{***}P < 0.001$, Mann-Whitney test, two-tailed).

and *GSKIP* led to a further decrease in CFU-MK size beyond that resulting from silencing with the shRNAs individually (**Fig. 7b**,**c**).

To investigate the role of these genes in the phenotypic consequences of the genomic duplication, the shRNAs were virally transduced, alone or in combination, into $CD41^+GPA^+$ megakaryocyte progenitors generated from P4-CNV iPSCs at day 12, and the progenitors were grown for 12 additional days in serum-free fibrin clots in the presence or absence of TPO (**Fig. 7d**). *ATG2B* and *GSKIP* shRNA alone had no impact on spontaneous CFU-MK formation, but the combination of these shRNAs resulted in a major inhibition of TPO-independent CFU-MK formation (**Fig. 7e**). The role of *ATG2B* and *GSKIP* in the spontaneous generation of megakaryocytes was further confirmed by virally transducing both shRNAs into $CD34^+$ hematopoietic progenitor cells from patients (**Fig. 7f**,**g**). Conversely, we failed to detect any effect of *TCL1A* downregulation by shRNA under the same conditions, and overexpression of *TCL1A* in $CD34^+$ progenitors did not result in the spontaneous generation of megakaryocytes (data not shown). Altogether, our findings indicate that both *ATG2B* and *GSKIP* are responsible for the megakaryocyte phenotype induced by the CNV predisposition locus.

Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of the genes differentially expressed by control- and patient-derived EBVCs identified eight significantly different gene sets and points to abnormal functioning of the endoplasmic reticulum with deregulated chaperoning and glycosylation (**Supplementary Table 4**). Transcriptome analysis showed a significant ($P < 0.001$) increase in the expression of *WIPI1* (*ATG18*), which encodes a protein that interacts with ATG2B, in patients with MPN. Moreover, autophagosome accumulation was found in platelets from patients or asymptomatic carriers, as shown by an increase in the levels of the autophagosome marker LC3-II, a phosphatidylethanolamine-conjugated form of LC3-I (**Supplementary Fig. 6a**,**b**).

DISCUSSION

We identified a 700-kb germline duplication that predisposes patients to myeloid malignancies, including AML, MPN, CMML and frequent essential thrombocythemia with progression to myelofibrosis and secondary AML. Two of the six duplicated genes, namely *ATG2B* and *GSKIP*, are overexpressed in hematopoietic progenitors. In turn, megakaryocyte progenitors demonstrate increased sensitivity to TPO. The overexpression of these genes cooperates with classical mutations in *JAK2*, *MPL* and *CALR* to generate the MPN phenotype. This germline duplication is, to our knowledge, the first identified genomic alteration that accounts for familial MPN.

Some susceptibility alleles, such as the one in the *JAK2* 46/1 haplotype, an intronic SNP in the *TERT* gene[9–13] and variants in the

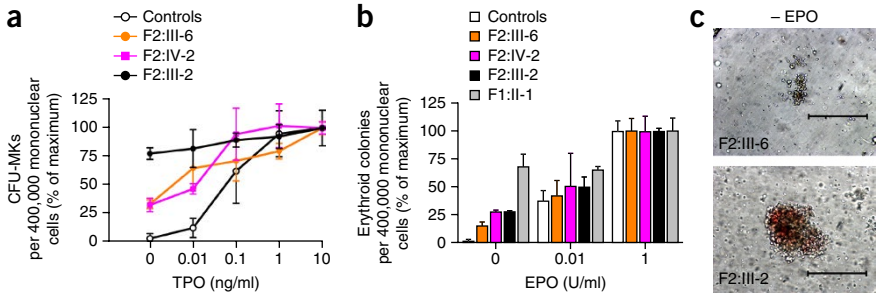


Figure 6 The CNV predisposition locus modifies the sensitivity to EPO and TPO in primary cells from patients. (**a**–**c**) Primary $CD34^+$cells from three control donors and four patients (F2:III-6, F2:IV-2, F2:III-2 and F1:II-1) were grown with SCF (25 ng/ml) in the presence or absence of various doses of TPO (**a**) or were grown with SCF (25 ng/ml) and IL-3 (10 µg/ml) in the presence or absence of various doses of EPO (**b**,**c**) and then grown under semisolid conditions either in plasma clots (**a**) or methylcellulose (**b**,**c**). The percentage of CFU-MKs (**a**) or erythroid colonies (**b**) was calculated in comparison to the condition with maximum growth (with either 10 ng/ml TPO or 1 U/ml EPO, respectively). Results are shown as means ±s.d. (**c**) Images represent EECs (obtained without EPO). Scale bars, 200 µm.

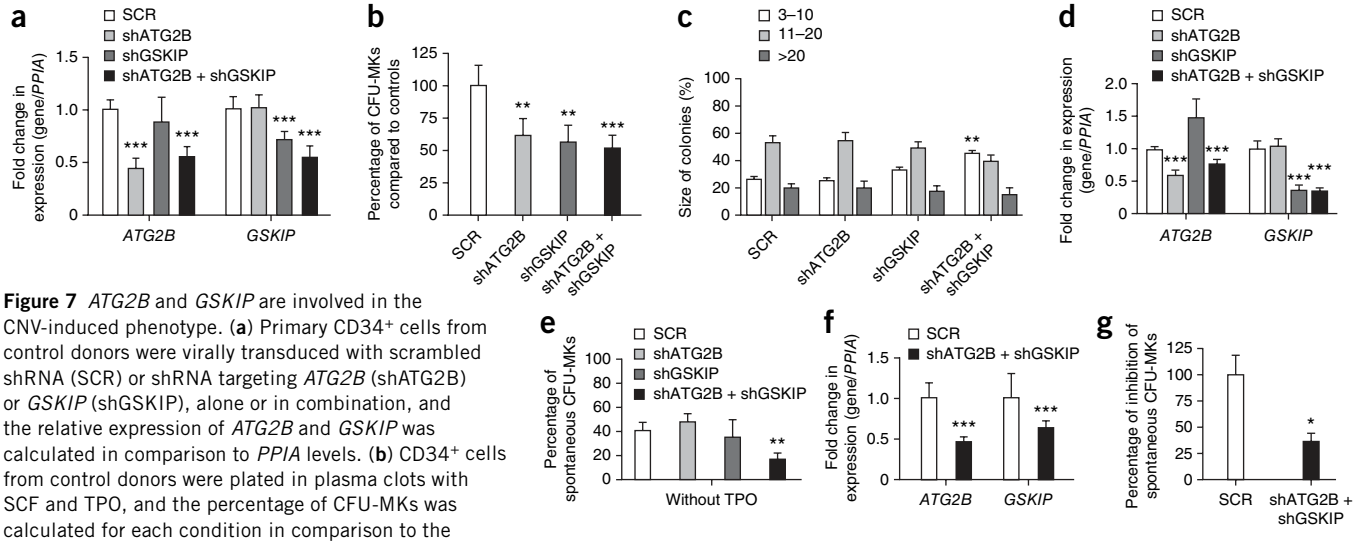


Figure 7 *ATG2B* and *GSKIP* are involved in the CNV-induced phenotype. (**a**) Primary CD34$^+$ cells from control donors were virally transduced with scrambled shRNA (SCR) or shRNA targeting *ATG2B* (shATG2B) or *GSKIP* (shGSKIP), alone or in combination, and the relative expression of *ATG2B* and *GSKIP* was calculated in comparison to *PPIA* levels. (**b**) CD34$^+$ cells from control donors were plated in plasma clots with SCF and TPO, and the percentage of CFU-MKs was calculated for each condition in comparison to the control with scrambled shRNA. (**c**) CFU-MKs were classified on the basis of megakaryocyte cluster size (corresponding to the number of megakaryocytes per colony). (**d**) CD41$^+$ cells derived from P4-CNV iPSCs were virally transduced with shRNAs to *ATG2B* and *GSKIP*, and the expression of these genes was calculated in comparison to *PPIA* levels. (**e**) iPSC-derived hematopoietic progenitors (CD41$^+$GPA$^+$) were plated in plasma clots with or without TPO, and CFU-MKs were counted at day 10 after indirect staining for CD41a. The percentage of spontaneous CFU-MKs was calculated for each condition in comparison to the condition with maximum growth. (**f**) CD34$^+$ cells purified from patients F2:III-2, F2:IV-4, F2:IV-5 and F2:III-10 were virally transduced with scrambled shRNA or shRNAs to *ATG2B* and *GSKIP*, and the expression of these genes was calculated in comparison to *PPIA* levels. (**g**) CD34$^+$ cells from patients (F2:III-2, F2:IV-4, F2:IV-5 and F2:III-10) were plated in plasma clots with or without SCF and TPO, and CFU-MKs were counted at day 10 after indirect staining for CD41a. The percentage of inhibition of spontaneous CFU-MK formation with combined shRNAs to *ATG2B* and *GSKIP* was calculated in comparison to the condition with scrambled control shRNA. Results are the means ±s.e.m. of three independent experiments for **b**, **c** and **e**, four independent experiments for **a**, **f** and **g**, and five independent experiments for **d**; $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, Student's *t* test.

ERCC2, *ATM*, *CCDC6* and *NR3C1* (*GR*) genes, are diversely present in the general population and favor the development or evolution of sporadic cases of MPN[21,22]. However, they appear to have limited roles in familial MPN. CNVs, involving either deletion or amplification, have been frequently related to cancer predisposition[23,24]. The 700-kb germline duplicated region identified here, which segregates among the 24 affected family members of the 4 families, is associated with a high penetrance level, above 80%. The predisposition locus is located in the 14q32.2 region, which is rarely affected by recurrent cytogenetic aberrations in chronic and acute phases of MPN evolution[25–27], although trisomy 14 has been associated with myeloid malignancies that develop in older individuals[28,29], in MDS, atypical chronic myeloid leukemia (aCML) and CMML[30,31]. Interestingly, one patient (F1:II-7) who directly developed an acute leukemia demonstrated mosaic trisomy 14 with up to five copies of the CNV, arguing for a gene dosage effect.

Our familial MPN cases were characterized by an earlier age of MPN onset in comparison to sporadic cases (41 years versus >60 years). Notably, the spectrum of acquired driver mutations leading to essential thrombocythemia included *JAK2* (p.Val617Phe), *MPL* and *CALR* mutations, with some cases triple negative for these mutations, similar to the spectrum of mutations in sporadic essential thrombocythemia cases[32,33]. Additional genetic events were detected in this setting, including the combination of a *BCR-ABL* fusion gene[34] and epigenetic regulator gene mutations affecting *TET2*, *IDH1*, *IDH2* and *ASXL1*, when the MPN progressed. The percentage of cases with *TET2* mutations (38%) was very high when compared to that described for cases of sporadic essential thrombocythemia and other familial clusters of MPN (10–15%)[7,35]. The detection of *TET2* mutations (including biallelic mutations) and *IDH1* and *IDH2* mutations was frequently associated with disease progression to acute leukemia, in agreement with some observations suggesting that, in MPN, the acquisition of a *TET2* mutation is predictive of a poor outcome and a high risk of transformation into leukemia[7,36,37]. No mutation of *TP53* was found in our cases, contrary to what was observed in AML evolving from MPN, suggesting two different pathways for leukemic transformation.

A well-studied chromosomal duplication that predisposes to acute leukemia is the trisomy 21 responsible for Down syndrome. Trisomy 21 predisposes to the acquisition of *GATA1* mutations resulting in a short form of the transcription factor, which induces a transient myeloproliferative disorder. In about 25% of cases, this transient proliferation evolves into megakaryoblastic leukemia through the acquisition of secondary mutations that target cohesin components, epigenetic modulators and signaling pathways, including *JAK2* and *MPL* mutations[38]. There is evidence that trisomy 21, by increasing the proliferation of erythroid and megakaryocyte progenitors in fetal liver, predisposes to leukemia. This effect has been modeled in iPSCs, as mouse models of trisomy 21 failed to fully recapitulate the disease predisposition[39,40]. We also established iPSC clones to explore the consequences of the gene duplication, alone or in combination with acquired mutations, on hematopoiesis, showing that the CNV (i) promotes the amplification of hematopoietic progenitors, including megakaryocytes, through increasing their sensitivity to TPO; (ii) consequently induces an increase in the size and ploidy of mature CD41$^+$ megakaryocytes; (iii) cooperates with *JAK2* (p.Val617Phe) mutation to increase the sensitivity of erythroid progenitors to EPO and promote constitutive activation of signaling pathways; and (iv) synergizes with *TET2* and *JAK2* (p.Val617Phe) mutations to promote erythroid cell proliferation and amplification. The hypersensitivity of hematopoietic precursors carrying the CNV to EPO and TPO was further confirmed in primary cells collected from patients. As with trisomy 21, the identified germline duplication did not result in overt genetic instability but rather increased the fitness of the *JAK2* mutation encoding p.Val617Phe by acting synergistically with it.

In agreement with this, we observed a similar occurrence of higher numbers of mutations (around 10–15) during iPSC reprogramming in cells derived from patients in comparison to control cells from normal donors[41]. Moreover, recent data show that the *JAK2* mutation encoding p.Val617Phe is quite frequent in the general aging population[42–45], suggesting that it rarely induces MPN development, as demonstrated in mice engrafted with a single hematopoietic stem cell (HSC) expressing Jak2 Val617Phe[46]. Therefore, the duplication could provide a favorable genetic background to facilitate the clonal dominance of a cell expressing JAK2 Val617Phe (**Supplementary Fig. 7**). In support of this hypothesis, (i) the duplication enhances the fitness of mutations but does not directly cause a mutant phenotype and (ii) the spectrum and distribution of signaling mutations in these families are similar to those of sporadic MPN cases.

Of the six duplicated genes in the CNV region, only the expression of *ATG2B* and *GSKIP* was reproducibly detected in hematopoietic cells, including $CD34^+$ hematopoietic progenitors, megakaryocytes and erythroid cells. *TCL1A*, whose expression was first described in mature T cell leukemia, was not expressed in $CD34^+$ cells or myeloid cells, nor was it overexpressed in patient-derived EBVCs. Strikingly, *ATG2B* cooperated with *GSKIP* to induce the spontaneous formation of CFU-MKs from patient-derived cells and iPSC-derived hematopoietic cells, and silencing of *ATG2B* and *GSKIP* reversed this phenotype. Loss-of-function mutations in *ATG2B*, which encodes an important player in autophagy[47], have been identified in gastric and colorectal cancers[48], and deregulated autophagy could affect HSC self-renewal capabilities, for example, in aging subjects[49,50]. Accordingly, we found increased autophagosome accumulation in platelets from patients and asymptomatic carriers of the CNV. The *GSKIP* gene encodes a negative regulator of GSK3β (refs. 51,52), and its mutation may thus mimic activation of the Wnt/β-catenin signaling pathway involved in HSC homeostasis and normal megakaryopoiesis[53,54], as well as transformation in chronic myeloid leukemia and the development of leukemia stem cells in AML[55,56]. Accordingly, the overexpression of these two genes may account for the progression of essential thrombocythemia to myelofibrosis and AML observed in the families studied here. Complementary studies will explore how *ATG2B* and *GSKIP* overexpression cooperates with *JAK2* (p.Val617Phe) mutation to provide a growth advantage to HSCs. It may also be useful to determine whether *ATG2B* or *GSKIP* amplification is a frequent event in a large cohort of patients with sporadic MPN.

In summary, analysis of four families with the same geographical origin identifies a locus whose duplication promotes the occurrence of severe MPN through overexpression of *ATG2B* and *GSKIP*, resulting in increased fitness for cells bearing somatic mutations such as mutations in *JAK2*, *MPL* and *CALR*. Together with the recent observation that these mutations are frequent in otherwise healthy aging individuals, these results pave the way for the search for other genetic events that promote the expansion of mutated cells and the generation of overt MPNs, with or without progression to myelofibrosis and AML.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The microarray data have been submitted to the ArrayExpress data repository at the European Bioinformatics Institute under accession E-MTAB-3570. CGH array data are available at the Gene Expression Omnibus (GEO) under accession GSE67938. The CNVs found in patients were deposited in ClinVar under accessions SCV000223992, SCV000224007, SCV000224008, SCV000224009 and SCV000224010.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We greatly thank all the patients and family members involved in the study. We also thank O. Bawa and P. Opolon for the histopathological analysis of teratomas. We thank B. Benyahia for cytogenetic analysis. We thank M. Vestris for recruitment of patients. We greatly thank B. Job and the genomic platform for transcriptome and CGH analysis and also the iPSC platform of Institut Gustave Roussy. We thank S. Saker and T. Larmonier from the Généthon DNA and Cell Bank (Evry, France) for the establishment of B-lymphoblastoid cells from 'NMP' patients. We thank the cytometry platform of Institut Gustave Roussy (P. Rameau and Y. Lecluse). We thank C. Marzac for clinical data. We are grateful to S. Constantinescu and J. Feunteun for critical reading of the manuscript. We are also very grateful to E. Schwartz for proofreading the manuscript.

This work was supported by grants from Agence Nationale de la Recherche (ANR) (Blanc Megon 2009, Thrombocytosis 2011; ANR-13-JVSV1-GERMPN-01), Association pour la Recherche contre le Cancer (ARC) (Fondation ARC Libre 2012-SL220120605292), Groupe Information Santé (GIS)–Institute for Rare Diseases for High-Throughput Sequencing (AO9102LS), Association de Recherche sur la Moelle Osseuse (ARMO), regional Programme Hospitalier de Recherche Clinique (PHRC) AOR07014, Association Laurette Fugain and INCa-DGOS-INSERM 6043. Labex GR-Ex (I.P. and W.V.) is funded by the program 'Investissements d'Avenir'. G. Lenglet was supported by a postdoctoral fellowship from Ile-de-France Cancéropôle and ANR Molecular Medicine in Oncology (MMO) (funded by the program 'Investissements d'Avenir'). F.P. was supported by ARC. L.S. and J.S. were supported by doctoral grants from the Ile-de-France region (Cancéropôle and DIM Cellule Souche) and from Fondation pour la Recherche Médicale (FRM). C.M. was supported by ANR-13-JVSV1-GERMPN-01.

AUTHOR CONTRIBUTIONS

W.V., I.P. and C.B.-C. designed and performed research, analyzed data, prepared figures and wrote the manuscript. J.S. performed research on iPSCs and primary cells. G. Lenglet, A.D.S. and L.S. performed research on iPSCs. C.S.-M., N. Droin and G. Leroy performed pangenomic analysis. C.M. performed immunoblot and qRT-PCR analyses. A.P. purified primary cells from donors. E.M. performed colony assays. F.P., J.-C.M., C.D.-D., P.F., F.I. and N.C. were involved in the clinical aspect of the study. A.N. was involved in the clinical aspect of the study and initiated the familial study of MPN. S.G. and S.C. were involved in the generation of EBVCs and their study. B.K. performed cytogenetic analysis. M'b.D., J.M. and P.D. carried out bioinformatics study of exome and transcriptomic data. N. Debili and H.R. provided experimental and/or intellectual input on iPSC culture and hematopoietic differentiation. V.D.V. performed TCL1 mouse modeling. E.S. and O.A.B. contributed intellectual input. All authors contributed to writing and editing.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Song, W.J. *et al.* Haploinsufficiency of *CBFA2* causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat. Genet.* **23**, 166–175 (1999).
2. Smith, M.L., Cavenagh, J.D., Lister, T.A. & Fitzgibbon, J. Mutation of *CEBPA* in familial acute myeloid leukemia. *N. Engl. J. Med.* **351**, 2403–2407 (2004).
3. Hahn, C.N. *et al.* Heritable *GATA2* mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat. Genet.* **43**, 1012–1017 (2011).
4. Pasquet, M. *et al.* High frequency of *GATA2* mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia. *Blood* **121**, 822–829 (2013).
5. Bluteau, D. *et al.* Thrombocytopenia-associated mutations in the *ANKRD26* regulatory region induce MAPK hyperactivation. *J. Clin. Invest.* **124**, 580–591 (2014).
6. Bellanné-Chantelot, C. *et al.* Genetic and clinical implications of the Val617Phe *JAK2* mutation in 72 families with myeloproliferative disorders. *Blood* **108**, 346–352 (2006).
7. Saint-Martin, C. *et al.* Analysis of the ten-eleven translocation 2 (*TET2*) gene in familial myeloproliferative neoplasms. *Blood* **114**, 1628–1632 (2009).

8. Yamada, O. *et al.* Emergence of a *BCR-ABL* translocation in a patient with the *JAK2*V617F mutation: evidence for secondary acquisition of *BCR-ABL* in the *JAK2*V617F clone. *J. Clin. Oncol.* **32**, e76–e79 (2014).
9. Jones, A.V. *et al. JAK2* haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat. Genet.* **41**, 446–449 (2009).
10. Kilpivaara, O. *et al.* A germline *JAK2* SNP is associated with predisposition to the development of *JAK2*^{V617F}-positive myeloproliferative neoplasms. *Nat. Genet.* **41**, 455–459 (2009).
11. Olcaydu, D. *et al.* A common *JAK2* haplotype confers susceptibility to myeloproliferative neoplasms. *Nat. Genet.* **41**, 450–454 (2009).
12. Jäger, R. *et al.* Common germline variation at the *TERT* locus contributes to familial clustering of myeloproliferative neoplasms. *Am. J. Hematol.* **89**, 1107–1110 (2014).
13. Oddsson, A. *et al.* The germline sequence variant rs2736100_C in *TERT* associates with myeloproliferative neoplasms. *Leukemia* **28**, 1371–1374 (2014).
14. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
15. Saliba, J. *et al.* Heterozygous and homozygous *JAK2*^{V617F} states modeled by induced pluripotent stem cells from myeloproliferative neoplasm patients. *PLoS ONE* **8**, e74257 (2013).
16. Takayama, N. *et al.* Generation of functional platelets from human embryonic stem cells *in vitro* via ES-sacs, VEGF-promoted structures that concentrate hematopoietic progenitors. *Blood* **111**, 5298–5306 (2008).
17. Klimchenko, O. *et al.* A common bipotent progenitor generates the erythroid and megakaryocyte lineages in embryonic stem cell–derived primitive hematopoiesis. *Blood* **114**, 1506–1517 (2009).
18. Vodyanik, M.A., Bork, J.A., Thomson, J.A. & Slukvin, I.I. Human embryonic stem cell–derived CD34^{+} cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. *Blood* **105**, 617–626 (2005).
19. Prchal, J.F. & Axelrad, A.A. Letter: bone-marrow responses in polycythemia vera. *N. Engl. J. Med.* **290**, 1382 (1974).
20. James, C. *et al.* A unique clonal *JAK2* mutation leading to constitutive signalling causes polycythaemia vera. *Nature* **434**, 1144–1148 (2005).
21. Jones, A.V. & Cross, N.C. Inherited predisposition to myeloproliferative neoplasms. *Ther. Adv. Hematol.* **4**, 237–253 (2013).
22. Harutyunyan, A.S. & Kralovics, R. Role of germline genetic factors in MPN pathogenesis. *Hematol. Oncol. Clin. North Am.* **26**, 1037–1051 (2012).
23. Krepischi, A.C., Pearson, P.L. & Rosenberg, C. Germline copy number variations and cancer predisposition. *Future Oncol.* **8**, 441–450 (2012).
24. Kuiper, R.P., Ligtenberg, M.J., Hoogerbrugge, N. & Geurts van Kessel, A. Germline copy number variation and cancer risk. *Curr. Opin. Genet. Dev.* **20**, 282–289 (2010).
25. Klampfl, T. *et al.* Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. *Blood* **118**, 167–176 (2011).
26. Rice, K.L. *et al.* Analysis of genomic aberrations and gene expression profiling identifies novel lesions and pathways in myeloproliferative neoplasms. *Blood Cancer J.* **1**, e40 (2011).
27. Rumi, E. *et al.* Identification of genomic aberrations associated with disease transformation by means of high-resolution SNP array analysis in patients with myeloproliferative neoplasm. *Am. J. Hematol.* **86**, 974–979 (2011).
28. Cui, W. *et al.* Trisomy 14 as a sole chromosome abnormality is associated with older age, a heterogenous group of myeloid neoplasms with dysplasia, and a wide spectrum of disease progression. *J. Biomed. Biotechnol.* **2010**, 365318 (2010).
29. Mancini, M. *et al.* Trisomy 14 in hematologic diseases. Another non-random abnormality within myeloid proliferative disorders. *Cancer Genet. Cytogenet.* **66**, 39–42 (1993).
30. Mertens, F. *et al.* Trisomy 14 in atypical chronic myeloid leukemia. *Leukemia* **4**, 117–120 (1990).
31. Toze, C.L., Barnett, M.J., Naiman, S.C. & Horsman, D.E. Trisomy 14 is a non-random karyotypic abnormality associated with myeloid malignancies. *Br. J. Haematol.* **98**, 177–185 (1997).
32. Bellanne-Chantelot, C., Jego, P., Lionne-Huyghe, P., Tulliez, M. & Najman, A. The *JAK2*^{V617F} mutation may be present several years before the occurrence of overt myeloproliferative disorders. *Leukemia* **22**, 450–451 (2008).
33. Rumi, E. *et al. CALR* exon 9 mutations are somatically acquired events in familial cases of essential thrombocythemia or primary myelofibrosis. *Blood* **123**, 2416–2419 (2014).
34. Cabagnols, X., Cayuela, J.M. & Vainchenker, W. A *CALR* mutation preceding *BCR-ABL1* in an atypical myeloproliferative neoplasm. *N. Engl. J. Med.* **372**, 688–690 (2015).
35. Delhommeau, F. *et al.* Mutation in *TET2* in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
36. Abdel-Wahab, O. *et al.* Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Res.* **70**, 447–452 (2010).
37. Lundberg, P. *et al.* Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood* **123**, 2220–2228 (2014).
38. Yoshida, K. *et al.* The landscape of somatic mutations in Down syndrome–related myeloid disorders. *Nat. Genet.* **45**, 1293–1299 (2013).
39. Chou, S.T. *et al.* Trisomy 21–associated defects in human primitive hematopoiesis revealed through induced pluripotent stem cells. *Proc. Natl. Acad. Sci. USA* **109**, 17573–17578 (2012).
40. Maclean, G.A. *et al.* Altered hematopoiesis in trisomy 21 as revealed through *in vitro* differentiation of isogenic human pluripotent cells. *Proc. Natl. Acad. Sci. USA* **109**, 17567–17572 (2012).
41. Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63–67 (2011).
42. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
43. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
44. McKerrell, T. *et al.* Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep.* **10**, 1239–1245 (2015).
45. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).
46. Lundberg, P. *et al.* Myeloproliferative neoplasms can be initiated from a single hematopoietic stem cell expressing JAK2-V617F. *J. Exp. Med.* **211**, 2213–2230 (2014).
47. Kishi-Itakura, C., Koyama-Honda, I., Itakura, E. & Mizushima, N. Ultrastructural analysis of autophagosome organization using mammalian autophagy-deficient cells. *J. Cell Sci.* **127**, 4089–4102 (2014).
48. Kang, M.R. *et al.* Frameshift mutations of autophagy-related genes *ATG2B*, *ATG5*, *ATG9B* and *ATG12* in gastric and colorectal cancers with microsatellite instability. *J. Pathol.* **217**, 702–706 (2009).
49. Mortensen, M., Watson, A.S. & Simon, A.K. Lack of autophagy in the hematopoietic system leads to loss of hematopoietic stem cell function and dysregulated myeloid proliferation. *Autophagy* **7**, 1069–1070 (2011).
50. Warr, M.R. *et al.* FOXO3A directs a protective autophagy program in haematopoietic stem cells. *Nature* **494**, 323–327 (2013).
51. Chou, H.Y. *et al.* GSKIP is homologous to the Axin GSK3(interaction domain and functions as a negative regulator of GSK3β. *Biochemistry* **45**, 11379–11389 (2006).
52. Lin, C.C. *et al.* GSKIP, an inhibitor of GSK3β, mediates the N-cadherin/β-catenin pool in the differentiation of SH-SY5Y cells. *J. Cell. Biochem.* **108**, 1325–1336 (2009).
53. Li, D., August, S. & Woulfe, D.S. GSK3β is a negative regulator of platelet function and thrombosis. *Blood* **111**, 3522–3530 (2008).
54. Soda, M., Willert, K., Kaushansky, K. & Geddis, A.E. Inhibition of GSK-3β promotes survival and proliferation of megakaryocytic cells through a β-catenin–independent pathway. *Cell. Signal.* **20**, 2317–2323 (2008).
55. Abrahamsson, A.E. *et al.* Glycogen synthase kinase 3β missplicing contributes to leukemia stem cell generation. *Proc. Natl. Acad. Sci. USA* **106**, 3925–3929 (2009).
56. Wang, Y. *et al.* The Wnt/β-catenin pathway is required for the development of leukemia stem cells in AML. *Science* **327**, 1650–1653 (2010).

ONLINE METHODS

Patients. All participants in this study gave their written informed consent in accordance with the Declaration of Helsinki, and the study was approved by the local research ethics committee from Saint-Antoine Hospital (Paris, France). MPN was defined according to World Health Organization (WHO) criteria for essential thrombocythemia and myelofibrosis[57]. Clinical features and hematological parameters were collected at diagnosis and during the course of the disease and recorded in an Access database approved by the French computer commission (CNIL 815419).

We studied 4 families originating from the West Indies and analyzed as controls 98 unrelated familial cases collected through a national network previously described as controls[6,58]. We also used geographically matched DNA controls from 39 sporadic MPN cases recruited from the Department of Hematology (Fort de France Hospital) and 199 healthy controls of West Indies origin collected by the Department of Genetics (Pitié-Salpêtrière Hospital).

iPSC generation and culture. $CD34^+$ and $CD3^+$ cells were purified from blood mononuclear cells. $CD34^+CD38^-$ and $CD34^+CD38^+$ cells were cultured in serum-free medium with cytokines for 5 d before being infected with VSV-G pseudotyped retroviruses encoding *Oct4*, *Myc*, *Klf4* and *Sox2* (ref. 14). Six days later, cells were seeded on irradiated mouse embryonic fibroblasts (MEFs) in ES medium[59]. Colonies with an ES-like morphology were picked from day 20 to day 30 and expanded. Cells were routinely screened for mycoplasma using the Plasma Test kit (Invivogen).

Hematopoietic differentiation was performed on OP9 stromal cells in the presence of VEGF (20 ng/ml; Peprotech)[16]. The mouse OP9 cell line was from the RIKEN Institute, Osaka University. On day 7, EPO (1 U/ml; Amgen), TPO (20 ng/ml; a generous gift from Kirin), SCF (25 ng/ml; Biovitrum AB) and IL-3 (10 ng/ml; Miltenyi Biotec) were added, and on days 11 and 12 cells were enzymatically dissociated. The recovered cells were cultured or sorted on the basis of the expression of GPA and CD41. Clonal differentiation of iPSCs was also performed on OP9 cells.

Quantification of clonogenic progenitors in semisolid culture. Cells were plated either in methylcellulose to quantify erythroid and granulomonocytic (CFU-GM) progenitors or in serum-free fibrin clot assays to quantify CFU-MKs[60]. Cultures were scored after 12–14 d for all colonies[17]. CFU-MKs were enumerated at day 10 after labeling by an indirect alkaline phosphatase–based immunostaining technique using a monoclonal antibody to CD41a (Becton Dickinson, clone HIP8), as previously described[60].

Teratoma assays and embryoid bodies. iPSCs (1×10^6) were scraped off plates and resuspended in 140 µl of ES medium. Undiluted Matrigel (60 µl) was added before subcutaneous injection into $Rag2^{-/-}\gamma C^{-/-}$ mice. After 8–12 weeks, tumors were isolated and fixed in 10% formalin. Sections were stained for germ layer analysis. Spontaneous differentiation was achieved by embryoid body formation[17].

Antibodies and flow cytometry analysis. Monoclonal antibodies directly conjugated to fluorescent dye were used for iPSCs (SSEA4, eBioscience; TRA-1-81, BD) and for the sorting and characterization of hematopoietic cells (CD34, Beckman; CD43 (clone L10), CD42 (clone GRP-P) and GPA (clone CLB-ery-1(AME-1)), Invitrogen; CD41 (clone HIP2) and CD14 (clone M5E2), Pharmingen). Cells were sorted on an Influx flow cytometer (BD) and analyzed on a FACSCanto II (BD). iPSC colonies were stained by an alkaline phosphatase reaction (Stemgent).

Quantitative RT-PCR and gene expression array analysis. Total RNA was isolated using the RNeasy Mini kit (Qiagen), and cDNA was synthesized by SuperScript II Reverse Transcriptase (Invitrogen). PCR reactions were carried out on the ABI Prism GeneAmp 7500 Sequence Detection System (Applied Biosystems, Life Technologies), using Power SYBR Green PCR Master Mix (Invitrogen) and TaqMan gene expression assays for *TCL1A*, *BDKRB1*, *BDKRB2*, *ATG2B* and *GSKIP* (Applied Biosystems, Life Technologies). The expression for all genes was calculated relative to the levels of *PPIA* or *HPRT1*. For microarray analysis, RNA was hybridized to Agilent 4X44K arrays according to the manufacturer's protocols. Analysis was performed using Bioconductor and Rosetta Resolver (Microsoft). Gene set class comparison was performed with KEGG to classify regulated genes.

Karyotypes and CGH arrays. CGH arrays for $CD34^+$ cells and iPSCs were conducted on the human CGH 2x400K platform (G4448A, Agilent Technologies) by hybridization of sample versus matched normal commercially available reference, and hierarchical clustering was performed. Karyotype analysis was performed using standard procedures on R-banded metaphase spreads (450–600 bands).

Linkage analysis, microsatellite genotyping and single-nucleotide array analysis. Families F1 and F2 were subjected to a genome scan using the 6 K Illumina SNP genotyping Linkage IVb panel on a BeadStation system (Illumina). Allele detection and genotype calling were performed using BeadStudio software (Illumina). Genotype data from the genome scan were subjected to parametric linkage analyses in MERLIN using a dominant model with incomplete penetrance (0.8), a disease allele frequency of 0.000005 and no phenocopy.

A total of 16 microsatellite markers spanning the chromosome 14 candidate regions were added to the analysis. All microsatellites were analyzed by electrophoresis on the ABI3730 Genetic Analyzer (PE Applied Biosystems), and data were collected and analyzed using GENESCAN software Version 4.0 (PE Applied Biosystems). Multipoint parametric linkage analysis was performed in ALLEGRO using a dominant model with incomplete penetrance (0.8), a disease allele frequency of 0.000005 and no phenocopy. Microsatellite allele frequencies were deduced from a matched control population ($n = 33$).

SNP array analysis. Five affected cases (F1:II-2, F1:II-5, F1:II-7, F2:III-2 and F2:III-7; **Fig. 1**) and one unaffected control (F1:I-2) were genotyped using the Illumina Human CNV 370 BeadChip. Results were analyzed with the genome viewer in Illumina GenomeStudio software. The duplicated region was further confirmed by qRT-PCR based on SYBR Green I fluorescence using the ABI 7900 Sequence Detection System (Life Technologies, Thermo Scientific). Primers were designed mapping to two genes in proximal and distal parts of the duplicated region, and PCR was carried out in a 20-µl reaction using SYBR Green I PCR Master Mix (Life Technologies) together with 300 nmol/l of each primer and 20 ng of DNA. We used as reference the β-globin gene (*HBB*). The number of copies was determined as previously described[61]. Primer sequences are available in **Supplementary Table 2**.

Characterization of the duplication. The breakpoints of the duplicated motif were determined by quantitative RT-PCR based on SYBR Green I fluorescence. Oligonucleotides were designed for 12 amplicons between 100 and 200 bp in length (**Supplementary Table 2**). Refinement of the breakpoints and determination of the orientation of the two duplicated motifs were carried out by PCR and agarose gel electrophoresis (Chr14_B2C2.1F and Chr14_B1C2.3R; **Fig. 2c**) followed by sequencing analysis (Chr14_B2C2.1F/Chr14dup_joncR and Chr14dup_joncF/Chr14_B1C2.3R; **Fig. 2d**).

Targeted sequencing and whole-exome sequencing. Primers were designed with Primer3plus. The coding sequence and exon-intron boundaries of candidate genes were amplified, and PCR products were sequenced in both directions with the ABI PRISM BigDye Terminator v1.1 Ready Reaction Cycle Sequencing kit (Life Technologies) on an ABI PRISM 3730 Genetic Analyzer. Sequences were analyzed with Seqscape software v 2.2 (Life Technologies), and variants identified were analyzed with Alamutv2 software (Interactive Biosoftware). Signaling mutations of *JAK2* (c.1849G>T, p.Val617Phe), *MPL* (c.1544G>T, p.Trp515Leu) and *CALR* (c.1099_1150del, p.Leu367Thrfs*46 and c.1154_1155ins, p.Lys385Asnfs*47) were analyzed as previously described[53]. The spectrum of other acquired events was determined by Sanger sequencing of *IDH1* and *IDH2* (exons 4 and 5), *ASXL1* (exons 12 and 13) and the entire coding regions of *TET2*, *EZH2*, *DNMT3A* and *TP53*.

Whole-exome sequencing was performed using the HiSeq 2000 platform after capture with the Agilent kit (SureSelect v4). We analyzed the results by comparing $CD3^+$ non-tumoral cells to either $CD34^+$ cells or iPSCs and by

comparing $CD34^+$ cells to iPSCs ($CD3^+$ and $CD34^+$ cells were purified by CD34 or CD3 microbead kit (Miltenyi Biotec, 130-046-702 and 130-050-101)).

Immunoblot analysis. Signaling studies were performed on cultured erythroblasts after overnight cytokine deprivation in serum-free medium. Stimulation by EPO (10 U/ml) for 15 min served as a positive control. Samples were subjected to immunoblot analysis using polyclonal antibodies against the phosphorylated forms of STAT5 (Tyr694, 9359), ERK1 and ERK2 (Thr202/Tyr204, 9101) and AKT (Ser473, 9271) and the LC3I/II form of LC3 (12741) (Cell Signaling Technology). HSC70 was used as a loading control, and the corresponding antibody was from Stressgen.

Constructs and viral particle production. The sequences for shRNAs targeting human *ATG2B* and *GSKIP* (**Supplementary Table 2**) were vectorized either in PLKO.1-puromycin-Ubc-turboGFP (Sigma-Aldrich) or PRRLsin-PGK-eGFP-WPRE for *GSKIP* shRNA or in PRRLsin-PGK-mCherry-WPRE for *ATG2B* shRNA (Généthon). Lentivirus particles were produced as previously described[62]. iPSC-derived $CD34^+$ cells or $CD34^+$ cells were transduced with lentivirus encoding shRNA for *GSKIP* or *ATG2B* and sorted for GFP or mCherry, respectively, on the BD Influx sorter. Alternatively, we used selection with puromycin.

Statistics. Data are presented as means (±s.d. or s.e.m.). Statistical significance was determined by Student's *t* test or Mann-Whitney Wilcoxon test. $P < 0.05$ was considered statistically significant. A Bonferroni test was performed in instances involving testing of multiple parameters. All the tests were two-tailed and unpaired.

57. Tefferi, A. *et al.* Proposals and rationale for revision of the World Health Organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: recommendations from an ad hoc international expert panel. *Blood* **110**, 1092–1097 (2007).
58. Malak, S., Labopin, M., Saint-Martin, C., Bellanne-Chantelot, C. & Najman, A. Long term follow up of 93 families with myeloproliferative neoplasms: life expectancy and implications of $JAK2^{V617F}$ in the occurrence of complications. *Blood Cells Mol. Dis.* **49**, 170–176 (2012).
59. Mali, P. *et al.* Improved efficiency and pace of generating induced pluripotent stem cells from human adult and fetal fibroblasts. *Stem Cells* **26**, 1998–2005 (2008).
60. Debili, N. *et al.* Characterization of a bipotent erythro-megakaryocytic progenitor in human bone marrow. *Blood* **88**, 1284–1296 (1996).
61. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).
62. Plo, I. *et al.* JAK2 stimulates homologous recombination and genetic instability: potential implication in the heterogeneity of myeloproliferative disorders. *Blood* **112**, 1402–1412 (2008).