UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN
SCIENZE BIOMEDICHE

CICLO XXX

COORDINATORE Prof. Dello Sbarba Persio

Somatic mutation rate and cancer: an exploratory study of mutant frequency in classical Philadelphia-negative myeloproliferative neoplasms

Settore Scientifico Disciplinare MED/04

Dottorando
Dott.ssa Nannelli Caterina

Tutore
Prof. Dello Sbarba Persio

Coordinatore
Prof. Dello Sbarba Persio

Anni 2014/2018

CONTENTS

ABSTRACT

Oncogenesis is tightly related to the occurrence of somatic mutations. Somatic mutations are constantly produced by DNA replication processes, as the activity of DNA polymerases and DNA repair systems is highly efficient, but not perfect. The great part of spontaneous somatic mutations is irrelevant to cancer, but a very small fraction of them is oncogenic. Therefore, the individual predisposition to develop somatic mutations may be considered as a risk factor for tumor development.

Mutation rate (μ) and mutant frequency (f) are the main parameters used to quantify the occurrence of somatic mutations. The flow cytometry *PIGA* mutant assay is one of the available methods to measure these two parameters and is based on the use of the sentinel gene *PIGA*, which is involved in the first step of GPI-anchor biosynthesis.

Philadelphia-negative classical Myeloproliferative Neoplasms (MPNs) are chronic hematopoietic stem cell disorders mainly associated with a constitutive activation of the JAK/STAT pathway. MPN patients have an increased risk to progress into acute myeloid leukemia and to develop other (second) primary tumors.

In order to investigate whether there is any association between the rate of occurrence of somatic mutations and the individual risk of developing cancer, we used the flow cytometry *PIGA* mutant assay to measure the parameter f in peripheral blood granulocytes of 177 healthy subjects and of a cohort of 195 patients affected by Philadelphia-negative classical MPN.

Our results suggest that higher f values are strongly associated with an increased risk to develop one, or even more primary tumors. As regarding MPN, our findings indicate that an increased risk to develop a clinically more severe form of myeloproliferative disease is associated with higher f

values. Finally, we did not find evidence of a general genomic instability associated with deregulation of *JAK2*.

INTRODUCTION

Mutations and cancer

Cancer is a highly complex and heterogeneous disease. However, a common pathogenesis is thought to be at the basis of all type of cancers, which is the result of the interaction of two processes: progressive acquisition of somatic genetic (and epigenetic) alterations in individual cells and natural selection that may favor specific mutated cells with improved capabilities to survive and proliferate compared with their neighbors. Adult multicellular organisms themselves can be described as kind of genetic mosaics, because they are constituted by thousands of clones originated from individual cells that acquired a set of genetic differences from their progenitors and that passed the natural selection process. Most of them have limited proliferation potential and do not develop malignant features, but when a single cell acquires a sufficient number of advantageous mutations that allow uncontrolled growth and invasion of tissues, then cancer initiate (Stratton et al., 2009).

Cells that eventually will become malignant progressively accumulate additional somatic mutations that may or may not contribute to cancer development. Somatic mutations are classified as “driver” and “passenger” mutations. Driver mutations confer a growth advantage to malignant cells and are positively selected at some point during cancer development; passenger mutations are simply inherited from cancer cell precursors, even before the acquisition of driver mutations, and they do not play any specific function in oncogenesis. However, passenger mutations may become driver mutations at some point, thus conferring growth advantages to mutated cells, if changes in the tumor microenvironment occur (Luzzatto and Pandolfi, 2015, Stratton et al., 2009).

The number of driver mutations required to develop different types of cancer is not yet completely understood, but it is thought to be 3 to 6 in most cases (Luzzatto and Pandolfi, 2015). On the other end, it is widely

accepted that passenger mutations constitute the large majority of mutations found in cancer genomes (Stratton et al., 2009, Tomasetti et al., 2013).

Mutation rate and mutant frequency

To quantify occurrence of mutations in any cell population, two parameters are commonly used: mutation rate (μ) and mutant frequency (f). Mutation rate was firstly defined in Bacteria as "a fixed small chance *per time unit*" for each cell to undergo a mutation, where time is usually measured in "*units of division cycles"* of the cell (Luria and Delbrück, 1943). In modern terms, mutation rate is generally defined as the probability of a new mutation occurring in a specific gene per cell division. Mutant frequency, instead, represents the fraction of cells in a population harboring a mutation in a specific gene (Green et al., 1995).
Provided that mutations do not affect cell growth, that both the wild type and the mutated populations grow exponentially, and that a large cell population is studied, it can be applied the simple formula

$$f = \mu \times d$$

where d is the number of cell divisions occurred to produce the entire cell population (Araten et al., 2005, Araten et al., 2013, Peruzzi et al., 2010).

The cancer risk

Endogenous metabolic products (such as reactive oxygen species) and exogenous mutagens continuously damage DNA in normal cells. As expected, high and continuous exposure to exogenous mutagens, which are known in some cases (e.g. cigarette smoking, radiation, oncogenic viruses...), increases the rate of acquired somatic mutations and, as a consequence, raise the risk to develop specific tumors, which often show the mutational signature typically associated with a specific mutagen (Stratton et al., 2009).

However, the great part of DNA mutations in pre-malignant cells is thought to simply occur by chance during cell duplication, due to the intrinsic error rate associated with the DNA replication process itself that leads to misincorporation of wrong nucleotides during DNA synthesis or while repairing damaged DNA (Luzzatto and Pandolfi, 2015).

For instance, Polδ and ε, which are the main enzymes responsible for DNA replication, have error rates ranging between ~1 and 2 per 10^5 nucleotides synthesized, whereas Pol-α makes even more mistakes. However, Polδ and ε have a “proof reading” capacity that allows them to exert exonuclase activity by degrading the newly synthesized DNA strand from its 3’-OH end and then change misincorporated nucleotides, thus reducing error rates by 4- to 24-fold (DePamphilis, 2016). Nevertheless, it has been postulated that even in the presence of effective DNA repair systems which correct single mistakes made during DNA synthesis, DNA duplication itself may be sufficient to determine genomic instability (DePamphilis, 2016).

Other mechanisms contribute to the accumulation of somatic mutations, even in the absence of DNA replication. For instance, spontaneously oxidative deamination of methylated cytosines at CpG sites takes place in a replication-independent manner and increases with time. However, the contribution of this mechanism, even if consistent, is limited to specific CpG-enriched DNA regions that result hypermutated compared with other genomic traits (Lynch, 2010).

From a biological point of view, the fact that accumulation of mistakes has never been completely suppressed by natural selection demonstrates its fundamental role to drive evolution. On the other hand, this is tightly associated with individual disease development (DePamphilis, 2016).

In ~5-10% of malignancies, a hereditable genetic predisposition to cancer development is known (Tomasetti and Vogelstein, 2015). Some of them are related to defects in the DNA repair systems. For instance, it has been shown that cells from patients affected by inherited cancer-prone syndromes with impaired DNA repair, such as Fanconi anemia and ataxia

telangiectasia, have an increased predisposition to develop mutations (Araten et al., 2005). Moreover, colorectal, endometrial and gastric cancers with mutations in gene of the Mismatch Repair (MMR) system show a “mutator phenotype” because of the higher acquisition rate of genetic mutations caused by the impaired ability of cells to correct DNA errors (Lengauer et al., 1998).

However, whether a mutator phenotype is always at the basis of cancer development is still controversial (Luzzatto and Pandolfi, 2015, Stratton et al., 2009). Several studies demonstrated that hypermutability is not a universal feature of cancer cells (Araten et al., 2010, Elmore et al., 1983, Lengauer et al., 1998). Actually, the rate at which mutations occur in normal cells may be sufficient to promote, by chance, the accumulation of a specific set of oncogenic genetic (and epigenetic) alterations in an individual cell, thus determining clonal expansion and transformation (Luzzatto and Pandolfi, 2015). Then, during cancer progression there may be a sudden acceleration in the accumulation of somatic mutations in transformed cells, probably favored by the inhibition of restriction mechanisms and by the activation of mutagenic-prone processes (Mutter-Rottmayer et al., 2016), which would explain the high mutational complexity and heterogeneity of cancer genomes (Lengauer et al., 1998, Stratton et al., 2009).

Provided that the first source of genetic errors is DNA replication, the more a cell replicates during its lifetime, the higher is the risk to acquire potentially dangerous mutations and to become malignant. This is in perfect accordance with the observation that the lifetime risk to develop specific type of cancer is proportional to the number of stem-cells with self-renewal capacity found in the tissue (Tomasetti and Vogelstein, 2015) and that it dramatically increases with age (DePinho, 2000).

However, the individual risk to develop cancer is highly variable and it may depend on a combination of the intrinsic error rate of DNA replication processes, which is represented by the parameter μ and is genetically determined, and external factors (Luzzatto and Pandolfi, 2015).

If the number of independent driver mutations in various genes, required to transform a normal cell of a specific tissue into a malignant cell, is indicated with n, and the frequency of cells of that tissue harboring a mutation in a specific gene 1, 2, 3… is indicated with f_1, f_2, f_3…, respectively, then the probability (P) that all n driver mutations occur by chance in the same cell can be calculated as:

$$\mathbf{P = f_1 \times f_2 \times f_3 \ldots f_n}$$

Assuming that μ is constant, then

$$\mathbf{f_n = \mu_n \times d}$$

where μ_n is the rate of new mutations in n specific genes and d the total number of cell divisions occurred.
By substitution,

$$\mathbf{P = (\mu_1 \times d) \times (\mu_2 \times d) \times (\mu_3 \times d) \ldots (\mu_n \times d).}$$

If the geometric mean of μ_1, μ_2, $\mu_3 \ldots \mu_n$ is indicated with $\underline{\mu}$, then

$$\mathbf{P = (\underline{\mu} \times d)^n}$$

Since n is expected to be higher than 1 in most cases (Luzzatto and Pandolfi, 2015), a greater than linear relationship between development of cancer and both number of cell divisions and μ is predicted. As a consequence, even small changes of μ would have big effects on the probability that a normal cell becomes malignant (Araten et al., 2005, Araten et al., 2013). This renders the parameter μ a very interesting one and many efforts have been made during the past years to find ways to measure it in human cells.

The measurement of mutation rate and mutant frequency

The first attempts in finding a way to estimate mutation rate come from the past, when studies on bacterial cultures were performed (Luria and Delbrück, 1943, Novick and Szilard, 1950, Ryan and Wainwright, 1954): they paved the way to the elaboration of several mathematical models for

estimating mutation rates not only in bacterial cultures, but also in yeast and animal cell cultures (Crump and Hoel, 1974, Kendal and Frost, 1988, Magni and Von Borstel, 1962, Stewart et al., 1990, Wu et al., 2009, Zheng, 1999). However, quantification of mutation rates in animal cells has been proven more difficult for many reasons. For instance, elimination of pre-existing mutant cells is not easy and extensive counting and cell culture work is required to identify new mutants. Moreover, the number of cell divisions and the cell culture density are much lower than those of bacteria, therefore the application of methods derived from observations in bacterial cultures has been questioned. Finally, animal cells, at variance with bacteria, are diploid (Araten et al., 2013, Kendal and Frost, 1988, Peruzzi et al., 2010). For these reasons, the great part of tests applied in mammalian systems allows only the measurement of mutant frequency. The “spot-test” was the first method developed to quantify somatic cell gene mutations *in vivo* in mammals (Russell and Major, 1957). As for human cells, the first promising assays were subsequently discarded because it was not possible to demonstrate that the phenotypic changes were due to gene mutations (Albertini et al., 1990).

Eventually, a few assays for measuring f and in some cases also μ have been developed. They are based on the use of “sentinel genes” that are easily identified when mutated, such as hemoglobin genes, glycophorin A (*GPA*), *XK*, *HPRT*, T cell receptors, *HLA*, ouabain resistance and *PIGA*.

Hemoglobin genes

Hemoglobin genes are organized in two clusters, α and β, of strictly linked genes mapped on chromosome 16p and chromosome 11p, respectively. Hemoglobin is a heterotetrameric protein very abundant in red blood cells which is made up by two subunits encoded by one of the member of the α-cluster and two subunits encoded by one of the member of the β-cluster. The main human postnatal hemoglobin is hemoglobin A (HbA), which is made of two α and two β chains. Many mutation are known in hemoglobin genes. One of these is a mutation in the β gene which causes a A→T

substitution that results in the replacement of a valine with a glutamine at position 6. This mutation produces the HbS found in congenital sickle-cell anemia. The use of fluorescent anti-HbS polyclonal antibodies on fixed red cells has been used to measure the frequency of rare red blood cells containing the HbS variant because of a somatic mutation (Albertini et al., 1990, Tates et al., 1989).

GPA gene

The *GPA* gene is on chromosome 4q and encodes a glycosylated protein that carries the M and N blood group antigens on the cell surface of red blood cells. Approximately 10^5 molecules of GPA are exposed on the surface of each red blood cell. M and N are the only two alleles of the *GPA* locus and the related proteins, which differ by two non-adjacent amino acids, are codominantly expressed. In M/N heterozygous individuals, specific anti-M and anti-N antibodies labeled with two different fluorochromes can be used to detect the rare mutant erythrocytes that are stained with only one fluorescence, in contrast to wild-type M/N heterozygous cells with double fluorescence. Therefore, the *GPA* assay allows to measure the frequency of hemizygous cells, where either M or N antigen expression is lost (M0, N0), and homozygous cells, with loss of expression of one antigen and double expression of the other (MM, NN) as a result of gene conversion or somatic crossing-over (Akiyama et al., 1995, Albertini et al., 1990, Kyoizumi et al., 1989, Langlois et al., 1993).

XK gene

The *XK* gene is located on chromosome Xp and only one copy is active in both males and females. Inherited mutations in the *XK* gene cause the McLeod syndrome (De Franceschi et al., 2014).

The XK protein is a big transmembrane protein which spans ten times across the membrane of different cell types, including red blood cells. Its function is still not known, but it has structural characteristics of membrane

transporter proteins. On red blood cells, XK proteins are covalently linked to Kell glycoproteins by a disulfide bond, which is required for the expression of Kell antigens on the cell surface. Inactivating mutations that lead to the absence of the XK protein on red blood cells are associated with a reduced expression of the Kell blood group antigen and with mild compensated hemolysis (Araten et al., 2009, De Franceschi et al., 2014). The frequency of red blood cells lacking the expression of XK proteins, which do not seem to be subjected to premature removal from the blood, can be easily detected by cytofluorimetric techniques with specific antibodies against a non-polymorphic Kell protein determinant (Araten et al., 2009).

The three methods described above are performed in red blood cells, which are terminally differentiated cells that during the maturation process lose their nucleus. Therefore, the mutations that are detected may have occurred in both the stem-cell compartment or in committed red cell precursors and further molecular analyses of mutants cannot be performed (Albertini et al., 1990, Green et al., 1995).

HPRT gene

The *HPRT* gene is on chromosome Xq. Hereditary mutations in the *HPRT* gene are associated with the Lesch-Nyhan syndrome (Bell et al., 2016). This gene encodes the hypoxanthine-guanine phosphoribosyltransferase (HPRT) enzyme, which is constitutively and ubiquitously expressed, but is not necessary in mammalian cells. The HPRT enzyme is involved in the purine salvage pathway; since it is able to phosphoribosylate and to activate, besides its normal substrates, also cytotoxic purine analogues, only the cells where it is not expressed are resistant in the presence of these compounds. Only one copy of the *HPRT* gene is active in both male and female cells, therefore a single inactivating mutation is required to originate the mutant phenotype (Albertini et al., 1990).

The *HPRT* mutant assays have been developed from studies on mutant hamster cell lines obtained in selective medium (Littlefield, 1969). Most of them take advantage of the cytotoxicity of the purine analogue 6-thioguanine (TG) on cells expressing the HPRT enzyme, whereas mutant cells lacking its activity are TG resistant (TG^r). In one assay variant, the incorporation of ^{3}H-thymidine is detected by autoradiography in resistant T-lymphocytes cells that survive TG cytotoxicity when DNA synthesis is stimulated *in vitro* by the mitogen phytohemagglutinin (PHA). In a modification of the assay, incorporation of 5-bromodeoxyuridine (BrdU) is followed by staining with the Hoechst 33258 dye to distinguish mutant cells by detection of fluorescence differences. These short-term variants of the assay allow to easily measure TG^r frequency (Albertini et al., 1990).

An important advantage of using lymphocytes, or other cell types (i.e. fibroblasts), compared with red blood cells, is that they can be grown in culture. Indeed, other variants of the *HPRT* mutant assay are based on cell cloning, which allows recovery and further analysis of resistant clones.

Human T-lymphocytes are an extremely heterogeneous population: each T-cell, when differentiates in the thymus, becomes able to recognizes a specific antigen by progressive rearrangements of the T-cell receptor (TCR) genes that lead to the expression of a unique TCR on the plasma membrane, which will be inherited only by the clonal progeny of that specific cell. Thus, analysis of TCRs allow to recognize the independent or common origin of mutant T-cells observed in an individual (Nicklas et al., 1986).

This specificity is conveniently used in another variant of the *HPRT* mutant assay. Limited dilutions of peripheral blood T-lymphocytes are prepared in microtiter dish wells with and without TG, in the presence of the T-cell growth factor interleukin-2, the mitogen PHA and X-irradiated *HPRT*$^-$ feeder cells. Resistant cells lacking functional HPRT enzyme can proliferate and form colonies in the presence of TG. Then, they can be isolated and propagated *in vitro* for further analyses, including *HPRT* molecular studies and determination of the TCR-gene rearrangement

patterns to establish whether two clones derive from the same T-lymphocyte, or result from two independent mutational events. This allows to calculate, and not simply estimate, the actual *HPRT* mutant frequency *in vitro*, which is assumed to be the same as *in vivo* since no cell division takes place before the limited dilution preparation (Albertini et al., 1990, Green et al., 1995).

In order to measure μ using the *HPRT* gene, different strategies have been applied: from the use of the HAT medium (which contains hypoxanthine, aminopterin and thymidine) to remove pre-existing mutant cells (Rossman et al., 1995), to the application of fluctuation analyses (Elmore et al., 1983, Seshadri et al., 1987) originally performed in bacteria cultures (Luria and Delbrück, 1943).

The *HPRT* mutant assays have been the first successful methods for the measurement of mutant frequency and somatic mutation rate in human cells *ex vivo* and they have been widely employed (Akiyama et al., 1995, Albertini et al., 1982, Green et al., 1995, Oller et al., 1989, Rossman et al., 1995, Seshadri et al., 1987, Tates et al., 1989).

HLA complex

The *HLA* complex, which encodes cell surface molecules important in immune responses, is on chromosome 6p and includes several linked loci containing two classes of genes. The class I loci comprises the *HLA-A*, *HLA-B*, *HLA-C* genes, which are almost ubiquitously expressed, and several other genes. Numerous genetic variants of *HLA-A*, *HLA-B*, and *HLA-C* genes have been identified and the alleles are codominantly expressed. Inactivating mutations in one allele that lead to loss of one HLA antigen can be easily detected in heterozygous individuals by using a combination of specific anti-HLA monoclonal antibodies and complement factors. Since the *HLA* genes are autosomal, also gene conversion or mitotic recombination can be studied with this system. (Albertini et al., 1990).

In the *HLA* T-lymphocyte assay, T-cells are plated in microtiter dishes and then immunoselected. The method allows to measure the frequency of *HLA*-mutants which survive the immunoselection, that can be further analyzed (Albertini et al., 1990, Grist et al., 1992, McCarron et al., 1989).

Ouabain resistance

Ouabain resistance (Oua^R) develops from alterations in the membrane ATPase involved in the Na^+/K^+ transport. The Na^+/K^+ ATPase plays a key role in establishing and maintaining the electrochemical gradients of Na and K ions across the cell membrane. This enzyme is made by two subunits: a large catalytic subunit (α) and a smaller glycoprotein subunit (β) (Morth et al., 2011). When the enzyme is mutated, its affinity for the steroid ouabain decreases and it is no longer blocked by it, resulting in Oua^R. The genes encoding the Na^+/K^+ ATPase are autosomal, but the Oua^R phenotype is codominantly expressed and Oua^R mutants can be selected in cell cultures in the presence of ouabain. However, the Na^+/K^+ ATPase activity is required for cell survival, therefore the types of mutations valuable by this method are limited (Buchwald, 1977, Elmore and Barrett, 1982, Elmore et al., 1983).

PIGA gene

PIGA is the phosphatidylinositol glycan anchor biosynthesis class A gene and is located on chromosome Xp (Miyata et al., 1993).

The *PIGA* gene encodes the subunit of one of the numerous proteins required for the biosynthesis of glycosylphosphatidylinositol (GPI) anchor, which is ubiquitously expressed on cell surfaces. Biosynthesis of GPI anchors is a complex process performed through 11 reaction steps in the endoplasmic reticulum; once synthesized, GPI anchors are transferred onto precursor GPI-anchored proteins, which have a GPI-attachment signal peptide at their carboxyl terminus that is then cleaved (Figure 1). It is interestingly to notice that *PIGA* participates to the first step of this

process and, among all genes involved, it is the only one located on chromosome X (Kinoshita and Fujita, 2016, Kinoshita et al., 2008, Takeda et al., 1993).

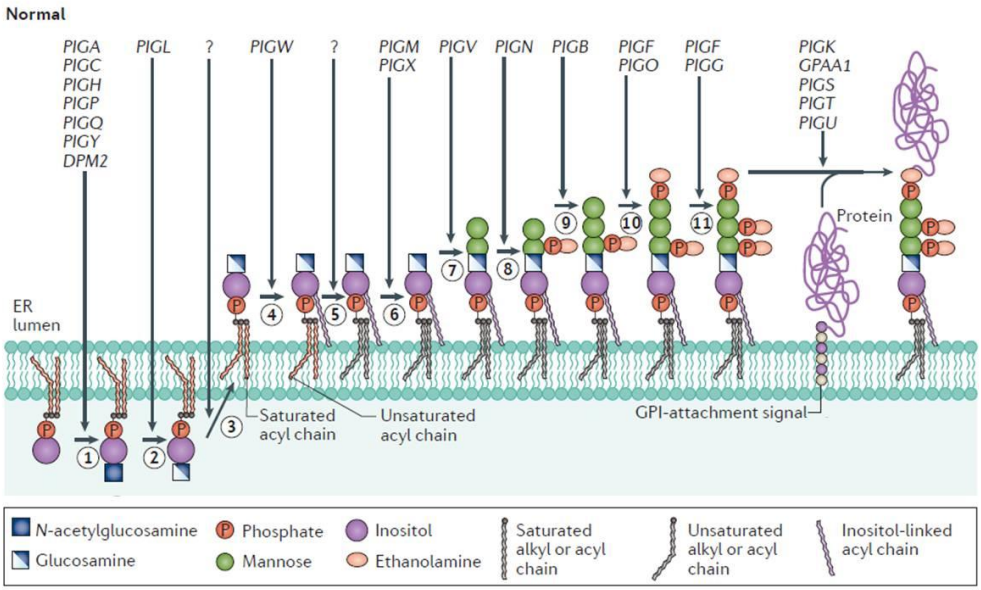

Figure 1. Glycosylphosphatidylinositol-anchor biosynthesis.
GPI molecules are synthesized through 11 reaction steps in the Endoplasmic Reticulum (ER). Once synthesized, they are linked to precursor GPI-anchored proteins that have a GPI-attachment signal peptide, which is then cleaved. Many genes participate to GPI-biosynthesis and some of them are still not known (as indicated by question marks). *PIGA* is involved in the first step of the GPI biosynthetic pathway (Hill et al., 2017).

After further remodelling reactions in both the endoplasmic reticulum and the Golgi apparatus, the GPI-protein complexes are transported via secretory vescicles to the plasma membrane. Once on the plasma membrane, GPI-linked proteins are mainly localized at lipid rafts, which are dynamic microdomains enriched in glycosphingolipids and cholesterol (Kinoshita and Fujita, 2016, Kinoshita et al., 2008). Numerous different GPI-anchored proteins are found on hematopoietic cells, such as blood group antigens, adhesion molecules, complement regulatory proteins, enzymes and receptors (Luzzatto and Notaro, 2003).

Acquired mutations in the *PIGA* gene are the main responsible for the hematological disease Paroxysmal Nocturnal Hemoglobinuria (PNH). PNH is a rare acquired disease affecting the multipotent Hematopoietic Stem Cell (HSC), which undergoes clonal expansion and produces an altered progeny that co-exists with variable fractions of normal hematological cells (Rosse and Dacie, 1966).

When somatic inactivating mutations occur in the *PIGA* gene, the first step of the GPI synthesis is defective and GPI-linked proteins cannot be properly tethered on the cell surface of affected cells and are degraded intracellularly, probably by proteasome (Hill et al., 2017). Since the *PIGA* gene is on chromosome X (Miyata et al., 1993) and only one allele is active, a single mutation is sufficient to originate the mutated phenotype.

Several *PIGA* mutations have been described in PNH patients, the majority of which are small insertions or deletions (indels) involving a single base or several bases, and single base substitutions, while large deletions are rare. Many mutations are frameshift and could result in early termination of transcription (Nafa et al., 1998).

Interestingly, rare GPI-defective cells have been detected also in many healthy individuals, with the same type of *PIGA* mutations found in PNH patients. Moreover, GPI-defective cells could persist for months in the blood of the healthy subjects, thus proving that they originated from hematological progenitors (Araten et al., 1999). Studies on PNH found no evidence of an intrinsic growth advantage of *PIGA*-mutated cells over non-mutated cells (Araten et al., 2002, Luzzatto, 2016, Rosti et al., 1997) and proved that GPI-deficient cells are not affected by hypermutability (Araten and Luzzatto, 2006, Araten et al., 2014).

The use of *PIGA* as a sentinel gene

As mentioned above, studies on PNH led to the discovery that rare *PIGA* mutated cells are present in healthy individuals (Araten et al., 1999).
From such studies, *PIGA* appeared to be an appealing candidate as sentinel gene for detecting somatic mutation occurrence (Araten et al., 1999) because it fulfilled important requirements.

- *PIGA* is ubiquitously expressed and is located on chromosome X (Miyata et al., 1993), therefore a single inactivating mutation is sufficient to originate the mutated phenotype.
- Moreover, it is involved in GPI-anchor biosynthesis (Miyata et al., 1993) and numerous proteins are GPI-linked to the plasma membrane of every cell type (Luzzatto and Notaro, 2003): this is very convenient for easy detection of mutant cells.
- Finally, several lines of evidence suggest that there is no strong selection in favor or against *PIGA* mutated cells (Araten et al., 2002, Rosti et al., 1997). The only exceptions are the *PIGA*-mutated red blood cells which are susceptible to complement-mediated lysis (Luzzatto and Notaro, 2003).

The *PIGA* mutant assay, firstly developed in 1999 (Araten et al., 1999), is a flow cytometry method based on the use of a combination of antibodies specific for GPI-linked proteins (or the GPI molecule itself), which are not able to stain the GPI-negative cells. This method allows to detect very rare cells that harbor inactivating *PIGA* mutations and are, as a consequence, GPI-deficient (Figure 2) (Peruzzi et al., 2010).

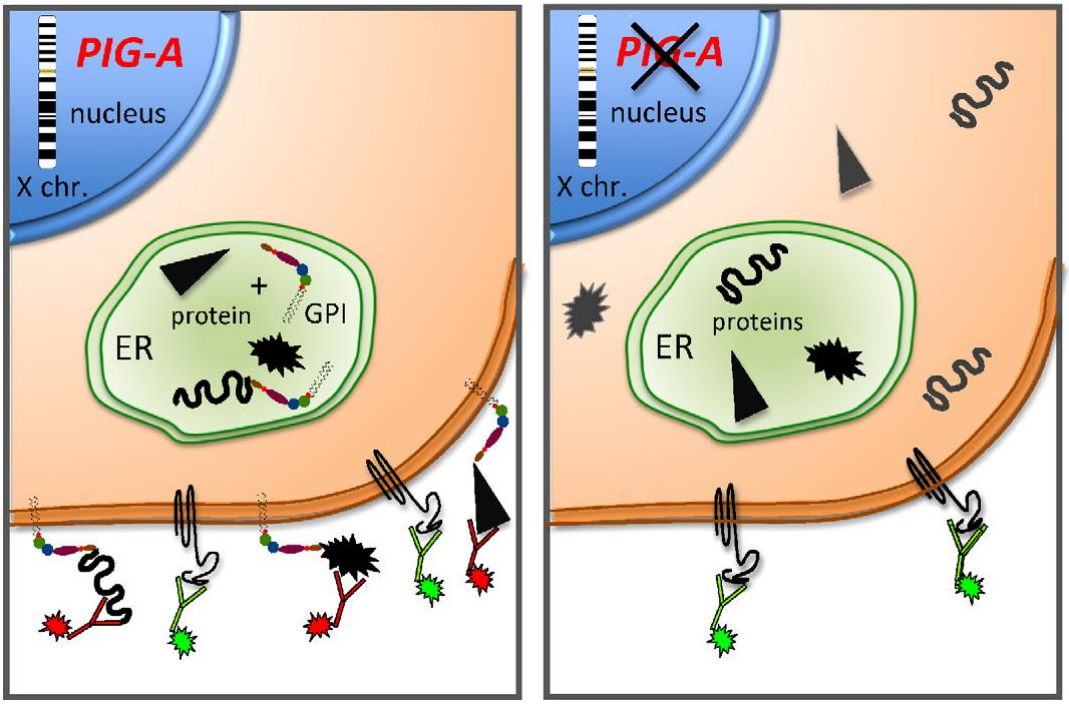

Figure 2. The flow cytometry *PIGA* mutant assay.
In normal cells (*left panel*) numerous proteins are synthesized in the endoplasmic reticulum and then exposed on the cell surface through GPI molecules, together with transmembrane proteins. If both types of cell membrane proteins are stained with specific antibodies linked to different fluorochromes, they can be easily detected by flow cytometry. In cells harboring *PIGA* inactivating mutations (*right panel*), the GPI anchor cannot be correctly synthesized and GPI-linked proteins are retained in the cytoplasm or degraded. However, transmembrane proteins are correctly exposed on the cell surface. Therefore, *PIGA* mutant cells can be easily identified by flow cytometry because they are still recognized by antibodies specific for transmembrane proteins, but not by those specific for GPI-linked proteins (Peruzzi et al., 2010).

Many research groups have employed the *PIGA* mutant assay to study mutant frequency (f) in different types of human blood cells and cell lines (Olsen et al., 2017). In addition, since the biosynthetic pathway of GPI anchors is highly conserved in mammalian cells (Kinoshita et al., 2008), *PIGA* has been extensively used in genotoxicity studies in several animal species, such as rats, mice and even monkeys (Dobrovolsky et al., 2010, Olsen et al., 2017, Phonethepswath et al., 2008).

The *PIGA* mutant assay has been employed also to measure mutation rate (μ) in human cell cultures. To remove pre-existing mutant GPI-

negative cells from cell cultures, GPI-positive cells have been initially selected either by flow sorting (Araten et al., 2005, Araten et al., 2010, Araten et al., 2013), or by means of cell dishes coated with antibodies against GPI-linked proteins (Krüger et al., 2015) and then re-cultured.

As regarding blood cells, f has been investigated in reticulocytes, red blood cells, lymphocytes and granulocyets (Olsen et al., 2017).

Red blood cells are non-dividing cells which are the most abundant in blood and can be easily obtained as few millions are contained in 1 microliter of blood. These features should make the use of red blood cells for the *PIGA* mutant assay very convenient. However, *PIGA*-mutant red blood cells are exquisitely sensitive to lysis by activated-complement and complement activation may be extremely variable throughout time in the same individual: for these reasons, the measurement of f in red blood cells is not reliable (Peruzzi et al., 2010). Despite of these very critical issues, red blood cells are widely employed to study f in animals (Dobrovolsky et al., 2010, Olsen et al., 2017, Phonethepswath et al., 2008) and an assay variant has been adapted to human red blood cells as well (Dertinger et al., 2015).

The parameter f has been studied also in lymphocytes (Olsen et al., 2017). However, the great heterogeneity of blood lymphocyte populations poses limitations to their use in this type of studies because differences in life-span and proliferation potential may affect the measurement of f by founder effects and bottleneck effects (Peruzzi et al., 2010).

Granulocytes are terminally differentiated myeloid cells which cannot proliferate and do not seem to undergo negative selection in the absence of GPI-linked proteins, at least in humans (Yamamoto et al., 2002). Thus, they appear to be a better cell system than others to measure f (but not μ, because they are non-dividing cells), even if a greater amount of blood is needed and their handling is more laborious (Peruzzi et al., 2010).

As for other sentinel genes, the *PIGA* mutant assay allows to detect only genetic mutations (or possibly epigenetic alterations) which cause a reduced or absent activity of the PIGA protein, whereas synonymous

mutations, whole gene translocations, inversions, or duplications are not identified. This inevitably leads to underestimate the absolute number of mutants. However, it may have little relevance in the case of comparative studies where, for instance, the effect of specific mutagens has to be assessed, because a general increase of the number of mutants over the background should be quite precisely measurable (Peruzzi et al., 2010).
Another limitation, as in other assays, is that when a one-time analysis of f is performed *in vivo*, the detection of mature GPI-deficient cells does not give indications about the hierarchical stage at which the *PIGA* mutation has occurred, which greatly affects the final number of mutants. Only when repeated analyses can be performed after periods of time, thus following the kinetics of mutant cell numbers or even sequencing *PIGA* to detect the specific mutation, it may be possible (in theory) to distinguish between mutations that occurred in stem cells or downstream in the differentiation process (Araten et al., 1999). Moreover, the maturation level of the first cell where the mutation occurred may alter the time of appearance of mature GPI-deficient cells *in vivo* (Peruzzi et al., 2010). However, when the frequency of GPI-deficient granulocytes is measured in the blood, it should represent the sum of both early and late mutational events; therefore, it seems reasonable to consider, by approximation, that a balance may take place (Rondelli et al., 2013).

Despite the advantages of *PIGA*, compared to other sentinel genes, the measurement of μ in humans is a still laborious process because it requires to obtain a cell line (i.e. B-lymphoblastoid cell line) from each subject of interest, which need to be grown and then be purged from preexisting GPI-negative cells before the analysis of μ could take place (Peruzzi et al., 2010). On the contrary, f is more readily measured with the *PIGA* mutant assay.

In order to verify whether the parameter f may be used as a surrogate of μ, the correlation between the frequency of mutant peripheral blood granulocytes and rate of mutations in B-lymphoblastoid cell lines of 31 subjects has been tested in our laboratory. As reported in Figure 3, a

significant correlation between the two parameters was found (Pearson's r = 0.52, p = 0.0026, Notaro et al., unpublished results), thus confirming the possibility to conveniently investigate f in place of μ.

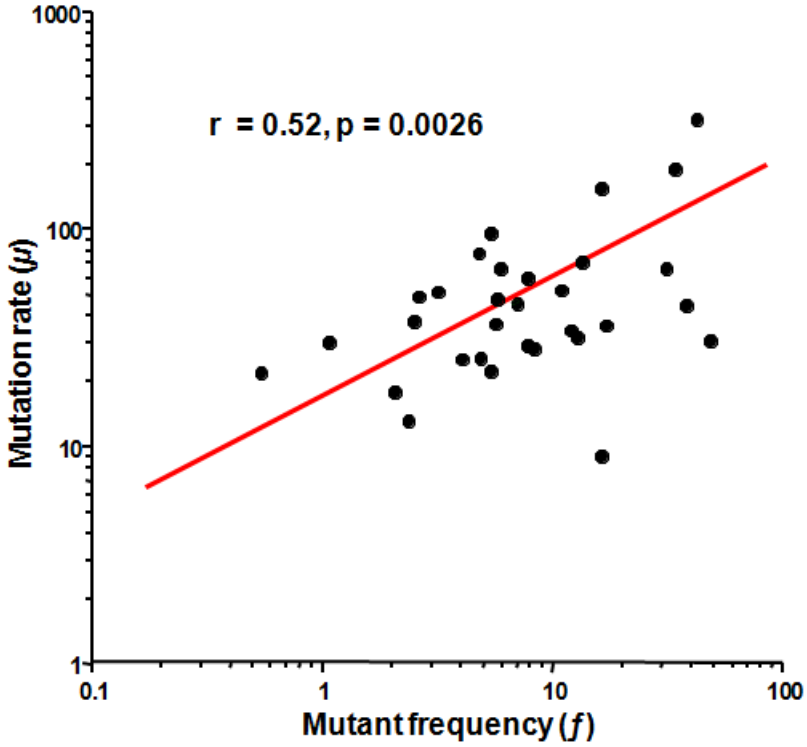

Figure 3. Correlation between f and μ.

The distribution of the parameter f was then studied in 142 healthy subjects, where f values were found to range from less than 1 to 37.5 x 10^{-6}. Moreover, log(f) was proven to have a normal distribution, as other quantitative traits (Rondelli et al., 2013).

In the interest of exploring whether the individual μ, estimated through the measurement of f, is effectively related with a predisposition to develop cancer, f was measured also in 49 patients with oral squamous cell carcinoma diagnosed at the Outpatient Service of the "Head and Neck Surgery" Clinic of our Institution. The median value of f distribution in cancer patients resulted similar to that found in the healthy subjects, but an excess of cancer patients with high f values was observed. Indeed, the proportion of patients with f values greater than the value that defined the

90^{th} percentile in the healthy population, resulted significantly higher (20%) compared to the healthy subjects (Notaro et al., unpublished results).

Philadelphia-negative classical Myeloproliferative Neoplasms

Classification of myeloid neoplasms

Myeloid neoplasms are divided into Acute Myeloid Leukemia (AML) and chronic myeloid neoplasms on the basis of the percentage of peripheral blood or bone marrow blasts (Barbui et al., 2018b).

Philadelphia-negative classical Myeloproliferative Neoplasms (MPN), also known as *JAK2-CALR-MPL* mutation-related MPNs, are chronic myeloid neoplasms and include Polycythemia Vera (PV), Essential Thrombocythemia (ET) and Primary Myelofibrosis (PMF) (Barbui et al., 2018b). Two stages of PMF are recognized: a prefibrotic/early stage, called pre-PMF, and an overt fibrotic stage (Rumi and Cazzola, 2017, Zimran et al., 2018) (Box 1).

Some ET and especially PV cases may progress into secondary myelofibrosis, sharing similar clinical features and management with PMF (Rumi and Cazzola, 2017, Zimran et al., 2018). In the present text, they have been indicated as Post Essential Trombocythemia - Myelofibrosis (PET-MF) and Post Polycythemia Vera - Myelofibrosis (PPV-MF), respectively.

Box 1. The 2016 WHO classification of chronic myeloid neoplasms

- **Myeloproliferative Neoplasms (MPN)**
  - Chronic Myeloid Leukemia (CML), BCR-ABL1^{+}
  - Polycythemia Vera (PV)
  - Essential Thrombocythemia (ET)
  - Primary Myelofibrosis (PMF)
    - PMF, prefibrotic/early stage (pre-PMF)
    - PMF, overt fibrotic stage
  - Chronic Neutrophilic Leukemia (CNL)
  - Chronic Eosinophilic Leukemia, Not Otherwise Specified (CEL-NOS)
  - MPN, Unclassifiable (MPN-U)
- **Myelodisplastic Syndromes (MDS)**
- **Myelodisplastic/Myeloproliferative Neoplasms (MDS/MPN)**
- **Myeloid/lymphoid neoplasms with eosinophilia and rearrangement of *PDGFRA*, *PDGFRB*, or *FGR1*, or with *PCM1-JAK2***
- **Mastocytosis**
- **Myeloid neoplasms with germline predisposition**

The Philadelphia-negative classical Myeloproliferative Neoplasms

Philadelphia-negative classical MPNs are chronic hematopoietic stem cell disorders characterized by altered count of terminally differentiated cells of one or more cell lineages in the peripheral blood and bone marrow alterations (Rumi and Cazzola, 2017, Zimran et al., 2018). Morphogenic dysplasia, including dyserythropoiesis, dysgranulopoiesis and monocytosis, is usually absent (Barbui et al., 2018b).

Splenomegaly, microcirculatory symptoms, bleeding, venous and arterial thromboembolic events are common. Moreover, Philadelphia-negative classical MPNs are often associated with systemic inflammation, which lead to fatigue and night sweats (Zimran et al., 2018).

The peak of incidence of Philadelphia-negative classical MPNs is around 60-70 years of age, but also much younger people are affected (Zimran et al., 2018).

All MPNs may progress into a blast phase indistinguishable from AML, which is sometimes preceded by a myelodysplastic phase (Rumi and Cazzola, 2017). Moreover, they have been associated with an overall increased risk of developing second primary tumors, besides AML (Chattopadhyay et al., 2018, Landtblom et al., 2018).

The identification of molecular alterations in Philadelphia-negative classical MPNs is becoming increasingly important, but each disease entity is still primarily defined on the basis of clinical and histopathological features (Rumi and Cazzola, 2017).

Polycythemia Vera

PV often presents as a panmyelocytosis, with increased white blood cell count, hemoglobin, red blood cell mass and platelets (Meier and Burton, 2014). Bone marrow is characterized by hypercellularity, with the typical pattern of panmyelocytosis (Rumi and Cazzola, 2017).

The incidence of PV has been estimated around 0.4-2.8 per 100000 persons per year, with median age of insurgence of 65 years (Rumi and Cazzola, 2017).

The most common clinical features of PV are headache, lightheadedness, weakness, pruritus often exacerbated by bathing and skin irritation, dizziness, and diaphoresis, which are non-specific symptoms that are detected also in many other diseases. Vascular complications, such as erythromelalgia which could progress to ulcerative lesions, transient ischemic attacks, myocardial infarctions or pulmonary emboli, are often the first events detected in PV patients. It has been estimated that arterial and venous thrombotic events, also in unusual locations, at the time of initial diagnosis range from 34% to 39%; the risk of thrombotic complications increases with age older than 60 years and previous history

of thrombosis (Meier and Burton, 2014, Rumi and Cazzola, 2017). Splenomegaly is found in up to 70% of PV patients (Meier and Burton, 2014). Thrombotic complications are the main cause of morbidity and death of PV patients, followed by hematologic transformation mainly into myelofibrosis, but also into a blast phase similar to AML (Figure 4), and development of solid tumors (Meier and Burton, 2014, Rumi and Cazzola, 2017). Duration of disease and age are considered risk factors for the transition to myelofibrosis, while age older than 70 years, leucocytosis and abnormal karyotype are risk factors for developing leukemia (Meier and Burton, 2014, Tefferi and Barbui, 2017).

Essential Thrombocythemia

ET is characterized by an elevated platelet count, generally normal hemoglobin levels and usually normal or slightly increased white blood cell count (Meier and Burton, 2014). Bone marrow cellularity is normal or even slightly reduced, but an increased number of enlarged and hyperlobulated megakaryocytes is found (Rumi and Cazzola, 2017).

The ET incidence is estimated to be 0.38-1.7 cases per 100000 people per year (Rumi and Cazzola, 2017). There is an increase of ET cases with age, with a peak between 50 and 70 years and median age of insurgence of 68 years; it affects more women than men (Meier and Burton, 2014, Rumi and Cazzola, 2017).

Most cases of ET are incidentally diagnosed during other laboratory exams because patients are often asymptomatic. However, the most common symptoms are headache, visual disturbance and dizziness. Vascular complications, such as easy bleeding and bruising, erythromelalgia, ocular migraines and transient ischemic attacks, are relatively common. Splenomegaly is found in ~50% of patients and is relatively mild. Thrombosis events also in unusual locations, more often arterial than venous, are more common than hemorrhage, but less frequent than in PV (Figure 4). Increased risk of thrombotic events is

associated with age older than 60 years, history of thrombosis, *JAK2* mutations and common cardiovascular risk factors (i.e. smoking habit, diabetes mellitus, arterial hypertension and hypercholesterolemia) (Barbui et al., 2018a, Meier and Burton, 2014, Rumi and Cazzola, 2017).

ET is the more benign of MPNs for survival, which is comparable to that of the sex- and age-standardized European population (Rumi and Cazzola, 2017). Moreover, the risk of transformation into myelofibrosis or leukemia is lower than in other MPNs (Figure 4), but it is associated with an extremely poor prognosis (Meier and Burton, 2014, Zimran et al., 2018).

Primary Myelofibrosis

PMF differs from PV and ET in that there is no proliferation of a specific cell line. Most frequently normocytic and normochromic anemia is detected. White blood cell count is usually normal, but it could be also markedly abnormal; platelets may be increased or decreased: thrombocytosis is frequently found in pre-PMF patients, while thrombocytopenia in advanced myelofibrosis has been associated with poor survival and progression to a blast phase (Meier and Burton, 2014, Rumi and Cazzola, 2017, Zimran et al., 2018). Pancytopenia is preferentially associated with disease progression (Meier and Burton, 2014). Bone marrow of pre-PMF is often characterized by increased cellularity, with atypical megakaryocytic and granulocytic proliferation; in overt PMF, besides megakaryocytic proliferation and atypia, reticulin and/or collagen fibrosis are typically detected (Rumi and Cazzola, 2017).

PMF is the less common of Philadelphia-negative classical MPNs. Its incidence ranges between 0.1 and 1.0 per 100000 people per year; it affects more frequently men older than 50 years and the median age of insurgence is 70 years (Meier and Burton, 2014, Rumi and Cazzola, 2017).

The main features of PMF are bone marrow fibrosis and extramedullary hematopoiesis, which determines many clinical findings such as

lymphadenopathy, peripheral edema, ascites and pulmonary edema. Up to 25% of patients are initially asymptomatic. Symptoms such as night sweats, fatigue and weight loss are more common than in other MPNs. Splenomegaly, often associated with decreased appetite and abdominal fullness, is frequent and most prominent in PMF. In two-thirds of patients hepatomegaly is found, often in association with splenomegaly. Only a small percentage (<10%) of PMF patients develop thrombotic complications, mainly venous thromboembolism (Figure 4) (Meier and Burton, 2014).

PMF is associated with a poorer prognosis than PV and ET, with a five-year survival of 40% and a median survival time of ~69 months. The main cause of death is transformation into leukemia (Figure 4) (Meier and Burton, 2014).

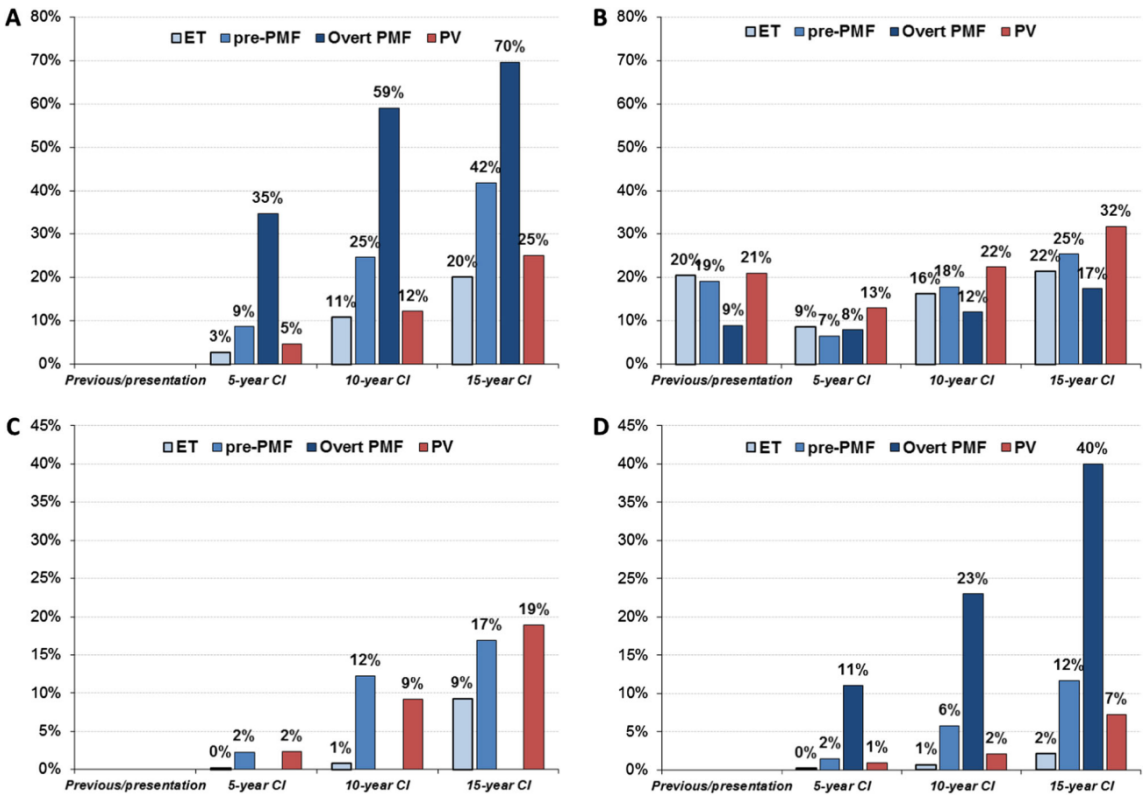

Figure 4. A) Mortality; B) Major arterial and venous thrombotic complications; C) Myelofibrosis; D) Blast transformation in ET, PV, pre-PMF and overt PMF.
Cumulative Incidence (CI) during follow-up is calculated at 5, 10 and 15 years from diagnosis (Barbui et al., 2018b).

Genetic abnormalities

The pathogenesis of Philadelphia-negative classical MPNs is mainly related to a constitutive activation of the JAK/STAT pathway by specific driver mutations, detected in the majority of patients (Zimran et al., 2018).

Mutations in the tyrosine kinase *JAK2*, which is on chromosome 9p and encodes the Janus kinase 2, are prevalent. In particular, the *JAK2*V617F in exon 14 is the most prevalent in Philadelphia-negative classical MPNs and is found in ~95% of PV and ~60% of ET and PMF cases (Zimran et al., 2018). The expansion of a dominant homozygous *JAK2*V617F subclone is almost exclusively found in PV patients and an allele burden >50% has been identified as a risk factor for PV progression to myelofibrosis (Rumi and Cazzola, 2017). The presence of the *JAK2*V617F mutation in ET is tightly associated with a higher risk of thrombosis and a lower risk of myelofibrosis, whereas the identification of the same mutation in pre-PMF has an unfavorable prognostic significance (Barbui et al., 2018b, Tefferi and Barbui, 2017). *Indels* in exon 12 of *JAK2* are detected in ~3-5% of PV patients lacking the *JAK2*V617F mutation, with a low mutation allelic burden; they are rare in ET or PMF. Other *JAK2* mutations have been found in some PV patients which do not harbor the canonical *JAK2* mutations (Barbui et al., 2018b, Tefferi and Barbui, 2017).

MPL (Myeloproliferative Leukemia virus oncogene), which is on chromosome 1p and encodes the thrombopoietin receptor, is mainly mutated at exon 10 at codon 515 (W>K, L, A or R) and more rarely at codon 505 (S>N) in ~4% of ET and ~8% of PMF cases, but almost not in PV (Barbui et al., 2018b, Tefferi and Barbui, 2017). Heterozygous *MPL*-mutant clones may become homozygous during disease progression (Rumi and Cazzola, 2017).

In 2013, mutations in *CALR*, which is on chromosome 19p and encodes a multi-functional Ca^{2+} binding protein chaperone mainly localized in the endoplasmic reticulum (calreticulin), have been described; they are found in ~25-35% of PMF patients and ~15-25% of ET patients (Barbui et al., 2018b, Tefferi and Barbui, 2017). *CALR* mutations are highly

heterogeneous *indels* in exon 9; the most prevalent are mutation type 1 (a 52-bp deletion, p.L367fs*46) and type 2 (a 5-bp insertion, p.K385fs*47), whereas the others (<20%) are defined as type 1-like and type 2-like. The type 1 *CALR* mutations are more frequently associated with PMF, while the type 2 *CALR* mutations are usually found in ET cases; *CALR* mutations are almost not detected in PV patients (Barbui et al., 2018b). *CALR*-mutant clones in MPNs are mainly heterozygous and usually undergo a progressive expansion to eventually become fully dominant during disease progression (Rumi and Cazzola, 2017). *CALR* mutations in ET are associated with a lower risk of thrombotic events, but a relative high risk of progression to myelofibrosis; *CALR* mutations in PMF have a better prognostic significance than *JAK2*V617F or *MPL* mutations, or especially the absence of mutations in any of the three driver genes (*JAK2*, *MPL* and *CALR*) (Rumi and Cazzola, 2017).

Driver mutations described above are mutually exclusive in almost all patients (Zimran et al., 2018). In the 2016 version of WHO classification, *JAK2*V617F and exon 12 mutations have been included as major diagnostic criteria for PV, whereas *JAK2*V617F, *CALR* and *MPL* mutations are indicated as major criteria for ET and PMF. However, no driver mutations are found in up to 20% of ET and 10-15% of PMF patients, which are categorized as “triple-negative”. Only a small percentage (up to 10%) of them harbor non-canonical mutations in *MPL* and *JAK2* (Barbui et al., 2018b, Zimran et al., 2018).

Additional mutations involved in Philadelphia-negative classical MPN pathogenesis have been described in *ASXL1*, *EZH2*, *TET2*, *IDH1/IDH2*, *SRSF2*, *SF3B1*, *DNMT3A*, *U2AF1*, *SH2B3/LNK*, *CBL*, *IKZF1*, *NFE2* and *TP53*, which participate to epigenetic regulation, splicing and signalling pathways (Barbui et al., 2018a, Zimran et al., 2018). They are often acquired during progression of disease, but they can also precede the occurrence of driver mutations. Mutations in *ASXL1*, *EZH2*, *IDH1/IDH2* and *SRSF2* are classified as “high-molecular risk mutations” (Zimran et al., 2018).

In triple-negative PMF cases, also investigation of other non-driver “most frequent” mutations (*ASXL1*, *EZH2*, *TET2*, *IDH1/IDH2*, *SRSF2*, *SF3B1*), which are marker of clonality when present, is recommended by WHO (Barbui et al., 2018b). The finding of two or more of these mutations and the absence of type 1 *CALR* mutations, which have a better survival significance, are considered as negative prognostic factors in myelofibrosis (Zimran et al., 2018). Mutations in *ASXL1*, *EZH2*, *IDH1/IDH2* and *SRSF2* have been associated with high risk of leukemic transformation or premature death in PMF; mutations in *ASXL1*, *SRSF2* and *IDH2* in PV and mutations in *SH2B3*, *SF3B1*, *U2AF1*, *TP53*, *IDH2* and *EZH2* in ET have been related with lower survival (Rumi and Cazzola, 2017, Tefferi and Barbui, 2017).

In triple-negative PMF cases, also chromosomal abnormalities can be searched as marker of clonality, although not indicated in the WHO document. The finding of additional mutations, which is common, is not currently included by WHO as criteria of clonality in triple-negative PV or ET (Barbui et al., 2018b).

Germ line MPN-predisposition alleles have been identified in the general population (Rumi and Cazzola, 2017). Genetic predisposition may explain the insurgence of familial cases of MPNs, but the molecular basis of genetic predisposition has been identified only in few families so far (Rumi and Cazzola, 2017).

AIM OF THE STUDY

The association between somatic mutations and cancer is, by now, a well established one. Since the acquisition of a specific set of driver mutations is required for the transformation of a normal into a malignant cell, the individual predisposition to develop somatic mutations may be considered as a risk factor for tumor development.

Mutation rate (μ) and mutant frequency (f) are the main parameters used to quantify the occurrence of somatic mutations. There are several methods to measure them, which take advantage of the use of “sentinel genes”. One of such methods is the flow cytometry *PIGA* mutant assay.

The measurement of μ is extremely laborious and does not lend itself for the analysis of large numbers of samples. However, in our laboratory it has been proven that the parameter f, measured in peripheral blood granulocytes, significantly correlates with μ measured in B-lymphoblastoid cell lines established from the same subjects. Thus, f can be conveniently used as surrogate of μ.

Philadelphia-negative classical Myeloproliferative Neoplasms (MPNs) are chronic hematopoietic stem cell disorders mainly associated with mutations in *JAK2*, *MPL* and *CALR*. MPN patients have an increased risk to progress into acute myeloid leukemia and to develop other (second) primary tumors.

The aim of the present study was to investigate whether there is any association between f, used as surrogate of μ, and individual cancer risk in a cohort of MPN patients.

MATERIALS AND METHODS

Subjects

Mutant frequency (f) quantification was performed on peripheral blood granulocytes of 35 healthy volunteers enrolled during the present study and 142 healthy volunteers included in a work previously published by our laboratory (Rondelli et al., 2013), for a total of 177 healthy volunteers (75 females and 102 males, with age ranging from 22 to 63 years).

Moreover, f was measured in peripheral blood granulocytes collected from 245 patients, which were referred to the "Center of Research and Innovation of Myeloproliferative Neoplasms" (CRIMM at Azienda Ospedaliero-Universitaria Careggi, Firenze, directed by Prof. Alessandro Vannucchi) for suspected or previously diagnosed ET, PV, PMF (including both pre-PMF and overt PMF), PET-MF or PPV-MF.

Both patients and healthy volunteers signed informed consent in accordance with an IRB approved protocol before blood sample collection. Peripheral blood samples (7-11 ml) were collected in EDTA-containing tubes and stored at room temperature. Experimental procedures were started within two hours from the time of collection.

Purification of peripheral blood granulocytes

A double density centrifugation method was used to obtain highly purified granulocytes from peripheral blood samples, as previously described (Rondelli et al., 2013).

Briefly, each blood sample was diluted ~1:3 with pre-warmed sterile PBS. A double density Ficoll gradient was prepared in 15 ml Falcon tubes by carefully layering 3.5 ml of low density Ficoll (LymphoSep® - Lymphocyte Separation Media, Biowest, with density of 1,077 g/ml) over 3.5 ml of high density Ficoll (Lympholyte® - poly Cell Separation Media, Cedarlane, with density of 1,113 g/ml); the PBS-diluted blood was then carefully layered over the double density Ficoll in a 1:1 ratio.

Samples were centrifuged at 500 x *g* for 35 minutes at room temperature. After centrifugation, two rings were formed: the upper one, at the plasma - low density Ficoll interphase, mainly contained mononuclear cells; in the lower one, at the low density Ficoll - high density Ficoll interphase, granulocytes were preferentially found (Figure 5). The majority of red blood cells settled at the bottom, but in some cases they slightly contaminated the lower ring without affecting subsequent steps.

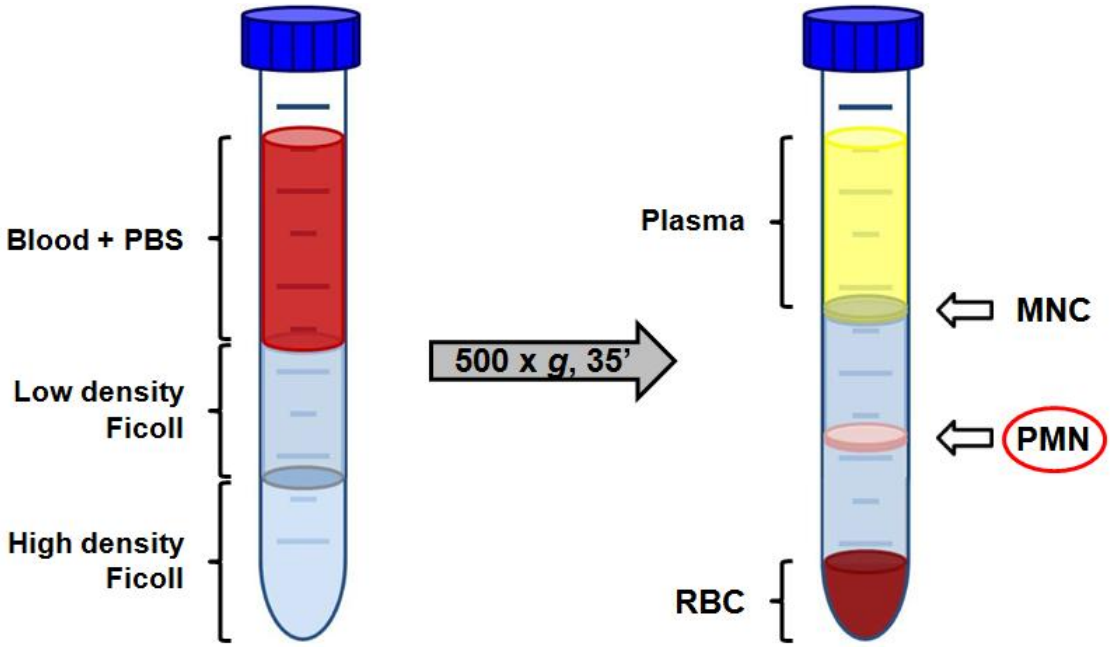

Figure 5. Double density centrifugation method.
MNC = Mononuclear Cells; PMN = Polymorphonuclear Cells; RBC = Red Blood Cells.

The lower ring was gently collected in 50 ml Falcon tubes and washed with pre-warmed PBS in order to eliminate any residue of Ficoll. Samples were centrifuged at 400 x *g* for 5 minutes at room temperature and the supernatant was then removed by pipetting. Pellets were re-suspended in a cold solution of NH_4Cl 0.157 M - $KHCO_3$ 0.01 M - EDTA 100 mM in order to lyse any contaminating red blood cell. When solutions became visibly transparent, samples were centrifuged at 300 x *g* for 5 minutes at 4°C and the supernatant was removed by pipetting. Pellets containing the purified granulocytes where then gently re-suspended in 4 ml of cold PBS and

cells were counted in Neubauer chambers after staining with Trypan Blue to distinguish living cells. Aliquots containing 3 x 10^6 granulocytes were subsequently transferred into polystyrene tubes (BD Bioscience) for the antibody staining step.

Staining of granulocytes

Cell suspensions were centrifuged at 300 x *g* for 5 minutes at 4°C and supernatant was discarded. Pellets were then stained with 5 antibodies: three PE-coniugated antibodies recognizing three different GPI-linked proteins (CD24, CD55, CD59) and two antibodies specific for transmembrane proteins (the leukocyte-specific CD45 and the granulocyte-specific CD11b) conjugated to APC and FITC, respectively.

CD24 is a small glycoprotein expressed by a variety of cell types. In the hematopoietic compartment, CD24 is expressed by B cells, T cells, neutrophils, eosinophils, dendritic cells and macrophages (Fang et al., 2010).

CD55, which is also known as Decay-Accelerating Factor (DAF), belongs to the Regulators of Complement Activation (RCA) family of proteins (Kim and Song, 2006). CD59 is another complement regulator. Both CD55 and CD59 are ubiquitously expressed by human cells (Kim and Song, 2006).

As previously recommended (Peruzzi et al., 2010), the expression of three GPI-linked proteins on the cell surface was assessed to discriminate between *PIGA*-mutated/GPI-deficient cells and cells lacking for any other reason the expression of one or even two specific GPI-linked proteins (because it would be unlikely that a cell did not express on its surface three different GPI-linked proteins at the same time if the GPI was correctly produced).

CD45, which is also known as Ly-5 or leukocyte common antigen, is a large transmembrane glycoprotein that functions as receptor protein tyrosine phosphatase. It is expressed by all nucleated hematopoietic cells, whereas mature erythrocytes and platelets are CD45-deficient (Rheinländer et al., 2018).

CD11b is the Integrin Alpha M (ITGAM), a protein subunit that, in association with CD18, constitutes the phagocytic receptor *α*M*β*2, also known as Complement Receptor 3 (CR3) or Macrophage-1 antigen (Mac-1). It is expressed by polymorphonuclear leukocytes (mostly neutrophils), Natural Killer (NK) lymphocytes and mononuclear phagocytes such as macrophages (Lim and Hotchin, 2012).

The following antibodies were used: PE mouse anti-human CD24 - 5 µl/10^6 cells (BD Pharmigen, clone ML5); PE mouse anti-human CD55 - 1 µl/10^6 cells (BD Pharmigen, clone IA10); PE mouse anti-human CD59 - 1 µl/10^6 cells (Bio-Rad, clone MEM-43); APC mouse anti-human CD45 - 3 µl/10^6 cells (BD Pharmigen, clone HI30) and FITC mouse anti-human CD11b - 5 µl/10^6 cells (Miltenyi Biotec, clone M1/70.15.11.5).

Antibodies were added directly onto the pellets, which were then carefully mixed by flicking the tubes to ensure that all cells took contact with the antibodies. The incubation was performed for 30 minutes at 4°C. After incubation, cells were washed with cold PBS and centrifuged at 300 x *g* for 5 minutes at 4°C. Supernatant was discarded, pellets were re-suspended in ~ 100 µl of cold PBS and stored at 4°C until analyzed by flow cytometry.

New staining

In order to improve the detection of GPI-deficient granulocytes, a different combination of antibodies was tested. Investigated antibodies were as follows: PE mouse anti-human CD24 - 5 µl/10^6 cells (BD Pharmigen, clone ML5); PE mouse anti-human CD66b - 6 µl/10^6 cells (BD Pharmigen, clone G10F5); PE mouse anti-human CD157 - 6 µl/10^6 cells (BD Pharmigen, clone SY/11B5); APC mouse anti-human CD45 - 3 µl/10^6 cells (BD Pharmigen, clone HI30); Pacific Blue mouse anti-human CD11b - 1 µl/10^6 cells (Beckman Coulter, clone Bear1). Five µl of PE Mouse IgG1, κ Isotype Control (BD Pharmigen, clone MOPC-21) were used to stain negative controls.

Flow cytometry analysis

Flow cytometry was performed with a BD FACSCanto II and FACS Diva Software was used for the analysis.

The gating strategy was performed as follows. Living granulocytes were initially selected on the basis of physical features (SSC and FSC). In order to exclude any erythrocyte and platelet, cells were selected according to the positive expression of the “pan-leukocyte” marker CD45, in order to be sure to discard any other type of cell. Subsequently, CD45-positive cells were evaluated for the expression of the granulocyte-specific marker CD11b and the 5% of cells expressing lower levels of CD11b was discarded, whereas the remaining 95% was identified as the granulocyte population of interest. Granulocytes were finally tested for the expression of the GPI-linked proteins CD24, CD55 and CD59. Only those cells that tested negative for the expression of the three GPI-linked proteins at the same time were identified as *PIG-A* mutants.

Since fluorescence intensity of GPI-linked proteins varies within individuals and within the experimental procedure, in order to obtain comparable results it was decided to use the 4% of the geometric mean of GPI-linked fluorescence values of all events from each experiment as cutoff to discriminate between positive and negative events, as previously described (Peruzzi et al., 2010).

In the interest of accuracy, an acquisition flow rate lower than 3000 events per second was used.

Since mutations are stochastic events and GPI-deficient cells are usually very rare, it was previously estimated that, in order to obtain an accurate measurement of the rare mutant cells and to reduce as much as possible the drift effect, it is necessary to screen at least 1 million of events (Araten et al., 2005, Peruzzi et al., 2010). In practice, at least 1.5 million of events, selected according to physical parameters, were acquired per each experiment.

Determination of *ƒ*

The parameter *ƒ* was calculated by counting the number of granulocytes lacking all three GPI-linked proteins, identified by the procedure described above, per million of events. In order to compare the distribution of *ƒ* without censoring the cases in which no mutant granulocytes were detected because *ƒ* was below the limit of detection of the method (*ƒ* = 0), zero values were replaced by the reciprocal of the number of cells counted [*e.g.* $1/(1.5 \times 10^6)$] divided by root square of 2 (Hornung and Reed, 1990).

Statistical analyses

Normality of distributions of *ƒ* values measured in different groups of subjects has been tested by the Shapiro-Wilk test. The Mann-Whitney U test for unpaired data has been used to compare distributions of *ƒ* values. Differences between number of subjects with *ƒ* values higher than a specific cut-off within specific subgroups have been tested by the Chi-Square test. Correlation analyses have been carried out according to Pearson. Statistical significance has been accepted for $p<0.05$. All statistical analyses have been performed with GraphPad Prism 5.

RESULTS

The distribution of *ƒ* in healthy subjects

The frequency of GPI-negative granulocytes per million of events (*ƒ*) was measured in the peripheral blood of 35 healthy subjects, in addition to the 142 previously studied in our laboratory (Rondelli et al., 2013).

The distribution of *ƒ* values in these 35 new subjects was not different from that found in the previously studied 142. Overall, in this cohort of 177 healthy donors (75 females and 102 males, with age ranging from 22 to 63 years) *ƒ* values ranged from 0.52 to 37.51, with a median of 4.37; the mean was 6.39, with a standard deviation of 5.6 (Figure 6).

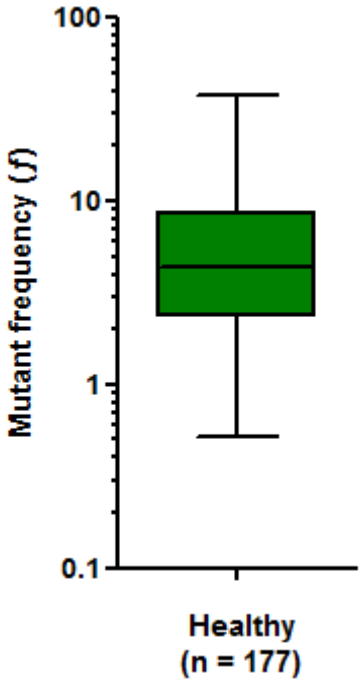

Figure 6. Distribution of *ƒ* values in a population of healthy subjects.
Box represents the middle 50% of *ƒ* values. The horizontal line inside the box indicates the median value of the distribution of *ƒ*. The upper and lower ends of the box represent the approximate upper and lower quartiles of the distribution of *ƒ*. Whiskers indicate the highest and the lowest data points.

The distribution of *ƒ* in MPN patients

The parameter *ƒ* was measured in the peripheral blood of 245 patients referred to the "Center of Research and Innovation of Myeloproliferative Neoplasms" (Azienda Ospedaliero-Universitaria Careggi, Firenze) for suspected or previously diagnosed Philadelphia-negative classical MPN. Forty-two patients were then excluded because the final diagnosis was different from PV, ET or MF, and other 8 were excluded for technical issues during the experimental phase. The demographic of the remaining 195 patients was as follow: 82 females and 113 males, with age ranging from 24 to 90 years. 58 patients were affected by ET, 65 by PV, 48 by PMF and 24 by PET-MF or PPV-MF. All available clinical and molecular features of patients are summarized in Table 1.

Patient characteristics	**ET**	**PV**	**PMF**	**PET/ PPV-MF**	**Total**
Males/Females	24/34	41/24	34/14	14/10	**113/82**
Average age (range)	58 (25-87)	66 (33-83)	61 (24-90)	67 (50-88)	**63 (24-90)**
Diagnosis/follow-up	30/25*	27/33*	17/26*	10/14	**84/98***
JAK2 mutated/tested	44/58	62/64	34/46	19/24	**159/192**
MPL mutated/tested	2/46	0/32	2/35	0/18	**4/131**
CALR mutated/tested	6/27	0/24	5/21	1/7	**12/79**
Abnormal cytogenetics/tested	2/44	6/43	6/30	2/14	**16/131**
Second primary tumor	3	11	4	4	**22**

*The date of diagnosis of 13 patients (3 ET, 5 PV and 5 PMF) was unknown.

Table 1. Patient characteristics.

In these 195 MPN patients, *f* values ranged from 0.36 to 1093.34, with a median of 4.75; the mean was 21.5, with a standard deviation of 104 (Figure 7).

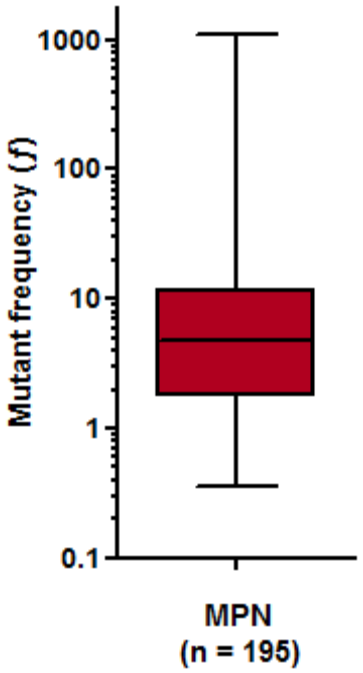

Figure 7. Distribution of *f* values in MPN patients.
Box represents the middle 50% of *f* values. The horizontal line inside the box indicates the median value of the distribution of *f*. The upper and lower ends of the box represent the approximate upper and lower quartiles of the distribution of *f*. Whiskers indicate the highest and the lowest data points.

In healthy subjects, gender does not seem to affect *f* values (Rondelli et al., 2013) and this was confirmed also in MPN patients. Indeed, in female patients (n = 82), *f* values ranged from 0.48 to 196.54, with median of 3.3, mean of 11.26 and standard deviation of 24.66 (Figure 8); in male patients (n = 113), *f* values ranged from 0.36 to 1093.34, with median of 5.08, mean of 28.93 and standard deviation of 134.83 (Figure 8), with no significant difference (Mann-Whitney U test, p = 0.13).

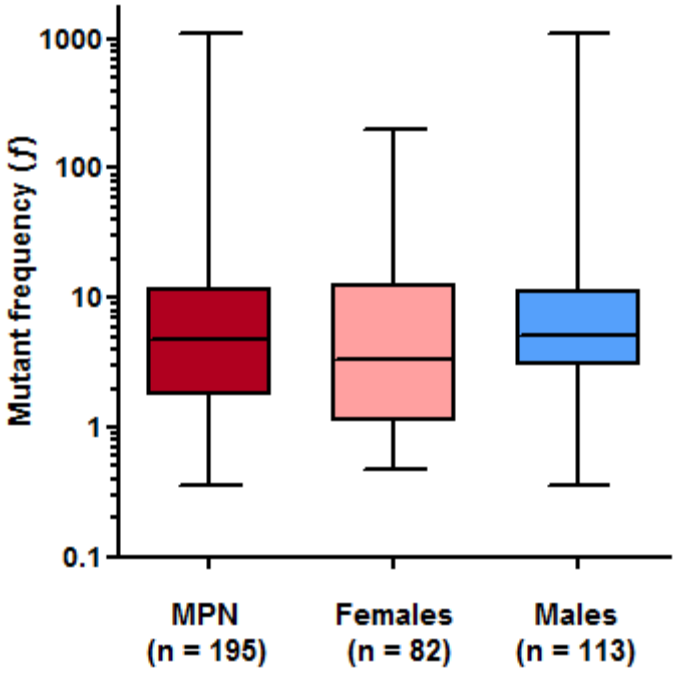

Figure 8. Distribution of f values in MPN patients stratified by gender.
Boxes represent the middle 50% of f values. The horizontal lines inside the boxes indicate the median values of the distributions of f. The upper and lower ends of each box represent the approximate upper and lower quartiles of the distributions of f. Whiskers indicate the highest and the lowest data points.

It has been also investigated whether there was any association between f values and age in MPN patients, but no correlation was found between the two parameters (R^2 = 0.0009, p = 0.68) (Figure 9).

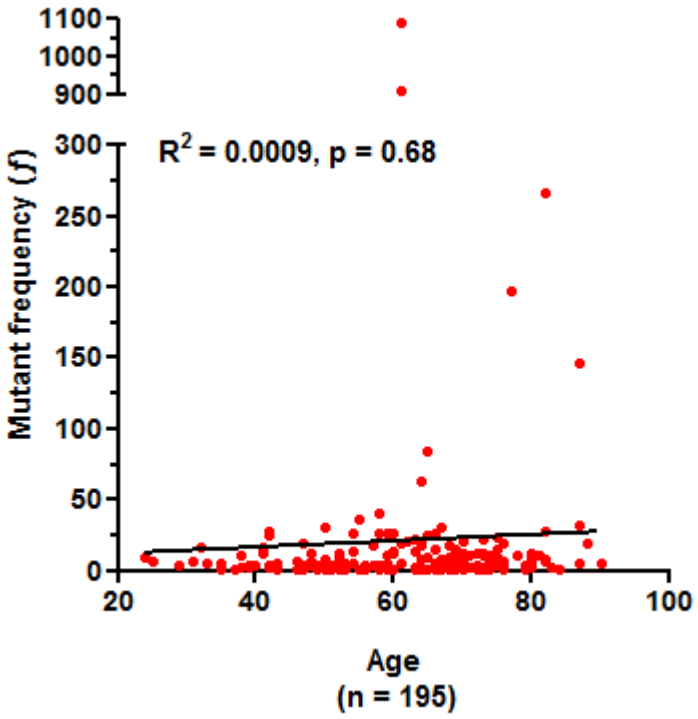

Figure 9. Correlation between f values and age in MPN patients.

The distribution of *f* values in MPN patients in comparison to healthy subjects

The distributions of *f* values in MPN patients and healthy subjects were compared and no statistically significant differences were found (Mann-Whitney U test, p = 0.64) (Figure 10).

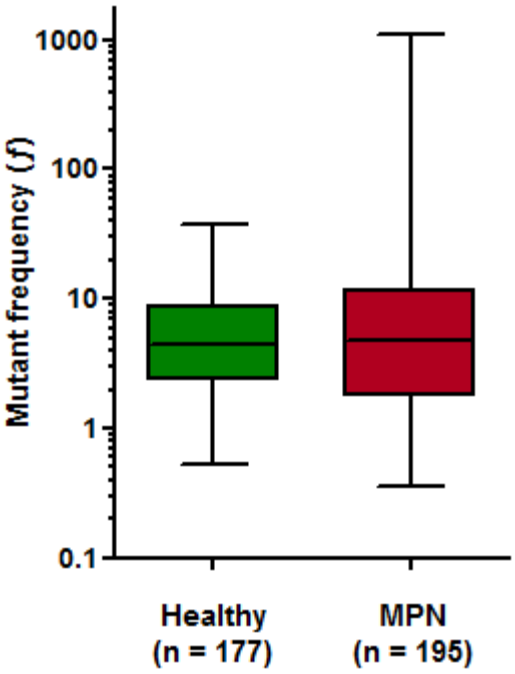

Figure 10. Distribution of *f* values in healthy subjects and MPN patients.
Boxes represent the middle 50% of *f* values. The horizontal lines inside the boxes indicate the median values of the distributions of *f*. The upper and lower ends of each box represent the approximate upper and lower quartiles of the distributions of *f*. Whiskers indicate the highest and the lowest data points.

However, the visual inspection of data distributions suggested that there was an excess of high *f* values in the group of MPN patients compared to the group of healthy subjects. This is somehow similar to what was previously observed in a group of 49 oral squamous cell carcinoma patients, in which the proportion of cancer patients with *f* values higher than the value that defined the 90th percentile of *f* distribution in 142 healthy subjects was 20% (Notaro et al., unpublished results).

After performing a similar analysis in MPN patients, we found that the "outliers" (subjects with *ƒ* values higher than the value that defines the 90th percentile of *ƒ* distribution in the healthy population) were 41 out of 195 (21%) (Chi-Square test, p = 0.004).

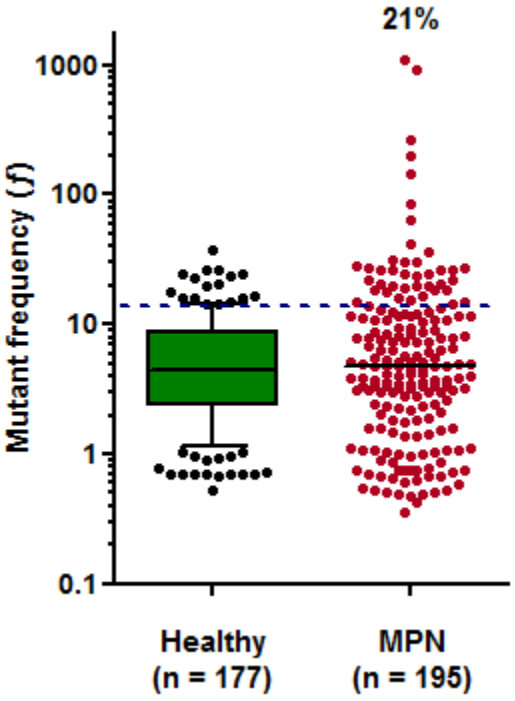

Figure 11. Outliers in healthy subjects and MPN patients. Distribution of *ƒ* values in the group of healthy subjects is represented as a box, where the internal horizontal line indicates the median value and the upper and lower ends of the box represent the approximate upper and lower quartiles of the distribution of *ƒ*. In this case, whiskers indicate the 90th and the 10th percentiles of the distribution of *ƒ*. Black circles represent *ƒ* values above and below the 90th and the 10th percentiles, respectively. Each *ƒ* value in the group of MPN patients is represented as a purple red circle and the black line indicates the median value of the distribution of *ƒ*. The dashed blue line highlights the 90th percentile of the distribution of *ƒ* values in the group of healthy subjects, which represents the cut-off to identify outliers in both groups.

The influence of the *JAK2* mutational status on *ƒ* in MPN patients

Further analyses of *ƒ* values were performed after stratification of patients according to the mutational status of *JAK2*, which was known in 192 cases out of 195. In the subgroup of 159 patients with mutations in *JAK2*, *ƒ* values ranged from 0.36 to 910.84, with median of 4.55, mean of 16.22 and standard deviation of 74.57 (Figure 12). The distribution of *ƒ* values in the 33 patients with unmutated *JAK2* (4 of which were proven to harbor mutations in *MPL* and 12 in *CALR*) ranged from 0.48 to 1093.34, with

median of 4.99, mean of 48.33 and standard deviation of 193.05 (Figure 12). No significant differences between f distributions in the two subgroup of patients were found (Mann-Whitney U test, p = 0.44).

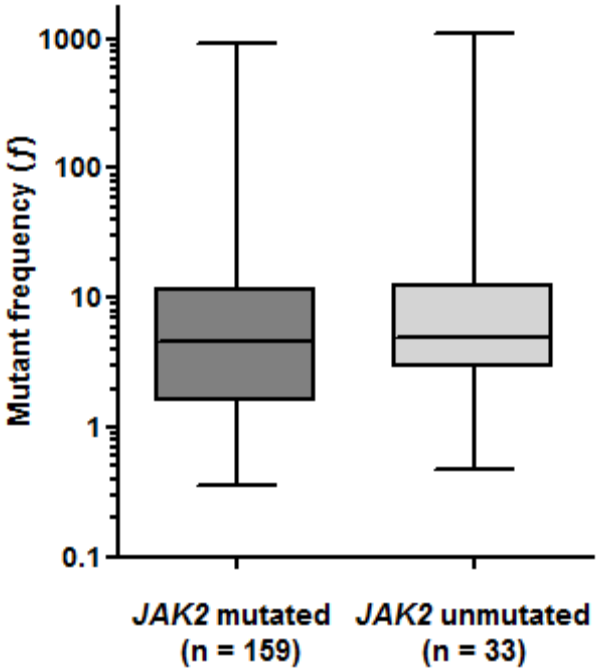

Figure 12. Distribution of f values in MPN patients stratified according to the mutational status of *JAK2*. Boxes represent the middle 50% of f values. The horizontal lines inside the boxes indicate the median values of the distributions of f. The upper and lower ends of each box represent the approximate upper and lower quartiles of the distributions of f. Whiskers indicate the highest and the lowest data points.

In addition, there was no correlation between f values and *JAK2* allele burden in the 155 *JAK2*-mutated patients about which this information was known (R^2 = 0.019, p = 0.09) (Figure 13).

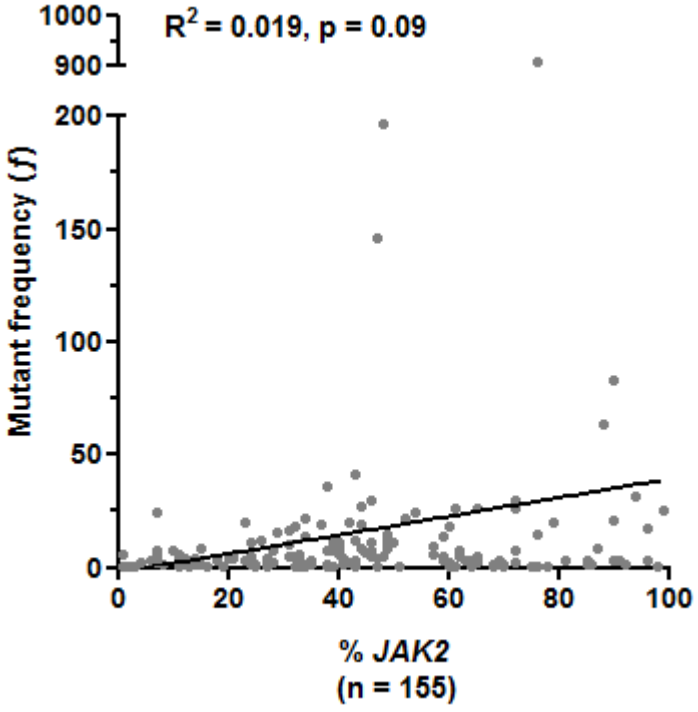

Figure 13. Correlation between f values and burden of mutated *JAK2*.

The influence of the time from diagnosis on *f* in MPN patients

The distribution of *f* was compared between the patients who were diagnosed in the same year of *f* analysis ("diagnosis" subgroup) and those in which *f* was measured during a follow-up visit ("follow-up" subgroup). In the 84 patients at diagnosis, *f* values ranged from 0.36 to 63.41, with median of 3.62, mean of 6.72 and standard deviation of 9.44 (Figure 14). In the 98 follow-up patients, *f* values ranged from 0.48 to 1093.34, with median of 5.21, mean of 32.01 and standard deviation of 143.07 (Figure 14). The diagnosis date was not known in 13 cases, therefore they were discarded from this analysis. Distributions of *f* values in the two subgroup of patients were compared and differences resulted statistically significant (Mann-Whitney U test, p = 0.006).

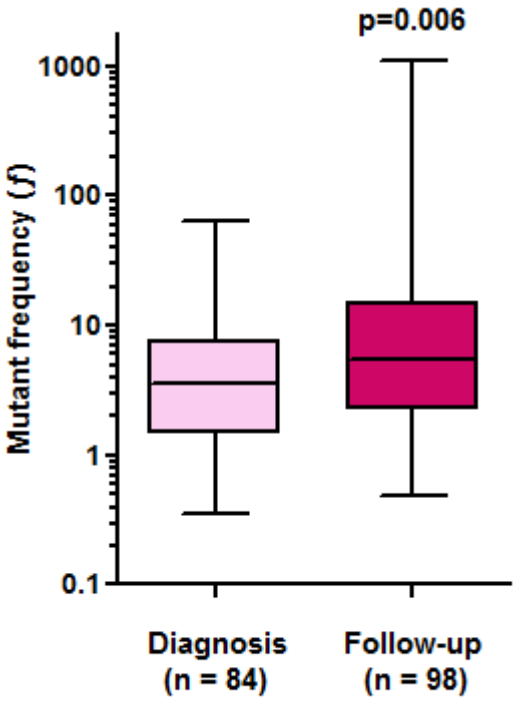

Figure 14. Distribution of *f* values in MPN patients whose *f* was measured in the same year of diagnosis or during follow-up.
Boxes represent the middle 50% of *f* values. The horizontal lines inside the boxes indicate the median values of the distributions of *f*. The upper and lower ends of each box represent the approximate upper and lower quartiles of the distributions of *f*. Whiskers indicate the highest and the lowest data points.

However, in the subgroup of patients studied during follow-up there was no correlation between *f* values and years from diagnosis (R^2 = 0.009, p = 0.36) (Figure 15).

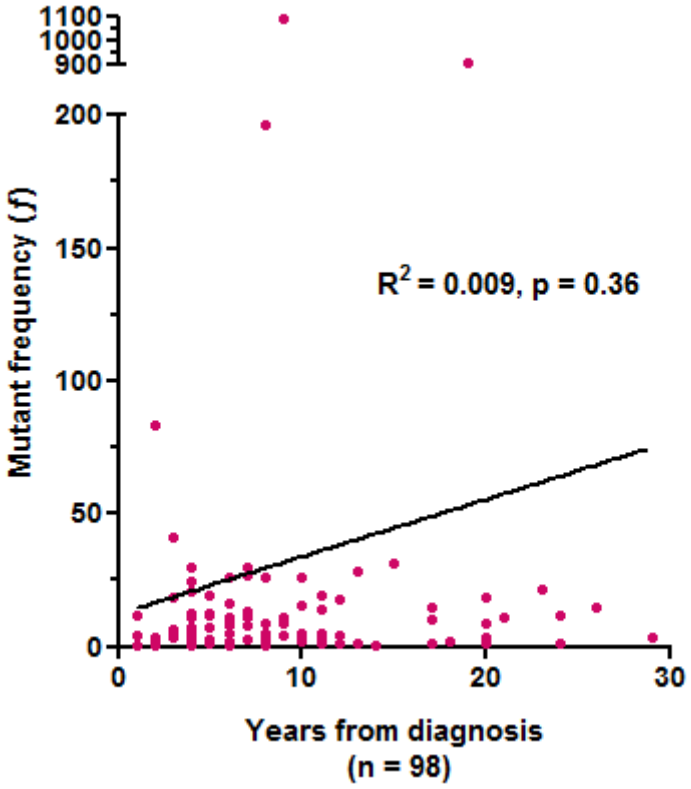

Figure 15. Correlation between *f* values and years from diagnosis.

The distribution of *f* according to the specific MPN disease

In order to deepen the analysis of *f* distribution in MPN patients, they were categorized according to the specific disease: ET, PV, PMF and PET/PPV-MF. In the subgroup of 58 ET patients, *f* values ranged from 0.48 to 35.98, with median of 3.85, mean of 6 and standard deviation of 6.65 (Figure 16). In the 65 PV patients, *f* values ranged from 0.36 to 910.84, with median of 3.39, mean of 24.49 and standard deviation of 116.47 (Figure 16). The PMF subgroup of 48 patients was characterized by *f* values ranging from 0.58 to 196.54, with median of 6.34, mean of

15.76 and standard deviation of 34.34 (Figure 16). Finally, f values in the 24 PET/PPV-MF patients ranged from 0.51 to 1093.34, with median of 12.73, mean of 62.36 and standard deviation of 220.47 (Figure 16).

Comparison of distributions of f values between healthy subjects and each subgroup of MPN patients yielded statistically significant differences in PMF (Mann-Whitney U test, p = 0.04) and PET/PPV-MF (Mann-Whitney U test, p = 0.0003) cases, but not in ET (Mann-Whitney U test, p = 0.21) and PV (Mann-Whitney U test, p = 0.14) cases (Figure 16).

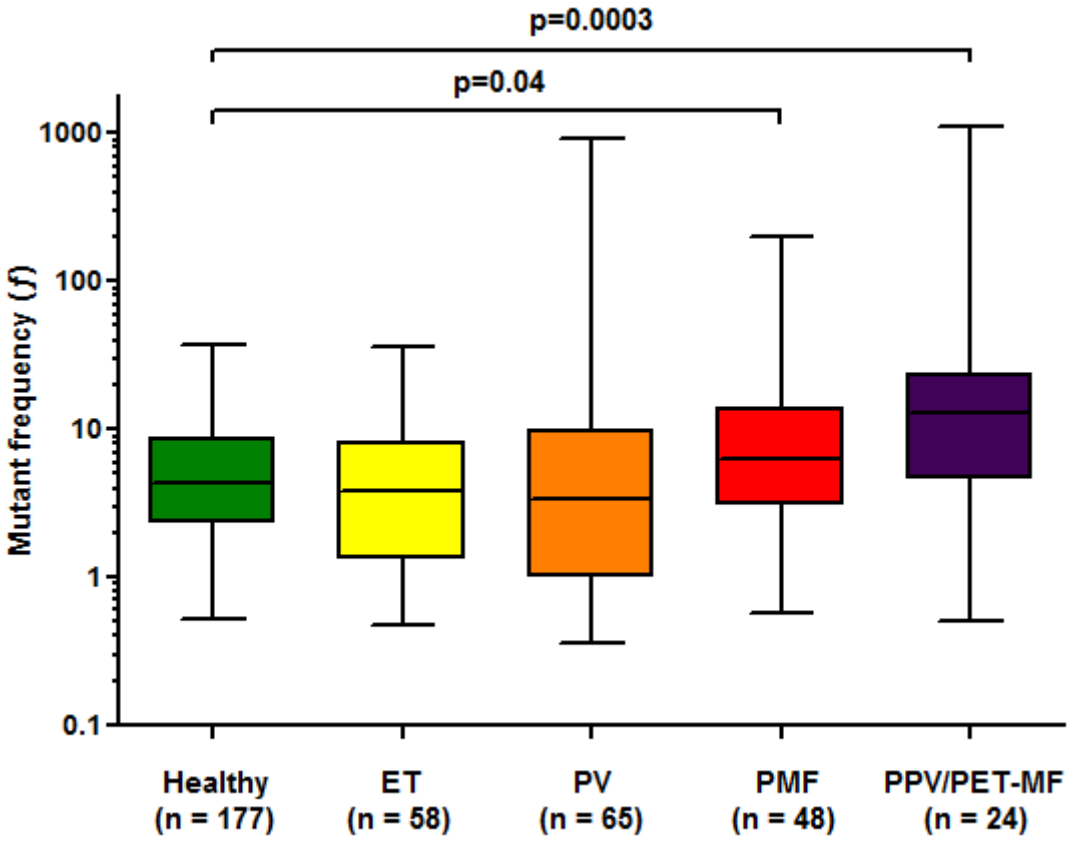

Figure 16. Distribution of f values in healthy subjects and MPN patients classified according to the specific disease.
Boxes represent the middle 50% of f values. The horizontal lines inside the boxes indicate the median values of the distributions of f. The upper and lower ends of each box represent the approximate upper and lower quartiles of the distributions of f. Whiskers indicate the highest and the lowest data points.

When the proportion of outliers was assessed, as described above, it resulted significantly higher in the PV subgroup (13 out of 65), the PMF

subgroup (12 out of 48) and the PET/PPV-MF subgroup (10 out of 24) of patients compared to the healthy subjects (Chi-Square test, p = 0.04, p = 0.007 and p<0.0001, respectively), while that in the ET subgroup was comparable (6 out of 58) (Chi-Square test, p = 0.97) (Figure 17).

We noticed that, when ordering MPN diseases according to a rough estimate of their clinical severity from the mildest to the most severe (ET, PV, PMF and PET/PPV-MF), there was an apparent increase in the frequency of outliers. Interestingly, this increase resulted statistically significant when tested with the robust Chi-Square test for trend (p<0.002).

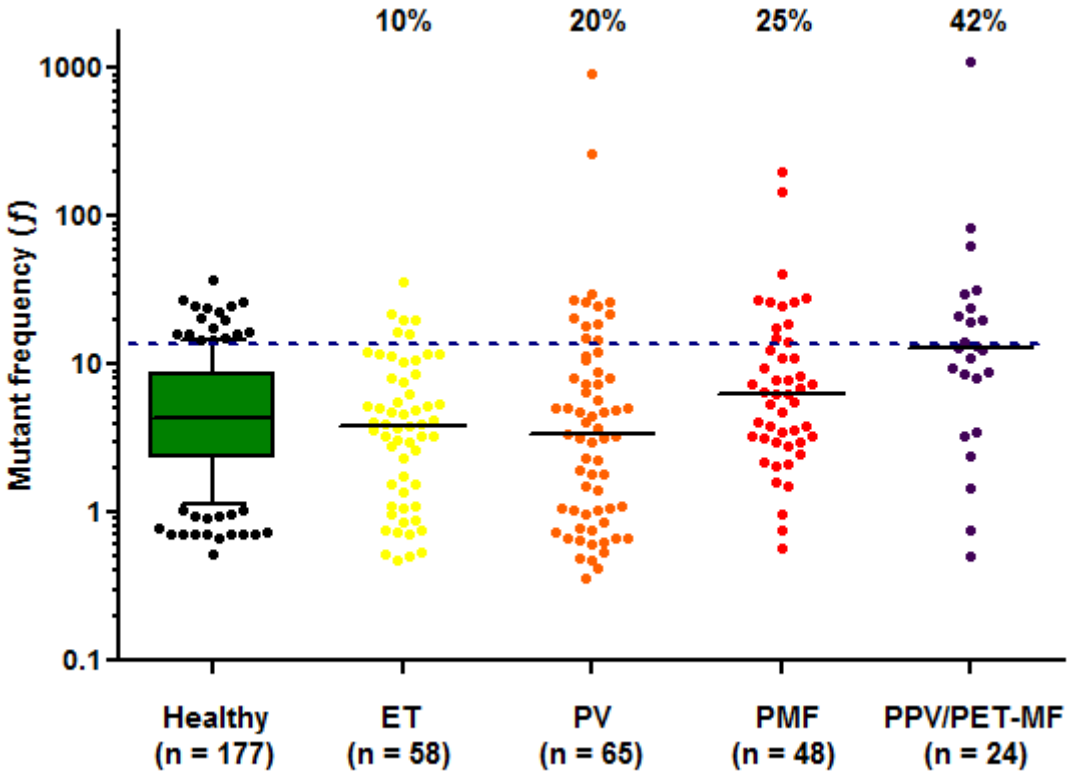

Figure 17. Number of outliers in healthy subjects and MPN patients classified according to the specific disease.

The distribution of f values in the group of healthy subjects is represented as a box, where the internal horizontal line indicates the median value and the upper and lower ends of the box represent the approximate upper and lower quartiles of the distribution of f. In this case, whiskers indicate the 90th and the 10th percentiles of the distribution of f. Black circles represents f values above and below the 90th and the 10th percentiles, respectively. Each f value in the subgroups of MPN patients is represented as a colored circle and the black lines indicate the median values of the distributions of f. The dashed blue line highlights the 90th percentile of the distribution of f values in the group of healthy subjects, which represents the cut-off to identify outliers in each group.

The distribution of *ƒ* in patients affected by MPN and a second primary tumor

A comparison between the distribution of *ƒ* values in a subgroup of MPN patients which had developed at least one second primary tumor throughout their life and that in patients affected only by MPN was performed. The distribution of *ƒ* values in the subgroup of 173 patients affected only by MPN ranged from 0.36 to 1093.34, with median of 4.13, mean of 21.2 and standard deviation of 108.81 (Figure 18). In the subgroup of 22 patients with more than one primary tumor *ƒ* values ranged from 0.53 to 266, with median of 13.78, mean of 23.85 and standard deviation of 54.96 (Figure 18). Differences between the two distributions of *ƒ* values resulted statistically significant (Mann-Whitney U test, p = 0.03).

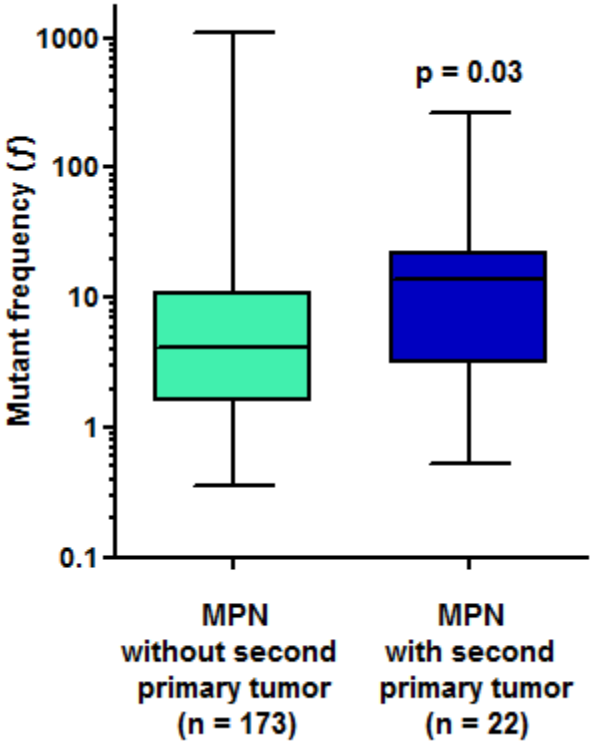

Figure 18. Distribution of *ƒ* values in patients affected only by MPN and patients with at least one second primary tumor.
Boxes represent the middle 50% of *ƒ* values. The horizontal lines inside the boxes indicate the median values of the distributions of *ƒ*. The upper and lower ends of each box represent the approximate upper and lower quartiles of the distributions of *ƒ*. Whiskers indicate the highest and the lowest data points.

DISCUSSION

Somatic mutations are stochastic events whose occurrence is the result of a combination of genetic factors, endogenous and environmental factors that may threaten genome integrity, rate of cell divisions and chance (Luzzatto and Pandolfi, 2015). Somatic mutations are a constant of cell replication, as the activity of DNA polymerases and DNA repair systems is highly efficient, but not perfect (DePamphilis, 2016). Multicellular organisms are constituted by thousands of cell clones which originated from common progenitors and, during cell replication, acquired a set of genetic changes from the original genome. Luckily for us, the great part of spontaneous somatic mutations does not confer to cells any growth advantage or malignant feature; however, a very small fraction of them is oncogenic (Luzzatto and Pandolfi, 2015, Stratton et al., 2009). Thus, the "chance" of occurrence of a set of oncogenic mutations within a cell and, as a consequence, the risk of developing cancer should be higher in individuals with a general greater predisposition to develop somatic mutations.

The main parameters used to quantify the occurrence of somatic mutations are mutation rate (μ) (Luria and Delbrück, 1943) and mutant frequency (f) (Green et al., 1995). The flow cytometry *PIGA* mutant assay (Araten et al., 1999) is one of the available methods to measure these two parameters; it is based on the X-linked *PIGA* gene (Miyata et al., 1993), which is involved in the first step of GPI-anchor biosynthesis (Kinoshita and Fujita, 2016).

In order to investigate whether there is any association between the rate of occurrence of somatic mutations and the individual risk of developing cancer, we used the flow cytometry *PIGA* mutant assay to measure the parameter f in peripheral blood granulocytes of 177 healthy subjects and of a cohort of 195 patients affected by Philadelphia negative classical MPN.

Interestingly, we observed that the distribution of *ƒ* values in MPN patients and healthy subjects is similar (Figure 10). This finding suggests that this type of tumors may develop in subjects whose rate of somatic mutation, and consequently *ƒ*, is within the normal range, in contrast with the expression of a mutator phenotype (Loeb, 2011). This is not surprisingly, since what really matters for cancer development is not the overall amount of somatic mutations, but a combination of factors that includes, but are not limited to, their functional effect and the role of the mutated gene in the specific tissue. Accordingly, in a recent study on 197 MPN patients, a panel of 105 cancer-related genes has been investigated by targeted Next Generation Sequencing (NGS) and mutations in only one MPN driver gene were detected in 55% of patients (Lundberg et al., 2014).

In our series of MPN patients we have also studied whether there was any association between *ƒ* distribution and mutations in *JAK2*. We have to notice that, even though we did not apply any type of selection during the enrollment of MPN patients and PV cases in our cohort were only 33%, a selection bias occurred since 159 out of 192 (83%) MPN patients were proven to be *JAK2* mutated, whilst only 33 were not. Anyway, we observed that *ƒ* distribution does not change when MPN patients are stratified according to the mutational status of *JAK2* and we did not find any correlation between *ƒ* values and *JAK2* allele burden (Figure 12 and 13). Even if we cannot exclude that other genes, in a different genomic context, may behave differently from *PIGA*, our results contrast with the notion of a general genomic instability associated with a deregulated *JAK2* (Scott and Rebel, 2012).

When we compared the distribution of *ƒ* in MPN patients at diagnosis at the time of *ƒ* measurement with that in patients whose *ƒ* was assessed during a follow-up visit, we observed a significant increase in the second group ($p = 0.006$) (Figure 14). This finding seem to suggest that *ƒ* may consistently increase over time, which would be in disaccord with the results of a previous study (Lundberg et al., 2014). However, in apparent contrast, we also observed that *ƒ* values in the follow-up group of patients

do not correlate with the number of years from diagnosis, as it would be reasonable to expect if *f* really increased over time (Figure 15). A possible explanation of this apparent discrepancy may be the fact that many of our follow-up patients underwent therapeutic treatments, including chemotherapy, which could have raised the frequency of mutant granulocytes detected in their blood by acting as additional environmental mutagenic factors. The therapeutic management of patients was not addressed in the work of Lundberg and colleagues (Lundberg et al., 2014) and, at the same time, our information about duration and type of therapies are incomplete and do not allow a more in-depth analysis. Therefore, we cannot completely rule out that the difference we observed was simply due to chance. In order to definitively demonstrate whether *f* increases in patients over time would be to perform a prospective study with repeated measurements of this parameter.

An intriguing observation we made is that, even if the distribution of *f* values in MPN patients resulted comparable to that of the normal population, an accumulation of subjects with high *f* values was evident in cancer patients. More precisely, in a consistent proportion (21%) of MPN patients (which we defined “outliers”), *f* values were higher than those measured in the 90% of healthy subjects (Figure 11). It is noteworthy that very similar results were obtained in a study previously performed in our laboratory on patients affected by a solid tumor (Notaro et al., unpublished results). Indeed, distribution of *f* values in a group of 49 oral squamous cell carcinoma patients was not different from the healthy population, but the proportion of outliers within the cancer patients group was 20%. Altogether, these findings strongly supports the hypothesis that, even if tumors may develop also in the presence of a “normal” mutation rate, an increased predisposition to develop mutations is more frequently found in subjects who eventually develop a tumor. As explained before, we cannot completely rule out that, at least in MPN, the higher *f* values that we observed in some patients were the consequence of tumor development, rather than the cause. However, it would be much more complicated to

explain why similar results were obtained by measuring f in the blood of subjects whose tumor had a completely different localization. Moreover, we found further support to this hypothesis when we observed that the distribution of f values in 22 MPN patients which developed, throughout their life, at least one second primary tumor (it is noteworthy that, in our set of patients, 3 had hematological malignancies of the lymphocyte compartment and 19 different types of solid tumors) was significantly higher than that of the other MPN patients ($p = 0.03$) (Figure 18). It has been often speculated that, in the absence of hereditary syndromes associated with mutations in genes involved, for instance, in DNA repair mechanisms (including ataxia telangiectasia, Fanconi anemia…) and which are known to cause cancer predisposition (Vogelstein and Kinzler, 2004), development of multiple primary tumors in the same individuals may depend on a higher mutation rate. However, to our knowledge this is the first study which provided direct evidence of that.

As regarding, more specifically, the MPN diseases, it is interesting to note that, when compared to the healthy population, a significantly higher distribution of f is found in PMF and PET/PPV-MF cases ($p = 0.04$ and $p = 0.0003$, respectively) (Figure 16). From a clinical point of view, PMF and secondary MF are the most severe of MPN diseases: they are associated with a poorer prognosis and with an increased risk of progression to AML (Meier and Burton, 2014). In addition, we also observed that the proportion of outliers, as previously defined, is lower in ET, the mildest of MPNs, and progressively increases, almost in parallel to a rough estimate of the clinical severity of the disease, in PV, PMF and PET/PPV-MF (Figure 17). The molecular basis underlying the phenotypic differences of MPN diseases are still controversial (Jones and Cross, 2013, Scott and Rebel, 2012). However, the presence of additional mutations other than those affecting driver genes has been associated, in every MPN disease, with a poorer prognosis and a higher risk of progression into a more severe form of disease [reviewed in (Rumi and Cazzola, 2017, Tefferi and Barbui, 2017, Zimran et al., 2018)]. Keeping that in mind, it would be very

interesting to discover whether a greater proportion of outliers in our ET and PV groups will, at some point, progress into MF and, similarly, a greater proportion of outliers in MF groups will develop AML. At the same time, it could be also argued that the outliers in the healthy subjects group have an increased risk to develop, at some point, any kind of tumor compared to those with lower f.

Altogether, the results of our exploratory study suggest that higher f values are strongly associated with an increased risk to develop one, or even more primary tumors. Moreover, as regarding MPN, our findings indicate that an increased risk to develop a clinically more severe form of myeloproliferative disease is associated with higher f values. However, observational prospective studies where the parameter f is repeatedly measured throughout time are needed to properly validate our conclusions.

BIBLIOGRAPHY

Akiyama, M., Kyoizumi, S., Hirai, Y., Kusunoki, Y., Iwamoto, K. S. and Nakamura, N. (1995) 'Mutation frequency in human blood cells increases with age', *Mutat Res,* 338(1-6), pp. 141-9.

Albertini, R. J., Castle, K. L. and Borcherding, W. R. (1982) 'T-cell cloning to detect the mutant 6-thioguanine-resistant lymphocytes present in human peripheral blood', *Proc Natl Acad Sci U S A,* 79(21), pp. 6617-21.

Albertini, R. J., Nicklas, J. A., O'Neill, J. P. and Robison, S. H. (1990) 'In vivo somatic mutations in humans: measurement and analysis', *Annu Rev Genet,* 24, pp. 305-26.

Alfinito, F., Del Vecchio, L., Rocco, S., Boccuni, P., Musto, P. and Rotoli, B. (1996) 'Blood cell flow cytometry in paroxysmal nocturnal hemoglobinuria: a tool for measuring the extent of the PNH clone', *Leukemia,* 10(8), pp. 1326-30.

Araten, D. J., Bessler, M., McKenzie, S., Castro-Malaspina, H., Childs, B. H., Boulad, F., Karadimitris, A., Notaro, R. and Luzzatto, L. (2002) 'Dynamics of hematopoiesis in paroxysmal nocturnal hemoglobinuria (PNH): no evidence for intrinsic growth advantage of PNH clones', *Leukemia,* 16(11), pp. 2243-8.

Araten, D. J., Golde, D. W., Zhang, R. H., Thaler, H. T., Gargiulo, L., Notaro, R. and Luzzatto, L. (2005) 'A quantitative measurement of the human somatic mutation rate', *Cancer Res,* 65(18), pp. 8111-7.

Araten, D. J., Krejci, O., Ditata, K., Wunderlich, M., Sanders, K. J., Zamechek, L. and Mulloy, J. C. (2013) 'The rate of spontaneous mutations in human myeloid cells', *Mutat Res,* 749(1-2), pp. 49-57.

Araten, D. J. and Luzzatto, L. (2006) 'The mutation rate in PIG-A is normal in patients with paroxysmal nocturnal hemoglobinuria (PNH)', *Blood,* 108(2), pp. 734-6.

Araten, D. J., Martinez-Climent, J. A., Perle, M. A., Holm, E., Zamechek, L., DiTata, K. and Sanders, K. J. (2010) 'A quantitative analysis of genomic instability in lymphoid and plasma cell neoplasms based on the PIG-A gene', *Mutat Res,* 686(1-2), pp. 1-8.

Araten, D. J., Nafa, K., Pakdeesuwan, K. and Luzzatto, L. (1999) 'Clonal populations of hematopoietic cells with paroxysmal nocturnal hemoglobinuria genotype and phenotype are present in normal individuals', *Proc Natl Acad Sci U S A,* 96(9), pp. 5209-14.

Araten, D. J., Sanders, K. J., Pu, J. and Lee, S. (2009) 'Spontaneously arising red cells with a McLeod-like phenotype in normal donors', *Mutat Res,* 671(1-2), pp. 1-5.

Araten, D. J., Zamechek, L. and Halverson, G. (2014) 'No evidence of hypermutability in red cells from patients with paroxysmal nocturnal hemoglobinuria using the XK gene', *Haematologica,* 99(8), pp. e142-4.

Barbui, T., Tefferi, A., Vannucchi, A. M., Passamonti, F., Silver, R. T., Hoffman, R., Verstovsek, S., Mesa, R., Kiladjian, J. J., Hehlmann, R., Reiter, A., Cervantes, F., Harrison, C., Mc Mullin, M. F., Hasselbalch, H. C., Koschmieder, S., Marchetti, M., Bacigalupo, A.,

Finazzi, G., Kroeger, N., Griesshammer, M., Birgegard, G. and Barosi, G. (2018a) 'Philadelphia chromosome-negative classical myeloproliferative neoplasms: revised management recommendations from European LeukemiaNet', *Leukemia,* 32(5), pp. 1057-1069.

Barbui, T., Thiele, J., Gisslinger, H., Kvasnicka, H. M., Vannucchi, A. M., Guglielmelli, P., Orazi, A. and Tefferi, A. (2018b) 'The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion', *Blood Cancer J,* 8(2), pp. 15.

Bell, S., Kolobova, I., Crapper, L. and Ernst, C. (2016) 'Lesch-Nyhan Syndrome: Models, Theories, and Therapies', *Mol Syndromol,* 7(6), pp. 302-311.

Blaha, J., Schwarz, K., Fischer, C., Schauwecker, P., Höchsmann, B., Schrezenmeier, H. and Anliker, M. (2018) 'The Monoclonal Anti-CD157 Antibody Clone SY11B5, Used for High Sensitivity Detection of PNH Clones on WBCs, Fails to Detect a Common Polymorphic Variant Encoded by BST-1', *Cytometry B Clin Cytom,* 94(4), pp. 652-659.

Buchwald, M. (1977) 'Mutagenesis at the ouabain-resistance locus in human diploid fibroblasts', *Mutat Res,* 44(3), pp. 401-11.

Chattopadhyay, S., Zheng, G., Sud, A., Yu, H., Sundquist, K., Sundquist, J., Försti, A., Hemminki, A., Houlston, R. and Hemminki, K. (2018) 'Risk of second primary cancer following myeloid neoplasia and risk of myeloid neoplasia as second primary cancer: a nationwide,

observational follow up study in Sweden', *Lancet Haematol,* 5(8), pp. e368-e377.

Crump, K. S. and Hoel, D. G. (1974) 'Mathematical models for estimating mutation rates in cell populations', *Biometrika,* 61, pp. 237-252.

De Franceschi, L., Bosman, G. J. and Mohandas, N. (2014) 'Abnormal red cell features associated with hereditary neurodegenerative disorders: the neuroacanthocytosis syndromes', *Curr Opin Hematol,* 21(3), pp. 201-9.

DePamphilis, M. L. (2016) 'Genome Duplication: The Heartbeat of Developing Organisms', *Curr Top Dev Biol,* 116, pp. 201-29.

DePinho, R. A. (2000) 'The age of cancer', *Nature,* 408(6809), pp. 248-54.

Dertinger, S. D., Avlasevich, S. L., Bemis, J. C., Chen, Y. and MacGregor, J. T. (2015) 'Human erythrocyte PIG-A assay: an easily monitored index of gene mutation requiring low volume blood samples', *Environ Mol Mutagen,* 56(4), pp. 366-77.

Dobrovolsky, V. N., Miura, D., Heflich, R. H. and Dertinger, S. D. (2010) 'The in vivo Pig-a gene mutation assay, a potential tool for regulatory safety assessment', *Environ Mol Mutagen,* 51(8-9), pp. 825-35.

Elmore, E. and Barrett, J. C. (1982) 'Measurement of spontaneous mutation rates at the Na+/K+ ATPase locus (ouabain resistance) of human fibroblasts using improved growth conditions', *Mutat Res,* 97(5), pp. 393-404.

Elmore, E., Kakunaga, T. and Barrett, J. C. (1983) 'Comparison of spontaneous mutation rates of normal and chemically transformed human skin fibroblasts', *Cancer Res,* 43(4), pp. 1650-5.

Fang, X., Zheng, P., Tang, J. and Liu, Y. (2010) 'CD24: from A to Z', *Cell Mol Immunol,* 7(2), pp. 100-3.

Funaro, A., Ortolan, E., Ferranti, B., Gargiulo, L., Notaro, R., Luzzatto, L. and Malavasi, F. (2004) 'CD157 is an important mediator of neutrophil adhesion and migration', *Blood,* 104(13), pp. 4269-78.

Green, M. H., O'Neill, J. P. and Cole, J. (1995) 'Suggestions concerning the relationship between mutant frequency and mutation rate at the hprt locus in human peripheral T-lymphocytes', *Mutat Res,* 334(3), pp. 323-39.

Grist, S. A., McCarron, M., Kutlaca, A., Turner, D. R. and Morley, A. A. (1992) 'In vivo human somatic mutation: frequency and spectrum with age', *Mutat Res,* 266(2), pp. 189-96.

Hernández-Campo, P. M., Almeida, J., Sánchez, M. L., Malvezzi, M. and Orfao, A. (2006) 'Normal patterns of expression of glycosylphosphatidylinositol-anchored proteins on different subsets of peripheral blood cells: a frame of reference for the diagnosis of paroxysmal nocturnal hemoglobinuria', *Cytometry B Clin Cytom,* 70(2), pp. 71-81.

Hill, A., DeZern, A. E., Kinoshita, T. and Brodsky, R. A. (2017) 'Paroxysmal nocturnal haemoglobinuria', *Nat Rev Dis Primers,* 3, pp. 17028.

Hornung, R. and Reed, L. (1990) 'Estimation of average concentration in the presence of nondetectable values', *Applied Occupational and Environmental Hygiene*, pp. 46-51.

Jones, A. V. and Cross, N. C. (2013) 'Inherited predisposition to myeloproliferative neoplasms', *Ther Adv Hematol,* 4(4), pp. 237-53.

Kendal, W. S. and Frost, P. (1988) 'Pitfalls and practice of Luria-Delbrück fluctuation analysis: a review', *Cancer Res,* 48(5), pp. 1060-5.

Kim, D. D. and Song, W. C. (2006) 'Membrane complement regulatory proteins', *Clin Immunol,* 118(2-3), pp. 127-36.

Kinoshita, T. and Fujita, M. (2016) 'Biosynthesis of GPI-anchored proteins: special emphasis on GPI lipid remodeling', *J Lipid Res,* 57(1), pp. 6-24.

Kinoshita, T., Fujita, M. and Maeda, Y. (2008) 'Biosynthesis, remodelling and functions of mammalian GPI-anchored proteins: recent progress', *J Biochem,* 144(3), pp. 287-94.

Krüger, C. T., Hofmann, M. and Hartwig, A. (2015) 'The in vitro PIG-A gene mutation assay: mutagenicity testing via flow cytometry based on the glycosylphosphatidylinositol (GPI) status of TK6 cells', *Arch Toxicol,* 89(12), pp. 2429-43.

Kuroki, M., Abe, H., Imakiirei, T., Liao, S., Uchida, H., Yamauchi, Y. and Oikawa, S. (2001) 'Identification and comparison of residues critical for cell-adhesion activities of two neutrophil CD66 antigens, CEACAM6 and CEACAM8', *J Leukoc Biol,* 70(4), pp. 543-50.

Kyoizumi, S., Nakamura, N., Hakoda, M., Awa, A. A., Bean, M. A., Jensen, R. H. and Akiyama, M. (1989) 'Detection of somatic mutations at the glycophorin A locus in erythrocytes of atomic bomb survivors using a single beam flow sorter', *Cancer Res,* 49(3), pp. 581-8.

Landtblom, A. R., Bower, H., Andersson, T. M., Dickman, P. W., Samuelsson, J., Björkholm, M., Kristinsson, S. Y. and Hultcrantz, M. (2018) 'Second malignancies in patients with myeloproliferative neoplasms: a population-based cohort study of 9379 patients', *Leukemia,* 32(10), pp. 2203-2210.

Langlois, R. G., Akiyama, M., Kusunoki, Y., DuPont, B. R., Moore, D. H., Bigbee, W. L., Grant, S. G. and Jensen, R. H. (1993) 'Analysis of somatic cell mutations at the glycophorin A locus in atomic bomb survivors: a comparative study of assay methods', *Radiat Res,* 136(1), pp. 111-7.

Lengauer, C., Kinzler, K. W. and Vogelstein, B. (1998) 'Genetic instabilities in human cancers', *Nature,* 396(6712), pp. 643-9.

Lim, J. and Hotchin, N. A. (2012) 'Signalling mechanisms of the leukocyte integrin αMβ2: current and future perspectives', *Biol Cell,* 104(11), pp. 631-40.

Littlefield, J. W. (1969) 'Hybridization of hamster cells with high and low folate reductase activity', *Proc Natl Acad Sci U S A,* 62(1), pp. 88-95.

Loeb, L. A. (2011) 'Human cancers express mutator phenotypes: origin, consequences and targeting', *Nat Rev Cancer,* 11(6), pp. 450-7.

Lundberg, P., Karow, A., Nienhold, R., Looser, R., Hao-Shen, H., Nissen, I., Girsberger, S., Lehmann, T., Passweg, J., Stern, M., Beisel, C., Kralovics, R. and Skoda, R. C. (2014) 'Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms', *Blood,* 123(14), pp. 2220-8.

Luria, S. E. and Delbrück, M. (1943) 'Mutations of Bacteria from Virus Sensitivity to Virus Resistance', *Genetics,* 28(6), pp. 491-511.

Luzzatto, L. (2016) 'Recent advances in the pathogenesis and treatment of paroxysmal nocturnal hemoglobinuria', *F1000Res,* 5.

Luzzatto, L. and Notaro, R. (2003) 'Paroxysmal Nocturnal Hemoglobinuria', in Handin, R., Lux, S. & Stossel, T. (eds.) *Blood: Principles and Practice of Hematology.* 2nd ed. Philadelphia: Lippincot Williams & Wilkins, pp. 319-334.

Luzzatto, L. and Pandolfi, P. P. (2015) 'Causality and Chance in the Development of Cancer', *N Engl J Med,* 373(16), pp. 1579.

Lynch, M. (2010) 'Rate, molecular spectrum, and consequences of human mutation', *Proc Natl Acad Sci U S A,* 107(3), pp. 961-8.

Magni, G. E. and Von Borstel, R. C. (1962) 'Different Rates of Spontaneous Mutation during Mitosis and Meiosis in Yeast', *Genetics,* 47(8), pp. 1097-108.

McCarron, M. A., Kutlaca, A. and Morley, A. A. (1989) 'The HLA-A mutation assay: improved technique and normal results', *Mutat Res,* 225(4), pp. 189-93.

Meier, B. and Burton, J. H. (2014) 'Myeloproliferative disorders', *Emerg Med Clin North Am,* 32(3), pp. 597-612.

Miyata, T., Takeda, J., Iida, Y., Yamada, N., Inoue, N., Takahashi, M., Maeda, K., Kitani, T. and Kinoshita, T. (1993) 'The cloning of PIG-A, a component in the early step of GPI-anchor biosynthesis', *Science,* 259(5099), pp. 1318-20.

Morth, J. P., Pedersen, B. P., Buch-Pedersen, M. J., Andersen, J. P., Vilsen, B., Palmgren, M. G. and Nissen, P. (2011) 'A structural overview of the plasma membrane Na+,K+-ATPase and H+-ATPase ion pumps', *Nat Rev Mol Cell Biol,* 12(1), pp. 60-70.

Mutter-Rottmayer, E., Gao, Y. and Vaziri, C. (2016) 'Cancer cells activate damage-tolerant and error-prone DNA synthesis', *Mol Cell Oncol,* 3(6), pp. e1225547.

Nafa, K., Bessler, M., Castro-Malaspina, H., Jhanwar, S. and Luzzatto, L. (1998) 'The spectrum of somatic mutations in the PIG-A gene in paroxysmal nocturnal hemoglobinuria includes large deletions and small duplications', *Blood Cells Mol Dis,* 24(3), pp. 370-84.

Nicklas, J. A., O'Neill, J. P. and Albertini, R. J. (1986) 'Use of T-cell receptor gene probes to quantify the in vivo hprt mutations in human T-lymphocytes', *Mutat Res,* 173(1), pp. 67-72.

Novick, A. and Szilard, L. (1950) 'Experiments with the Chemostat on spontaneous mutations of bacteria', *Proc Natl Acad Sci U S A,* 36(12), pp. 708-19.

Oller, A. R., Rastogi, P., Morgenthaler, S. and Thilly, W. G. (1989) 'A statistical model to estimate variance in long term-low dose mutation assays: testing of the model in a human lymphoblastoid mutation assay', *Mutat Res,* 216(3), pp. 149-61.

Olsen, A. K., Dertinger, S. D., Krüger, C. T., Eide, D. M., Instanes, C., Brunborg, G., Hartwig, A. and Graupner, A. (2017) 'The Pig-a Gene Mutation Assay in Mice and Human Cells: A Review', *Basic Clin Pharmacol Toxicol,* 121 Suppl 3, pp. 78-92.

Ortolan, E., Augeri, S., Fissolo, G., Musso, I. and Funaro, A. (2018) 'CD157: From immunoregulatory protein to potential therapeutic target', *Immunol Lett*.

Peruzzi, B., Araten, D. J., Notaro, R. and Luzzatto, L. (2010) 'The use of PIG-A as a sentinel gene for the study of the somatic mutation rate and of mutagenic agents in vivo', *Mutat Res,* 705(1), pp. 3-10.

Phonethepswath, S., Bryce, S. M., Bemis, J. C. and Dertinger, S. D. (2008) 'Erythrocyte-based Pig-a gene mutation assay: demonstration of cross-species potential', *Mutat Res,* 657(2), pp. 122-6.

Pouyan, M. B., Jindal, V., Birjandtalab, J. and Nourani, M. (2016) 'Single and multi-subject clustering of flow cytometry data for cell-type identification and anomaly detection', *BMC Med Genomics,* 9 Suppl 2, pp. 41.

Rheinländer, A., Schraven, B. and Bommhardt, U. (2018) 'CD45 in human physiology and clinical medicine', *Immunol Lett,* 196, pp. 22-32.

Rondelli, T., Berardi, M., Peruzzi, B., Boni, L., Caporale, R., Dolara, P., Notaro, R. and Luzzatto, L. (2013) 'The frequency of granulocytes with spontaneous somatic mutations: a wide distribution in a normal human population', *PLoS One,* 8(1), pp. e54046.

Rosse, W. F. and Dacie, J. V. (1966) 'Immune lysis of normal human and paroxysmal nocturnal hemoglobinuria (PNH) red blood cells. I. The sensitivity of PNH red cells to lysis by complement and specific antibody', *J Clin Invest,* 45(5), pp. 736-48.

Rossman, T. G., Goncharova, E. I. and Nádas, A. (1995) 'Modeling and measurement of the spontaneous mutation rate in mammalian cells', *Mutat Res,* 328(1), pp. 21-30.

Rosti, V., Tremml, G., Soares, V., Pandolfi, P. P., Luzzatto, L. and Bessler, M. (1997) 'Murine embryonic stem cells without pig-a gene activity are competent for hematopoiesis with the PNH phenotype but not for clonal expansion', *J Clin Invest,* 100(5), pp. 1028-36.

Rumi, E. and Cazzola, M. (2017) 'Diagnosis, risk stratification, and response evaluation in classical myeloproliferative neoplasms', *Blood,* 129(6), pp. 680-692.

Russell, L. B. and Major, M. H. (1957) 'Radiation-Induced Presumed Somatic Mutations in the House Mouse', *Genetics,* 42(2), pp. 161-75.

Ryan, F. J. and Wainwright, L. K. (1954) 'Nuclear segregation and the growth of clones of spontaneous mutants of bacteria', *J Gen Microbiol,* 11(3), pp. 364-79.

Scott, L. M. and Rebel, V. I. (2012) 'JAK2 and genomic instability in the myeloproliferative neoplasms: a case of the chicken or the egg?', *Am J Hematol,* 87(11), pp. 1028-36.

Seshadri, R., Kutlaca, R. J., Trainor, K., Matthews, C. and Morley, A. A. (1987) 'Mutation rate of normal and malignant human lymphocytes', *Cancer Res,* 47(2), pp. 407-9.

Stewart, F. M., Gordon, D. M. and Levin, B. R. (1990) 'Fluctuation analysis: the probability distribution of the number of mutants under different conditions', *Genetics,* 124(1), pp. 175-85.

Stratton, M. R., Campbell, P. J. and Futreal, P. A. (2009) 'The cancer genome', *Nature,* 458(7239), pp. 719-24.

Sutherland, D. R., Acton, E., Keeney, M., Davis, B. H. and Illingworth, A. (2014) 'Use of CD157 in FLAER-based assays for high-sensitivity PNH granulocyte and PNH monocyte detection', *Cytometry B Clin Cytom,* 86(1), pp. 44-55.

Sutherland, D. R., Illingworth, A., Marinov, I., Ortiz, F., Andreasen, J., Payne, D., Wallace, P. K. and Keeney, M. (2018) 'ICCS/ESCCA consensus guidelines to detect GPI-deficient cells in paroxysmal nocturnal hemoglobinuria (PNH) and related disorders part 2 - reagent selection and assay optimization for high-sensitivity testing', *Cytometry B Clin Cytom,* 94(1), pp. 23-48.

Sutherland, D. R. and Musani, R. (2018) 'Re: Blaha J et al.: The monoclonal anti-CD157 antibody clone SY11B5, used for high sensitivity detection of PNH clones on WBCs, fails to detect a

common polymorphic variant encoded by BST-1', *Cytometry B Clin Cytom*.

Takeda, J., Miyata, T., Kawagoe, K., Iida, Y., Endo, Y., Fujita, T., Takahashi, M., Kitani, T. and Kinoshita, T. (1993) 'Deficiency of the GPI anchor caused by a somatic mutation of the PIG-A gene in paroxysmal nocturnal hemoglobinuria', *Cell,* 73(4), pp. 703-11.

Tates, A. D., Bernini, L. F., Natarajan, A. T., Ploem, J. S., Verwoerd, N. P., Cole, J., Green, M. H., Arlett, C. F. and Norris, P. N. (1989) 'Detection of somatic mutants in man: HPRT mutations in lymphocytes and hemoglobin mutations in erythrocytes', *Mutat Res,* 213(1), pp. 73-82.

Tefferi, A. and Barbui, T. (2017) 'Polycythemia vera and essential thrombocythemia: 2017 update on diagnosis, risk-stratification, and management', *Am J Hematol,* 92(1), pp. 94-108.

Tomasetti, C. and Vogelstein, B. (2015) 'Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions', *Science,* 347(6217), pp. 78-81.

Tomasetti, C., Vogelstein, B. and Parmigiani, G. (2013) 'Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation', *Proc Natl Acad Sci U S A,* 110(6), pp. 1999-2004.

Vogelstein, B. and Kinzler, K. W. (2004) 'Cancer genes and the pathways they control', *Nat Med,* 10(8), pp. 789-99.

Wu, X., Strome, E. D., Meng, Q., Hastings, P. J., Plon, S. E. and Kimmel, M. (2009) 'A robust estimator of mutation rates', *Mutat Res,* 661(1-2), pp. 101-9.

Yamamoto, T., Shichishima, T., Shikama, Y., Saitoh, Y., Ogawa, K. and Maruyama, Y. (2002) 'Granulocytes from patients with paroxysmal nocturnal hemoglobinuria and normal individuals have the same sensitivity to spontaneous apoptosis', *Exp Hematol,* 30(3), pp. 187-94.

Yoon, J., Terada, A. and Kita, H. (2007) 'CD66b regulates adhesion and activation of human eosinophils', *J Immunol,* 179(12), pp. 8454-62.

Zheng, Q. (1999) 'Progress of a half century in the study of the Luria-Delbrück distribution', *Math Biosci,* 162(1-2), pp. 1-32.

Zimran, E., Hoffman, R. and Kremyanskaya, M. (2018) 'Current approaches to challenging scenarios in myeloproliferative neoplasms', *Expert Rev Anticancer Ther,* 18(6), pp. 567-578.

APPENDIX

Development of a new staining panel for the flow cytometry PIGA mutant assay

The panel of antibodies used in the original *PIGA* mutant assay was developed in 1999 during studies on PNH, to detect rare GPI-deficient granulocytes in the blood of healthy subjects (Araten et al., 1999). Therefore, the two GPI-anchored proteins chosen as target for the staining (CD55 and CD59) were those used at that time for the diagnosis of PNH. Later on, it was decided to increase the number of GPI-linked proteins to three by including CD24 in the analysis (Rondelli et al., 2013). In the present study, other GPI-anchored proteins and different combination of antibodies have been tested and compared to identify a staining strategy that, possibly, could result in a better discrimination between GPI-positive and GPI-negative cells. After performing a literature research, two interesting GPI-linked proteins expressed by granulocytes were identified (Hernández-Campo et al., 2006): CD66b (Alfinito et al., 1996) and CD157 (Sutherland et al., 2014). CD66b, also known as Carcinoembryonic Antigen-related Cell Adhesion Molecule 8 (CEACAM8), is a member of the immunoglobin (Ig) superfamily and more specifically of the CEA family. It is specifically expressed by granulocytes and is recognized as a granulocyte “activation marker” because its expression on the cell surface is up-regulated during cell activation (Kuroki et al., 2001, Yoon et al., 2007). CD157, which is also known as Bone marrow Stromal cell antigen 1 (BST1), is a glycoprotein receptor that belongs to the ADP-ribosyl Cyclases (ARC) gene family. CD157 is prevalently expressed by cells of the myelomonocytic lineage, mainly monocytes and neutrophils, and it has been found also in a variety of other cell types (Funaro et al., 2004, Ortolan et al., 2018).

It is of notice that both anti-CD66b and anti-CD157 are included in the recently published guidelines for the diagnosis of PNH (Sutherland et al., 2018).

Single and double staining of GPI-linked proteins

To preliminary test the capacity of each antibody to distinguish between GPI-positive and GPI-negative cells, we used the peripheral blood of PNH patients with two clear granulocyte populations. Since there was no need to obtain highly purified granulocytes, 100 µl of PNH peripheral blood were used for each test. Multiple amounts of blood were mixed and incubated with 10-15 ml of cold NH_4Cl 0.157 M - $KHCO_3$ 0.01 M - EDTA 100 mM solution for ~10 minutes to lyse red blood cells. Samples were then centrifuged at 300 x *g* for 5 minutes at 4°C and the supernatant was discarded. Pellets were re-suspended in a proper amount of cold PBS (depending on the number of samples to be stained) and aliquoted into polystyrene tubes, which were centrifuged at 300 x *g* for 5 minutes at 4°C. The supernatant was discarded and pellets were stained with excess amounts of each one or of combinations of any of the following antibodies against GPI-anchored proteins: 5 µl of PE mouse anti-human CD59 (clone MEM-43), 5 µl of PE mouse anti-human CD55 (clone IA10), 5 µl of PE mouse anti-human CD24 (clone ML5), 5 µl of PE mouse anti-human CD66b (clone G10F5) and 5 µl of PE mouse anti-human CD157 (clone SY/11B5[1]). Negative controls were stained with 5 µl of PE Mouse IgG1, κ

[1] Recently, it has been reported that a common polymorphism (p.Arg145Gln) in exon 3 of *BST1* gene (overall homozygous frequency of 0.27%) prevents the binding of the clone SY/11B5 of the anti-CD157 antibody (Blaha et al., 2018). In addition, it has been also reported that, occasionally, apparently CD157- negative cells are present irrespectively of the above polymorphism (Sutherland and Musani, 2018). The failure of a specific clone of a monoclonal antibody in binding its target because of a genetic variation is a general problem. However, in the specific case of our *PIGA* mutant assay this kind of failure case is relatively irrelevant because, for the sake of robustness of our test, we target multiple GPI-linked proteins. This greatly reduces the chance to count a false GPI-negative cell.

Isotype Control. Antibodies were added directly onto the pellets, which were then carefully mixed by flicking the tubes. The incubation was performed for 30 minutes at 4°C. After incubation, cells were re-suspended in cold PBS and centrifuged at 300 x *g* for 5 minutes at 4°C. Supernatant was discarded, pellets were re-suspended in ~ 30 µl of cold PBS and stored at 4°C until analyzed by flow cytometry.

In these preliminary experiments, living granulocytes were selected only on the basis of physical features (SSC and FSC). The expression of GPI-anchored proteins was then assessed to evaluate the capacity of each antibody, alone or in combination with others, to discriminate between GPI-positive and GPI-negative granulocytes.

Examples of single staining results are reported in Figure 19 and Figure 20, whereas histograms in Figure 21 represent PE-fluorescence resulting from the combination of two antibodies specific for GPI-anchored proteins.

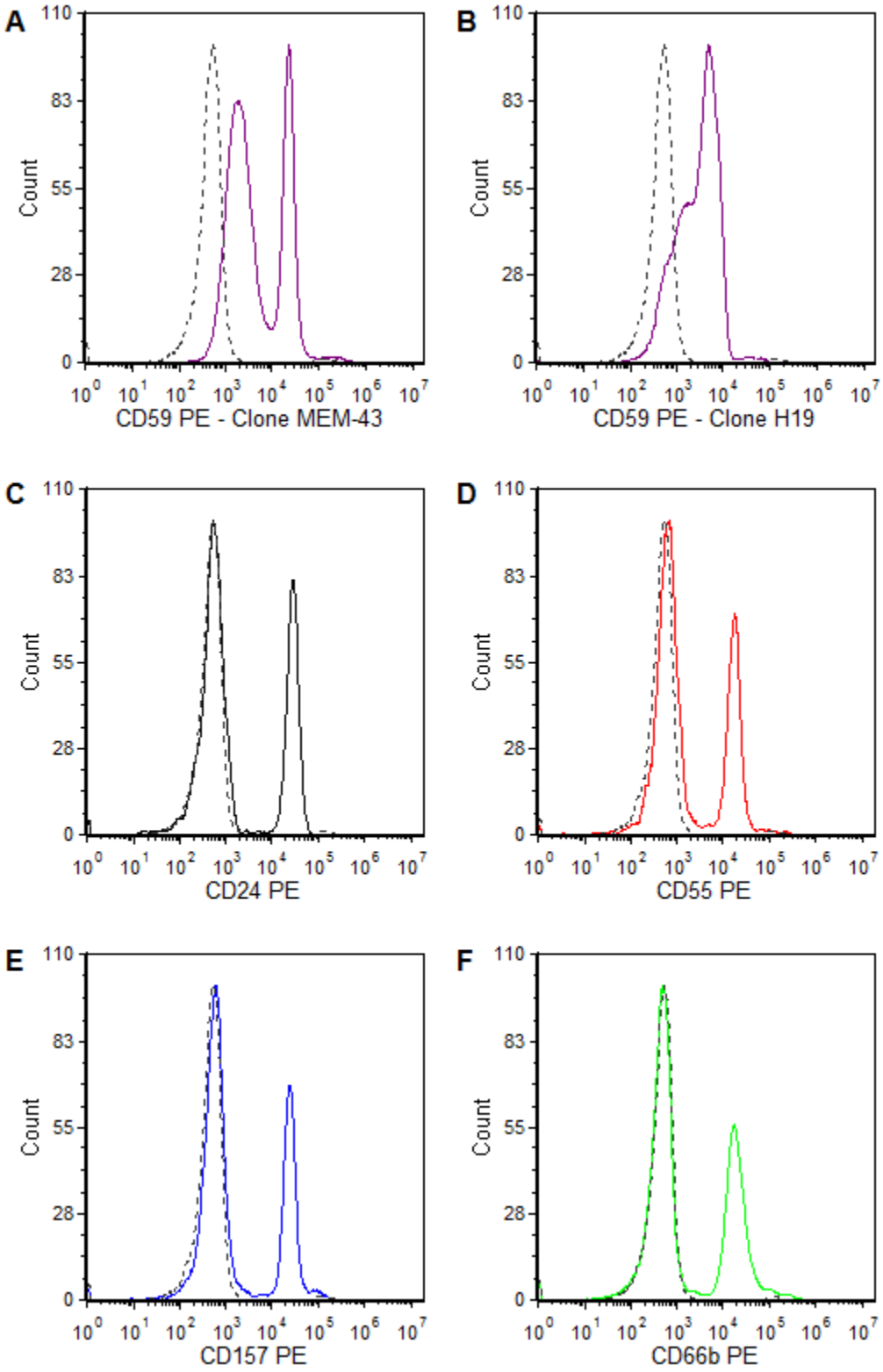

Figure 19. Single staining histograms.
Negative controls are represented as dashed gray histograms.

In these preliminary experiments, staining of cells with the Bio-Rad PE mouse anti-human CD59, clone MEM-43, which was previously included in the standard panel of antibodies, repeatedly yielded not satisfactory results. Indeed, the fluorescence signal of the GPI-negative population was always higher than negative controls and a substantial background of cells with intermediate levels of fluorescence did not allow a clear separation between GPI-positive and GPI-negative cells (Figure 19A). In order to check whether the problem was related to the specific anti-CD59 antibody, the BD Pharmingen PE mouse anti-human CD59, clone p282 (H19), was tested as well with the same procedure. In this case, results were even worse because there was no separation between GPI-positive and GPI-negative populations of granulocytes since fluorescence ranged from low to high as a continuum (Figure 19B). Therefore, it was decided to discard anti-CD59 from the panel of antibodies further investigated (Figure 20).

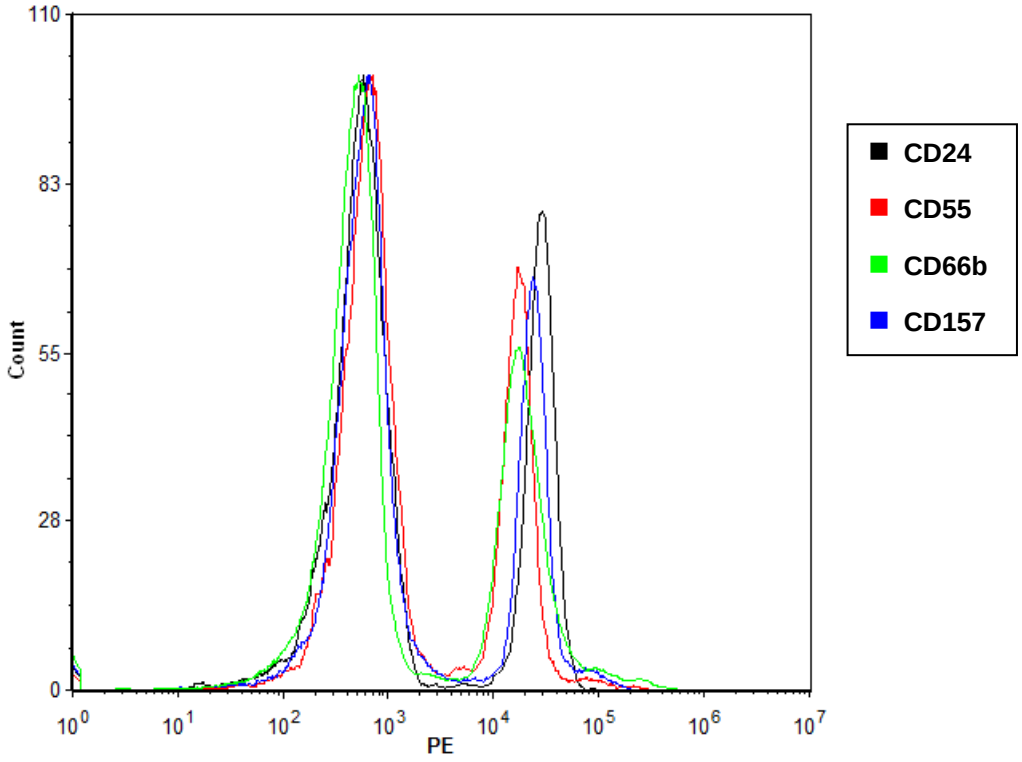

Figure 20. Overlays of single staining histograms without anti-CD59.

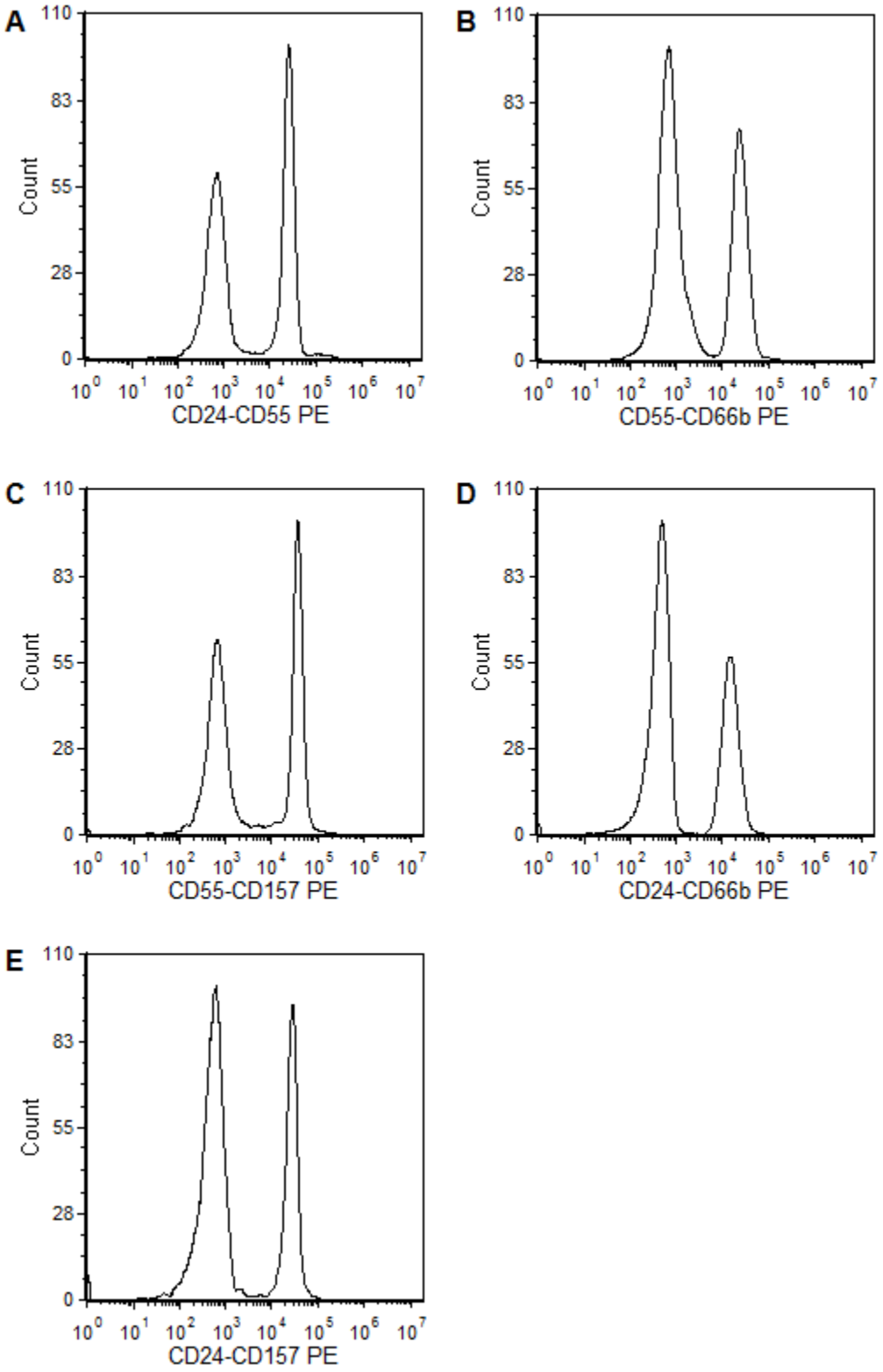

Figure 21. Double staining histograms.

Single stainings with the other four antibodies were good. However, fluorescence levels of GPI-negative granulocytes stained with PE mouse anti-human CD55 were slightly higher than those obtained with the other antibodies, therefore the GPI-positive and the GPI-negative populations were closer (Figure 19B and 20); moreover, the background of cells with intermediate levels of fluorescence between the two populations was bigger in both single (Figure 19B and 20) and double (Figure 21A, B and C) stainings with anti-CD55. For these reasons we decided to discard also the anti-CD55 antibody and to continue testing the other three.

Comparison between the standard and the new panel of antibodies

Tests to compare the combination of antibodies specific for GPI-linked proteins used hitherto (indicated as “standard”) (Figure 22A) with the proposed new panel (indicated as “new”) (Figure 22B) were subsequently performed.

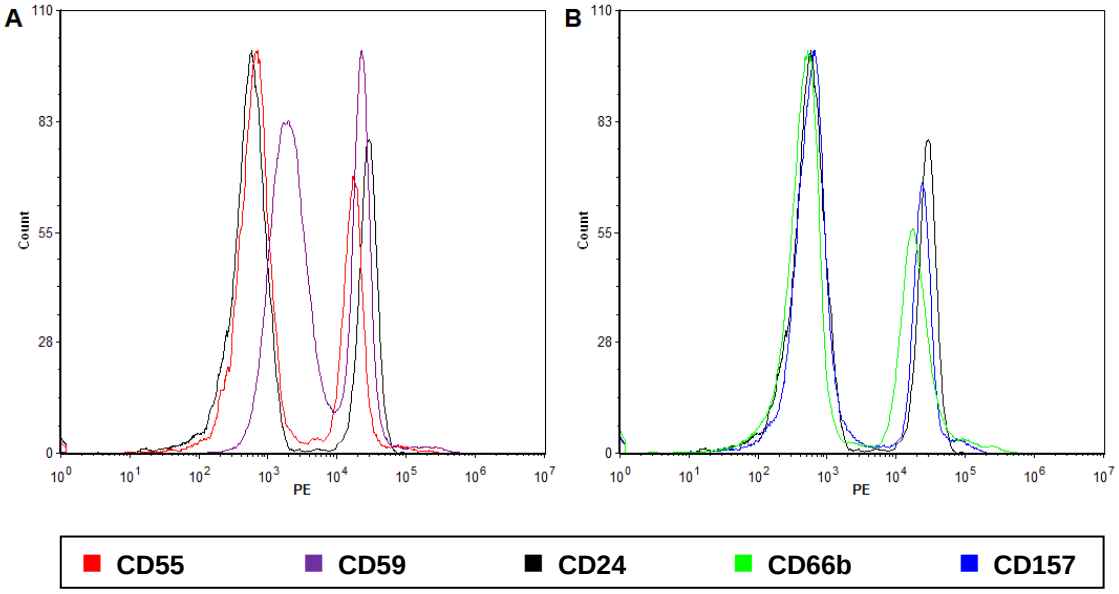

Figure 22. Overlays of single staining histograms: A) standard staining; B) new staining.

For preliminary analyses, granulocytes were purified from 6 ml of peripheral blood of PNH patients with known percentages of GPI-positive and GPI-negative granulocytes, as described in “Materials and methods”. Two aliquots containing 0.5 - 1 x 10^6 granulocytes for each PNH patient were stained in parallel with the following combination of antibodies:

1) standard staining: PE mouse anti-human CD24 (clone ML5); PE mouse anti-human CD55 (clone IA10); PE mouse anti-human CD59 (clone MEM-43); APC mouse anti-human CD45 (clone HI30) and FITC mouse anti-human CD11b (clone M1/70.15.11.5);

2) new staining: PE mouse anti-human CD24 (clone ML5); PE mouse anti-human CD66b (clone G10F5); PE mouse anti-human CD157 (clone SY/11B5); APC mouse anti-human CD45 (clone HI30) and Pacific Blue (PB) mouse anti-human CD11b (clone Bear1). The decision to change the antibody specific for CD11b was taken just because the current availability on the market of a wider choice of antibodies conjugated to fluorochromes, compared with the past, allowed to select a fluorescent dye whose emission spectrum almost did not interfere with that of the PE antibodies against GPI-linked proteins in use, thus requiring only minimal compensation adjustments. A significant example of comparison between the two staining strategies, performed on granulocytes purified from the peripheral blood of a PNH patient, is represented in Figure 23. In flow cytometry analysis, cells were firstly selected on the basis of physical parameters (SSC and FSC) and accordingly to the positive expression of CD45; CD11b and GPI-anchored proteins expression was then assessed. Both panels of antibodies allowed to distinguish two comparable populations of GPI-positive and GPI-negative granulocytes, but the combination of antibodies currently in use gave a much higher background of cells with intermediate levels of PE fluorescence (Figure 23A) compared with the new panel under investigation (Figure 23B), thus confirming single and double staining results.

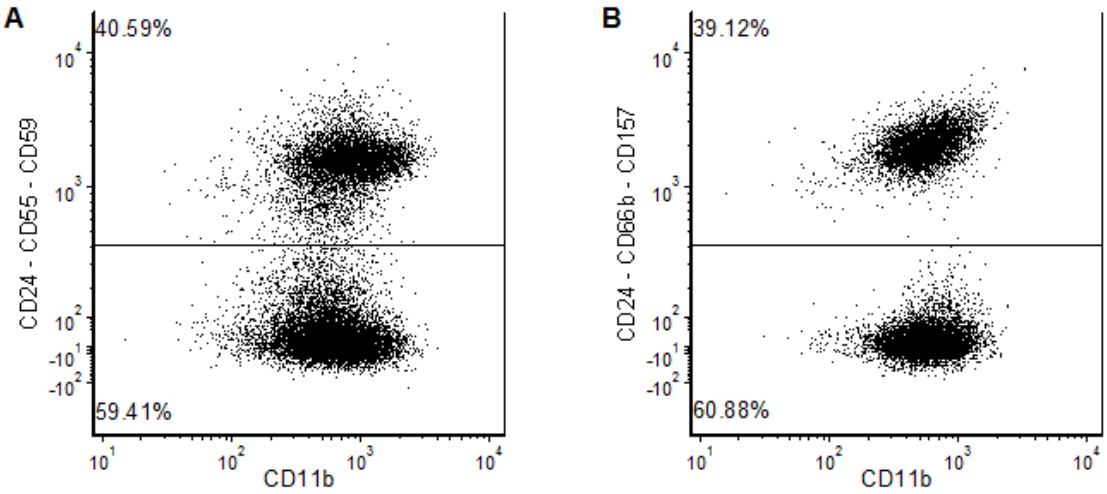

Figure 23. Dot plots of peripheral blood granulocytes from a PNH patient: A) standard staining; B) new staining.

Spiked controls

To further compare the two panel of antibodies, three experiments with artificial controls were performed. To prepare artificial controls, granulocytes were purified from the peripheral blood of PNH patients with >90% of GPI-negative cells and counted; known percentages of GPI-negative granulocytes were then spiked into granulocyte aliquots purified from healthy subjects. Once prepared, spiked controls were divided in two aliquots, stained in parallel with the two combinations of antibodies and analyzed by flow cytometry, as described previously. Mutant frequency (f) was calculated as the number of GPI-negative granulocytes detected per 10^6 cells.

In the first experiment, three aliquots of $6x10^6$ healthy granulocytes were mixed with 0.75%, 1.5% and 2.25% of GPI-negative granulocytes. In the samples stained with the standard combination of antibodies, measured f values were 7879.1, 16590.2 and 23084, respectively; those obtained in the samples stained with the new panel of antibodies were 7757.2, 16335.9 and 24521, respectively (Table 2).

In the second experiment, two sets of spiked controls were made. In the first set, 0.2% and 1% of GPI-deficient granulocytes were added to two aliquots of 10^6 normal granulocytes; in the second set, 2.5% and 5% of GPI-negative granulocytes were mixed with $4x10^5$ normal granulocytes. In the samples stained with the standard panel of antibodies, assessed *ƒ* values were 2235.2, 10904.2, 27149.9 and 50300.1, respectively; those obtained in the samples stained with the new combination of antibodies were 2405.8, 10944.7, 29680.5 and 58417, respectively (Table 2).

In the third experiment, 0,001% and 0,005% of GPI-negative granulocytes were spiked into two aliquots of $3x10^6$ normal granulocytes. In the samples stained with the standard combination of antibodies, measured *ƒ* values were 12.3 and 36.2, respectively; those obtained in the samples stained with the new panel of antibodies were 13.3 and 44.9, respectively (Table 2).

Expected mutant frequency ($x10^{-6}$)	**Measured mutant frequency ($x10^{-6}$), standard staining**	**Measured mutant frequency ($x10^{-6}$), new staining**
10 (0.001%)	12.3	13.3
50 (0.005%)	36.2	44.9
2000 (0.2%)	2235.2	2405.8
7500 (0.75%)	7879.1	7757.2
10000 (1%)	10904.2	10944.7
15000 (1.5%)	16590.2	16335.9
22500 (2.25%)	23084	24521
25000 (2.5%)	27149.9	29680.5
50000 (5%)	50300.1	58417

Table 2. Comparison between *ƒ* values measured in spiked controls stained in parallel with the standard and the new staining.

Correlation between *f* values assessed in samples stained in parallel with the two combination of antibodies was very good (R^2 = 0.9965, p<0.0001) (Figure 24), thus confirming that the two staining strategies yield comparable results.

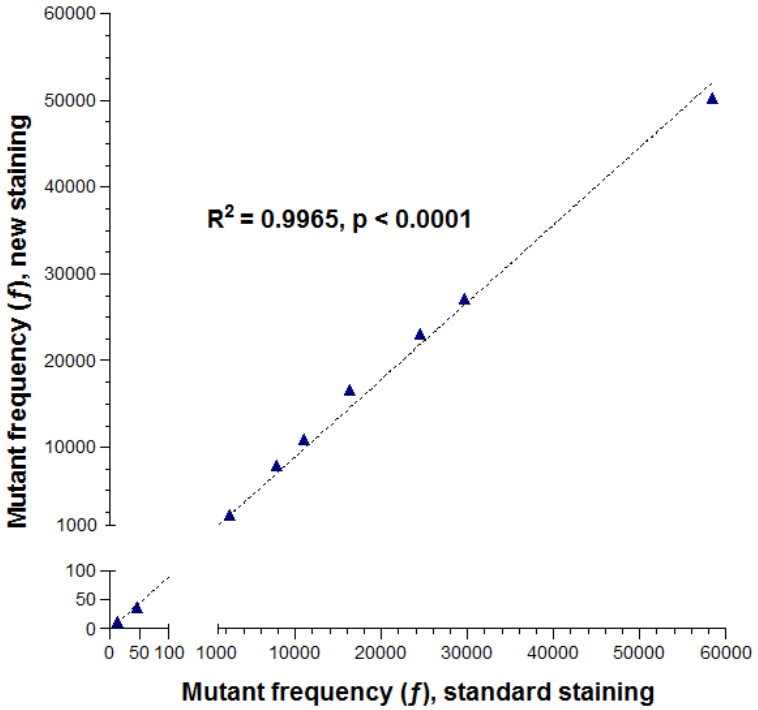

Figure 24. Correlation between *f* values measured in spiked controls stained in parallel with the standard and the new staining.

Comments and conclusions

This investigation of single and multiple stainings for granulocyte GPI-linked markers has resulted in few findings that could potentially improve the flow cytometry *PIGA* mutant assay.

Both anti-CD55 and, especially, anti-CD59 have not performed well in discriminating GPI-positive and GPI-negative granulocytes because they produced an intermediate intensity fluorescence in a small, but potentially significant, fraction of cells. This background of intermediately stained granulocytes is not found with other antibodies specific for GPI-linked proteins, such as CD24, CD66b and CD157.

In addition, we found that the FITC-labeled anti-CD11b antibody, whose emission partially overlaps with the PE signal of antibodies against GPI-linked proteins, can be effectively replaced with a PB anti-CD11b that is now commercially available.

Therefore, we propose for the flow cytometry *PIGA* mutant assay a new panel of antibodies slightly different from that previously developed (Araten et al., 1999, Rondelli et al., 2013). This new panel includes PE-labeled antibodies specific for the GPI-linked proteins CD24, CD66b and CD157, APC-labeled anti-transmembrane CD45 and PB-labeled anti-transmembrane CD11b.

This panel yielded comparable results to the standard panel on peripheral blood granulocytes from PNH patients and on artificial spiked controls containing known amounts of GPI-negative granulocytes. However, the new panel of antibodies has allowed the better discrimination of GPI-positive and GPI-negative cells by reducing the number of events with intermediate GPI-related fluorescence. Therefore, not only it was proven able to provide clearer results, but it may also allow to overcome the use of the 4% of the geometric mean of the GPI-linked fluorescence values of all events from each experiment, which was previously adopted as cut-off to discriminate GPI-positive and GPI-negative cell populations (Peruzzi et al., 2010). Indeed, even though such a cut-off was certainly a useful tool to

standardize flow cytometry data analyses, in some cases it turned out to be not appropriate because it divided in two otherwise well-defined cell populations.

Our thoroughly analysis of different suitable granulocyte GPI-linked markers has resulted in the definition of a new panel of antibodies for the flow cytometry *PIGA* mutant assay. This new combination of antibodies, once validated in a proper comparison study, should improve the discrimination between GPI-positive and GPI-negative granulocytes and may allow the adoption of automate clustering strategies, which are being increasingly developed to overcome the manual gating step that is one of the major weak points of flow cytometry data analyses (Pouyan et al., 2016).