GENETIC STUDIES OF HEREDITARY THROMBOCYTHEMIA

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Annalisa Pianta
aus Santo Stefano Ticino, Milano
Italy

Basel, 2013

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Radek Skoda

Prof. Christoph Handschin

Basel, den 16 Oktober 2012

Prof. Dr. Jörg Schibler
Dekan der Philosophisch-
Naturwissenschaftlichen
Fakultät

TABLE OF CONTENTS

ABBREVIATION

- 5'UTR: 5 prime untranslated region
- BFU-MK: burst-forming unit-megakaryocyte
- BM: bone marrow
- C: cysteine
- CDKI: cyclin-dependent kinase inhibitor
- CFU-MK: colony-forming unit-megakaryocyte
- CLP: common lymphoid precursor
- CMP: common myeloid precursor
- CRM: cytokine receptor homology module
- CXCR4: CXC chemokine receptor 4
- DMS: demarcation membrane system
- DNA: deoxyribonucleic acid
- ECM: extracellular matrix
- EPO: erythropoietin
- ET: essential thrombocythemia
- ETP: earliest thymic progenitors
- FGF4: fibroblast growth factor 4
- FOG: friend of GATA
- G: glycine
- GMP: granulocyte−macrophage progenitor
- GSN: gelsolin
- HGF: hematopoietic growth factor
- HSC: hematopoietic stem cell
- HT: hereditary thrombocytosis
- IBD: increased identical by descendent
- IFNγ: interferon gamma
- IL: interleukin
- JAK: janus kinase
- K: lysine
- M: methionine
- MAP: mitogen activated protein
- MEP: MK/Erythro progenitor
- MK: megakaryocyte
- MPD: myeloproliferative disorder
- MPN: myeloproliferative neoplasm

- MPP: multipotent progenitor
- N: asparagine
- NF-E2: nuclear factor erythroid 2
- NGS: next generation sequencing
- NO: nitric oxide
- PCR: polymerase chain reaction
- PFCP: primary familial and congenital polycythemia
- PI3K: phosphatidylinositol 3 kinase
- RFLP: restriction fragment length polymorphism
- RGS16: regulator of G protein signaling
- SCF: stem cell factor
- SDF1: stromal cell-derive factor 1
- SIFT: sorting tolerant from tolerant
- SNP: single nucleotide polymorphism
- SOCS: suppressor of cytokine signaling
- STAT: signal transducer and activator of transcription
- STR: short tandem repeat
- TF: transcription factor
- THPO: thrombopoietin gene
- TNFα: Tumor Necrosis Factor alpha
- TPO: thrombopoietin
- uATG: upstream initial codon
- uORF: upstream open reading frame
- V: valine
- Y: tyrosine

1. SUMMARY

Hereditary thrombocythemia (HT) is a familial myeloproliferative disorder characterized by an elevated platelet count in peripheral blood. Thrombocytosis is due to a genetic alteration that can be transmitted to the offspring. Recently, major progress has been made in understanding the biology of HT thanks to the discovery of mutations in two genes: thrombopoietin (THPO) and its receptor MPL. Interestingly, the analysis of these mutations has provided more insights in the physiological regulation of platelet homeostasis [1]. However, not all the HT pedigrees carry mutations in the THPO or MPL genes. In 80-90% of these pedigrees the disease-causing gene remains unknown [2] and it is likely that hereditary thrombocytosis can be caused by alterations in other genes, not yet identified.

The focus of my PhD studies was a large US family affected by HT. In this family, THPO and MPL were excluded as disease-causing genes. Therefore, genome-wide linkage analysis was performed to identify co-segregating regions shared by the affected family members as a target in search of possible candidate mutations responsible for the thrombocytosis phenotype. One region with significant logarithm of odds (LOD) score values has been located using microsatellites and SNP chip arrays. One novel candidate mutation was found in the gelsolin gene by next generation sequencing and confirmed by capillary sequencing in all the 12 affected family members. Gelsolin is a Ca^{2+} regulated actin filament severing, capping and nucleating protein abundant in platelets. It is involved in the regulation of cell structure and metabolism. Interestingly, it has a key-role in apoptosis regulation and modulation of platelets. Computational predictions showed that this alteration can probably affect protein function and the structural analysis indicated that the alteration is located at the interface with actin. The platelets-biogenesis *in vitro* assay showed that the candidate alteration can increase the release of platelets-like particles in DAMI cell line stably transfected with the mutant gelsolin. To study the *in vivo* role of the candidate mutation in the pathogenesis of HT, different mouse models have been established. In lethally irradiated recipient mice transplanted with BM cells transduced with retrovirus expressing the human mutant gelsolin, variations

in platelet counts in peripheral blood have been observed. Transgenic mice expressing the human mutant gelsolin were generated to fully characterize the new discovered alteration. These mice developed a tendency to elevated platelet counts compared to their wild type littermates.

Taken together, these data illustrate the discovery of a new candidate mutation associated with the pathogenesis of HT. Until now, mutations in gelsolin gene were never described except for a mutant (D187N/Y) plasma gelsolin responsible for familial amyloidosis of Finnish type (FAF). My work contributed to further characterize this gene and to link it with the pathogenesis of HT.

2. INTRODUCTION

2.1 HEMATOPOIESIS

Hematopoiesis is the lifelong process by which all the blood cells are produced. This dynamic process is highly regulated in order to fulfil the requirements of the body for the transport of oxygen, blood coagulation and immune response.

In humans, hematopoiesis starts in the yolk sac during the first few weeks of gestation and then moves to fetal liver and spleen until 6 to 7 months, when the bone marrow (BM) becomes the main site of blood production and remains the major source of new blood cells throughout normal life [3].

Hematopoiesis is a hierarchical system. At the top, hematopoietic stem cells (HSCs) reside in specialized microenviroments, known as bone marrow niche. In humans, HSCs are defined by the expression of the CD34, CD133, Thy-1 and c-Kit antigens and the absence of lineage markers and CD38 [4]. HSCs are maintained in a quiescent state in order to minimize stresses due to cellular respiration and genome replication helping to HSC longevity and function [5]. In this way, HSCs can persist for a lifetime and give rise to progenitor cells that become increasingly lineage restricted and ultimately differentiate into all lineages of mature blood cells. Thanks to its self-renewal capability, the pool of HSC remains constant in a normal healthy steady state. HSCs reside at the apex of hematopoietic hierarchy and they are connected to mature cells by a complex roadmap of progenitor intermediates. HSCs differentiate into multipotent progenitors (MPPs) that further segregate along two fundamental branches: myeloid and lymphoid. The earliest myelo-lymphoid split gives rise to common myeloid precursors (CMP) and immature lymphoid precursors (MLP) and each of these undergo further commitment steps. CMPs give rise to GMPs, which become committed to the granulocyte-monocyte fate, and MEPs, which only produce erythroid and megakaryocyte cells. On the lymphoid side, MLPs give rise to B and NK cell precursors and the earliest thymic progenitors (ETPs) committed to the T lineage (Fig.1). The molecular mechanisms that regulate the balance between the self-renewal and differentiation are typically associated with changes in gene expression and are driven by transcription factors [6].

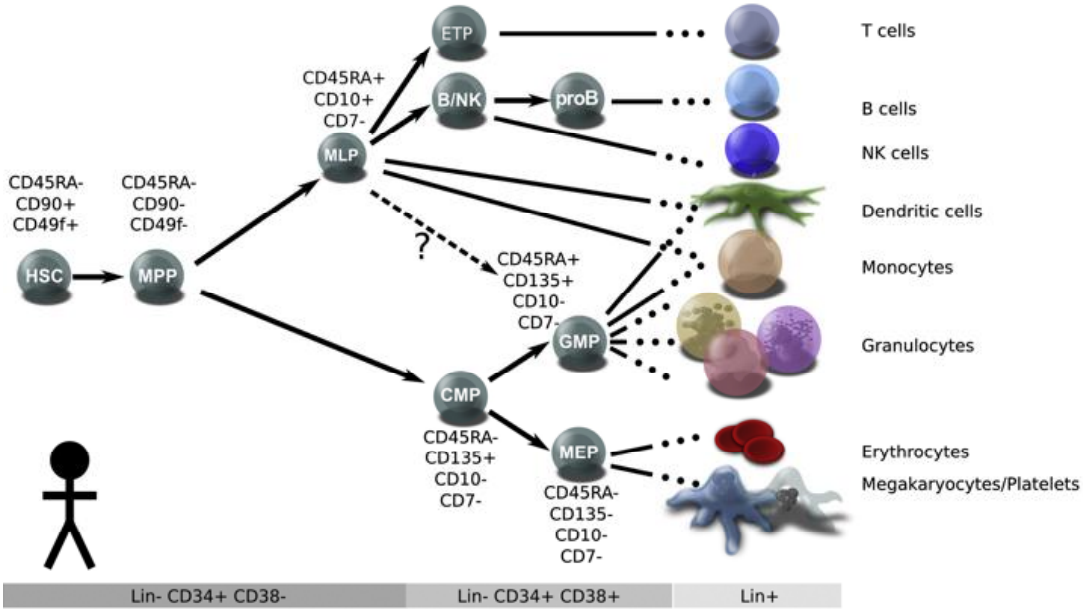


Fig.1: The hierarchy of hematopoietic cells in human. HSCs are defined by the expression of the CD34 and the absence of lineage markers and CD38. MPP, multipotent progenitor; CMP, common myeloid progenitor; MEP, megakaryocyte/erythroid progenitor; GMP, granulocyte−macrophage progenitor; MLP, immature lymphoid progenitors; ETP, earliest thymic progenitors. Adapted from [6]

Hematopoiesis is tightly regulated through interactions between progenitor cells and various growth factors. Hematopoietic growth factors are key external regulators of HSCs. They sustain survival, proliferation, differentiation and maturation of hematopoietic cells at all stages. These growth factors are glycoprotein hormones that can act locally at the site where they are produced or circulate in plasma. They also bind to the extracellular matrix to form niches to which stem and progenitor cells can adhere. The biological activity of growth factors is mediated by specific receptors on the cell surface. Most of these receptors are from the haematopoietin receptor superfamily. Upon the binding with their ligand, the receptors dimerize and give rise to a series of intracellular signal transduction pathway. The major ones are: JAK/STAT, the mitogen activated protein (MAP) kinase and the phosphatidylinositol 3 kinase (PI3K) pathways [3].

2.2 MEGAKARYOCYTOPOIESIS

Megakaryocytopoiesis is the process that leads to the production of platelets. It involves the commitment of hematopoietic stem cells and the proliferation, maturation and terminal differentiation of megakaryocytic progenitors. This process is

characterized by DNA endoreduplication, cytoplasmic maturation and expansion, and release of cytoplasmic fragments as circulating platelets. An overview of megakaryocyte (MK) production of platelets is shown in figure 2 (Fig.2).

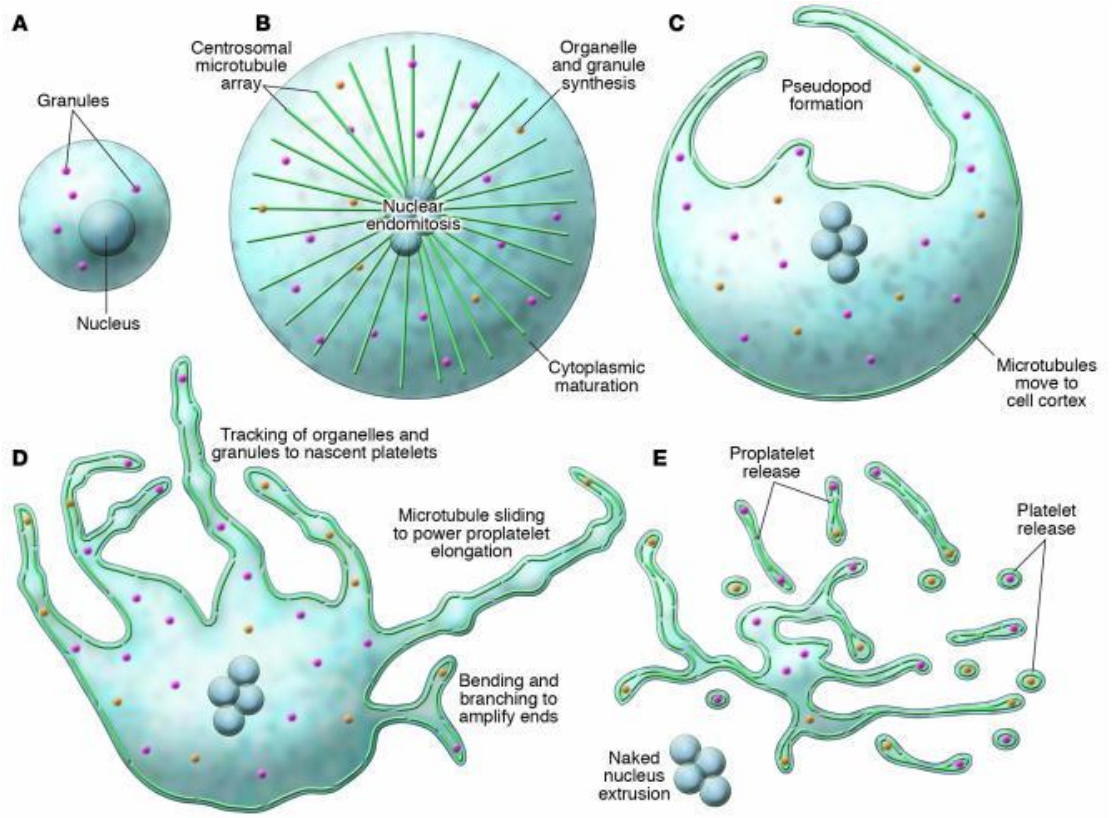


Fig.2: Overview of platelets biogenesis. As megakaryocytes transition from immature cells (**A**) to released platelets (**E**), a systematic series of events occurs. (**B**) The cells first undergo nuclear endomitosis, organelle synthesis, and cytoplasmic maturation and expansion, while a microtubule array, emanating from centrosomes, is established. (**C**) Prior to the onset of proplatelet formation, centrosomes disassemble and microtubules translocate to the cell cortex. Proplatelet formation initiates with the development of thick pseudopods. (**D**) Sliding of overlapping microtubules drives proplatelet elongation as organelles are tracked into proplatelet ends, where nascent platelets assemble. Proplatelet formation continues to expand throughout the cell while bending and branching amplify existing proplatelet ends. (**E**) The entire megakaryocyte cytoplasm is converted into a mass of proplatelets, which are released from the cell. The nucleus is eventually extruded from the mass of proplatelets, and individual platelets are released from proplatelet ends. [7]

Each day in every human, approximately 1×10^{11} platelets are produced. It has been estimated that a mature MK release around 2000-3000 platelets [8]. Platelets are anucleated cells that play an essential role in thrombosis and hemostasis. Blood platelets appear as small oval discs with average size of 2.5 x 5.0 μm. Platelets contain multiple cellular organelles including mitochondria, lysosomes, granules and peroxisomes.

2.2.1 From MK/Erythro progenitor (MEP) to MK

The MK and erythroid lineages are closely linked and have a common bipotent progenitor called MEP. In humans, the MEP is defined as Lin^{-} $CD34^{+}$ $CD38^{+}$ $IL3Ra^{-}$ $D45RA^{-2}$. This bipotent progenitor commits to the MK lineage through MK progenitors capable of proliferating and to give rise *in vitro* to MK colonies. BFU-MK are the most primitive MK-committed progenitors with the highest aptitude to proliferate. CFU-MK have a lower proliferation capacity. In humans, MK progenitors have been characterized by the presence of the surface markers CD34, CD31 and the CD133. CD41a (α integrin chain gpIIb/IIIa complex) and CD41b (β integrin chain of glycoprotein IIb) are specific for the MK lineage. Only about 3% of the CD34+ marrow cells express CD41. This fraction of cells includes MK progenitors, but does not contain all the CFU-MK (Fig.3). CD41 expression precedes CD42, the expression of which corresponds to a later differentiation step and correlates with the presence of other molecules such as Mpl, GPVI (collagen receptor), $\alpha 1\beta 2$ integrin, CD36 and proteins contained in the α-granules (PF4, vWF) [9].

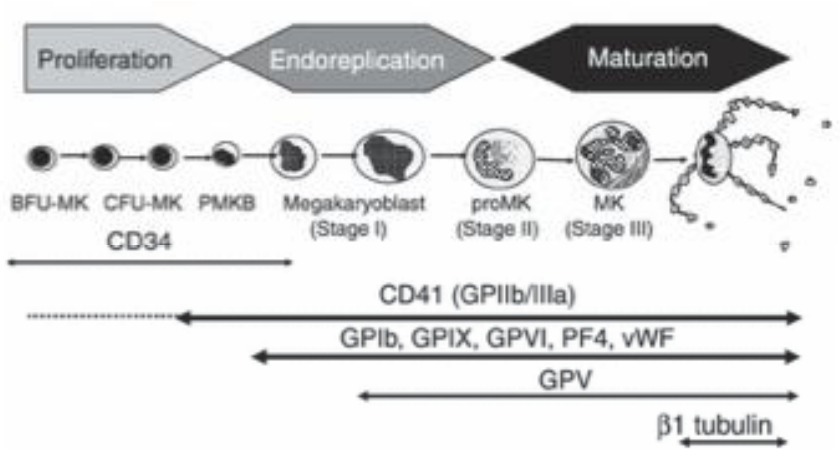


Fig.3: expression of differentiation markers along the human megakaryocytic differentiation. BFU-MK, burst-forming unit-megakaryocyte; CFU-MK, colony-forming unit-megakaryocyte; PMKB, pro-megakaryoblast; proMK, pro-megakaryocyte; MK, megakaryocyte. Adapted from [9].

2.2.2 Megakaryocytes development and platelets formation

After the initiation of the synthesis of platelet proteins, the MK precursor (also called promegakaryoblast) begins to enlarge and to increase its ploidy through a process of *endomitosis*. The MK is one of the rare cells that are polyploid during normal

differentiation. MKs can stop DNA duplication at any stage between 2N and 64N, and possibly 128N. In humans, the modal ploidy is 16N (about 50% of MKs). Interestingly, the MK ploidy increases during ontogeny with low ploidy MKs in primitive hematopoiesis (embryos) compared to adult MKs [10]. An endomitosis, similarly to mitosis, begins with a duplication of the centrosomes; a normal prophase with development of a mitotic spindle, chromatin condensation and the rupture of the nuclear envelope; alignment of the chromosomes on the equatorial plate occurs during the metaphase and finally the sister chromatids separate at anaphase. The spindle of a poliploid MK is multipolar with the number of poles corresponding to the ploidy level. The endomitosis fails of late cytokynesis leading to the formation of a MK containing a single nucleus with a single nuclear membrane. In MKs the cell cycle is composed of a succession of G1, S, G2 and M phases, but the M phase is incomplete. After M phase, MKs re-enter into G1 to initiate a subsequent cell cycle [9].

In addition to expansion of DNA, megakaryocytes experience significant maturation as internal membrane system, granules and organelles. The invagination of the MK plasma membrane creates a demarcation membrane system (DMS) that constitutes a membrane reservoir for platelets formation. The DMS associates with both the microtubules and the actin filaments and it is evaginated to form pseudopodal processes during proplatelet formation. Platelets are formed by fragmentation of the proplatelet protrusion and this process occurs in the blood circulation. Together with the DMS, microtubules and acto-myosin complex are the main determinants of platelets shedding [11]. Microtubules constitute the protrusion forces that allow proplatelet formation. The driving force for proplatelet elongation is not microtubules polymerization but microtubule sliding. Proplatelets fail to form in megakaryocytes treated with agents that inhibit microtubule assembly [12]. Microtubules also play an important role in organelle transport in proplatelets [9]. The role of acto-myosin is not completely understood. Actin polymerization does not play a major role in proplatelets extension. However, it is required for proplatelets branching and thus may be important for the regulation of platelets production. MKs treated with one of the actin toxins (cytochalasin, latrunculin) can extend long proplatelets but fail to branch [13 11].

2.2.3 Apoptosis in platelets biogenesis

Apoptosis is a programmed form of cell death. Morphologically, it is defined by cell membrane blebbing, cell shrinkage, chromatin condensation and DNA fragmentation. In metazoan cells, apoptosis is regulated by two highly conserved pathways: the extrinsic one (triggered by ligands such as Fas ligand, TNFα) and the intrinsic or mitochondrial one. Both share a common end point: the activation of proteolytic enzymes called caspases, which mediate the rapid dismantling of cells. In viable cells, caspases reside in the cytosol as inactive precursors and they will be activated only upon precise stimuli.

Since the initial observation that platelets release from megakaryocytes resembles the onset of apoptosis [14], an increasing body of evidence has suggested that platelets shedding is an apoptotic process. Megakaryocytes possess both an intrinsic and an extrinsic apoptosis pathway, which they might restrain in order to survive and at the same time they may need it to facilitate platelets shedding. Once released in circulation platelets fate depends on the Bcl-2 family proteins. In particular, Bcl-XL is the key player for platelets survival (Fig.4) [15].

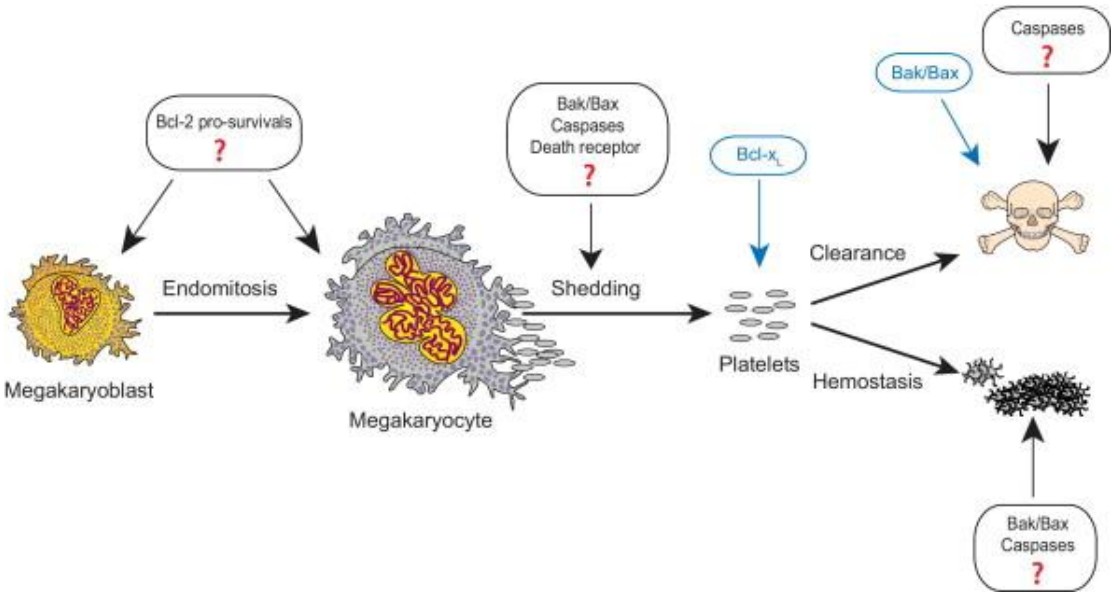


Fig.4: Apoptotic pathways and processes in the megakaryocyte lineage. While the critical players have yet to be identified, current evidence suggests that megakaryocytes possess both an intrinsic and extrinsic apoptosis pathway. Once released into the circulation, platelets depend on the Bcl-2 family protein Bcl-x_L for their survival. Bcl-x_L restrains the pro-death protein Bak until the end of the platelet's life span, when it is presumed apoptosis is initiated to facilitate clearance from the circulation. Whether caspases are required, and whether this apoptotic pathway intersects at all with agonist-driven activation pathways, remains to be established. [16].

Overexpression of pro-survival Bcl-XL in megakaryocytes inhibits proplatelets formation [17]. Pro-apoptotic factors, including caspase and nitric oxide (NO), are also expressed in MKs. Evidence indicating a role for caspases in platelet assembly is strong. Caspase activation has been established as a requirement of proplatelet formation. Caspase-3 and caspase-9 are active in mature megakaryocytes and inhibition of these caspases blocks proplatelet formation [18]. As a matter of fact, in cultured human megakaryocytes it has been observed cleavage of substrates such as gelsolin in cells that had begun to shed platelets [19]. NO, as well as pro-apoptotic cytokines such as TNF-α and IFN-γ, can trigger increased release of platelet-like bodies from megakaryocytic cell line Meg-01 [20]. Conversely, a range of pathophysiological stimuli, including chemotherapy, is thought to cause thrombocytopenia by inducing the apoptotic death of megakaryocytes and their progenitors. Recently, Josefsson *et al* proved that MKs do not activate the intrinsic pathway to generate platelets, but they must restrain it to survive and progress safely through proplatelet formation and platelets shedding. Their studies demonstrate that deletion of pro-apoptotic factors Bak and Bax, the gatekeepers of the intrinsic pathway had no adverse effect on MK number, ploidy or proplatelet formation. Moreover, genetic deletion of Bak and Bax could block death and rescue proplatelet formation in the presence of ABT-737 (proapoptotic agent) and completely restored platelets production in mice lacking Bcl-XL. Thus, megakaryocytes require Bcl-XL to restrain the activity of Bak and Bax during platelet production. They confirm that MKs become dependent on Bcl-XL just as they enter into proplatelet formation and it is not required for their growth and development [21].

Once released into the circulation, platelets encounter one of the two possible fates: consumption in a hemostatic process or removal by the reticuloendothelial system in liver or spleen. Since only a fraction of the circulating platelets population is required to maintain hemostasis at steady state, the majority of platelets die via the second way, being cleared after 10 days in humans and 5 days in mice [16 19]. First postulated by Vanags *et al* [22], it is now well established that platelets depend on pro-survival Bcl-XL to stay alive. The role of Bcl-XL is to restrain pro-death Bak. Deletion of Bak nearly doubles circulating platelets life span [15]. Thus, the balance between these two Bcl-2 family members dictates whether a platelet lives or dies. Differently from MKs, only caspase-3 is abundant in platelets, while caspase-9 is absent [7]. These data

support differential mechanism for programmed cell death in platelets and megakaryocytes. However, it remains unclear how this apoptotic program is controlled.

2.2.4 Regulation of megakaryocytopoiesis

MK differentiation is regulated by extracellular and intracellular mechanism. Cytokines and numerous components of the bone marrow microenvironment are involved in the regulation of MK terminal differentiation. Among cytokines, thrombopoietin (TPO), the key player, IL6 and IL11 have a role in MK differentiation. In addition, proplatelet formation is tightly regulated by interactions with the stroma and the extracellular matrix (ECM). In the marrow, collagen inhibits proplatelet formation. When a megakaryocyte starts its migration through the endothelium it interacts with some components of the ECM such as fibrinogen, which may induce and increase proplatelets formation [23]. MK marrow localization and migration are regulated by FGF4 and SDF-1. Stromal cell-derived factor 1 (SDF-1) is a CXC chemokine whose main receptor is CXCR4, a seven-transmembrane receptor coupled to G-protein. SDF-1 is produced locally by stromal cells located in the marrow and promotes the migration and contact of immature MK with a permissive, endothelial rich BM microenvironment [24]. CXCR4 is expressed along the entire MK differentiation pathway from early progenitors to platelets. Albeit the increase in CXCR4 expression, mature MK and platelets are poorly reactive to SDF-1. This phenomenon is related to a reduced function of the receptor caused by an overexpression of RGS16 (regulator of G protein signaling). This induces a decrease in retention forces and may explain the MK egress from the marrow [25].

MK development is controlled by numerous transcription factors (TFs), which form complexes that coordinately regulate the chromatin organization to specifically activate the genes of MK lineage and concurrently repress gene expression that support other cell type. Many MK-specific genes are co-regulated by GATA and friend of GATA (FOG) together with acute myeloid leukemia/runt related TF1 (AML/RUNX1) and ETS proteins. The zinc-finger protein GATA-1 is the principle TF directing MK development by functioning both as an activator or repressor depending of the protein complex. One of the initial events during MK/E lineage restriction is

downregulation on PU.1, the main TF responsible for myeloid cell differentiation and upregulation of GATA1. GATA-1 possesses a robust MK-specific genetic program, regulating all stages of MK development. In humans, GATA-1 mutations lead to severe diseases involving both erythroid cells and MK [26,27]. Nuclear factor erythroid 2 (NF-E2) is a heterodimeric leucine zipper TF that comprises an MK-erythroid specific 45-kDa subunit (P45) and a non-lineage specific p18 Maf family subunit which controls terminal MK maturation, proplatelet formation and platelet release [28] [26]. Maf or P45 mutations in mice result in severe impairment of megakaryocytopoiesis. NF-E2 deficient mice have profound thrombocytopenia with MK maturation arrest and severe platelets deficit [29].

2.2.4.1 Thrombopoietin

Thrombopoietin (TPO), also known as c-Mpl ligand, is the primary physiological growth factor for the MK lineage, which also plays a central role in the survival and proliferation of HSC [30]. Human thrombopoietin gene (THPO) is located on chromosome 3q26.3-3q27. Abnormalities (inversion or deletion) at its chromosomal locus are often found in megakaryocytic leukemia and other myeloproliferative disorders associated with thrombocytosis [31]. The gene contains 5 coding exons and 2 upstream noncoding exons, which result in a long 5 prime untranslated region (5'UTR) with additional 7 upstream initiation codons (uATG). The presence of 7 uATG inhibits translation by causing premature initiation and thereby preventing the ribosome from initiating at the physiological start codon [1]. These types of uORF only exist in 10% of mRNA transcripts in human and are often found in highly regulated genes [32]. Different mutations in the 5'UTR regulatory region leading to an overexpression of TPO are associated with some cases of familial essential thrombocythemia or familial thrombocytosis [33]. *TPO* mRNA produces a 353 amino acid precursor protein. The mature molecule is composed of 332 amino acid (95 kDa) and it is acidic and heavily glycosylated. The TPO protein consists of 2 domains: the N-terminal and the C-terminal domains. The N-terminal portion (residue1-153) has high homology with erythropoietin (EPO) and represents the receptor-binding domain of the hormone. The C-terminal part of TPO protein (residue 154-332) contains a carbohydrate-rich domain that is highly glycosylated and it is important in maintaining protein stability [34].

TPO is mainly produced in the liver and to a minor degree in the kidney, spleen, and other organs. The normal serum concentration of TPO is very low, ranging between 0.5 and 2 pmol/L [35]. Its production is constitutive and it is independent from the actual platelet concentration in blood. Plasma TPO concentration is regulated by the targets, because megakaryocytes and platelets metabolize the hormone via binding to the receptor. Levels of TPO are usually inversely related to BM MK mass and platelet counts [33]. Thus, megakaryocytopoiesis is regulated by plasma levels of unbound TPO, which reflects the balance between constitutive production and rate of destruction that is generally dictated by the overall platelets production [36].

2.2.4.2 TPO receptor: c-Mpl

The TPO receptor, c-Mpl, is a typical type I hematopoietic growth factor (HGF) receptor and contains 2 cytokine receptor homology modules (CRMs). Biochemical and crystallographic data show that TPO binds only the distal CRM (CRM1) and thereby initiates signal transduction. Therefore, CRM1 acts as an inhibitor of c-Mpl that is relieved upon the binding with TPO. Whether TPO causes dimerization of the receptor or simply stabilize dimers is unclear. The thrombopoietin receptor is expressed primarly in hematopoietic tissues, specifically in MKs, their precursors and their progeny (platelets). C-Mpl is constitutively expressed on the surface of these cells and its display is regulated by the thrombopoietin binding and receptor internalization [37]. Upon binding ligand, the receptor is activated to transmit numerous biochemical signals (Fig.5). The HGF receptors exist in a homodimeric state in the absence of the ligand and in this conformation the cytoplasmic domains are separated. After the binding with the hormone, there is a conformational change bringing the domains in contact with each other. Several studies indicate that the cytoplasmic domains of the receptor bind the JAK kinases also in the inactive state. But the activation of the kinases, through cross-phosphorylation, occurs only with the closer juxtaposition of the enzymes reached after the onset of the signal transduction [37].

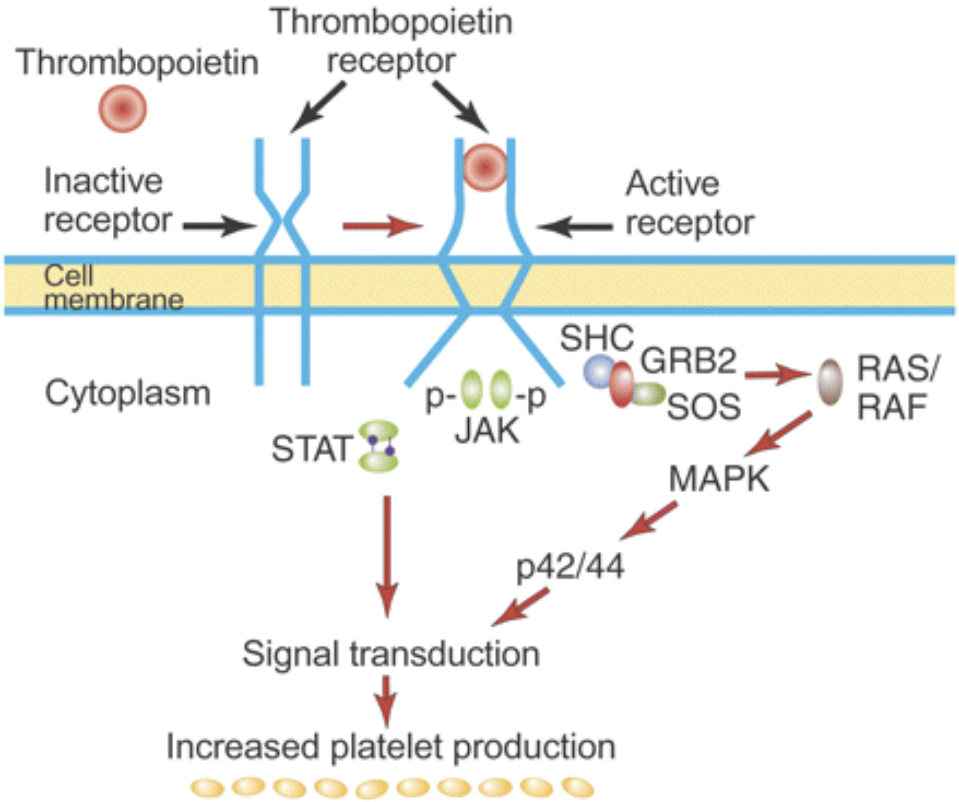


Fig.5: Mechanism of activation of the TPO receptor. TPO binds the distal CRM (CRM 1) of the inactive TPO (c-Mpl) receptor and inducing a conformational change that initiates many downstream signal transduction events. STAT, signal transducer and activator of transcription; JAK, janus kinase; MAPK, mitogen activated protein kinase; RAS, GTPase protein; RAF, serine/threonine-specific protein kinases; SHC-GRB2, adaptor proteins; SOS, son of sevenless (guanine nucleotide exchange factor). [34].

Once JAK kinases are active, they phosphorylate different tyrosine residues (Y) on the receptor itself creating docking sites (P-Y) for signaling molecules that contain Shc homology (SH)2 or phosphotyrosine-binding (PTB) motifs. Numerous signaling molecules mediate the signal transduction events: signal transducer and activators of transcription (STATs) and phosphoinositol-3-kinase (PI3K) leading to cell survival and proliferation; the mitogen-activated protein kinases (MAPKs) promoting differentiation events and activation of anti-apoptotic pathways that promote cellular viability. After stimulation, the induced signal must be limited through different mechanism. First, the docking sites at the receptor can recruit suppressor molecules such as SOCS proteins and some phosphatases like SHP1, SHIP1 and PTEN [38] [39]. SOCS proteins are induced by STAT-mediated transcription and once translated they can bind to the P-Y on the receptor and JAK kinases in order to preclude binding of additional signaling molecules and induce the proteolytic destruction. Phosphatases act removing the P-Y sites on the receptor and signaling molecules. Moreover, c-Mpl

can be removed from the cell surface after TPO-binding in a clathrin-mediated process. The receptor, in fact, bears two YRRL sequences necessary for the recognition of this protein as a substrate of the clathrin-mediated endocytosis. Glycosylation of c-Mpl appears to play an important role in cell surface expression of the receptor. There are four sites of potential N-linked glycosylation: N_{117}, N_{178}, N_{298} and N_{358}. In some cases of polycythemia vera, c-Mpl has been found to be underglycosylated and surface display of the receptor is decreased [40]. In addition, c-Mpl might go to degradation through the proteasome since it contains two intracellular lysines (K) that are potential targets for the ubiquitination (K_{553}, K_{573}) [37,41].

2.3 HEREDITARY THROMBOCYTHEMIA

Hereditary Thrombocythemia (HT) is a familial form of myeloproliferative disease (MPD) characterized by sustained proliferation of megakaryocytes and overproduction of platelets, which is often clinically indistinguishable from the sporadic essential thrombocythemia (ET). Platelets count normally ranges from 100 to 400 x 10^9/L [2,42-44]. In clinical practice, the term "thrombocytosis" refers to platelets counts above 450 x 10^9/L. Thrombocytosis may be classified into mild (platelets counts: 450-700 10^9/L), moderate (700-900 x 10^9/L) or severe (>900 x 10^9/L) [42]. Thrombocytosis can be considered as primary, if caused by a defect intrinsic to the hematopoietic progenitors, or as secondary (or reactive), if consequence of a disease that persistently stimulate the otherwise normal magakaryocytopoiesis. Primary thrombocytosis includes both acquired (essential thrombocythemia) and hereditary form [45]. In the first ones, abnormalities are detectable only in cells belonging to the hematopoietic system and around 50% of the acquired cases present a gain-of function mutation in JAK2 gene (JAK2 V617F) [46-49]. In the hereditary forms genetic defects are present in both somatic and germ line cells and are transmitted as a hereditary character. None of the hereditary form presents the JAK2 V617F alteration [45]. However, it has been recently reported a germline *JAK2* V617I mutation associated with hereditary thrombocytosis [50]. HT is often a polyclonal disease affecting selectively the megakaryocytic lineage, the inheritance is autosomal dominant and the penentrance is close to 100% [51]. In some case of HT, mutations of thrombopoietin or MPL gene are the disease-causing defects (Table1). Up to date, four different THPO mutations have been described. All mutations are located in the 5'UTR region and delete the untranslated open reading frame (uORF). This causes increased translation of the THPO mRNA resulting in an overproduction of TPO and hence thrombocytosis. THPO mutations that cause HT have not been found in patients with sporadic ET [52]. Three germ line MLP mutations have been reported so far. The MPL-S505N mutation involves the transmembrane domain of MPL inducing an autonomous dimerization of the receptor activating the downstream signaling pathways in a TPO-independent way [53]. This mutation appears functionally similar to MPL-W515K/L acquired mutation reported in some patients with myeloproliferative neoplasm (MPN). MPL-K39N is a polymorphism restricted to African Americans. About 7% of this population is heterozygous for this mutation. In homozygosity, this

mutation causes a severe thrombocytosis [54]. The MPL-P106L mutation was first described in an Arabic family. The frequency of this alteration is about 6% among Arabic individuals. The homozygous state is associated with mild to severe thrombocytosis [55]. Both these last two mutation involve the CRM1 domain of the MPL, most probably affecting the receptor's ability to bind TPO and resulting in reduced clearance of the thrombopoietin and in the over-stimulation of megakaryocytopoiesis [45]. Thus, in many families with hereditary thrombocythemia the disease-causing gene remains unknown.

Involved GENE	Molecular alteration	Autosomal inheritance	TPO levels
THPO	G to C mutation in the splice donor site of intron 3	Dominant	High
THPO	G deletion in 5'UTR	Dominant	High
THPO	G to T substitution in exon 3	Dominant	High
THPO	A to G mutation in intron 3	Dominant	High
MPL	S505N	Dominant	-
MPL	K39N	Dominant (low penetrance)	-
MPL	P106L	Recessive	High

Table 1: molecular alteration in hereditary thrombocythemia

2.4 GENETIC STUDIES FOR THE IDENTIFICATION OF THE DISEASE CAUSING GENE

2.4.1 Linkage analysis

Familial forms of MPD, such as HT, are polyclonal and caused by germ-line mutations inherited in the Mendelian way among family members. The use of linkage analysis to search for the disease gene has been quite successful in some families with HT and with primary familial and congenital polycythemia (PFCP) [52]. The principle of linkage analysis is simple. The genome differs in several positions for different genetic markers, which are DNA sequences that show polymorphism (variations in size or sequence) in the population. Thanks to these variants it is possible to distinguish the maternal and paternal alleles and in the case of the disease gene, the alternative alleles will be the normal allele and the disease allele. In this condition the different alleles can be distinguished by looking for occurrences of the disease in the pedigree. These studies are based on the fact that recombination rarely occurs between two loci that are close to each other on the same chromosome. By searching for genetic markers segregating with the disease phenotype, the potential disease genes can be identified from their proximity to the marker's location. Any mendelian character that can distinguish the paternal and maternal allele in one individual can be used as a genetic marker. However, a good genetic marker needs to be sufficiently polymorphic and densely located throughout the whole genome (<20cM). The first generation of genetic markers was restriction fragment length polymorphism (RFLP). The limitation of RFLPs is that they have only two alleles, the restriction site is present or it is absent, and therefore not very informative. Construction of dense human genetic maps could be possible with the introduction of microsatellite markers. Short tandem repeat polymorphisms (STRs) are multiallelic and more informative. STRs can be identified by PCR making the linkage analysis fast and easy. Moreover, many compatible sets of microsatellite markers have been developed that can cover the whole genome with 400 markers. Microsatellite marker analysis can be complemented with the use of single nucleotide polymorphism (SNPs). Nowadays, a gene chip technology is available that can integrate thousands of SNPs in one single chip and genotype them at once. SNPs are bi-allelic markers and less informative than microsatellites. However the large

amount of SNPs that can be genotyped in a single chip make them a very powerful tool [56].

Linkage studies rely on statistical evaluation of the evidence in favour of the co-segregating marker loci with a trait [57]. The analysis can be parametric or non-parametric. Parametric linkage analysis is pedigree-based and defines explicit relationship between phenotypic and genetic similarity. This approach is typically used for single gene disorders and Mendelian forms of complex disorders. It requires a model for the disease, including the mode of inheritance, the disease penetrance and the allele frequency. In principle, this type of analysis estimates the recombination fraction between two or more loci. Such estimation can be biased if the model of inheritance has been mis-specified. Therefore, a non-parametric analysis is required when the disease model is unknown. Non-parametric methods test for increased sharing among affected individuals and evaluate whether segregation at specific location is “not-random”. Specifically, the objective is to show increased identical by descendent (IBD) sharing among sets of affected individuals. Without making any assumptions about the genetics of the disease, it has been used as the main tool for studying common nonmendelian diseases. However nonparametric methods decrease the power of mapping, candidate regions defined by this method are usually large [58].

Although powerful for detecting genetic loci in single gene disorder, linkage analysis has only limited success in finding genes for multifactorial diseases such as diabetes, asthma and heart disease. With the rapid progress of genotyping technology and the identification of a constantly increasing number of DNA variants, genetic association studies became the preferable for mapping complex diseases. Genetic association approaches assess correlation between genetic variants and trait difference on a population scale. Association relies on the retention of adjacent DNA variants over many generations. Thus, association studies can be regarded as very large linkage analysis of unobserved hypothetical pedigrees. They draw from historic recombination so disease-associated regions are very small, including one gene or gene fragment. In fact, through subsequent generations, recombination will occur and separate the disease mutation from the specific alleles of its original haplotype. Still, there are various limitations in these studies dependent on the particular design, study aims and analytical framework adopted. Therefore, it is crucial to critically

design the approach to use in order to avoid mistakes and maximize the potential to identify new components of disease [59].

2.4.2 Characterization of candidate mutation

Identification of genetic variation within a pedigree is quite relevant to study cases of familial disease, such as hereditary thrombocytemia. The automated Sanger sequencing method, considered as “first-generation” technology, has been replaced with newer methods referred to as next-generation sequencing (NGS). These newer technologies constitute various strategies that rely on a combination of sample preparation, sequencing, genome alignment and assembly methods. NGS platforms are very useful for many applications, including the sequencing of targeted region of interest. Once an alteration is detected, it is needed to confirm that is a candidate mutation. The sequence work-up is summarized in figure 6 (Fig.6). The first step is excluding that the alteration is a known polymorphism. The mutation must be present in all affected family member of the pedigree and must be absent in at least 100 normal controls analyzed for the same alteration. When the confirmation step is concluded, the relevance of the candidate mutation must be validated with functional assays. The first approach could be to predict the impact of the amino acid change on protein function with computational program such as Sorting Intolerant From Tolerant (SIFT) [60] or PolyPhen-2 [61] [62]. This is only the first screening that must be complemented with proper in-vitro and in-vivo assay. The generation of transgenic mouse model could be a powerful tool to directly demonstrate the role of this mutation for induction of the disease.

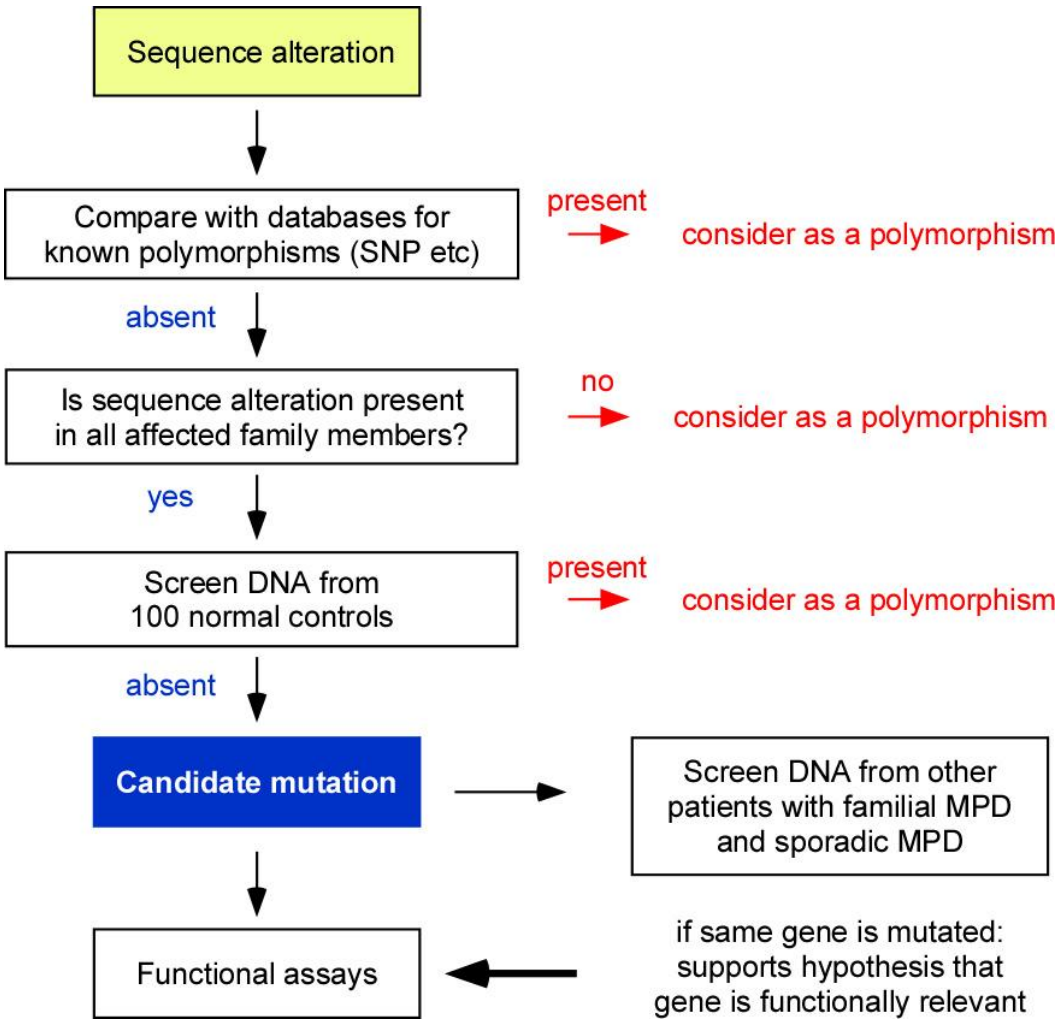


Fig. 6: sequence work-up

3. RESULTS

3.1 FAMILY WITH HEREDITARY THROMBOCYTHEMIA

Hereditary Thrombocythemia is reported in 16 family members in five successive generations (Fig.1). The inheritance is autosomal dominant with early onset of the disease. The pathological phenotype is characterized by persistent elevation of the platelet counts ranging from 700 x 10^9/L to 1200 x 10^9/L, occasional giant platelets with abnormal aggregation, isolated hyperplasia of enlarged megakaryocytes and splenomegaly. White blood cell count, hemoglobin level, thrombopoietin and bone marrow iron storage are normal in affected patients [63]. No mutations were detected in MPL and THPO.

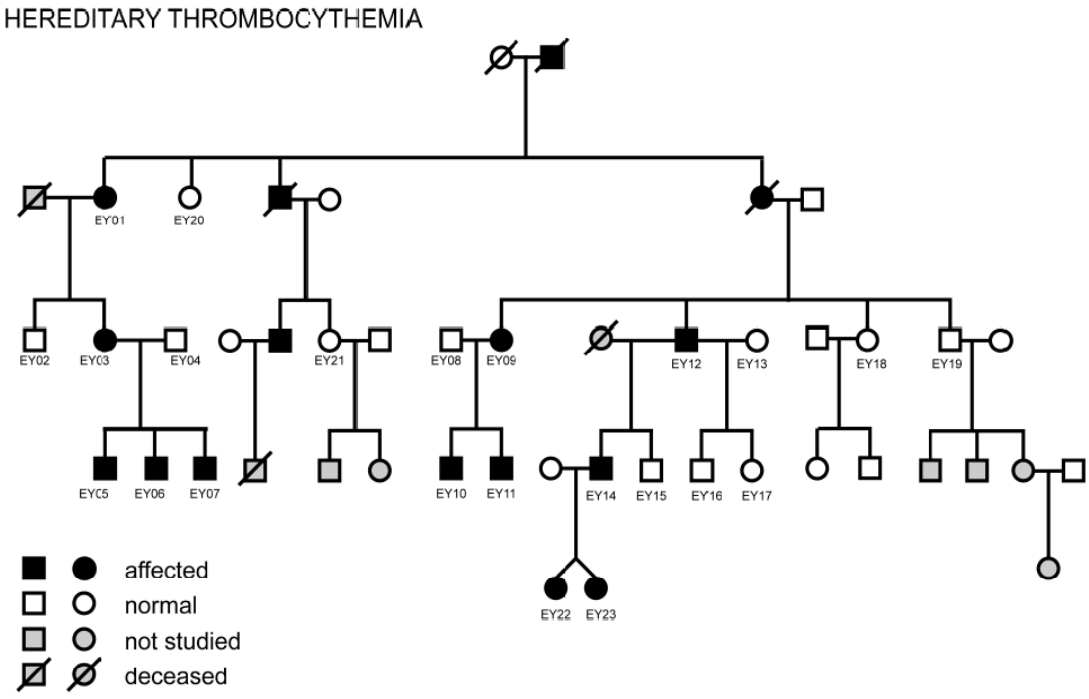


Fig.1 Pedigree of the family affected by hereditary thrombocythemia. The inheritance is autosomal dominant with a high penetrance of the disease. Filled black symbols, affected individuals; open symbols, normal individuals; filled grey symbols, not studied; crossed symbols, deceased.

3.2 LINKAGE ANALYSIS

Since in the HT family it was expected that a genetic component was responsible for the phenotype, parametric linkage analysis was used to detect the chromosomal location of the disease-gene. This has been the traditional approach for mapping Mendelian disease since the 1970's. To determine if there was significant evidence for linkage, LOD scores were calculated. LOD stands for Log of the ODds. A LOD score greater than 3 gives the evidence of linkage with 5% chance of error. Linkage can be rejected if the LOD score is lower than -2. Values between -2 and 3 are inconclusive [64].

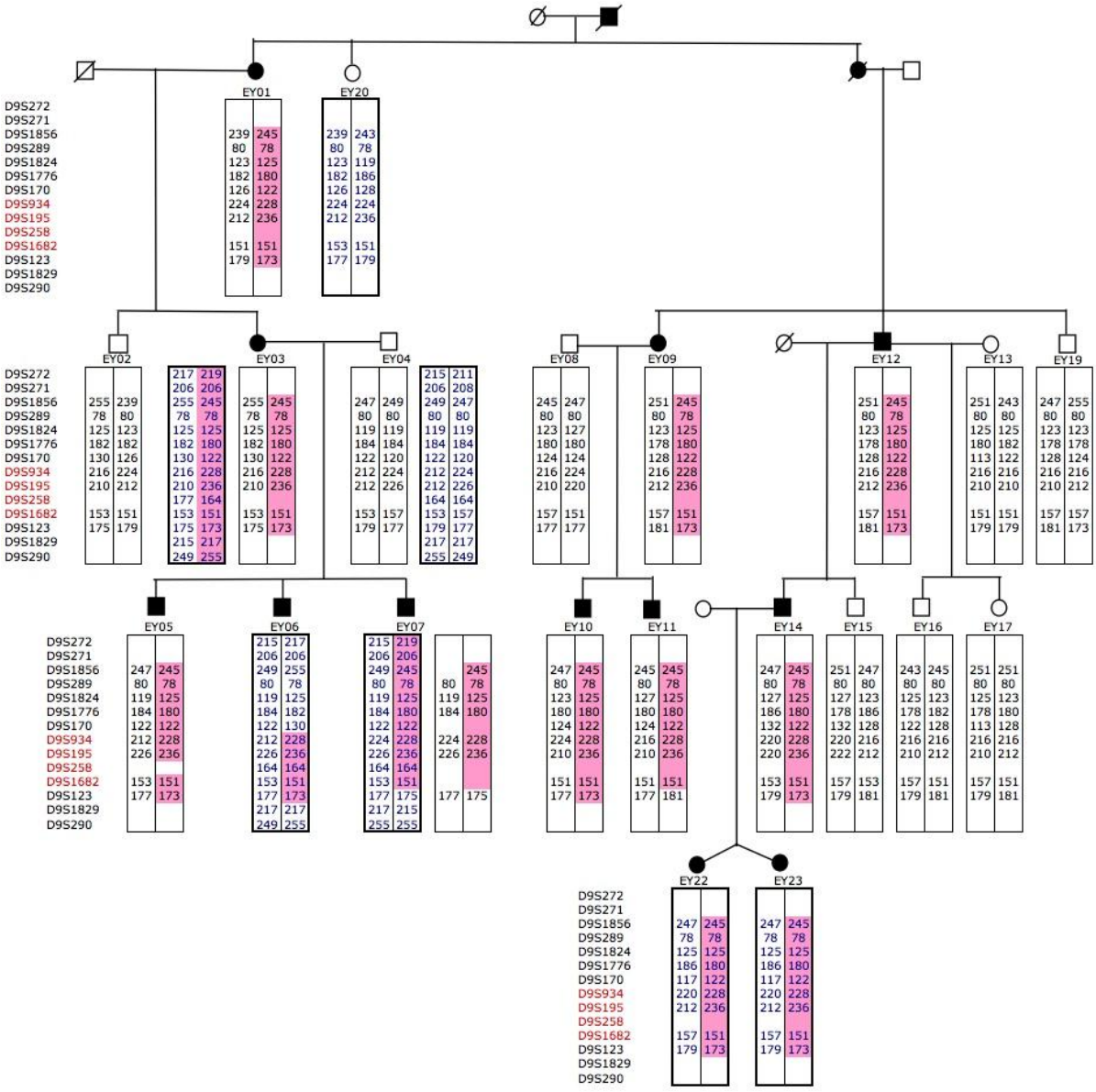


Fig 2 Segregation of microsatellite markers. Microsatellite markers are positioned according to physical distance (measured in Mb). Co-segregating microsatellites are reported in red. The disease allele is marked in pink. Data reported in blue are the ones I generated.

A former PhD student, Liu Kun, performed genome-wide linkage analysis of 20 family members (10 affected and 10 unaffected) with microsatellite markers [65]. When I joined the lab, I filled some gaps of the analysis and I characterized three additional patients EY6, EY22 and EY23 (Fig.2). From this segregation analysis the haplotypes were built as shown in figure 3. A co-segregating region shared between the 12 affected family members and not shared by the non-affected family members was found on chromosome 9p with the highest LOD score of 4.3 at theta = 0 (Fig.3). The co-segregating region found was approximately 4 megabase (Mb) in size.

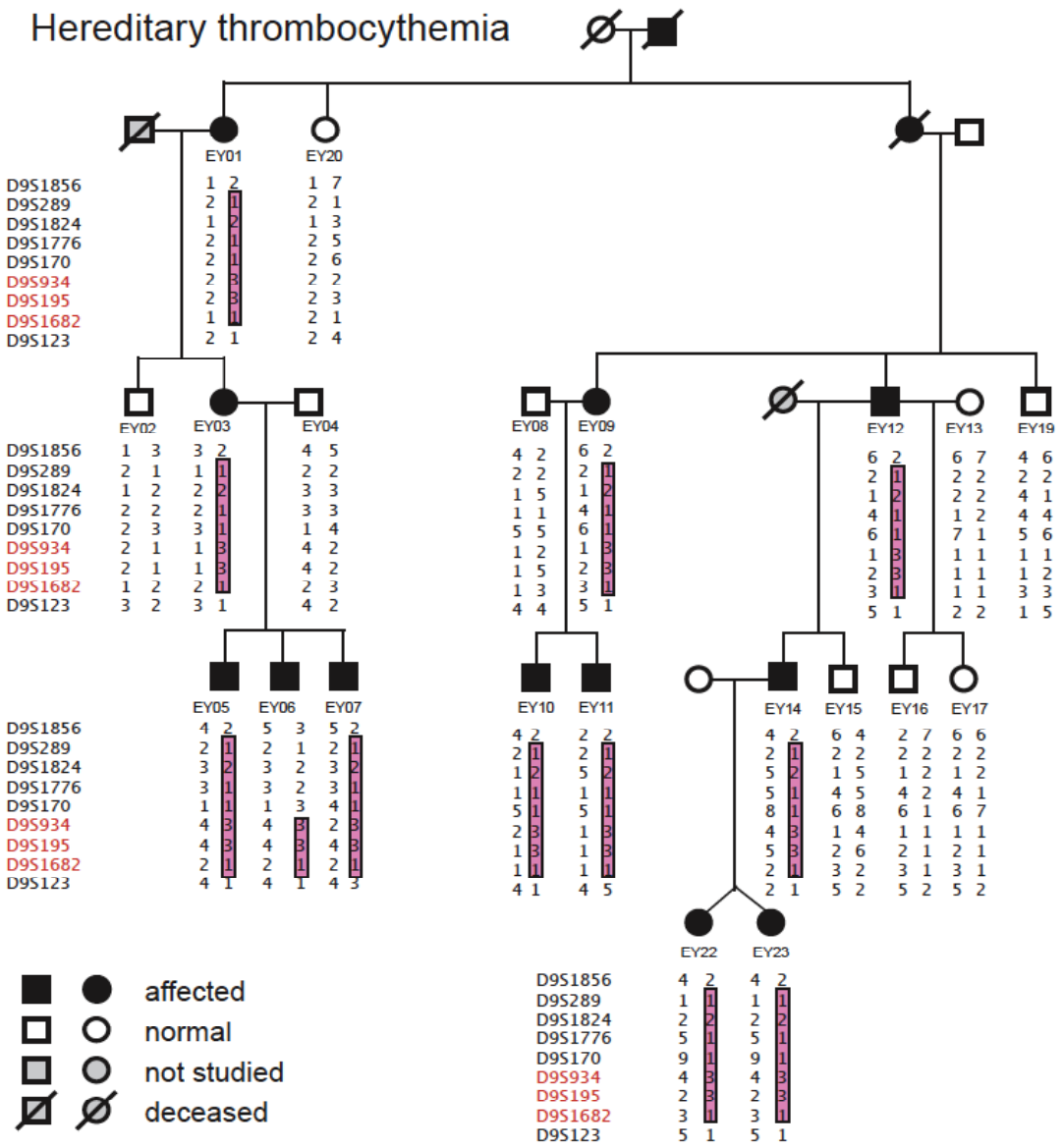


Fig.3 Haplotype analysis of nine microsatellite markers located on chromosome 9 from marker D9S1856 to marker D9S123. Microsatellite markers are positioned according to physical distance (measured in Mb). Individuals with HT are indicated by filled symbols and unaffected individuals by open symbols. Haplotypes for these markers are shown with pink boxes.

To further investigate the inherited traits in the family a SNP chip array was performed on 14 family members: 7 affected and 7 unaffected. The major part of the data were obtained from the former PhD student [65], after I took over the project the SNP analysis was remade with the addition of a new patient. The data obtained confirmed a co-segregating region on chromosome 9p of 4 Mb (positions: 119.9-124.0, NCBI36/hg18 assembly) with the highest LOD score of 3 at theta=0 (Fig.4).

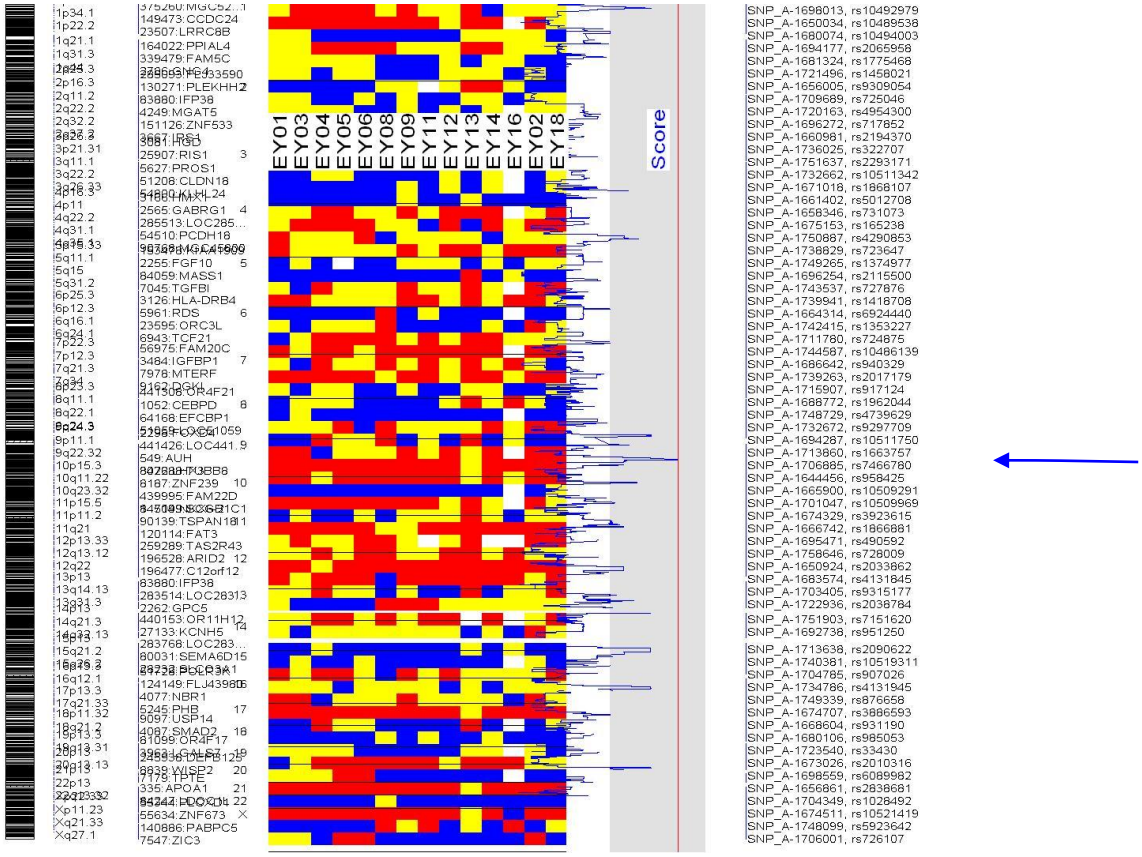


Fig. 4 Linkage analysis of hereditary thrombocythemia for Affimetrix 50K SNP array Xba240 is shown for all chromosomes. 14 family members (7 affected and 7 unaffected) were included in the analysis. Homozygous SNPs are shown in red or blue, heterozygous SNPs are shown in yellow. The maximum LOD peak was 3 (θ=0) at chromosome 9 assuming an autosomal dominant model. LOD score of 3 is indicated by a red line.

According to this data, the co-segregating region on chromosome 9p was defined from position 113 Mb to 124 Mb (hg18, UCSC). This region contains 93 genes listed in Table1.

STR marker	Gene ID	Gene Name	Position (Mb)
D9S1856	OR2K2	Olfactory receptor, family 2, subfamily K, member 2	113.8
	KIAA0368	KIAA1958	
	ZNF483	zinc finger protein 618	
	PTGR1	Prostaglandin reductase 1	
	LRRC37A5P	Leucine-rich repeat-containing 37 member A5 pseudogene	
	DNAJC25-GNG10	DNAJC25-GNG10	
	DNAJC25	DnaJ (Hsp40) homolog, subfamily C , member 25	
	GNG10	Guanine nucleotide binding protein (G protein), gamma 10	
	C9orf84	Chromosome 9 open reading frame 84	
	UGCG	UDP-glucose ceramide glucosyltransferase	
	SUSD1	Sushi domain containing 1	
	PTBP3	Polypyrimidine tract binding protein 3	
	HSDL2	Hydroxysteroid dehydrogenase like 2	
	KIAA1958	KIAA1958	
	INIP	SSB-interacting protein 1	
	SNX30	Sorting nexin family member 30	
	SLC46A2	Solute carrier family 46, member 2	
	ZNF883	Zinc finger protein 883	
	ZFP37	Zinc finger protein 37	
	FAM225B	Family with sequence similarity 225, member B	
	FAM225A	Family with sequence similarity 225, member A	
	SLC31A2	Solute carrier family 31 (copper transporters), member 2	
	FKBP15	FK506 binding protein 15	
	SLC31A1	Solute carrier family 31 (copper transporters)	
	CDC26	Cell division cycle 26 homolog	
	PRPF4	PRP4 pre-mRNA processing factor 4 homolog	
	RNF183	Ring finger protein 183	
	WDR31	WD repeat domain 31	
	BSPRY	B-box and SPRY domain containing	
	HDHD3	Haloacid dehalogenase-like hydrolase domain containing 3	
	ALAD	Aminolevulinate dehydratase	
	POLE3	Polymerase (DNA directed), epsilon 3, accessory subunit	
	C9orf43	chromosome 9 open reading frame 43	
D9S289	RGS3	regulator of G-protein signaling 3	115.3
	ZNF618	zinc finger protein 618	
	AMBP	alpha-1-microglobulin/bikunin precursor	
D9S1824	KIF12	kinesin family member 12	115.8
	COL27A1	collagen, type XXVII, alpha 1	
	ORM1	orosomucoid 1	
	ORM2	orosomucoid 2	
	AKNA	AT-hook transcription factor	
	DFNB31	deafness, autosomal recessive 31	
	ATP6V1G1	ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G1	
	C9orf91	chromosome 9 open reading frame 91	

	TNFSF15	tumor necrosis factor (ligand) superfamily, member 15	
	TNFSF8	tumor necrosis factor (ligand) superfamily, member 8	
D9S1776	TNC	tenascin C	116.9
	PAPPA	pregnancy-associated plasma protein A, pappalysin	
D9S170	ASTN2	astrotactin 2	118.0
	TRIM32	tripartite motif containing 32	
	TLR4	toll-like receptor 4	
D9S934	DBC1	deleted in bladder cancer 1	120.0
D9S195	CDK5RAP2	CDK5 regulatory subunit associated protein 2	121.0
	MEGF9	Multiple EGF-like-domains 9	
	FBXW2	F-box and WD repeat domain containing 2	
	PSMD5	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 5	
	PHF19	PHD finger protein 19	
	TRAF1	TNF receptor-associated factor 1	
	C5	Complement component 5	
	CEP110	Centriolin	
	RAB14	Member RAS oncogene family	
	GSN	Gelsolin	
	STOM	Stomatin	
	GGTA1P	Glycoprotein, alpha-galactosyltransferase 1 pseudogene	
	DAB2IP	DAB2 interacting protein	
	TTLL11	Tubulin tyrosine ligase-like family, member 11	
	NDUFA8	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 8	
	MORN5	MORN repeat containing 5	
	LHX6	LIM homeobox 6	
	RBM18	RNA binding motif protein 18	
	MRRF	Mitochondrial ribosome recycling factor	
	PTGS1	Prostaglandin-endoperoxide synthase 1	
	OR1J1	olfactory receptor, family 1, subfamily J, member 1	
	OR1J2	olfactory receptor, family 1, subfamily J, member 2	
	OR1J4	olfactory receptor, family 1, subfamily J, member 4	
	OR1N1	olfactory receptor, family 1, subfamily N, member 1	
	OR1N2	olfactory receptor, family 1, subfamily N, member 2	
	OR1L8	olfactory receptor, family 1, subfamily L, member 8	
	OR1Q1	olfactory receptor, family 1, subfamily Q, member 1	
	OR1B1	olfactory receptor, family 1, subfamily B, member 1	
	OR1L1	olfactory receptor, family 1, subfamily L, member 1	
	OR1L3	olfactory receptor, family 1, subfamily L, member 3	
	OR1L4	olfactory receptor, family 1, subfamily L, member 4	

	OR1L6	olfactory receptor, family 1, subfamily L, member 6	
	OR5C1	olfactory receptor, family 5, subfamily C, member 1	
	OR1K1	olfactory receptor, family 1, subfamily K, member 1	
D9S1682	PDCL	phosducin-like	123.8
	RC3H2	ring finger and CCCH-type domains 2	
	ZBTB6	zinc finger and BTB domain containing 6	
	ZBTB26	zinc finger and BTB domain containing 26	
	RABGAP1	RAB GTPase activating protein 1	
	GPR21	G protein-coupled receptor 21	
D9S123	STRBP	spermatid perinuclear RNA binding protein	124.0

Tab 1 List of the genes present in the co-segregating region. STR, short tandem repeat. Microsatellites markers in yellow represent the maximal shared region. Microsatellites markers in red represent the minimal co-segregating region.

3.3 NEXT GENERATION SEQUENCING (NGS)

To sequence the co-segregating region of interest, a next generation sequencing based on “target-enrichment” was performed. This “target-enrichment” method allows the selective capture of genomic fragments using RNA baits with complementary sequences to the DNA regions of interest. The enrichment of the DNA was performed with two different platforms (Fig 5). In the Nimblegen platform 20 μg of genomic DNA are fragmented and then attached to 454-adaptor without PCR amplification before the enrichment step. The Agilent platform, instead, allows the enrichment of smaller quantity of genomic DNA (1-5 μg), which is amplified by a PCR step after the fragmentation and adapter-ligation steps. The enriched exonic region on chromosome 9 had a total size of 450 kb and it covered around 900 exons (including UTR and intron-exon boundaries).

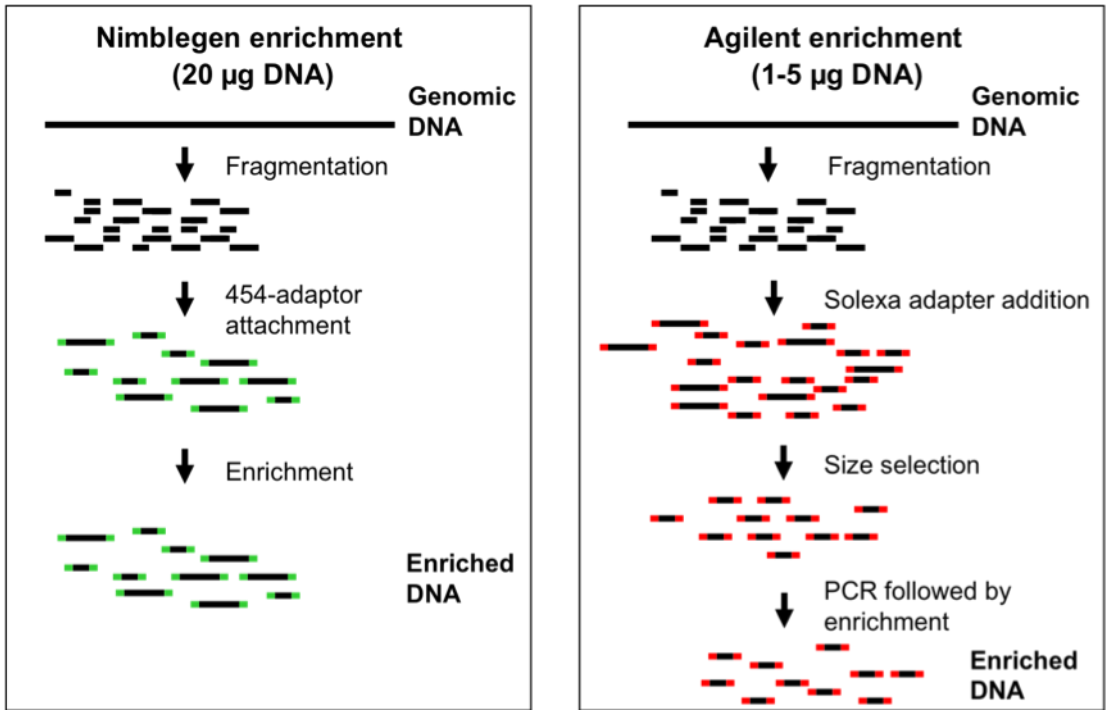


Fig 5 Platforms used for the enrichment step.

The sequencing step was performed using both Illumina and Roche 454 technologies. The statistics of this step are summarized in Table 2. The sequence capture followed by NGS was very efficient and only 11 exons were not covered. Five of these exons were successfully analyzed by capillary sequencing and the remaining 6 could not be sequenced probably due to a high GC content.

	Illumina (Nimblegen)	**Illumina (Agilent)**	**454 (Nimblegen Titanium)**
Reads	14,798,154 (two PE seq)	5,058,820 (one SE seq)	613,227 (1/2 plate)
bp	497,487,433	186,988,134	218,899,568
% of exons (> 10 coverage)	96.7 %	94.8 %	95,3 %
Average Coverage	114	45	46

Tab 2 Statistics about the coverage of the sequencing step. PE seq, pair-end sequencing; SE seq, single end sequencing.

3.4 SEQUENCING DATA ANALYSIS AND VALIDATION

3.4.1 Filtering and analysis of NGS data in the family affected by HT

The next generation sequencing was performed on two patients: EY1 and EY11. These patients were chosen for sequencing since they are located far apart in the pedigree, thus decreasing the amount of regions shared by chance. In both patients, around 1200 sequence alterations were found. These alterations were filtered out according to the fact that they were not described as known polymorphisms, they alter the coding region of a gene causing a nonsynonymous amino acid change and they are common heterozygous alterations. Only five SNPs matched these criteria and they were confirmed by capillary sequencing. Among these polymorphisms only two were confirmed in all the 12 affected family members (Tab 3).

	EY01	EY11
SNPs in target regions	1297	1229
Non-described SNPs	384	340
In a gene	235	227
In coding regions	43	50
Nonsynonymous	15	21
Common heteroz. alterations	5	
Confirmed by capillary seq.	5	
Confirmed in all affected patients	2	

Tab 3 Filtering of the next generation sequencing data.

The first SNP was found in the gene centriolin (CEP110) causing a leucine to serine change at position 954 (L954S) and the second one in the gene gelsolin (GSN) causing a glycine to cysteine change at position 254 (G254C). In order to exclude

that these alterations are rare polymorphisms, a screening of the 2 SNPs was performed in normal controls (NC). The alteration L954S in CEP110 was found in 7 NC on 307 screened, suggesting that it is a rare polymorphism with an allele frequency in the population of 1.1%. On the contrary, the alteration G254C in gelsolin was not detected in 443 normal controls screened and can be considered as a candidate mutation rather than a polymorphism.

3.4.2 Screening of GSN mutation in HT families and sporadic ET patients

Since GSN was the only candidate gene fitting all the criteria, 13 other families affected by Hereditary Thrombocythemia and 240 sporadic MPD patients were screened for mutations in this gene. The screening of the 13 families was done by capillary sequencing and it revealed the presence of sequence alterations in 3 of them. In one family, 2 SNPs were found in exon 13 (T563S) and exon 14 (T616M). These alterations were homozygous in the father (the propositus) and heterozygous in the two children. Another pedigree showed a sequence alteration in the intron between exon 6 and 7. However, all these variations had already been reported as known SNPs. In another family, a SNP was found in exon 16 (T695M), which was not present in all the affected family members. The cohort of 240 sporadic MPD patients was analyzed by next generation sequencing. The screening identified 7 alterations (Fig. 7). Among these, only 3 alterations were confirmed by capillary sequencing: V179M, V460M and V606M. The first two were recurrent alteration because they were observed in two different ET patients, whereas the third one was present only in one patient. These SNPs have been already reported in the common SNP database with a frequency in the population lower than 1%. In some patients, the alteration was found to be of germline origin (Tab 4). Interestingly, the variation V179M can be present in a patient as a germline mutation or can occur as a somatic event.

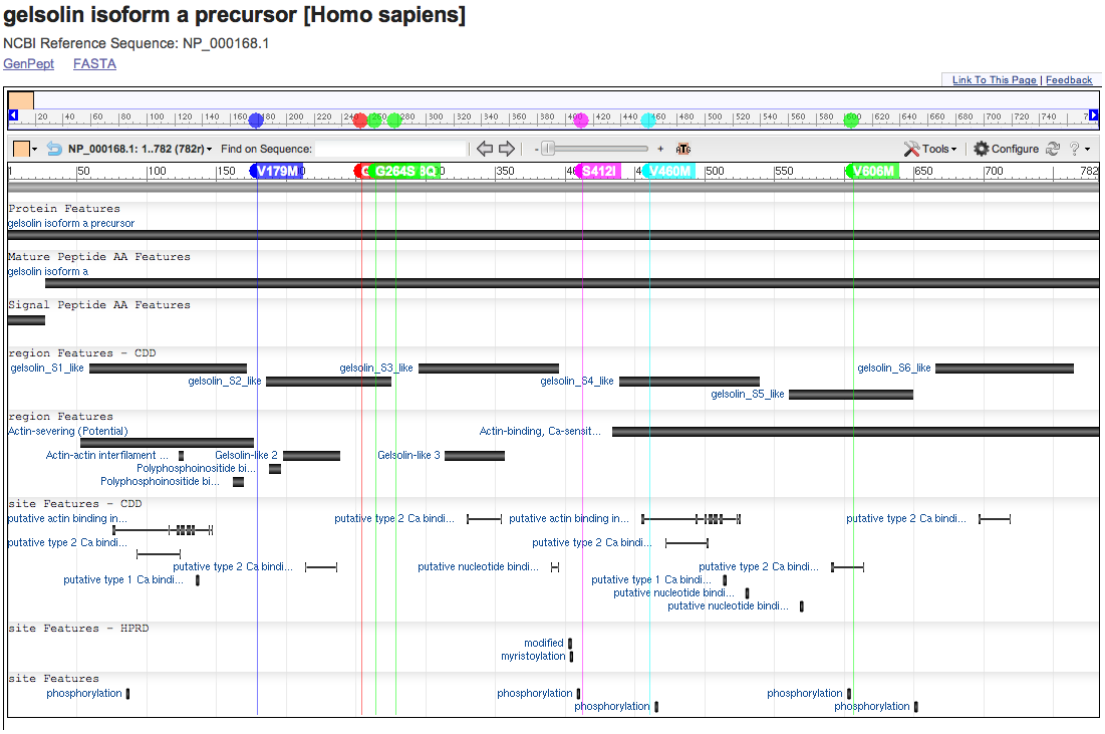


Fig. 6 Position of the 7 different alterations found in sporadic ET patients by NGS. In order: V179M, G254C, G264S, P278Q, S412I, V460M, V606M.

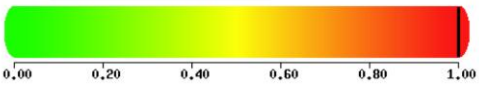
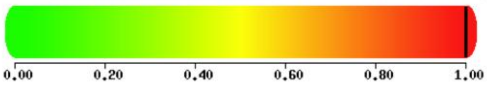
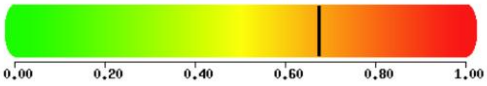
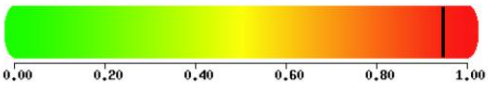
Patient (diagnosis)	JAK2 V617F status	Alteration (granulocytic DNA)	Presence (germline DNA)	SNP frequency
p060 (ET)	-	V179M	YES	0.003708
p163 (ET)	-	V179M	NO	
p273 (ET)	97%	V460M	Not available	0.000659
p325 (PMF)	-	V460M	YES	
p291 (PV)	87%	V606M	YES	0.000660

Tab 4 Alterations found by next generation sequencing and confirmed by capillary sequencing in sporadic patients. Sequencing was performed on granulocytes DNA and germline DNA. ET, essential thrombocytemia; PV, Polycythemia vera; PMF, primary myelofibrosis.

3.4.3 Characterization of the GSN variations found

In order to predict if the amino acid substitution can potentially affect protein function, computational predictions were performed with two different software: SIFT (Sorting Intolerant From Tolerant) and PolyPhen-2. SIFT uses sequence homology to predict

whether an amino acid change will affect protein function and potentially alter the phenotype. This algorithm considers the position at which the change occurred and the type of amino acid change. SIFT predicts substitutions with score less than 0.05 as deleterious. Polyphen-2 uses eight sequence-based and three structure-based predictive features to evaluate the impact of the amino acid change. The SNP alteration in GSN gene were analyzed with both software and the results are summarized is Table 5. Interestingly, the alteration G254C identified in the HT family is predicted to alter the protein function with both algorithms and with the most significant score values. Another interesting alteration is V179M, which is predicted to be probably damaging with both software, even if the score values are less impressive than the ones of G254C.

Mutation	**SIFT**	**PholyPhen**
G254C	0.00	This mutation is predicted to be **PROBABLY DAMAGING** with a score of **1.000** (sensitivity: **0.00**; specificity: **1.00**) 
V179M	0.04	This mutation is predicted to be **PROBABLY DAMAGING** with a score of **0.999** (sensitivity: **0.14**; specificity: **0.99**) 
V460M	0.01	This mutation is predicted to be **POSSIBLY DAMAGING** with a score of **0.675** (sensitivity: **0.86**; specificity: **0.92**) 
V606M	0.22	This mutation is predicted to be **POSSIBLY DAMAGING** with a score of **0.947** (sensitivity: **0.79**; specificity: **0.95**) 

Tab 5 Computational predictions of the different amino acid change in GSN gene with the SIFT and PholyPhen software.

Computational analysis must be complemented with structural analysis. Using the Consurf server is possible to estimate the evolutionary conservation of amino acid positions in a protein molecule based on the phylogenetic relations between homologous sequences and to visualize the desired position in the structure of the molecule. The degree to which an amino acid position is evolutionarily conserved is strongly dependent on its structural and functional importance; rapidly evolving positions are variable while slowly evolving positions are conserved. Usually, the

functionally important regions of the protein are highly conserved. The crystal structure of gelsolin domains G1-G3 bound to actin was analysed with the Consurf server. Gelsolin was visualized using a space-filled model and actin with sticks structure. The mutation G254C and the polymorphism V179M were localized and marked with yellow haloes. Interestingly, the positions of both alterations are conserved. The position 254 has a conservation score of 9 (bounds of the confidence interval 9-9) with few residues variety (S,A,G) and the position 179 has a conservation score of 8 (bounds of the confidence interval 8-7) with six residues variety (A,Q,T,P,I,V). Both the variations are located at the interface with actin.

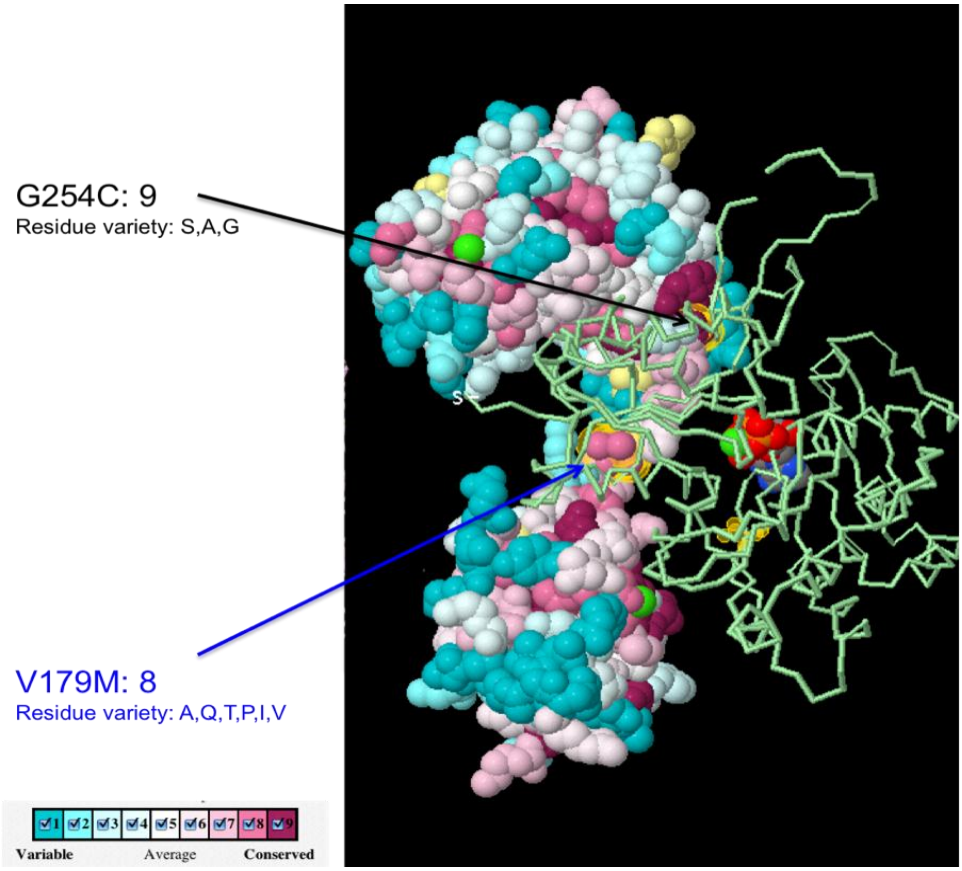


Fig 7 Crystal structure of human GSN domains G1-G3 bound to actin analyzed with Consurf server. The 3D structure of gelsolin bound to actin is presented using a space-filled model. The amino acids are colored by their conservation grades using the color-coding bar, with turquoise-through-maroon indicating variable-through-conserved. Positions, for which the inferred conservation level was assigned with low confidence, are marked with light yellow. Amino acid of the variations G254C and V179M are marked with yellow haloes.

3.4.4 Gelsolin expression in sporadic ET patients

In sporadic patients affected by essential thrombocytosis, where the disease is not caused by an inherited mutation, it could be possible that alterations in gelsolin gene could affect not only the sequence itself, but also the expression level. Therefore in addition to the sequencing data, GSN expression in platelets of ET patients was investigated. For the analysis, 57 sporadic ET patients were included and subdivided according to their JAK2 V617F status. Fifteen normal controls were used as reference in the study. GSN expression in JAK2 V617F negative and JAK2 V617F positive ET patients was not significantly different from the normal control (Fig. 8a). Also patients EY11 (affected) and EY18 (unaffected) were included in the analysis. Interestingly, patient EY11 had a higher GSN expression (value 10.98) compared to patient EY18 (value 0.08). Unfortunately, it was not possible to include more EY patients in the analysis because of the difficulties to get viable material from the patients who live in the United States. No correlation between GSN expression and JAK2 V617F allele burden was found (Fig. 8b).

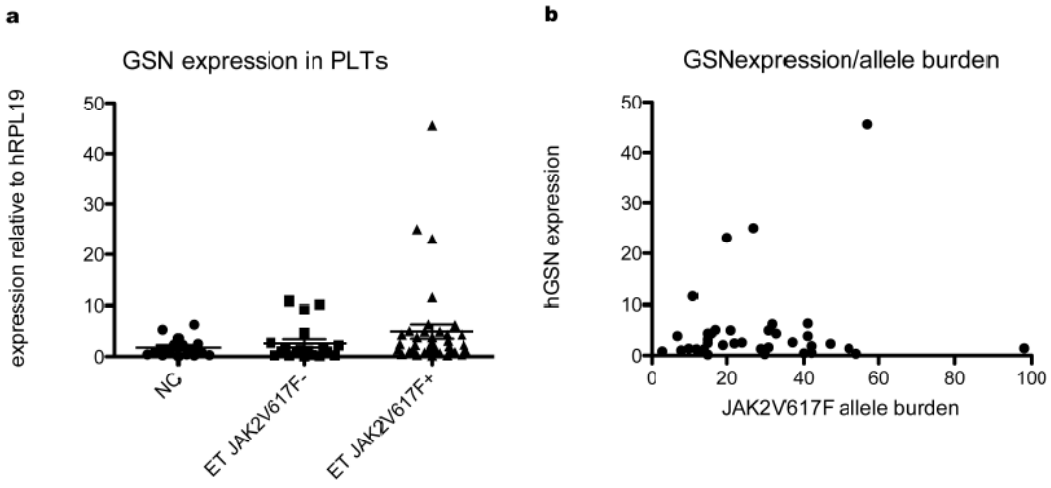


Fig 8 (a) GSN expression in platelets of sporadic ET patients relative to the housekeeping gene RPL19. 15 normal controls, 19 sporadic ET patients JAK2 V617F negative, 38 sporadic ET patients JAK2 V617F positive have been included in the study. (b) Correlation between GSN expression and JAK2 V617F allelic burden.

3.5 GSN: a potential candidate gene

3.5.1 Structural analysis of the mutation G254C found in the HT family

Gelsolin is a Ca^{2+} regulated actin filament severing, capping and nucleating protein. It is involved in the regulation of cell structure and metabolism. Interestingly, it has a key-role in regulating apoptosis and platelet modulation. Gelsolin is composed of six domains named (from the N-terminus) as G1-G6. Each domain contains a Ca^{2+}-binding site. A unique feature of GSN is that apart from the cytoplasmic protein found in most cell types, a secreted plasma form can be generated by alternative splicing in muscle cells. Cytoplasmic as well as secreted gelsolin are the most potent actin filament severing proteins identified to date [66]. From the analysis of the crystallography structure of GSN (domains G1-G3) and actin, it is possible to identify the position of the alteration G254C (Fig. 9a). GSN protein is marked in violet and actin in green. The glycine at position 254, marked in yellow, is localized at the interphase with actin molecule. The distance between the $C\alpha$ of the amino acid glycine of gelsolin and the $C\alpha$ of the serine on actin is 4.2 Angstrom (Fig 9b). The change of glycine to a cysteine (G254C) in the position 254 can potentially affect the interaction between the two molecules.

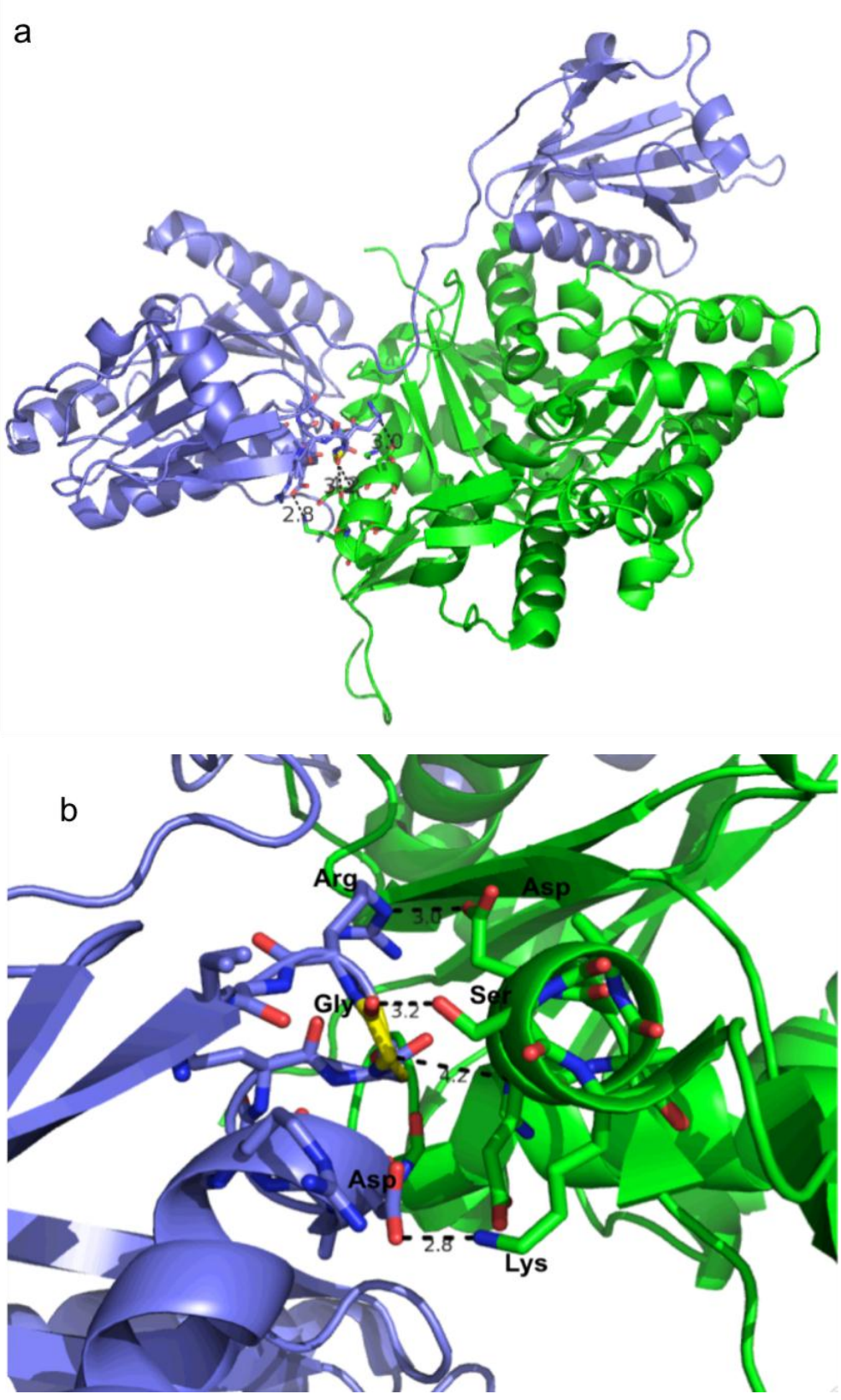


Fig 9 (a) Crystal structure of gelsolin protein domains G1-G3 (violet) bound to a molecule of actin (green). (b) Zoom on the glycine (yellow) at position 254 of the gelsolin protein.

3.5.2 Caspase-3 cleavage of GSN protein

Apoptosis is a fundamental physiological process required for the development and homeostasis of different tissues. Caspase-3, an important effector of apoptosis, cleaves gelsolin in two fragments: N-terminal gelsolin and C-terminal gelsolin. The N-terminal GSN fragment not only severes actin in a Ca^{2+}-independent manner, but also contributes to morphological changes of apoptosis. De Botton *et al* found active forms of caspase-3 and caspase-9 in cultured human megakaryocytes and observed cleavage of substrates such as gelsolin in cells that had begun to shed platelets [18]. To investigate whether the presence of the mutation G254C could affect the caspase-3 activity, an *in vitro* cleavage assay was performed. This test showed that GSN G254C is cleaved more efficiently compared to GSN wild type. The C-terminal fragment of GSN G254C was visible after 30 minutes of caspase-3 incubation, whereas for the wild type GSN the fragment appeared only after one hour (Fig 10).

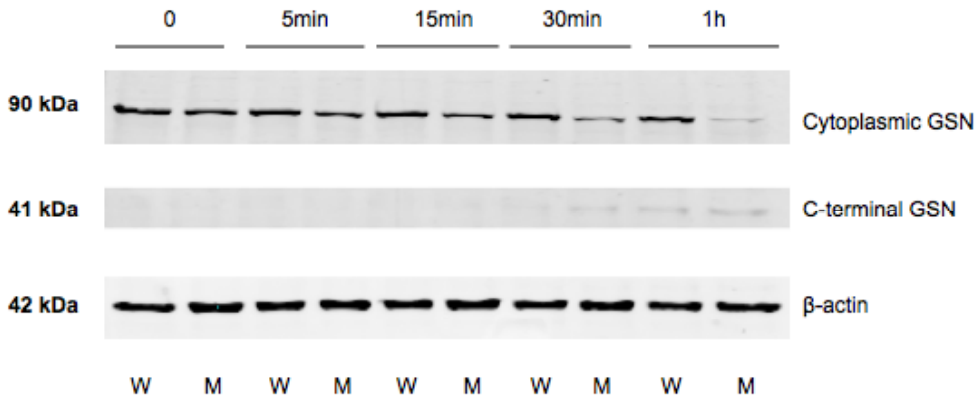


Fig 10 Time curve of caspase cleavage assay. W: human gelsolin cytoplasmic wild type, M: human gelsolin cytoplasmic G254C. Actin: loading control.

3.5.3 Gelsolin translocation into the nucleus

Gelsolin can translocate into the nucleus and participate in signal transduction. Caspase-3 cleavage products or full-length gelsolin can act as inhibitory or stimulatory factors in apoptosis. To investigate whether the alteration G254C can influence the ability of GSN to translocate into the nucleus, nuclear and cytoplasmic fractions of gelsolin stable transfected cell line were analyzed by western blot. No

alterations in GSN translocation were observed between cell line stable transfected with the wild type gene and with the altered gene (Fig 11).

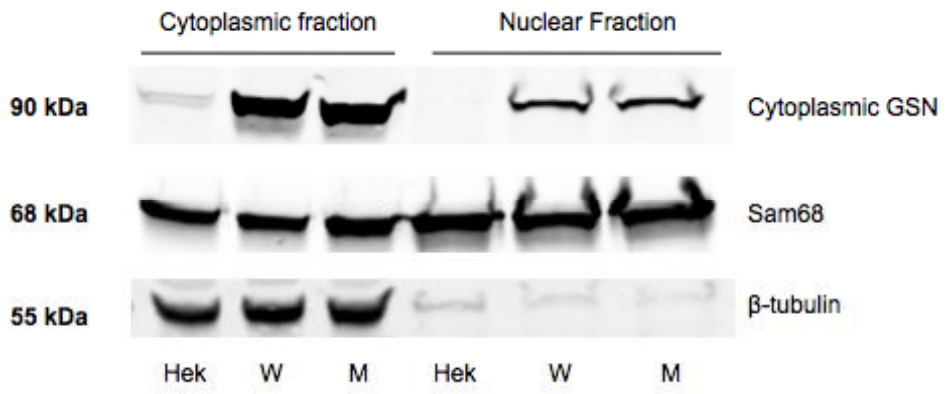


Fig 11 Gelsolin translocation into the nucleus. Hek: HEK 293T cells (negative control), W: HEK 293T cells stable transfected with human gelsolin cytoplasmic wild type, M: HEK 293T cells stable transfected with human gelsolin cytoplasmic G254C. Sam68: nuclear marker, Tubulin: cytoplasmic marker.

3.5.4 Platelets biogenesis assay

The megakaryoblastic cell line DAMI is derived from the peripheral blood of a patient with megakaryoblastic leukemia. These cells can be induced to differentiate along the megakaryocytic lineage when stimulated with phorbol myristate acetate (PMA) and TPO. After differentiation, DAMI cells can release platelets-like particles in the culture supernatant [67]. DAMI cell line stably transfected with wild type GSN and GSN G254C were established and used for the platelets biogenesis assay. During the PMA and TPO stimulation, DAMI cells stopped to proliferate and started to increase their adherence to the plastic culture plates and to spread (Fig 12a). After 7 days of stimulation, DAMI cells started to release platelets-like particles in the supernatant and they could be quantified and identified with the markers CD61 and CD41 by FACS. DAMI cells expressing the mutant GSN (G254C) released higher amounts of CD61+CD41+ particles in the culture supernatant compared to untransfected DAMI cells and DAMI cells expressing wild type GSN (Fig 12b).

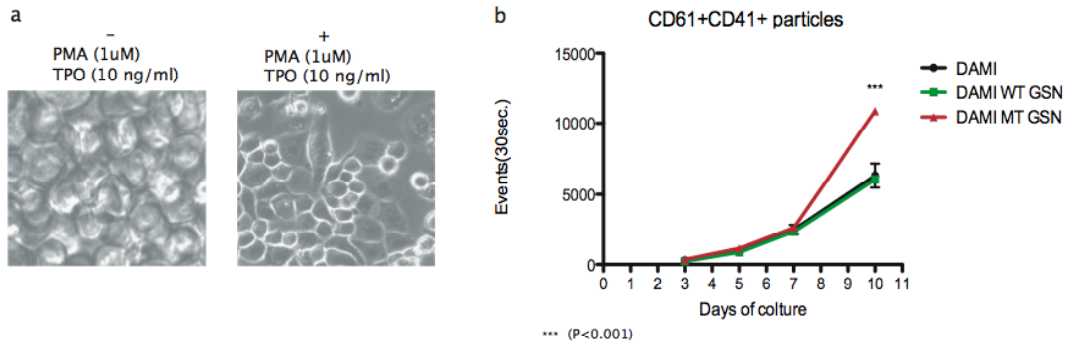


Fig 12 (a) DAMI cell line differentiates along megakaryocytic lineage upon stimulation with PMA (1uM) and TPO (10 ng/ml). Cells stop proliferating and start to adhere to the plastic culture dish. (b) PLTs-like particles released in the culture supernatant. DAMI, cell line not transfected; DAMI WT GSN, DAMI cells stable transfected with wild type GSN; DAMI MT GSN, DAMI cells stable transfected with mutant GSN. Results were compared with 2way ANOVA (***p<0.001).

3.6 ANIMAL MODEL

3.6.1 Transplantation with bone marrow cells transduced with GSN

To study the potential in vivo role of GSN G254C in the pathogenesis of HT, lethally irradiated recipient mice were transplanted with BM cells transduced with retrovirus expressing human cytoplasmic mutant GSN-IRES-GFP. As a control, a virus overexpressing human cytoplasmic wild type GSN-IRES-GFP was used. In each group, 4 recipient mice were transplanted with 15% GFP positive BM cells. Mice were monitored 30 days after transplant and followed for 8 months. At each time point, chimerism and blood counts were investigated. The chimerism was calculated as the percentage of GFP positive cells for different cell lineages: erythrocytes (Ter119+), platelets (CD61+), granulocytes (Gr-1+) and monocytes (Mac1+). The percentage of GFP positive cells in each mouse showed a high variability. The mouse MT2, with stable GFP positive platelets close to 90 % from day 60 post-transplant, showed a mild increase in platelet counts in the group of animals transplanted with BM cells transduced with human cytoplasmic mutant GSN-IRES-GFP whereas the other animal from this group exhibited entirely normal blood counts (Fig.13). No mice from the control group had variations in platelet counts and any other blood parameters.

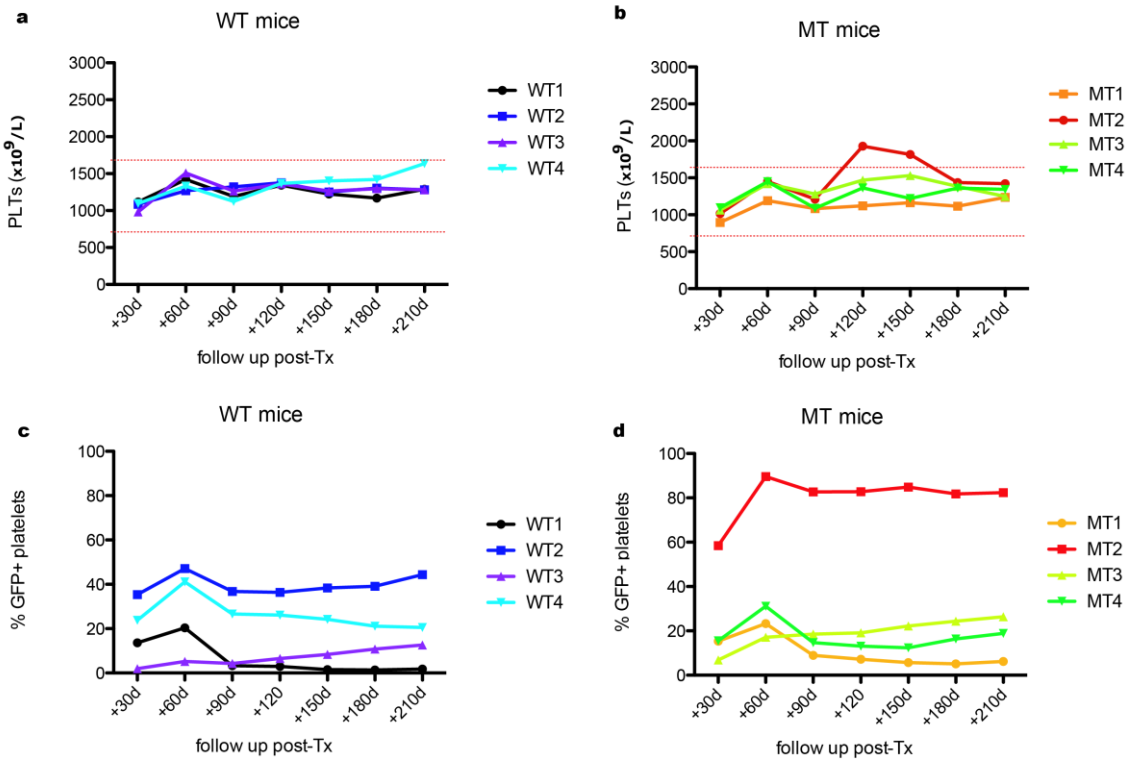


Fig. 13 (a) Platelet counts in the group of mice transplanted with human cytoplasmic wild type GSN-IRES-GFP. (b) Percentage of GFP positive circulating platelets in mice transplanted with human cytoplasmic wild type GSN-IRES-GFP. Platelet chimerisms were measured according to the double-positive population CD61+/GFP+. (c) Platelets count in the group of mice transplanted with human cytoplasmic mutant GSN-IRES-GFP. (b) Percentage of GFP positive circulating platelets in mice transplanted with human cytoplasmic mutant GSN-IRES-GFP. Platelet chimerisms were measured according to the double-positive population CD61+/GFP+.

GSN expression was analyzed in peripheral blood (120 days after transplant) and in platelets (210 days after transplant) by real time PCR with primers human specific (Fig. 14). Human gelsolin expression in each mouse was quite variable as expected from the GFP signal. The mouse MT2 showed a high relative human gelsolin expression in peripheral blood and platelets. This data correlated with the high GFP chimerism observed by FACS analysis.

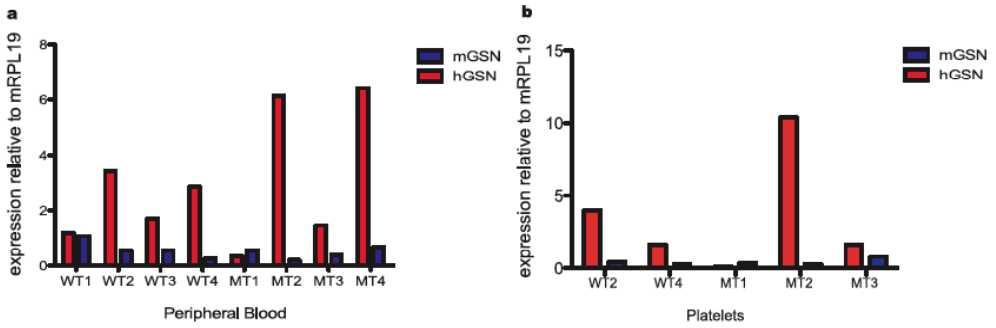


Fig. 14 (a) Human gelsolin expression (red bar) and mouse gelsolin expression (blue bar) for each mouse in peripheral blood 120 days after transplant. (b) Human gelsolin expression (red bar) and mouse gelsolin expression (blue bar) in mouse platelets at 210 days after transplant.

Cytoplasmic gelsolin can be involved in apoptosis regulation. To assess whether G254C GSN can induce a longer platelets survival in transplanted mice, platelets clearance was analyzed in vivo. By tracking the survival of biotin-labeled platelets in vivo at different time point, it was possible to analyze the life span of GFP positive and GFP negative platelets in mice (Fig. 15). No differences in platelets life span were observed between GFP positive and negative platelets in both groups of mice.

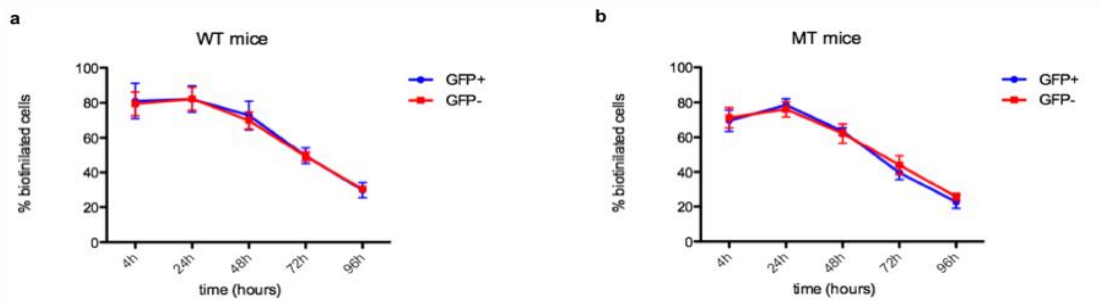


Fig. 15 (a) Life span of biotinylated platelets in mice group transplanted with human cytoplasmic wild type GSN-IRES-GFP. (b) Life span of biotinylated platelets in mice group transplanted with human cytoplasmic wild type GSN-IRES-GFP. Peripheral blood samples were taken at 4, 24, 48, 72 and 96 hours after injection with NHS-biotin.

3.6.2 Generation of GSN G254C transgenic mice

To establish a more stable animal model expressing both cytoplasmic and plasma gelsolin, BAC-transgenic mice that express the human GSN-G254C driven by the endogenous human GSN promoter were generated. The 140-kb BAC clone contained 70 kb of GSN 5′-upstream region that does not include other known genes, as well as exons 1-17 and a part of intron 17 of the human GSN gene. Using homologous recombination in bacteria, the missing part of intron 17, exon 18 and 2.4 kb of 3'-downstream region were added. The fragment used for homologous recombination was excised from the BAC clone RP11-269G20 with BglII and XbaI digestion (Fig. 16). The insertion of the mutation was performed with a BAC recombineering approach using galK selection cassette. This targeting step introduced the G to T transversion at the desired position of GSN (Fig. 17).

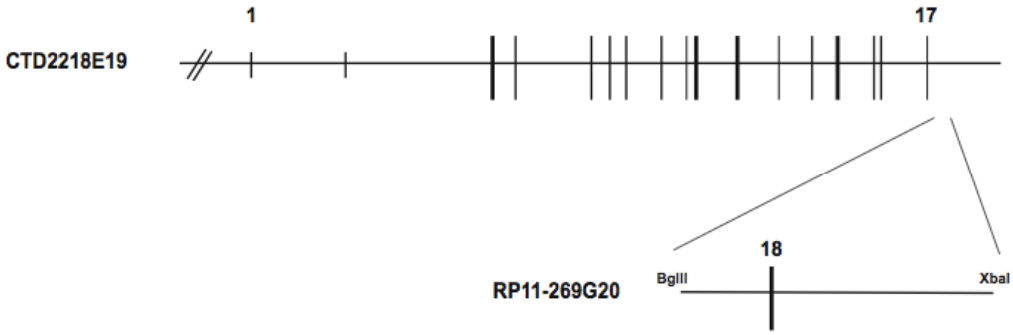


Fig.16 Transgenic construct. BAC clone CTD2218E19 was modified by the addition of the last exon and 3' UTR sequence of GSN. Homolougus recombination was induced with the excised fragment cut with BglII and XbaI from BAC clone RP11-269G20.

The repaired BAC clone CTD-2218E19 was linearized with the unique enzyme cutter PmeI and microinjected into oocytes from inbred C57BL/6 mice. Oocyte injection of this modified BAC construct, yielded 5 transgenic lines in the inbred C57BL/6 background named from G1 to G5.

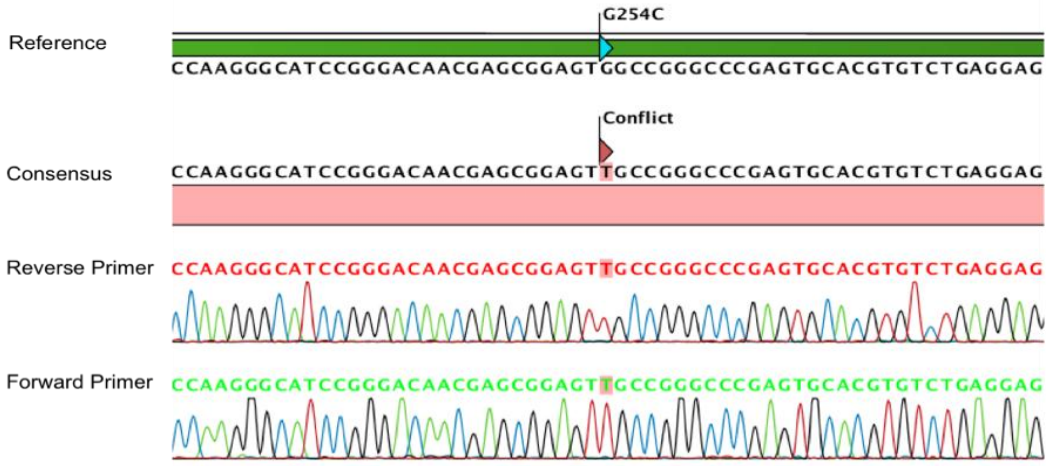


Fig. 17 Insertion of the G to T transversion at the desired position of GSN gene.

To verify that the repair of the BAC clone at the 3' end did not affect the correct transcription of the transgene, the integrity of the human gelsolin messenger RNA was checked in each founder. A PCR reaction was designed using two different human specific primer sets annealing in the middle and at the end of the transcript. Amplification was obtained with both primer sets showing that the splicing process occured correctly and therefore the human transgene was fully transcribed in mice (Fig.18).

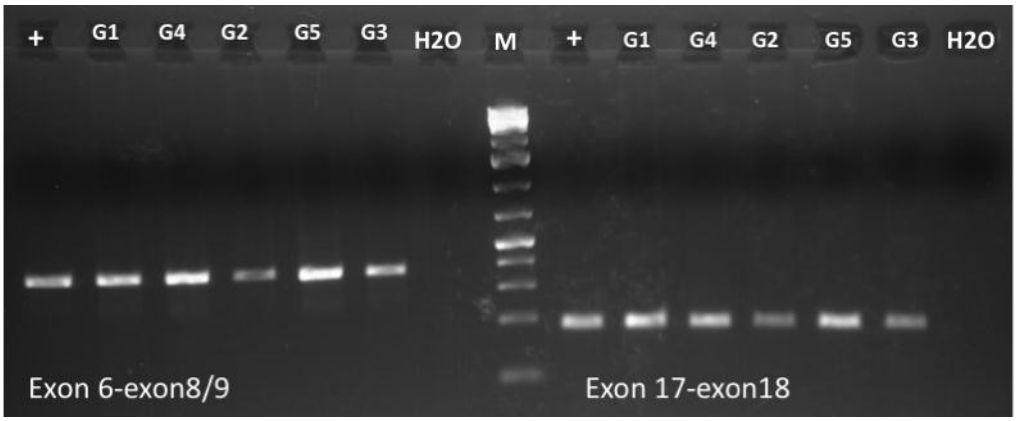


Fig.18 Amplification of the middle part of human gelsolin transcript (region between exon 6 and exon 8/9) and the last part of the transcript (region between exon 17 and exon 18). Primers were designed to be human specific and to not cross-react with the mouse endogenous GSN gene.

The 5 transgenic lines had different transgene copies as well as human GSN expression (Fig.19). The transgene was not passed to the progeny in the G1 line, G2 and G3 founders had a too low copy number, G4 founder was sterile, while G5 founder bred efficiently with a very high transgene expression.

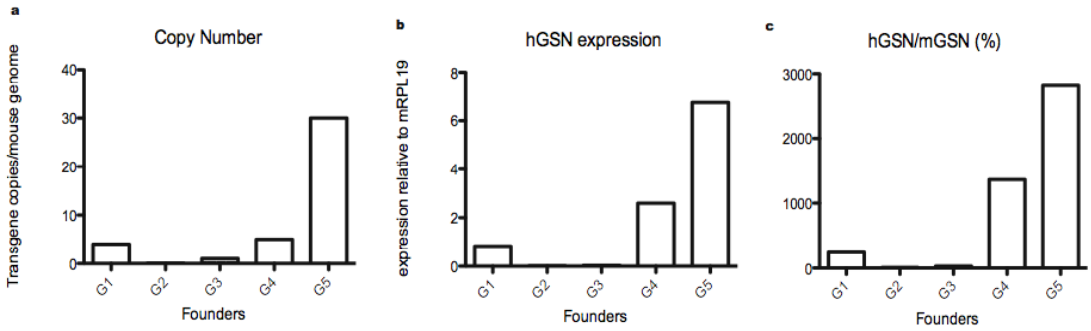


Fig.19 Transgene copy number and mRNA expression. (a) Transgene copy number per mouse genome. The numbers were assessed in the “Absolute Quantification” setup with standard curves made from human genomic DNA isolated from a normal control. (b) Relative expression of human gelsolin (hGSN). Real-time PCR was performed with primers specific for the human GSN transgene. The numbers represent relative expression values calculated by the ΔCT method after normalization to the mRNA of mouse RPL19 (c) Human/mouse GSN mRNA ratio. Each GSN primer pair was tested to have the same efficiency. Therefore, ratios of human and mouse GSN were assessed in the “Relative quantification” setup using the $2^{-\Delta CT}$ method.

The difference in gelsolin expression between G3 and G5 transgenic line was evident also in the platelets lysates isolated from the first progeny derived from the two different founders (Fig.20). G5 mice express high quantity of human gelsolin compared to wild type mice either C57BL/6 or G5 wild type littermate.

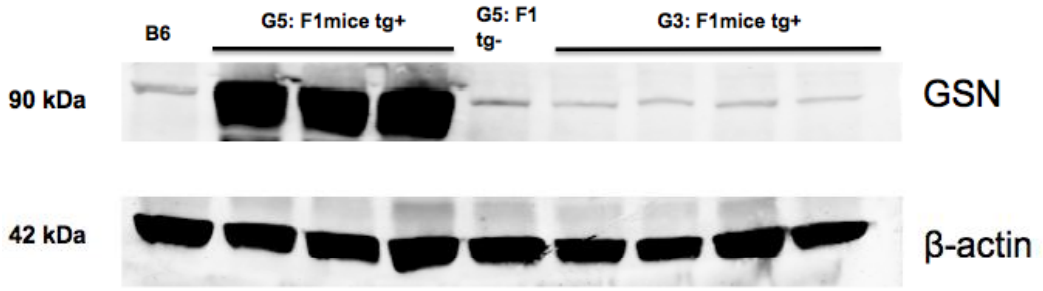


Fig. 20 GSN protein in platelets lysate from the first progeny of G5 (G5: F1mice tg+) and G3 transgenic line (G3: F1mice tg+). A wild type C57BL/6J mouse (B6) and a G5 wild type littermate (G5: F1 tg-) were included in the blot as reference. Actin was used as loading control.

Transgenic mice derived from G3 strain did not show any abnormalities in different blood parameters as expected from the low expression of the transgene (Fig. 21a). Therefore the analysis of this strain was stopped at week 22. The G5 derived mice, instead, developed a mild phenotype showing a tendency to elevated platelet counts compared to their wild type littermates (Fig. 21b). The platelet counts were found to be significantly higher in G5 transgenic mice compared to wild type littermates (Mann-Whitney test, P = 0.004). No other alterations were observed in blood counts in these mice.

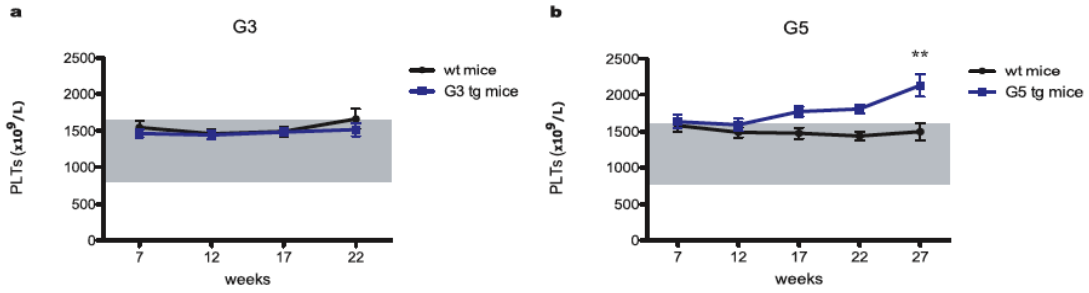


Fig 21 Time course of platelet counts in transgenic and control mice. Platelets counts of mice at 7 weeks, 12 weeks, 22 weeks and 27 weeks of age are shown. Black symbols, control mice; blue symbols, transgenic mice. Physiological ranges are indicated with a gray box. (a) G3 transgenic mice, (b) G5 transgenic mice. Results were compared with Mann-Whitney test (**p=0.004)

Expression of human gelsolin in G5 transgenic mice was consistent, even if some variations were observed in different mice (Fig 22a). Also the expression of endogenous gelsolin was variable within the mice. Therefore the ratio between human GSN and mouse GSN did not correlate with the expression of the transgene (Fig 22b).

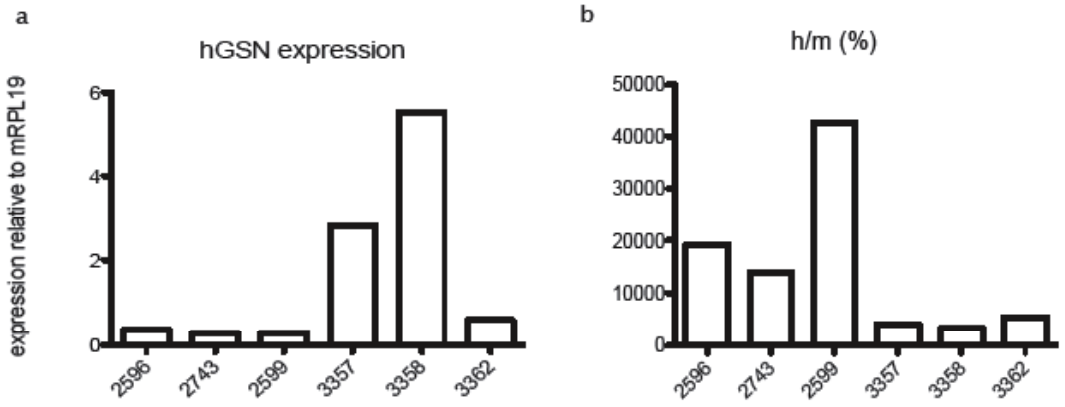


Fig 22 (a) Human transgene expression in whole blood from 6 mice of G5 strain. (b) Human/mouse GSN mRNA ratio.

4. DISCUSSION

My PhD studies focused on a large US family affected by hereditary thrombocythemia. The patients present with persistent elevation of platelet counts, occasional giant platelets with abnormal aggregation, isolated hyperplasia of enlarged megakaryocytes and splenomegaly. So far, HT has been described as a rare benign disorder possibly caused by mutations in the THPO or the c-Mpl receptor genes. Interestingly, in the pedigree studied no alterations were described in these two genes. Therefore, the aim of my work was to identify new germ line mutations causing HT, which can be also relevant in understanding MPD pathogenesis and eventually represent a new target for drug development. Using linkage analysis and subsequent targeted next generation sequencing, a novel candidate mutation in gelsolin gene (GSN) was discovered in the affected family members. This alteration affects the position 254 of the protein causing a major amino acid change from glycine to cysteine, thus altering a highly conserved position of the protein. Structural analysis revealed that the mutant position is at the interface with actin and the distance between the Cα of the glycine on GSN and the Cα of the serine on actin is only 4.2 Angstrom. The substitution of a glycine with a cysteine at this position can easily influence the interaction between the two molecules. In addition, gelsolin is abundant in platelets and therefore an interesting candidate gene. Even though the gelsolin superfamily proteins are mainly involved in the regulation of actin dynamics, they can be multifunctional. Gelsolin, for example, can be involved in apoptosis regulation, modulation of platelets, signal transduction and transcriptional activation [66]. To functionally characterize the GSN mutation, *in vitro* and *in vivo* studies have been performed. DAMI cell line can differentiate into megakaryocytic lineage when stimulated with PMA and TPO providing a good system to evaluate cell maturation and platelets production. In my experiment, DAMI cells overexpressing GSN G254C released more platelet-like particles in the supernatant compared to DAMI overexpressing wild type GSN. The in vivo role of the mutation was investigated with the generation of BAC transgenic mice expressing the human GSN-G254C. These

mice showed a mild tendency to thrombocytosis compared to their wild type littermates.

Up to date, the only mutation reported in gelsolin gene was found to be the cause of the familial amyloidosis of Finnish type (FAF). This mutation causes an amino acid substitution of aspartic acid at residue 187 by either asparagine or tyrosine (in a danish FAF family) creating a conformation that predisposes GSN to proteolysis. This aberrant cleavage generates the FAF amyloid precursor fragment of 68 kDa. Further cleavage at position 244 is required for the generation of FAF amyloid protein [68]. A transgenic mouse model accurately recapitulates the proteolytic cascade that yields the 8- and 5-kDa amyloidogenic peptides putatively causing FAF. Notably, the amyloidogenesis is restricted to the muscle tissues where D187N gelsolin is synthesized [69]. However, so far no mutation in gelsolin have been described in hematological disorders.

Even though there is no evidence showing gelsolin involvement in thrombocytosis, it is likely that it can play a role in the disease by influencing the apoptotic processes in platelets biogenesis and/or favoring the platelets release. The role of apoptosis in the megakaryocyte lineage is quite intriguing. Megakaryocytes presumably need to prevent the activation of the apoptotic machinery in order to survive and produce platelets. Various insults, such as chemotherapy and autoantibodies, have been suggested to induce apoptotic death of megakaryocytes and progenitors. Conversely, apoptotic processes have been implicated in megakaryocyte development and platelet production. Since the initial observation that platelets release from megakaryocytes resembles the onset of apoptosis [14], an increasing body of evidence suggested that platelets shedding is an apoptotic process. It is clear that megakaryocytes possess both an intrinsic and an extrinsic apoptosis pathway. However, the control of the apoptosis and its role in MK and platelets biogenesis remain to be established. On the contrary, it is well-founded that platelets possess an intrinsic apoptosis pathway and can undergo apoptosis [70 71]. Bcl-XL and Bak are the major components of a “molecular clock” that determines platelets life span. Degradation of Bcl-XL triggers Bak-mediated apoptosis and clearance from circulation [15]. Interestingly, gelsolin is a substrate of caspase-3 and activated forms of caspase-3 were detected in maturing megakaryocytes [18]. Caspase-3 cleaves gelsolin

between residue Asp352 and Gly353 resulting in the generation of two fragments: N-terminal GSN and C-terminal GSN. Deoxyribonuclease (DNase I) is a key enzyme involved in the DNA degradation process occurring in apoptosis. Gelsolin-actin-DNase I form a ternary complex sequestering the enzyme and therefore showing an anti-apoptotic function. N-terminal gelsolin can competitively bind to actin causing the release of DNase I from actin to enhance apoptotic activity. On the contrary, the C-terminal GSN is anti-apoptotic and has a function similar to Bcl-XL, the major regulator of platelets survival [66]. In my studies, the *in vitro* caspase-3 cleavage assay showed a more efficient cleavage of mutant gelsolin compared to the wild type. Since proplatelets formation is regulated by caspase activation and GSN cleavage is a hallmark of maturing MKs, the mutation G254C could potentially lead to a more efficient proplatelets formation by enhancing caspase-3 gelsolin cleavage. In addition, it could also be postulated that in the cytoplasm of platelets expressing the mutant GSN there is a higher amount of N-terminal GSN and C-terminal GSN. Considering that N-terminal GSN function is based on its capacity to bind actin and this binding might be altered by the mutation (as shown from the structural analysis), it is possible that the anti-apoptotic function of C-terminal GSN prevails the one of N-terminal GSN. According to this model, platelets half-life should be increased in patients expressing mutant gelsolin. However, the *in vivo* platelets clearence assay in transplanted mice did not support this hypothesis. Life span of platelets overexpressing human cytoplasmic wild type GSN or mutant GSN was not increased compared to normal platelets.

On the other hand, gelsolin can play a role in actin polymerization during platelets shedding promoting proplatelet branching. Even though, the role of actomyosin in platelets biogenesis is not completely understood, actin polymerization is required for proplatelets branching and thus may be important for the regulation of platelets production [10]. It has been shown that megakaryocytes treated with one of the actin toxins cytochalasin and latrunculin can extend long proplatelets, but fail to branch [13]. In addition, Zunino *et al* showed the involvement of scinderin, a Ca^{2+}-dependent filamentus actin severing protein similar to gelsolin, in megakaryocytes apoptosis and platelets fragmentation. Scinderin was found to be expressed in human BM cells together with the megakaryocytes-specific markers, but absent from megakaryoblastic leukemia cells and megakaryocytic cell lines (Meg-01, HEL, K562

and HL-60) incapable of producing platelets in culture. Expression of scinderin in the Meg-01 cell line decreased cell proliferation and induced cell polyploidization, differentiation and apoptosis inducing the release of platelet-like particles [72]. According to these observations, mutant gelsolin could increase platelets production by playing a role in the actin dynamics. In support of this model, the *in vitro* platelets biogenesis assay showed that DAMI cells expressing the mutant GSN release higher amounts of platelets-like particles in the culture supernatant compared to DAMI cells not transfected and DAMI cells expressing wild type GSN. The role of the mutant gelsolin and the actin polymerization is still not completely understood. However, the generation of transgenic mice expressing the mutant gelsolin supported a role of this alteration also *in vivo*. Mice expressing the mutant GSN developed a mild phenotype showing an increase in platelet counts without any alteration in other cell lineages.

In the affected family members HT shows an early onset in life with only a mild thrombocytosis. Therefore, the moderate phenotype observed in the transgenic mice fits into this clinical picture. One factor that can influence the severity of the phenotype could be the genetic background of the mice. This is a possibility that is worth to investigate since it has been reported that the genetic landscape plays a role in the development of myeloproliferative disorders in humans. As a matter of fact, on the background of a JAK2 or MPL mutation and/or of any other pre-existing or additional molecular abnormality, individual-related characteristics contribute to the phenotypic pleiotropy of the MPDs. This was supported also by results in animal models. Depending on the genetic background of mice transplanted with JAK2V617F transfected cells, the characteristics of their myeloproliferative disease were somewhat different. In addition to polycythemia, the JAK2V617F allele induced marked leukocytosis and marrow reticulin fibrosis in Balb/c mice, while leukocytosis was minimal and fibrosis was limited to the spleen in case of C56Bl/6 strain [73]. It has been proved that the genetic background can cause considerable variation in the phenotype of genetically engineered mice, possibly due to the presence of multiple functional interactions of different alleles at modifier loci [74]. Indeed, the phenotype of $GSN^{-/-}$ mice can vary significantly depending on the genetic background. To investigate the critical function of gelsolin in a vertebrate organism in the presence of numerous other protein that also regulate actin assembly, transgenic gelsolin null mice ($GSN^{-/-}$) were generated in the C57BL/6 outbred genetic background and found

to have normal development and longevity. However lack of GSN caused physiological dysfunction during hemostasis and activation of platelets, inflammation response and leukocytes mobility. Interestingly, knockout mice displayed decrease of platelet size and impaired function of platelets leading to prolonged bleeding time [75]. A recent finding showed that transferring the null gelsolin allele into the Balb/c inbred genetic background resulted in almost complete loss of embryos, with very few $GSN^{-/-}$ mice escaping embryonic lethality. The lack of gelsolin in BALB/c mice resulted in a defective erythroid maturation, with persistence of circulating nucleated cells in embryonic life and in impaired red blood cell production under stress condition in adult mice. In these mice, the gelsolin deficiency alters the actin polymerization/depolymerization equilibrium in the erythrocyte causing impaired terminal maturation. These data provide a new link between gelsolin and the hematological system and suggest a non-redundant role of the protein in terminal erythroid differentiation [76]. Since the influence of the genetic background could play a major role on the phenotype, it could be possible that backcrossing G5 transgenic mice with BALB/c mice would lead to an exasperation of the phenotype in these mice.

Furthermore, the presence of two endogenous copies of wild type gelsolin in the mice could compete with the human mutant gene and thus repress the phenotype. Therefore, it would be interesting to cross G5 mice with $GSN^{-/-}$ mice in order to reproduce the genetic condition present in patients with the heterozygous mutation.

Another possibility to take into consideration is that the mechanisms regulating platelet production may not be the same between mice and humans. In 2010, Bender *et al* proved that ADF/n-cofillin-dependent actin turnover determines platelets formation and sizing in mice. Their results showed that platelet counts in mice lacking both ADF and n-cofillin in megakaryocytes were dramatically reduced indicating that *in vivo* actin turnover is a critical step in the terminal phase of platelets formation. They conclude that no other actin severing/depolymerizing protein in platelets such as gelsolin or adseverin can functionally compensate for the role of ADF/n-cofillin in platelets morphogenesis [77]. In humans, evidence supports an important role of gelsolin and related proteins in the regulation of platelet formation [72]. Therefore it could be possible that in mice ADF/cofillin family proteins are the major key-player in

platelets formation, whereas in humans this process is mainly regulated by gelsolin family proteins.

5. CONCLUSION

My studies demonstrate that genetic linkage analysis combined with the use of targeted next generation sequencing is a powerful tool to detect the chromosomal location of disease genes involved in the pathogenesis of hereditary thrombocythemia. The discovery of a new mutation G254C in gelsolin gene was possible because of solid results of these genetic analyses and facilitated also by the large size of the family. In my work, I characterized this alteration involved in the pathogenesis of HT. In order to investigate the functional relevance of the mutation, *in vitro* and *in vivo* studies have been performed. The platelets biogenesis assay *in vitro* showed that DAMI cells overexpressing the mutant GSN release higher amounts of platelets-like particles in the supernatant compared to DAMI cells overexpressing wild type GSN. To further characterize the role of the mutation *in vivo*, BAC-transgenic mice expressing the human GSN G254C were generated. Transgenic mice developed a mild phenotype showing a tendency to elevated platelet counts without alterations in any other blood parameters. Taken together, these results provide new evidence that GSN can play a role in platelets biogenesis and that the mutation GSN G254C might be the genetic alteration responsible of HT in the pedigree studied.

Even though GSN is the most prominent candidate gene found and is now shown to confer the mild phenotype observed in HT, there could be additional alterations not detected in the studies. There is a strong likelihood that the co-segregating region found on chromosome 9 is correct, but the genetic analysis of this region might be incomplete. There could be, in fact, mutations or deletion/insertion that might have been missed with the NGS approach. In addition, my studies did not cover the regulatory regions of different genes and the possibility that micro RNA could be present in the region and dysregulated. Therefore, the analysis of the co-segregating region found could be extended. For this purpose, the platelets transcriptome could be analyzed by NGS obtaining information not only about the sequence itself, but also about the gene expression profile. The RNA sequencing can identify and quantify both rare and common transcripts and it is possible to characterize all

transcriptional activity, coding and non-coding [78]. With this approach it would be possible to detect alterations that have been missed with the DNA next generation sequencing and to check the gene expression profile in platelets.

Further investigations are needed to better characterize the G5 transgenic mice and obtain better insights about the pathological mechanism of mutant GSN. It would be interesting to investigate the number of MKs in the bone marrow of the G5 mice, the ploidy status of the marrow MKs and the number of MK precursors in order to gain new inputs of how gelsolin can contribute to the development of thrombocytosis. In addition, it would be interesting to investigate if the treatment of the transgenic mice with the TPO would affect MK expansion and platelet levels to the same extent as it does in wild type mice.

6. MATERIAL and METHODS

Patients and clinical feature

This hereditary thrombocytosis family was reported for the first time in 1986 [63]. Up to now, thrombocythemia has been documented in 16 family members in five successive generations. The diagnostic criteria of the World Health Organization (WHO) were followed for all patients. The study was approved by the local ethics committees and all samples were obtained after subjects provided written informed consent. 12 affected and 11 unaffected family members were enrolled in the study.

13 families affected by hereditary thrombocytosis with no mutation detected as disease-causing gene and 240 sporadic MPD patients were included for further genetic analysis. 443 normal controls were used as reference.

Separation of blood cells and extraction of DNA and RNA

Blood cells were separated by standard protocols using Histopaque (Sigma, St. Louis, MO, USA) gradient centrifugation. Granulocytes and peripheral blood mononuclear cells were collected. Platelets were collected using Sepharose (Amersham Pharmacia Biotech AB, Uppsala, Sweden) gel filtration method [79]. DNA was extracted using the QIAamp DNA Mini Kit (QUIAGEN AG, Hombrechtikon CH). RNA was isolated using the TRIfast reagent (peqLab Biotechnology GmbH, Erlangen, Germany).

Microsatellite analysis

DNA was amplified by PCR using dye-labeled primers for microsatellite markers. The conditions were 94°C for 15 s, 55°C for 15 s, 72°C for 30 s for 10 cycles, 89°C for 15 s, 55°C for 15 s and 72°C for 30 s for 20 cycles. The PCR products were analyzed using the ABI 3100 genetic analyzer and the Genemapper software package version 3.5 (Applied Biosystems, Foster City, CA, USA). Linkage analysis was carried out with FASTLINK software package version 4.1p assuming equal allele frequencies for

the marker alleles and an autosomal dominant inheritance model with 100% penetrance.

SNP CHIP Array

The Affymetrix GeneChip Human Mapping 50K Xba 240 was used to genotype single nucleotide polymorphisms (SNPs) in genomic DNA according to the Affymetrix GeneChip Mapping Assay Manual (Affymetrix Inc., Santa Clara, CA, USA). The SNP calls were generated by GeneChip DNA Analysis Software. dCHIP version 2005 was used to perform parametric linkage analysis [80] . An autosomal dominant inheritance model with a 100% penetrance was used.

Next Generation Sequencing

The enrichment of DNA of the region of interest was performed with two different platforms: Nimblegen enrichment and Agilent enrichment. Briefly, genomic DNA was fragmented using Covaris instrument (Covaris, Woburn MA USA). Fragments were blunt-end repaired and modified with dA-overhang in order to ligate them with paired-end adapters. The DNA prepped was subsequently hybridized to a specific biotinylated RNA library to enrich it for the region of interest on cromosome 9. The sequencing step was performed with Illumina and Roche454 Technology according to the manufacturer's protocol. The data were analyzed using CLC DNA Workbench 4 (CLC bio, Aarhus Denmark).

Genomic DNA sequencing

Candidate genes were sequenced from PCR fragments amplified from genomic DNA. PCR conditions were 95ºC for 2 m, 94ºC for 30 s, 60ºC for 30 s and 72ºC for 45 s for 35 cycles. Sequencing was performed on Applied Biosystems 3700 DNA sequencer (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocols.

cDNA synthesis, and Quantitative RT-PCR

Total RNA (2 µg) was reverse transcribed to generate cDNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA). Quantitative PCR was performed with the Power SYBR green master mix on a 7500 Fast machine (Applied Biosystems, Foster City, CA). The differences of threshold

cycles (ΔC_T) were derived by subtracting the C_T value for the internal reference, ribosomal protein L19 (RPL19), from the C_T values of the marker gene gelsolin (GSN). Relative expression level for each reaction was calculated as $2^{-\Delta CT}$. The primers were designed across exon-intron junctions [81]. Sequences of the primers are listed in Table 1.

Human GSN_F	GGTTCCAACAGCAATCGGTA
Human GSN_R	GTTTGCCTGCTTGCCTTTC
Human RPL19_F	GATGCCGGAAAAACACCTTG
Human RPL19_F	TGGCTGTACCCTTCCGCTT

Tab 1: Primers for gelsolin expression analysis in plateles cDNA from normal controls and ET patients

DNA construct

Human gelsolin cDNA clone in pOTB7 vector was purchased from RZPD (Germany). The full length secreted and cytoplasmic gelsolin cDNA wild type and mutant (G254C) were cloned into retroviral vector pMSCVpuro and MSCV2.2iresGFP encoding green fluorescent protein (kindly provided by Dr. J. Cools, Flanders Interuniversity Institute for Biotechnology, Leuven, Belgium).

Stable transfected cell line

Cell lines (293T and DAMI) were transfected with vectors containing cytoplasmic wild type and mutant GSN cDNA using Turbofect following the manufacturer's protocol (Fermentas Life Science, St. Leon-Rot Germany). After transfection, cells were grown in puromycine selection media. Clones were analysed for gelsolin expression by Western Blot (WB).

Protein extraction and Western blot

Whole cell lysates were extracted in RIPA buffer (25mM Tris-HCl, pH 7.6, 150 mM NaCl, 1% NP40, 0.1% sodium deoxycholate, 0.1% SDS). Preparation of cytoplasmic and nuclear cell extracts was obtained with the use of two different solutions: hypo-osmotic buffer (20mM Hepes pH 7.9, 10mM KCl, 0,1mM NaVO4, 1mM EDTA, 1mM

EGTA, 0.2% NP40, 10% glycerol) to isolate cytoplasmic proteins and hyper-osmotic buffer (420mM NaCl, 20mM Hepes pH 7.9, 10mM KCl, 0,1mM NaVO4, 1mM EDTA, 1mM EGTA, 20% glycerol) to isolate the nuclear extracts.
Indicated amount of protein were separated by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane. The following antibodies were used as primary antibodies: goat polyclonal anti-human gelsolin antibody (clone C20, Santa Cruz Biotechnology, CA, USA), mouse monoclonal anti-βactin antibody (Clone AC-15, Sigma-Aldrich, Buchs CH). IRDye fluorescent secondary antibodies were used and membranes were scanned with the Odyssey infrared imaging system (LI-COR Bioscience, Nebraska USA).

In-vitro Caspase3 cleavage assay

Protein lysates were obtained from GSN wild type and mutant stable transfected 293T cell lines by a step of sonication in Caspase3 reaction buffer (50mM Hepes, pH 7.2, 50 mM NaCl, 0.1% Chaps, 10 mM EDTA, 5% glycerol and 10 mM DTT). A reaction containing 50 μg of proteins and 2 units of active recombinant human Caspase-3 (BioVision, Milpitas, CA USA) was incubated at 37°C. At 0, 5min, 15min, 30min, 1hour 10 μg of proteins have been removed, cooked at 95°C 5 min and analyzed by SDS-PAGE.

Platelets biogenesis in DAMI cells

DAMI cells were cultured in IMDM supplemented with 10% Horse Serum (Gibco) and penicillin-streptomycin (100units/ml and 100μg/ml, respectively, Gibco) at 37°C in a humidified atmosphere with 5% CO_2. DAMI cells differentiation trough megakaryocytic lineage was induced with 1 μM PMA and 10 ng/ml TPO during 7 days and cultured for another 3 days without stimulation to evaluate the characteristics of the particles produced [67]. Platelets-like particles were quantified in the supernatant by flow cytometry. Particles were identified by their positivity to CD41 and CD61 and by their size and granulation. Counts were obtained during 30sec in a continuous flow (60μl/min) (FACSCalibur, BD Bioscience).

Retroviral transduction and bone marrow transplantation

GFP retroviral vectors containing human cytoplasmic GSN cDNA or human cytoplasmic GSN G254C cDNA were used for bone marrow transplantation as previously described [73]. Briefly, the viral titer was estimated by transducing BaF3 cells with MSCV_GFP retroviral supernatant plus polybrene (10μg/ml) and analyzed for the percentage of green fluorescent positive cells by FACS two days after transduction. C57BL/6 donor mice (Charles River, Germany) were treated with 5-fluorouracil (150 mg/kg) 5 days prior to harvest. Bone marrow cells from donor mice were harvested by crashing femurs and tibias and cultured 1 day in transplant media (RPMI 10%FCS, 6ng/ml IL3, 10ng/ml IL6 and 10ng/ml stem cell factor). Cells were treated with spin infection with 1ml retroviral supernatant and centrifuge 1800 g 90 minutes at 30°C, 24 hours before and on day of transplantation. Recipient mice were lethally irradiated (2x6Gy) and transplanted with 1 x 10^6 whole BM cells. Mice were maintained on water plus addition of Nopil (Mepha, Aesch BL, CH).

Platelets Clearence

Mice were injected intravenously with 600 μg N-hydroxysuccinimido-biotin (NHS-biotin) (Sigma). At different time point 10μl of whole blood were collected and mixed in BSGC buffer (116 mM NaCl, 13.6 mM tri-sodium citrate, 8.6 mM Na_2HPO_4, 1.6 mM KH_2PO_4, 0.9 mM EDTA, 11.1 mM glucose). 2μl of blood were washed in balanced salt solution (BSS: 149 mM NaCl, 3.7 mM KCl, 2.5 mM $CaCl_2$, 1.2 mM $MgSO_4$, 7.4 mM HEPES, 1.2 mM KH_2PO_4, 0.8 mM K_2HPO_4, 3% bovine calf serum), pelleted at 1210 g for 10 min, and stained with PE-coniugated anti-CD61 (BD) and APC-conjugated streptavidin (BD) for 1 hr on ice. Samples were washed again in BSS and flow cytometry was performed on FACSCalibur (BD Bioscience) [15].

Transgenic mice

The human GSN-G254C transgene was constructed using the BAC CTD2218E19 (Invitrogen human BAC clones, Grand Island, NY USA), which is approximately 130-kb long and contains part of the GSN gene reaching from 70 kb upstream of exon 1 to 700 bp of the last intron. The construct was repaired with a fragment containing the last exon 18 derived from the BAC RP11269G20 and an ampicillin resistance cassette flanked by Frt sites (from the plasmid

phCre.myc.nuc.FRT.AMP.FRT, kindly provided by Dr Günther Schütz, DKFZ, Heidelberg, Germany), and 100 bp of sequence homologous to the BAC vector pBeloBAC11. This construct was inserted into the BAC by homologous recombination in the bacterial strain SW105 (kindly provided by National cancer Institute NCI Frederick, Maryland USA), and the ampicillin cassette was subsequently removed by induction of Flpe recombinase [82]. The insertion of the desired mutation was obtained by a new recombineering-based *E.coli* BAC modification system using *galK* selection [83]. For oocyte injection, BAC DNA was linearized with PmeI (NEB, Ipswich MA) and purified over a Sepharose CL4b column (GE Healthcare, Little Chalfont, United Kingdom). Genotyping of transgenic mice was performed using three sets of primers pair covering all the transgene. A set of primers pair for β-globin was used to check the DNA quality and exclude false negative results (Tab 2).

5'UTR_F	TAGCGTGCCATGTTTCTCTG
5'UTR_R	ACAGCAAATTGGGGTCACTC
Exon6_F	GGCGGGGCTTATAGGAAG
Exon6_R	CAAATTCACCTCAAATTGCC
Exon18_F	TGGGCAGTAGGGACAGTAGG
Exon18_R	AGGGACACAGGCAGAGAAGA
β-globin_F	GCTTTCCAGCAGGCACTAAC
β-globin_R	AGAATAGCCAGGGGAAGGAA

Tab 2: primers used for genotyping

Transgene copy number analysis was performed with a quantitative PCR using the following primers: TAGCGTGCCATGTTTCTCTG, ACAGCAAATTGGGGTCACTC. The number of transgene copy per genome was assessed in the "Absolute Quantification" setup with standard curves made from human genomic DNA isolated from a normal control.

Expression level of the human gelsolin gene in mice was performed with Real Time PCR using GSN primers specific for human sequence in order to distinguish the expression of human gelsolin in transgenic mouse samples. Murine GSN expression was detected with primers selective for the mouse transcript. Each GSN primers pair

was tested to have the same efficiency. Therefore, ratios of human and mouse GSN were assessed in the “Relative quantification” setup using the $2^{-\Delta CT}$ method. Primers pair are summarizes in Table 3.

All mice used in this study were kept under specific pathogen-free conditions and in accordance to Swiss federal regulations.

HumanGSN_F	GGTTCCAACAGCAATCGGTA
HumanGSN_R	GTTTGCCTGCTTGCCTTTC
MouseGSN_F	GTACCTTTCGATGCTGCTACG
MouseGSN_R	AGGTGCCACCCTTGTAGATG
MouseRPL19_F	ATCCGCAAGCCTGTGACTGT
MouseRPL19_R	TCGGGCCAGGGTGTTTTT

Tab 3: primers used for hGSN expression in transgenic mice

1 Ghilardi, N., Wiestner, A. & Skoda, R. C. Thrombopoietin production is inhibited by a translational mechanism. *Blood* **92**, 4023-4030 (1998).

2 Skoda, R. C. Thrombocytosis. *Hematology Am Soc Hematol Educ Program*, 159-167, doi:10.1182/asheducation-2009.1.159 (2009).

3 Hoffbrand A. V., M. P. A. H. a. P. J. E. *Essential Haematology*. (Blackwell Publishing, 2006).

4 Manz, M. G., Miyamoto, T., Akashi, K. & Weissman, I. L. Prospective isolation of human clonogenic common myeloid progenitors. *Proc Natl Acad Sci U S A* **99**, 11872-11877, doi:10.1073/pnas.172384399 (2002).

5 Pietras, E. M., Warr, M. R. & Passegue, E. Cell cycle regulation in hematopoietic stem cells. *J Cell Biol* **195**, 709-720, doi:10.1083/jcb.201102131 (2011).

6 Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: a human perspective. *Cell Stem Cell* **10**, 120-136, doi:10.1016/j.stem.2012.01.006 (2012).

7 Patel, S. R., Hartwig, J. H. & Italiano, J. E., Jr. The biogenesis of platelets from megakaryocyte proplatelets. *J Clin Invest* **115**, 3348-3354, doi:10.1172/JCI26891 (2005).

8 Long, M. W. Megakaryocyte differentiation events. *Semin Hematol* **35**, 192-199 (1998).

9 Chang, Y., Bluteau, D., Debili, N. & Vainchenker, W. From hematopoietic stem cells to platelets. *Journal of thrombosis and haemostasis : JTH* **5 Suppl 1**, 318-327, doi:10.1111/j.1538-7836.2007.02472.x (2007).

10 Bluteau, D. *et al.* Regulation of megakaryocyte maturation and platelet formation. *J Thromb Haemost* **7 Suppl 1**, 227-234 (2009).

11 Hartwig, J. H. & Italiano, J. E., Jr. Cytoskeletal mechanisms for platelet production. *Blood Cells Mol Dis* **36**, 99-103, doi:10.1016/j.bcmd.2005.12.007 (2006).

12 Tablin, F., Castro, M. & Leven, R. M. Blood platelet formation in vitro. The role of the cytoskeleton in megakaryocyte fragmentation. *J Cell Sci* **97 (Pt 1)**, 59-70 (1990).

13 Italiano, J. E., Jr., Lecine, P., Shivdasani, R. A. & Hartwig, J. H. Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes. *J Cell Biol* **147**, 1299-1312 (1999).

14 Zauli, G. *et al.* In vitro senescence and apoptotic cell death of human megakaryocytes. *Blood* **90**, 2234-2243 (1997).

15 Mason, K. D. *et al.* Programmed anuclear cell death delimits platelet life span. *Cell* **128**, 1173-1186, doi:10.1016/j.cell.2007.01.037 (2007).

16 White, M. J. & Kile, B. T. Apoptotic processes in megakaryocytes and platelets. *Semin Hematol* **47**, 227-234, doi:10.1053/j.seminhematol.2010.03.006 (2010).

17 Kaluzhny, Y. *et al.* BclxL overexpression in megakaryocytes leads to impaired platelet fragmentation. *Blood* **100**, 1670-1678, doi:10.1182/blood-2001-12-0263 (2002).

18 De Botton, S. *et al.* Platelet formation is the consequence of caspase activation within megakaryocytes. *Blood* **100**, 1310-1317, doi:10.1182/blood-2002-03-0686 (2002).

19 Kile, B. T. The role of the intrinsic apoptosis pathway in platelet life and death. *Journal of thrombosis and haemostasis : JTH* **7 Suppl 1**, 214-217, doi:10.1111/j.1538-7836.2009.03366.x (2009).

20 Battinelli, E., Willoughby, S. R., Foxall, T., Valeri, C. R. & Loscalzo, J. Induction of platelet formation from megakaryocytoid cells by nitric oxide. *Proc Natl Acad Sci U S A* **98**, 14458-14463, doi:10.1073/pnas.241427398 (2001).

21 Josefsson, E. C. *et al.* Megakaryocytes possess a functional intrinsic apoptosis pathway that must be restrained to survive and produce platelets. *J Exp Med* **208**, 2017-2031, doi:10.1084/jem.20110750 (2011).

22 Vanags, D. M., Orrenius, S. & Aguilar-Santelises, M. Alterations in Bcl-2/Bax protein levels in platelets form part of an ionomycin-induced process that resembles apoptosis. *Br J Haematol* **99**, 824-831 (1997).

23 Larson, M. K. & Watson, S. P. A product of their environment: do megakaryocytes rely on extracellular cues for proplatelet formation? *Platelets* **17**, 435-440, doi:10.1080/09537100600772637 (2006).

24 Avecilla, S. T. *et al.* Chemokine-mediated interaction of hematopoietic progenitors with the bone marrow vascular niche is required for thrombopoiesis. *Nat Med* **10**, 64-71, doi:10.1038/nm973 (2004).

25 Riviere, C. *et al.* Phenotypic and functional evidence for the expression of CXCR4 receptor during megakaryocytopoiesis. *Blood* **93**, 1511-1523 (1999).

26 Schulze, H. *et al.* Interactions between the megakaryocyte/platelet-specific beta1 tubulin and the secretory leukocyte protease inhibitor SLPI suggest a role for regulated proteolysis in platelet functions. *Blood* **104**, 3949-3957, doi:10.1182/blood-2004-03-1179 (2004).

27 Muntean, A. G. & Crispino, J. D. Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. *Blood* **106**, 1223-1231, doi:10.1182/blood-2005-02-0551 (2005).

28 Lecine, P. & Shivdasani, R. A. Cellular and molecular biology of megakaryocyte differentiation in the absence of lineage-restricted transcription factors. *Stem Cells* **16 Suppl 2**, 91-95, doi:10.1002/stem.5530160712 (1998).

29 Shiraga, M. *et al.* Primary megakaryocytes reveal a role for transcription factor NF-E2 in integrin alpha IIb beta 3 signaling. *J Cell Biol* **147**, 1419-1430 (1999).

30 Kaushansky, K. Lineage-specific hematopoietic growth factors. *N Engl J Med* **354**, 2034-2045, doi:10.1056/NEJMra052706 (2006).

31 Yamamoto, K., Nagata, K., Tsurukubo, Y., Morishita, K. & Hamaguchi, H. A novel translocation t(3;22)(q21;q11) involving 3q21 in myelodysplastic syndrome-derived overt leukemia with thrombocytosis. *Leuk Res* **24**, 453-457 (2000).

32 Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**, 8125-8148 (1987).

33 Kaushansky, K. The molecular mechanisms that control thrombopoiesis. *J Clin Invest* **115**, 3339-3347, doi:10.1172/JCI26674 (2005).

34 Kuter, D. J. New thrombopoietic growth factors. *Blood* **109**, 4607-4616, doi:10.1182/blood-2006-10-019315 (2007).

35 Kosugi, S. *et al.* Circulating thrombopoietin level in chronic immune thrombocytopenic purpura. *Br J Haematol* **93**, 704-706 (1996).

36 Deutsch, V. R. & Tomer, A. Megakaryocyte development and platelet production. *Br J Haematol* **134**, 453-466, doi:10.1111/j.1365-2141.2006.06215.x (2006).

37 Kaushansky, K. Molecular mechanisms of thrombopoietin signaling. *Journal of thrombosis and haemostasis : JTH* **7 Suppl 1**, 235-238, doi:10.1111/j.1538-7836.2009.03419.x (2009).

38 Solberg, L. A., Jr. Biologic aspects of thrombopoietins and the development of therapeutic agents. *Curr Hematol Rep* **4**, 423-428 (2005).

39 Kaushansky, K. Thrombopoietin and the hematopoietic stem cell. *Ann N Y Acad Sci* **1044**, 139-141, doi:10.1196/annals.1349.018 (2005).

40 Moliterno, A. R., Hankins, W. D. & Spivak, J. L. Impaired expression of the thrombopoietin receptor by platelets from patients with polycythemia vera. *N Engl J Med* **338**, 572-580, doi:10.1056/NEJM199802263380903 (1998).

41 Saur, S. J., Sangkhae, V., Geddis, A. E., Kaushansky, K. & Hitchcock, I. S. Ubiquitination and degradation of the thrombopoietin receptor c-Mpl. *Blood* **115**, 1254-1263, doi:10.1182/blood-2009-06-227033 (2010).

42 Dame, C. & Sutor, A. H. Primary and secondary thrombocytosis in childhood. *Br J Haematol* **129**, 165-177, doi:10.1111/j.1365-2141.2004.05329.x (2005).

43 Tefferi, A., Hanson, C. A. & Inwards, D. J. How to interpret and pursue an abnormal complete blood cell count in adults. *Mayo Clin Proc* **80**, 923-936, doi:10.4065/80.7.923 (2005).

44 Cazzola, M. Molecular basis of thrombocytosis. *Haematologica* **93**, 646-648, doi:10.3324/haematol.13194 (2008).

45 Teofili, L. & Larocca, L. M. Advances in understanding the pathogenesis of familial thrombocythaemia. *Br J Haematol* **152**, 701-712, doi:10.1111/j.1365-2141.2010.08500.x (2011).

46 Kralovics, R. *et al.* A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* **352**, 1779-1790, doi:10.1056/NEJMoa051113 (2005).

47 Rumi, E. Familial chronic myeloproliferative disorders: the state of the art. *Hematol Oncol* **26**, 131-138, doi:10.1002/hon.863 (2008).

48 Levine, R. L., Pardanani, A., Tefferi, A. & Gilliland, D. G. Role of JAK2 in the pathogenesis and therapy of myeloproliferative disorders. *Nat Rev Cancer* **7**, 673-683, doi:10.1038/nrc2210 (2007).

49 Smith, C. A. & Fan, G. The saga of JAK2 mutations and translocations in hematologic disorders: pathogenesis, diagnostic and therapeutic prospects, and revised World Health Organization diagnostic criteria for myeloproliferative neoplasms. *Hum Pathol* **39**, 795-810, doi:10.1016/j.humpath.2008.02.004 (2008).

50 Mead, A. J., Rugless, M. J., Jacobsen, S. E. & Schuh, A. Germline JAK2 mutation in a family with hereditary thrombocytosis. *N Engl J Med* **366**, 967-969, doi:10.1056/NEJMc1200349 (2012).

51 Percy, M. J. & Rumi, E. Genetic origins and clinical phenotype of familial and acquired erythrocytosis and thrombocytosis. *Am J Hematol* **84**, 46-54, doi:10.1002/ajh.21313 (2009).

52 Skoda, R. & Prchal, J. T. Lessons from familial myeloproliferative disorders. *Semin Hematol* **42**, 266-273, doi:10.1053/j.seminhematol.2005.08.002 (2005).

53 Ding, J. *et al.* Familial essential thrombocythemia associated with a dominant-positive activating mutation of the c-MPL gene, which encodes for the receptor for thrombopoietin. *Blood* **103**, 4198-4200, doi:10.1182/blood-2003-10-3471 (2004).

54 Moliterno, A. R. *et al.* Mpl Baltimore: a thrombopoietin receptor polymorphism associated with thrombocytosis. *Proc Natl Acad Sci U S A* **101**, 11444-11447, doi:10.1073/pnas.0404241101 (2004).

55 El-Harith el, H. A. *et al.* Familial thrombocytosis caused by the novel germ-line mutation p.Pro106Leu in the MPL gene. *Br J Haematol* **144**, 185-194, doi:10.1111/j.1365-2141.2008.07430.x (2009).

56 T. Strachan, A. R. *Human Molecular Genetics 4ed*. (Garland Science Publishing, 2010).

57 Pulst, S. M. Genetic linkage analysis. *Arch Neurol* **56**, 667-672 (1999).

58 Ott, J., Kamatani, Y. & Lathrop, M. Family-based designs for genome-wide association studies. *Nat Rev Genet* **12**, 465-474, doi:10.1038/nrg2989 (2011).

59 Cardon, L. R. & Bell, J. I. Association study designs for complex diseases. *Nat Rev Genet* **2**, 91-99, doi:10.1038/35052543 (2001).

60 Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812-3814 (2003).

61 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).

62 Maathuis, M. H., Colombo, D., Kalisch, M. & Buhlmann, P. Predicting causal effects in large-scale systems from observational data. *Nat Methods* **7**, 247-248, doi:10.1038/nmeth0410-247 (2010).

63 Eyster, M. E. *et al.* Familial essential thrombocythemia. *The American journal of medicine* **80**, 497-502 (1986).

64 Read, T. S. a. A. P. *Human Molecular Genetics. 4th Edition*. (Wiley-Liss 2010).

65 Kun, L. *PhD Thesis: Genetic Studies of Familial Myeloproliferative Disorders*, University of Basel, (2007).

66 Li, G. H., Arora, P. D., Chen, Y., McCulloch, C. A. & Liu, P. Multifunctional roles of gelsolin in health and diseases. *Med Res Rev*, doi:10.1002/med.20231 (2010).

67 Lev, P. R. *et al.* Production of functional platelet-like particles by the megakaryoblastic DAMI cell line provides a model for platelet biogenesis. *Platelets* **22**, 26-36, doi:10.3109/09537104.2010.515271 (2011).

68 de la Chapelle, A. *et al.* Gelsolin-derived familial amyloidosis caused by asparagine or tyrosine substitution for aspartic acid at residue 187. *Nat Genet* **2**, 157-160, doi:10.1038/ng1092-157 (1992).

69 Page, L. J. *et al.* Secretion of amyloidogenic gelsolin progressively compromises protein homeostasis leading to the intracellular aggregation of proteins. *Proc Natl Acad Sci U S A* **106**, 11125-11130, doi:10.1073/pnas.0811753106 (2009).

70 Brown, S. B., Clarke, M. C., Magowan, L., Sanderson, H. & Savill, J. Constitutive death of platelets leading to scavenger receptor-mediated phagocytosis. A caspase-independent cell clearance program. *J Biol Chem* **275**, 5987-5996 (2000).

71 Pereira, J. *et al.* Platelet aging in vivo is associated with activation of apoptotic pathways: studies in a model of suppressed thrombopoiesis in dogs. *Thromb Haemost* **87**, 905-909 (2002).

72 Zunino, R. *et al.* Expression of scinderin in megakaryoblastic leukemia cells induces differentiation, maturation, and apoptosis with release of plateletlike particles and inhibits proliferation and tumorigenesis. *Blood* **98**, 2210-2219 (2001).

73 Wernig, G. *et al.* Expression of Jak2V617F causes a polycythemia vera-like disease with associated myelofibrosis in a murine bone marrow transplant model. *Blood* **107**, 4274-4281, doi:10.1182/blood-2005-12-4824 (2006).

74 Doetschman, T. Influence of genetic background on genetically engineered mouse phenotypes. *Methods Mol Biol* **530**, 423-433, doi:10.1007/978-1-59745-471-1_23 (2009).

75 Witke, W. *et al.* Hemostatic, inflammatory, and fibroblast responses are blunted in mice lacking gelsolin. *Cell* **81**, 41-51 (1995).
76 Cantu, C. *et al.* Defective erythroid maturation in gelsolin mutant mice. *Haematologica* **97**, 980-988, doi:10.3324/haematol.2011.052522 (2012).
77 Bender, M. *et al.* ADF/n-cofilin-dependent actin turnover determines platelet formation and sizing. *Blood* **116**, 1767-1775, doi:10.1182/blood-2010-03-274340 (2010).
78 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63, doi:10.1038/nrg2484 (2009).
79 Kovacsovics, T. J. *et al.* Phosphoinositide 3-kinase inhibition spares actin assembly in activating platelets but reverses platelet aggregation. *J Biol Chem* **270**, 11358-11366 (1995).
80 Lin, M. *et al.* dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**, 1233-1240, doi:10.1093/bioinformatics/bth069 (2004).
81 Kralovics, R. *et al.* Comparison of molecular markers in a cohort of patients with chronic myeloproliferative disorders. *Blood* **102**, 1869-1871, doi:10.1182/blood-2003-03-0744 (2003).
82 Lee, E. C. *et al.* A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65, doi:10.1006/geno.2000.6451 (2001).
83 Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res* **33**, e36, doi:10.1093/nar/gni035 (2005).

Acknowledgements

First, I wish to say a heartfelt thank you to Professor Radek Skoda because he gave me the opportunity to work in his lab. He trusted me and gave me the freedom to develop my project according to my way, but helping me constantly with nice suggestions and encouragement throughout the course of my PhD. I have benefited from his guidance, great knowledge and kindness. I thank Professor Christoph Handschin for being my Co-referee showing great interest in my work. I thank Professor Ed Palmer for joining my Thesis committee.

I would like to express my special gratitude to all the members of my lab. When I came to Basel I found a nice environment and a lovely place where to work. I can say I found a second family here. I would like to thank all the previous and current members of our great team because everybody had something to teach me about my professional and private life. Without all of them my work would not be possible: Adrian, Alexey, Axel, Franz, Hui, Imane, Jean, Liu Kun, Lucia, Pan, Pontus, Renate, Ronny, Sabina, Sai Lii, Silvia, Sophie and Takafumi. In particular, I thank Liu Kun for starting the project and providing the majority of the linkage and genetic data; Pontus for helping me with the linkage analysis, for his precious work with the NGS and his kind suggestions about my project; Hui and Takafumy for their kind help with the mouse work. I thank Axel and Jean for reading and correcting my Thesis.
In addition, I would like to thank Professor Torsten Schwede and Morena Spreafico for their help with the structural analysis.

A dear *"grazie"* is for my parents and my sister Ilaria. They have followed me in all these years, giving me all the love, support and encouragement. They were always next to me whenever I needed, in the good and difficult moments.

I wish to give an affectionate thanks to Jonas. For all the nice moments we spent together and for all the time he was next to me giving me joy and support.

A great and precious thank is for my friends. It is impossible to mention everybody one by one, but inside them they know how much I love them and I appreciate their company. I thank them for showing me their friendship and affection. We had really great time inside and outside Basel and I will never forget it. Without them this achievement would not be so special.

Finally, I thank the Swiss National Foundation for supporting my study.

Grazie di cuore
Annalisa

Annalisa Pianta

Socinstrasse 71, 4051 Basel, Switzerland
Mobile: 0041 78 8629884
E-mail: annalisa.pianta@unibas.ch

Personal Information

Date of birth:	07. June 1984
Place of birth:	Magenta (Milan), Italy
Citizenships:	Italian

Professional Aim

I am keen to begin a career in biomedical science. I am a PhD student, close to my final exam. In achieving this, I have shown myself to be self-motivated, committed and determined in achieving my goals. During this experience I acquired planning skills, a good sense of responsibility and capacity to work hard under pressure. These aspects combined to my passion for research will make me suitable for a Post-Doc experience.

Professional Experience

November 2008 – present	**PhD student** in the Lab of Prof. Radek Skoda, Experimental Hematology, Department of Biomedicine, University Hospital Basel, Basel, Switzerland Field of interest: research on myeloproliferative disease.
March 2007 – October 2008	**Master student** in the Lab of Prof. Paolo Corradini, Hematology and Allogeneic Stem Cell Transplantation Unit, International cancer institute, Milano, Italy Field of interest: diagnosis and research on hematologic malignancies. Deep interest in the effect of graft-versus-host disease and graft-versus-tumor in patients treated with allogeneic stem cells transplant.
February 2006 – July 2006	**Bachelor student** in the Lab of Prof. Paolo Corradini, Hematology and Allogeneic Stem Cell Transplantation Unit, International cancer institute, Milano, Italy Field of interest: immune reconstitution in patients after allogeneic transplant.

Technical Skills

- Molecular and Cellular biology: isolation of PBMCs from blood by Ficoll-Hypaque density gradient, selection of cell population by magnetic cell sorter autoMACS, culture of cell line, FACS analysis. DNA and RNA isolation, cDNA synthesis, PCR, Real-Time PCR, electrophoresis, cloning and sequencing, microsatellites and SNP screening, chimerism analysis. Western Blot, fluorescent microscopy.
- Mice *in vivo* experiments: BM transplant in irradiated mice, tail bleeding of mice, measuring of platelets half-life in vivo, creation and characterization of transgenic mouse model.
- Computer skills: Microsoft Office Package, Mac Office Package, GraphPad Prism, FlowJo, Illustrator, EndNote.

Education

October 2006 - October 2008	**Master in Biotechnology and Molecular Medicine (Oncology)** University of Milan Final Exam grade: 110/110 Subjects: Biology, Genetic, Pathology, Immunology, Lab. Technology, Oncology, Diagnostic, Innovative Therapy
September 2003 - July 2006	**Bachelor in Biotechnology for Medical Research** University of Milan Final Exam grade: 110/110 cum laude Subjects: Biology, Genetics, Maths, Physics, Chemistry, Physiology, Pathology, Medical applications of Biotechnology.
September 1998 - June 2003	**High School diploma** Liceo classico statale S. Quasimodo Magenta (Milano) Final Exam grade: 100/100 Subjects: Italian, English, German, French, Maths, Science, Latin, History, Philosophy .

Training courses

May 2012	**WIN: women into industry program** Joint project of the University of Basel and Novartis. Program to support and encourage young and highly qualified female academics to proactively plan and work on their career.
February 2012 – September 2012	**Global Perspective Program (GPP): preparing future academic leaders** Exchange program between students from Switzerland and the US. The program is made to favour the development of a deep knowledge of different countries and their higher education institutions.
30 September 2010 –2 October 2012	**International Conference MYELOPROLIFERATIVE NEOPLASMS** European School of Haematology
June 2010	**Soft Skill: making an effective oral presentation, conference poster and writing an effective journal paper** University of Basel
September 2009 – December 2009	**Tutorial course for first year biology students** Biozentrum, University of Basel, Switzerland
August 2009	**Introductory Course in Laboratory Animal Science, LTK1 Modul 1** Institute of Laboratory Animal Science, University of Zurich Irchel, Switzerland
28-29 May 2007	**National course of Clinical Hematology** Italian Society of Hematology, Milano, Italy
18 May 2007	**New drugs in Oncology: theoretical basis and clinical development** The European School of Oncology, Milano, Italy

Social Skills	Working with people with different personalities in a stressed and competitive work environment where it is essential to have a balanced personality and good personal relationships with every colleague. I like classical music: I love going to

concerts and ballet. I myself attend a ballet class. I love also arts, going to exhibition and reading. As an Italian, I love cooking.

Languages

Italian (native language),

English (very good),

French (quite good),

German (basic knowledge)

References

Prof. Radek Skoda, radek.skoda@unibas.ch

Prof. Jürg Schwaller, j.schwaller@unibas.ch

Prof. Christoph Handschin, christoph.handschin@unibas.ch

Prof. Paolo Corradini, paolo.corradini@istitutotumori.mi.it