Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
BIOCHIMICA, BIOLOGIA MOLECOLARE E BIOTECNOLOGIE

CICLO XXVII

COORDINATORE Prof. Francesco Bernardi

Development of cellular and animal models of coagulation factors deficiencies for the assessment of innovative therapeutic approaches acting on transcriptional and post-transcriptional regulation

Settore Scientifico Disciplinare BIO/11

Dottorando
Dott. Barbon Elena

Tutore
Prof. Pinotti Mirko

Co-tutore
Prof. Mingozzi Federico

Anni 2012-2014

Contents

Ringraziamenti

Questo lavoro, e soprattutto l'esperienza e l'arricchimento scientifico e personale che ne sono scaturiti, non sarebbe stato possibile senza l'aiuto delle persone con cui ho lavorato in questi tre anni. Ringrazio Mirko e il Prof.Bernardi per avermi introdotto e guidato nel fantastico mondo della biologia molecolare, sostenendomi in tutte le fasi del mio percorso di dottorato e trasmettendomi i loro insegnamenti. Grazie a Matteo, un buon capo ma soprattutto un buon amico, per la sua grande disponibilità e per aver condiviso con me la sua esperienza, le idee e gli spunti per lavorare assieme sinergicamente. Grazie anche a Silvia e Mattia, perché parte di questo lavoro è merito del loro impegno, che ho apprezzato soprattutto quando non ero fisicamente presente a seguirli. Un grazie enorme va a Federico, che è diventato per me un ottimo mentore oltre che mente scientifica da ammirare, per la sua grande disponibilità verso di me e il mio progetto fin dai primi momenti, e per avermi insegnato che la condivisione delle conoscenze e lo scambio scientifico sono l'unica, onesta e produttiva strada da seguire. Merci beaucoup à Giuseppe, Christian, Fanny, Séverine, Amine et Francesco, parce-que chez eux j'ai trouvé une équipe fantastique du travail, unie et compacte. Grazie al mitico gruppo Cartel, per aver contribuito a rendere magnifica la mia esperienza a Parigi. Grazie a Silvia, perché avere la presenza di una cara amica come compagna di laboratorio ha reso questi anni di dottorato più pazzi e allegri. Grazie come sempre ai miei genitori ed al loro incondizionato sostegno in tutte le scelte.

Grazie ad Alessandro, che ha reso perfetto il layout di questa tesi... ed anche tutto il resto.

Introduction

In the last decades, enormous efforts have been pushed toward the development of therapeutic approaches for human genetic diseases, and the research all over the world has obtained remarkable achievements. Advances in biotechnology have brought gene therapy to the forefront of medical research. Gene replacement, in which a normal copy of the defective gene is efficiently delivered and expressed into target cells, represents one of the most advanced strategies that produced encouraging results in patients with diseases caused by single gene recessive disorders (like haemophilia, muscular dystrophy, sickle cell anemia, cystic fibrosis etc.), some viral infections and inherited genetic diseases such as cancer and ADA deficiency [13]. Notwithstanding, the intense research also led to potential therapeutic strategies based on the correction of the specific disease-causing defects, which might circumvent some limitations (i.e. gene size, gene regulation) of gene replacement therapy. These approaches are of great interest for patients with coagulation deficiencies, since they would benefit from even small increase in functional protein levels. The low therapeutic threshold makes the rare bleeding disorders (RBDs) an ideal target for investigating therapeutic strategies based on counteracting the pathogenic molecular mechanisms. This work propose the development of *in–vitro* and *in–vivo* models of RBDs in order to explore corrective molecular approaches acting on the specific disease-causing defects, both at transcriptional and post-transcriptional level.

The chapter 1 describes the usage of engineered transcription factors (eTFs) as potential therapeutic strategy for FVII deficiency caused by promoter mutations. Transcription impairment by mutations in the promoter regions represents indeed a small, but considerable cause of severe coagulation factor defects and of all inherited human diseases. These mutations often affect only the untranslated control regions impairing to different extent the expression of completely unaffected genes. Different from gene therapy approaches inserting exogenous sequences that drive the expression of the missing factor, the induced restart of the transcription would maintain the natural chromosomal structure and context restoring gene expression. Engineered TFs have been shown to have wide-range potential in modulating desired gene expression through targeting their promoters [109]. A recent breakthrough with transcription activator-like effector proteins (TALEs) makes it possible to establish universal types of engineered TFs that can potentially target any selected gene. Hence, fusion of a TALE-derived binding domain to a transcriptional activator, such as VP64 [21], could generate TALE-transcription factors (TALE-TFs) that can target selected promoter regions and

modulate expression of corresponding genes [55]. Intensive studies on stem-cell maintenance and differentiation have been recently carried out with TALE-TFs [55, 27, 46, 56, 72], but to our knowledge only one example of these eTFs has been used for the recovery of transcription impaired by human disease-causing mutations [86]. As a model to exploit TALE-TF transcription restoration in coagulation factor deficiencies, we chose two severe promoter mutations impairing FVII transcription, the -94 C>G and -61 T>G transversions, falling in the binding site for the hepato-specific HNF-4 and the ubiquitous Sp1 transcription factors, respectively [18, 29]. The patients homozygous for these mutations experience life-threatening hemorrhagic symptoms and require replacement therapy. Through the expression of gene reporter plasmids we first created a cellular model for the two F7 promoter variants. Then, we assembled 4 TALE-TFs (TF1-4) designed to target different regions on the F7 proximal promoter in order to test their efficacy in stimulating transcriptional activity on the target gene. The treatment with the different TALE-TFs demonstrated that TF4, targeting a sequence between the mutations, induced a robust increase of gene transcription in the presence of the defective promoter. Interestingly, TF4 appreciably increased the endogenous F7 transcription and mRNA levels in HepG2 cells and induced F7 expression in Hek293 cells that do not virtually express FVII.

The chapter 2 describes the exploitation of the Sleeping Beauty Transposon System (SBTS) to develop cellular and mouse model of haemophilia B (HB) caused by splicing mutations, in order to assess the efficacy of an RNA-based therapeutic approach. HB is a prototypical example of disease with a heterogeneous mutational pattern (www.factorix.org) and a relatively high frequency of splicing mutations (>15%) in severe disease forms. Intervention at the pre-mRNA splicing level is emerging as a promising therapeutic strategy for genetic disorders. Increasing attention has been given to the U1 small nuclear RNA (U1snRNA) that, in the earliest splicing step, mediates the recognition of the donor splice site (5'ss) by the ribonucleoprotein U1snRNP. Studies in various cellular models of human disease indicated the potential therapeutic effect of engineered U1snRNAs to rescue aberrant splicing caused by mutations at 5'ss, a relatively frequent cause of severe forms. In the last years, modified U1snRNAs have been exploited to correct splicing mutations causing severe coagulation factor VII deficiency and HB [37, 74, 75]. The rescue has been proven in cellular models by using F7/F9 minigenes. However, the F9 mutational pattern is extremely heterogeneous, with several different splicing mutations. Therefore, the evaluation of the U1snRNA-mediated correction strategy *in–vivo* implies the creation of mouse models, which are not yet available. Here we used the SBTS to develop cellular/mouse models of HB caused by the FIXex5-2C splicing variant, and subsequently assess the U1-mediated correction in chromatin context. Indeed, in contrast to other approaches the transposon enables the transgene to be integrated into the genome, which represents the preferred physiologic situation for subsequent splicing studies. Initially we have performed experiments to test the hyperactive transposase SB100X-mediated integration of the splicing-competent human FIX transgene into the genome using human embryonic kidney cells. We have generated Hek293

stable clones expressing the normal or mutated human splicing-competent FIX cassettes integrated into the genome as a result of the transposase activity. These preliminary studies provided us with optimized experimental protocol to create in a relatively short time cellular models of human disease caused by splicing mutations. Moreover, it permitted the assessment of the modified U1 snRNA-mediated rescue *in–vitro* in a genomic expression context instead of a transient episomal system. This also provided with the rationale for the creation of mouse models through hydrodynamic injection of the transposon plasmids and of SB100X transposase in C57BL/6 *wt* mice [62] in order to assess the efficacy of the U1 snRNAs-mediated rescue *in–vivo*.

Chapter 1

TALE TFs

1.1 Brief overview of coagulation process

Haemostasis is a physiological process that depicts a delicate balance between pro-coagulant and anti-coagulant activities. From many years researchers have been interested in deepening the mechanisms underlying the complex series of biochemical reactions occurring to maintain these equilibrium in-vivo. According to the classical coagulation waterfall/cascade model, described in 1964 by two independent groups [96, 112], several proteins with serine-protease activity are involved in the formation of blood clot whenever an injury in a vessel occurs. In physiological conditions these coagulation factors, circulating normally as zymogens, are converted by cleavage into their active form and interact both together and with the pro-coagulant surfaces in order to produce the fibrin formation [95]. In this traditional view a contact pathway (intrinsic) as well as a tissue-factor (TF) depending-pathway (extrinsic) have been described. The intrinsic pathway, so named because triggered by components already present in blood, is composed of FXII and pre-kallikrein (PK), with FXII being activated either by PK or plasmin to the active enzyme (FXIIa) when in contact with a negatively charged or artificial surfaces. In the intrinsic pathway the activation of a certain amount of FXII leads to the subsequent generation of FXIa from FXI, which then activates FIX. FIXa, together with phospholipids, calcium and FVIIIa, activates FX, which in turn converts prothrombin (FII) to thrombin (FIIa). A peculiar feature of FXII and PK zymogens is that they are calcium-independent, thereby able to "auto-activate" in a continuous manner even in the presence of anticoagulant molecules. In spite of this, it has been reported that individuals with deficiencies of FXII or PK do not experience any bleeding tendency in vivo, thereby suggesting that the contact pathway may not have such a fundamental role in blood coagulation, but rather in the thrombus formation [10, 97]. The first waterfall/cascade model has been implemented from the 1977, with the work of Osterud and Rapaport showing that FVIIa-TF can activate FIX and FX, so focusing attention on the role of the extrinsic pathway in interacting with the intrinsic one during clotting [3]. And in 1992 a group proposed the actual model of coagulation, called "cell-based" [105], which is composed of an initiation phase and a propagation phase (fig.1.1). The

clotting is triggered in the initiation phase when tissue factor (TF), also called thromboplastin or FIII, comes in contact with blood. TF is an integral membrane protein expressed on the surface of several cell types, like adventitial cell, smooth muscle cells and keratinocytes, that normally are not in contact with blood flow. When a lesion in the endothelial barrier occurs these cells become exposed to blood and consequently TF is able to bind the plasma serine protease FVIIa, forming a complex that is able to generate a modest amount of thrombin. In the following propagation phase, acting on the activated platelet surfaces, the tenase complex can be properly assembled together with the prothrombinase complex and calcium for massive production of thrombin. This "thrombin burst" arises at the end in the conversion of insoluble fibrinogen into soluble fibrin monomers forming the stable clot.

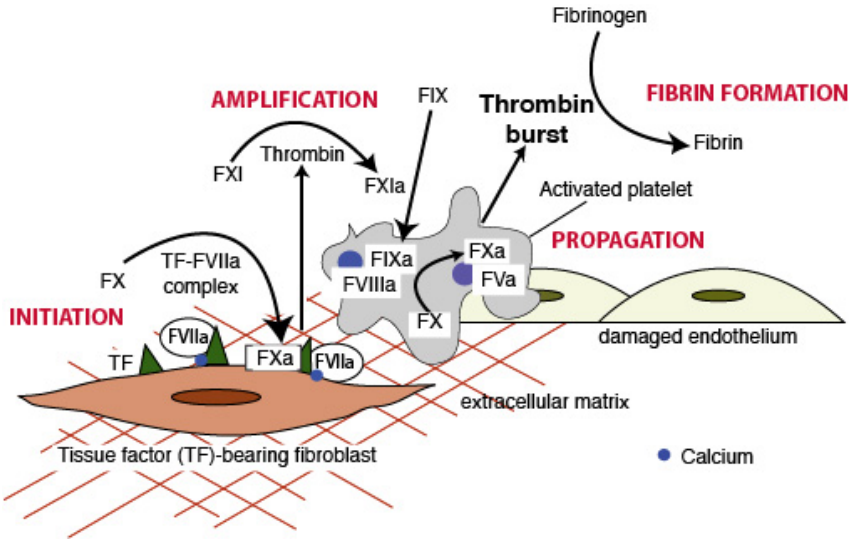

Figure 1.1: *Cell-based model of coagulation*

1.1.1 Initiation-FVII/FVIIa

Factor VII is a vitamin K-dependent serine protease produced in liver and circulating at a concentration of 500 ng/ml [33, 90]. It displays the shortest half-life as zymogen (approximately 5 hours) and is the only coagulation factor that is present in both the active and inactive forms. In humans, the active form (FVIIa) represents about 1% of total circulating FVII. FVII itself has a negligible activity and participates in the initiation phase principally after its activation to FVIIa. A number of coagulation enzymes including FIXa, FXa, FXIIa, thrombin, plasmin or FVII-activating proteases can activate FVII. However, FXa seems to be the most potent activator of FVII. Also FIX plays an important role in this activation because humans with haemophilia B display very low levels of FVIIa. Moreover, the FVII-TF complex can be auto-activated by FVIIa-TF complex. FVIIa has a long half-life in plasma because unless it is bound to TF it can not be inactivated by any protease inhibitor. The FVIIa-TF complex then activates FIX to

FIXa and FX to FXa. It is worth noting that this initiation phase only results in a trace amount of thrombin after FXa activation, because FXa and thrombin are rapidly neutralized by antithrombin (AT) and the complex FVIIa-TF-FXa by the tissue factor pathway inhibitor (TFPI). The TFPI/AT provides a good regulatory system in order to prevent a massive thrombin generation for a false alarm. The pro-coagulant signal is provided only when TF is exposed at high enough levels to overcome inhibition by TFPI an AT. If this is the case, thrombin is able to enhance its own generation through the activation of FXI to FXIa and FV to FVa on the platelets surface, and also through the activation of FVIII to FVIIIa by cleaving von Willebrand factor (vWF). In this context recruited activated platelets provide the pro-coagulant phospholipids membrane needed for the subsequent enzymatic reactions, and the further generation of FXa is dependent on the formation of the tenase complex.

1.1.2 Propagation–The tenase complex

Once FX, FIX and the cofactors FV and FVIII are activated after the generation of thrombin in the initiation phase the assembly of tenase complex can occurs. FIXa, in association with its cofactor FVIIIa, binds FX and calcium on negatively charged surfaces, mostly on platelets membrane, in order to form the tenase complex, leading to the generation of a large amount of FXa. Most of the FXa is physiologically produced in vivo by the tenase complex, which is considered to be around 50-fold more efficient in this process compared to the FVIIa-TF complex. Therefore, in the initiation phase the amount of FXa is insufficient to sustain hemostasis whereas during propagation a massive production of FXa is achieved through the formation of the tenase complex. FXa participates in the assembly of the prothrombinase complex together with FVa and calcium, which results in an explosive generation of thrombin and the subsequent stable clot formation. In physiological conditions the process is tightly controlled, in order to maintain the coagulative events just at the site of injury. As a matter of fact even if thrombin and FXa can diffuse from the site of damage to the adjacent healthy endothelium they are inhibited by the lack of pro-coagulant membranes and the presence of anticoagulants molecules like AT, TFPI and thrombomodulin, able to reduce their activity. It is worth noting that in the propagation phase both FVIII and FIX are required. Indeed, in the absence of FVIII and FIX the initiation occurs normally, but the propagation steps are severely diminished, resulting in a deficient fibrin clot formation and the incapacity of the hemostatic system to respond in repairing the damaged tissues.

1.2 Focus on FVII: gene and protein features

1.2.1 Factor FVII protein

Human factor FVII is a zymogen for a vitamin K-dependent serine protease synthesized in liver and circulating in plasma at a concentration of approximately

0.5μg/ml (10nmol/L). It is secreted into blood as a single chain glycoprotein of 48KDa. All vitamin K-dependent coagulation zymogens share a similar protein domain structure consisting an N-terminal gamma-carboxyglutamic acid-rich domain (Gla domain, residues 1-38), a short hydrophobic segment (residues 39-45) that often is considered as part of the Gla domain, two epidermal growth factor (EGF)-like domains (residues 47-84 and 85-131), and the C-terminal serine protease domain (residues 153-406). The Gla domain is responsible for the interaction of the protein with lipid membranes, while the EGF-like domain has a calcium ion binding site that to some degree mediates interaction with the TF exposed at the site of vessel injury (Tab 1.1). The major proportion of FVII

Summary of FVII features	
Synthesis and localization	Synthesized in the liver and circulating in the plasma as a zymogen
Half life	3-6 hours
Molecular weight	50,000 Da
Structure	N-terminal (light chain) Gla domain, 2 EGF-like domain, C-terminal (heavy chain) catalytic domain
Co-factor	Tissue factor
Substrate	Factor VIIa/tissue factor complex activates factors X and IX

Table 1.1: *Summary of factor VII features*

circulates in plasma in the inactive form. Activation of FVII occurs by the cleavage of a single peptide bond between Arg152 and Ile153, located in the region connecting the second EGF-like domain and the protease domain [111]. This results in the formation of a two-chain FVIIa molecule consisting of a light chain of 152 amino acids (containing the membrane-binding Gla domain) and a heavy chain of 254 amino acids (containing the catalytic domain) held together by a single disulfide bond (between Cys-135 and Cys-262). Conversion of factor VII to factor VIIa is catalyzed by a number of proteases, including thrombin, factor IXa, factor Xa, factor XIa, and factor XIIa [111]. Comparison of these proteins has shown that factor Xa, in association with phospholipids, has the highest potential to activate factor VII. Rapid activation also occurs when factor VII is combined with its cofactor, which is the tissue factor in the presence of calcium (autocatalysis). This reaction may be initiated by a small amount of preexisting factor VIIa. FVIIa is a trypsin-like enzyme and catalyses the hydrolysis of peptide bonds within a polypeptide chain to produce two new smaller peptides. In 1964, FVIIa proteolytic activity was observed first against FX, but the methods used did not reveal whether TF first reacted with FVII to form an intermediate that then activated FX or whether FVII directly activated FX. A few years later, it was demonstrated that TF interacted with FVII and the formed inter-

mediate was catalytically active [122]. The reaction product of TF and FVII is a potent activator of FX and also of FIX [3]. The interaction of FVIIa with TF is Ca^{2+}–dependent. Ca^{2+} saturation of the Gla domain is likely responsible for this increase in affinity, since deletion of Gla in FVIIa results in a loss of affinity for TF. Ca^{2+} may stabilize energetically important hydrophobic contacts of Gla with TF [7]. In addition, Ca^{2+} increases FVIIa affinity for FX by conformational changes in FVIIa and FX that are essential for the interaction of these proteins with phospholipids [120].

1.3 Factor FVII gene

The human factor VII gene (F7) is a single copy gene located on chromosome 13 (13q34) which consists of eight exons spread over 12 kb of genomic DNA and produces a 2.4-kb mRNA encoding a secreted protein of 406 amino acids. The entire nucleotide sequence was reported for the first time in 1987 [70]. The introns ranged in size from 68 nucleotides (intron C) to nearly 2.6 kb (intron A). The exons also varied considerably in size, ranging from 25 nucleotides (exon 3) to 1.6 kb (exon 8). In 1986 different cDNAs coding for factor VII were isolated from a cDNA library prepared from a HepG2 cell line [45], by which it was possible to establish that 1b is an optional exon not always included in the mature mRNA. Its inclusion determines the length of the pre-pro leader sequence (from 38

Exons	Coding region	Dimension (bp)
1a	Pre-pro leader sequence	100
1b	Pre-pro leader sequence	66
2	Pre-pro leader, Gla domain	161
3	Gla domain	25
4	EGF-like domain	139
5	EGF-like domain	141
6	Activation site	110
7	Catalytic domain	124
8	Catalytic domain	1622

Table 1.2: *Factor VII gene exons organization*

aminoacids when the mRNA lack 1b to 60 aminoacids when 1b is incorporated). In both cases the product is an in frame-mRNA that results in a functional transcript that codes for a biologically active FVII. In normal liver the most represented form is mRNA lacking exon 1b. As for other coagulation factor genes, F7 exons encode for discrete domains of the protein (Tab. 1.2): the pre-pro leader sequence (exon 1a, 1b and part of exon 2), the γ–carboxilase region (exons 2 and 3), the two EGF-like domains (exons 4 and 5), the activation domain (exon 6)and finally the catalytic domain (exons 7 and 8). The conservation of domains, exons

and introns positions among the members of the vitamin K-dependent protein family supports the theory of their evolution as modular protein by exon shuffling [100].

1.4 FVII deficiency

FVII, or proconvertin, deficiency was first recognized in 1951 [16]. Considered the most common of rare bleeding disorders its incidence is estimated at 1 per 300,000–500,000 [59]. It is inherited in an autosomal recessive fashion and affects males and females equally. Factor VII deficiency is very rare, but like all autosomal recessive disorders, it is found more frequently in areas of the world where marriage between close relatives is common. Studies conducted on F7 *knock-out* mice suggest that a complete lack of FVII is incompatible with life [77], so it is thought that FVII deficiency can not be associated with complete absence of functional FVII. Triplett et al. [87] have classified FVII deficiency in CRM- (activity and antigen proportionally reduced), CRM+ (reduced activity, antigen normal) and CRMred (antigen is reduced but not as much as activity). This haemorragic disorder displays a variable clinical heterogeneity [65], which ranges from lethal to mild or even asymptomatic forms (in general when FVII:C >10%). The most frequent symptoms, like epistaxis and early bruising, indicate that the disease is mild in the majority of cases. However, severe to very severe cases are not infrequent, and are characterized by hemartrosis, muscle hematomas, or even central nervous system (CNS) and gastrointestinal (GI) bleeding. Nose bleeding is by far the most frequent symptom and is not gender-related. Other very common symptoms are post–operative, skin and gum bleeds. As for gender, women are more prevalent among bleeders: this is mainly attributable to menorrhagia, which occurs in about two-thirds of women of fertile age but gum bleeding and easy bruising are also more frequent in females. Severe and life-threatening haemorrhages are rare in general (about 5% of the bleeds) and occur most frequently during the first 6 months of life. In newborns (< 1 month) presenting bleeding manifestations were, as ranked for frequency, central nervous system (CNS), gastro-intestinal(GI), cephalohaematoma and umbilical bleeding. Several therapeutic options are available for patients experience FVII deficiency, which may result in a very effective correction of the disease [99]. Fresh-frozen plasma is still used in developing countries, the most important drawback being the blood volume overload and a relatively high risk of transmitting blood-borne viruses. Plasma-derived FVII concentrates are essentially prothrombin complex concentrates with a higher content of FVII. They are used for prophylactic treatment, as well as for controlling serious bleeding episodes, and bleeding during surgery. However, plasma-derived concentrates carry the risk of potential transmission of blood-borne pathogens. Recombinant FVIIa is to be considered nowadays the optimal protein replacement therapy. It is free of human plasma and albumin, so there is no risk of human viral transmission, but it is very expensive and not available for all patients, particularly in developing countries.

A considerable number of mutations have been reported to date in FVII gene (F7 mutation database: `http://www.hgmd.cf.ac.uk/ac/index.php`). Missense mutations represent the most frequent portion (fig.1.2) and occur in the 68% of cases, followed by splicing-site (13%), promoter (8%) and nonsense (6%) mutations, small insertions and deletions (6%) [99, 60, 61]. A number of these mutations have been identified as the cause of FVII deficiency but only a limited number of them have been well characterized.

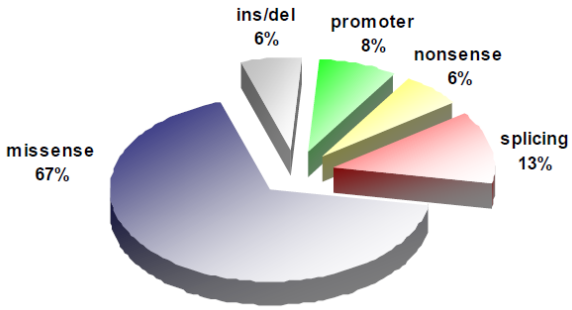

Figure 1.2: *Pie chart reporting the percentages of each specific mutation type found in F7 gene.*

1.4.1 Background and rationale

In the last years new technologies have been developed in order to control the expression of human genes by engineering synthetic transcription factors that can theoretically target any DNA sequence in a specific manner. The capability of regulating gene expression in the native chromosomal context may represent a way to address several genetic diseases. This approach also circumvent some of the traditional challenges of gene replacement, like gene delivery, specific activation and immune response related-aspects. Over the past 20 years several platforms have been developed in order to target a specific DNA sequence and the most widely used are represented by zinc finger-based transcription factor/nucleases, the transcription activator-like effectors and the CRISPR/Cas9 systems (figure 1.3). All these systems are based on the exploitation of particular DNA binding domains (DBDs) that can be associated with several effectors to target a DNA sequence in order to stimulate a specific biological effect [6]. The first synthetic transcription factors designed to modulate the expression of human genes were engineered from zinc-finger proteins [114], able to recognize a specific target through their modular DBD. Despite many achievements in using these modified zinc finger-based DBDs linked to several actuators in different biological application [22, 8, 121, 88], there are many issues related to their engineering that have limited the widespread adoption of this platform. More recently, another system protein-based has been developed, starting from a more detailed knowledge of the structure of the bacterial transcription activator-like effectors (TALEs). Similar to zinc finger proteins, they show a modular DBD that can be modified in order

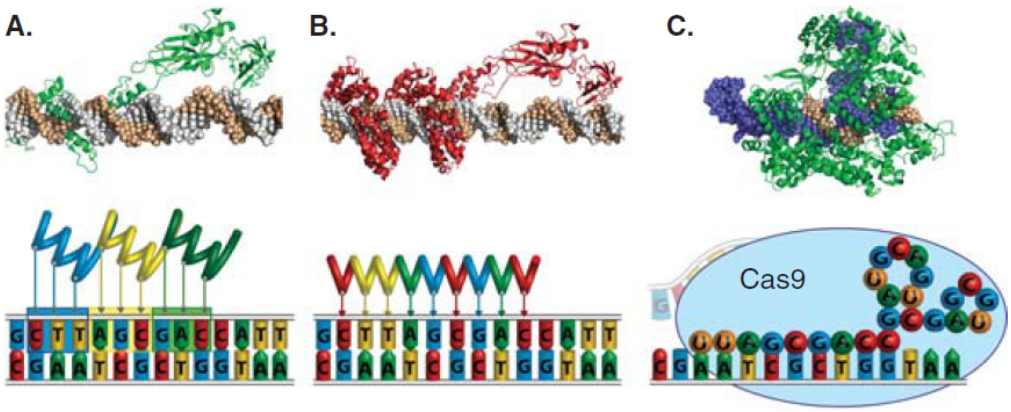

Figure 1.3: *Technologies for engineering programmable DNA-binding proteins, including (A) zinc finger proteins, (B) TALEs and (C) CRISPR/Cas9. Representative crystal structures of a (A) zinc finger protein or (B) TALE fused to the p65 transcriptional activation domain or (C) Cas9 (green) bound to a gRNA (blue) and the corresponding DNA target site (brown).*

to determine the specificity for a particular sequence in the genome. The possibility in engineering these TALE proteins in a relatively simpler way compared to zinc fingers has determined their broad usage in the last years [94]. Moreover, a third technology has emerged based on the RNA- guided DNA endonuclease Cas9 from the type II bacterial adaptive immune system CRISPR [58]. It differs from the previous ones because the targeting of a specific DNA loci is due to the complementary pairing with a guide RNA (gRNA) that form a complex with Cas9 endonuclease, meaning that the system requires only the exchange of the nucleotides into the gRNA expression-cassette, and not the design and assembly of new proteins. Overall, despite the differences in the way of recognizing the DNA target site, the possibility to deliver a specific actuator to it makes the spectrum of action of these platform wide, spanning from the gene editing to the gene activation/suppression [110].
Transcription impairment by mutations in the promoter regions represents a small, but considerable cause of severe coagulation factor defects and of all inherited human diseases. These mutations often affect only the transcriptional regulatory sequences affecting at different extent the expression of completely unaffected genes. The enhancement of the transcriptional activity could potentially restore gene expression at therapeutic levels, above all in the case of monogenic diseases. A recent breakthrough with transcription activator-like effector proteins (TALEs) makes it possible to establish universal types of engineered TFs (eTFs) that can potentially target any selected promoter. TALEs are natural effectors found in the gram-negative *Xanthomonas* plant pathogen bacteria, secreted into the host plant cell cytoplasm and then transported into the nucleus where they function as eukaryotic-like transcription factors. Recognition of specific promoters and the interaction with the transcriptional machinery lead to the selective expression of specific host plant genes [102]. TALEs structure consists in a N-terminal

translocation domain (which permits the protein to be delivered into the cell cytoplasm), a C-terminal region composed of a nuclear localization signal (NLS) and a transcriptional activation domain (AD), and a unique central domain that determines the DNA binding specificity. Typically, the central region of the protein contains tandem repeats of 34 amino acids (aa) in length termed monomers (ranging normally from 7 to 34). Although the sequence of each monomer is highly conserved, they differ essentially in two positions termed RVDs (repeat variable di-residues) in position 12 and 13. These two amino acids are responsible for preferential binding of the repeat module to a single specific nucleotide and a simple cipher specifies the target base of each RVD, where HD, NG/HG and NI recognize almost exclusively cytosine, thymine and adenine, respectively [24, 107]. On the contrary, NN has more degenerated specificity (adenine or guanine). Therefore, the linear sequence of the monomers in the protein determines the DNA target sequence in 5' to 3' orientation, such that each repeat contacts one specific DNA base pair via the RVD. The natural DNA binding site within plant genomes always start with a thymine, which is contained in a cryptic repeat of the N-terminal region (NTR) determining the overall basic charge of the TALE protein [40]. The tandem repeat DBD always ends with an half-repeat, consisting of 20aa. Hence, the length of the DNA binding site correspond to the number of full repeat plus two. This simple coding principle has enabled assembly of TALE repeat arrays for targeting almost any given DNA sequence. The fusion of TALEs to transcriptional activator, such as the monomer or oligomer of the VP16 acidic transactivator [21, 12], could generate TALE-TFs that may be used to specifically induce the expression of a gene as a potential treatment for haplo-insufficiency. Recently, synthetic TALE- based transcription factors (TALE-TFs) targeted to the frataxin gene have been used to enhance its expression in Friedreich ataxia (FRDA) patients fibroblast, down-regulated in presence of a trinucleotide repeat expansion in the first intron [31]. Moreover, TALE-TFs has been applied in the development of genetic reprogramming of induced pluripotent stem cells (iPSCs), thus providing a means for enhancing tissue regeneration and directing cell lineage specification [27]. However, to our knowledge there are no examples of TALE-TFs being used to enhance gene expression impaired by promoter mutations leading to human disease. Interestingly, coagulation factor deficiencies represent a good target for this type of therapeutic approach, as they would significantly benefit even from a modest increase in the production of the functional protein.

1.5 Aim of the study

As a model to exploit TALE-TF transcription restoration in coagulation factor deficiencies, we considered natural promoter mutations impairing F7 gene transcription. It is worth noting that in the F7 mutational pattern a relatively abundant percentage of the cases is represented by the regulatory variants (∼8%), all reported in the Human Gene Mutation Database (`http://www.hgmd.cf.ac.uk/ac/index.php`). In this study we chose two natural promoter mutations found in patients experience severe FVII deficiency (tab 1.3), the -94C>G and -61T>G

transversions relative to the initiation codon [18, 29]. The patients homozygous for these mutations show life-threatening hemorrhagic symptoms and require replacement therapy.

Sequence change	F7 gene region	Molecular defect	Clinical severity	Number of cases
-94C>G	5'-flanking region	Sp1 binding site	Severe	1
-61T>G	5'-flanking region	HNF-4 binding site	Severe	1

Table 1.3: *Synopsis of F7 promoter mutations considered in the study: the relative position respect the transcription starting site, the transcription factors affected, clinical severity and number of cases are listed. All the regulatory variants are reported in HGMD-Human Gene Mutation Database.*

The main scope of this project was to explore the ability of engineered TALE-Transcription Factors (eTFs) to enhance the expression of F7 gene by acting at the transcriptional level, and to rescue promoter mutations that are associated to severe FVII deficiency. More specific objectives were the identification of eTFs able to enhance transcriptional activity on F7 promoter in presence of the mutations impairing gene expression, the assessment of the efficacy and the specificity of the promoter targeting and the evaluation of their capability in modulating F7 expression in the endogenous genomic context of human liver and non-liver cell lines. All these studies were conducted in-vitro to establish the feasibility of this molecular therapeutic approach, in order to subsequently assess the efficacy of eTF mediated-rescue of gene expression in specific mouse models.

1.6 Materials and Methods

1.6.1 Design and assembly of TALE-TFs

The 4 customized TALE-TFs targeting the FVII promoter were assembled using the protocol described by Zhang and colleagues [79] and reagents distributed by Addgene (Cambridge, MA). TALE-repeat domains were synthesized by hierarchical ligation of individual repeat monomers, and cloned into TALE-TF cloning backbones (available by Addgene). The final TALE-TF expressing-plasmids contain the TALE-derived DNA binding domain, the nuclear localization signal (NLS), a transcriptional activation domain (VP64), and a 2A sequence followed by an enhanced EGFP-coding gene.

1.6.2 Construction of the reporter plasmids

– pGL3 FVII *wt* vector was obtained by amplification of a 520bp human FVII promoter region from genomic DNA, by using Pfu DNA polymerase (Thermo Scientific) and with primers containing Hind III recognition site

at 5'ends: FVII prom F,

5'-AAA AAGCTT CTGTGGCTCACCTAAGAAACCAG-3'

and FVII prom R,

5'-AAA AAGCTTGATG AAATCTCTGCAGTGCTGC-3'

. This region was subsequently cloned into a commercial pGL3 Basic Vector (Promega) through HindIII sites.

– pGL3 FVII -94G and pGL3 FVII -61G were obtained by site-directed mutagenesis with the primers FVII -94G F

5'-TCCTCCCCTCC**G**CCATCCCTCTG-3'

and R

5'-CAGAG GGATGG**C**GGAGGGGAGGA-3'

FVII -61G F

5'-GCAGAGAAC**G**TTGCC CGTCAGTC-3'

and R

5'-GACTGACGGGCAA**C**GTTCTCTGC-3'

respectively. The mutagenesis reaction was performed using QuickChange II Site-Directed Mutagenesis Kit (Agilent Technologies). All the restriction reactions were performed with enzymes from NEB (New England Biolabs).

1.6.3 Cell culture and reporter activation assay

The human hepatocyte carcinoma HepG2 and the human embryonic kidney Hek293 cell lines were maintained under 37°C, 5% CO2 using Dulbecco's modified Eagle's Medium supplemented with 10% FBS, 2mM GlutaMAX (Invitrogen), 100U/ml penicillin and 100 μg/ml streptomycin. Firefly luciferase reporter activation was assessed by co-transfecting HepG2 cells with plasmids carrying Luciferase reporters and TALE-TFs. HepG2 cells were seeded into 12-well plates the day before transfection at densities of 0.2×10^6 cells/well. Approximately 24h after initial seeding, cells were transfected using Lipofectamine 2000 (Invitrogen). For 12-well plates we used 1 μg of each plasmid TALE-TF and Reporters, and 100ng of Renilla Luciferase. After about 48h from transfection the reporter activity was established by luminometer measurement in Dual-Luciferase assay (Promega). All fold-induction values were normalized to the activity level of Renilla Luciferase for each transfection. All experiments were performed according to manufacturer's recommended protocol.

1.6.4 FACS analysis

HepG2 cells were seeded into a T75 flask 2 days before transfection at density of 2.1 $\times 10^6$ cells/flask. Approximately 72h after initial seeding, cells were transfected with 20 μg of the plasmid encoding TF4 using Lipofectamine 2000 (Invitrogen). After 24h the transfection efficiency was established by evaluating the eGFP expression by a fluorescence microscope. After about 48h post-transfection HepG2 cells were trypsinized, resuspended in a PBS-5 EDTA solution and filtered through a 50 μm cell strainer to eliminate clumps and debris. eGFP-positive cells were recovered in DMEM F12 medium after FACS analysis performed with a BD Facs Aria II sorter (BD Biosciences). The total RNA was extracted with Trizol Reagent (Invitrogen). The cDNA was generated by iScript cDNA Synthesis Kit (Bio-Rad) according to the manufacturer's recommended protocol. F7 transcript was detected by PCR with primer F7 ex5 F

5'-GAGAACG GCGGCTGTGAG-3'

and F7 ex7 R

5'-GTTCCTCCAGTTCTTGATTTTGTCG-3'

1.6.5 RT-PCR and qPCR for FVII mRNA

HepG2 and Hek293 cells were seeded in 12-well plates the day before transfection at density of 0.2 $\times 10^6$ cells/well. After 24h from the seeding cells were transfected with 2 μg of each TALE-TF. Approximately 48h after transfection the total RNA was extracted with Trizol Reagent (Invitrogen). The cDNA was generated by iScript cDNA Synthesis Kit (Bio-Rad) according to the manufacturer's recommended protocol. F7 transcript was detected by PCR with primer F7 ex5 F and F7 ex7 R. GAPDH, 18S and FVII mRNAs were detected by qPCR using SsoFast EvaGreen Supermix (Bio-Rad).

1.7 Results

1.7.1 Assembly of engineered TALE-TFs and reporter plasmids

The two natural variants we considered, the -94 C>G and the -61 T>G transversions relative to the initiation codon, are reported to falling in the binding site for the hepato-specific HNF-4 and the ubiquitous Sp1 transcription factors, respectively (Fig. 1.4). The proximal promoter region of FVII gene has been shown to maintain the maximal promoter activity in HepG2 cells [76], with the HNF-4 and Sp1 binding sites included in the first 185 bp upstream the translation start site (TSS). We have therefore generated the reporters vectors for the F7 promoter, wt and mutants, by cloning 520 bp human F7 promoter region upstream the coding sequence for firefly luciferase (Fig.1.4), so obtaining the pGL3 FVII wt, pGL3

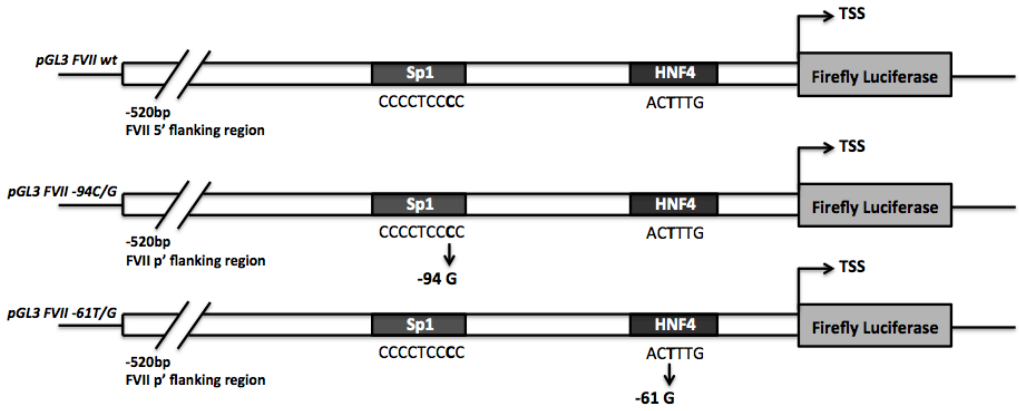

Figure 1.4: *Scheme representing the reporter plasmids pGL3 FVII wt, -94G and -61G. In each construct the firefly luciferase expression cassette is cloned downstream the F7 proximal promoter (520 bp). The reported Sp1 and HNF-4 binding sites with the point mutations are included in the first 185bp before the TSS.*

FVII -94G and pGL3 FVII -61G plasmids. Based on computational analyses of the F7 promoter and of the transcription factor binding sites in relation with the position of the nucleotide changes, we performed a tailing design for a panel of TALE-TFs targeting the functional sequences and resulting in 4 TALE-TF platforms (TF 1, 2, 3 and 4) covering a region of 70 bp ranging from -122 to -52 of the F7 5'-flanking region (Fig.1.5). In order to reduce off-target activation we targeted sequences from 18 to 26 bp in length spanning the region between Sp1 and HNF-4 binding site on FVII promoter predicted to be unique sites in the human genome by performing an *in–silico* analysis. All the constructs are

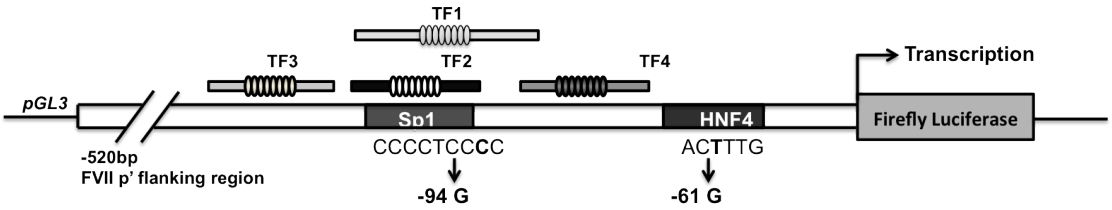

Figure 1.5: *Scheme representing the reporter system. Firefly luciferase gene is regulated by the proximal promoter region of FVII gene (520bp),* wt *or mutated, targeted by four different TALE-TFs (TF1, TF2, TF3 and TF4).*

based on a specific TALE-TF cloning backbone which permits the expression of the customized TALE protein fused with a transcription activator that is, four VP16 peptides (i.e.VP64), able to recruit the transcriptional machinery on the target promoter (Fig1.6a). The final TALE-TF expressing-plasmids contain also a 2A sequence followed by an enhanced EGFP-coding gene (Fig.1.6b).

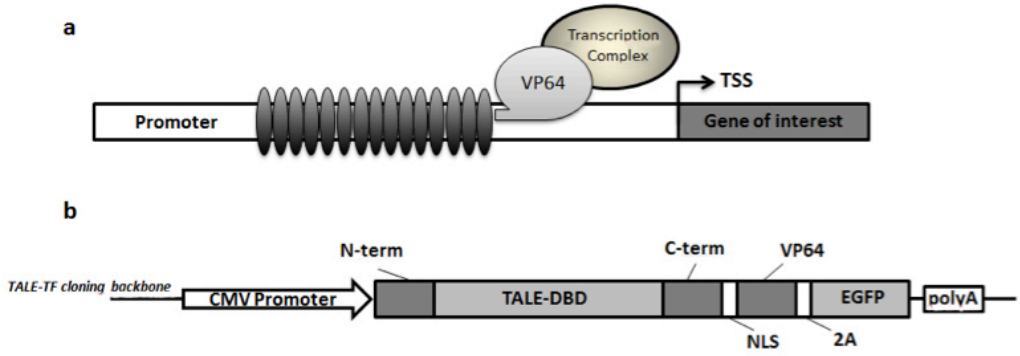

Figure 1.6: *a) Scheme representing the TALE-TF platform. Each synthetic TALE-TFs is expressed as a fusion protein composed of a TALE-derived DNA modular binding-domain and a VP64 transactivation domain, able to recruit the transcriptional machinery on a specific target promoter to stimulate the transcription of the gene of interest. b) The TALE-TF expression plasmid is composed of the TALE-derived DNA binding domain, the nuclear localization signal (NLS), a transcriptional activation domain (VP64), and a 2A auto-cleavage peptide followed by an enhanced EGFP-coding gene. The expression cassette is driven by the CMV strong promoter.*

1.7.2 In vitro models: effect of promoter mutations and validation of the efficacy of eTFs

The expression of the reporter plasmids pGL3 -94G and pGL3 -61G in HepG2 cell line showed the causative effect of the mutations, with an abolishment of the firefly luciferase transcriptional levels compared to the pGL3 FVII *wt* (Fig.1.7). Most importantly, we observed a robust increase (25±4.9 fold increase for pGL3 -61G and 5.9±0.6 for pGL3 -94G) in the transcriptional levels of the reporter genes guided by promoter mutants by co-transfecting the cells with TALE-TF 4 (fig.1.7), and thus was confirmed also when it was co-expressed with the pGL3 FVII *wt* (36.6±17.3 fold increase). Conversely, TALE-TF 1, 2 and 3 did not show a significant transcriptional activity, either with the mutants and the F7 promoter *wt*. A negative control was represented by a pBasic vector carrying the firefly luciferase gene without an upstream promoter sequence (fig.1.7).

1.7.3 Assessment of the specificity of TF4

To investigate more in detail the specificity of TF4, our best candidate, we performed a series of experiments in HepG2 cells by modulating some in-cis and in-trans elements of the reporter genes transcriptional system. First we co-transfected TF-4 expressing-vector with the pBasic vector lacking the 5' promoter region and, in parallel, with a pSlug vector containing the firefly luciferase guided by the ubiquitous Slug promoter derived from a Snail gene family of zinc-finger transcription factors [32]. In both conditions transcriptional levels of the reporter gene remained unaltered (Figure 1.8). A positive control was represented by the

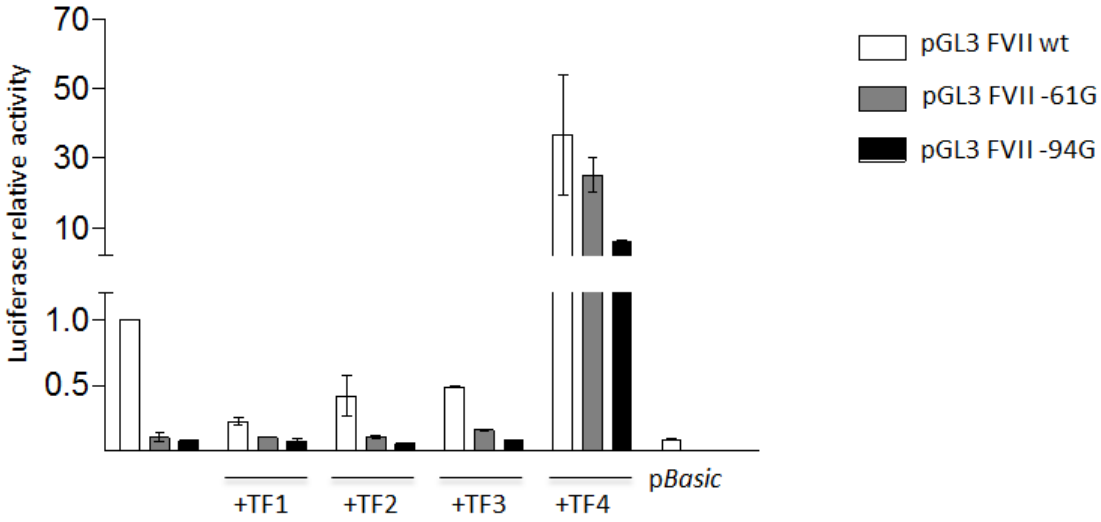

Figure 1.7: *Co-expression of the reporter plasmids* wt*, pGL3 -94G or -61G with each synthetic TALE-TFs in HepG2 cell line. Histograms report luciferase relative activity as change fold relative to pGL3 FVII* wt *alone. The negative control is represented by a plasmid pBasic carrying firefly luciferase gene without an upstream promoter. All the results are expressed as mean ±SD derived from three independent experiments.*

co-transfection of TF-4 with pGL3 FVII *wt*.

Moreover, to test the DNA binding specificity we introduced 5 mismatches in

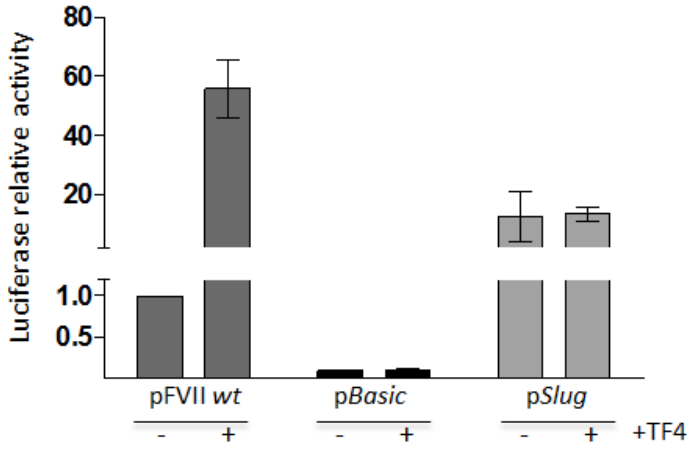

Figure 1.8: *Co-expression of the TALE-TF4 with pGL3 FVII* wt*, pBasic or pSlug reporter plasmids in HepG2 cell line. Histograms of the dual assay report firefly luciferase relative activity as change fold relative to pGL3 FVII* wt *alone. All the results are expressed as mean ±SD derived from three independent experiments.*

the TF4 target sequence by site-specific mutagenesis of the plasmids pGL3 FVII *wt* and pGL3 FVII -61G, thus creating the pFVII *wt* ΔT4 and the pFVII -61G ΔT4 vectors. In presence of 5bp changes TF4 DNA binding and transactivation

capability is virtually abolished, as shown in Figure 1.9. In addition, we tested

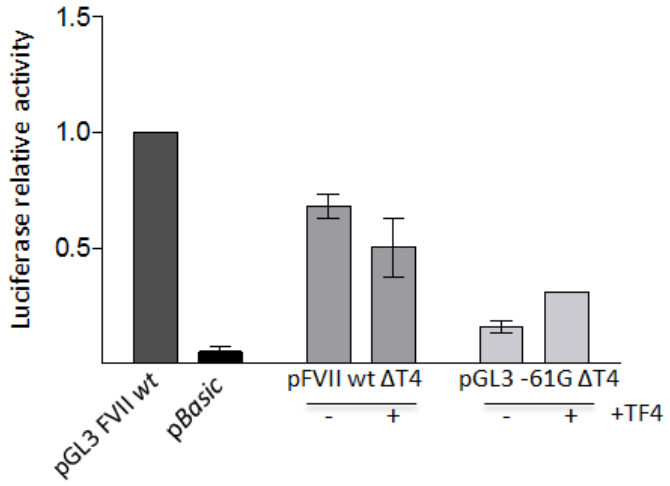

Figure 1.9: *Effect of TF4 co-transfected in HepG2 with pGL3 FVII* wt *ΔT4 or pGL3-61 ΔT4 reporter plasmids, expressing firefly luciferase guided by FVII proximal promoter carrying 5 mismatches in TALE-TF4 binding site. The negative control is represented by pBasic vector. All the results are expressed as mean ±SD derived from three independent experiments*

the TALE-DNA binding domain-play role by co-transfecting pGL3 FVII *wt*, -61G and -94G variants with a pHD plasmid expressing the VP64 transactivator without the TALE-derived DBD, and even in this case any substantial enhancement in firefly relative activity was appreciable (Figure 1.10).

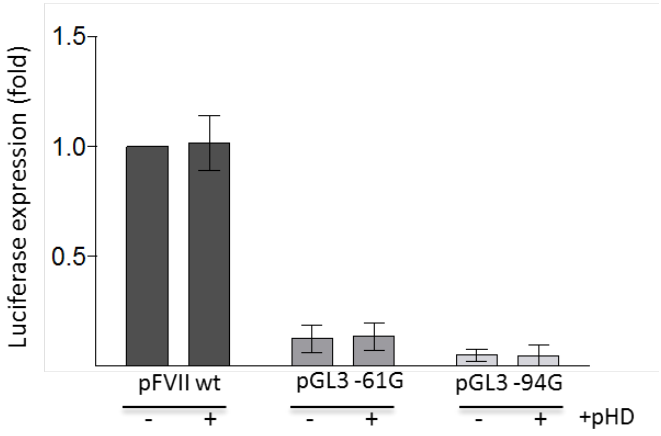

Figure 1.10: *Co-transfection of HepG2 cells with the pGL3 FVII* wt*, -61G or -94G with a pHD backbone vector expressing the VP64 transactivation domain alone. Histograms of the dual assay report firefly luciferase relative activity as change fold relative to pGL3 FVII* wt *alone. All the results are expressed as mean ±SD derived from three independent experiments.*

1.7.4 Effect of TF4 in the endogenous context of HepG2 cells

To assess the capability of TF4 in enhancing gene expression in the genomic context we targeted the endogenous F7 promoter in HepG2 cells. Because of the low transfection efficiency assessed in this cell line by using lipofectamine method (<10%), we took advantage of the eGFP expression from the TF4 coding-vector (fig.1.6b) thus selecting eGFP positive-HepG2 post-transfection by FACS analysis. Total RNA was then extracted from both eGFP-positive and negative cells, followed by a specific RT-PCR in order to detect F7 mRNA. In the eGFP-positive HepG2 (+TF4) the F7 transcript amount was increased compared to eGFP-negative cells (fig.1.11a). The quantification of F7 mRNA levels through qPCR showed an increase by 2.3±0.2 fold (P <0.01) after treatment with TF4 (fig.1.11b). Data analyses were performed with T-Student test comparing TF4-treated cells with untrasfected mock cells; * $p < 0.05$; ** $p < 0.01$.

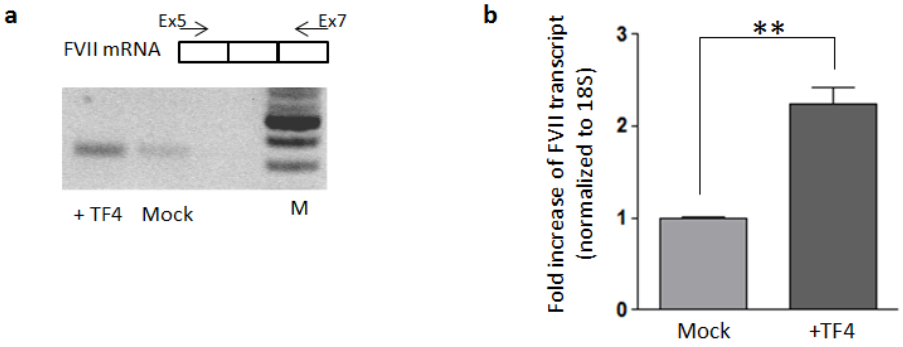

Figure 1.11: *Effect of TF4 on the FVII transcriptional levels in HepG2 endogenous context. a) RT-PCR performed on cDNA from HepG2 eGFP positive and negative (mock) cells. F7 transcript was amplified using a primer F on exon 5 (Ex5) and a primer R on exon 7 (Ex7). PCR products (312bp) were detected by gel electrophoresis on agarose gel 2%. b) Levels of FVII mRNA in transfected GFP positive-HepG2 cell compared to the mock, as determined by relative qPCR. All the results are expressed as mean ±SD derived from three independent experiments. The change folds are reported to 18S housekeeping gene expression.*

1.7.5 Effect of TF4 in the endogenous context of Hek293 cells

We tested TF4 transcriptional activity also in the endogenous context of Hek293 cells that normally do not express F7 gene (`http://webserver.mbi.ufl.edu/~shaw/293.html`), as confirmed from our experimental data. Hek293 cells were treated with TF4, and the presence of F7 mRNA was confirmed by specific RT-PCR. Conversely, in the mock non-treated Hek293 cells no F7 mRNA was detectable, as expected (fig.1.12a). The increase in F7 mRNA levels was quantified

by qPCR and established at 6.3±1.3 fold (P <0.05) compared to the mock cells (fig.1.12b). Data analyses were performed with T-Student test comparing TF4-treated cells with untrasfected mock cells; * P < 0.05.

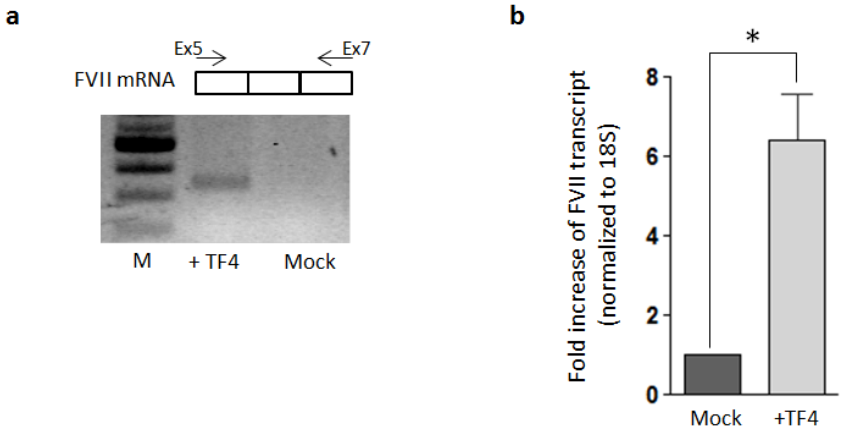

Figure 1.12: *Effect of TF4 on the FVII transcriptional levels in HeK293 endogenous context. a) RT-PCR performed on cDNA from Hek293 TF4-treated and non-treated cells. F7 transcript was amplified using a primer F on exon 5 (Ex5) and a primer R on exon 7 (Ex7). PCR products (312bp) were detected by gel electrophoresis on agarose gel 2%. b) Levels of FVII mRNA in TF4-treated Hek293 cells compared to the mock (transfected with the lipofectamine vehicle), as determined by relative qPCR. All the results are expressed as mean ±SD derived from three independent experiments. The change folds are reported to 18S housekeeping gene expression.*

1.7.6 Effect of the TF1, TF2 and TF3 in the endogenous context of HepG2 and Hek293 cells

As experimental control HepG2 and Hek293 cells were transfected with TF1, 2 or 3, even if they did not show appreciable transcriptional activity in the gene reporter assays (fig1.7). F7 mRNA levels were quantified by qPCR and normalized to the untreated cells. The F7 mRNA was increased by 1-1.5 fold (P value=ns) in HepG2 cells (fig.1.13a) and by 2-3 fold in Hek293 cells (fig.1.13b), a non-significant effect compared to TF4 treatment (fig.1.12a). Data analyses were performed with one-way ANOVA and Bonferroni correction for multiple comparisons (TFs compared to the mock).

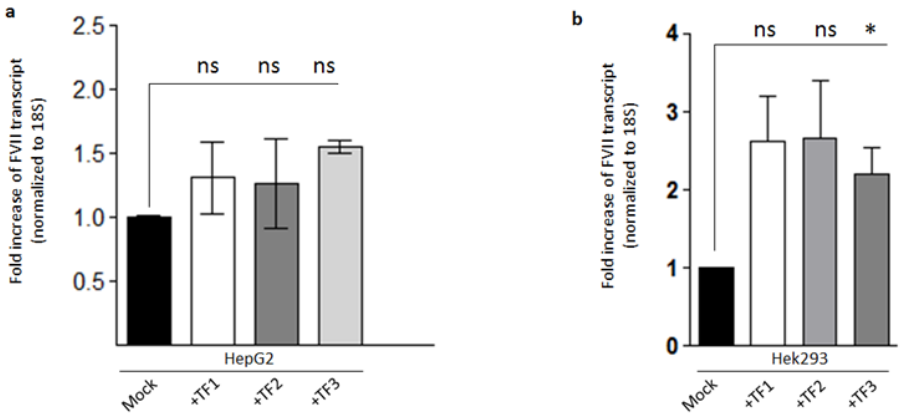

Figure 1.13: *Effect of TF1, TF2 and TF3 on the FVII transcriptional levels in HepG2 and HeK293 endogenous context. a) Levels of FVII mRNA in HepG2 cells treated with TF1, 2 and 3 compared to the mock untreated cells, as determined by relative qPCR. All change folds are reported to 18S housekeeping gene expression. b) Levels of FVII mRNA in Hek293 cells treated with TF1, 2 and 3 compared to the mock untreated cells, as determined by relative qPCR. All the results are expressed as mean ±SD derived from three independent experiments. All change folds are reported to 18S housekeeping gene expression.*

1.7.7 Discussion and future perspectives

The in-vitro study is part of a wider project focused on the genetically-determined deficiencies of blood coagulation factors that, in the most severe cases, are associated to life-threatening bleeding symptoms, sometimes also fatal. It is worth noting that in these pathologies even modest increase of functional protein levels in plasma significantly ameliorates the bleeding phenotype. FVII deficiency due to promoter mutations represents a good model to investigate the ability of engineered TFs of enhancing gene transcription for therapeutic purposes. As a matter of fact, increased expression of F7 gene may reduce or completely reverse the symptoms of this deficiency, given that patients have been shown to receive benefit from restored level of proteins up to 6% [98]. In this in-vitro study the model of FVII deficiency has been validated for the promoter variants -94C>G and -61T>G by exploiting gene reporter assays in HepG2 cell line, due to the impossibility of conducting the experiments in primary hepatocytes from patients, which are not available. These mutations, by altering the DNA binding sites for the natural transcription factors Sp1 and HNF-4 [18, 29], lead to a consequent decrease in the functionality of the promoter and reduce the transcriptional levels, as confirmed by the firefly luciferase reporter assays (fig.1.7). The screening with four engineered TFs designed to recognize different target sequences spanning from Sp1 to HNF-4 binding sites on F7 promoter allowed the identification of TF4, that showed a robust enhancement of the transcriptional activity (from ∼6 to 50 fold increase, depending on the variant considered, fig.1.7). The different changes in the expression levels for the mutants -94C>G and -61T>G (fig.1.7) could be explained by a different DNA-protein interaction framework governing

transcription in the different mutant context. Besides, it is important to note that the increased expression is observable also in the presence of the F7 promoter *wt*, while the treatment with the TF1, 2 and 3 is no effective in any case (fig.1.7), thus giving important information on the promoter accessibility. The reason why only TF4 is effective could be at least in part attributable to the position of its target sequence between the Sp1 and HNF-4 binding sites. This probably represents the best positioning from the TSS in close proximity to the endogenous TFs binding sites in order to recruit the transcriptional machinery, and to avoid the competition of the TF4 with the other TFs on their own DNA binding sites. An important issue that has to be considered in the correction approach with engineered TFs is the specificity, since there is a potential risk arising from non-specific DNA-TF interactions that would lead to the alteration of off-target gene expression. One way to improve the specificity is a careful design of the protein DBD. For this purpose the four TFs were assembled to recognize DNA sequences from 18 to 26bp in length that, by in-silico analyses, are predicted to be unique into the genome. We focused in particular on the potential off-target effects that could be induced by the treatment with TF4, the only effective, testing its binding specificity by modulating some elements in the fusion protein itself and in the target sequence. The co-transfection of TF4 expressing-plasmid with the pSlug vector carrying firefly luciferase under the control of the Slug promoter showed no increase in the reporter activity (fig.1.8), thus supporting the specificity for F7 promoter. Most importantly, we found that 5 mismatches in the TF4 target sequence (fig.1.9) are able to abolish the TF4 transactivation capability, thus strengthening a mechanism strictly dependent on the highly specific target sequence, as previously reported [41]. We confirmed that the recruitment of the transcriptional machinery by eTF4 on the promoter is specifically due to its binding with the DNA by testing the activity of VP64 lacking the TALE-derived DBD. The co-expression of VP64 alone in presence of the F7 promoter *wt*, as well as the two mutants, did not showed any change in the reporter relative activity (fig.1.10). Data obtained from the gene reporter assays showed the efficacy of TF4 that is able to efficiently recruit the transcriptional machinery and rescue the expression of the reporter otherwise abolished in presence of the disease-causing mutations. However, it is worth noting that the reporter genes episomal system may not be representative of the physiological context, because it does not permit to take into account the chromatin context. To address this issue we assessed the activity of the four eTFs on the endogenous F7 gene expression in HepG2 and Hek293, which express or not express FVII, respectively. In HepG2, a human hepatoma cell line, the treatment with TF1, 2 and 3 had a minor impact on F7 mRNA levels, with the maximum change fold measured around 1.5 (fig.1.13a), reflecting the previous data obtained with the gene reporter system. On the other hand, the treatment with eTF4 showed a more robust increase in F7 mRNA levels (fig.1.11a) established to be 2.3 ± 0.2 fold. Surprisingly, this fold increase was markedly less pronounced as compared to that observed in the episomal plasmid context with the F7 promoter *wt* vector (fig.1.7). It is therefore possible to argue that chromatin environment has a

strong impact on the capability of the eTF to bind efficiently F7 promoter and/or to enhance transcription via VP64. This finding is supported by other literature data in which the magnitude of gene activation by eTFs at native chromosomal loci was relatively modest [66, 93].

Nevertheless, it has also to be considered that the enhancement achieved is referred to its action on the F7 promoter *wt*, thus supporting the idea that a major improvement could be reachable in presence of promoter mutations. Experiments conducted in human embryonic kidney Hek293 background showed F7 mRNA levels increased from 2.2±0.3 (+TF3, fig.1.13b) to 6.3±1.3 times (+TF4, fig.1.12b). It is worth noting that Hek293 cells normally do not express FVII (as revealed by RT-PCR and qPCR experiments in the transfected cells), though after treatment with TF4 F7 mRNA was clearly detectable by RT-PCR (fig.1.12a). This data confirms the capability of TF4 in actively stimulating F7 gene expression in Hek293 cells. The different behavior in HepG2 and Hek293, with a surprising activity in Hek293 cells in which the transcription on F7 promoter is normally down-regulated, may underlie a competition of TF4 with the liver specific HNF-4 in HepG2 (`http://www.cisreg.ca/cgi-bin/tfe/articles.pl?tfid=140&tab=expression`).
Overall, these results suggest that TF4 should be able to increase expression of the F7 gene in patient cells. This may potentially reduce or completely reverse the symptoms of FVII deficiency, as patients with FVII deficiency have been shown to receive benefit from restored levels of protein up to 6% [98]. We are conscious that the study in vitro, by exploiting reporter gene assay, does not permit the evaluation of eTFs effect on the proper chromosomal context and thereby the efficacy of TF4 has to be validated in a proper model of FVII deficiency caused by promoter mutations. Future experiments will be addressed to test the detrimental impact of human F7 promoter mutations in the murine environment by transfecting expression vectors, either *wt* or mutated, in Hepa1-6 murine cells. Normal or impaired expression of the human FVII gene will be evaluated by qRT-PCR. Mutations that will affect transcription to similar extent in the human and mouse liver cell lines will provide the rationale for the subsequent experiments in-vivo. These will be conducted in specific mouse models for promoter mutations created by zygote injection of plasmids encoding CRISPR-Cas9 system to target the F7 murine promoter or a safe locus and insert the human F7 (hF7) gene under the control of the human promoter, either wild-type (positive control) or bearing the natural mutations. In our hypothesis, the mutations would prevent the appreciable expression of the hFVII in mouse liver, and thus the presence of circulating protein. Injection of liver-specific AAV vectors expressing TF4 in these new animal models and evaluation of hFVII expression at the mRNA and protein level will permit the assessment of the correction efficacy in vivo. Moreover, a proper animal model will permit the determination of the optimal delivery method, the characterization of the specificity in the gene regulation and the evaluation of the immunogenicity of the synthetic proteins, very important issues when proposing an innovative therapy.

Chapter 2

SB Transposon

2.1 Focus on FIX: gene and protein features

2.1.1 Factor FIX protein

Human coagulation factor IX, also called Christmas factor or antiemophilic B factor, is a single chain glycoprotein synthesized in liver and circulating in blood like a zymogen at a physiologic concentration of 5 μg/ml [1]. As FVII it features a N–terminal γ–carboxyglutamic acid–rich domain followed by two EGF–like domains and the C–terminal serine protease domain. The mature protein consists of 415 aminoacids of 57 KDa molecular weight. The first 40 aminoacids constitute the γ–carboxyglutamic (Gla) domain, which contains post-translational modifications of many glutamate residues by vitamin K–dependent carboxylation to form γ–carboxyglutamate (Gla). The Gla residues are responsible for the high-affinity binding of calcium ions [101, 117, 54], so conferring to FIX its biological activity. Then the protein consists of a short hydrophobic segment (residues 41-46), two EGF–like domains (residues 47-84 (EGF1) and 85-127 (EGF2)), an activation peptide region (residues 146-180), and a serine protease module (residues 181-415). Activation of FIX involves proteolytic cleavages at Arg145-Ala146 and Arg180 and Val181 bonds with a concomitant release of a 35-residue activation peptide (3, 40). The FIXa thus formed contains a light chain (residues 1-145) and a heavy chain (residues 181-415) held together by a single disulfide bond. The light chain consists of the Gla, EGF1, and EGF2 domains whereas the heavy chain contains the serine protease domain that features the catalytic triad of residues His-c57221, Asp-c102 269 and Ser-c195 365 [118, 103, 2].

2.1.2 Factor FIX gene

The gene of human factor IX (F9) is localized in q27.1-q27.2 region of chromosome X. Isolated and sequenced in 1985 [91] it is approximately 34kb in lenght and contains eight exons (spanning from 25bp to 548bp) and seven introns. The transcript is 2803 bases in length and comprises a short 5‘ UTR (29 bp), an open reading frame (1383 bp) and a 3‘ UTR (1390 bp). Exon 1 and part of exon

Exons	Coding region	Dimension (bp)
1	Pre-pro leader sequence	88
2	Pre-pro leader, Gla domain	164
3	Gla domain	25
4	EGF-like domain	114
5	EGF-like domain	129
6	Activation site	203
7	Catalytic domain	115
8	Catalytic domain	571

Table 2.1: *Factor IX gene exons organization*

2 encode for the pre-pro leader sequence, which is removed during the protein biosynthesis, while the last part of exon 2 together with exon 3 encode for the Gla domain containing 12 γ–carboxiglutamic acid residues. The γ–carboxilation is a modification needed for the proper folding of the FIX and for its calcium binding capacity. Exons 4 and 5 encode respectively for the first and the second epidermal growth factor-like domains (EGF-like 1 and 2). Exon 6 codes for the activation domain, in which factor IXa and the complex FVIIa-TF cut in order to activate FIX. The last two exons encode for the catalytic domain of the protein (tab.2.1) [17].

2.2 FIX deficiency-Haemophilia B

Haemophilia B, or Christmas disease, is an inherited, X-linked, recessive disorder that results in deficiency of functional plasma coagulation factor IX [11]. It has a prevalence of around 1 in 30,000 live births (about five times rarer than haemophilia A). There are usually carrier females and affected males. The classification of the severity of haemophilia B is based on either clinical bleeding symptoms or plasma pro-coagulant levels; the latter are the most widely used criteria.

- Severe disease occurs with a factor IX level below 1% of the reference and accounts for about 50% of cases.
- Moderate severity occurs with a level of 1-5% and accounts for around 30% of cases.
- Mild disease is with levels of 6-30% and accounts for around 20% of cases.

People with mild haemophilia B typically experience bleeding only after serious injury, trauma or surgery and in many cases the disease is not diagnosed until an injury, surgery or tooth extraction results in prolonged bleeding. Women with

mild haemophilia often experience menorrhagia, heavy menstrual periods and can hemorrhage after childbirth. People with moderate haemophilia tend to bleed after injuries, and finally people with a severe phenotype may have also frequent spontaneous bleeding episodes, often into their joints and muscles. The current treatment is based on the intravenous administration of FIX (replacement therapy), either plasma derived or recombinant [106, 119]. Since even small increase in FIX levels (>2%) would result in significant amelioration of the clinical phenotype, HB represents a model to investigate innovative therapeutic approaches in a quantitative manner, by virtue of functional and protein assays in plasma. Enormous efforts have been pushed on substitutive gene therapy, and very recently it has been demonstrated that intravenous infusion of a AAV vector encoding FIX in HB patients resulted in FIX expression ranging from 1% to 6% for periods of 2 years, with amelioration of bleeding phenotypes [104, 68]. DNA editing by using zinc finger nucleases has been also exploited in HB mouse models [53]. However, there are still limitations regarding the delivery method to the target site and safety of gene therapy that encourage research towards alternative therapeutic approaches.

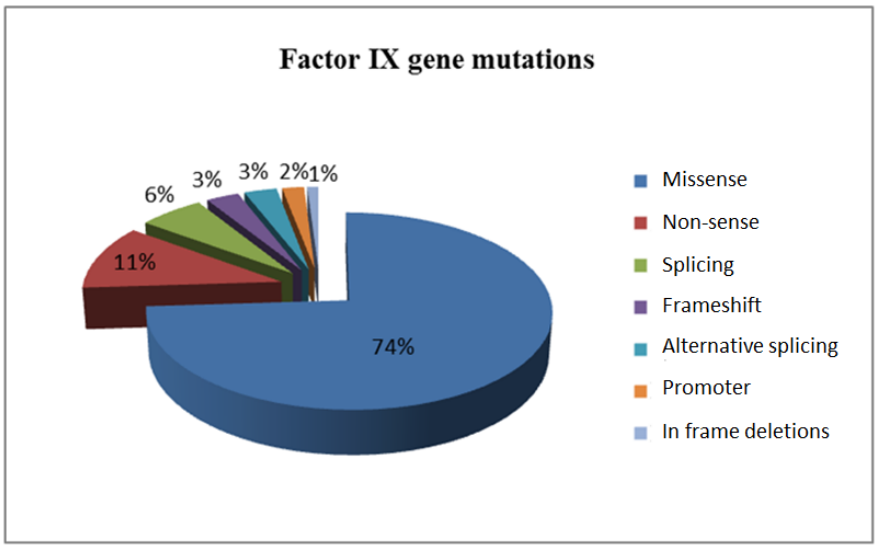

Figure 2.1: *Pie chart showing the F9 mutational pattern reported in the Haemophilia B Mutation Database.*

2.3 Background and rationale

In the last years enormous efforts have been made to develop therapeutic approaches for human genetic diseases alternative to protein replacement therapy. The most remarkable results arise from the gene therapy field that has shown its huge potential in the cure of diseases such as primary immunodeficiencies [15, 14, 44], Leber's congenital amaurosis [57], cancer [9], Fanconi's anemia [85]

and, recently, Haemophilia B [68, 69]. The concept behind gene replacement is very simple, concerning the possibility of efficiently transferring and expressing a normal copy of a defective gene into target cells [115]. Fascinatingly, the advances in biotechnology have allowed the therapeutic technique to move from the conceptual stage to its application in patients for clinical trials for a variety of disorders. To date, from 1989 to 2014 over one thousand gene therapy clinical trials have been completed, are ongoing or have been approved worldwide (`http://www.abedia.com/wiley`). However, there are still issues that limit the wide applicability of these approaches. Limitations are related to the transgene size, the long-term expression and the regulation of the therapeutic gene, to the immune response, to the nature of the DNA vehicle and to harmful effects such as insertional mutagenesis when the transgene is integrated into the genome [5]. Therefore, research is still pushed towards alternative and/or complementary therapeutic strategies based on the correction of the specific disease-causing defects, in order to try to circumvent some of these limitations. These approaches are of great interest especially for patients that would benefit even by small increase in functional protein levels, as it happens in the case of coagulation disorders (`http://www.haemophilia.org`). HB is a prototypical example of disease with a heterogeneous mutational pattern (`www.factorix.org`) and a relatively high frequency of splicing mutations (>15%) in severe disease forms. Mutations inducing aberrant splicing, at splice junctions or within exons, can be addressed by intervention at the pre-mRNA level and this therapeutic approach have been intensively studied by us to correct splicing mutations causing severe HB and FVII deficiency [37, 74, 75, 19]. In particular, we proposed the development of modified U1 small nuclear RNA (U1 snRNA), fundamental component of the U1small nuclear ribonuclearparticle (U1snRNP) with an essential role in exon definition [108], to restore exon inclusion during pre-mRNA maturation impaired by exon-skipping causing mutations.

2.4 Exon Specific U1s (ExSpeU1s) as therapeutic tool

An early event in exon definition is the recognition of the 5'ss by the U1 snRNP, which is composed by a 164bp-long U1 snRNA associated to U1-specific proteins (U1-70K, U1A and U1C) and Sm proteins that are also present in the other snRNPs [108]. The gene coding for the U1 RNA is present in multiple copies in the human genome, is highly evolutionary conserved and possesses its own specific Pol II promoter and polyA site in a short gene long $\sim$ 600bp . The assembly of the spliceosome in the early phase of the splicing process begins with the recognition of the 5' splice site (5'ss) sequence by the U1snRNP through its 5' RNA tail [78, 89], which U1 snRNP-associated proteins U1-70K and U1 stabilize this transient interaction. Another important step following the U1 snRNP-5'ss binding is the recognition of the 3'ss mediated by the U2 auxiliary factor [81]. The subsequent establishment of multiple interactions between the 3'ss and the

5'ss defines an exon, and constitutes the commitment step towards the splicing pathway [116, 4]. U1snRNP, through its U1A component, is also involved in the polyadenilation of the pre-mRNA and modified U1snRNAs with a tail that hybridize to the 3'-UTR of target transcripts have been tested for modulating gene expression [23, 52, 51]. The 5'-tails of U1 snRNAs has been previously altered in order to deliver antisense sequences. A modified U1snRNA targeting mouse DMD gene exon 23 3'ss and 5'ss was systemically delivered using AAV vectors to the dystrophin-deficient mouse model of DMD, and Body-wide dystrophin restoration was observed in treated mice. [34]. More recently, investigators used U1 snRNAs to correct splicing through exon skipping in human DMD pre-mRNA in primary patients fibroblasts [30]. A limited number of studies have explored the role and potential therapeutic effects of U1 snRNAs in correcting donor splice site mutations in cellular models [74, 20, 43, 73, 80, 83, 84]. Recently, the first "proof of concept" that a modified U1 snRNA is able to rescue mRNA processing and protein biosynthesis of coagulation factor VII (FVII) has been provided in mouse models expressing the splicing defective 5'ss human FVII transgene [19]. In all cases, the modified tails of the U1snRNAs have few nucleotide changes in comparison to the *wt* sequence and base pair to the mutant donor sites. However, this approach has limitations that reduce its potential applicability in therapy. First of all, it implies the use of a specific modified U1 snRNA for each pathogenic mutation, hardly conceivable for the development of therapies for the majority of human genetic disorders in which the mutational pattern is highly heterogeneous with several different splicing mutations. Moreover, modified U1 snRNAs complementary to the 5'ss might target by complementarity similar donor splicing sites and thus potentially interfere with splicing of other pre-mRNAs with toxic effects. As a matter of fact, in our in-vivo studies we observed liver toxicity in mice injected with the highest doses of adeno-associated viral vector (AAV) expressing the modified U1 RNAs[19], which might point toward off-target effects. The use of Exon Specific U1snRNAs (ExSpeU1s) recently identified in our group [37] represents a novel promising strategy to overcome the above mentioned limitations. ExSpeU1s are designed to bind by complementarity to intronic sequences downstream of the exon donor splice site and rescue different types of splicing defects associated to exon skipping, a very common pathogenic mechanism. Binding of the ExSpeU1s to non-conserved intronic sequences could significantly reduce the possibility of off-target events. ExSpeU1s are active on several 5'ss mutations, on mutations in the polypyrimidine tract and on defective splicing caused by exonic variants [37]. Most importantly, we demonstrated that a unique ExSpeU1 is able to correct several mutations inducing exon skipping of a given exon, thus providing a potential therapeutic tool for a cohort of patients and not only for very few of them. We exploited ExSpeU1s in cellular models of coagulation factor IX (FIX) deficiency, Cystic Fibrosis and Spinal Muscular Atrophy and demonstrated that they are able to rescue gene expression to a significant extent [37]. The ExSpeU1 therapeutic approach acting at pre-mRNA level has the advantages of maintaining the transcriptional gene regulation and of guaranteeing the correction in the physiologic tissues only. Moreover, it is based on the use of small sequences that

can be packaged virtually in any viral vector so far successfully tested for replacement gene therapy purposes. This previous data encouraged studies aimed at further improving this approach and at testing it in animal models.

2.5 Creation of a stable model: the Sleeping Beauty transposon system (SBTS)

The creation of proper cellular and mouse models of HB caused by splicing mutation was performed by exploiting the Sleeping Beauty transposon system (SBTS), one of the promising technology platform for gene transfer in vertebrates [115]. SB is a synthetic transposable element that has been generated by reverse engineering from defective copies of an ancestral Tc1/mariner-like active vertebrate element from "dead" transposon fossils found in salmonid fish genomes, which was so named Sleeping Beauty [48]. In its natural form, SB transposon consists of a single gene encoding the transposase enzymatic factor of transposition, which is flanked by terminal inverted repeats (ITRs) containing binding sites for the transposase itself (fig 2.2) [113]. Specific binding to the ITRs is mediated by an N-terminal, paired-like DNA binding domain of the transposase [48, 47, 49] that overlaps with a nuclear localization signal (NLS) [47]. An AT-hook motif has also been identified as a functional subdomain that is involved in the DNA binding [49]. The catalytic domain of the transposase mediates the DNA cutting and the subsequent integration reactions and is characterized by the DDE signature, an evolutionary conserved domain also found in some bacterial transposases, retrotransposon/retrovirus integrases and the RAG1 immunoglobulin gene recombinase [35]. The transposase gene can be physically separated from

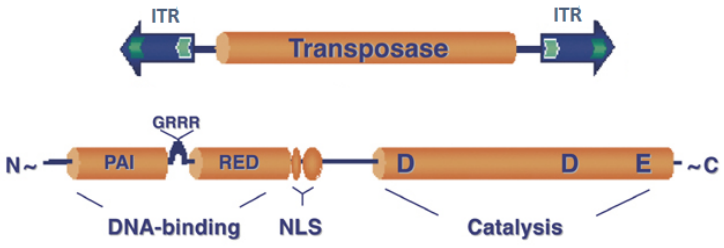

Figure 2.2: *The Sleeping Beauty transposable element and the transposition mechanism. (A) Components of the element. On top, a schematic drawing of the transposon is shown. The element has a single gene encoding the transposase, which is flanked by terminal inverted repeats (IR/DRs, blue arrows), each containing two binding sites for the transposase (small green arrows). The transposase has an N-terminal, bipartite, paired-like DNA-binding domain containing a GRRR AT-hook motif, a nuclear localization signal (NLS), and the DDE catalytic domain.*

the ITRs and replaced by other DNA sequences, because transposase enzyme can mobilize transposons in trans, as long as they retain the ITRs. The transposase gene can be located on the same DNA molecule as the transposon [64], or supplied

by another DNA molecule [124], or supplied in the form of either mRNA [36] or protein. The SB technology used in this study is based on a two-vector system (fig. 2.3). The transgene of interest is cloned into the transposon plasmid within

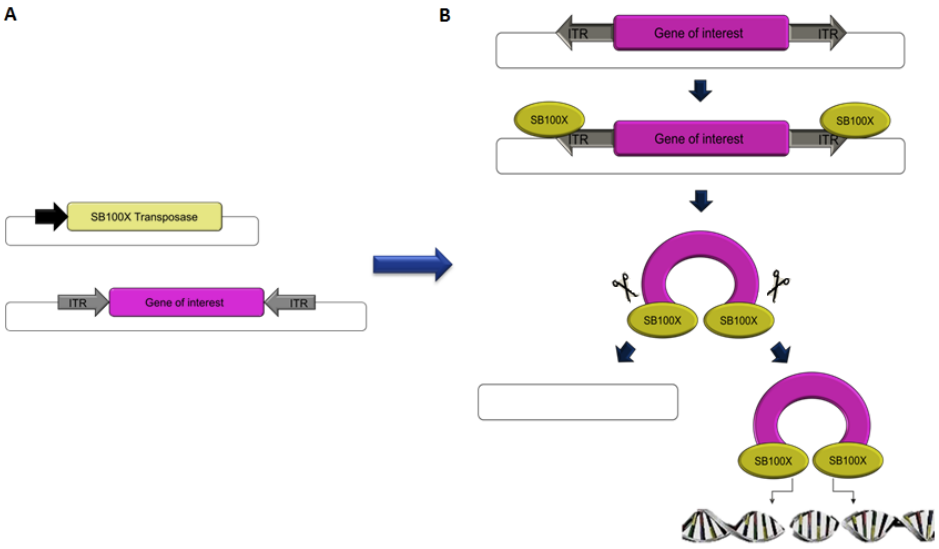

Figure 2.3: *SBTS two-vectors platform. A)The transgene of interest is cloned into the transposon plasmid within the inverted terminal repeats (ITRs), which are recognized by the hyperactive transposase SB100X, encoded by a plasmid provided in trans. B)SB100X hyperactive transposase binds to the ITRs and cuts the transposon out of the plasmid. The transgene can subsequently be integrated in a new site into the recipient DNA.*

the ITRs recognized by the transposase SB100X, encoded by a plasmid provided in trans (fig.2.3A). The SB100X is an hyperactive transposase recently derived from a large-scale genetic screen in mammalian cells, with ∼100-fold enhancement in efficiency when compared to the first-generation transposase [48, 67]. SB100X-mediated transposition is a cut-and-paste mechanism (fig.2.3B), during which the transposable element located within the ITRs is excised from its original location and the two ends of the elements are paired and held together by transposase molecules to form a synaptic complex [123]. Finally the transposable element is reintegrate at a target site, which for SB100X is preferentially a TA dinucleotide that is duplicated upon transposition and flanks the integrated element [49, 124]. The gene transfer mediated by SB100X transposase has been proven to be highly efficient in cellular and mouse models, and to permit the long–lasting expression of human coagulation factor FIX (hFIX) *in–vivo* [62].

2.6 Aim of the study

The aim of this project was to exploit the Sleeping Beauty transposon system for the development of cellular and mouse models of HB caused by F9 splicing mutations, and then to assess the efficacy of the ExSpeU1s-mediated correction approach in the chromatin expression context instead of a transient episomal system. Indeed, in contrast to other approaches, the SBTS non-viral system enables the transgene to be integrated into the genome, which represents the preferred physiologic situation for subsequent splicing studies. As a model, we chose the splicing mutation FIXex5-2C, a single point change falling in the exon 5 donor splice site and causing exon skipping [37]. Data obtained by exploiting minigene assays revealed that this variant can be rescued by a modified U1snRNA designed to have full complementarity to the mutant 5'ss named U1FIX-2C (fig.2.4 B), as well as by some modified U1 snRNA complementary to downstream intronic sequences, named exon-specific U1 (ExSpeU1) and able to redirect the spliceosome machinery on the mutated donor splice site in order to restore the exon 5 definition. Here we consider the ExSpeU1 sh9, one of the most effective in rescuing the FIXex5-2C variant and base pairing at 9 bp downstream of the 5'ss of F9 exon 5 (fig.2.4C). The usage of the ExSpeU1, like U1sh9, permits

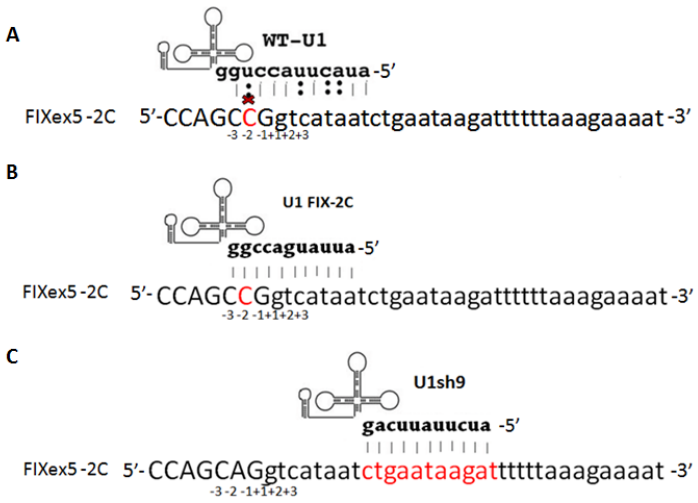

Figure 2.4: *Schematic representation of binding sites of the U1* wt*, U1FIX-2C and U1sh9 in F9 exon 5 context. Donor site and downstream intronic region are shown and exonic and intronic sequences are in upper and lower case, respectively. A) Binding of the U1* wt *on the exon 5 donor splice site in presence of the FIXex5-2C variant, marked in red. B) Binding of the modified U1 FIX-2C, designed to perfectly match the exon 5 mutated 5'ss. C) Binding of the ExSpeU1 sh9, designed to base pair at 9 bp downstream of the 5'ss of F9 exon 5.*

to target non-conserved intronic sequences downstream of the 5'ss, in order to improve specificity and reduce potential off-target effects on other pre-mRNAs. Moreover, it permits the exploitation of a unique ExSpeU1 to correct a panel of different exon-skipping mutations, thus extending its applicability to a cohort of patients sharing the same pathogenic mechanism (i.e. skipping of a given exon)

[37]. However, the efficacy of the ExSpeU1s have been demonstrated in cellular models only and in an episomal expression system using plasmids, and not in experimental system in which the splicing defective transgene in integrated into the genome. The final goal of our pioneer approach with the SB is to integrate the *wt* or mutated human splicing-competent FIX cassettes (SC-FIX) into the genome of human and mouse hepatocytes to induce the stable expression of the FIX pre-mRNA, in order to validate the efficacy of the ExSpeU1 sh9 in both the *in–vitro* and emphin–vivo models.

2.7 Materials and Methods

IN-VITRO STUDY

2.7.1 Vectors and cloning of the transposon plasmids

All the vectors used for the creation of the cellular models are composed of the expression cassettes cloned into a commercial pT2-SVNeo plasmid (available from Addgene) containing the ITR sequences recognized by the transposase SB100X. The transposase enzyme is codified by a commercial plasmid, pCMV(CAT)T7-SB100 (abbreviated pSB100X, available from Addgene).

- pT2 SChFIX *wt* and pT2 SChFIX -2C were obtained by PCR amplification of the splicing-competent hFIX cassette from the plasmid pBSK-FIX IVS 4-5 and pBSK-FIX IVS5-2C [37], by using Pfu DNA polymerase (Thermo Scientific) and with primers containing XmaI recognition site at 5'ends: F9 XmaI F,

  5'-AAACCCGGGACCACTTTCACAATCTGCTAGC-3'

  and F9 XmaI R

  5'-AAACCC GGGTCCTAACACATTATATTCCTCTAGTGG-3'

  The amplification products were subsequently cloned into a commercial pT2-SVNeo (Promega) through XmaI sites.

- pT2-eGFP plasmid was obtained by PCR amplification of the eGFP coding-sequence from a previous pPKCβ–eGFP plasmid, by using Pfu DNA polymerase (Thermo Scientific) and with primers containing XmaI recognition site at 5'ends: EGFP F,

  5'-AAA CCCGGGATGGTGAGCAAGGGCGAGGAG-3'

  and EGFP R

  5'-AAACCCGGGTTACTT GTACAGCTCGTCCATGC-3'

  This region was subsequently cloned into a commercial pT2-SVNeo (Promega) through XmaI sites.

2.7.2 Cell culture and generation of Hek293 stable clones

The human embryonic kidney Hek293 cell lines were maintained under 37°C, 5% CO2 using Dulbecco's modified Eagle's Medium supplemented with 10% FBS, 2mM GlutaMAX (Invitrogen), 100U/ml penicillin and 100 μg/ml streptomycin. Hek293 cells were seeded into T75 flasks 3-4 day before transfection at densities of 2 $\times 10^6$ cells/flask. For the generation of Hek293 stable clones cells have been co-transfected using Lipofectamine 2000 (Invitrogen) with pT2-splicing competent hFIX variants and pT2-eGFP at a molar ratio of 5:1, in combination with the transposase expressing-plasmid pCMV(CAT)T7-SB100 in ratio 1/20 of the total DNA (25 μg). Ten days post-transfection, the GFP-positive cells have been isolated by FACS analysis through a BD Facs Aria II sorter (BD Biosciences) and expanded in a T25 flask. Polyclonal cells from the selection step were then plated at a density of 10cells/ml on 148cm^2 petri dishes for stable clones isolation. When the size of the cell colonies formed on the petri dish reached about 2-3 mm in diameter (about two weeks after plating), a number of eGFP expressing clones were picked for each construct by using cloning cylinders in which 0.2 ml of the 5% trypsin-PBS solution was added. Single colonies were transferred to a 24-well and expanded for subsequent analyses.

2.7.3 Evaluation of the hFIX stable expression

The hFIX expression in Hek293 eGFP positive clones has been evaluated at the RNA and secreted protein level. The total RNA was extracted from cells at confluency in a 6-well plate (around 1.2 $\times 10^6$ cells) with Trizol Reagent (Invitrogen) and treated with DNase (Ambion-Life Technologies). The cDNA was generated by RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) according to the manufacturer's recommended protocol. F9 transcript was detected by PCR with primer hFIX Ex4F (5'-ATTCCTATGAATGTTGGTGTCCCT-3') and hFIX Ex6R (5'-GGGTGCTTTGAGTGATG TTATCCAA-3'). The conditions used for the PCR were 94°C for 5 min for the initial denaturation, 94°C for 30s, 56°C for 30s, 72°C for 30s for 30 cycles and 72°C for 10 min for the final extension. The evaluation of the protein levels was performed by a Factor IX Antigen Kit (Affinity Biologicals) on the culture medium samples according to the manufacturer's recommended protocol.

2.7.4 Evaluation of the gene copy number in Hek293 stable clones by Southern blot

The copy number of the hFIX transgene in Hek293 stable clones was examined by Southern blot analysis. Briefly, genomic DNA was extracted from the cells by Gentra Puregene Cell Kit (Qiagen), and digested over night by restriction enzyme XbaI. After electrophoresis on agarose gel 0.7% at 35V 4-5 h the digested DNA was transferred to a Nylon membrane through iBlot Dry blotting System (Invitrogen). The membrane was subsequently denatured in a 1.5M NaCl/0.5M NaOH solution and maintained in a hybridation buffer to

block non-specific sites. The probe was made by amplification of a 390 bp sequence of the hFIX transgene with AccuPrime Taq polymerase (Life Technologies) and with primer (Probe F (5'-TGCAGCGCGTGAACATGATC-3') and R (5'-CTAATTCACAGTTCTTTCCTTCAA-3'). The denatured DNA product was labeled with alkaline phosphatase trough Amersham AlkPhos Direct Labeling and Detection Systems (GE Healthcare) according to the manufacturer's instructions. The hybridization with the membrane was carried out overnight at 37°C. The day after membrane was washed and hybridized blots were detected by adding ECF substrate and by scanning at Storm Imager 840 (Molecular Dynamics) in chemifluorescence mode, 100microns, 650V.

2.7.5 Transfection with U1snRNA sh9

Hek293 cellular clones were seeded into 6-well plates the day before transfection at densities of 0.3 $\times 10^6$ cells/well. After 24h from the seeding cells were transfected with 2μg of the plasmid expressing modified U1sh9. Approximately 48h after transfection the total RNA was extracted with Trizol Reagent (Invitrogen) and treated with DNase (Ambion-Life Technologies). The cDNA was generated by RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) according to the manufacturer's recommended protocol. F9 transcript was detected by PCR with primer hFIX Ex4F and hFIX Ex6R. The evaluation of the protein levels was performed by a Factor IX Antigen Kit (Affinity Biologicals) on the culture medium samples according to the manufacturer's recommended protocol.

2.7.6 AAV-expressing U1sh9

The recombinant AAV8-U1sh9 plasmid expressing the modified snRNA U1sh9 was created by cloning the U1 expressing-cassette from a previous pGem-U1sh9 into a pSMD2-ApoE-hAAT backbone containing the AAV8 ITRs sequences for the subsequent packaging into virion. After restriction digestion the insert and the backbone plasmid had no overhanging bases at their termini and thus they were treated with a T4 DNA polymerase (New England Biolabs) in order to perform a blunt-end cloning. The resulting clones were verified by the sequence analysis. The virus preparation was carried out in Genethon institute thanks to the stuff of the Immunology and Liver Gene Transfer section (PI: Federico Mingozzi), which performed the plasmid triple transfection in Hek293 E4 cells, the harvest of the cells, the sonication, the PEG precipitation and the rAAV CsCl purification and dialysis. The rAAV titration to calculate the number of copies per ml (vg/ml) was performed through qPCR analysis.

2.7.7 Transduction with AAV-U1sh9 in Hek293 stable clones

Hek293 cellular clones were seeded into 12-well plates the day before transduction at densities of 0.2 $\times 10^6$ cells/well and maintained in medium DMEM+10% FBS containing Ponasteron A 1μg/ml. After 24 hours from the initial seeding AAV

vectors (AAV-U1sh9 and a negative control represented by an AAV-luciferase) were added on the cells at different MOI (10^3, 10^4, 2×10^4) in a 25ul DMEM-solution. Cells in each well were incubated at 37°C overnight with the addition of 500ul DMEM+2%FBS. The day after total RNA was extracted from the cells and the presence of the hFIX transcript was assessed trough RT-PCR as previously described.

IN–VIVO STUDY

2.7.8 Generation of vectors

- pT2-ApoE pAAT-SChFIX *wt* and pT2-ApoE pAAT-SChFIX -2C were created by blunt-end subcloning the SC-hFIX cassettes, *wt* and mutated, from the previously described pT2 SChFIX *wt* and pT2 SChFIX -2C to a pAAV–hFIX backbone containing the human apolipoprotein E (ApoE) control region followed by the liver specific α1-antitrypsin promoter (hAAT) and the bovine growth hormone polyadenylation (BGH-pA) signal. The resulting "ApoE-hAAT promoter–SChFIX-BGH polyA" cassettes were subsequently obtained by restriction digestion from this plasmid and cloned through blunt–end junctions between the two ITRs of the pT2-SVNeo backbone. The FIX–Padua arginine–to–leucine mutation at residue 338 of FIX was introduced in both hFIX *wt* and -2C through site–specific mutagenesis by using QuickChange II XL Site–Directed Mutagenesis Kit (Agilent Technologies) with the primers R338L F:

  5'-GAGCCACATGTCTTC**T**ATCTACAAAGTTCAC-3'

  and R338L R

  5'-GTGAACTTTGTAG**A**TAGAAGACATGTGGCTC-3'

  All the resulting plasmids were verified by the sequence analysis.

- pT2 Empty was created by restriction digestion of the pAAT–SChFIX *wt* with ClaI and BstBI enzymes in order to eliminate the SChFIX cassette between the two ITRs. The backbone was subsequently closed through ligation of the compatible ends generated by ClaI and BstBI enzymes, thus creating the pT2 Empty negative control containing the ITRs without the hFIX expression cassette.

- pT2 cDNA FIX was generated by cloning the ApoE–hAAT–cDNA hFIX–BGHpA cassette from the pAAV–hFIX plasmid into the ITRs for the SB transposase. The cassette was obtained by restriction digestion of the pAAV–hFIX with MscI enzyme and blunt–cloned into the pT2–SVNeo backbone digested with HindIII restriction enzyme.

2.7.9 Animal procedures

Animal protocols were approved by the Genethon ethical committee and conducted by certified operators under agreement number CE12–037. Eight–weeks–old mice C57BL/6 mice were used for *in–vivo* delivery of plasmids by hydrodynamic injection of DNA (2ml phosphate buffer saline–PBS) or AAV vectors (0.2ml in PBS/5% sorbitol) via tail vein injection. While transposon and transposase–encoding vectors were co–injected at day 0, the AAV vectors have been administrated at a later stage, with the AAV8–U1sh9 and the AAV8–Luciferase injected 70 days after the first injection. Blood samples were collected from the retro-orbital plexus at day 0 and every two weeks after first injection.

2.7.10 Measurement of hFIX antigen and mRNA

Human FIX antigen levels in mouse plasma were evaluated by Factor IX Antigen Kit (Affinity Biologicals) on the mice blood samples according to the manufacturer's recommended protocol. To validate and optimize the assay, a standard curve was created by adding know amounts of human purified FIX. The sensitivity threshold of the assay was established at 1ng/ml of hFIX. Depending on the expected concentrations, mouse plasma samples were diluted from 1:5 to 1:120. To extract nucleic acids from mouse samples, entire livers were homogenized by using T10 Ultra Turrax Homogenizer (Hielscher-Ultrasound Technologies) in PBS-solution. The muscle samples were homogenized in PBS-solution by using MP FastPrep-24 Tissue and Cell Homogenizer (MP Biomedicals), for 60seconds at 6.5m/s. To detect hFIX transcripts, total RNA was isolated by using Trizol Reagent (Invitrogen) and treated with DNase (Ambion-Life Technologies). The cDNA was generated by RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) according to the manufacturer's recommended protocol. F9 transcript was detected by PCR with primer hFIX Ex4F and hFIX Ex6R as previously described.

2.7.11 Assessment of the transgene copy number

To evaluate the transgene copy number the genomic DNA was extracted from homogenate samples by the MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche Life Science), according to the manufacturer's recommended protocol. Samples were subsequently processed through the MagNA Pure System (Roche) in 96-well to obtain the gDNA mouse samples. The copy number of the hFIX transgene that had integrated with the genome of the transgenic animals was analysed by qPCR with SYBR Green Dye (Bio-Rad). The primers hAAT F

5'-GGCGGGCGACTCAGATC-3'

and hAAT R

5'-GGGAG GCTGCTGGTGAATATT-3'

were used to amplify 94 bp of the human AAT promoter, while the primers Ttn F

5'-AAAACGAGCGGTGACATGAGC-3'

and Ttn R

5'-TTC AGTCATGCTAGCGCTCC-3'

were used to amplify 98bp of the mouse TTN gene. The copy number of the transgene in the examined samples was established by plotting the CT values obtained on a standard curve, which had been set by the usage of the CT values for the successive standard dilutions of the plasmid harboring the transgene with known copy number. The concentration of the plasmid ranged from 5.0 initial copies to 5×10^6 final copies. The copy number determined by transgene (hAAT) specific real-time PCR assay was normalized to the level of one copy control TTN gene to calculate the final copy number relative to 1000 cells.

2.8 Results

2.8.1 Generation of stable Hek293 clones expressing human FIX

The creation of an *in–vitro* model of Haemophilia B caused by splicing mutation was performed by exploiting the SBTS to obtain stable cell clones expressing splicing-competent hFIX cassettes (SChFIX), either *wt* (SChFIX wt) or carrying the splicing variant FIXex5-2C (SChFIX-2C) in the exon 5 donor splice site [37]. Each splicing-competent construct is based on FIX cDNA region from exon 1 to 4, exon 5 flanked by a splicing-functional region of introns 4 and 5, and the remaining cDNA from exon 6 to 8. The transposon plasmids are constituted by the expression-cassettes included between the ITRs recognized by the hyperactive transposase SB100X for the integration into the genome and placed under the control of the SV40 strong promoter for the expression of the transgene in cultured cells (fig.2.5). These vectors have been validated through transient transfection experiments in Hek293 cells and subsequent RNA analysis and protein assays (not shown). An additional transposon plasmid containing the coding sequence for the enhanced green fluorescence protein (eGFP) was used as a marker to select the transfected cells. The transposase encoding plasmid pSB100X is provided in trans (fig.2.5).

Preliminary experiments to test the transposase SB100X-mediated transfer of the SChFIX transgene into the genome were performed in Hek293 cell line. Cells have been co-transfected with pT2-SChFIX variants together with pT2-eGFP plasmid at specific molar ratio (5:1), in order to assess the transfection efficiency and to subsequently select those cells carrying the transgene inserted into the genome. The plasmid pSB100X expressing the transposase enzyme was co-delivered in a ratio 1:20 of the total transfecting-DNA. Ten days post-transfection (to reduce confounding effects of cells carrying plasmids at episomal level) cells were selected and sorted for the eGFP expression by FACS analysis, and part of

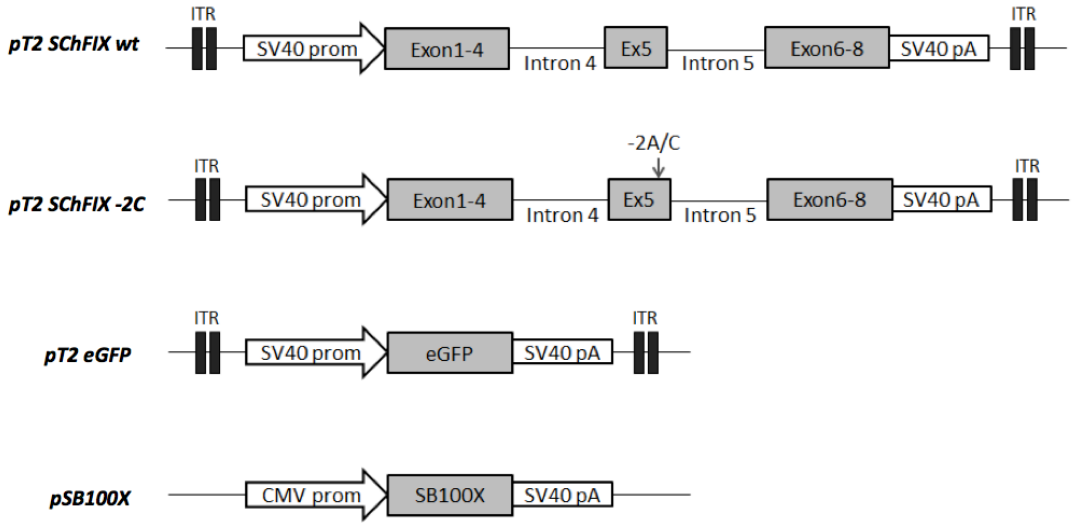

Figure 2.5: *Scheme representing the plasmids used for the generation of Hek293 stable clones expressing the splicing-competent human FIX expression cassettes. The pT2 plasmids harbor the SChFIX* wt*, -2C and eGFP cassettes between the two ITRs recognized by the transposase enzyme for the integration into the genome. All the expression cassettes are under the control of the SV40 strong promoter. The transposase is expressed from the pSB100X plasmid carrying CMV viral promoter.*

the recovered eGFP-positive cells were limiting diluted in Petri dishes to obtain single stable clones expressing eGFP and, probably, hFIX. A negative control was set up by transfecting Hek293 with the transposon plasmids without the SB100X encoding-vector. In this case, ten days post-transfection cells were observed by fluorescence microscopy and, in contrast of those treated with SB100X, no appreciable eGFP signal was detectable. This was further confirmed by FACS analysis (fig. 2.6).

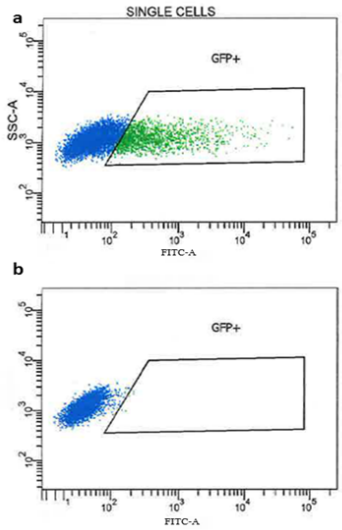

Figure 2.6: *Representative fluorescence-activated cell sorting (FACS) plots illustrating isolation of eGFP-positive Hek293 gated cells. a) Hek293 transfected with pT2 SChFIX* wt *or -2C in combination with pT2-eGFP and the pSB100X. b) Hek293 transfected with pT2 SChFIX* wt *or -2C in combination with pT2-eGFP without the pSB100X.*

Approximately two weeks after sorting the single clones were isolated and propagated for further analysis. The total RNA was extracted from a number of eGFP-positive Hek293 clones, followed by a specific RT-PCR to detect hFIX transcript. Noticeably, the expression of hFIX variants was clearly observable in the majority of the clones after electrophoresis separation of PCR products. On the contrary, any transcript was detected in the un-transfected cells and in cells transfected with the transposon plasmids but without the transposase enzyme (fig.2.7A). As expected from the poor definition of F9 exon 5 [37], even the wild-type construct produced two transcripts, the full-length one and the aberrant one showing exon 5 skipping (fig.2.7A lane pT2 SChFIX *wt*). On the other hand, cell clones expressing the mutated cassette from pT2 SChFIX -2C showed almost exclusively the presence of the aberrant form of the hFIX mRNA lacking exon 5 (fig.2.7A). The quantification of the exon inclusion was performed by densitometric analysis using ImageJ software and it is established $\sim$ 93% of the total transcript for the pT2 SChFIX *wt* and $\sim$ 4% for the pT2 SChFIX -2C (fig.2.7B). The production of secreted hFIX protein in clones expressing the

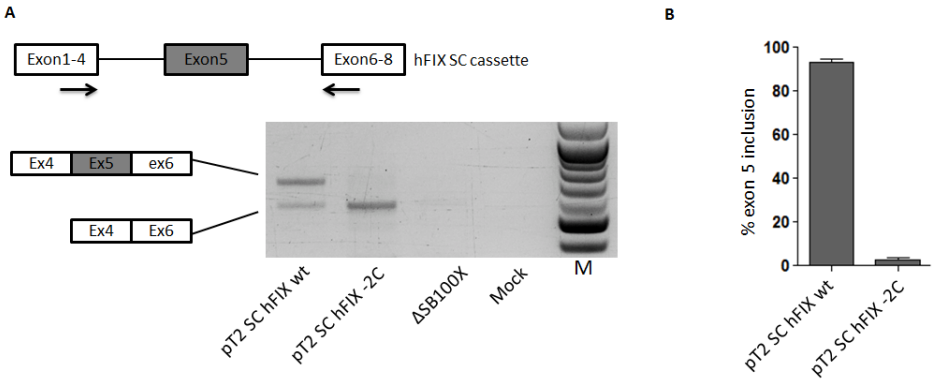

Figure 2.7: *Determination of hFIX expression in eGFP positive Hek293 stable clones. A) Evaluation of hFIX mRNA through RT-PCR on total RNA extracted from cell clones. The* wt *and the aberrant transcripts were resolved through gel electrophoresis on agarose gel 2%. Negative controls are represented by cells transfected with the transposon plasmids without the transposase enzyme (ΔSB100X) and by cells transfected with the lipofectamine vehicle alone (mock). B) Histogram reporting the quantification derived from densitometric analysis of the percentage of exon 5 inclusion. Results are expressed as mean ±SD derived from three independent experiments.*

SChFIX *wt* cassette has been evaluated by anti hFIX-specific ELISA on culture medium samples. Protein levels reflected mRNA levels assessed by RT-PCR on the same samples, varying from ∼3ng/ml to ∼50ng/ml in the positive clones, indicated as (+) in the figure 2.8a. It is also reported an example of an hFIX negative clone, indicated as (-), in which the transcripts, as well as the protein, was not detectable (fig.2.8).

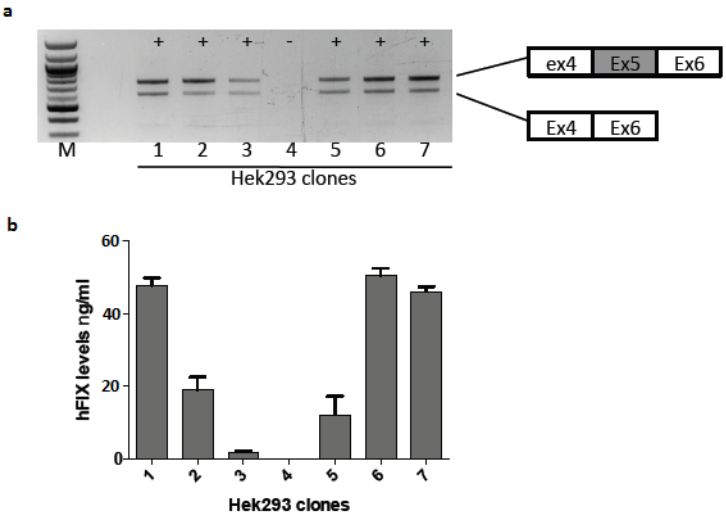

Figure 2.8: *Analysis in Hek293 SChFIX* wt *clones. a) Evaluation of hFIX mRNA through RT-PCR on total RNA extracted from cell clones. The* wt *and the aberrant transcripts were resolved through gel electrophoresis on agarose gel 2%. b) Histogram reporting hFIX protein levels in culture medium assessed by ELISA. Results are expressed as mean ±SD derived from three independent experiments.*

2.8.2 SB100X activity *in–vitro*: assessment of the gene copy number

The transgene copy number was evaluated in a number of Hek293 hFIX stable clones, *wt* and -2C, by performing Southern Blot analysis on the gDNA samples, which were extracted from cells and suitably digested with XbaI restriction enzyme. This presents a unique cutting site into the SChFIX cassettes, thus permitting the evaluation of the integration pattern profile for each clone analyzed. A specific hybridization probe was designed to base pair a 450bp region of the integrated cassettes spanning from hFIX coding sequence exon1 to exon 4 (fig.2.9). Both DNA from cells expressing the *wt* and the mutated cassette were run in

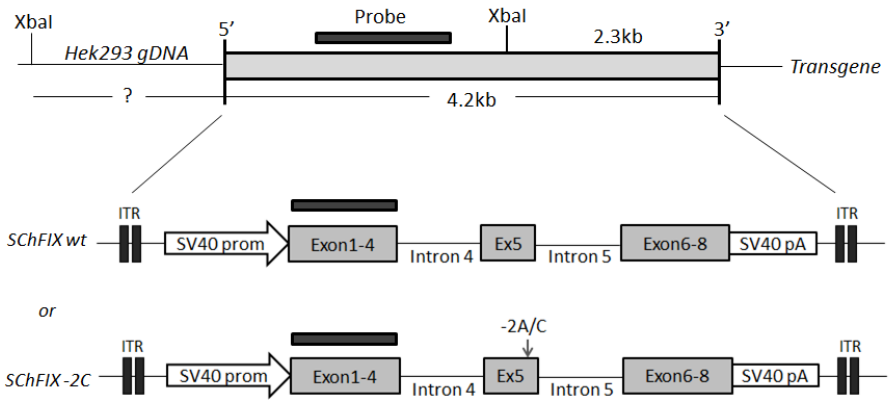

Figure 2.9: *Scheme representing the structure of the transgenes detected by Southern Blot analysis. Both the cassettes integrated by the SB100X transposase, the SChFIX* wt *and -2C, are 4.2 kb in length, and present an unique recognition site for XbaI restriction enzyme. The detection probe labeled with alkaline phosphatase is designed to base pair 450 bp spanning from hFIX cDNA exon1 to exon 4 region.*

the same agarose gel and blotted to the same membrane as the positive control represented by the linearized pT2 SChFIX plasmid (Ctrl+). The negative control (Ctrl -) is represented by genomic DNA extracted from untransfected Hek293 cells (fig.2.10). The southern blot transgene copy number analysis was performed on 4

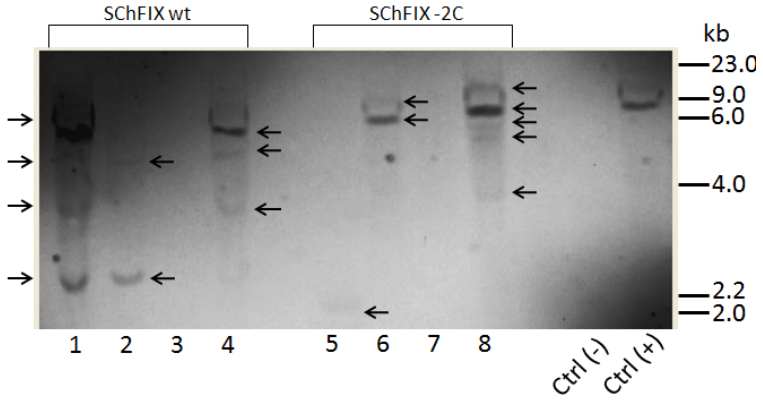

Figure 2.10: *Determination of the copy number of the splicing-competent hFIX transgene in Hek293 stable clones by genomic Southern Blotting. The genomic DNA (10µg) of Hek293 SChFIX* wt *and -2C was digested with XbaI restriction enzyme and subjected to electrophoresis on a 0.8% agarose gel. The DNA bands were then transferred to a Nylon membrane and hybridized with an alkaline phosphatase-labeled 450 bp probe generated from pT2 SChFIX* wt *plasmid by PCR. Both DNA from cells expressing the* wt *(lanes 1-4) and the mutated (lanes 5-8) cassettes were run in the same agarose gel and blotted to the same membrane as the positive control represented by the linearized pT2 SChFIX plasmid (lane Ctrl+). The position of molecular weight marker is indicated.*

clones for each type of transgene (SChFIX *wt* and -2C). In each group a negative control was also loaded (lanes 3 and 7), meaning these are Hek293 eGFP positive clones non-expressing hFIX. In these lanes there were no detectable bands, as expected. Vice versa, in the group of samples expressing SChFIX *wt* (lanes 1-4) some hybridization bands were detected, from a minimum of 2 (lane 2) to a maximum of 4 (lane 1), thus indicating a multiple copies integration profile. In the group of samples expressing SChFIX -2C (lanes 5-8) a single copy integration occurs in sample run on lane 5, because only a band of ∼2kb was detected. In the other samples integration number varies from 2 to 5 (lanes 6 and 8).

2.9 U1sh9 mediated correction: hFIX mRNA and protein rescue

To evaluate the capability of the modified snRNA U1sh9 to rescue hFIX mRNA splicing in the genomic context, Hek293 cells stable expressing the SChFIX -2C cassette (clones n.1 and 4 chosen as examples) were transiently transfected with the plasmid encoding the modified U1, pU1-sh9. Moreover, cells were treated with a modified U1 called U1 FIX*wt*, designed to perfectly match with exon 5 donor

splice site [37]. Transfection of hFIX variant expressing cells with either the U1 FIX*wt* or the U1sh9 resulted in appreciable rescue of correct FIX transcripts, from 0% to ∼20-45%, as revealed from mRNA RT-PCR (fig.2.11A) and densitometric analysis (fig.2.11B).

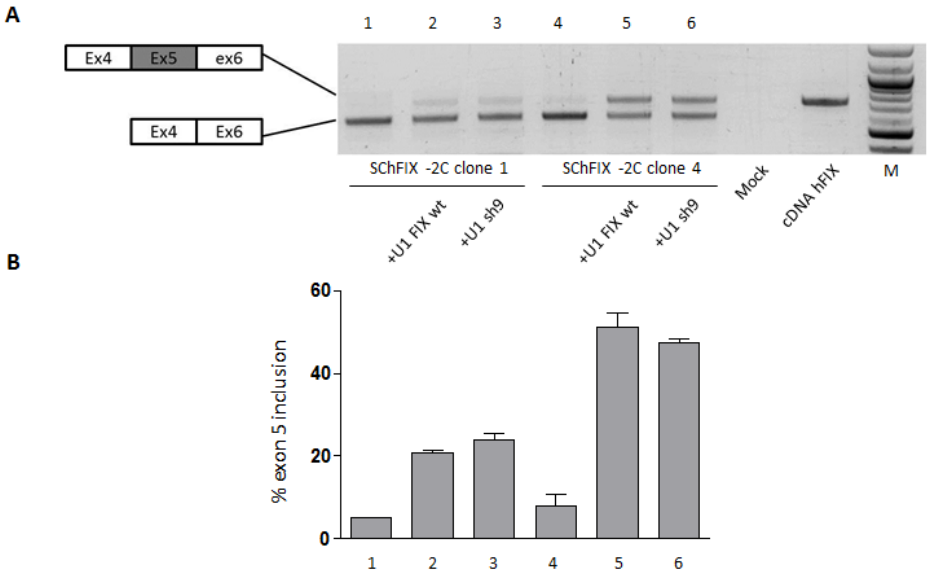

Figure 2.11: *Treatment of Hek293 SChFIX -2C clones with modified U1sh9. A) Evaluation of hFIX mRNA through RT-PCR on total RNA extracted from cell clones. The* wt *and the aberrant transcripts were resolved through gel electrophoresis on agarose gel 2%. Hek293SChFIX -2C clones 1 and 4 were treated with both U1FIX*wt *and U1sh9. Mock is represented by untransfected Hek293 cells. A positive control is represented by hFIX* wt *transcript. b) Histogram reporting the quantification of the percentage of exon 5 inclusion. Results are expressed as mean ±SD derived from three independent experiments.*

Moreover, to test the U1sh9-mediated rescue at the protein level an hFIX-specific ELISA was performed on media from Hek293 clones expressing the SChFIX -2C cassette, treated or not with the U1sh9. In the untreated cells antigen levels were very low, but after transfection with pU1sh9 a remarkable increase of secreted hFIX levels (fig.2.12A) in the medium was observable (from ∼3 to ∼45 ng/ml). This increase was paralleled by the relative splicing pattern correction (fig.2.12B) in which the exon 5 inclusion is stimulated up to ∼45%, as revealed by densitometric analysis.

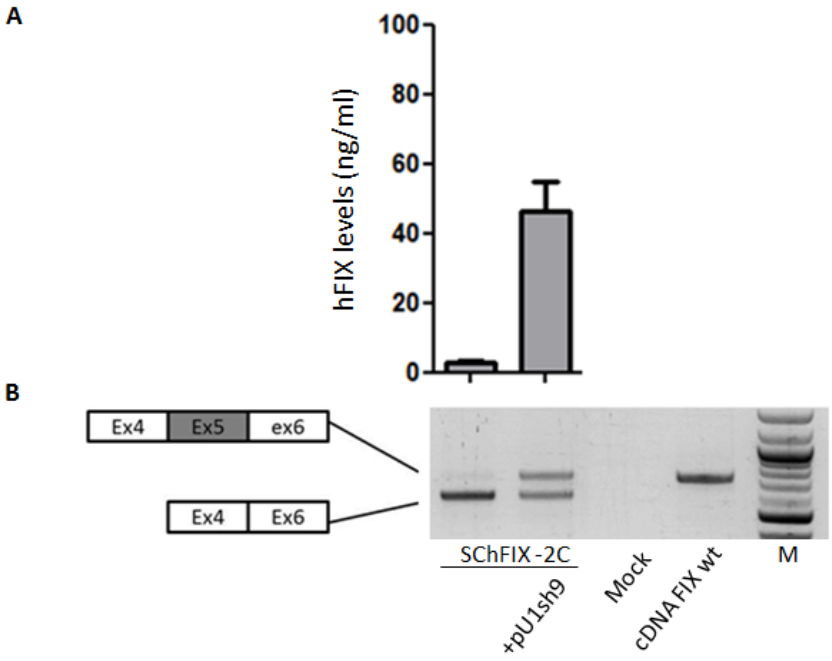

Figure 2.12: *Treatment of Hek293 cells expressing SChFIX -2C transgene with modified U1sh9. A) Evaluation of hFIX mRNA through RT-PCR on total RNA extracted from cells. The* wt *and the aberrant transcripts were resolved through gel electrophoresis on agarose gel 2%. Mock is represented by cells transfected with the lipofectamine vehicle alone. A positive control is represented by hFIX* wt *transcript. B) Histogram reporting the quantification of hFIX protein levels through ELISA assay conducted on the cell medium samples. Results are expressed as mean ±SD derived from three independent experiments.*

2.9.1 Cell transduction with an AAV8-vector expressing the modified U1sh9

To test the modified ExSpe U1 *in–vitro* efficacy by delivery through a viral vector, Hek293 FIX-variant expressing cells were transduced with an AAV8-expressing U1sh9 under the control of the U1snRNA gene endogenous promoter (fig.2.13). The infection of stable Hek293 clones was carried out at increasing MOI (from

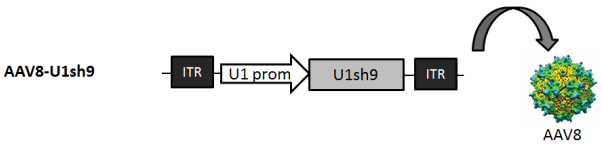

Figure 2.13: *Schematic representation of the expression cassette for the modified U1sh9 snRNA. The coding sequence for the snRNA is under the control of the endogenous promoter of the U1snRNA gene (U1 prom), and inserted between the two ITRs sequence for the packaging into an AAV8 serotype.*

10^3 to 2×10^4). The extraction of total RNA was performed after 72 hours post-infection, followed by RT-PCR to detect hFIX transcripts. In both cases, i.e. *wt*

and the mutated cassette, a dose-dependent rescue of the *wt* mRNA was obtained (fig.2.14A). Indeed, the densitometric analysis revealed an increase of the correct hFIX transcript up to ∼32% for the cells expressing SChFIX -2C and up to ∼85% for cells expressing SChFIX *wt* (fig.2.14B). As a further negative control, Hek293-

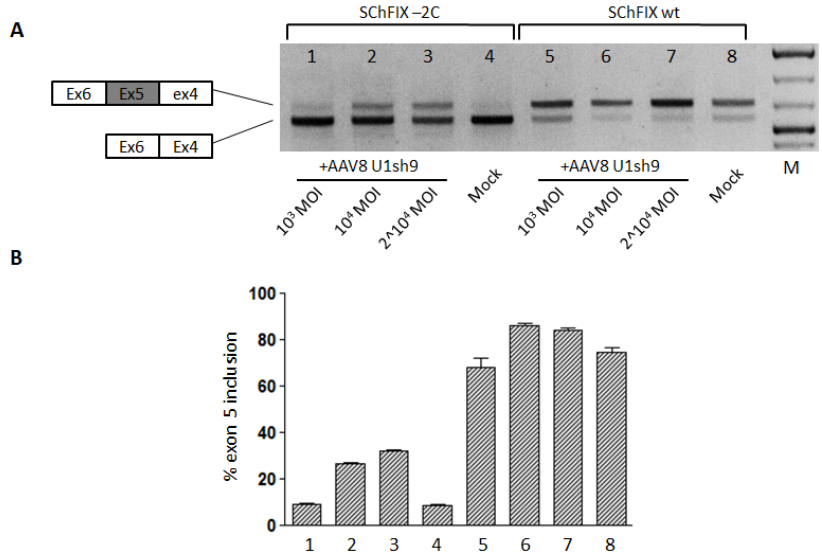

Figure 2.14: *Treatment of Hek293 SChFIX* wt *and -2C clones with AAV8-U1sh9. A) Evaluation of hFIX mRNA through RT-PCR on total RNA extracted from cell clones. The* wt *and the aberrant transcripts were resolved through gel electrophoresis on agarose gel 2%. Both Hek293SChFIX* wt *(lanes 1-4) and -2C clones (lanes 5-8) were treated with AAV8 U1sh9 at increasing MOI (from* 10^3 *to* 2×10^4*). Mock is represented by untransduced cells. b) Histogram reporting the quantification of the percentage of exon 5 inclusion. Results are expressed as mean* $\pm SD$ *derived from three independent experiments.*

variant expressing cells were transduced with an AAV8 expressing Luciferase gene (AAV8-Luc) at increasing MOI (as previously reported for AAV8-U1sh9). The extraction of total RNA was performed after 72 hours post-infection, followed by RT-PCR to detect hFIX transcripts. In this case any effect on splicing pattern was obtained in comparison with untransduced cells, as expected (fig.2.15).

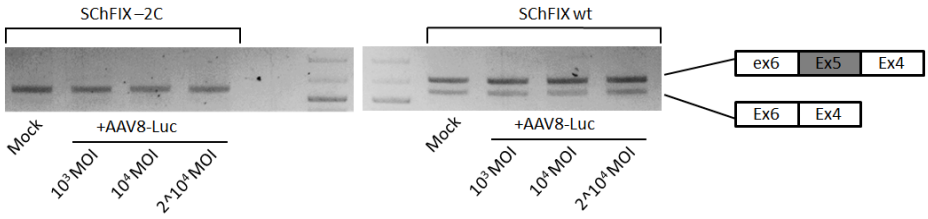

Figure 2.15: *Treatment of Hek293 SChFIX* wt *and -2C clones with AAV8-Luc. A) Evaluation of hFIX mRNA through RT-PCR on total RNA extracted from cell clones. The* wt *and the aberrant transcripts were resolved through gel electrophoresis on agarose gel 2%. Both Hek293SChFIX* wt *and -2C clones were treated with AAV8 Luc at increasing MOI (from* 10^3 *to* 2×10^4*). Mock is represented by untransduced cells.*

2.9.2 Generation and validation of the transposon plasmids for the creation of mouse models

The creation of mouse models of HB was performed, as for the cellular model, by designing a panel of trasposon vectors for the SB100X-mediated integration into the mouse genome. The transposon plasmids consisted of the expression-cassettes included into the ITRs and placed under the control of the hepato-specific human apolipoprotein E (ApoE) control region followed by the liver specific α1-antitrypsin promoter (hAAT) and the bovine gro*wt*h ormone polyadenylation (BGH-pA) signal (fig.2.16). The expression cassettes are represented by the splicing-competent hFIX *wt* and -2C, as well as by a cDNA hFIX positive control (pT2 cDNA hFIX) and a negative control (pT2 Empty) constituted of a plasmid carrying the ITRs for the transposase but lacking the expression cassette within them (fig.2.16). These vectors have been validated through transient

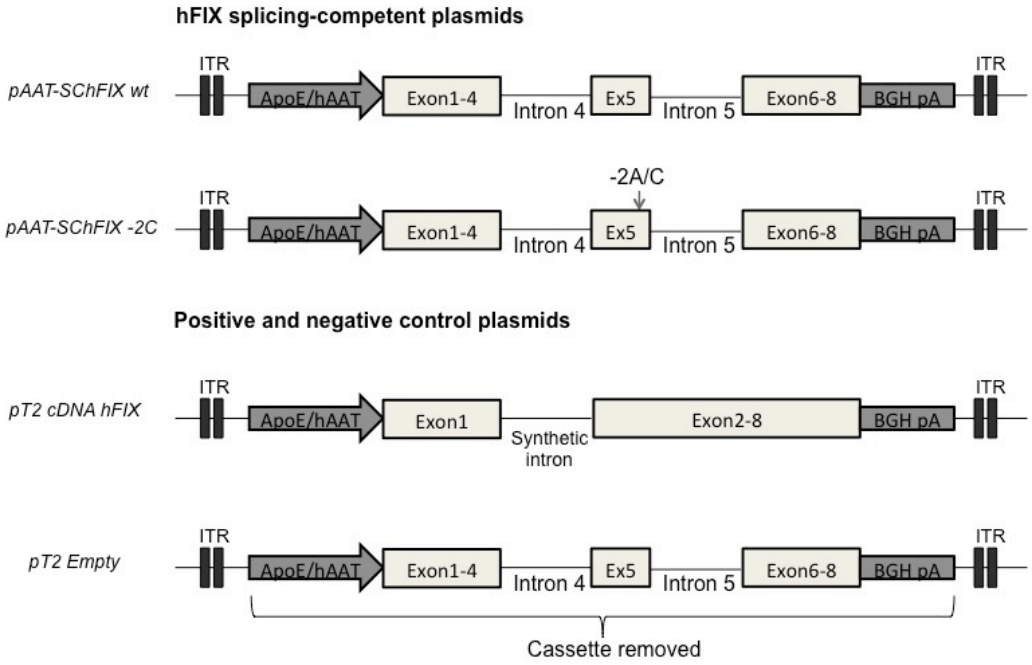

Figure 2.16: *Scheme representing the plasmids used for the generation of the mouse models expressing the splicing-competent human FIX expression cassettes. The positive control is represented by a plasmid expressing the cDNA of hFIX, carrying exon1, an optimized synthetic intron followed by the exon 2-8 remaining part of the coding sequence. The negative control is represented by an "empty" vector carrying the two ITRs without the interposed transgene. All the expression cassettes are guided by human ApoE control region followed by the liver specific hAAT promoter and the BGH-pA signal.*

transfection experiments in Huh7 human hepatocarcinoma cells and subsequent RNA analysis and protein assays. Cells were treated with 3μg of each vector, and the presence of hFIX mRNA was confirmed by specific RT-PCR (2.17A). Interestingly, Huh7 expressing the mutated cassette from pAAT-SChFIX -2C showed also the presence of the *wt* form of the hFIX mRNA, not detectable with the previous

construct (pT2-SChFIX -2C) expressed in Hek293 (see fig.2.7A). The corresponding protein levels were quantified in the medium samples by anti-hFIX ELISA assay (fig.2.17B), and range from ∼45ng/ml (pAAT-SChFIX -2C) to ∼950ng/ml (pT2 cDNA FIX). In the negative controls represented by cells treated with pT2 Empty vector and un-transfected cells (mock) no protein level was observable, as expected (2.17B).

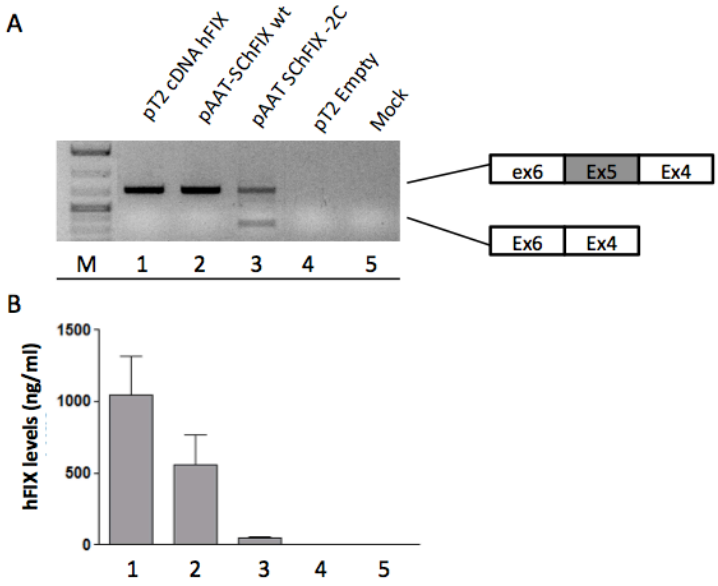

Figure 2.17: *Treatment of Huh7 cells with plasmids pT2 cDNA FIX, hFIX* wt*, -2C and pT2 Empty optimized for the generation of mouse models. A) Evaluation of hFIX mRNA through RT-PCR on total RNA extracted from the cells. The* wt *and the aberrant transcripts were resolved through gel electrophoresis on agarose gel 2%. Mock is represented by cells transfected with the lipofectamine vehicle alone. B) Histogram reporting the quantification of hFIX protein levels through ELISA assay conducted on the cell medium samples. Results are expressed as mean ±SD derived from three independent experiments.*

2.9.3 Generation of the mouse models

The pT2 transposon plasmids optimized for the liver restricted expression of hFIX were hydrodynamically injected into the tail vein of 8-weeks old C57BL/6J *wt* mice (50μg/mouse; n=10 per group) together with the pSB100X transposase-expressing vector (2μg/mouse). To assess the stable expression of the transgenes in mouse hepatocytes the plasma levels of circulating hFIX in treated mice were measured though a hFIX specific ELISA and monitored every 14 days for two months after the hydrodynamic injection. Positive controls were represented by mice treated with the cDNA hFIX expressing-cassette and showing stable protein levels between 100ng/ml (group named cDNA FIX low) and 2μg/ml (group named cDNA FIX high). Expression at low level was obtained for the SChFIX cassettes (<10ng/ml after 1 month), especially for the mutated cassette (SChFIX

-2C) with protein levels below the sensitivity limit of the ELISA at 2 months post-injection. The negative control is represented by the group co-injected with the pT2 Empty vector and pSB100X, in which protein levels are no detectable at each time point (fig.2.18).

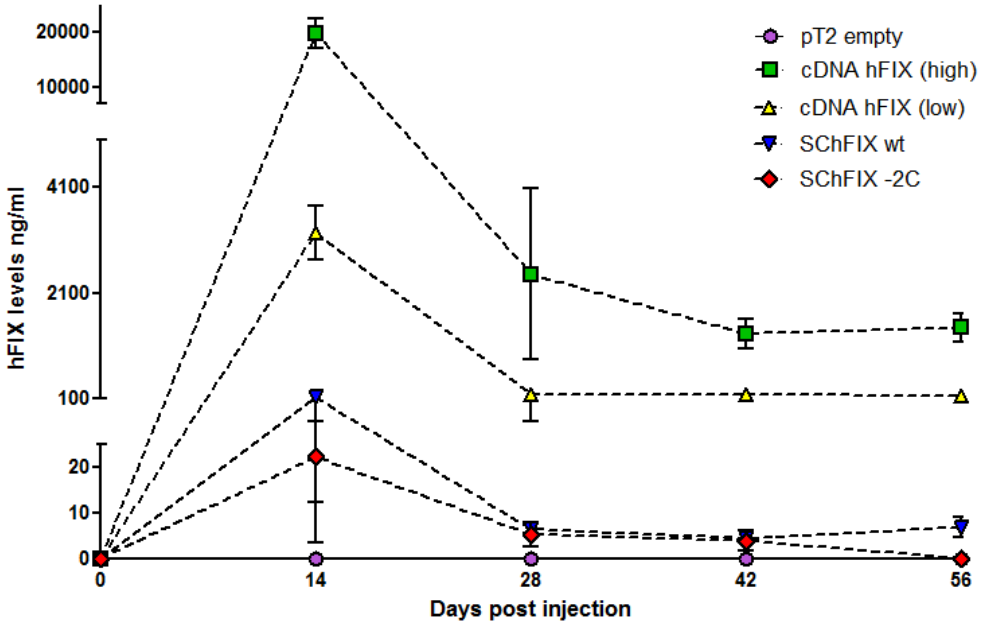

Figure 2.18: In–vivo *hepatic gene delivery using hyperactive SB100X transposase. The transposon-containing plasmids (50μg) were co-injected by hydrodynamic gene delivery with pSB100X (2μg) for the integration into the genome of the hepatocytes. Plasma hFIX concentrations were measured (n=10 mice per group) by anti-hFIX ELISA on plasma samples and analyzed in triplicate each. Mean values of hFIX concentration ± SD are shown for each time point (every 14 days post-injection).*

2.9.4 Assessment of U1sh9 efficacy *in–vivo*: RNA and protein analysis

To assess the in-vivo correction efficacy of the U1sh9 on the splicing-defective hFIX mutant we treated mice with the AAV8-U1sh9 two months after the hydrodynamic injection of plasmids. Mice previously injected with pAAT-SChFIX *wt* and pAAT-SChFIX -2C were split into two groups (5 mice each) to test the AAV8-U1sh9 or the AAV8-Luc as negative control. Each AAV vector was delivered at fixed dose of 5×10^{10} vg per mouse via tail vein injection. Mice were then monitored for circulating hFIX levels over a total period of 20 days. However, as shown in figure 2.19, the FIX antigen levels measured at day 6 and 14 days post AAV-injection did not indicate any appreciable increase. Mice were then sacrificed to evaluate the effect of the AAV8-U1sh9 delivery on hFIX mRNA splicing pattern in hepatocytes through RT-PCR analysis in liver tissue samples recovered after the sacrifice of the animals at day 90. As expected, analysis on liver

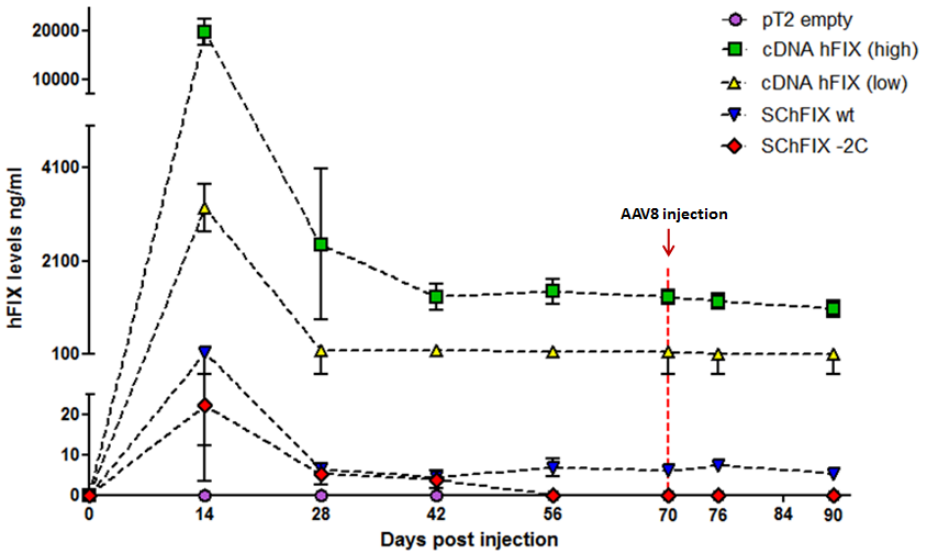

Figure 2.19: *Timeline of the hFIX levels in mice plasma assessed by ELISA assays. Plasma hFIX concentrations were measured (n=10 mice per group) by anti-hFIX ELISA on plasma samples and analyzed in triplicate each. Mean values of hFIX concentration ± SD are shown for each time point (every 14 days post-hydrodynamic injection, and after 6 and 14 days post AAV-injection). At day 70 AAV8-tail injection is marked with a red arrow.*

samples derived from mice injected with the Empty vector resulted in any detectable transcript (fig.2.20, lanes 1 to 5). On the contrary, analysis on the group cDNA hFIX *wt* revealed the presence of hFIX mRNA in all the samples, even if at different levels (fig.2.20, lanes 11 to 20). Consistently, almost all the mice injected with the SChFIX cassette *wt* express hFIX mRNA, with the exception of one animal in which the transcript is not detectable (fig.2.20, lanes 21 to 30).

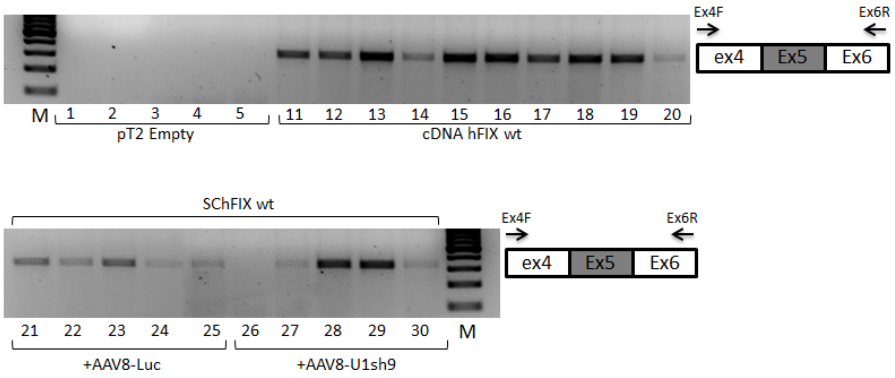

Figure 2.20: *RNA analyses on the mice liver samples. The Evaluation of hFIX mRNA through RT-PCR on total RNA extracted from the liver homogenates obtained from mice group pT2 Empty (lanes 1-5), cDNA hFIX* wt *(lanes 11-20) and SChFIX* wt *(lanes 21-30). The transcript was resolved through gel electrophoresis on agarose gel 2%.*

Most importantly, the RNA analysis on the mice group hydrodynamically injected with the mutated SChFIX cassette revealed no hFIX expression in the mice that received AAV8-Luc control vector (fig2.21, lanes 31 to 35), while 3 mice of 5 treated with AAV8-U1sh9 showed expression of full-length form of the hFIX mRNA (fig.2.21, lanes 36 to 40). Altogether these data indicated the long-lasting

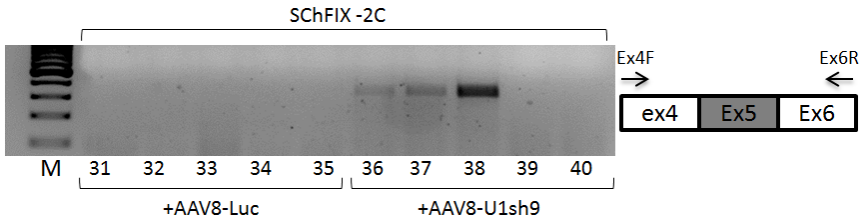

Figure 2.21: *RNA analyses on the mice liver samples. The Evaluation of hFIX mRNA through RT-PCR on total RNA extracted from the liver homogenates obtained from mice group pT2 SChFIX -2C (lanes 31-40). The transcript was resolved through gel electrophoresis on agarose gel 2%.*

expression of the mutant hFIX cassette, which points to an integration event, and the presence of the hFIX pre-mRNA that underwent splicing correction mediated by the ExSpe U1sh9, although to an extent that was not appreciable at the protein level.

As a further control to assess liver-restricted expression of the SC-hFIX transgene, we conducted the RT-PCR analysis in muscle triceps samples collected from mice in all the groups analyzed (cDNA hFIX, SChFIX *wt* and SChFIX -2C) and all of them, as for those injected with the pT2 Empty, resulted negative for the hFIX transcript (data not shown).

2.9.5 Evaluation of the integration profile: gene copy number

To unequivocally demonstrate the integration event into the mouse genome mediated by the SB100X *in–vivo* we evaluated the transgene copy number in mouse liver samples. Therefore, the genomic DNA was extracted from liver homogenate samples from all mice group treated with pT2 Empty, cDNA FIX, pAAT-SChFIX *wt* and pAAT-SChFIX -2C to perform the analysis. The number of copies of the transgene integrated into the genome was determined by a specific real time PCR designed to amplify a 94bp region of the human AAT promoter and a reference genomic DNA of known copy number. The final copy number referred to 1000 cells was determined after normalization to the level of the single-copy control mouse gene TTN and was reproducible among the all groups (means from ∼17 to 24). The group pT2 Empty negative control showed no copy number, as expected (fig.2.22).

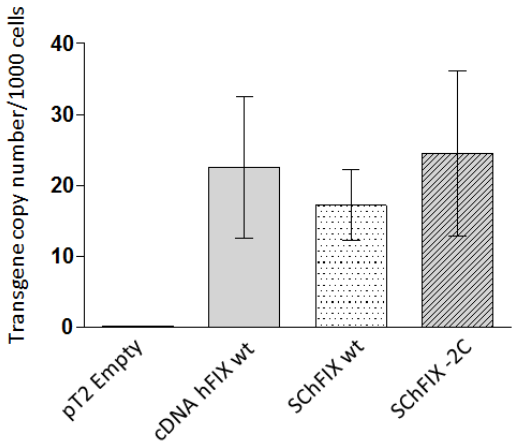

Figure 2.22: *Copy number determination of hFIX cassettes-expressing C57BL/6J mouse hepatocytes. The real time specific PCR assay to amplify the human AAT promoter, and used it together with the single–copy control gene TTN and a reference genomic DNA of known copy number. Values reported are means±SEM derived from samples analyzed in triplicate.*

2.9.6 Discussion and future plans

The study conducted has permitted to exploit the SBTS as an useful tool to obtain both cellular and animal models of Hemophilia B caused by a specific splicing mutation, the single point mutation IVS5-2C affecting F9 exon 5 donor splice site, in order to subsequently test the modified U1-snRNAs corrective approach in a genomic stable context instead of a transient episomal system, as previously described [37]. The *in–vitro* study performed on the human embryonic kidney 293 cells allowed us to evaluate hFIX transgene expression in a system in which the background signal is absent because this cells do not express FIX, as revealed from RT-PCR and ELISA data (fig.2.7). The generation of Hek293 stable clones expressing the SChFIX variants by taking advantage of the SBTS dual system has allowed to finely manage the ratio transposon-transposase (in this case 20:1) in order to avoid unwanted effects like overproduction inhibition (OPI), i.e. elevated concentration of transposase enzyme inhibit the transposition reaction [42, 92], and potential cytotoxicity due to residual integration events of transposase expression vectors that may lead to permanent transposase expression and resulting in uncontrollable transposition [39]. In our experimental setting the selection of the cells with the transgene positioned at chromosomal level was carried out by co-transfecting a second transposon vector expressing eGFP as a reporter. The usage of two transposon plasmids in each condition instead of a single transposon carrying the two cassettes in cis was aimed at maximizing the rate of transposition. Indeed, previous studies indicated that transposition efficiency decreases at approximately a logarithmic rate as a function of length [42] and that transposon around 6kb retain 50% of the maximal efficiency. Hence, we took advantage of the separate plasmids in order to work with trasposons <5kb in length. SB100X-mediated transposition occurred efficiently in Hek293 cells, as demonstrated by the FACS analysis conducted ten days after the initial transfection. As a matter of fact, just the cells receiving the transposons plasmids together with the transposase enzymes showed eGFP expression, which was stably maintained thereafter (fig.2.6), meaning that, once passed the transient episomal phase, the transgene expression is specifically due to its SB100X-mediated integration into the genome. The transposase efficiency *in–vitro* was clearly demonstrated by performing Southern Blot analysis on the Hek293 stable clones, which showed in most of the cases a multiple copies integration profile (fig.2.10). Further analyses at mRNA level demonstrated that Hek293 expressing the hFIX mutant cassette (SChFIX -2C) displayed the aberrant splicing pattern characterized by exon 5 skipping, and this reflected previous data obtained studying the molecular effect of this splicing variant through minigene assays [37]. The proper exon definition and inclusion was partially rescued by transfecting cells with the modified U1snRNA sh9 (fig.2.11A), able to redirect the spliceosome machinery on the mutated exon 5 donor splice site in order to increase the rate of exon inclusion (up to ∼45%, see fig.2.11B). This result was paralleled by increase of secreted hFIX protein in culture medium, as shown by ELISA antigen quantification (fig.2.12). Moreover, Hek293 cells expressing both the wt hFIX and mutated cassettes were transduced with an AAV-vector expressing snRNA U1sh9, at increasing MOI,

thus permitting to assess the dose-dependent rescue of the wt mRNA (fig.2.14). Interestingly, even the treatment of the cells carrying the SChFIX wt cassette showed some increase of the correct transcript, as demonstrated by densitometric analysis (fig.2.14B). The *in–vitro* study demonstrated the feasibility in the usage of SBTS in the generation of the model for a splicing mutation and, importantly, the efficacy of the U1sh9 in recovering the proper splicing pattern in the chromatin context. Altogether these data provided the ground for the creation of transposon-based mouse models. The SB gene transfer *in–vivo* was performed via hydrodynamic injection of plasmids carrying SC-hFIX variants together with the SB-100X in wild-type C57BL/6 mice. The rationale behind using wt mice was the availability of nucleic amplification and immunologic assays able to specifically detect the human FIX expression in the mouse background. The transposon vectors were specifically designed for optimizing the subsequent expression of the transgene in mouse liver by taking advantage of the ApoE control region, the hAAT promoter and a BGHpA sequence. The combination of these cis-regulatory regions has previously been shown to enhance high and stable serum hFIX expression in C57BL/6 background after injection in the liver of mice [63], thus used in this study to guarantee a tissue-specific regulation of the hFIX transgene. A lower amount of the transposase (2μg of pSB100X) compared to the transposon (50 μg) was hydrodynamically injected in each animal to avoid an excess of hyperactive SB100X that could diminish the transposition reaction, consistent with the over-production inhibition. The stable expression of hFIX was assessed by the measurement of hFIX plasma levels (ELISA) and by the evaluation of RNA expression in mice hepatocytes through RT-PCR. It is worth noting the trend of the hFIX expression levels during the entire time period of the in-vivo experiment, which is reproducible among all the mice groups injected, apart from the negative control pT2 Empty (fig.2.18). We can observe a high spike of expression at day 14 post-injection, followed by a decrease and a stabilization of the levels measured from day 28 (fig.2.18). Hence it is possible to speculate about a massive presence of the episomal component maintained for a few weeks after the hydrodynamic injection, which is probably completely lost after one month post-injection (fig.2.18). Interestingly, the different possibilities related to the integration mechanisms can be described by observing the trend in the hFIX levels for each animal. To simplify, the graph reported below (fig.2.23) represent all the hFIX measurement relative to the cDNA hFIX positive control group (n=10).

Noticeably, few mice injected with the cDNA hFIX showed low or no expression (as example n.17, marked with green symbol) at each time point. In these few cases we can speculate about an inefficient hydrodynamic procedure, which is one of the variables in these experimental conditions to be considered. Besides, some mice (as example n.13, marked with red symbol) showed very high expression in the supra-physiologic range at day 14 ($\sim$18μg/ml), which remained high and stable from day 28 for the whole duration of the experiment ($\sim$2.5 μg/ml), thus supporting good transposition efficiency. Moreover, there is also the case of an animal showing high expression level at day 14, $\sim$1μg/ml (see n.19, marked

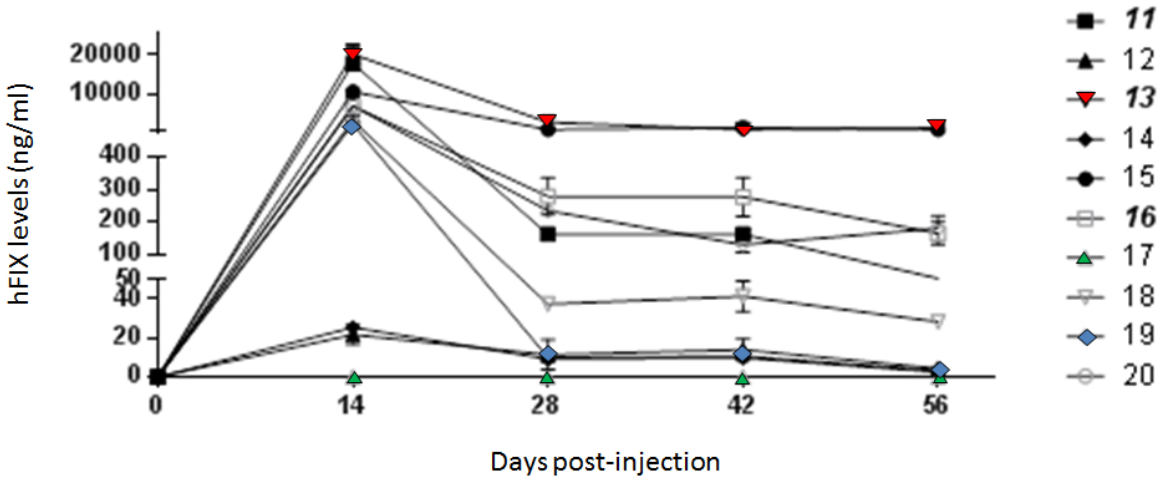

Figure 2.23: *Timeline of the hFIX levels in mice of the cohort injected with the pT2 cDNA hFIX, assessed by ELISA assays. Plasma hFIX concentrations were measured (n=10 mice per group) by anti-hFIX ELISA on plasma samples and analyzed in triplicate each. Mean values of hFIX concentration ± SD are shown for each time point (every 14 days post-hydrodynamic injection). The legend reports the number of each mouse injected.*

with blue symbol), with a dramatically decrease at day 28 (≤10 ng/ml), maintained over time. On the other hand, we observed mice treated with the same protocol that displayed with very low FIX levels (mice 19, 20), a finding that could be attributable to a very low transposition rate, or an unsuccessfully combination of the low transposition rate with some unproductive integration sites, such the delivery of the transgene in some genomic area subjected to epigenetic silencing-related regulation. As a matter of fact, data from literature have been shown that SB integration can be considered fairly random into chromosomes, with roughly 35% of the transposon insertions occurred in transcribed regions [123]. Importantly, it was possible to observe a good correlation between the gene copy number in mice hepatocytes and hFIX expression. Indeed, mice showing the higher circulating hFIX levels display also the major copy number (up to 100 copies/1000 cells) corresponding to ∼10% of the entire liver being integrated, while mice showing low hFIX levels display ∼1-4 copies/1000 cells. Overall, the mean copy number among the groups pT2 cDNA hFIX, pAAT SChFIX wt and -2 was reproducible and quantified around 20 copies/1000 cells, meaning that on average the integration rate in this experimental conditions is ∼2% of the entire liver (fig.2.22). At a constant integration rate among the groups no correspondence in hFIX levels was observable, in relationship with the nature of the transgene integrated. The cohort injected with the splicing-competent hFIX wt exhibit far lower hFIX levels in comparison with the cohort injected with the cDNA hFIX used as positive control (fig.2.18). Importantly, even if both of them produce a wt transcript, the splicing-competent cassette composed of the cDNA with a portion of the introns 4 and 5 (i.e. ∼300bp in proximity of the exon 5,

to maintain the main splicing regulatory sequences avoiding the too extensive complete introns) is much less efficient in producing a full-mature mRNA and, consequently, protein. We can argue that in this condition we are anyway losing some important in-cis regulatory regions, whereas the cDNA hFIX with a synthetic optimized intron 1 showed high expression efficiency [63, 26, 71]. The analysis on the RNA pattern in mice liver samples revealed that the treatment with the ExSPeU1 sh9 delivered by AAV8-vector after around 3 months after the first injection seems to be effective, because 3 of 5 mice expressing-hFIX mutated cassette expressed the wt hFIX RNA, while the animal in the same cohort injected with the AAV8-Luciferase control did not show any transcript, neither the aberrant form lacking exon 5 (fig.2.21). However, any appreciable change in the hFIX plasma levels were detected after the AAV8-U1sh9 tail vein injection, thus indicating that the mRNA rescue was not paralleled by an appreciable, in our experimental setting and with our ELISA system, rescue at the protein levels.

Therefore, our future objective will be the improvement of the experimental protocol in order to better assess the minigenes rescue *in–vivo*, i.e. by modulating the vector doses injected, by exploiting an optimized splicing-competent cassette to increase the basal hFIX levels and the evaluation of the eventual hepatocellular toxicity due to ExSPeU1 expression in mouse liver. To magnify the effect at the protein levels, and to better assess the impact of the ExSpeU1-mediated correction in mice, we have already inserted into the SCFIX expression cassettes the F9 R338L Padua mutation that confers a 8-fold increased activity both *in–vitro* and *in–vivo* [28, 38, 50, 82]. Once optimized the protocols in wild-type mice we plan the creation of HB mouse models for splicing mutations by applying the approach to F9 knock-out mice. This strategy will permit to assess the U1snRNA-mediated rescue of FIX antigen and activity levels and, most importantly, the impact of the treatment on the coagulation cascade efficiency (hFIX levels, aPTT assays) and the hemorrhagic phenotype (tail-clip assays).

2.10 General conclusions

The experimental activity on disease models such as factor VII deficiency and haemophilia B enabled us to explore two pioneer approaches for the rescue of gene expression of coagulation factor, namely the TALE-TFs and the modified snRNA U1 acting at the transcriptional and post-transcriptional levels, respectively. For the first time we provided experimental evidence that synthetic TALE-TFs can efficiently restore promoter activity impaired by disease-causing mutations. This approach might also suggest an innovative therapeutic approaches for human diseases in which the increased expression of a compensatory gene would produce beneficial effects [6]. Worth noting that eTFs could be also exploited to enhance the efficacy of other well-established correction approaches acting at the post- transcriptional levels (splicing modulation, ribosome read-through, etc.) by increasing the amount of the target mRNA. On the other hand, we have successfully explored the usage of the SB transposon system, which represented a useful tool to obtain models of HB caused by splicing mutations. This platform has

found in the last decade wide application and, because of its versatile activity in many different tissues, has been explored for ex-vivo and in-vivo gene delivery [25], but with the associated risks related to genotoxicity and efficient delivery. Our approach, in which SBTS represents the method for creating a model, has permitted to conduct the splicing studies and the assessment of the snRNA U1-mediated therapeutic approach in a genomic physiologic context, both in vitro and in vivo, thus avoiding the possible confounding effects related to the vectors episomal systems, that do not take into account the chromatin environment. The proposed approach could be improved and extended to F9 molecular defects other than splicing changes such as nonsense and missense mutations to create additional HB disease models and permit the investigation of alternative therapeutic options. It is worth noting that mutations affecting transcription or splicing are relatively frequent in human disease, particularly in the most severe cases. Both strategies imply the delivery to the affected cells (hepatocytes for coagulation factor disorders) of small size cassettes that can be packaged virtually in any expression vector, thus permitting to overcome limitations related to vector-mediated delivery of large genes. Moreover, the use of engineered U1 has the great advantage of maintaining the physiological gene regulation and of rescuing gene expression in the affected cells only. Therefore, the results obtained in these studies as well as the protocols so far developed, could be extended to a number of other human genetic diseases sharing similar pathogenic mechanisms and could pave the way for the development of novel therapeutic strategies where the current ones are difficult or poorly effective.

Bibliography

[1] Osterud B., B.N. Bouma, and J.H. Griffin. Human blood coagulation factor IX. purification, properties, and mechanism of activation by activated factor XI. *J Biol Chem*, 253(17):5946–51, 1978.

[2] Osterud B., K. Laake, and H. Prydz. The activation of human factor IX. *Thromb Diath Haemorrh*, 33(3):553–63, 1975.

[3] Osterud B. and S.I. Rapaport. Activation of factor IX by the reaction product of tissue factor and factor VII: additional pathway for initiating blood coagulation. *Proc Natl Acad Sci U S A*, 74(12):5260–4, 1977.

[4] Robberson B.L., G.J. Cote, and S.M. Berget. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol*, 10(1):84–94, 1990.

[5] Wu C. and C.E. Dunbar. Stem cell gene therapy: the risks of insertional mutagenesis and approaches to minimize genotoxicity. *Front Med*, 5(4):356–71, 2011.

[6] Gersbach C.A. and P. Perez-Pinera. Activating human genes with zinc finger proteins, transcription activator-like effectors and CRISPR/Cas9 for gene therapy and regenerative medicine. *Expert Opin Ther Targets*, 18(8):835–9, 2014.

[7] Kelly C.R., C.D. Dickinson, and W. Ruf. Ca2+ binding to the first epidermal growth factor module of coagulation factor VIIa is important for cofactor interaction and proteolytic function. *J Biol Chem*, 272(28):17467–72, 1997.

[8] Carroll D. Genome engineering with zinc-finger nucleases. *Genetics*, 188(4):773–82, 2011.

[9] Cross D. and J.K. Burmester. Gene therapy for cancer treatment: past, present and future. *Clin Med Res,*, 4(3):218–27, 2006.

[10] Gailani D. and G.J. Broze Jr. Factor XI activation in a revised model of blood coagulation. *Science*, 253(5022):909–12, 1991.

[11] Lillicrap D. The molecular basis of haemophilia b. *Haemophilia*, 4(4):350–7, 1998.

[12] Hall D.B. and K. Struhl. The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein cross-linking in vivo. *J Biol Chem*, 277(48):46043–50, 2002.

[13] Knoell D.L. and I.M. Yiu. Human gene therapy for hereditary diseases: a review of trials. *Am J Health Syst Pharm*, 55(9):899–904, 1998.

[14] Aiuti A. et al. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N Engl J Med*, 360(5):447–58, 2009.

[15] Aiuti A. et al. Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science*, 341(6148):1233151, 2013.

[16] Alexander B. et al. Congenital SPCA deficiency: a hitherto unrecognized coagulation defect with hemorrhage rectified by serum and serum fractions. *J Clin Invest*, 30(6:596–60, 1951.

[17] Anson D.S. et al. The gene structure of human anti-haemophilic factor IX. *EMBO J*, 3(5):1053–60, 1984.

[18] Arbini A.A. et al. Severe factor VII deficiency due to a mutation disrupting a hepatocyte nuclear factor 4 binding site in the factor VII promoter. *Blood*, 89(1):176–82, 1997.

[19] Balestra D. et al. An engineered u1 small nuclear RNA rescues splicing-defective coagulation F7 gene expression in mice. *J Thromb Haemost*, 12(2):177–185, 2014.

[20] Baralle M. et al. Identification of a mutation that perturbs NF1 agene splicing using genomic DNA samples and a minigene assay. *J Med Genet*, 40(3):220–2, 2003.

[21] Beerli R.R. et al. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *T. Proc Natl Acad Sci U S A*, 95(25):14628–33, 1998.

[22] Bibikova M. et al. Enhancing gene targeting with designed zinc finger nuclease. *s. Science*, 300(5620):764, 2003.

[23] Blazquez L. et al. Increased in vivo inhibition of gene expression by combining RNA interference and U1 inhibition. *Nucleic Acids Res*, 40(1):e8, 2012.

[24] Boch J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959):1509–12, 2009.

[25] Boehme P. et al. The Sleeping Beauty transposon vector system for treatment of rare genetic diseases: an unrealized hope? *Curr Gene Ther*, 2015.

[26] Brinster R.L. et al. Introns increase transcriptional efficiency in transgenic mice. *Proc Natl Acad Sci U S A*, 85(3):836–40, 1988.

[27] Bultmann S. et al. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res*, 40(12):5368–77, 2012.

[28] Cantore A. et al. Hyperfunctional coagulation factor IX improves the efficacy of gene therapy in hemophilic mice. *Blood*, 120(23):4517–20, 2012.

[29] Carew J.A. et al. Severe factor VII deficiency due to a mutation disrupting an Sp1 binding site in the factor VII promoter. *Blood*, 92(5):1639–45, 1998.

[30] Cazzella V. et al. Exon 45 skipping through U1-snRNA antisense molecules recovers the Dys-nNOS pathway and muscle differentiation in human DMD myoblasts. *Mol Ther*, 20(11):2134–42, 2012.

[31] Chapdelaine P. et al. A potential new therapeutic approach for friedreich ataxia: Induction of frataxin expression with TALE proteins. *Mol Ther Nucleic Acids*, 2:e119, 2013.

[32] Cohen M.E. et al. Human SLUG gene organization, expression, and chromosome map location on 8q. *Genomics*, 51(3):468–71, 1998.

[33] Cooper D.N. et al. Inherited factor VII deficiency: molecular genetics and pathophysiology. *Thromb Haemost*, 78(1):151–60, 1997.

[34] Denti M.A. et al. Chimeric adeno-associated virus/antisense U1 small nuclear RNA effectively rescues dystrophin synthesis and muscle function by local treatment of mdx mice. *Hum Gene Ther*, 17(5):565–74, 2006.

[35] Doak T.G. et al. A proposed superfamily of transposase genes: transposon–like elements in ciliated protozoa and a common D35E motif. *Proc Natl Acad Sci U S A*, 91(3):942–6, 1994.

[36] Dupuy A.J. et al. Mammalian germ-line transgenesis by transposition. *Proc Natl Acad Sci U S A*, 99(7):4495–9, 2002.

[37] Fernandez Alanis E. et al. An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum Mol Genet*, 21(11):2389–98, 2012.

[38] Finn J.D. et al. The efficacy and the risk of immunogenicity of fIX Padua (R338L) in hemophilia B dogs treated by AAV muscle gene therapy. *Blood*, 120(23):4521–3, 2012.

[39] Galla M. et al. Avoiding cytotoxicity of transposases by dose-controlled mRNA delivery. *Nucleic Acids Res*, 39(16):7147–60, 2011.

[40] Gao H. et al. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res*, 22(12):1716–20, 2012.

[41] Garg A. et al. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res*, 40(15):7584–95, 2012.

[42] Geurts A.M. et al. Gene transfer into genomes of human cells by the sleeping beauty transposon system. *Mol Ther*, 8(1):108–17, 2003.

[43] Glaus E. et al. Gene therapeutic approach using mutation-adapted U1 snRNA to correct a RPGR splice defect in patient-derived cells. *Mol Ther*, 19(5):936–41, 2011.

[44] Hacein-Bey-Abina S. et al. Efficacy of gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med*, 363(4):355–64, 2010.

[45] Hagen F.S. et al. Characterization of a cDNA coding for human factor VII. *Proc Natl Acad Sci U S A*, 83(8:2412–6, 1986.

[46] Hu J. et al. Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Res*, 42(7):4375–90, 2014.

[47] Ivics Z. et al. Identification of functional domains and evolution of Tc1-like transposable elements. *Proc Natl Acad Sci U S A*, 93(10):5008–13, 1996.

[48] Ivics Z. et al. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell*, 91(4):501–10, 1997.

[49] Izsvak Z. et al. Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in Sleeping Beauty transposition. *J Biol Chem*, 277(37):34581–8, 2002.

[50] Kao C.Y. et al. Incorporation of the factor IX Padua mutation into fIX-triple improves clotting activity in vitro and in vivo. *Thromb Haemost*, 110(2):244–56, 2013.

[51] Koornneef A. et al. AAV-mediated in vivo knockdown of luciferase using combinatorial RNAi and U1i. *Gene Ther*, 18(9):929–35, 2011.

[52] Koornneef A. et al. Apolipoprotein B knockdown by AAV-delivered shRNA lowers plasma cholesterol in mice. *Mol Ther*, 19(4):731–40, 2011.

[53] Li H. et al. In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature*, 475(7355):217–2, 2011.

[54] Li L. et al. Refinement of the NMR solution structure of the gamma–carboxyglutamic acid domain of coagulation factor IX using molecular dynamics simulation with initial ca2+ positions determined by a genetic algorithm. *Biochemistry*, 36(8):2132–8, 1997.

[55] Li Y. et al. Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci Rep,*, 2:897, 2012.

[56] Maeder M.L. et al. Robust, synergistic regulation of human gene expression using TALE activators. *Nat Methods*, 10(3):243–5, 2013.

[57] Maguire A.M. et al. Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N Engl J Med*, 358(21):2240–8, 2008.

[58] Mali P. et al. RNA-guided human genome engineering via Cas9. *Science*, 339(6121):823–6, 2013.

[59] Mariani G. et al. Molecular and clinical aspects of factor VII deficiency. *Blood Coagul Fibrinolysis*, 9 Suppl 1:S83–8, 1998.

[60] Mariani G. et al. Clinical manifestations, management, and molecular genetics in congenital factor VII deficiency: the international registry on congenital factor VII deficiency (irf7). *Blood*, 96(1):374, 2000.

[61] Mariani G. et al. Clinical phenotypes and factor VII genotype in congenital factor VII deficiency. *Thromb Haemost*, 93(3):481–7, 2005.

[62] Mates L. et al. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet*, 41(6):753–61, 2009.

[63] Miao C.H. et al. Inclusion of the hepatic locus control region, an intron, and untranslated region increases and stabilizes hepatic factor IX gene expression in vivo but not in vitro. *Mol Ther*, 1(6):522–32, 2000.

[64] Mikkelsen J.G. et al. Helper–Independent Sleeping Beauty transposon-transposase vectors for efficient nonviral gene delivery and persistent gene expression in vivo. *Mol Ther*, 8(4):654–65, 2003.

[65] Millar D.S. et al. Molecular analysis of the genotype-phenotype relationship in factor VII deficiency. *Hum Genet*, 107(4):327–42, 2000.

[66] Miller J.C. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, 29(2):143–8, 2011.

[67] Miskey C. et al. The Frog Prince: a reconstructed transposon from Rana pipiens with high transpositional activity in vertebrate cells. *Nucleic Acids Res*, 31(23):6873–81, 2003.

[68] Nathwani A.C. et al. Adenovirus-associated virus vector-mediated gene transfer in haemophilia B. *N Engl J Med*, 365(25):2357–65, 2011.

[69] Nathwani A.C. et al. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N Engl J Med*, 371(21):1994–2004, 2014.

[70] O'Hara P.J. et al. Nucleotide sequence of the gene coding for human factor VII, a vitamin k-dependent protein participating in blood coagulation. *Proc Natl Acad Sci U S A*, 84(15):5158–62, 1987.

[71] Palmiter R.D. et al. Heterologous introns can enhance expression of transgenes in mice. *Proc Natl Acad Sci U S A*, 88(2):478–82, 1991.

[72] Perez-Pinera P. et al. Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat Method*, 10(3):239–42, 2013.

[73] Pinotti M. et al. U1-snRNA-mediated rescue of mRNA processing in severe factor VII deficiency. *Blood*, 111(5):2681–4, 2008.

[74] Pinotti M. et al. Rescue of coagulation factor VII function by the U1+5A snRNA. *Blood*, 113(25):6461–4, 2009.

[75] Pinotti M. et al. RNA-based therapeutic approaches for coagulation factor deficiencies. *J Thromb Haemost*, 9(11):2143–52, 2011.

[76] Pollak E.S. et al. Functional characterization of the human factor VII 5'-flanking region. *J Biol Chem*, 271(3):1738–47, 1996.

[77] Rosen E.D. et al. Mice lacking factor VII develop normally but suffer fatal perinatal bleeding. *Nature*, 390(6657):290–4, 1997.

[78] Rossi F. et al. Involvement of u1 small nuclear ribonucleoproteins (snRNP) in 5' splice site-U1 snRNP interaction. *J Biol Chem*, 271(39):23985–91, 1996.

[79] Sanjana N.E. et al. A transcription activator-like effector toolbox for genome engineering. *Nat Protoc*, 7(1):171–92, 2012.

[80] Schmid F. et al. U1 snRNA-mediated gene therapeutic correction of splice defects caused by an exceptionally mild BBS mutation. *Hum Mutat*, 32(7):815–24, 2011.

[81] Sickmier E.A. et al. Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol Cell*, 23(1):49–59, 2006.

[82] Simioni P. et al. X-linked thrombophilia with a mutant factor IX (factor IX padua). *N Engl J Med*, 361(17):1671–5, 2009.

[83] Susani L. et al. TCIRG1-dependent recessive osteopetrosis: mutation analysis, functional identification of the splicing defects, and in vitro rescue by U1 snRNA. *Hum Mutat*, 24(3):225–35, 2004.

[84] Tanner G. et al. Therapeutic strategy to rescue mutation-induced exon skipping in rhodopsin by adaptation of U1 snRNA. *Hum Mutat*, 30(2):255–63, 2009.

[85] Tolar J. et al. Gene therapy for Fanconi anemia: one step closer to the clinic. *Hum Gene Ther*, 23(2):141–4, 2012.

[86] Tremblay J.P. et al. Transcription activator–like effector proteins induce the expression of the frataxin gene. *Hum Gene Ther*, 23(8):883–90, 2012.

[87] Triplett D.A. et al. Hereditary factor VII deficiency: heterogeneity defined by combined functional and immunochemical analysis. *Blood*, 66(6):1284–7, 1985.

[88] Urnov F.D. et al. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, 11(9):636–46, 2010.

[89] Will C.L. et al. In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res*, 24(23):4614–23, 1996.

[90] Wion K.L. et al. Distribution of factor VIII mRNA and antigen in human liver and other tissues. *Nature*, 317(6039:726–9, 1985.

[91] Yoshitake S. et al. Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). *Biochemistry*, 24(14):3736–50, 1985.

[92] Zayed H. et al. Development of hyperactive sleeping beauty transposon vectors by mutational analysis. *Mol Ther*, 9(2):292–304, 2004.

[93] Zhang F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol*, 29(2):149–53, 2011.

[94] Zhang M. et al. TALE: a tale of genome editing. *Prog Biophys Mol Biol*, 114(1):25–32, 2014.

[95] Davie E.W. A brief historical review of the waterfall/cascade of blood coagulation. *J Biol Chem*, 278(51):50819–32, 2003.

[96] Davie E.W. and O.D. Ratnoff. Waterfall sequence for intrinsic blood clotting. *Science*, 145(3638):1310–2, 1964.

[97] Muller F. and T. Renne. Novel roles for factor XII-driven plasma contact activation system. *Curr Opin Hematol*, 15(5):516–21, 2008.

[98] Mariani G. *Congenital Factor VII Deficiency*. 2 edition, 2010.

[99] Mariani G. and F. Bernardi. Factor VII Deficiency. *Semin Thromb Hemost*, 35(4):400–6, 2009.

[100] Long G.L. Structure and evolution of the human genes encoding protein C and coagulation factors VII, IX, and x. *Cold Spring Harb Symp Quant Biol*, 51 Pt 1:525–9, 1986.

[101] Amphlett G.W., W. Kisiel, and F.J. Castellino. The interaction of Ca2+ with human factor IX. *Arch Biochem Biophys*, 208(2):576–85, 1981.

[102] Boch J. and U. Bonas. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol*, 48:419–36, 2010.

[103] Lawson J.H. and K.G. Mann. Cooperative activation of human factor IX by the human extrinsic pathway of blood coagulation. *J Biol Chem*, 266(17):11317–27, 1991.

[104] High K.A. Gene therapy for hemophilia: the clot thickens. *Hum Gene Ther*, 25(11):915–22, 2014.

[105] Mann K.G., Krishnaswamy S., and J.H. Lawson. Surface-dependent hemostasis. *Semin Hematol*, 29(3):213–26, 1992.

[106] Carcao M.D. and L. Aledort. Prophylactic factor replacement in hemophilia. *Blood Rev*, 18(2):101–13, 2004.

[107] Moscou M.J. and A.J. Bogdanove. A simple cipher governs DNA recognition by TAL effectors. *Science*, 326(5959):1501, 2009.

[108] Green M.R. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu Rev Cell Biol*, 7:559–99, 1991.

[109] Blancafort P., L. Magnenat, and 3rd C.F. Barbas. Scanning the human genome with combinatorial transcription factor libraries. *Nat Biotechnol*, 21(3):269–74, 2003.

[110] Perez-Pinera P., D.G. Ousterout, and C.A. Gersbach. Advances in targeted genome editing. *Curr Opin Chem Biol*, 16(3-4):268–77, 2012.

[111] Radcliffe R. and Y. Nemerson. Activation and control of factor VII by activated factor X and thrombin. isolation and characterization of a single chain form of factor VII. *J Biol Chem*, 250(2):388–95, 1975.

[112] Macfarlane R.G. An enzyme cascade in the blood clotting mechanism, and its function as a biochemical amplifier. *Nature*, 202:498–9, 1964.

[113] Plasterk R.H, Z. Izsvak, and Z. Ivics. Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet*, 15(8):326–32, 1999.

[114] Beerli R.R., B. Dreier, and C.F. Barbas 3rd. Positive and negative regulation of endogenous genes by designed transcription factors. *Proc Natl Acad Sci U S A*, 97(4):1495–500, 2000.

[115] Misra S. Human gene therapy: a brief overview of the genetic revolution. *J Assoc Physicians India*, 61(2):127–33, 2013.

[116] Berget S.M. Exon recognition in vertebrate splicing. *J Biol Chem*, 270(6):2411–4, 1995.

[117] Bajaj S.P. Cooperative ca2+ binding to human factor IX. effects of ca2+ on the kinetic parameters of the activation of factor IX by factor XIa. *J Biol Chem,*, 257(8):4127–32, 1982.

[118] Bajaj S.P., S.I. Rapaport, and W.A. Russell. Redetermination of the rate-limiting step in the activation of factor IX by factor XIa and by factor VIIa/tissue factor. Explanation for different electrophoretic radioactivity profiles obtained on activation of 3H- and 125I-labeled factor IX. *Biochemistry*, 22(17):4047–53, 1983.

[119] Pipe S.W. Recombinant clotting factors. *Thromb Haemost*, 99(5):840–50, 2008.

[120] Bom V.J. and R.M. Bertina. The contributions of Ca2+, phospholipids and tissue-factor apoprotein to the activation of human blood-coagulation factor X by activated factor VII. *Biochem J*, 265(2):327–36, 1990.

[121] Gommans W.M., H.J. Haisma, and M.G. Rots. Engineering zinc finger protein transcription factors: the therapeutic relevance of switching endogenous gene expression on or off at command. *J Mol Biol*, 354(3):507–19, 2005.

[122] Nemerson Y. The reaction between bovine brain tissue factor and factors VII and X. *Biochemistry*, 5(2):601–8, 1966.

[123] Izsvak Z. and Z. Ivics. Sleeping beauty transposition: biology and applications for molecular therapy. *Mol Ther*, 9(2):147–56, 2004.

[124] Izsvak Z., Z. Ivics, and R.H. Plasterk. Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *J Mol Biol*, 302(1):93–102, 2000.