



Università degli Studi di Cagliari

## **DOTTORATO DI RICERCA**

Terapia Pediatrica e Farmacologia dello Sviluppo

Ciclo XXVIII

**Terapia genica della  $\beta$  Talassemia mediante *editing* del DNA**

Settore/i scientifico disciplinari di afferenza

[MED/03]

Presentata da:	Dott.ssa Maura Mingoia
Coordinatore Dottorato	Prof. Paolo Moi
Tutor/Relatore	Prof. Paolo Moi
Correlatore	Dott.ssa Maria Giuseppina Marini

Esame finale anno accademico 2014 – 2015

# Indice

1. Abstract	3
2. Introduzione	4
2.1 Emoglobina	4
a. I geni globinici umani	6
b. La regolazione dei geni globinici	8
c. L'espressione dei geni globinici durante l'ontogenesi	11
d. Lo switching HbF→HbA	14
e. Le emoglobinopatie	24
f. $\beta$ Talassemia	25
i. Caratteristiche cliniche	28
ii. Genetica molecolare della $\beta$ talassemia	30
iii. Modificatori genetici	33
g. Le Terapie	42
2.2 Il <i>Genome-Editing</i>	47
a. Il Sistema CRISPR/Cas	50
i. gRNA	56
ii. Cas9	57
iii. Donor	58
3. Scopo della tesi	59
4. Materiali e Metodi	60
4.1 Coltura della linea cellulare $\beta$ -K562	60
4.2 Design dell'RNA guida	61
4.3 Clonaggio in vettore CRISPR	61
4.4 Elettroporazione delle cellule $\beta$ -K652	63
4.5 FACS <i>sorting</i>	63

4.6	Estrazione DNA, RNA e proteine	64
4.7	Sequenziamento di Sanger	64
4.8	PCR screening	65
4.9	RT-qPCR	65
4.10	Western blot	66
<b>5.</b>	<b>Risultati</b>	<b>67</b>
5.1	DSB RNA-guidato	67
5.2	Analisi sull'efficacia dell' <i>editing</i>	69
5.3	Caratterizzazione molecolare dell' <i>editing</i> RNA-guidato nella sequenza target	73
5.4	PCR screening	80
5.5	Caratterizzazione dell'effetto dell' <i>editing</i> sull'espressione dei geni globinici	84
<b>6.</b>	<b>Discussione</b>	<b>89</b>
<b>7.</b>	<b>Bibliografia</b>	<b>94</b>

# 1. Abstract

$\beta$  thalassemia is a hereditary disease caused by mutations in the human hemoglobin  $\beta$  (HBB) gene, resulting in absence ( $\beta^0$ ) or deficiency ( $\beta^+$ ) of  $\beta$  globin chain synthesis. This genetic disorder occurs most frequently in people from Mediterranean countries, such as Italy. In particular, the data indicates that about 12.6% of the Sardinian subjects are carriers of  $\beta$  thalassemia and these are among the highest frequencies of thalassemia genes found in a Caucasian population. In Sardinia, the disease is generally determined by a nonsense mutation at codon 39 (E39X) of exon 2 causing the interruption of  $\beta$  globin synthesis. Patients homozygous with E39X mutation have a severe anemia and require frequent transfusions and iron chelation. The only definitive cure today possible for  $\beta$  chain hemoglobinopathies is the hematopoietic stem cells transplantation, but it is limited by availability of HLA matched donors. However, in the last few years new therapeutic approaches for this genetic disease are taking place. The correction of disease-causing mutation through the technique of *Genome-Editing* in patient-specific stem cells and subsequent autologous transplantation would be the ideal approach for the treatment of monogenic diseases such as  $\beta$  thalassemia. However, due to difficulties in obtaining sufficient homologous recombination percentages for therapeutic purposes, the aim of my PhD project is to reproduce artificially the HPFH mutations identified in non-coding regions of the  $\beta$  globin cluster, using the system CRISPR/Cas9 associated with NHEJ pathway. In this way, we hope to restore at therapeutic levels the expression of HBG genes and

consequently the synthesis of a functional HbF in order to ameliorate the phenotype of  $\beta$  thalassemia.

## 2. Introduzione

### 2.1 L'emoglobina

La presenza dell'ossigeno nell'atmosfera della Terra e il suo ruolo nella combustione delle sostanze organiche ha favorito l'evoluzione di diversi meccanismi che ne consentono la conservazione e il trasporto. Negli organismi più semplici il trasporto dell'ossigeno e il rifornimento ai processi respiratori dipendono esclusivamente del fenomeno della diffusione per gradiente. Diversamente gli organismi più complessi utilizzano dei sistemi di trasporto, che forniscono ai tessuti l'apporto di ossigeno necessario al metabolismo cellulare tramite specifiche proteine in grado di legarlo e trasportarlo. Nei mammiferi la proteina preposta a tale compito si trova all'interno dei globuli rossi e prende il nome di emoglobina (Hb). L'Hb è la prima proteina la cui funzione è stata identificata con esattezza e di cui è nota nel dettaglio la struttura molecolare. Essa infatti è una cromo-proteina tetramerica con struttura quaternaria costituita da due catene globiniche di tipo  $\alpha$  ( $\zeta$  e  $\alpha$ ), di 141 residui amminoacidici, e due catene di tipo  $\beta$  ( $\epsilon$ ,  $G\gamma$ ,  $A\gamma$ ,  $\delta$  e  $\beta$ ), di 146 amminoacidi. Le quattro subunità si assemblano spontaneamente mediante interazioni elettrostatiche e idrofobiche

organizzandosi in dimeri di protomeri  $\alpha\beta$ . Ciascuna globina è a sua volta legata ad un gruppo prostetico protoporfirinico chiamato EME a cui è legato un singolo atomo di ferro nello stato di ossidazione ferroso ( $\text{Fe}^{2+}$ ), in grado di legare reversibilmente l'ossigeno (FIG.1) (1, 2). Nonostante le numerose conoscenze acquisite in questo campo, ancora oggi vengono condotti numerosi studi sull'Hb e sui geni globinici in tutto il mondo. Il motivo principale risiede non solo nel fatto che le emoglobinopatie rappresentano le malattie congenite più diffuse nel mondo. Inoltre la comprensione dei processi alla base della regolazione dell'espressione dei geni globinici nella loro complessità può essere utile nella realizzazione di modelli di studio applicabili ad altri sistemi genici.

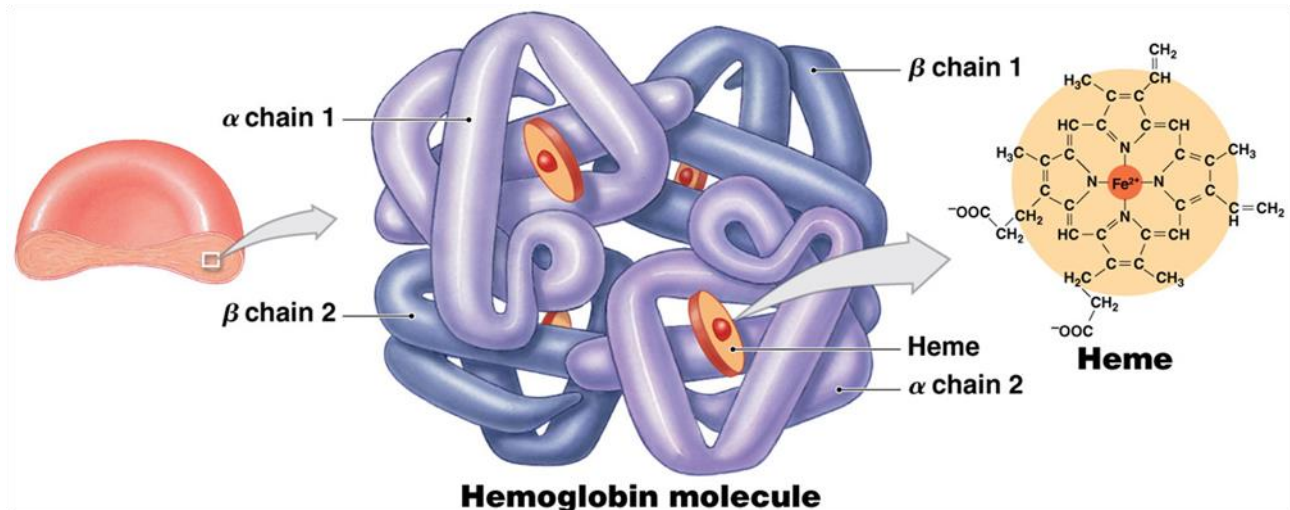


Fig. 1: Struttura dell'emoglobina e del gruppo eme.

### a. I geni globinici umani

Nel caso specifico dell'uomo, le emoglobine presenti nel sangue in età adulta sono rappresentate per il 96% da una componente maggiore, HbA, composta da due catene  $\alpha$  e due catene  $\beta$ , da una componente minore l'HbA<sub>2</sub>, che rappresenta il 2,5% circa del totale ed è formata da due catene  $\alpha$  e due catene  $\delta$  e da una piccola quantità residua (circa lo 0,8%) di emoglobina fetale (HbF), propria della vita intrauterina, costituita da due catene  $\alpha$  e due catene  $\gamma$ . I geni deputati all'espressione delle globine sono organizzati in due raggruppamenti della dimensione di circa 50 Kb chiamati *cluster*  $\alpha$  e  $\beta$ , localizzati rispettivamente all'estremità del braccio corto del cromosoma 16 (16p13.3) e nella regione 11p15.15 del braccio corto del cromosoma 11. In entrambi i gruppi i geni sono disposti dal 5' al 3' secondo l'ordine con cui vengono attivati durante l'ontogenesi. Il cluster  $\alpha$  contiene i due geni  $\alpha$ -globinici duplicati (HBA1 e ed HBA2), il gene  $\alpha$ -simile embrionario  $\zeta 2$  e tre pseudogeni ( $\psi\zeta 1$ ,  $\psi\alpha 2$  e  $\psi\alpha 2$ ) disposti nel seguente ordine: 5'- $\zeta 2$ -  $\psi\zeta 1$ -  $\psi\alpha 2$ -  $\psi\alpha 1$ - $\alpha 2$ - $\alpha 1$ -3' (3). Il cluster  $\beta$  globinico è costituito da un pseudogene ( $\psi\beta 1$ ) non funzionante e cinque geni  $\beta$ -simili disposti nel seguente ordine: 5'-  $\epsilon$  -  $G\gamma$  -  $A\gamma$  -  $\delta$  -  $\beta$  - 3'. I geni  $\beta$ -simili hanno 3 esoni che codificano per 146 amminoacidi e due introni di 122-130 e 850-900 bp, tra i codoni 30-31 e 104-105 (FIG.2). La struttura dei geni  $\alpha$ -simili è analoga, ma gli introni sono più piccoli. Gli pseudogeni hanno una marcata omologia di sequenza con i geni funzionali, ma contengono diverse mutazioni che ne impediscono l'espressione (4, 5).

Gli pseudogeni possono perciò essere considerati dei residui vestigiali di geni che sono stati funzionanti in certi periodi dell'evoluzione della specie. I due geni  $\alpha$  codificano per catene globiniche  $\alpha$  identiche. Il gene  $\alpha 1$  ha una maggiore espressione del gene  $\alpha 2$  (rapporto 3:1). I due geni  $\gamma$  (HBG) codificano anch'essi per catene  $\gamma$  identiche ad esclusione del polimorfismo glicina/alanina in posizione 136 (6). Il rapporto di sintesi tra gene  $G\gamma$  (HBG2) e gene  $A\gamma$  (HBG1) è di 3:1 nel feto e di 2:3 nell'adulto (7). I due geni HBA e HBG sono il prodotto di un processo di duplicazione e conversione genica. L'elevata espressione del gene 5' nei due geni duplicati HBA e HBG è dovuta probabilmente alla loro interazione preferenziale, per ragioni di vicinanza, con le sequenze regolatrici. I geni globinici vengono regolati durante lo sviluppo mediante un fine meccanismo di controllo dell'espressione a livello trascrizionale, post-trascrizionale e post-traduzionale. Quando è richiesta l'attivazione di un gene globinico l'interazione tra determinati fattori di trascrizione e le sequenze regolatrici presenti nel cluster permettono il posizionamento del macchinario di trascrizione sul promotore della specifica globina consentendone l'espressione.



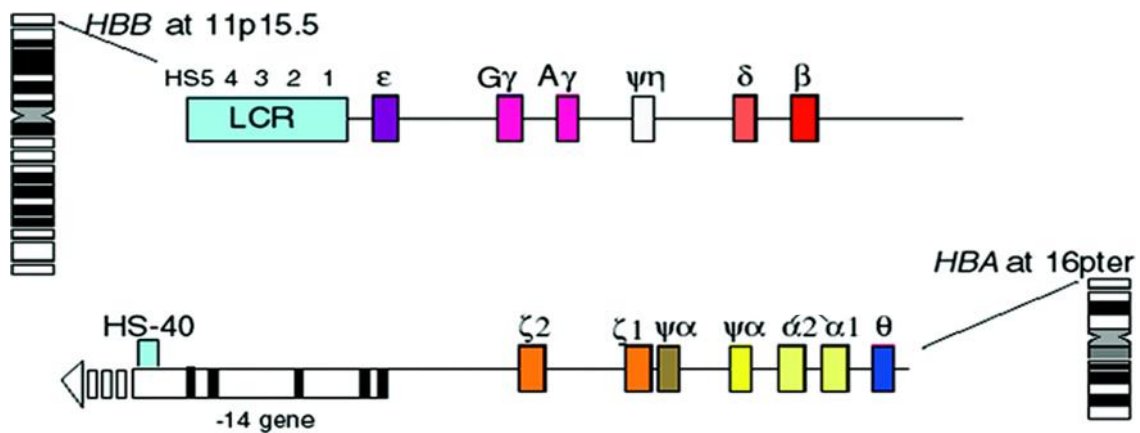


FIG.2: Localizzazione cromosomica e organizzazione del cluster  $\beta$  e  $\alpha$  globinico.

## b. La regolazione dei geni globinici

La regolazione dell'espressione dei geni globinici è affidata alla presenza di elementi in grado di interagire con i fattori di trascrizione eritroidi-specifici e ubiquitari situati generalmente nella regione *promoter*, da sequenze *enhancer* o *silencer* e da un'importante sequenza situata in posizione 5' rispetto ai geni la cui delezione comporta la totale inattivazione dell'espressione genica (8). Nel cluster  $\alpha$  questa sequenza è denominata HS40 e si trova a circa 40 Kb a monte dei geni (9). Nel cluster  $\beta$  tale regione è rappresentata dalla Locus Control Region (LCR), localizzata nella regione tra 5 e 25 Kb a monte del gene  $\epsilon$  e a 50 Kb dal gene  $\beta$  (10, 11). Essa è costituito da cinque siti caratterizzati da ipersensibilità alla digestione con DNAsi (HS), eritroidi specifici (HS1, HS2, HS3, HS4) a parte il sito HS5. Tali sequenze sono tipiche di quelle strutture cromatiniche cosiddette "aperte", caratteristica dei geni trascrizionalmente attivi in cui è facilitato l'accesso degli elementi di regolazione necessari per

l'attivazione dell'espressione dei geni legati. Ciascun sito HS contiene un elemento "core" di 200-300 bp costituito da specifiche combinazioni di siti di legame per fattori di trascrizione sia ubiquitari che eritroidi specifici. L'LCR rappresenta un complesso elemento regolatore con funzione di *enhancer* eritroide specifico, di isolatore e di rimodellamento della struttura cromatinica.

Dalla scoperta della LCR sono stati condotti una pleora di esperimenti col fine di chiarirne il funzionamento. Nonostante le diverse opinioni e i modelli spesso contrastanti proposti per spiegare la modalità di azione dell'LCR, il modello corrente più accreditato suggerisce che l'LCR agisca come una singola unità funzionale. Secondo tale modello proposto per la prima volta da Grosveld, i singoli HS interagiscono fra loro formando un olocomplexo in grado di attivare un singolo gene globinico alla volta durante le varie fasi dello sviluppo eritroide. L'espressione genica di un specifico gene globinico sarebbe così il risultato di una competizione dinamica per il legame tra i geni e l'LCR. La stabilità e la durata di tali interazioni sembrano dipendere dall'ordine dei geni rispetto al cluster, dalla loro distanza e orientamento dall' LCR e fondamentalmente dall'ambiente di fattori di trascrizione presenti durante i diversi stadi dell'eritropoiesi.

A livello del promoter  $\beta$  globinico sono state identificate altre sequenze regolatrici TATA, CAAT, e CACCC box. Questi elementi, se pur con variazioni di sequenza minime, sono presenti in tutti i promotori dei geni globinici. Tuttavia i promotori oggetto di un'indagine più intensa sono quelli relativi ai geni  $\gamma$  e  $\beta$  (HBB) per via del loro

coinvolgimento nel fenomeno dello *switching* da Hb fetale ad adulta. Infatti da questi studi è stata evidenziata una differenza sostanziale tra i due promotori in quanto nel  $\gamma$  è presente una duplicazione del CAAT box (prossimale e distale) ed un unico CACCC, al contrario nel  $\beta$  sono i CACCC box ad essere presenti in duplicato (prossimale e distale) e CAAT è singolo. Queste differenze possono avere delle implicazioni nella regolazione dell'espressione di tali geni. Diversi fattori ubiquitari sono risultati in grado di legare queste sequenze in condizioni sperimentali, ma i fattori realmente fondamentali sono quelli definiti eritroidi-specifici. Perciò il livello di trascrizione di un gene  $\beta$  globinico è il risultato della frequenza con cui si verifica l'interazione tra LCR e il promotore del gene specifico e dalla sua stabilità (12, 13). L'ipotesi più accreditata con cui l'LCR svolge il suo ruolo prevede l'interazione diretta tra i siti HS e i promotori prossimali dei geni globinici mediante la formazione di *loop*, consentendogli di regolarne l'attività durante le diverse fasi dello sviluppo eritroide (FIG.3).

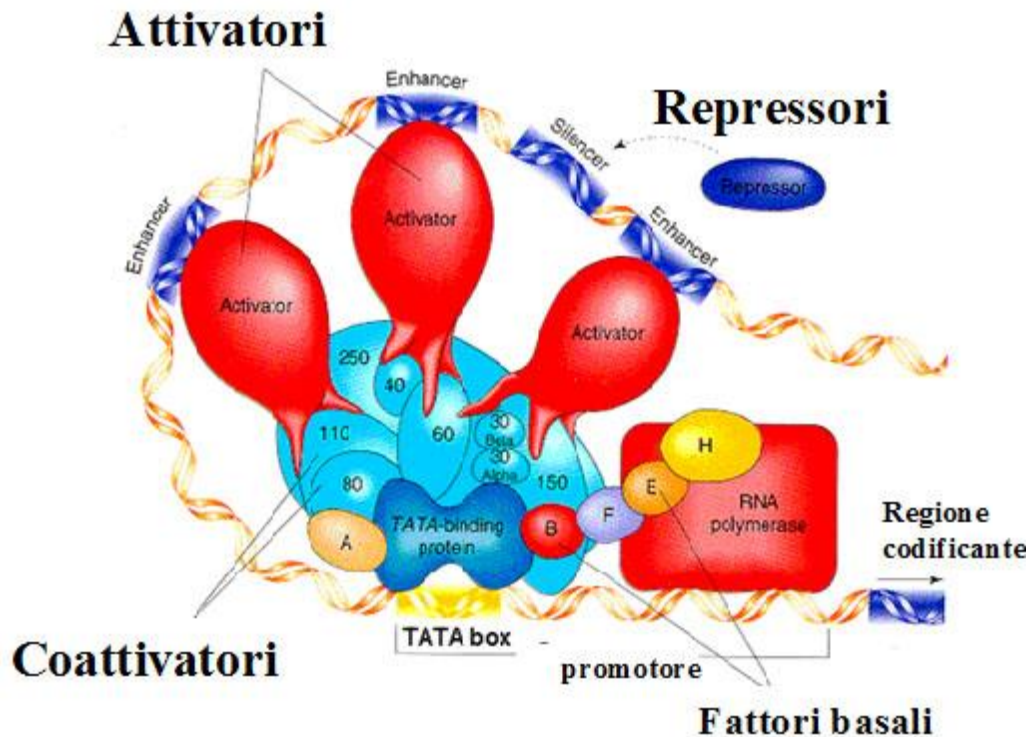


FIG 3: Esempio di una struttura a "loop".

### c. L'espressione dei geni globinici durante l'ontogenesi

I loci globinici agiscono in concerto nel produrre differenti sottotipi di emoglobina in risposta alle diverse richieste di ossigeno durante le varie fasi dello sviluppo.

Durante i tre stadi dello sviluppo ontogenico (embrionale, fetale ed adulto) si osserva l'espressione di sette differenti tipi di Hb, la quale è governata da meccanismi di regolazione spazio-temporale che nel loro insieme prendono il nome di *switching* globinico. Nell'uomo si osservano due eventi di *switching*: il primo avviene dopo la quinta settimana di gestazione e determina la variazione dell'espressione dai geni embrionali ( $\zeta$  ed  $\epsilon$ ) a quelli fetali ( $\alpha$  e  $\gamma$ ) mentre il secondo, nonché più importante,

avviene in periodo peri-natale e porta all'attivazione dei geni  $\beta$ -globinici di tipo adulto ( $\beta$  e  $\delta$ ) e al graduale spegnimento dei geni  $\gamma$  (FIG.4).

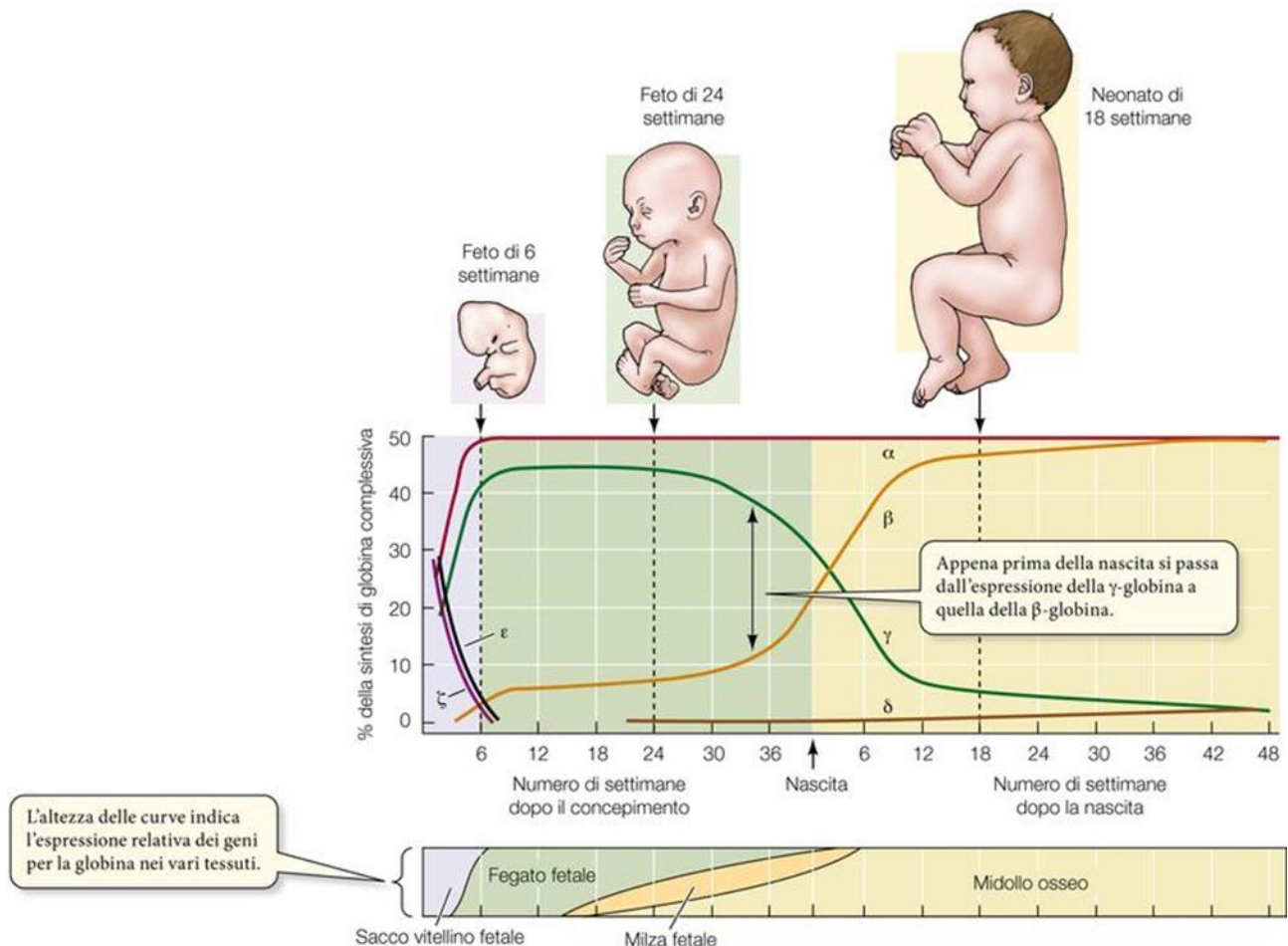


FIG.4: Nel corso dello sviluppo umano, in momenti diversi e in tessuti diversi vengono espressi differenti geni globinici.

Le Hb embrionali sono i primi sottotipi di emoglobina espresse, nei megaloblasti del sacco vitellino. Esistono tre differenti tipi di emoglobine embrionali: l'Hb *Gower 1* formata dalla catena  $\alpha$ -simile  $\zeta$  e  $\beta$ -simile  $\epsilon$  ( $\zeta_2\epsilon_2$ ), l'Hb *Gower 2* che ha struttura  $\alpha_2\epsilon_2$  e l'Hb *Portland* con la formula  $\zeta_2\gamma_2$ . L'espressione dell'Hb embrionale prosegue fino

al 35-42 giorno di vita intrauterina quando gradualmente viene sostituita dall'emoglobina fetale (HbF), prodotta prevalentemente nel fegato fetale per poi spostarsi a livello splenico. L'HbF è costituita da due catene  $\alpha$  e due catene  $\gamma$ , rappresenta la principale Hb presente durante la vita intrauterina e della quale sono distinguibili due tipi a seconda che la  $\gamma$  globina porti in posizione 136 della catena polipeptidica l'aminoacido glicina (G $\gamma$ ) o l'aminoacido alanina (A $\gamma$ ). Poco prima della nascita si verifica il secondo *switching* globinico quando la sintesi dell'HbF incomincia a ridursi e gradualmente ad incrementare la sintesi della emoglobina adulta HbA ( $\alpha_2\beta_2$ ). L'espressione dell'HbA inizia progressivamente già dalla decima settimana di gestazione e ha luogo prima nei normociti del fegato poi della milza e infine nella sede definitiva dell'eritropiesi, ossia il midollo osseo. La produzione di HbA<sub>2</sub> ( $\alpha_2\delta_2$ ) invece risulta essere più tardiva, infatti la sua espressione comincia lentamente nel periodo perinatale e si rende visibile dopo i primi due mesi di vita. Mentre i livelli di espressione delle globine di tipo  $\gamma$  rappresentano il 90% del totale intorno alla 35<sup>a</sup> settimana di vita intrauterina, il 60-80% alla nascita e lo 0,5%, a due anni circa dalla nascita. Tale riduzione indica come lo spegnimento dei geni HBG a favore dell'espressione dei geni  $\beta$  e  $\delta$  globinici sia un fenomeno che avviene in maniera graduale (FIG.5) (4, 14).

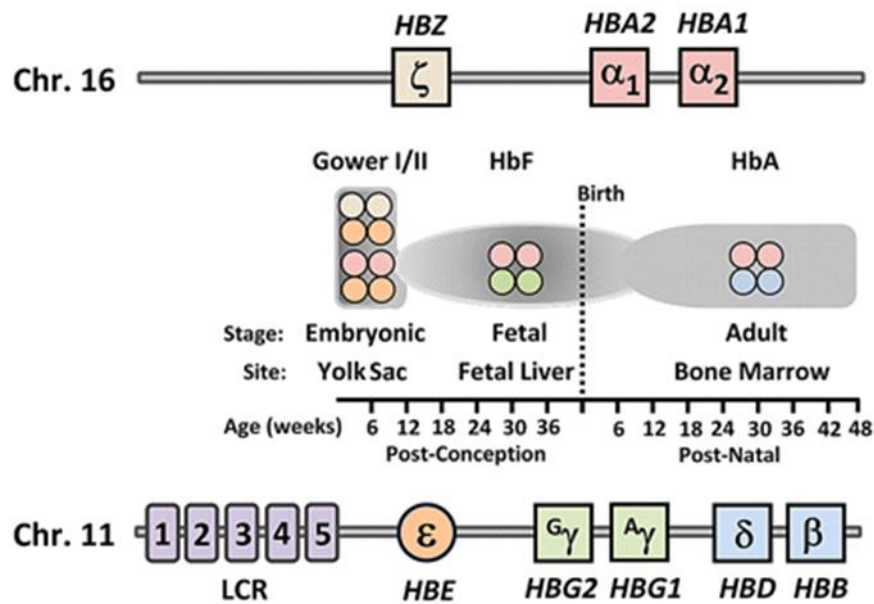


FIG.5: differenti tipi di globine espresse durante lo sviluppo ontogenico.

#### d. Lo *switching* HbF $\rightarrow$ HbA

La transizione da HbF a HbA rappresenta uno degli eventi più importanti e più studiati nella biologia dell'Hb. I geni umani HBG e HBB mostrano un'espressione strettamente correlata allo stadio dello sviluppo ontogenico. Negli ultimi due decenni, è stato dimostrato che numerosi fattori di trascrizione sono in grado di legarsi a diverse regioni regolatrici presenti all'interno dell'LCR e in ogni singolo promotore (cis-elementi). I meccanismi trascrizionali descritti prevedono il legame di fattori transattivanti sequenza-specifici (fattori di trascrizione, TFS) a particolari sequenze di DNA, i cis-elementi, seguita dall'interazione con dei fattori co-regolatori tramite interazioni dirette proteina-proteina. I fattori co-regolatori tipicamente formano grandi complessi multiproteici e mediano l'attivazione o la repressione in modo dipendente

dal contesto eritroide studiato. Comprendere nel dettaglio questo processo ha evidenti implicazioni nella terapia di  $\beta$  talassemia e anemia falciforme. Infatti i geni HBG possono funzionalmente sostituire il gene HBB, che in queste patologie risulta essere difettoso. Perciò ogni progresso nella comprensione dello *switching* può portare allo sviluppo di terapie alternative basate sulla riattivazione del gene HBG (15). Le scoperte più importanti in questo senso sono legate all'identificazione dei fattori di trascrizione che regolano l'espressione fase-specifica dei geni HBG e HBB. Tra le varietà di fattori trascrizionali e co-regolatori del differenziamento eritroide, alcuni sono di particolare importanza.

GATA1 è una DNA *binding-protein* in grado di legare tramite due motivi *zinc-finger* la sequenza consenso (A/T)GATA(A/G) (16). Tale sequenza è presente sia negli HS dell'LCR che nei promotori, condizione che suggerisce una regolazione del tipo *promoter-enhancer looping*. Nel caso  $\gamma$  promoter intorno alla posizione -175 è presente un motivo GATA associato a *down-regulation* della  $\gamma$  globina (17). Diversamente nel gene HBB è stato identificato una regione *enhancer* in cui è presente un sito di riconoscimento per il fattore GATA1 a valle, associata all'aumento dell'espressione di  $\beta$  globina. Inoltre risulta essenziale per il differenziamento eritroide in quanto il topo *knock-out* per GATA1 manifesta un arresto della maturazione allo stadio di proeritroblasto (FIG.6) (18, 19).



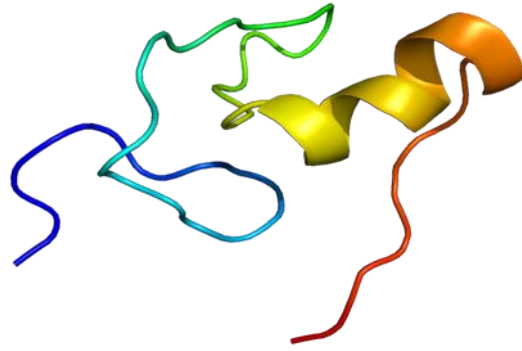


FIG. 6: la struttura del fattore GATA1.

FOG1 (Friend Of GATA) è un cofattore in grado di interagire e formare eterodimeri con i fattori di trascrizione appartenenti alla famiglia GATA (1,2,3). Tale eterodimero può sia attivare che reprimere la trascrizione a seconda del contesto cellulare in cui si trova e del promotore con cui interagisce. Nelle cellule eritroidi è in grado di attivare l'espressione dei geni NF-E2, HBA e HBB (FIG.7) (20, 21, 22).

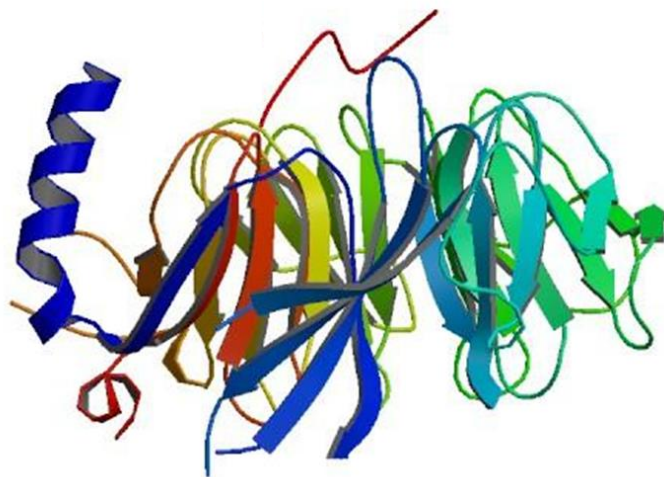


FIG.7: la struttura del cofattore FOG1.

NF-E2 (fattore nucleare eritroide 2) è un importante regolatore della sintesi emoglobinica in complesso con altri fattori. Nello specifico è in grado di legarsi alla sequenza consenso (T/C)GCTGA(C/G)TCA(T/C) presente nel sito HS2 dell'LCR. Oltre all'LCR interagisce con il promotore  $\beta$  attivando la trascrizione del gene HBB (FIG.8) (23, 24).

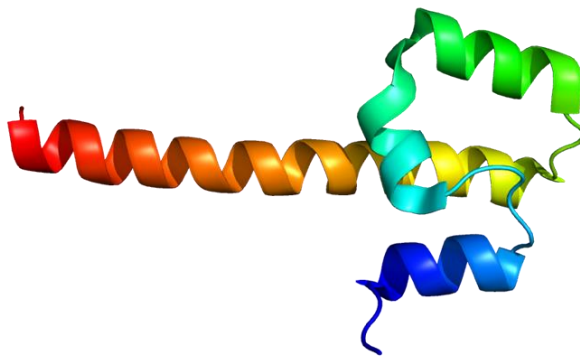


FIG.8: la struttura del fattore NF-E2.

KLF1 (Kruppel-like-factor) è un fattore eritroide-specifico in grado di legare la sequenza consenso CCACACCCT tramite tre motivi *zinc-finger*. Esso è un potente regolatore dei geni  $\beta$ -like nell'adulto in quanto possiede un doppio ruolo di regolatore durante lo *switching*. È in grado di attivare direttamente il gene  $\beta$  globinico interagendo con i siti HS1 HS3 dell'LCR e i CACCC box del promotore  $\beta$ , in particolare il CACCC prossimale, favorendo la formazione del *loop* (25, 26). Inoltre interagisce con il promotore del gene BCL11A, attivando l'espressione di un potente repressore dei geni HBG1/HBG2 (27). È stato osservato che durante il periodo fetale i livelli di KLF1

risultano essere bassi, consentendo l'interazione dell'LCR con i promotori dei geni  $\gamma$ , che vengono espressi per la sintesi di HbF. Nel periodo post-natale si rileva invece un aumento dei livelli di KLF1 associato ad un incremento dell'espressione dei geni HBB e BCL11A, con conseguente diminuzione della sintesi di  $\gamma$  globina (FIG.9) (28).

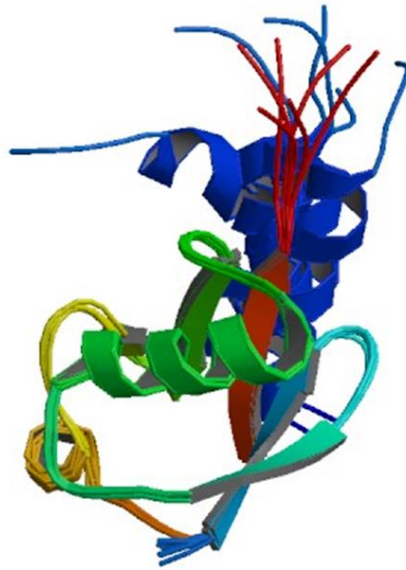


FIG.9: la struttura del fattore KLF1.

BCL11A (B-Cell Lymphoma/Leukemia 11 A) è un regolatore negativo dei geni HBG di tipo zinc-finger. Scoperto tramite studi di associazione GWAS con i livelli di HbF (29). La sua espressione è attivata da KLF1 e tramite fenomeni di *splicing* alternativo del pre-mRNA esistono quattro differenti trascritti denominati XL (5,9 Kb; 125 KDa), L (3,8 Kb; 100 KDa), S (2,4 Kb; 35 KDa) e XS (1,5 Kb; 2,5 KDa) (30).

Oltre a ricoprire un ruolo essenziale nella linfopoiesi T e B è in grado di silenziare i geni  $\gamma$  globinici nella fase adulta legandosi, nei precursori eritroidi, all'HS3 dell'LCR e alla regione intergenica  $A\gamma$  - $\delta$  e interagendo con il complesso NuRD, in grado di

rimodellare la cromatina, e con altri fattori di trascrizione come GATA1 e FOG1. Recentemente è stata scoperta anche la cooperazione di BCL11A e SOX6 nello spegnimento dell'espressione della  $\gamma$  globina (FIG.10) (31, 32).

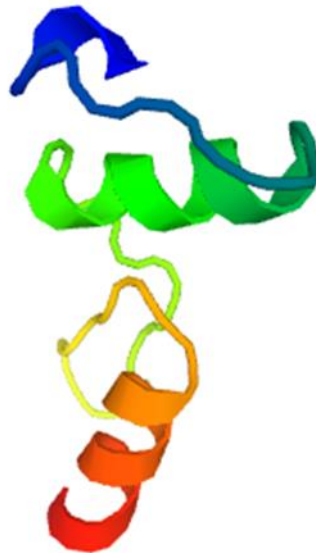


FIG.10: la struttura del fattore BCL11A.

Il modello corrente di *switching* globinico HbF→HbA mostra come l'espressione dei geni globinici dipenda dalla concentrazione di KLF1. Nello stadio fetale dello sviluppo eritroide o in presenza di una bassa concentrazione di KLF1 (aploinsufficienza), la bassa espressione di BCL11A non permette la repressione dei geni HBG (33). Klf1 in questo caso perciò andrà ad interagire preferenzialmente con i geni HBG, i quali sono stimolati dal reclutamento dell'LCR. Nello stadio adulto, in presenza di alte concentrazioni KLF1, BCL11A viene espresso ad alti livelli e di conseguenza i geni HBG

vengono repressi. In questo caso KLF1 legherà preferenzialmente i geni globinici adulti che verranno espressi a seguito del reclutamento dell'LCR (FIG.11) (34).

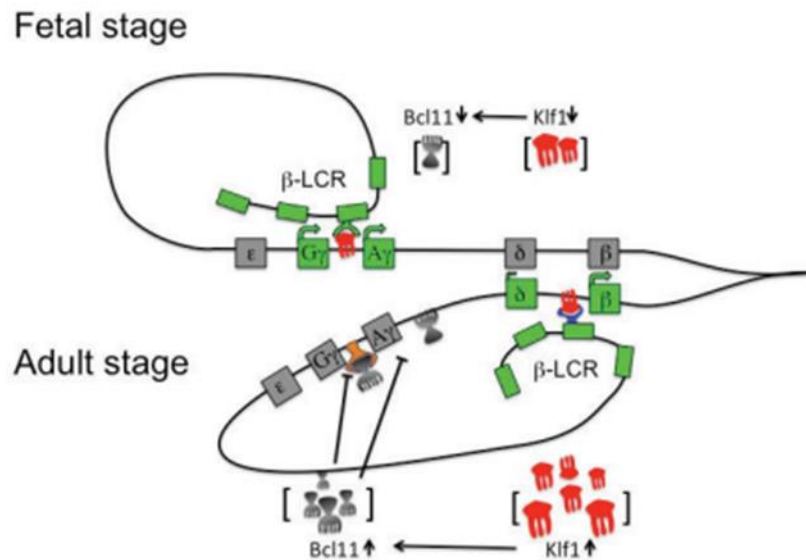


FIG.11: modello di *switching* globinico mediato da KLF1 e BCL11A.

SOX6 è un membro della famiglia di fattori di trascrizione Sry-related HMG box. Agisce legandosi al solco minore del DNA determinando un cambio conformazionale che promuove l'assemblaggio di complessi multiproteina biologicamente attivi (35). La sequenza consenso di tale fattore è (A/T)(A/T)CAA(A/T)G e può fungere da attivatore o repressore a seconda del tipo di proteine e promotori con cui interagisce (36, 37). A livello dei precursori eritroidi stimola la proliferazione e la loro maturazione finale. Inoltre, come già accennato, coopera con BCL11A nel silenziamento dei geni HBG, in quanto è stato osservato che Sox6 è in grado di legare fortemente i promotori dei geni HBG (38).

L'interazione proteina-proteina tra Sox6 e BCL11A può essere coinvolta nel *looping* della cromatina che avviene durante lo *switching* tra le sequenze regolatrici presenti nel cluster  $\beta$  (FIG.12).

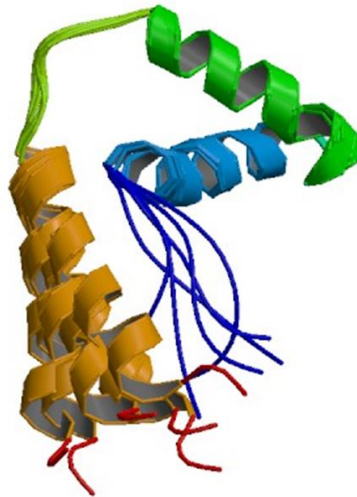


FIG.12: la struttura del fattore SOX6.

COUP-TFII è una proteina in grado di omodimerizzare e di legarsi a sequenza direttamente ripetute del motivo AGGTCA con diversi orientamenti (39). Nei precursori eritroidi sarebbe un candidato repressore dei geni HBE e HBG quando interagisce con le doppie regioni CCAAT box presenti nei promotori (40, 41). A validare tale ipotesi studi di *knockdown* hanno mostrato un aumento dei livelli di HbF e anomalie nel legame di COUP-TFII al promotore  $\gamma$  determina persistenza di HbF in età adulta (FIG.13) (42).

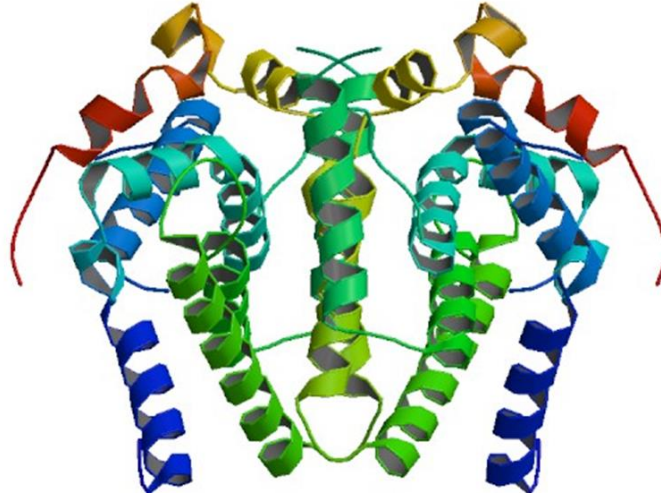


FIG.13: la struttura del fattore COUP-TFII.

La regione non codificante tra i geni HBS1L e MYB è stata individuata mediante studi GWAS volti all'identificazione di polimorfismi associati alla variabilità dell'espressione di HbF (43). HBS1L appartiene ai geni cosiddetti *housekeeping*, i quali sono espressi costitutivamente in tutte le cellule e la sua inibizione non determina nessuna variazione dell'espressione dei geni globinici. MYB invece è un fattore cruciale per lo sviluppo eritroide e studi di *knockdown* hanno mostrato un'associazione della riduzione di proteina MYB con l'aumento della sintesi di HbF. Mentre la sua over-espressione è correlata all'inibizione dell'espressione di  $\gamma$  globina (FIG.14) (44, 45).

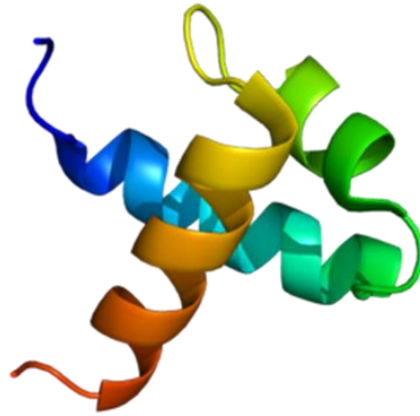


FIG.14: la struttura del fattore MYB.

Nella figura 15 è raffigurato il modello secondo il quale i principali fattori che regolano lo *switching* HbF→HbA interagiscono tra di loro e con gli elementi regolatori presenti nell'LCR e nei promotori dei geni globinici, affinché avvenga il graduale spegnimento dei geni HBG a favore dell'espressione dei geni globinici adulti (34, 38).

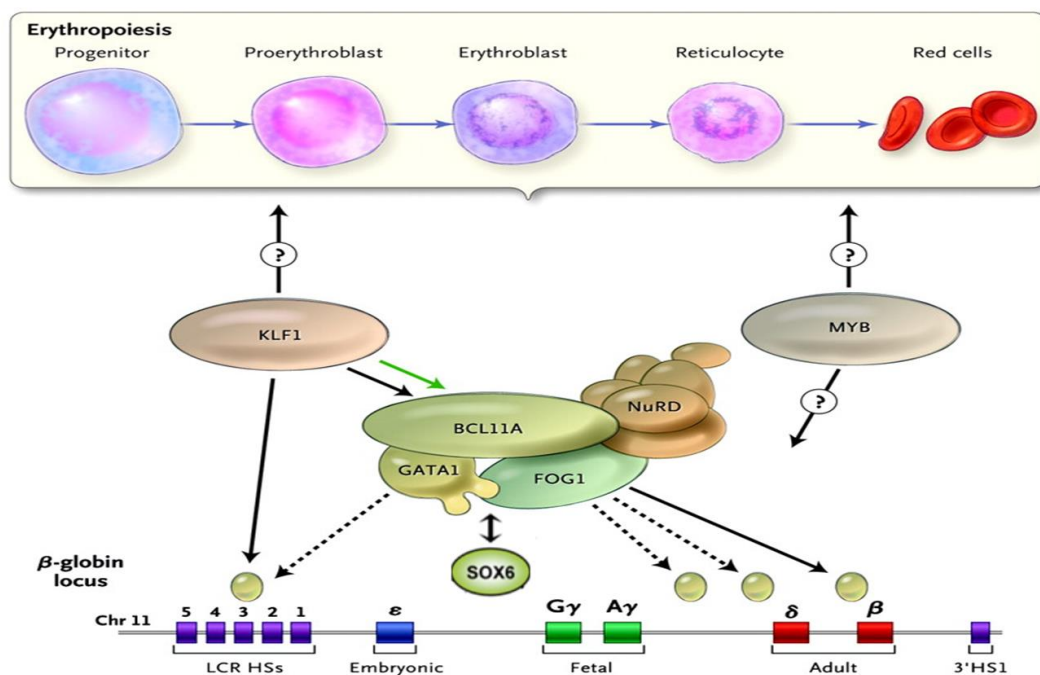


FIG.15: modello di silenziamento dei geni  $\gamma$  globinici e attivazione dei geni  $\beta$  e  $\delta$ .

“Terapia genica della  $\beta$  Talassemia mediante editing del DNA” - Università degli Studi di Cagliari  
Dot.ssa Maura Mingoa – Dottorato di Ricerca in “Scienze Pediatriche e Terapia dello Sviluppo”



## e. Le emoglobinopatie

Durante il differenziamento eritroide l'espressione di catene  $\alpha$  e  $\beta$  globiniche nel giusto rapporto è fondamentale per la sintesi di un HbA stabile. Squilibri o alterazioni di questi componenti a causa di mutazioni o delezioni in geni delle globine e/o dei loro regolatori possono essere deleteri per la fisiologia dei globuli rossi e dei loro precursori e possono determinare stati patologici che prendono il nome di emoglobinopatie (46). Le emoglobinopatie sono disordini caratterizzati da anomalie della struttura, della funzione o della produzione di emoglobina. Queste condizioni sono principalmente a carattere ereditario e le manifestazioni cliniche ad esse correlate possono essere molto differenti per gravità, che va da anomalie asintomatiche a quadri clinici severi quali la morte intrauterina del feto o anemie emolitiche gravi. Tale variabilità clinica è determinata dal genotipo, in quanto le forme eterozigoti risultano essere quasi sempre asintomatiche mentre le forme omozigoti e le doppie eterozigosi possono presentarsi con quadri clinico-ematologici piuttosto severi. Le emoglobinopatie costituiscono un problema di salute importante in più di 160 paesi nel mondo, ogni anno nascono oltre 330.000 bambini affetti da patologie a carico dell'emoglobina (CDC 2012). L'individuazione dei portatori di varianti emoglobiniche ha assunto oggi maggiore importanza a causa del fenomeno dell'immigrazione che sta modificando le caratteristiche della popolazione che afferrisce ai servizi sanitari. Inoltre lo screening per emoglobinopatie può essere

mirato ed effettuato nei familiari di soggetti portatori di emoglobinopatia. Le recenti acquisizioni, soprattutto in campo genetico, e la disponibilità di metodologie diagnostiche sempre più sofisticate ha permesso di approntare protocolli di *screening* per le emoglobinopatie al fine di selezionare, fra la popolazione che si sottopone ad indagini ematologiche di laboratorio, i soggetti per cui è necessario avviare alle indagini più complesse (47).

#### f. $\beta$ Talassemia

L' Emoglobinopatia più diffusa nei paesi mediterranei, tra cui l'Italia, è la  $\beta$  talassemia, una grave anemia ereditaria di tipo monogenico ed autosomico recessivo causata da mutazioni che determinano la riduzione ( $\beta^+$ ) o l'assenza ( $\beta^0$ ) di espressione del gene HBB (FIG.16). Tali mutazioni causano uno sbilanciamento del rapporto fisiologico di 1:1 esistente tra catene  $\alpha$  e non- $\alpha$  a livello di precursori eritroidi (48, 49). Poiché l'espressione fisiologica del gene HBA resta invariata, l'anomala sintesi di catene  $\beta$  in caso di talassemia comporta un eccesso di  $\alpha$  globina che non si trova coniugata nel complesso  $\alpha_2\beta_2$  della HbA. Questa condizione determina la conseguente formazione di aggregati molecolari di catene  $\alpha$  scarsamente solubili e la loro possibile precipitazione, la quale genera un danno ossidativo a carico delle membrane, già allo stadio di precursori eritroidi, i quali vanno incontro a prematura apoptosi con conseguente eritropoiesi inefficiente. Solo pochi proeritroblasti saranno dunque in

grado di completare la maturazione ed essere quindi immessi nel torrente circolatorio come eritrociti. Da queste alterazioni nella sintesi delle catene che costituiscono l'HbA ne deriva una riduzione del contenuto di Hb all'interno dei globuli rossi (anemia), che a seconda della gravità richiede la somministrazione più o meno frequente di trasfusioni di sangue associate ad una terapia ferro-chelante.

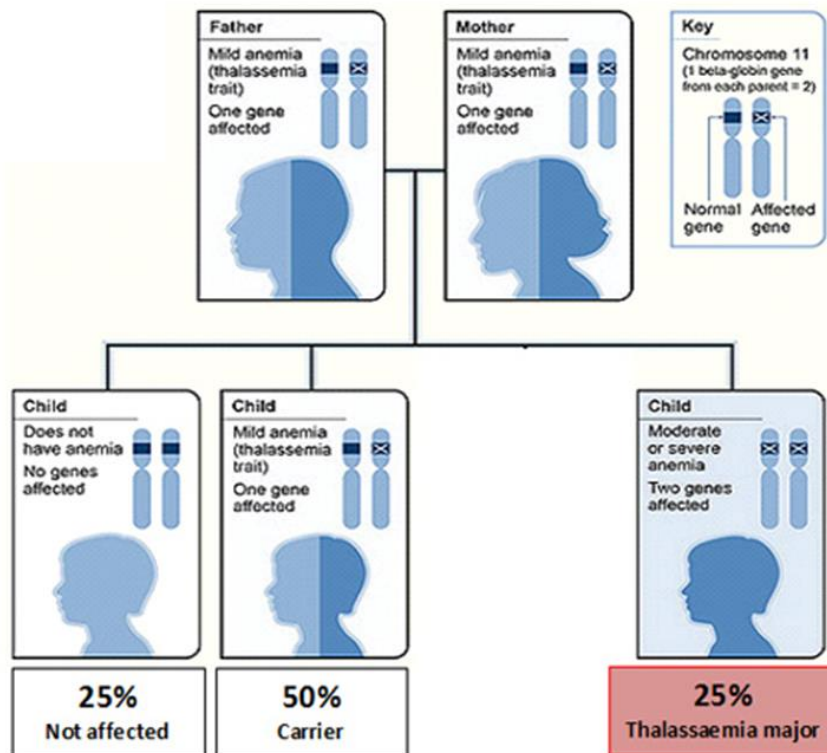


FIG.16: meccanismo di trasmissione ereditaria della  $\beta$  talassemia.

La diagnosi di  $\beta$  talassemia viene effettuata in primo luogo grazie all'osservazione clinica che viene successivamente confermata da analisi biochimiche sulla quantità e il tipo di catene globiniche presenti nel sangue (esame emocromocitometrico ed "Terapia genica della  $\beta$  Talassemia mediante editing del DNA" - Università degli Studi di Cagliari Dot.ssa Maura Mingoia – Dottorato di Ricerca in "Scienze Pediatriche e Terapia dello Sviluppo"

elettroforesi dell'emoglobina), accompagnate poi dalla conferma molecolare mediante l'analisi genetica per l'individuazione della precisa mutazione causa della patologia. L'incidenza annuale totale di individui sintomatici è ad oggi stimata in 1 su 100.000 in tutto il mondo e 1 persona su 10.000 nell'Unione Europea (FIG.17) (50).

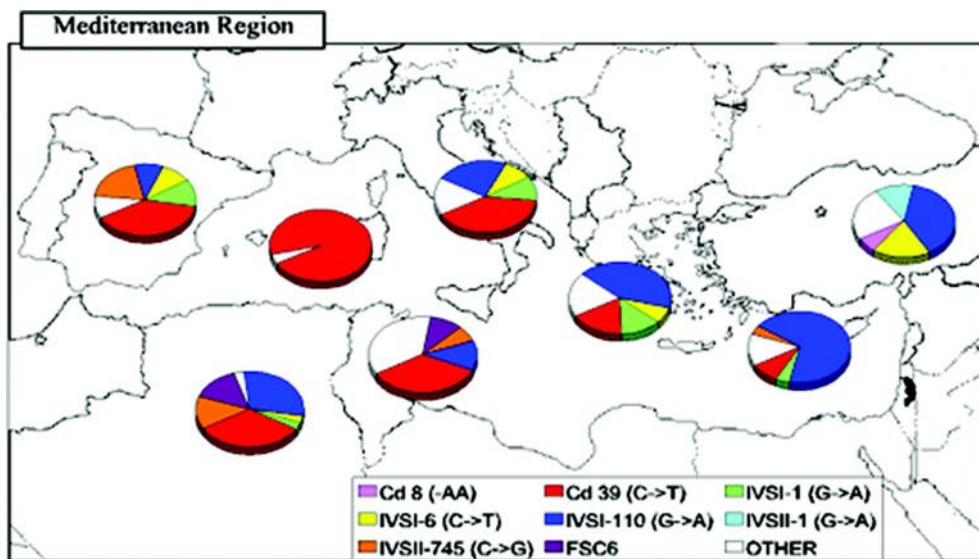


FIG.17: distribuzione geografica delle più comuni mutazioni  $\beta$  talassemiche nella regione mediterranea.

La riduzione dell'incidenza di questa anemia ereditaria è stata resa possibile grazie al programma di prevenzione, effettuato soprattutto negli ultimi trent'anni, volto all'identificazione dei portatori di mutazioni  $\beta$  talassemiche, la consultazione genetica e la diagnosi prenatale (FIG.18) (47).

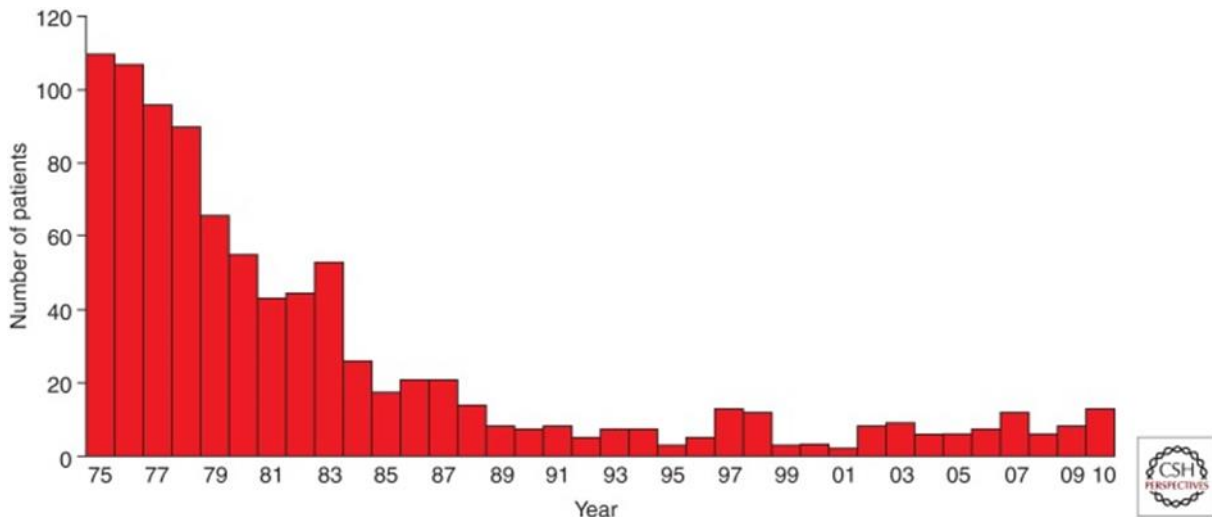


FIG.18: diminuzione del tasso di natalità di individui affetti da  $\beta$  talassemia in Sardegna dal 1975 al 2010.

#### i. Caratteristiche cliniche

In base alla severità del quadro clinico che gli individui con mutazioni  $\beta$  talassemiche manifestano è possibile riconoscere tre condizioni cliniche ed ematologiche differenti: lo stato di portatore di  $\beta$  talassemia, la talassemia major e la talassemia intermedia.

I portatori di  $\beta$  talassemia a livello genetico possiedono la mutazione patologica in eterozigosi e perciò risultano essere clinicamente asintomatici. Tuttavia mostrano alcune caratteristiche ematologiche, quali: microcitosi, ipocromia e variazione nel pattern di Hb presente all'interno degli eritrociti (92-95% HbA, >3.8 HbA<sub>2</sub> e una quantità variabile di HbF 0.5-4%).

Gli individui affetti da talassemia major (detta anche Morbo di Cooley o anemia mediterranea) di solito presentano entro i primi due anni di vita un'anemia grave trasfusione-dipendente associata ad un aumento del numero dei globuli rossi, il volume corpuscolare medio (MCV) e l'Hb corpuscolare media (MCH) bassi. Tale situazione è generalmente associata alla condizione di omozigosi o di composti eterozigoti per le mutazioni  $\beta$  talassemiche. Il profilo emoglobinico è variabile a seconda del tipo di  $\beta$  talassemia. Nella  $\beta^0$  talassemia l'HbA è assente, l'HbF rappresenta il 95-98% del totale, mentre l'HbA<sub>2</sub> raggiunge il 2-5%. Nella  $\beta^+$  talassemia o nei composti eterozigoti  $\beta^0/\beta^+$  l'HbA è sintetizzata tra il 10 e il 30%, l'HbF è nell'ordine del 70-90% e l'HbA<sub>2</sub> del 2-5%. Se non trattati gli individui affetti da talassemia major mostrano ritardo nella crescita, pallore, ittero, epato-splenomegalia, ulcere alle gambe, ematopoiesi extramidollare, modifiche scheletriche derivate dall'espansione del midollo osseo a causa dell'eritropoiesi inefficace fino alla morte che sopraggiunge in età infantile.

I pazienti con talassemia intermedia mostrano una moderata anemia e un quadro ematologico eterogeneo che a seconda dei casi può essere simile allo stato di portatore sano come alla condizione di talassemia major. Tale variabilità è generalmente associata alla presenza di mutazioni  $\beta$  talassemiche lievi o silenti in omozigosi o in coeredità con mutazioni severe oppure alla coeredità di fattori genetici in grado di modificare e migliorare il quadro clinico. Infatti la coeredità con mutazioni

$\alpha$  talassemiche oppure con mutazioni in grado di sostenere una continua espressione di catene  $\gamma$  nella vita adulta determina la riduzione dello sbilanciamento tra le catene  $\alpha$ /non- $\alpha$ , dal quale dipende la severità clinica di questa anemia ereditaria (50). Tali forme di  $\beta$  talassemia si verificano in presenza di uno stato di omozigosi per delle particolari mutazioni lievi, come ad esempio la IVS-1-6 (T→C) e quelle a carico del  $\beta$  promoter -88 (C→T) e -29 (A→G) (51-54).

## ii. Genetica molecolare della $\beta$ talassemia

La  $\beta$  talassemia è una malattia molto eterogenea a livello molecolare e finora sono state identificate più di 200 mutazioni che ne determinano l'insorgenza (55, 56). Tali mutazioni possono essere suddivise a loro volta in tre differenti categorie: mutazioni trascrizionali (a carico della regione promotrice, del sito CAP o della 5' UTR); mutazioni che interferiscono sulla maturazione dell'RNA messaggero (difetti di splicing e mutazioni nelle regioni di poliadenilazione e 3' UTR); mutazioni che portano a difetti nella traduzione dell'mRNA. La maggior parte delle mutazioni appartenenti a quest'ultima categoria è determinata da mutazioni puntiformi, ossia delezioni o inserzioni di nucleotidi che portano a frameshift o da singole sostituzioni nucleotidiche invariati o nonsenso a livello del gene  $\beta$  globinico. Raramente le  $\beta$  talassemie sono causate da delezioni geniche, condizione invece più comune per quanto riguarda le  $\alpha$  talassemie. Tuttavia alcune delezioni possono coinvolgere oltre al gene HBB anche il gene  $\delta$  ( $\delta$ - $\beta^0$  talassemia) oppure sia il gene HBD che quelli HBG

(G $\gamma$ -A $\gamma$ - $\delta$ - $\beta^0$  talassemia). Anche la delezione parziale o totale della regione regolatrice LCR può inattivare l'espressione della  $\beta$  globina nonostante il gene HBB sia intatto. In generale le  $\beta^0$  talassemie sono determinate da delezioni e mutazioni a livello del codone iniziale, mutazioni nonsense, che determinano frameshift o che comportano difetti nello splicing. Mentre le  $\beta^+$  talassemie si manifestano soprattutto in presenza di mutazioni nel promotore, nella regione del sito poliA e nelle 5' o 3' UTR o per difetti di splicing. Le mutazioni  $\beta^+$  possono essere classificate a loro volta come severe, lievi e silenti a seconda del quadro ematologico che da esse ne deriva (TAB.1) (57, 58).

POPULATION	B-GENE MUTATION	SEVERITY
INDIAN	-619 del	$\beta^0$
MEDITERRANEAN	-101 CTT	$\beta^{++}$
BLACK	-88 CTT	$\beta^{++}$
MEDITERRANEAN; AFRICAN	-87 CTG	$\beta^{++}$
JAPANESE	-31 ATG	$\beta^{++}$
AFRICAN	-29 ATG	$\beta^{++}$
SOUTHEAST ASIAN	-28 ATC	$\beta^{++}$
MEDITERRANEAN; ASIAN INDIAN	IVS1-nt1 GTA	$\beta^0$
EAST ASIAN; ASIAN INDIAN	IVS1-nt5 GTC	$\beta^0$
MEDITERRANEAN	IVS1-nt6 TTC	$\beta^{+}/^{++}$



<b>MEDITERRANEAN</b>	IVS1-nt110 GTA	$\beta^+$
<b>CHINESE</b>	IVS2-nt654 CTT	$\beta^+$
<b>MEDITERRANEAN</b>	IVS2-nt745 CTG	$\beta^+$
<b>MEDITERRANEAN</b>	codon 39 CTT	$\beta^0$
<b>MEDITERRANEAN</b>	codon 5 -CT	$\beta^0$
<b>MEDITERRANEAN; AFRICAN-AMERICAN</b>	codon 6 -A	$\beta^0$
<b>SOUTHEAST ASIAN</b>	codon 41/42 -TTCT	$\beta^0$
<b>AFRICAN-AMERICAN</b>	AATAAA to AACAAA	$\beta^{++}$
<b>MEDITERRANEAN</b>	AATAAA to AATGAA	$\beta^{++}$
<b>MEDITERRANEAN</b>	codon 27 GTT Hb (Hb Knossos)	$\beta^{++}$
<b>SOUTHEAST ASIAN</b>	codon 79 G>A (Hb E)	$\beta^{++}$
<b>MALAYSIA</b>	Codon 19 G>A (Hb Malay)	

$\beta^0$ :complete absence of beta globin on the affected allele

$\beta^+$ :residual production of beta globin (around 10%)

$\beta^{++}$ :very mild reduction in beta globin production

TAB. 1: le mutazioni  $\beta$  talassemiche più comuni.

Le diverse mutazioni  $\beta$  talassemiche identificate sembrerebbero essersi originate indipendentemente nelle varie popolazioni e siano state poi selezionate positivamente a causa della protezione relativa degli individui eterozigoti verso la

malaria da *Plasmodium falciparum*. Per tale motivo la popolazione Sarda possiede una frequenza di portatori del circa il 12,6%, rappresentando la maggiore frequenza di geni talassemici tra la popolazione caucasica (59). In Sardegna, la malattia è determinata per il 96% dei casi da un'unica mutazione nonsense a livello del codone 39 (E39X) dell'esone 2 che causa l'interruzione della sintesi della  $\beta$  globina. Nello specifico si tratta di una transizione C>T in posizione 118, che comporta una modifica della tripletta codificante per l'aminoacido Glutamina (CAG) in un segnale di stop (TAG→UAG nell'RNA) determinando così la sintesi di una proteina tronca e non funzionale (60).

### iii. Modificatori genetici

Come già accennato, la coeredità di modificatori genetici in grado di ridurre lo sbilanciamento del rapporto tra la sintesi di catene  $\alpha$  e non- $\alpha$  comporta un miglioramento del fenotipo  $\beta$  talassemico. Il quadro ematologico che deriva da tale condizione risulta essere notevolmente variabile e non prevedibile, in quanto può andare dal fenotipo normale, o moderatamente simile, a quello talassemico.

- Talassemia  $\beta^+$ : un esempio di determinante genetico capace di migliorare il fenotipo è rappresentato dalle mutazioni  $\beta$  talasemiche lievi o silenti, le quali sono associate al mantenimento di un'espressione residua del gene HBB accettabile. Anche la coeredità in eterozigosi di mutazioni lievi differenti o di
- “Terapia genica della  $\beta$  Talassemia mediante editing del DNA” - Università degli Studi di Cagliari  
Dot.ssa Maura Mingoia – Dottorato di Ricerca in “Scienze Pediatriche e Terapia dello Sviluppo”*

una mutazione lieve con una severa può dar luogo a forme meno gravi che richiedono solo qualche sporadica trasfusione (TAB.2) (34, 50).

### MUTAZIONI LIEVI E SILENTI NEL GENE HBB

TIPO DI MUTAZIONE O LOCALIZZAZIONE	Talassemia intermedia $\beta^+$	Silente
<b>MUTAZIONI NEL CACC BOX PROSSIMALE</b>	-90 C>T -88 C>T -88 C>A -87 C>T -87 C>A -86 C>T -86 C>G	- 101 C>T -92 C>T
<b>TATA BOX</b>	-31 A>G -30 T>A -29 A>G	
<b>5' UTR</b>	+22 G>A +10 -T +33 C>G	+1' A>C
<b>SPLICING ALTERNATIVO</b>	Cd19 A>C (Hb Malay) Cd24 T>A	Cd27 G>T (Hb Knossos)

<b>SPLICING CONSENSO</b>	IVS-6 T>C	
<b>INTRONE</b>	IVS2-844 C>G	
<b>3'UTR</b>	+6 C>G	
<b>SITO POLY-A</b>	AACAAA	AATAAG
	AATGAA	
<b>LIEVE BETA<sup>0</sup> – FRAMESHIFT</b>	cd6-AA	
	cd8-AA	

TAB. 2: le mutazioni  $\beta^+$  talassemiche più comuni.

- $\alpha$  talassemia: un altro modificatore genetico che risulta in manifestazioni cliniche attenuate è la coeredità di geni che determinano  $\alpha$  talassemia. Questa condizione è stata descritta nelle popolazioni in cui è presente un'elevata incidenza sia di geni  $\beta$  che  $\alpha$  talassemici. Le  $\alpha$  talassemie sono causate dalla produzione difettosa o assente di catene  $\alpha$ , determinato generalmente da una delezione di segmenti di dimensione variabile a livello dei *loci*  $\alpha$ . Il genotipo normale ha un assetto a quattro geni ( $\alpha\alpha/\alpha\alpha$ ), due su ciascun cromosoma 16. La forma più lieve di  $\alpha$ -tal, l' $\alpha$ 2-tal ( $-\alpha/\alpha\alpha$ ), è determinata dalla delezione (o

altra anomalia) di uno dei quattro *loci*  $\alpha$  ed è, nel complesso, asintomatica. L' $\alpha$ 1-tal ( $-\alpha/-\alpha$  o  $--/\alpha\alpha$ , a seconda che i geni  $\alpha$  mutati siano *in cis* o *in trans* rispetto al cromosoma) deriva dalla compromissione di due geni su quattro. In Sardegna il 33% della popolazione è portatore silente ( $-\alpha/\alpha\alpha$ ), l'1,5% doppio eterozigote ( $-\alpha/-\alpha$ ), e lo 0,8% e portatori del tipo  $--/\alpha\alpha$ . Un genotipo più grave di  $\alpha$  talassemia è rappresentato dalla presenza della delezione patologica in 3 dei 4 geni  $\alpha$  ( $--/-\alpha$ ) caratterizzato da una notevole riduzione della sintesi di  $\alpha$  globina e conseguente eccesso di catene  $\beta$ , condizione denominata "Malattia da HbH". Nel contesto  $\beta$  talassemico è stato osservato che la coeredità con forme  $\alpha$ 2-tal non modifica la severità clinica della malattia, mentre la presenza di due o tre geni HBA non funzionanti determina un notevole miglioramento fenotipico soprattutto in concomitanza con una mutazione del tipo  $\beta^+$  (49) (FIG.19).

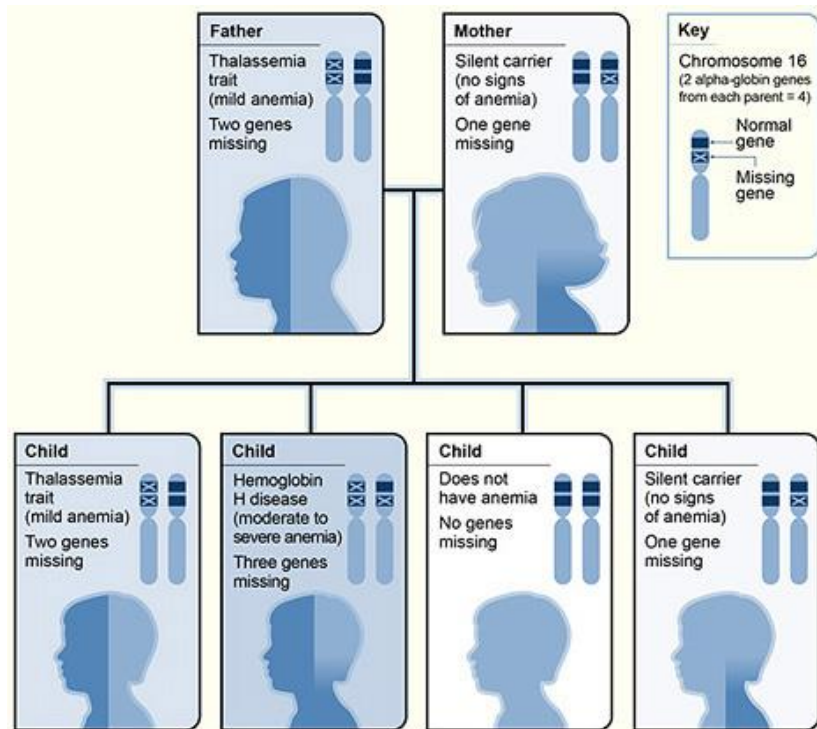


FIG.19: meccanismo di trasmissione ereditaria dell'  $\alpha$  talassemia.

- Persistenza della sintesi di HbF: un terzo meccanismo molecolare in grado di modificare il quadro patologico della  $\beta$ -talassemia è rappresentato dalla coeredità di un determinante genico in grado di sostenere la produzione continuativa di catene  $\gamma$  durante la vita adulta. Livelli incrementati di HbF nella vita adulta possono essere dovuti ad effetti primari o secondari di patologie acquisite o ereditarie. Le condizioni ereditarie, contraddistinte da un'elevata espressione di HbF, possono essere causate da ampie delezioni che determinano  $\delta/\beta$  talassemia o da delezioni/mutazioni a carico delle regioni regolatrici dei geni globinici. In altri casi può essere dovuto alla co-trasmissione di specifici determinanti per la persistenza ereditaria di emoglobina fetale

(HPFH). Il meccanismo per cui le  $\delta/\beta$ talassemie e le HPFH determinano la produzione di livelli elevati di  $\gamma$  globina non è ancora stato compreso nei dettagli. Tuttavia sono stati ipotizzati diversi eventuali meccanismi come la delezione di sequenze regolatrici del cluster  $\beta$ , che agiscono come elementi regolatori positivi o negativi. Il fenotipo risultante deriverebbe da una riorganizzazione delle sequenze regolatrici che restano funzionali all'interno del cluster; la delezione impedisce il funzionamento di *enhancer* normalmente localizzati a valle del gene HBB. Ciò favorirebbe un incremento della espressione dei geni fetali; la delezione del gene HBB permette all'LCR di continuare a interagire con i geni HBG normalmente quiescenti già in età post-natale. Tuttavia, un unico meccanismo non può spiegare i differenti fenotipi associati alla presenza di tali modificatori. Si ritiene che tali meccanismi non siano esclusivi di un particolare tipo di delezione e che il loro contributo possa differire da caso a caso relativamente al particolare tipo di delezione (61). Le  $\delta/\beta$  talassemie sono caratterizzate da una ridotta o addirittura assente espressione dei geni HBD e HBB. Questa malattia genetica è determinata da delezioni più o meno estese degli elementi che costituiscono il cluster  $\beta$ , le quali coinvolgono i geni  $\delta$  e  $\beta$  globinici [ $G\gamma A\gamma$  ( $\delta\beta$ )<sup>0</sup>-talassemia] e anche il gene  $A\gamma$  [ $G\gamma$  ( $A\gamma\delta\beta$ )<sup>0</sup>-talassemia]. L'effetto associato a questa condizione patologica per via dell'esclusiva interazione dell'LCR con i promotori dei geni HBG è la persistenza della sintesi di  $\gamma$  globina con conseguente aumento dei livelli di HbF. Individui

eterozigoti per mutazioni del tipo  $G\gamma A\gamma (\delta\beta)^0$  talassemia presentano livelli variabili di HbF compresi tra 4,0% e il 18,6%, mentre coloro che possiedono mutazioni  $G\gamma (A\gamma\delta\beta)^0$  talassemiche mostrano concentrazioni di HbF che vanno dal 9,3% al 25% (TAB.3) (62).

Mutation type	Ethnicity	% of Hb F <sup>(16)</sup>
$\gamma^G(\gamma^A\delta\beta)^0$	Belgian	14.0 % - 23.0%
	Malay-1	25%
	Black	17.3% - 21.6%**
	Chinese	9.3% - 15.7%
	German	9.9% - 12.5%
	Turkish	10.0% - 13.0%
$\gamma^G(\gamma^A\delta\beta)^0$	Macedonian	6.7% - 12.8%
	Thai	9.9%
	East European	13.0% - 18.0%
	Sicilian	4.0% - 18.6%
	Black	24.8%**
	Spanish	5.0% - 13.0%
	Japanese	7.8%
	Turkish	6.0% - 15.0%

\*\* Mutations associated with Hb S

TAB.3: le mutazioni  $\delta/\beta$  talassemiche più comuni.

Nell'adulto normale la quota di HbF supera raramente l'1% dell'Hb totale. La sintesi delle globine fetali ( $G\gamma$  e  $A\gamma$ ) è confinata ad una linea di eritrociti, chiamate cellule F, il cui numero è geneticamente determinato (63). L'HPFH è una condizione benigna caratterizzata da un aumento sostanziale dell'espressione di catene  $\gamma$ , determinando di conseguenza la persistenza di alti livelli di HbF e cellule F in età adulta (64). Rappresenta un gruppo molto



eterogeneo di condizioni benigne caratterizzate, in eterozigosi, da livelli di HbF compresi fra il 4 e il 30% con indici globulari ed di ematocrito vicini al normale. Le HPFH possono essere suddivise in due categorie differenti a seconda alla distribuzione di HbF nei globuli rossi, che può essere pancellulare (distribuzione uniforme) o eterocellulare (distribuzione eterogenea). Le HPFH si distinguono in HPFH non-delezionali (ndHPFH) e HPFH da delezione (dHPFH) a seconda che siano causate da mutazioni puntiformi oppure da delezioni più o meno estese a carico dei geni HBG e/o di altre porzioni del cluster  $\beta$  globinico coinvolte nella regolazione dell'espressione dei geni fetali (65). Le ndHPFH sono generalmente determinate da mutazioni puntiformi a livello della regione promotrice dei entrambi i geni HBG, la maggior parte delle quali sono localizzate intorno alle posizioni -114,-175 and -200 rispetto al sito Cap. È stato osservato come alcune di mutazioni puntiformi modificano elementi regolatori cis-acting alterando il sito di legame per i vari fattori trascrizionali *trans-acting*, eritroidi specifici o ubiquitari (66) (TAB.4).

Table 1 - Mutations associated with high Hb F levels, ethnic groups and Hb F levels		
Mutation type	Mutation	% of Hb F <sup>16</sup>
Single point mutation	- 110 (A → C) $\gamma^G$ (Czech)	0.6% - 0.8%
	- 114 (C → G) $\gamma^G$ (Australian)	8.6%
	- 114 (C → T) $\gamma^G$ (Japanese)	11.0% - 14.0%
	- 175 (T → C) $\gamma^G$	22.1%
	- 202 (C → G) $\gamma^G$	18.0% - 23.5%
	- 114 (C → T) $\gamma^A$ (Georgia)	4.7% - 4.8%
	- 117 (G → A) $\gamma^A$ (Black-Greek)	10.9% - 15.9%
	- 117 (G → A) $\gamma^A$ (Greek-Italian)	7.1% - 19.0%
	- 158 (C → T) $\gamma^A$ (Greek)	2.9% - 5.1%
	- 175 (T → C) $\gamma^A$ (Black)	36.7% - 38.5%
	- 195 (C → G) $\gamma^A$ (Brazilian)	4.5% - 7.0%
	- 196 (C → T) $\gamma^A$ (Italian)	12.0% - 16.0%
	- 198 (T → C) $\gamma^A$ (British)	1.8% - 13.0%
	- 202 (C → 'T) $\gamma^A$	1.6% - 3.9%
Deletional	HPFH-1 (Black)	21.4% - 28.2%
	HPFH-2 (Ghanaian)	21.6% - 27.2%
	HPFH-3 (Indian)	21.6% - 23.6%
	HPFH-4 (Italian)	21.0% - 30.0%
	HPFH-5 (Italian)	16.0% - 20.0%
	HPFH-6 (Thai)	17.2% - 20.0%

$\gamma^G$ : gamma glycine gene;  $\gamma^A$ : gamma alanine gene; HPFH: Hereditary Persistence of Fetal Hemoglobin

TAB.4: le mutazioni e delezioni HPFH più comuni.

E' ipotizzabile che tali mutazioni HPFH possano ridurre l'affinità di legame di un fattore di repressione per la specifica sequenza consenso oppure potrebbero favorire il legame di fattori trascrizionali attivatori che normalmente non si legherebbero al promotore HBG in ambiente eritroide adulto (61, 67, 68, 69).

## g. Le terapie

Negli ultimi anni le terapie per la  $\beta$  talassemia hanno fatto passi da gigante. Progressi continui nella terapia emo-trasfusionale e nella terapia ferro-chelante hanno portato notevoli miglioramenti sia nella durata che nella qualità di vita dei soggetti talassemici, così che oggi tale malattia può essere considerata a “prognosi aperta”. Ciononostante la malattia impone tuttora continui sacrifici ai pazienti e ai familiari e rappresenta un carico pesante per la società e il sistema di assistenza sanitaria nazionale. Al momento l'unica arma terapeutica in grado di risolvere alla radice la malattia è rappresentata dal trapianto allogenico di cellule staminali ematopoietiche (CSE), il quale è limitato dalla disponibilità di donatori di midollo osseo HLA compatibili (FIG.20). Questa pratica terapeutica offre buone prospettive di guarigione soprattutto per i pazienti che ricevono il trapianto da un fratello donatore (70, 71). Purtroppo c'è solo una possibilità su quattro che un dato fratello abbia un sistema di istocompatibilità identico, con una probabilità per un paziente talassemico di avere un donatore compatibile variabile in rapporto al numero dei componenti della famiglia. Poiché la maggior parte dei pazienti talassemici non ha un fratello compatibile, si sta cercando di utilizzare donatori non familiari ma compatibili. Sfortunatamente, le complicanze legate al trapianto usando donatori compatibili non familiari sono generalmente molto più alte di quelle riscontrate con i fratelli compatibili.

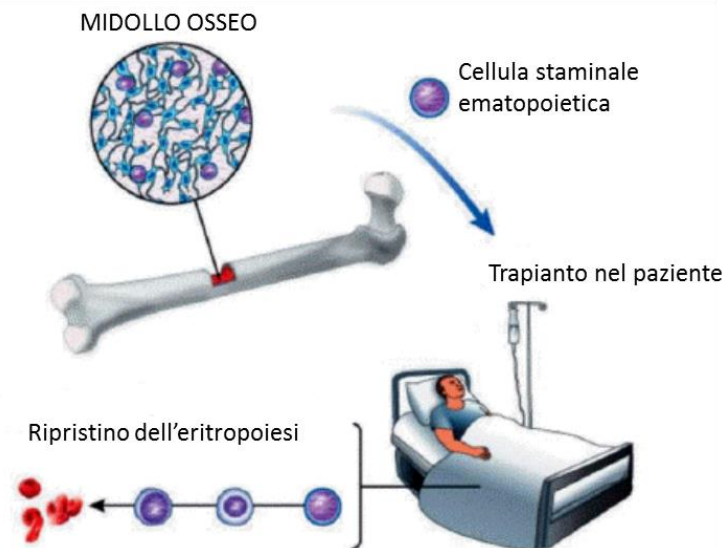


FIG.20: rappresentazione schematica del trapianto allogenico di CSE.

Tuttavia, il 60% degli individui affetti da talassemia è privo di un donatore correlato o non correlato e perciò non ha attualmente speranze di guarigione definitiva. Questa categoria di pazienti ripone ogni speranza di cura definitiva nello sviluppo di terapie alternative.

Tra questi nuovi approcci terapeutici, attualmente la terapia genica basata sul trapianto autologo di CSE geneticamente corrette rappresenta la più promettente alternativa alle terapie convenzionali.

Lo sviluppo di vettori lentivirali (LVs) in grado di integrare una nuova copia del gene HBB normale nel DNA di cellule staminali e progenitrici ematopoietiche specifiche del paziente e l'ottimizzazione delle condizioni di trasduzione ha portato alla loro recente applicazione in differenti trials clinici (72, 73, 74) (FIG.21).

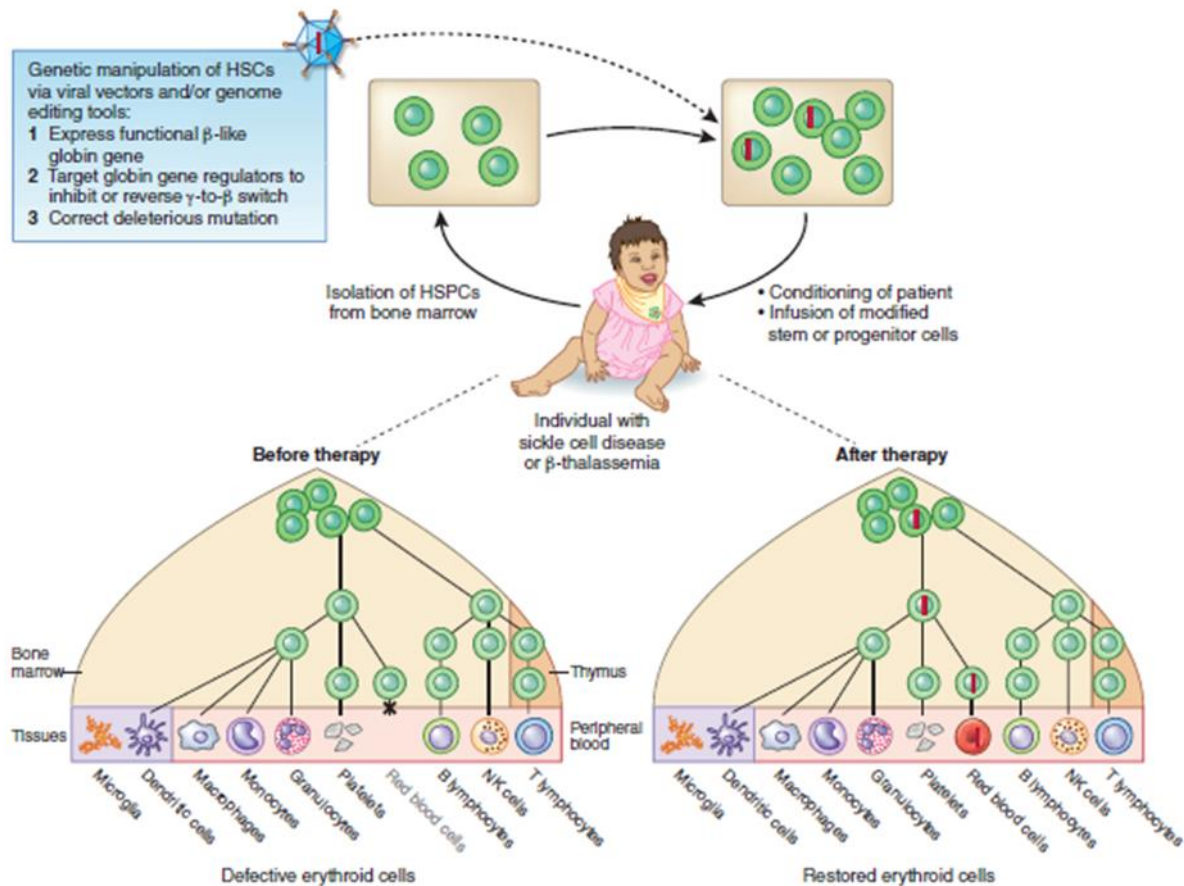


FIG.21: rappresentazione schematica del ripristino dell'espressione del gene HBB mediante LVs.

Il successo di tale integrazione dovrebbe portare i parametri ematologici dei pazienti trattati ai livelli normali, in maniera tale da non dover più richiedere trasfusioni di sangue. Nonostante ciò perché la terapia genica con LVs sia efficace e risolutiva i vettori utilizzati per garantire il ripristino della sintesi "de novo" della catena  $\beta$  globinica devono ancora affrontare delle sfide importanti per quanto riguarda la loro efficacia nell'integrazione del transgene e nella stabilità della sua espressione al fine di garantirne livelli terapeutici. In particolare la maggiore problematica concerne l'eventuale integrazione casuale del transgene in siti multipli del genoma dell'ospite

*"Terapia genica della  $\beta$  Talassemia mediante editing del DNA"* - Università degli Studi di Cagliari  
 Dot.ssa Maura Mingoia – Dottorato di Ricerca in "Scienze Pediatriche e Terapia dello Sviluppo"

che potrebbe portare all'attivazione o la repressione di geni fiancheggianti il sito di integrazione, alterando così il ciclo e la sopravvivenza delle cellule trasdotte fino alla loro possibile trasformazione in cellule cancerogene (75, 76).

L'approccio ideale per curare una malattia monogenica come la  $\beta$  talassemia, evitando le problematiche causate dall'inserimento random di un gene esogeno, sarebbe quello di correggere direttamente la mutazione patologica. Questa pratica prende il nome di *Gene-Editing*, un'innovativa biotecnologia in cui una sequenza specifica di DNA detta target viene modificata attraverso l'utilizzo di endonucleasi ingegnerizzate artificialmente (FIG.22).

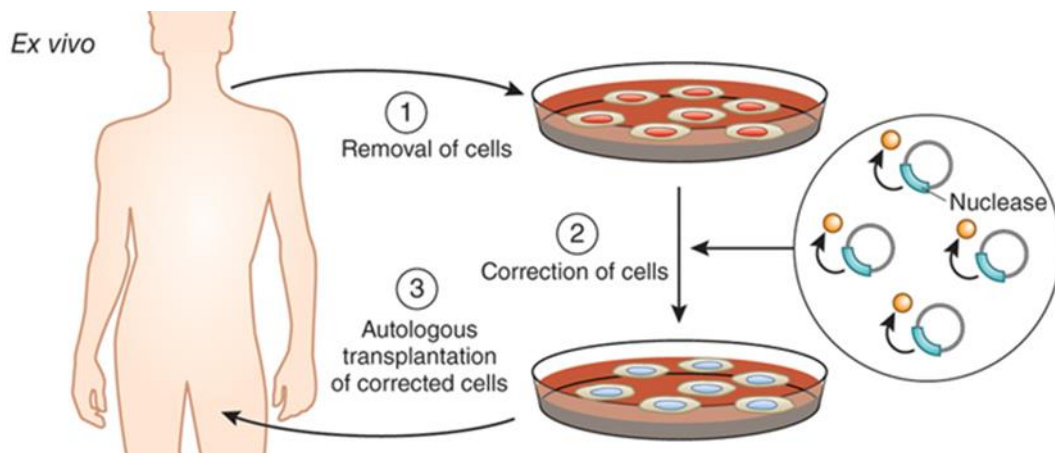


FIG.22: rappresentazione schematica della correzione di un difetto genetico mediante Gene-Editing.

Il principale vantaggio dell'utilizzo di tale approccio per il trattamento della  $\beta$  talassemia è quello di evitare sia i problemi di rigetto nel caso del trapianto allogenico sia la possibilità di mutazioni associate all'utilizzo dei vettori lentivirali, poichè non

necessita dell'integrazione di geni esogeni all'interno del genoma, rendendo così le cellule geneticamente corrette più sicure per la salute dello stesso paziente.

Sono disponibili diverse tecnologie in grado di compiere il *Gene-Editing*, tra queste quella basata sul sistema CRISPR/Cas9 ha dimostrato di essere realmente all'altezza delle aspettative in essa riposte grazie ai risultati straordinari ottenuti in breve tempo, promettendone in futuro degli altri ancor più significativi. Questa tecnologia è stata infatti ideata soltanto nel 2013 negli Stati Uniti dal gruppo di ricerca guidato da F. Zhang ed è stata testata su diversi tipi cellulari e organismi con un'efficienza di *editing* mai ottenuta prima e proprio per questo motivo è stata inserita al primo posto nella TOP TEN della scienza 2015, classifica stilata ogni anno dalla prestigiosa rivista scientifica *Science* (77, 78, 79). Il meccanismo d'azione di questo sistema prevede l'impiego di un RNA chimerico chiamato RNA guida (gRNA) in grado di appaiarsi con una determinata sequenza *target*, in corrispondenza o vicino al sito della mutazione, e di richiamare la Nucleasi CRISPR-associata (Cas) che permette di programmare in questo punto specifico un taglio dando luogo all'*editing* mediante l'innescò dei sistemi di riparo cellulari. Nello specifico, la correzione del difetto genetico avviene mediante il processo cellulare di ricombinazione omologa (HR) favorito dal DSB, quando al sistema CRISPR/Cas9 viene associato un DNA donatore che possiede delle sequenze corrette omologhe alla sequenza bersaglio, sede della mutazione. Tale correzione è quindi in grado di ripristinare fisiologicamente l'espressione del gene

HBB e la sintesi di una HbA funzionale nelle cellule staminali pluripotenti indotte (iPSCs) specifiche del paziente che potrebbero essere differenziate in CSE per l'eventuale trapianto autologo. Inoltre con l'ottenimento di una elevata percentuale di HR nelle cellule trattate si conta di riuscire a convertire, nel caso della  $\beta$  talassemia come altre malattie monogeniche, lo stato di omozigosi per il difetto genetico responsabile della malattia in uno stato di eterozigosi nella maggior parte delle cellule trattate, condizione più simile allo stato di portatore sano trasfusione indipendente (80).

## 2.2 Il *Genome Editing*

L'*editing* del DNA è un meccanismo fisiologico di revisione e correzione degli errori che possono verificarsi durante la duplicazione del genoma. In ingegneria genetica il termine si riferisce ad un'innovativa biotecnologia in cui una sequenza specifica di DNA può essere mutata, inserita, sostituita o rimossa attraverso l'utilizzo nucleasi ingegnerizzate artificialmente. Sono disponibili diverse tipologie di proteine chimeriche in grado di compiere il DNA *editing*:

- le Meganuclease: versioni ingegnerizzate degli enzimi di restrizione in grado di riconoscere sequenze di DNA estese (14-40 bp) e di tagliare l'estremità della sequenza target in maniera specifica (81).



- le Zn finger nuclease (ZFNs) costituite dalla fusione di un dominio di legame al DNA del tipo Zn finger (Zf) con un dominio ad attività nucleasica aspecifica derivata dall'enzima di restrizione FokI. Questa metodica si basa sulla capacità del motivo Zf nel riconoscere una determinata tripletta nucleotidica; mediante l'inserimento in tandem da 3 a 6 motivi Zf è possibile ottenere proteine chimeriche ZNFs in grado di legare una specifica sequenza target (82). La prima tecnologia utilizzata nell'*editing* del DNA nel 2012 (83), ma che tuttavia risulta molto complicata da utilizzare e manipolare.

- le TALENs similmente alle ZFNs possiedono il dominio nucleasico di FokI fuso con DNA *binding domain* costituito da dei motivi TALE ripetuti ciascuno dei quali è costituito da 33-35 aa in cui due residui detti ipervariabili le conferiscono la specificità di legame con una diversa base azotata. Le TALENs sono generate attraverso la modifica di queste regioni ipervariabili e concatenando i motivi TALE in successione secondo la sequenza target. Per via della elevata specificità del DNA *binding domain* gli eventi di off targets sono rari, tuttavia il dominio nucleasico è aspecifico. È perciò una tecnologia molto dispendiosa e difficile da ingegnerizzare (84).

- le Mega-TALs rappresentano una combinazione di due tecnologie, in quanto possiedono il DNA binding domain del tipo TALENs e il dominio nucleasico specifico proveniente da una meganucleasi. Essendo una nucleasi sito specifica dovrebbe

avere una percentuale di off targets più bassa delle TALENs con il dominio aspecifico fok, tuttavia la loro efficacia resta ancora da testare (85).

Tutte e tre le tecnologie lavorano come dimero in quanto ogni dominio nucleasico è in grado di tagliare un solo filamento di DNA, ne consegue la necessità di costruire due geni codificanti per ciascuna di queste proteine chimera.

- la CRISPR/Cas, che consiste in una endonucleasi RNA guidata molto efficace e soprattutto di facile sintesi rispetto alle precedenti, è risultata ad oggi la tecnologia con maggior efficienza di *editing* su diversi tipi cellulari e organismi, presentando però anche una maggiore percentuale di off targets rispetto alle TALENs. Inoltre, questa è l'unica biotecnologia in grado di operare sotto forma di monomero, grazie alla presenza di due domini nucleasici ciascuno deputato al taglio di un filamento del DNA. L'utilizzo di diversi tipi di RNA guida rende possibile inoltre ottenere un *multiplex genome editing* su diversi *loci* contemporaneamente (86).

Nella TAB.5 sono riassunte le differenze tra tutti i sistemi in grado di effettuare l'Editing del DNA in base al tipo di riconoscimento della sequenza target, sensibilità alla metilazione, la modalità di legame, il tipo di sito target riconosciuto, la possibilità di *multiplex editing* e le diverse problematiche a cui si va in contro a seconda del sistema scelto.

Proprietà	Meganucleases	ZFNs	TALENs	CRISPR/Cas9
Tipo di riconoscimento	Proteina-DNA	Proteina-DNA	Proteina-DNA	RNA-DNA
Metilazione	Sensibile	Sensibile	Sensibile	Non sensibile
Modalità di legame	Combinazione di DNA binding domains di diverse endonucleasi	Un ZF lega 3-4 nucleotidi	Un RDV riconosce un nucleotide	Un nucleotide si appaia con un nucleotide
Sito Target	2 x 14- 40 nucleotidi	2 x 12 nucleotidi o più	2 x 16 nucleotidi o più	20 nucleotidi
Multiplexing	No	No	No	Si
Problematiche	Costose e sequenze target limitate	Difficili da ingegnerizzare e costose	Difficili da ingegnerizzare e costose	Molti potenziali off-targets

TAB.5: le diverse tecnologie per il Gene-Editing a confronto.

### a. Il sistema CRISPR/Cas

Per via della sua semplicità nella progettazione, la sua efficienza e la sua versatilità le CRISPR/Cas rappresentano il sistema che meglio si presta alle diverse applicazioni dell'*editing* del DNA. In natura il sistema CRISPR/Cas costituisce un tipo di immunità adattativa usato da molti organismi procarioti per proteggersi dall'ingresso di acidi nucleici estranei. Quando un DNA invasore fa il suo ingresso per la prima volta nella cellula, esso viene segmentato e incorporato all'interno del cromosoma batterico a livello dei loci chiamati CRISPR (*Clustered regularly interspaced short palindromic repeat*), costituiti da sequenze ripetute intervallate da piccoli segmenti di DNA di

derivazione virale o plasmidica, chiamate *SPACERS*, provenienti da precedenti esposizioni a DNA esogeni. Tale *locus* è in grado di trascrivere un RNA non codificante chiamato *crRNA* complementare a una porzione di DNA esogeno chiamata *PROTOSPACER*. Il *crRNA* assieme ad un altro RNA ausiliario, il *tracrRNA*, si appaia alle sequenze ripetute provenienti dal CRISPR formando un ibrido a RNA in grado di richiamare la Nucleasi CRISPR accociata (Cas) sulla sequenza di DNA invasore da tagliare. All'estremità 3' della sequenza del DNA target di ogni *protospacer* è inoltre associato un "*protospacer adjacent motif*" (PAM), elemento essenziale per il riconoscimento da parte della Cas del sito di taglio, il quale avverrà a circa 2-3 bp a monte della PAM (FIG.23).

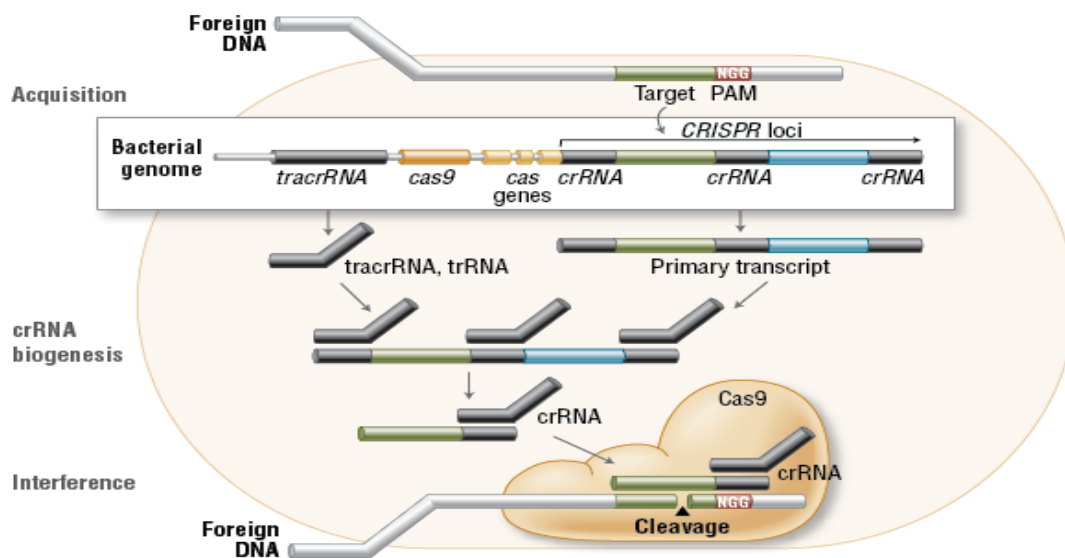


FIG.23: il sistema CRISPR/Cas nelle cellule batteriche.

Nella versione ingegnerizzata utilizzata per l'*editing* il crRNA e il tracrRNA sono fusi in un'unica struttura chimerica chiamata RNA guida (gRNA). Il gRNA è composto da un sequenza di 20 nt target-specifica e dall'RNA scaffold che comprende la sequenza palindromica ripetuta del crRNA e il tracrRNA che assumerà una struttura secondaria a *tetraloop*. Tale struttura è in grado di appaiarsi ad una sequenza di circa 20 nucleotidi del DNA target e di richiamare la Cas per collocarla nel sito adiacente alla sequenza PAM dove, grazie ai suoi due domini nucleasici, sarà in grado di tagliare entrambi i filamenti del DNA target introducendo un Double Strand Breake (DSB) (FIG.24) (87).

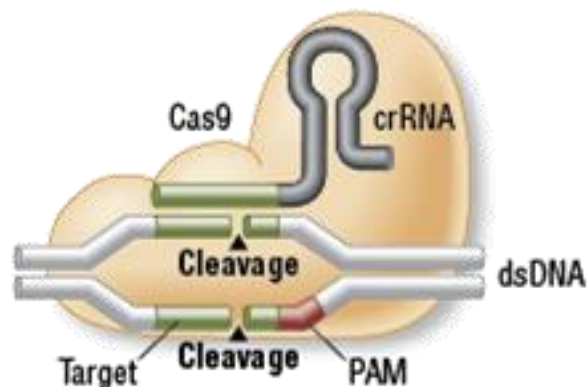


FIG.24: versione ingegnerizzata del sistema CRISPR/Cas.

Il DSB provocato dall'azione della Cas innesca dei meccanismi di riparo cellulari che possono dar luogo all'*editing* del DNA attraverso principalmente due *pathway* (FIG.25) (88):

- Nel NHEJ (*NonHomologous End Joining repair*) il taglio a doppio filamento del DNA viene direttamente riparato in maniera del tutto casuale determinando la formazione di mutazioni indels, che eventualmente possono dar luogo a DSB multipli in zone vicine causando in questi casi delezioni anche di un centinaio di paia basi. Questo meccanismo viene utilizzato soprattutto per il *knockout* genico e per indurre delezioni su sequenze promotrici o non codificanti.
- Nel HDR (*Homology Directed Repair*) la riparazione del DSB avviene mediante ricombinazione omologa, in quanto viene fornito un DNA donatore esogeno a singolo o a doppio filamento che possiede delle zone di omologia con il sito target facilitandone così l'ingresso nella regione di interesse. Questo sistema viene utilizzato soprattutto per correggere un difetto genetico o per effettuare un *knockin*, cioè per inserire uno o più transgeni nel sito di taglio. L'*editing* con questa via di riparo è possibile solo in cellule in divisione e la sua efficacia dipende dal tipo cellulare, dalla sequenza target e dalla sequenza del donor.

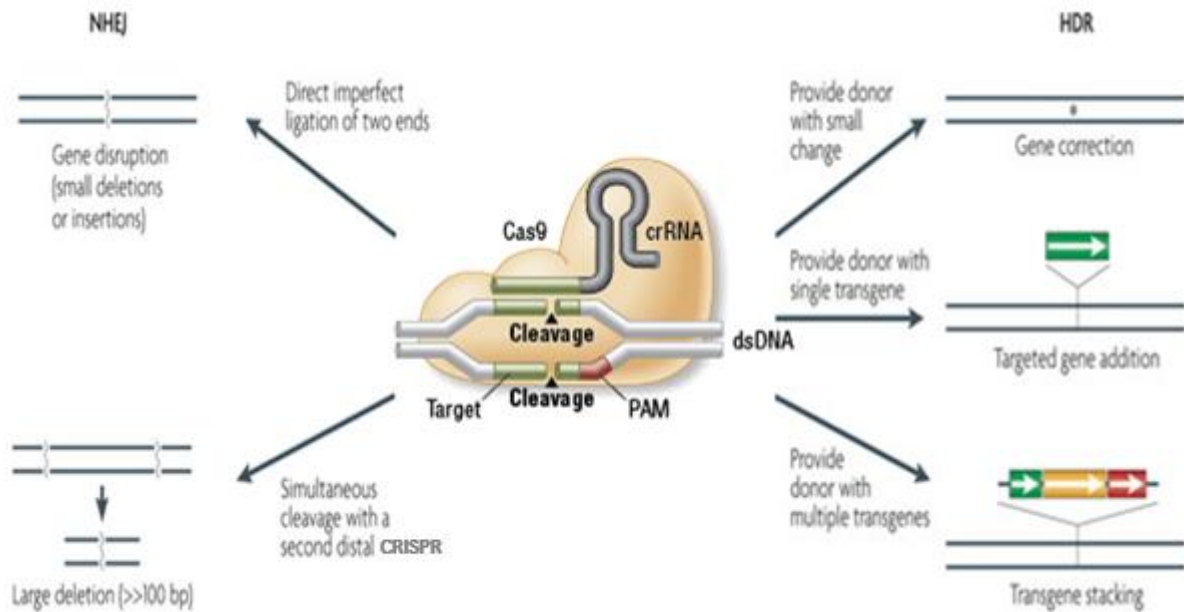


FIG.25: i due principali meccanismi di riparo cellulari che possono dar luogo all'editing del DNA.

L'efficienza di *editing* osservata favorendo l'una o l'altra via di riparo del taglio è molto differente. È stato riportato in letteratura che il riparo del DSB mediante il *pathway* NHEJ ha ottenuto un'efficienza di *editing* genomico pari circa al 92% in cellule di rene embrionale umano HEK293T (89). Per quanto riguarda il riparo attraverso HDR l'efficienza è notevolmente ridotta a causa della difficoltà nell'ottenere eventi di ricombinazione omologa in grado di dar luogo all'*editing* con una resa sufficiente. I migliori risultati sono stati ottenuti con un particolare DNA donatore appartenente alla categoria degli elementi trasponibili mediante il quale è stato ottenuto una percentuale del 23,5 di ricombinazione omologa (80).

Esistono 3 sistemi CRISPR/CAS basati su differenti Cas in grado di riconoscere specifiche sequenze PAM che conferiscono al taglio una maggiore precisione e specificità, diminuendo gli eventuali off-targets. Tuttavia il sistema più dettagliatamente descritto e utilizzato nel *genome editing* è il sistema CRISPR/Cas9 di tipo due derivato da *S.pyogenes* che riconosce come sequenza PAM 5'-nGG e con minor affinità 5'-nAG, che va comunque tenuta in considerazione nella valutazione degli eventuali off-targets (FIG.26) (77).

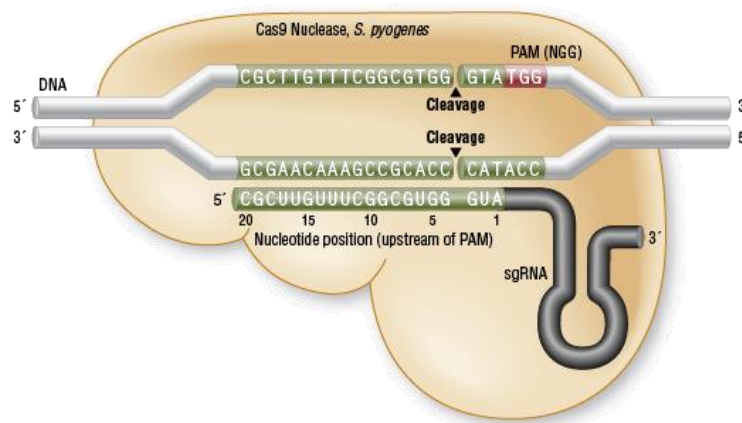


FIG.26: il sistema CRISPR/Cas9 di *S.pyogenes*.

Questo sistema è stato testato in diverse linee cellulari, in cellule primarie tra cui cellule embrionali e le cellule staminali pluripotenti e recentemente in organismi inclusi uomo, topo, zebrafish, *Drosophila melanogaster* e *Cernorhabditis elegans* (90-95).



## i. gRNA

Una volta selezionata la sequenza target gli gRNA possono essere disegnati attraverso l'utilizzo di alcuni programmi disponibili online, in grado inoltre di classificarli in base a un punteggio (SCORE) che va da 0 a 100 che tiene conto del tipo di sequenza e degli eventuali off-target che possono interessare altre regioni codificanti. Un buon RNA guida solitamente dovrebbe rientrare in un *range* di score che va da 60 a 100, ma non trattandosi comunque di valori assoluti anche guide con score più bassi si sono rivelate ugualmente efficaci. L'RNA guida può essere inserito all'interno delle cellule direttamente sotto forma di semplice RNA, oppure come DNA, all'interno di un vettore di espressione eucariota, che verrà poi trascritto in RNA all'interno della cellula. In quest'ultimo caso il DNA codificante per la guida viene inserito in un plasmide di espressione sotto la guida del promotore U6 riconosciuto dalla RNA polimerasi III detta U6gRNA (per renderlo quindi attivo costitutivamente). La cassetta di espressione del RNA guida può essere poi cotransfettata con un plasmide di espressione per la Cas9, oppure inserita direttamente all'interno di un vettore CRISPR "*all-in-one*" che possiedono oltre al promotore U6 anche la cassetta per l'espressione della Cas9, e i geni codificanti per proteine repoters fluorescenti o di resistenza ad antibiotici che permettono la selezione dei cloni positivi (77).

## ii. Cas9

La struttura cristallografica della Cas9 e del complesso gRNA/Cas9 ha permesso di comprendere meglio il meccanismo d'azione di questa endonucleasi RNA-guidata e ha reso possibile lo sviluppo di alcune versioni mutanti di Cas9 in grado di compiere diverse funzioni rispetto alla proteina wild type (wt) (96). La versione wt è chiamata Cas9 nuclease, possiede entrambi i domini nucleasici funzionanti ed è in grado di introdurre il DSB sotto forma di monomero nel DNA target, perciò particolarmente indicata per il knockout genico. La Cas9 nickase (Cas9n) possiede un dominio nucleasico mutato (generalmente la versione più utilizzata è la D10A dove il dominio catalitico RuvC viene mutato nel residuo aa 10 sostituendo l'acido aspartico con l'alanina), questa versione mutante è in grado di tagliare solo un filamento di DNA target, quindi per ottenere un DSB è necessario l'utilizzo di due Cas9n e 2 gRNA uno per ogni filamento. Questo tipo di Cas9 è utilizzata per ottenere con maggiore efficienza una riparazione del taglio attraverso la via HDR rispetto alla NHEJ. Nonostante la *nickase* sia risultata essere meno efficace nell'editing rispetto alla versione wt essa ha come vantaggio la riduzione % dei possibili off-targets, in quanto un solo *nick* non è in grado di innescare i sistemi di riparo NHEJ e la probabilità che entrambi gli RNA guida si leghino in sequenze off-target vicine è relativamente bassa. La Cas9 *Nickase* è perciò particolarmente indicata per effettuare correzioni geniche. Una terza variante è chiamata defective Cas9 (dCas9) in cui entrambi i domini

nucleasici sono inattivi. Questo tipo di Cas9 può essere utilizzato per studiare le proprietà dinamiche del genoma. Infatti chimere di dCas9 sono state utilizzate per attivare così come reprimere l'espressione genica, qualora fosse associata ad attivatori o repressori della trascrizione. Oppure come piattaforma RNA-guidata che per ingombro sterico che va a interferire con i macchinari di trascrizione (CRISPRi) silenziando l'espressione del gene target. O ancora possono essere associate a delle proteine fluorescenti o a specifici epitopi o ad enzimi modificatori della cromatina utili per poter studiare l'attività genomica correlata ad eventi di sviluppo come la proliferazione cellulare, il differenziamento e la morfogenesi (97).

La Cas9 può essere espressa costitutivamente in particolari cellule ingegnerizzate, oppure introdotta al loro interno sottoforma di mRNA che verrà tradotto in proteina dal macchinario di traduzione cellulare oppure attraverso plasmidi di espressione in associazione o meno con guida e geni reporters.

### iii. Donor

Il *donor* fornisce la sequenza di DNA stampo che si vuole inserire nella regione target mediante HDR. Questa sequenza è inserita tra due braccia omologhe alle sequenze adiacenti al DSB generato dalla Cas9 le quali favoriranno l'ingresso del *donor*. Il DNA *template* può essere un ssODNs quando si vogliono effettuare piccole modificazioni nella sequenza oppure a dsDNA o plasmide quando si vogliono inserire sequenze più

grandi (77). Oltre alle correzioni geniche il DNA *donor* può essere utilizzato per poter introdurre diverse cassette come ad esempio per l'espressione di geni *reporters* o *marker* di resistenza antibiotica, che possono essere utili nelle successive analisi di efficienza dell'*editing*. Per ottenere un'efficienza di ricombinazione omologa più elevata possono essere utilizzati delle tipologie di *donor* che già per natura riescono ad integrarsi autonomamente all'interno di sequenze geniche e in presenza di un DSB questa loro capacità è resa ancora più efficace. Un esempio sono gli adenovirus associati (AAVs) che sfruttando la capacità di integrazione del ssDNA mediante le sequenze ITR facilita la ricombinazione omologa (98). Oppure i trasposoni *Piggy BAC* che sono in grado di effettuare una modificazione allelica senza lasciare tracce (*footprint*) (80).

### 3. Scopo della tesi

La correzione della mutazione che determina una patologia ereditabile mediante la tecnica del *Genome-Editing* su cellule staminali specifiche del paziente e successivo trapianto autologo sembrerebbe l'approccio ideale per la cura di malattie monogeniche come la  $\beta$  talassemia. Tuttavia, a causa delle difficoltà nell'ottenere delle percentuali di ricombinazione omologa sufficienti ai fini terapeutici, lo scopo del mio progetto di dottorato, svoltosi presso il laboratorio di Terapia Genica del Prof.

*"Terapia genica della  $\beta$  Talassemia mediante editing del DNA"* - Università degli Studi di Cagliari  
Dot.ssa Maura Mingoia – Dottorato di Ricerca in "Scienze Pediatriche e Terapia dello Sviluppo"

Paolo Moi, è quello di riprodurre artificialmente mediante l'utilizzo del sistema CRISPR/Cas9 associato a NHEJ le mutazioni HPFH identificate in regioni non codificanti del cluster  $\beta$  globinico. In tal modo si conta di ripristinare a livelli terapeutici l'espressione dei geni HBG e di conseguenza la sintesi di una HbF funzionale al fine di migliorare il fenotipo  $\beta$  talassemico.

## 4. Materiali e Metodi

### 4.1 Coltura della linea cellulare $\beta$ -K562

I risultati riportati in questo lavoro sono stati ottenuti mediante l'utilizzo delle cellule  $\beta$ -K562 (*Human erythromyeloblastoid leukemia cell line*), una variante della linea di leucemia mieloide cronica umana in grado di esprimere oltre alla  $\gamma$  globina, in proporzione minore, anche la  $\beta$  (99). La coltura è stata mantenuta con terreno RPMI 1640 (Life Technologies), complementato con 10% di Siero Bovino Fetale (FBS, Life Technologies), 1% di penicillina (100 U/ml), 1% di streptomina (100  $\mu$ g/ml) ed è stata mantenuta alla temperatura di 37°C in un incubatore (Hereaus HERA CELL150) ad atmosfera umidificata al 5% di CO<sub>2</sub>.

## 4.2 Design dell'RNA guida

Il software utilizzato per designare le guide è quello fornito dal Massachusetts Institute of Technology (MIT) disponibile online chiamato CRISPR Design Tool (<http://tools.genome-engineering.org>). Il CRISPR Design Tool fornisce le sequenze per tutti gli oligo necessari alla preparazione dell'sgRNA, esprimendo tramite un numero di SCORE l'efficienza presunta delle modificazioni al sito target e i potenziali tagli in siti off target.

## 4.3 Clonaggio in vettore CRISPR

Per clonare la sequenza dell'sgRNA nel vettore *all-in-one* pSpCas9(BB)-2A-GFP (PX458) ([www.addgene.org](http://www.addgene.org) Plasmid #48138) (FIG.27) sono stati disegnati degli oligo top e bottom strands con le estremità protrudenti 5'-CACC e 5'-AAAC rispettivamente (Sigma Aldrich).

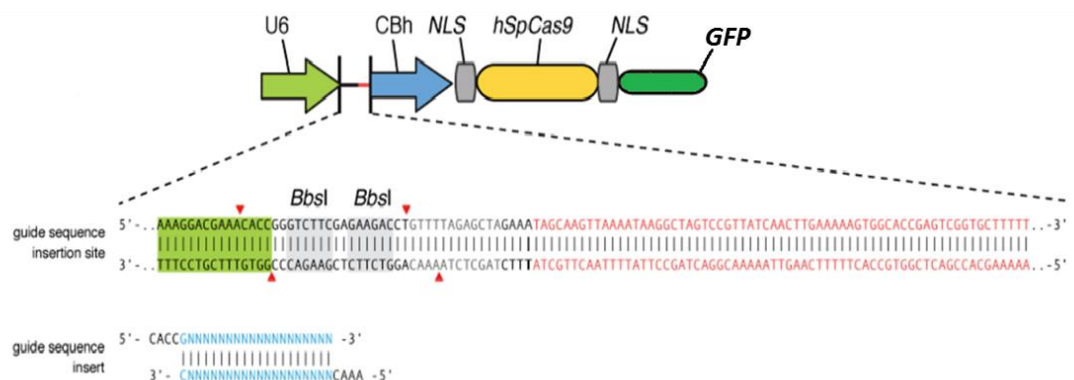


FIG.27: il vettore pSpCas9(BB)-2A-GFP (PX458).

L'inserto a doppio filamento dell'sgRNA è stato ottenuto mediante *annealing* degli oligo top e bottom (100  $\mu$ M) a 95°C per 5' e successivo raffreddamento fino a 25°C. Il vettore CRISPR è stato linearizzato mediante digestione con l'enzima di restrizione BbsI (NEB cat. no. R0539S). La reazione di clonaggio è stata eseguita utilizzando la T4 ligase (Invitrogen cat. no. 15224-017) e il seguente protocollo:

Componenti	Quantità
<b>pSpCas9(BB)-2A-GFP</b>	100ng
<b>dsOligos (dil. 1:200)</b>	2 $\mu$ L
<b>10X ligation buffer</b>	1,5 $\mu$ L
<b>T4 Ligase</b>	1,5 $\mu$ L
<b>H<sub>2</sub>O</b>	To 15 $\mu$ L

Le reazioni di *ligation* sono state incubate a 22°C per circa due ore e successivamente 3  $\mu$ L sono stati trasformati in cellule competenti *E.coli* Stbl3 ONE SHOT (Life technologies, cat. no. C7373-03) seguendo il protocollo fornito dalla ditta produttrice. Le colonie ottenute sono state inoculate in 3 mL di LB medium e ampicillina (100  $\mu$ g/mL) e incubate in agitazione a 37°C overnight per la mini preparazione di DNA plasmidico. Il plasmide estratto dalle cellule batteriche è stato sequenziato a partire dall'U6 promoter utilizzando il primer U6-Fwd (5'-GAGGGCCTATTTCCCATGATTCC-3') utilizzando il 3130XL Genetic

Analyser (Applied Biosystem). I cloni positivi sono stati inoculati per la maxi preparazione di DNA plasmidico mediante il kit PureLink HiPure Plasmid Filter Maxyprep Kit (Invitrogen cat. no. K2100-17).

#### 4.4 Elettroporazione delle cellule $\beta$ -K562

$1 \times 10^7$  cellule K562 sono state elettroporate con 3  $\mu\text{g}$  di vettore PX458 mediante l'utilizzo del Neon Transfection System (Invitrogen) con il seguente settaggio: 1350 volt, 10 msec, 3 pulse. Le cellule sono state messe in coltura con terreno RPMI 1640 addizionato di FBS e privo di Pen/Strep e tenute in incubazione a 37°C e 5% CO<sub>2</sub> per 48 ore. Per ottenere singoli cloni una aliquota delle cellule transfettate è stata diluita alla concentrazione finale di 2 cellule per 100  $\mu\text{L}$  e seminate in una piastra da 96 well (100  $\mu\text{L}$ /well) e lasciate espandere per circa tre settimane.

#### 4.5 FACS Sorting

Le cellule sono state raccolte, centrifugate e risospese in PBS 1X alla concentrazione finale di  $1 \times 10^7$ /mL. Il sorting è stato effettuato tramite FACS ARIA (Becton Dickinson) settando il macchinario con un controllo negativo di cellule transfettate con un vettore *empty*. La fluorescenza relativa all'espressione dal reporter GFP presente nel vettore ha permesso

l'identificazione di tre popolazioni:

*"Terapia genica della  $\beta$  Talassemia mediante editing del DNA"* - Università degli Studi di Cagliari  
Dot.ssa Maura Mingoia – Dottorato di Ricerca in "Scienze Pediatriche e Terapia dello Sviluppo"



- GFP negative: popolazione con assenza del vettore
- GFP high: popolazione con elevata espressione del vettore GFP
- GFP medium: popolazione mista GFP+ e GFP-

Di tutte e tre le popolazioni è stata verificata la purezza tramite il software FACSDiva 8.0.

## 4.6 Estrazione DNA, RNA e proteine

Il DNA genomico è stato ottenuto mediante il QuickExtract DNA extraction solution (Epicentre, cat. no. QE09050). L'RNA totale è stato estratto mediante l'RNeasy Mini Kit (Qiagen cat. no. 74106). L'estratto proteico totale è stato ricavato mediante RIPA lysis buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.4, 1% Triton X-100, 1% Sodium Deoxycholate, 0,1% SDS).

## 4.7 Sequenziamento di Sanger

La reazione di sequenza è stata effettuata mediante il termociclatore Gene Amp PCR System 9700 (Applied Biosystem) utilizzando 100 ng di DNA genomico, il BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystem cat. no. 4337455), il primer Ay1FOR (5'-TTTTCTTAGAAACCACTGCTAACTG-3') per il promotore Ay e il primer Gy1FOR (5'-TGAGACTAAGACGTGTCCCATC-3') per il promotore Gy attraverso le seguenti condizioni: 95°C 1', (95°C 30'', 50°C 15'', 60°C 4') per 25 cicli.

Le reazioni sono state successivamente purificate e lette al sequenziatore 3130XL Genetic Analyser (Applied Biosystem).

## 4.8 PCR screening

Le regioni target sono state amplificate utilizzando l'Herculase II fusion polymerase (Agilent Technologies cat. no. 600679) a partire da 100 ng di DNA genomico mediante le seguenti coppie di primers:

- A $\gamma$ 1FOR (5'-TTTTCCTTAGAAACCACTGCTAACTG-3') e A $\gamma$ REV (5'-GACCGTTTCAGACAGATATTTGCA-3') per il promotore A $\gamma$ .
- G $\gamma$ 1FOR (5'-TGAGACTAAGACGTGTCCCATC -3') e G $\gamma$ REV (5'-GACCGTTTCAGACAGATATTTGCA-3') per il promotore G $\gamma$ .
- G1G2gamma-intrag. FOR (5'-CAGGATGTGCAGGCTTGTTA-3') e G1G2Ggamma-intrag. REV (5'-TTGGTTCTGTTTTGCTTTTGG-3') per la regione intergenica.

La PCR è stata eseguita secondo le seguenti condizioni: 95°C 2', (95°C 20'', 60°C 20'', 72°C 30'') per 35 cicli, 72°C 3'.

## 4.9 RT-qPCR

Per determinare l'espressione dei geni HBG e HBB è stata effettuata una RT-qPCR su cDNA ottenuto per retrotrascrizione utilizzando il kit SuperScript III First-Strand Synthesis System (Invitrogen cat. no. 18080051). La mix di reazione è composta dalla TaqMan Universal PCR Master Mix (Applied Biosystems cat.

no. 4304437) e dai seguenti primers e probe: Gene target "gamma globin"  
*"Terapia genica della  $\beta$  Talassemia mediante editing del DNA"* - Università degli Studi di Cagliari  
 Dot.ssa Maura Mingoia – Dottorato di Ricerca in "Scienze Pediatriche e Terapia dello Sviluppo"

(Applied Biosystem cat. no. Hs00426283\_m1); Geni di riferimento “beta globin” (Applied Biosystem cat. no. 4319413E) e “ $\beta_2$ microtubulina” (Applied biosystem cat. no. 4308313). La RT-qPCR è stata eseguita utilizzando la macchina 7900 HT Fast Real Time PCR System (Applied Biosystems) e le condizioni universali (10 min a 95° C, dopo 15 s a 95° C e 1 min a 60° C per 40 cicli). Per valutare i livelli di espressione dei geni è stato utilizzato il metodo comparativo dei DDCT, Threschold Cycle (Ct), (100) e l’espressione è stata quantificata in fold change.

#### 4.10 Western blot

50  $\mu$ g di estratto proteico totale sono stati fatti migrare mediante SDS-PAGE utilizzando il Bolt 4-12% Bis-Tris Plus Gel (Invitrogen cat. no. NW04120BOX). Il blotting è avvenuto over night a 26 volts su una membrana di PVDF da 0,22  $\mu$ m (Amersham cat. no. 10600021) utilizzando il Transfer Buffer 1X (0,025 M Tris-Base, 0,192M Glicina, 15% Metanolo, 0.1% SDS). L’ibridazione con gli anticorpi primari è stata effettuata utilizzando i seguenti:

- per la  $\gamma$  globina l’Hemoglobin  $\gamma$  Antibody (51-7) (Santa Cruz cat. no. sc-21756)
- per la  $\beta$  globina l’Hemoglobin  $\beta$  Antibody (37-8) (Santa Cruz cat. no. sc-21757)
- per la  $\beta$  actina il  $\beta$ -Actin Antibody (C4) (Santa Cruz cat. no. sc-47778)

L’ibridazione con l’anticorpo secondario è stata effettuata utilizzando il goat anti-mouse IgG-HRP (Santa Cruz cat. no. sc-2031). Per la detection è stato

utilizzato il kit ECL Prime Western Blotting Detection Reagent (Amersham cat. no. RPN2232).

## 5. Risultati

### 5.1 DSB RNA-guidato

Il sistema CRISPR/Cas9 di tipo II ha dimostrato di essere efficiente e specifico nel taglio programmato sulla sequenza target in colture cellulari così come in piante e animali. Al fine di creare vettori CRISPR/Cas9 in grado di indurre specifiche mutazioni nella regione target dei promotori HBG, tramite il programma “CRISPR Design tool” abbiamo identificato e disegnato tre differenti guide in grado di tagliare la regione -202 -195 di entrambi i promotori dei geni HBG, regione che risulta identica e sede di un cluster di mutazioni HPFH in entrambi i promotori. L’RNA guida, costituita da 20 paia di basi, è stata disegnata in modo da potersi appaiare alla regione bersaglio adiacente alla sequenza contenente il trinucleotide PAM NGG canonico, rappresentato da: GGG per la guida 1; GGG per la guida 2; GGA per la guida 3 (FIG.28). Le tre guide presentano un valore di score di 23 e 25, dovuto al tipo di sequenza e al numero di off targets rilevati. Ogni singola guida, infine, è stata clonata all’interno di un vettore di espressione CRISPR *all-in-one* “PX458” sotto la guida del promotore U6.

Lo stesso plasmide contiene inoltre il gene per l'espressione della endonucleasi Cas9 di *S.pyogenes* umanizzata e il gene *reporter* di fluorescenza verde, GFP.

	score	sequence	
Guide #1	67	AATTAGCAGTATCCTCTTGG	GGG
Guide #2	61	AAATTAGCAGTATCCTCTTG	GGG
Guide #3	60	AAAATTAGCAGTATCCTCTT	GGG
Guide #4	56	AAAAATTAGCAGTATCCTCT	TGG
Guide #5	35	AAAAACTGGAATGACTGAAT	CGG
Guide #6	31	TATCTGTCTGAAACGGTCCC	TGG
Guide #7	31	CCCATGGGTGGAGTTTAGCC	AGG
Guide #8	31	GGAATGACTGAATCGGAACA	AGG
Guide #9	30	CCTGGCTAAACTCCACCCAT	GGG
Guide #10	29	CCATGGGTGGAGTTTAGCCA	GGG
Guide #11	29	ATTTGCATTGAGATAGTGTG	GGG
Guide #12	29	ACTGAATCGGAACAAGGCAA	AGG
Guide #13	28	TATTTGCATTGAGATAGTGT	GGG
Guide #14	28	CCCTGGCTAAACTCCACCCA	TGG
Guide #15	26	ATATTTGCATTGAGATAGTG	TGG
Guide #16	25	CATTGAGATAGTGTGGGGAA	GGG
Guide #17	25	GCATTGAGATAGTGTGGGGA	AGG
Guide #18	24	GTGGGGAAGGGGCCCAAG	AGG
Guide #19	23	ATTGAGATAGTGTGGGGAAG	GGG
Guide #20	18	ATGCAAATATCTGTCTGAAA	CGG

93..115 GUIDA 2

94..116 GUIDA 3

92..114 GUIDA 1

FIG.28: guide identificate con il CRISPR Design Tool.

## 5.2 Analisi sull'efficacia dell'*editing*

Al fine di valutare l'efficacia delle guide disegnate nell'indurre la formazione di indels casuali nella sequenza promotrice dei geni HBG1 e HBG2, i tre differenti vettori CRISPR/Cas9 in esame sono stati elettroporati nella linea cellulare  $\beta$ -K562 (leucemia eritromieloblastoide umana), che presenta un *pattern* di espressione globinica attribuibile al periodo fetale dello sviluppo eritropoietico. A 48 ore dalla transfezione le cellule GFP positive sono state isolate mediante FACS *sorting* e diluizioni seriali.

La selezione dei cloni positivi all'*editing* è infatti essenziale per caratterizzare le modificazioni generate dal DSB e la loro ripercussione a livello trascrizionale e traduzionale.

- FACS *sorting*: le cellule transfettate sono state analizzate al citofluorimetro (FACS ARIA); sfruttando le caratteristiche morfologiche del profilo di *forward and side scatter* sono stati preventivamente eliminati sia i detriti che gli agglomerati cellulari. A questo punto essendo le cellule transfettate fluorescenti per la presenza di GFP nel costrutto, è stato possibile identificare con chiarezza le cellule positive grazie al confronto con un campione transfettato con un vettore vuoto e non fluorescente con cui si è contrassegnata la popolazione negativa. Le cellule sono state poi suddivise, attraverso specifici

gates basati sulla intensità della fluorescenza, in tre popolazioni indicate come: GFP+ High (costituita da una popolazione cellulare GFP+), GFP+ Medium (popolazione mista GFP+ e GFP-) e GFP- Low (popolazione cellulare negativa alla GFP). Queste popolazioni cellulari sono state separate attraverso il FACS ARIA e raccolte in tre provette distinte (FIG.29).

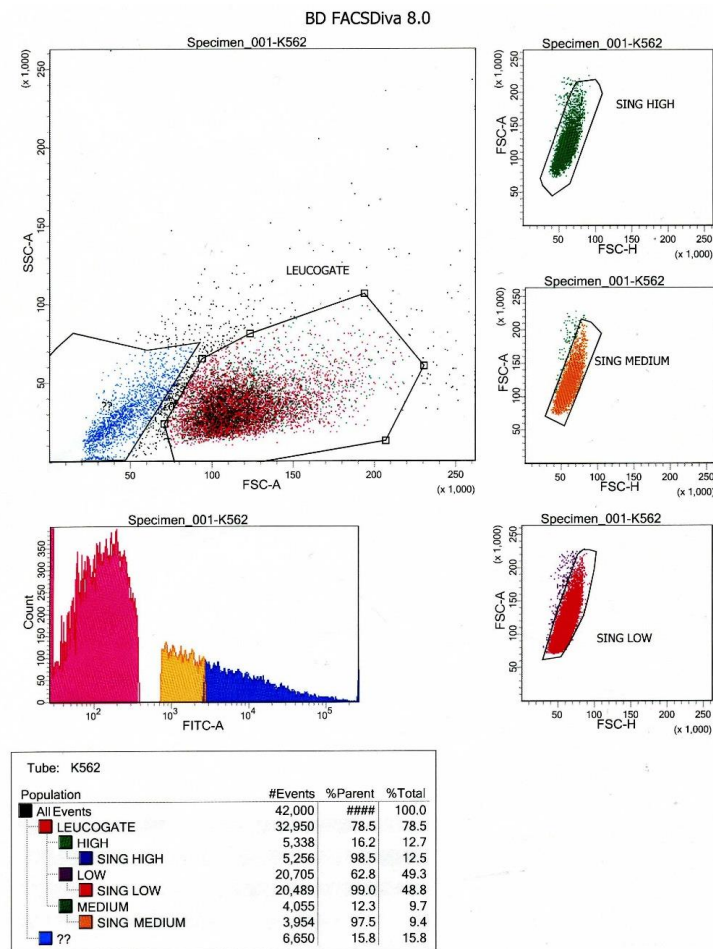
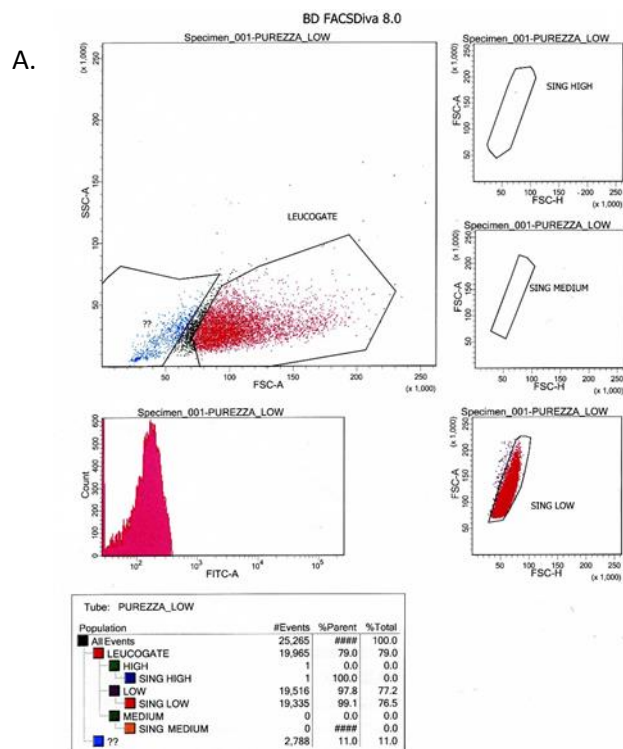


FIG.29: separazione dei gates corrispondenti alle tre popolazioni.

Ognuno dei tre campioni ottenuti è stato quindi analizzato nuovamente al citofluorimetro per valutarne la purezza. La popolazione cellulare “Low “ presenta una percentuale di cellule GFP – del 97,8% (FIG. 30 A), mentre la popolazioni cellulare “High” e “Medium” possiedono una percentuale di cellule GFP+ rispettivamente del 97,2% e 87,5% (FIG.30 B,C).





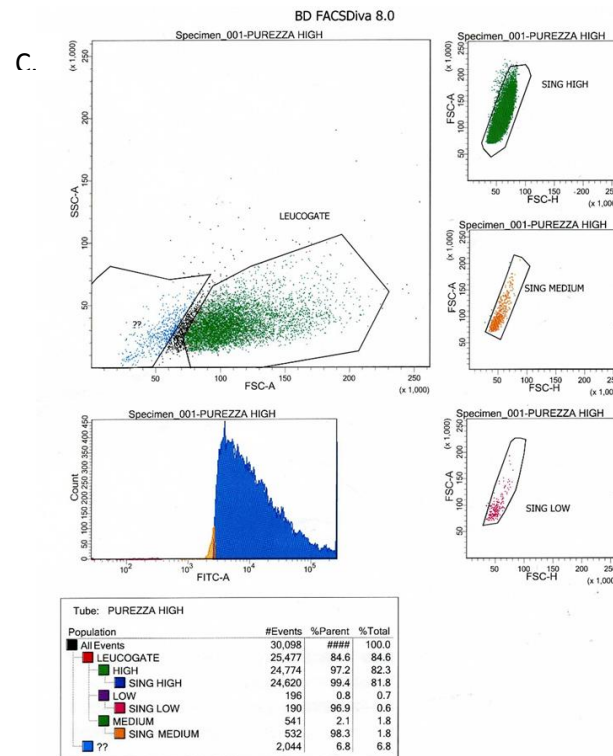
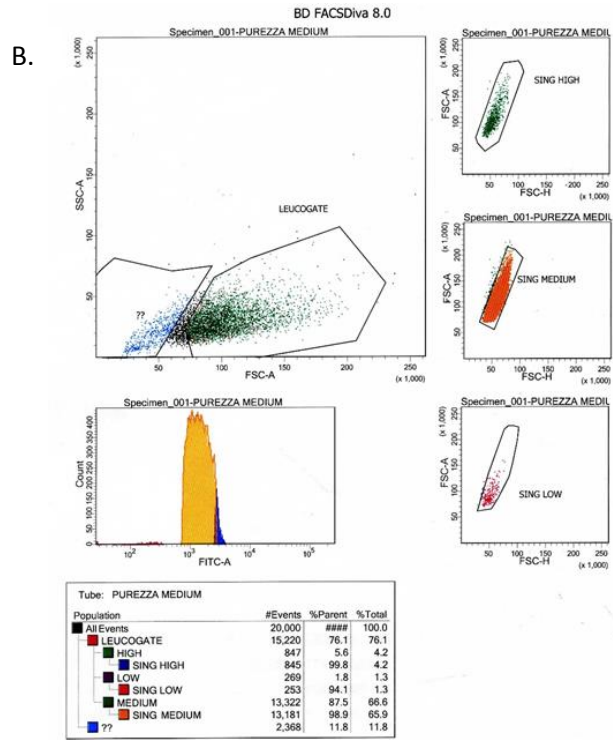


FIG.30: purezza delle popolazioni sortate; A. Low; B. Medium; C. High.

- Diluizioni seriali: la grande variabilità nel tipo di indel indotte dal sistema CRISPR/Cas9 potrebbe determinare differenti profili di espressione dei geni globinici tra i diversi cloni cellulari. Al fine di identificare e caratterizzare i differenti eventi di *editing* avvenuti nelle cellule elettroporate, abbiamo isolato singoli cloni cellulari positivi diluendo le  $\beta$ -K562 in modo da ottenere una singola cellula per pozzetto. Dopo circa tre settimane dai singoli cloni è stato isolato il DNA, RNA e le proteine. A partire da 96 pozzetti sono stati isolati in totale 8 singoli cloni, di cui il clone C1 proveniente dall'elettroporazione con la guida 1, i cloni D5 ed E12 con la guida 2 e i cloni G6, G8, H3, H10 e H11 con la guida 3.

### 5.3 Caratterizzazione molecolare dell'*editing* RNA-guidato nella sequenza target

Al fine di comprendere se gli *indels* creati artificialmente tramite le CRISPR/Cas9 siano state in grado di riprodurre il fenotipo HPFH, abbiamo analizzato l'esatta natura dell'*editing* tramite amplificazione e sequenziamento della regione target.

Tutti i cloni  $\beta$ -K562 positivi sono stati sequenziati intorno alla regione del promotore dei geni HBG1 e HBG2.

Degli otto cloni isolati sei erano come atteso mutati (D5, E12, G6, G8, H3 e H10) mentre gli altri due erano wild type (C1 e H11).

Come mostrato in figura 31 A e B, i cloni C1 e H11 risultano wt nella regione target del promotore HBG2 mentre la sequenza target del promotore HBG1 risulta eterozigote, dovuto al fatto che un allele presenta un'inserzione AAGC molto comune, ma riconducibile ugualmente ad una sequenza wt.



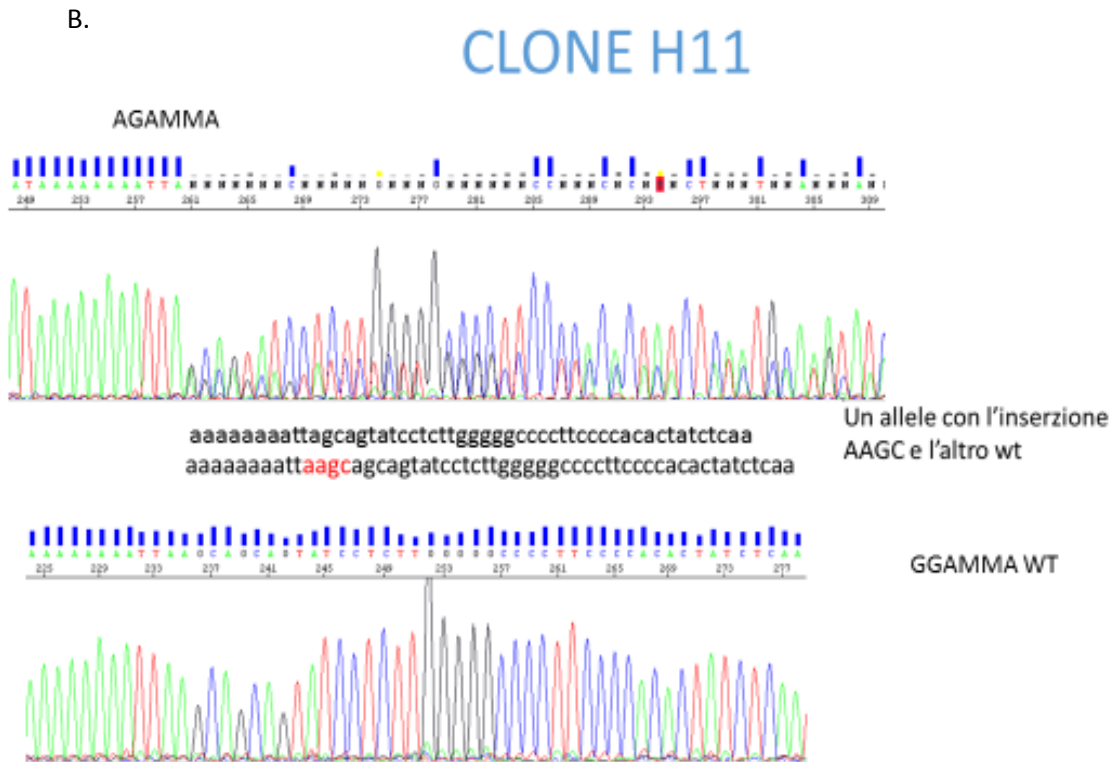


FIG.31: sequenziamento dei promotori  $\gamma$  e  $\text{A}\gamma$  nei cloni wt C1 e H11.

Il clone D5 presenta in  $\text{G}\gamma$  la sequenza wt, mentre sul  $\text{A}\gamma$  mostra una sequenza alterata e illeggibile in corrispondenza della sequenza target a partire dal nucleotide -226, a indicare che è avvenuto un riarrangiamento genico che determina una sovrapposizione e confusione delle sequenze nucleotidiche. (FIG.32).

## CLONE D5

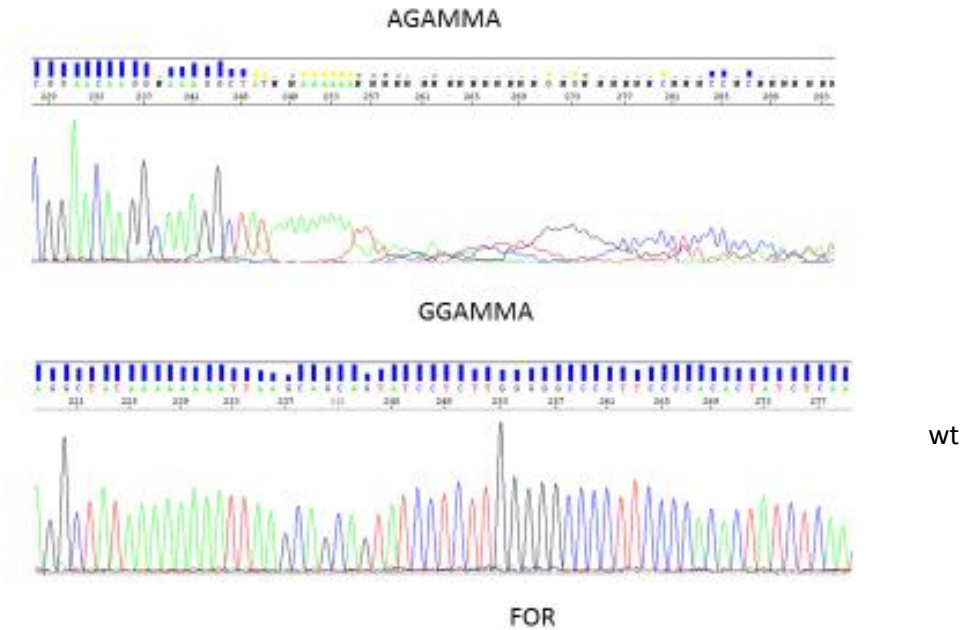


FIG.32: sequenziamento dei promotori G $\gamma$  e A $\gamma$  nel clone D5.

Il clone E12 in G $\gamma$  presenta una delezione in eterozigosi di 27 bp a partire dal nucleotide -211, mentre il promotore HBG1 ha una delezione di due basi C in posizione -196/197. In questo caso la sequenza risulta unica e tale risultato potrebbe essere riconducibile o alla presenza della stessa mutazione in omozigosi oppure all'assenza della sequenza di uno dei due alleli, che per via di una delezione estesa non è stato amplificato (FIG.33).

## CLONE E12

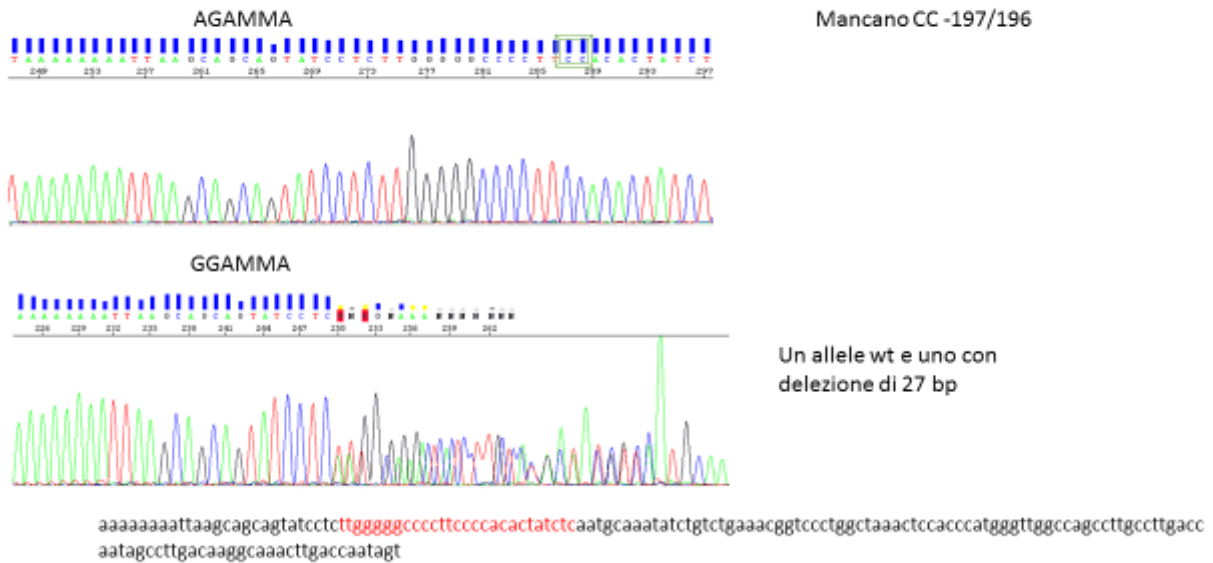


FIG.33: sequenziamento dei promotori Gy e Ay nel clone E12.

Il clone G6 presenta in Gy una doppia sequenza dovuta alla presenza di una delezione di 32 bp a partire dal nucleotide -215 in un allele e alla delezione di due C in posizione -203/204 accompagnata da una mutazione puntiforme (G o T al posto della C) in posizione -196 nell'altro. Nella regione promotrice del gene HBG1, come per il clone E12 la sequenza è singola e presenta la delezione delle due C in posizione -196/197 (FIG.34).

## CLONE G6

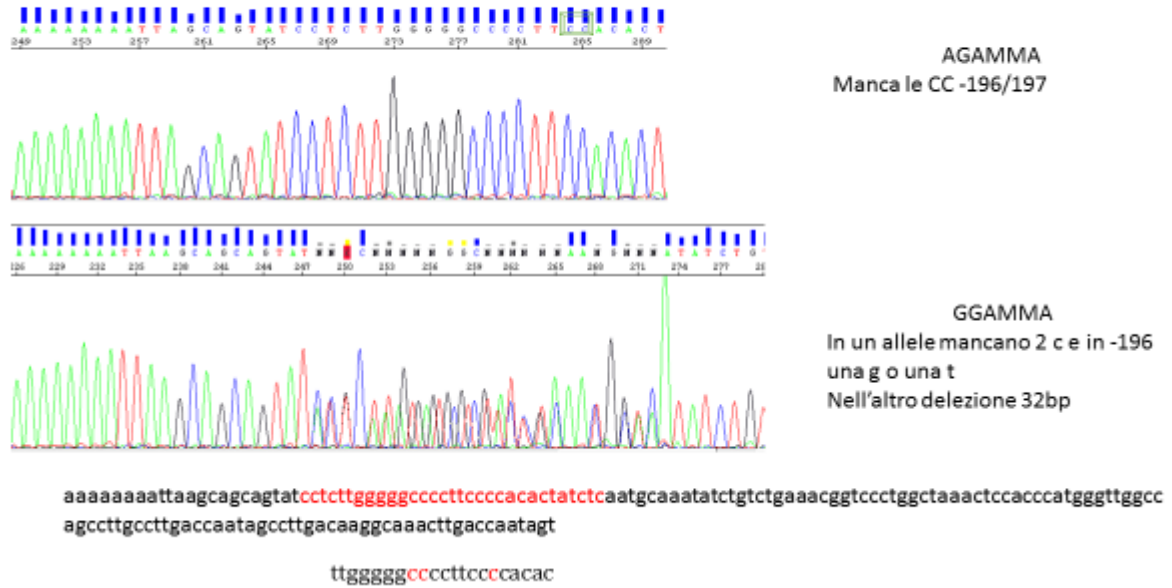


FIG.34: sequenziamento dei promotori Gy e Ay nel clone G6.

Il clone G8 possiede nel promotore di entrambi i geni HBG la delezione delle due C in posizione -196/197 in omozigosi o in un singolo allele per promotore (FIG.35).

## CLONE G8

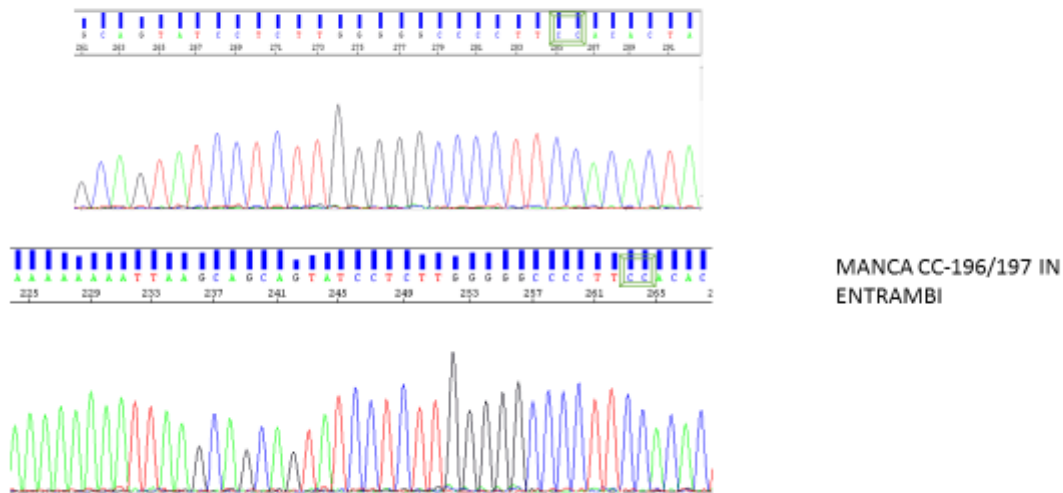


FIG.35: sequenziamento dei promotori Gy e Ay nel clone G8.

Il clone H3 mostra una delezione a partire dal nucleotide -203 di 19 bp in Gy, mentre non è stato possibile amplificare la regione Ay (FIG.36).

## CLONE H3

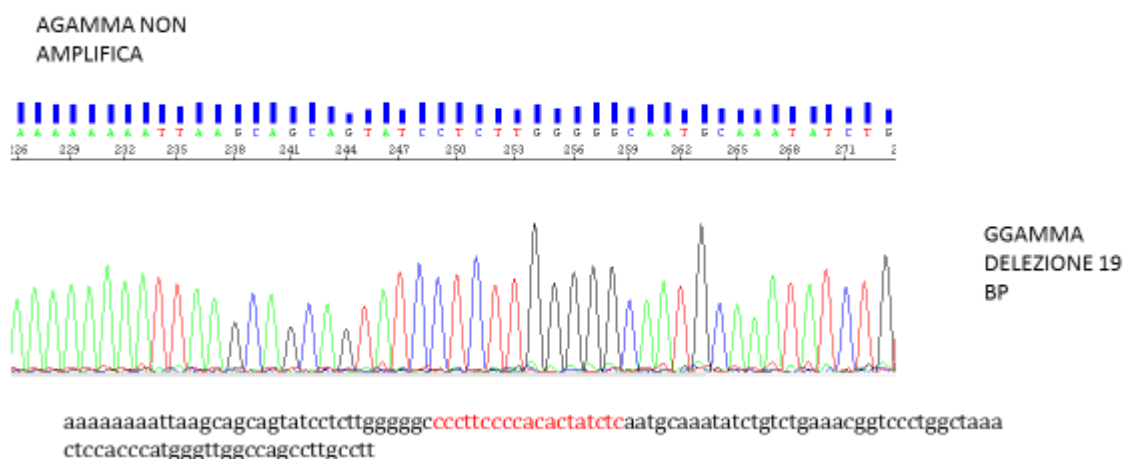


FIG.36: sequenziamento dei promotori Gy e Ay nel clone H3.



Similmente il clone H10 possiede una delezione in posizione -213 di 18 bp in Gy e anche in questo caso non è stato possibile sequenziare il promotore HBG1 (FIG.37).

## CLONE H10

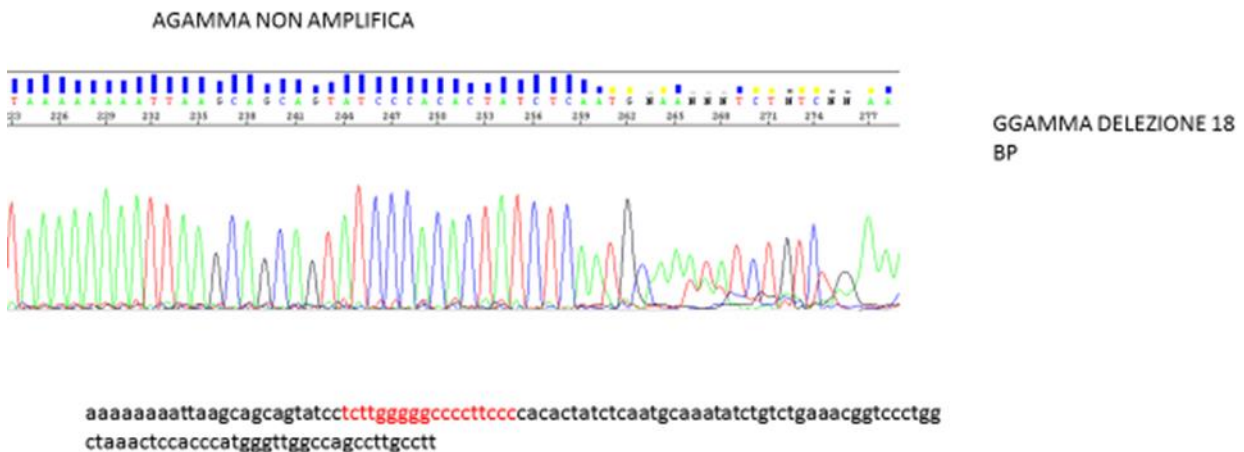


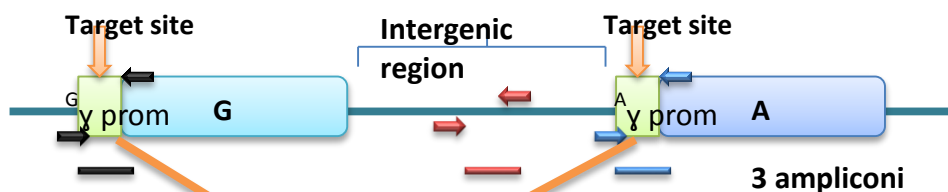
FIG.37: sequenziamento dei promotori Gy e Ay nel clone H10.

## 5.4 PCR screening

Poiché la regione target delle CRISPR/CAS9 da noi utilizzate risulta identica in entrambi i promotori Ay e Gy, questo potrebbe determinare, in base alla cinetica della reazione di riparo del DSB, la formazione di una grossa delezione che rimuoverebbe la parte prossimale del promotore, il gene HBG2, la regione intergenica tra Gy e Ay e parte distale del promotore HBG1. Per verificare se in alcuni dei cloni positivi si sia verificata la delezione della regione intergenica tra i due geni HBG, abbiamo eseguito

uno *screening* mediante PCR della sequenza del promotore del gene HBG2 e HBG1. Le coppie di primers utilizzati sono state le seguenti: G $\gamma$  FOR/A-G $\gamma$  REV che amplificano una sequenza del promotore HBG2 di 355 bp, A $\gamma$ 1FOR/A-G $\gamma$ REV che amplificano una sequenza del promotore HBG1 di 325 bp e i primers G1G2gamma-intrag.FOR/G1G2gamma-intrag.REV che amplificano una sequenza della regione intergenica tra G $\gamma$  e A $\gamma$  di 298 bp (FIG.38 A ).

**A**



**B**



FIG.38: rappresentazione schematica dello screening mediante PCR.

Come mostrato in figura 39 i cloni C1, D5, E12, G6, G8 e H11, presentano tutti e tre gli ampliconi del promotore HBG2, HBG1 e la regione intergenica G $\gamma$  A $\gamma$ , ad indicare l'assenza di grosse delezioni in questi campioni. Viceversa il clone H3 (FIG.39b) mostra soltanto l'amplicone relativo al promotore HBG2, mentre mancano quelli relativi alla

regione intergenica e al promotore HBG1. Riteniamo che in questo caso si sia verificata una grossa delezione che interessa la regione del promotore HBG2 a valle del sito di taglio, il gene HBG2, la regione intergenica tra G $\gamma$  e A $\gamma$  e la regione del promotore HBG1 a monte del sito di taglio (FIG.38b). Il clone H10 (FIG.39a) invece non mostra l'amplificato corrispondente al promotore HBG1, probabilmente a causa della presenza di una delezione che non permette l'appaiamento del primer reverse, il quale si trova in prossimità della sequenza target.

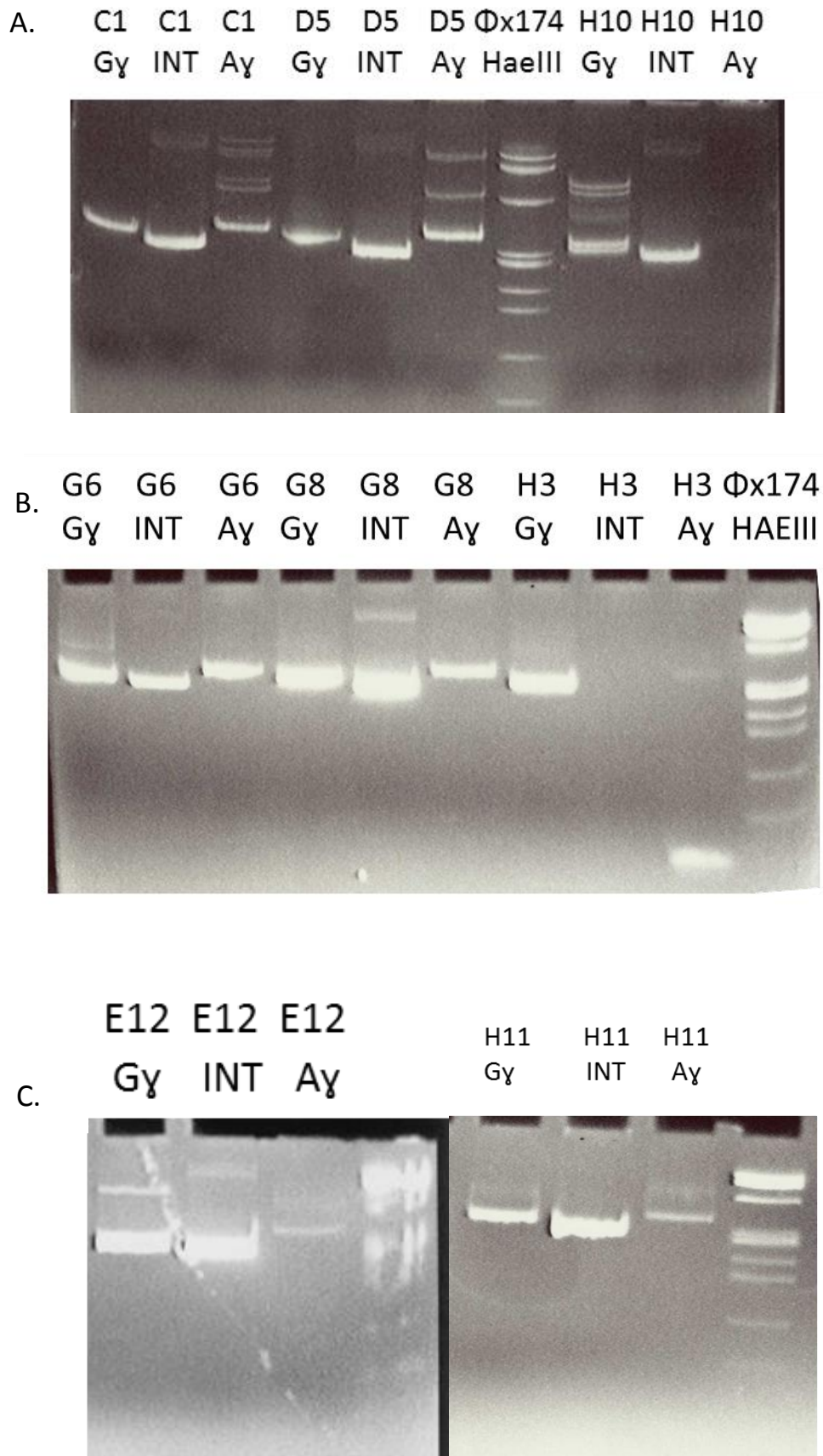


FIG.39: PCR screening al livello della regione target dei singoli cloni.

## 5.5 Caratterizzazione dell'effetto dell'*editing* sull'espressione dei geni globinici

Per quantificare l'effetto dell'*editing* sull'espressione dei geni HBG abbiamo condotto un'analisi di RT-qPCR sull'mRNA estratto dai campioni separati con il FACS e derivati dai singoli cloni. I dati così ottenuti sono stati in seguito confermati mediante western blot, il quale ha permesso di saggiare i livelli di proteina prodotta.

- RT-qPCR: l'RNA estratto dai campioni derivati dal FACS *sorting* e dai singoli cloni è stato successivamente retrotrascritto in cDNA ed analizzato mediante un saggio di Real Time qPCR utilizzando le sonde TaqMan specifiche per la  $\gamma$  globina, la  $\beta$  globina e il gene *housekeeping*  $\beta_2$ microtubulina, necessario per normalizzare i dati relativi ai geni target.

L'espressione relativa dei geni  $\gamma$  e  $\beta$  globinici è stata valutata inizialmente nel pool di cellule K562 separate al FACS. Come mostrato in figura 40 il pool cellulare HIGH GFP+ presenta una notevole riduzione dell'espressione della  $\gamma$  globina e  $\beta$  globina rispetto al pool di  $\beta$ -K562 GFP negativi (Low).

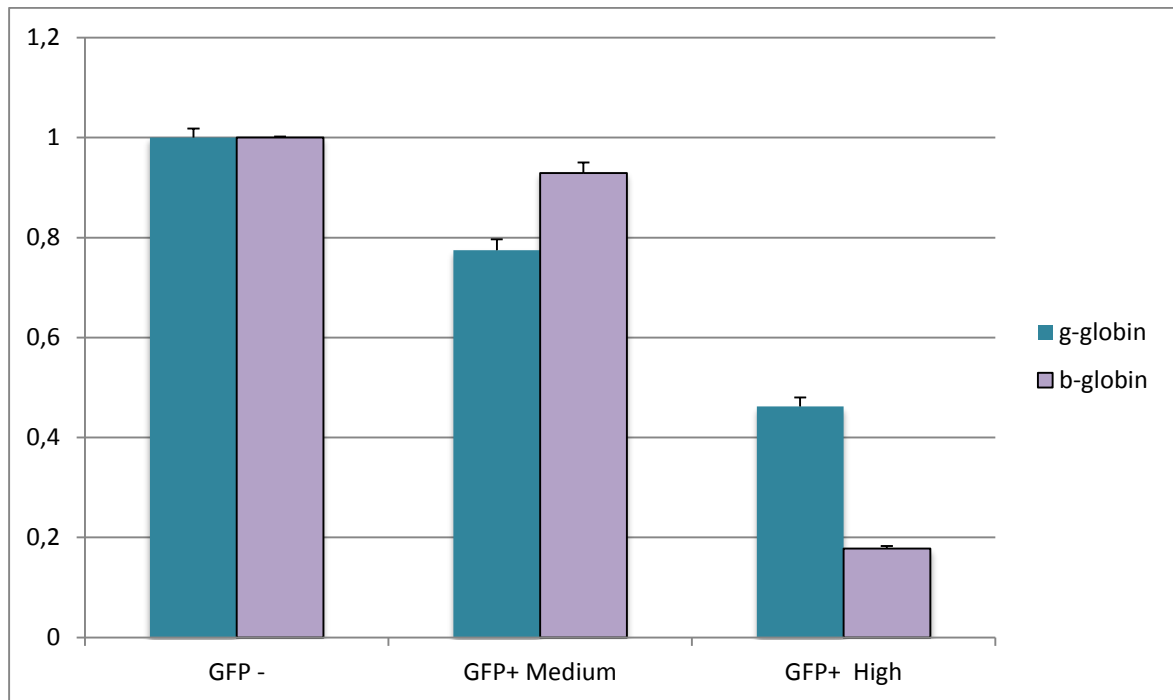


FIG.40: grafico relativo all'RT-qPCR delle popolazioni separate al FACS.

Successivamente abbiamo analizzato l'espressione dei geni HBG e HBB nei singoli cloni di cellule  $\beta$ -K562 dopo 35 giorni dalla transfezione. Dal grafico riportato nella figura 41 è possibile osservare un incremento dell'espressione della  $\gamma$  globina in tutti i cloni risultati positivi all'*editing*, fatta eccezione per il clone D5. Nello specifico i cloni G8, H3 e H10 mostrano un aumento dell'espressione della  $\gamma$  globina di circa 3,5 volte rispetto al controllo wt, il clone G6 mostra livelli di  $\gamma$  globina incrementati di circa 6,8 volte, mentre il clone E12 risulta quello con i livelli di  $\gamma$  globina più elevati (11,48 volte). I livelli di  $\beta$  globina risultano leggermente più variabili rispetto alla  $\gamma$ , ma questo sarebbe più attribuibile alla natura delle cellule utilizzate per questi esperimenti piuttosto che agli effetti dell'*editing*, le quali di natura mostrano un'espressione variabile del gene HBB. Questi risultati mostrano che indipendentemente dalle

differenti modificazioni ottenute, i cloni che possiedono delezioni intorno alla regione target HPFH (E12, G6, G8, H3, H10) presentano un aumento dell'espressione dei geni HBG. Da notare inoltre come il clone H3, che presenta una grossa delezione tra i due promotori bersaglio  $A\gamma$  e  $G\gamma$ , , mostra un *pattern* di espressione paragonabile agli altri cloni che non la possiedono, suggerendo che la perdita di parte del promotore HBG2, del gene HBG2, la regione intergenica tra  $G\gamma$  e  $A\gamma$  e parte del promotore HBG1 non comprometta la funzionalità del cluster  $\beta$  globinico nei saggi in  $\beta$ -K562. Tuttavia, se rapportato con i livelli di espressione della  $\beta$  globina l'aumento di espressione dei geni HBG risulta meno marcato. Infatti, considerando il rapporto tra  $\gamma$  e  $\beta$  globina i cloni E12, G8, H3 e H11 esprimono circa tre volte di più i geni HBG rispetto al gene HBB. Soltanto il clone G6 mostra dei livelli di  $\gamma$  globina sei volte superiori a quelli relativi alla  $\beta$  globina.

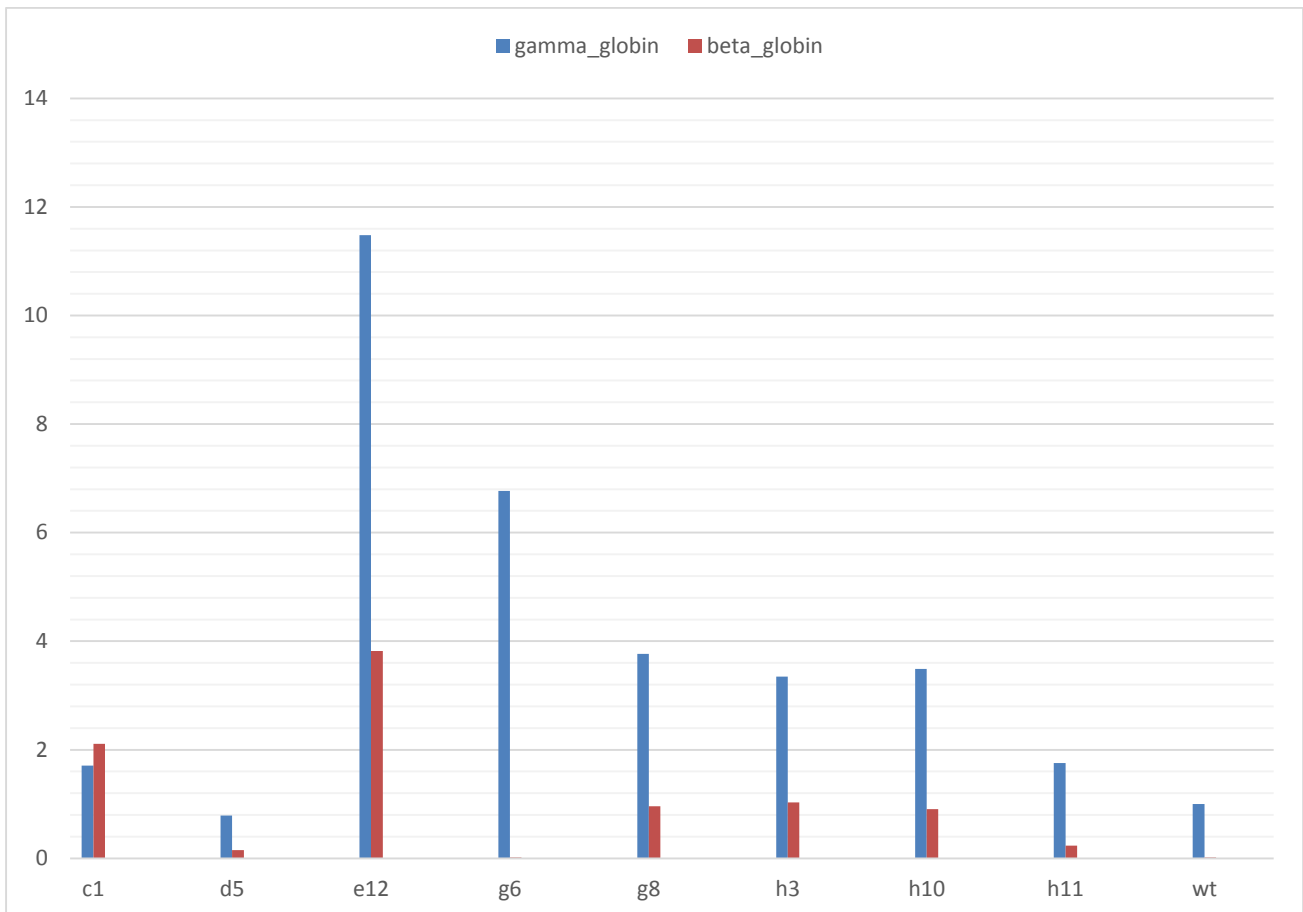


FIG.41: grafico relativo all'RT-qPCR dei singoli cloni ottenuti con le diluizioni seriali.

- Western blot: per confermare l'incremento nell'espressione dei geni HBG anche a livello traduzionale, l'estratto proteico totale proveniente dai singoli cloni sotto esame è stato analizzato mediante western blot e ibridazione anticorpo-specifica con anticorpo anti- $\gamma$  globina,  $-\beta$  globina e  $-\beta$  actina come controllo. Analogamente ai risultati ottenuti con la qPCR il saggio di western blot, mostra che i cloni E12, G8, H3, H10 presentano il maggior incremento nell'espressione  $\gamma$  globinica. Il segnale relativo al clone G6 risulta al contrario meno intenso (FIG.42a). Un risultato simile ai dati di qPCR è stato ottenuto anche per quanto concerne la  $\beta$  globina (FIG.42b).



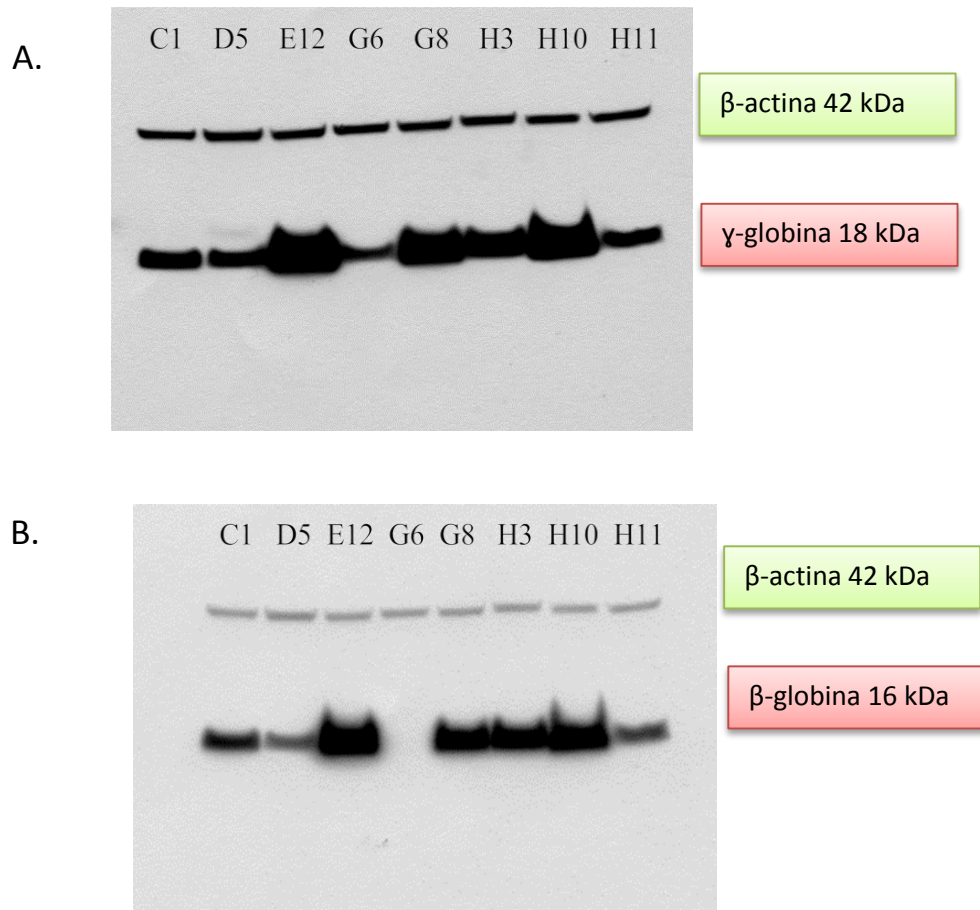


FIG.42: western blot della  $\gamma$  e  $\beta$  globina.

## 6. Discussione

La  $\beta$  talassemia è una delle malattie genetiche più comuni in tutto il mondo. Nel corso degli ultimi trent'anni si sono compiuti numerosi passi avanti nella comprensione della regolazione e funzionamento dei geni globinici e nel miglioramento delle terapie. Ad oggi l'unica arma terapeutica è rappresentata dal trapianto allogenico di CSE. Tuttavia la maggior parte dei pazienti talassemici è privo di un donatore HLA compatibile e perciò non ha attualmente speranze di guarigione definitiva. Risulta perciò di fondamentale importanza nella ricerca scientifica volta al trattamento di questa severa emoglobinopatia l'identificazione di nuovi approcci terapeutici. La terapia genica attraverso la manipolazione delle cellule staminali ematopoietiche autologhe ha dimostrato di essere la più promettente alternativa alle terapie convenzionali. Tre differenti trials clinici basati sull'utilizzo di vettori lentivirali in grado di trasferire una cassetta di espressione per il gene HBB, e garantire la sintesi *de novo* di catene  $\beta$  globiniche, sono al giorno d'oggi in fase di sperimentazione sull'uomo (referenze). Tuttavia dati di letteratura mostrano che anche una terapia genica basata sul ripristino dell'espressione dei geni HBG potrebbe essere indicata per il trattamento delle  $\beta$  emoglobinopatie, in quanto il fenomeno dell'HPFH rappresenta il principale fattore modificatore del fenotipo talassemico. È noto infatti che composti eterozigoti per la  $\beta$  talassemia e per dHPFH evidenziano un fenotipo talassemico lieve (58); composti eterozigoti per una mutazione  $\beta$  talassemica e per mutazioni HPFH non

delezionali presentano fenotipicamente una forma di  $\beta$  talassemia intermedia (101, 102). Inoltre, è stato osservato che Individui omozigoti per dHPFH, che possiedono esclusivamente HbF nel loro sangue, sono clinicamente normali e mostrano un ematocrito elevato con modesta alterazione degli indici eritrocitari (103, 104). Basandoci su tali presupposti in questo studio abbiamo cercato di sviluppare un nuovo approccio di terapia genica volto a riprodurre artificialmente i determinanti genetici responsabili della HPFH in età adulta, mediante l'utilizzo del sistema CRISPR/Cas9 associato al *pathway* di riparo NHEJ. Questo nuovo approccio eviterebbe la mutagenesi inserzionale attribuibile all'integrazione cromosomica casuale di un gene trasferito da vettori lentivirali, e il rischio di rigetto insito nel trapianto allogenico. Sfruttando la maggiore percentuale di *editing* che si ottiene mediante la via del NHEJ tale approccio ovvia al problema della bassa percentuale di eventi di *editing* ottenuta mediante ricombinazione omologa, che generalmente non arriva alla soglia minima di efficacia terapeutica richiesta. Con questa tecnologia abbiamo creato una serie di indels *in vitro* nella sequenza target localizzata a livello dei promotori dei geni HBG, sede di differenti mutazioni HPFH.

Tra i risultati ottenuti quelli più significativi sono stati conseguiti negli esperimenti coi singoli cloni ottenuti dalle diluizioni seriali. In questi cloni, derivate da singole cellule, è stato possibile genotipizzare gli eventi di *editing* al sito target, identificare la presenza di delezioni più o meno estese e saggiare le loro ripercussioni a livello

trascrizionale e traduzionale. Dei sei cloni in cui l'*editing* è stato confermato dal sequenziamento, tutti, eccetto il clone D5, hanno mostrato un incremento dell'espressione di  $\gamma$  globina, da un minimo di 3,48 ad un massimo di 11,5 volte il valore basale. Il clone D5, di cui non è stata possibile ottenere l'esatta genotipizzazione del promotore HBG1, probabilmente presenta delle modificazioni che compromettono la funzionalità del promotore stesso. I cloni positivi genotipizzati mostrano delezioni più o meno estese nel promotore del gene HBG2, mentre nella sequenza promotrice HBG1 sono state evidenziate delezioni dei nucleotidi -196/-197 (cloni E12, G6), fatta eccezione per il clone G8 che mostra la delezione -196/-197 in entrambi i promotori HBG2 e HBG1. La predisposizione alle delezioni sul gene HBG2 potrebbe essere attribuibile alla conformazione più aperta della sua cromatina in ragione della maggior espressione di quel gene nell'ambiente fetale delle  $\beta$ -K562 (rapporto d'espressione HBG2/HBG1: 3/1). I cloni H3 e H10 sono stati gli unici a non mostrare i tre ampliconi attesi dallo *screening* condotto sulla regione target mediante PCR. Il clone H10 mostra soltanto i due ampliconi derivanti dal promotore HBG2 e dalla regione tra i due geni HBG, probabilmente a causa della presenza di una delezione che non permette l'appaiamento del primer reverse a livello della regione del promotore HBG1, il quale si trova in prossimità della sequenza target. Il clone H3 mostra soltanto l'amplicone relativo al promotore HBG2, tale esito documenta l'eventuale presenza di una grossa delezione che comprende la perdita della parte 3' del promotore HBG2, dell'intero gene HBG2, della regione intergenica compresa tra i

due geni e la parte 5' del promotore HBG1. Tuttavia, dai dati ottenuti in PCR real time si osserva ugualmente un incremento dell'espressione di  $\gamma$  globina di 3,35 volte per il clone H3 e di 3,48 volte nel clone H10 rispetto ai controlli wt, suggerendo che la presenza di grosse delezioni comporti un arrangiamento del cluster  $\beta$  globinico in questa regione tale da non compromettere la funzionalità dei geni HBG. I dati d'espressione dei geni  $\gamma$  e  $\beta$  globinici nei singoli cloni ottenuti attraverso l'RT-qPCR sono stati confermati anche a livello proteico nei saggi di western blot.

Diversamente, le popolazioni di cellule GFP+ positive (HIGH), separate dopo 48 ore dall'elettroporazione mediante FACS *sorting* presentano una notevole riduzione nell'espressione sia di  $\gamma$ - sia di  $\beta$ - globina rispetto alla popolazione negativa GFP-. E' probabile che le differenze d'espressione dei geni globinici riscontrate tra i pool alle 48 ore e i singoli cloni alle tre settimane dalla transfezione siano attribuibili agli effetti dannosi dell'elettroporazione sullo stato di salute delle cellule dei pool esaminati a breve distanza dalla transfezione. Inoltre, è probabile che la riorganizzazione del cluster  $\beta$  globinico a livello delle sequenze promotrici dei geni HBG, a seguito dell'azione del sistema CRISPR/Cas9, richieda maggior tempo per il ripristino dell'espressione dei geni target. Tali supposizioni suggeriscono che 48 ore non siano sufficienti per poter osservare gli effetti sull'espressione di un evento di *editing* a carico di una sequenza regolatrice, diversamente da quanto accade per il *knockout* genico.

Tuttavia i risultati preliminari ottenuti in questo studio supportano la nostra ipotesi che la creazione di mutazioni indels, tramite il sistema CRISPR/Cas9 in specifiche regioni dei promotori dei geni HBG, siano in grado di stimolare la sintesi di HbF nella linea cellulare  $\beta$ -K562.

Contiamo di confermare il possibile ripristino dell'espressione dei geni HBG in precursori eritroidi BFU-E e in cellule staminali ematopoietiche CD34<sup>+</sup> derivate da pazienti affetti da  $\beta$  talassemia. Se i risultati su questi sistemi più fisiologici confermeranno quelli ottenuti nelle cellule  $\beta$ -K562, si potranno porre le basi per lo sviluppo di un'innovativa e promettente terapia alternativa basata sul trapianto autologo di CSE modificate a livello delle sequenze promotrici  $\gamma$  globiniche mediante *Genome-Editing* al fine di curare definitivamente questa grave anemia ereditaria.

## 7. Bibliografia

1. Charlotte W. Pratt, Donald Voet, e Judith G. Voet, Fondamenti di Biochimica, Zanichelli, 2007.
2. David L. Nelson, Michael M. Cox., I principi di biochimica di Lehninger, Zanichelli, 2006.
3. Higgs D. R., et al., A review of the molecular genetics of the human  $\alpha$  globin gene cluster. Blood 73:1081, 1989.
4. Cao A., Dallapiccola B., Notarangelo L.D., Malattie Genetiche, Piccin 2004.
5. Efstratiadis A, et al. The structure and evolution of the human  $\beta$ -globin gene family. Cell, 21, 653-668, 1980
6. Schroeder W. A., et al., Evidence for multiple structural genes for the  $\gamma$  chain of human fetal hemoglobin. Proc. Natl. Acad. Sci. USA 60:537, 1968.
7. Nute P. E., et al., The G $\gamma$  and A $\gamma$  hemoglobin chains during human fetal development. Am. J. Hum. Genet. 25:271, 1973.
8. Curtin P., et al., A distant gene deletion affects  $\beta$ -globin gene function in an atypical  $\gamma\delta\beta$ -thalassemia. J. Clin. Invest, 76, 1554, 1985.
9. Higgs D. R., A major positive regulatory region located far upstream of the human  $\alpha$ -globin gene locus. Genes Dev. 4:1588, 1990.
10. Li Q., et al., Locus control regions: coming of age at a decade plus. Trends Genet. 15:403-408, 1999.
11. Grosveld F., Activation by locus control regions? Curr. Opin. Genet. Dev. 9:152-157, 1999.
12. Stamatoyannopoulos G. and Grosveld F., Hemoglobin Switching. In: Stamatoyannopoulos G., Majerus P.W., Perlmutter R. M., et al. The molecular basis of blood diseases. Philadelphia: Saunders, 2001, 135-82.

13. Fraser P., et al., Each hypersensitive site of the human beta globin locus control region confers a different developmental pattern expression to globin genes. *Genes Dev.* 7:106, 1993.
14. S and Talbot D. The regulation of human globin gene switching. *Phil. Trans. R. Soc. Lond. B*, 339, 183-191, 1993.
15. Sankaran V. G. and Orkin S. H., The switch from fetal to adult hemoglobin, *Cold Spring Harb Perspect Med* 2013, 3:a011643.
16. Orkin S. H., GATA-binding transcription factors in hematopoietic cells. *Blood* 80:575, 1992.
17. Martin D. I. K., et al., Increased  $\gamma$ -globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor, *Nature* 338:435, 1989.
18. Munugalavadla V., Repression of c-kit and its downstream substrates by GATA-1 inhibits cell proliferation during erythroid maturation, *Mol. Cell Biol.* 25:6747-6759, 2005.
19. Rylski M., GATA-1-mediated proliferation arrest during erythroid maturation. *Mol. Cell Biol.* 23:5031-5042, 2003.
20. Miccio A., NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development, *EMBO J.* 2010 Jan 20; 29(2): 442–456.
21. Grass J. A. et al., GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling, *PNAS* vol. 100 no. 15 8811–8816, 2003.
22. Zhou M., Friend of GATA-1 Represses GATA-3–dependent Activity in CD4+ T Cells, *J Exp Med.* 194(10): 1461–1471, 2001.
23. Andrews N. C., et al., Globin locus control region enhancer binding factor (NF-E2): An hematopoietic-specific basic-leucine zipper protein, *Nature* 362:722, 1993.
24. Ney P. A., Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein, *Nucleic Acid Res.* 18:6011, 1990.



25. Bieker, J. J., Kruppel-like factors: three fingers in many pies, *J.Biol.Chem.* 276:34355-34358, 2001.
26. Im H. et al., Chromatin domain activation via GATA-1 utilization of a small subset of dispersed GATA motifs within a broad chromosomal region, *Proc Natl Acad Sci* 102: 17065–17070, 2005.
27. Hodge D., et al. A global role for EKLF in definitive and primitive erythropoiesis. *Blood*;107 8: 3359–70, 2006.
28. Zhou D., et al., KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching, *Nat Genet* 42:742-4, 2010.
29. Uda M., et al., Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci USA*, 105, 1620-5, 2008.
30. Liu H., et al., Functional studies of BCL11A: characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells. *Mol Cancer*, 5, 18, 2006.
31. Sankaran V. G., et al., Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* 322:1839-1842, 2008.
32. Sankaran V. G., et al., Developmental and species-divergent globin switching are driven by BCL11A. *Nature* 460:1093-1097, 2009.
33. Borg J., et al. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet.* 2010; 42: 801-5)
34. Cao A., et al., Recent advances in  $\beta$ -thalassemias, *Pediatric Reports* 3:e17, 2011.
35. Ferrari S., et al., SRY, like HMG1, recognizes sharp angles in DNA, *EMBO J* 11: 4497–4506, 1992.
36. Lefebvre V., et al., A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J.* 17:5718-5733 1998.
37. Murakami A., et al., SOX6 binds CtBP2 to repress transcription from the Fgf-3 promoter. *Nucleic Acids Res.* 29:3347-3355, 2001.

38. Xu J., et al., Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev*, 24, 783-98, 2010.
39. Cooney A.J., et al., Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTC A response elements, allowing COUP-TF to repress hormonal induction of the vitamin D<sub>3</sub>, thyroid hormone, and retinoic acid receptors. *Mol Cell Biol*, 12, 4153-63, 1992.
40. Ronchi A.E., et al., Differential binding of the NFE3 and CP1/NFY transcription factors to the human gamma- and epsilon-globin CCAAT boxes. *J Biol Chem*, 270, 21934-41, 1995.
41. Liberati C., et al., DNA binding of NF-Y: the effect of HMGI proteins depends upon the CCAAT box. *FEBS Lett*, 433, 174-8, 1998.
42. Liberati C., et al., Cooperation and competition between the binding of COUP-TFII and NF-Y on human epsilon- and gamma-globin gene promoters. *J Biol Chem*, 276, 41700-9, 2001.
43. Wahlberg K., et al., The HBS1L-MYB intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells, *Blood* 114:1254-62, 2009.
44. Sankaran V. G., et al., MicroRNA-15<sup>o</sup> and 16-1 crt via MYB to elevate fetal globin hemoglobin expression in human trisomy 13, *Proc Natl Acad Sci U S A*, 2011.
45. Lettre G., et al., DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell diseases, *Proc Natl Acad Sci U S A* 105:11869-74, 2008.
46. Olivieri N.F. and Weatherall D.J., The therapeutic reactivation of fetal haemoglobin. *Hum Mol Genet*, 7, 1655-8, 1998.
47. Cao A. and Kan Y. W., The prevention of thalassemia, *Cold Spring Harb Perspect Med* 3:a011775, 2012.
48. Cao A., et al. Genotype-phenotype correlations in beta-thalassemias, *Blood Rev* 8:1-12, 1994.
49. Thein S. L., Genetic modifiers of beta-thalassemia, *Haematologica* 90:649-60, 2005.

50. Cao A. and Galanello R., Beta-thalassemia, *Genet Med* 12(2):61-76, 2010.
51. Gonzalez-Redondo, J.H., et al. Clinical and genetic heterogeneity in black patients with homozygous  $\beta$ -thalassemia from the Southeastern United States, *Blood* 72:1007, 1988.
52. Hill A.V., et al.,  $\beta$ -Thalassemia in Melanesia: Association with malaria and characterization of a common variant, *Blood* 72:9, 1988.
53. Lapoumeroulie C., et al.,  $\beta$  thalassemia due to a novel mutation in IVS-1 sequence donor site consensus sequence creating a restriction site.
54. Tamagnini G. P., et al.,  $\beta^+$  Thalassemia-Portuguese type: clinical, hematological and molecular studies of a newly defined form of  $\beta$ thalassemia, *Br. J. Haematol.* 54:189, 1983.
55. Hardison R. C., et al., HbVar: A relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. *Human Mutation*, 19: 225-233, 2002.
56. Steinberg MH, Forget BG, Higgs D Nagel RL, *Disorders of Haemoglobin*, Cambridge University Press, Cambridge, 2001.
57. Galanello R. and Origa R., Beta-thalassemia, *Orphanet Journal of Rare Diseases* 2010, 5:11
58. Stamatoyannopoulos G. and Grosveld F., *Thalassemias*. In: Stamatoyannopoulos G., Majerus P.W., Perlmutter R. M., et al. *The molecular basis of blood diseases*. Philadelphia: Saunders, 2001, 157-205.
59. Cao A., et al., Thalassaemia types and their incidence in Sardinia, *J Med Genet* 1978;15:443- 447 doi:10.1136/jmg.15.6.443.
60. Rosatelli M. C., et al., Molecular characterisation of beta-thalassemia in Sardinian population, *Am J Hum Genet.* 1992;50:422–426.
61. Wood, W.G. (2001) Hereditary persistence and fetal hemoglobin and  $\delta\beta$  thalassemia. In *Disorders of Hemoglobin* (Steinberg, M.H., Forget, B.G., Higgs, D.R., and Nagel, R.L., eds.). pp. 356-388, Cambridge University Press, Cambridge.
62. De Souza G. C., et al., What influences Hb Fetal production in adulthood?, *Rev Bras Hematol Hemoter.* 2011;33(3):231-6

63. Garner C., et al., Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood*, 95: 342-346, 2000.
64. Nussbaum RL, McInnes RR, Willard HF. *Genética médica*. 7th ed. Rio de Janeiro: Guanabara koogan, 2007.
65. Olave I. A., et al., Purification and identification of proteins that bind to the hereditary persistence of fetal hemoglobin-198 mutation in the  $\gamma$ -globin gene promoter, *J Biol Chem*. 2007;282(2): 853-62.
66. Ottolenghi S, Mantovani R, Nicolis S, Ronchi A, Giglioni B (1989) DNA sequences regulating human globin gene transcription in nondeletional hereditary persistence of fetal haemoglobin. *Hemoglobin*, 13: 523-541.
67. Rochette J., et al., Fetal hemoglobin levels in adults. *Blood Reviews*, 8: 213-224, 1994.
68. Forget B. G., Molecular basis of hereditary persistence of fetal hemoglobin, *Annals of the New York Academy of Sciences*, 850: 38-44, 1998.
69. Jane S.M. and Cunningham J. M., Understanding fetal globin gene expression. A step towards effective HbF reactivation in haemoglobinopathies. *British Journal of Haematology*, 102: 415-422, 1998.
70. Dini G., et al., Transplantation of bone marrow from unrelated donors in Italy: activity and results. *Ann Ist Super Sanita*. 1999;35(1):7-11.
71. Gaziev J. and Lucarelli G., Stem cell transplantation for hemoglobinopathies. *Curr Opin Pediatr*. 2003; 15: 24-31.
72. Bank A, Dorazio R, Leboulch P., A phase I/II clinical trial of beta-globin gene therapy for beta-thalassemia. *Ann N Y Acad Sci*. 2005; 1054:308–316.
73. Sadelain M., et al., Strategy for a multicenter phase I clinical trial to evaluate globin gene transfer in beta-thalassemia. *Ann N Y Acad Sci*. 2010;1202:52–58.
74. Roselli E. A., Correction of beta-thalassemia major by gene transfer in haematopoietic progenitors of pediatric patients. *EMBO Mol Med*. 2010 Aug;2(8):315-28. doi: 10.1002/emmm.201000083.
75. Sankaran V. D. and Weiss M. J., Anemia: progress in molecular mechanisms and therapies, *Nat Med* doi:10.1038/nm.3814, 2015.

76. Dong A., et al., Gene therapy for hemoglobinopathies: progress and challenges. *Transl Res.* 2013 161(4): 293–306. doi:10.1016/j.trsl.2012.12.011.
77. Ran F. A., et al., Genome engineering using the CRISPR-Cas9 system, *Nat Protoc.* 8 (11):2281-2308, 2013.
78. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819–823.
79. Mali P., et al., RNA-guided human genome engineering via Cas9. *Science* 339: 823–826, 2013.
80. Xie F., et al., Seamless gene correction of  $\beta$ -thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggybac, *Genome Res.* 2014 Sep;24(9):1526-33.
81. Silva G., et al., Meganucleases and Other Tools for Targeted Genome Engineering: Perspectives and Challenges for Gene Therapy, *Curr Gene Ther.* 2011 Feb; 11(1): 11–27.
82. Urnov F. D. et al., Genome editing with engineered zinc finger nucleases, *Nature Reviews Genetics* 11, 636-646, 2010.
83. Provasi E. et al., Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer, *Nat Med* doi:10.1038/nm.2700, 2012.
84. Joung J. K. and Sander J. D., TALENs: a widely applicable technology for targeted genome editing, *Nature Reviews Molecular Cell Biology* 14, 49-55, 2013.
85. Boisser S., et al., megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering, *Nucleic Acids Res.* 2014 Feb;42(4):2591-601.
86. Peng X., et al., Both TALENs and CRISPR/Cas9 directly target the HBB IVS2-654 (C>T) mutation in  $\beta$ -thalassemia-derived iPSCs, *Nat Scien Rep* 5:12065, 2015.
87. Kirchner M. and Schneider S., CRISPR-Cas: From the Bacterial adaptive immune system to a versatile tool for genome engineering, *Angew. Chem. Int. Ed.* 54, 2-9, 2015.
88. Sander J. D. and Joung K., CRISPR-Cas systems for editing, regulating and targeting genomes, doi:10.1038/nbt.2842, 2014.

89. Shalem O. et al., Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells, 10.1126/science.1247005, 2013.
90. Cradick T. J., et al., CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* 41: 9584–9592, 2013.
91. Nekrasov V., Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat Biotechnol* 2013;31:691–3.
92. Wu Y., et al., Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9, *Cell Stem Cell* 13, 659–662, 2013.
93. Hwang W.Y., et al., Efficient genome editing in zebrafish using a CRISPR-Cas system *Nature Biotechnology* 31:3: 227-9, 2013.
94. Basset A. R., et al., Highly Efficient Targeted Mutagenesis of *Drosophila* with the CRISPR/Cas9 System, *Cell Rep.* 2013 Jul 11; 4(1): 220–228.
95. Friedland A. E., et al., Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system, 2013; doi:10.1038/nmeth.2532.
96. Nishimasu H., et al., Crystal structure of Cas9 in complex with guide RNA and target DNA, *Cell* 156, 935-949, 2014.
97. Harrison M. M., et al., A CRISPR view of development, *Genes Dev.* 2014 28: 1859-1872
98. Senis E., et al., CRISPR/Cas9-mediated genome engineering: an adeno-associated viral (AVV) vector toolbox, *Biotechnol. J.* 9, 1402-1412, 2014.
99. Durlak M., et al., A Novel High-Content Immunofluorescence Assay as a Tool to Identify at the Single Cell Level  $\gamma$ -Globin Inducing Compounds, *PLoS One.* 2015 Oct 28;10(10):e0141083
100. Livak K.J. and Schmittgen T.D., Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method (2001) *Methods*, vol 24, n°4, pp. 402-408.
101. Fessas P., and Stamatoyannopoulos G., Hereditary persistence of fetal hemoglobin in Greece: A study and a comparison. *Blood* 24, 223-240, 1964.

102. Tate V. E., et al., The british form of hereditary persistence of fetal hemoglobin results from a single base mutation adjacent to an S1 hypersensitive site 5' of the A $\gamma$  globin gene, Blood 68, 1389-1393, 1986.
103. Wheeler J. T., and Krevans J. R., The homozygous state of persistent fetal hemoglobin and interaction of persistent fetal hemoglobin with thalassemia, Bull. Johns Hopkins Hosp. 109, 217-233, 1961.
104. Charache S., et al., The negro variety of hereditary persistence of fetal hemoglobin in a mild form of thalassemia. Br. J. Haematol. 34, 527-534, 1976.