



SAPIENZA
UNIVERSITÀ DI ROMA

Emoglobinopatie: caratterizzazione genetica molecolare e flussi migratori in una regione non endemica.

Dottorato di Ricerca in Biologia Umana e Genetica Medica
Dipartimento di Medicina molecolare
XXXIII CICLO

Ylenia Barbanera
Matricola 1792353

Relatore Interno
Prof.ssa Paola Grammatico

Relatore Esterno
Dott. Paolo Gorello

Alla mia famiglia, il senso del mio tutto.

SOMMARIO

1.INTRODUZIONE.....	6
1.1.L'EMOGLOBINA.....	6
1.1.a.Caratteristiche dell'emoglobina.....	6
1.1.b. I geni globinici e la loro espressione.....	6
2. MALATTIE GENETICHE DELL'EMOGLOBINA.....	20
2.1. Classificazione delle emoglobinopatie.....	20
2.2. Emoglobine qualitative.....	20
2.2.a.1. Anemia falciforme.....	20
2.2.a.2. I diversi aplotipi e le differenze molecolari.....	23
2.2.b.1 HbC.....	24
2.3. HbS/HbC.....	25
2.4. Emoglobine M.....	25
2.5. Emoglobine chimeriche.....	26
7. L'ereditarietà e la variabilità.....	55
7.1. Ereditarietà materna.....	55
7.2 Tasso evolutivo e mutazioni.....	56
7.3. Gli studi evoluzionistici.....	58
7.3.1 L'orologio molecolare.....	58
7.3.2 La variabilità del mtDNA applicata agli studi evoluzionistici di popolazioni umane.....	59
7.3.3. Gli aplogruppi e la storia dell'uomo.....	62
7.3.4. Gli aplogruppi principali.....	63
7.3.5. Il DNA mitocondriale europeo.....	64
8. SCOPO DELLA TESI.....	67

9. MATERIALI E METODI DEL PRIMO STUDIO.....	68
9.1. Casistica.....	68
9.2. Pazienti ed estrazione del DNA.....	71
9.3. Analisi del DNA genomico.....	72
9.4. Analisi del DNA mitocondriale.....	73
10. MATERIALI E METODI DEL SECONDO STUDIO.....	74
10.1. Casistica.....	74
10.2. Quantificazione del DNA genomico.....	76
10.3. NGS: custom bundle solution by Sophia Genetics, messa a punto del pannello.....	76
10.4. NGS: custom bundle solution by Sophia Genetics.....	79
11. RISULTATI DEL PRIMO STUDIO.....	92
11.1. Analisi dei SNPs modulatori del fenotipo clinico e distribuzione geografica.....	92
11.2. Analisi delle variazioni della sequenza di mtDNA e correlazione con le varianti dei geni modificatori.....	98
12. DISCUSSIONE DEL PRIMO STUDIO.....	106
12.1. Analisi dei SNPs modulatori del fenotipo clinico e distribuzione geografica.....	106
12.2. Analisi delle variazioni della sequenza di mtDNA e correlazione con le varianti dei geni modificatori.....	107
13. RISULTATI DEL SECONDO STUDIO.....	111
13.1. Risultati della validazione del pannello NGS.....	111
13.2. Validazione dei genotipi dei campioni affetti da emoglobinopatie...123	
13.3. Varianti identificate nei fratelli algerini e successivo studio familiare.....	129
14. DISCUSSIONE DEL SECONDO STUDIO.....	145
15. CONCLUSIONI.....	150
BIBLIOGRAFIA.....	152
RINGRAZIAMENTI.....	168
Appendice A.....	169
Appendice B.....	173

Appendice C.....	174
Appendice D.....	177
Appendice E.....	181
Appendice F.....	185

1. INTRODUZIONE

1.1. L'EMOGLOBINA

1.1.a. Caratteristiche dell'emoglobina

L'emoglobina (indicata con il simbolo Hb) è una proteina globulare la cui struttura quaternaria consta di quattro subunità. È solubile, di colore rosso (cromoproteina) ed è presente nei globuli rossi del sangue dei vertebrati. È responsabile del trasporto dell'ossigeno molecolare da un compartimento ad alta concentrazione di O₂ ai tessuti che ne hanno bisogno. L'emoglobina viene sintetizzata inizialmente a livello dei proeritoblasti, precursori dei globuli rossi, che diventeranno poi eritrociti, con il nucleo, e quindi matureranno in globuli rossi perdendo il nucleo dalle cellule, per poi essere inviati dal midollo osseo al sangue.

La componente proteica è costituita da quattro catene polipeptidiche uguali a due a due (2 catene di tipo α e 2 catene di tipo non- α) permettendo una differenziazione in: emoglobine embrionali: Hb Gower I ($\zeta_2\varepsilon_2$), Hb Portland ($\zeta_2\gamma_2$), Hb Gower II ($\alpha_2\varepsilon_2$); emoglobina fetale HbF ($\alpha_2\gamma_2$) e emoglobine adulte HbA₂ ($\alpha_2\delta_2$) e HbA ($\alpha_2\beta_2$). Ognuna di esse è poi a sua volta legata covalentemente ad un gruppo prostetico detto "eme", costituendo la componente non proteica; esso è formato da una molecola di protoporfirina che coordina uno ione ferro Fe²⁺, il quale sporge leggermente dal piano della molecola. L'emoglobina è inoltre una proteina allosterica perché il legame dell'ossigeno ad un sito determina modificazioni conformazionali a carico degli altri siti di legame.

1.1.b. I geni globinici e la loro espressione

L'emoglobina più frequente nell'adulto è l'HbA ($\alpha_2\beta_2$). I geni che codificano per i due diversi tipi di catene globiniche, α e β , sono organizzati in due raggruppamenti, detti *cluster* globinici (*Figura 1*).

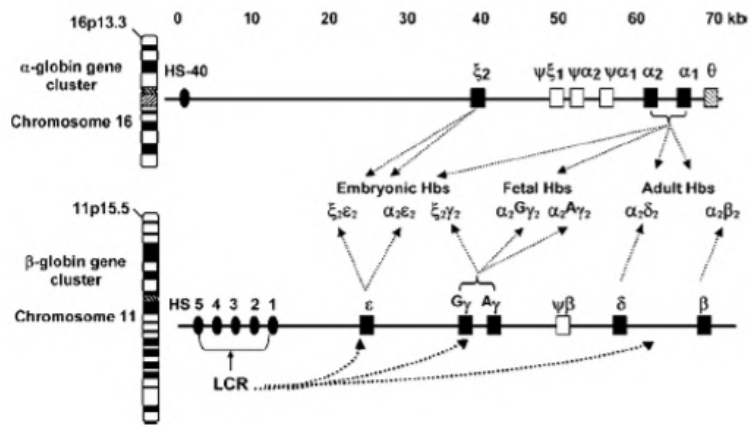


Figura 1. Cluster dei geni α e β globinici. Struttura ed espressione dei geni della globina durante ontogenesi. Viene anche schematizzata l'interazione specifica dello stadio tra la β-LCR e il cluster.

Le catene di tipo α e di tipo β mostrano una forte omologia di sequenza ad indicare una loro origine evolutiva comune (forse correlato al gene attuale per la mioglobina).

Si ipotizza che da un gene ancestrale, dopo duplicazione e diversificazione della funzione delle due copie, si siano originati i geni da cui si sono formate le famiglie α e β globinica. In base alla struttura dei geni attuali e al confronto con altre specie, si può ipotizzare che il gene che ha dato origine alla famiglia delle β-globine abbia incrementato la dimensione mediante aumento della dimensione del secondo introne molto precocemente nel corso dell'evoluzione.

Le duplicazioni e le delezioni geniche avvengono spesso a causa di un cattivo allineamento delle sequenze durante il crossingover, detto "crossingover ineguale".

Le duplicazioni possono insorgere anche per trasposizione. In seguito a duplicazione genica, le copie separate di un gene possono andare incontro a cambiamenti nella sequenza. In alcuni casi avvengono mutazioni che rendono non funzionale una copia del gene, creando uno pseudogene. In altri casi, variazioni della sequenza nucleotidica possono portare a funzioni differenti da parte della proteina prodotta [1,2].

Il **cluster di tipo α** , di circa 30Kb, è localizzato nella porzione distale del braccio corto del cromosoma 16, a livello della banda 16p13.3. Il *cluster* α globinico contiene due geni α -globinici duplicati ($\alpha 1$ e $\alpha 2$), un gene α -simile embrionale ($\zeta 2$) e tre pseudogeni ($\psi\zeta 1$, $\psi\alpha 2$, $\psi\alpha 1$) arrangiati dal telomero verso il centromero, nel seguente ordine: 5'- $\zeta 2$ - $\psi\zeta 1$ - $\psi\alpha 2$ - $\psi\alpha 1$ - $\alpha 2$ - $\alpha 1$ - $\theta 1$ -3'. All'estremo 3' si trova il gene θ a funzione non nota. Così come nel *cluster* β , anche nel *cluster* α il gene embrionale $\zeta 2$ è localizzato vicino alla regione 5' e i geni globinici adulti ($\alpha 2$ - $\alpha 1$) vicino alla regione 3'. Inoltre, il *cluster* α -globinico manca di un gene espresso solo nel periodo fetale, ma ha due geni pienamente funzionali ($\alpha 2$ - $\alpha 1$), che ricoprono l'espressione fetale e adulta, combinandosi con il gene γ -globinico, durante l'eritropoiesi fetale e con i geni δ e β globinici nello stadio adulto dello sviluppo. Anche se i geni derivano da un progenitore comune ed hanno una struttura piuttosto simile, la regolazione del gene per le α -globine differisce da quello β in molti aspetti [3].

La competizione tra i geni non influenza la regolazione del *cluster* α , e tutti i geni sono appropriatamente regolati quando singolarmente associati al β -LCR o all'elemento HS posizionato a -40. Tuttavia come il gene ϵ , il gene embrionale $\zeta 2$ è silenziato nell'adulto mediante sequenze vicine al proprio promotore, ma diverse da quelle inibitorie presenti a monte del gene ϵ [4].

Il *cluster* α occupa la posizione telomerica del cromosoma 16, una regione alquanto instabile e variabile, ricca in CG (54%), con una configurazione costitutivamente aperta della cromatina e un'alta densità di geni adiacenti non globinici costitutivamente espressi. Queste sequenze ripetute appartengono alla famiglia *Alu*, mentre il *cluster* β è ricco in elementi ripetuti del tipo LINE.

Studi di ibridazione *in situ* hanno permesso di definire la distanza del *locus* dal centromero durante il ciclo cellulare, rivelando un'ulteriore differenza tra i due *clusters*: mentre il *cluster* β , nelle cellule eritroidi, si sovrappone al centromero e si allontana da esso passando dallo stato inattivo a quello attivo dell'espressione, nelle cellule non eritroidi il *cluster* β risiede sempre in prossimità del centromero; d'altra parte il *cluster* α non si sovrappone mai al centromero, la sua espressione è sempre attiva e risulta indipendente sia dal ciclo cellulare, che dall'istotipo.

Il **cluster di tipo β** , di circa 70 kb, è collocato nella regione distale del braccio corto del cromosoma 11, all'interno della banda 11p15.5 e contiene sette geni globinici; questi sono i geni strutturali ϵ , γ , che sono presenti come coppia di geni $\epsilon\gamma$ e $\delta\gamma$, δ e β , e gli pseudogeni $\psi\beta 1$ e $\psi\beta 2$ [5]. Nel gene beta globinico la regione 5' non tradotta (5' UTR) è costituita da 50 nucleotidi ed ha per limite al 5' il sito "Cap" (sito di inizio della trascrizione) e come limite al 3' il codone di inizio della traduzione ATG. Nella regione 5' UTR vi sono due sequenze altamente conservate. La prima costituita dall'esanucleotide CTTCTG, è localizzata dalla posizione +8 alla +13 a valle del sito cap.

La seconda sequenza CACCATG comprende il codone iniziale ATG. La regione 3' UTR è costituita da 132 nucleotidi e si trova tra il codone di stop della traduzione e la coda poli A. Venti nucleotidi a monte del sito di attacco della coda poli-A si trova la sequenza conservata AATAAA che serve da segnale per il taglio della parte 3' terminale del trascritto primario e l'aggiunta della coda di poli-A, la quale conferisce al mRNA stabilità e ne potenzia la traduzione.

Il promotore del gene β -globinico, localizzato in posizione 5' rispetto al sito di inizio della trascrizione (Cap), comprende 3 elementi altamente conservati: il TATA box, situato tra la posizione nucleotidica -28 e la -31, il CAAT box dalla posizione -72 alla -76 ed il CACCC box che è duplicato (prossimale da -86 a -90; distale da -101 a 105) [3].

Esistono comunque delle differenze nei promotori dei diversi geni del cluster beta globinico, ad esempio il promotore gamma globinico presenta una duplicazione del motivo CAAT e un singolo CACCC, mentre il promotore Beta globinico presenta una duplicazione del motivo CACCC e un singolo CAAT.

Il promotore del gene δ rispetto al promotore del gene β presenta la mancanza del CACCC box, una mutazione del CCAAT box (GCAAT), e una ridotta distanza tra CAAT e TATA box. Questi specifici elementi sono dei regolatori della trascrizione che interagiscono con la LCR (*Locus Control Region*) e con fattori di trascrizione ubiquitari ed eritroide-specifici per promuovere la trascrizione [6].

LCR è localizzato circa 50kb a monte del gene Beta ed è costituito da quattro siti eritroidi specifici ipersensibili alla DNasi I (HS-1, HS-2, HS-3, HS-4), che sono dei marcatori specifici per l'interazione DNA-proteina. Ogni sito HS è costituito da una combinazione di sequenze di DNA necessarie per l'interazione con specifici fattori trascrizionali tra cui i più importanti sono GATA-1 (GATA indica la sequenza di DNA riconosciuta dal fattore GATA-1), NF-E2 (Nuclear factor erythroid 2), EKLF1 (Erythroid Kruppel-like factor) e FOG1 (Friend of GATA-1) [7].

LCR è fondamentale per l'attivazione della trascrizione: un'eventuale delezione di tale regione comporta l'inibizione dell'espressione di tutte le proteine codificate dai geni del cluster globinico contenuti nel cromosoma 11, provocando la β -talassemia [8,9]. Le funzioni svolte dall'LCR sono la costituzione di una regione più accessibile ai fattori di regolazione, detta regione cromosomica aperta e rappresenta il sito in cui sono presenti sequenze nucleotidiche che esercitano un'azione stimolatoria, inibitoria o azioni più complesse interagendo con fattori trascrizionali responsabili della regolazione dell'espressione dei geni globinici nelle cellule eritroidi, che avviene in modo differenziato durante le fasi dello sviluppo embrio-fetale [10].

Nonostante le diverse e, spesso contrastanti opinioni sulle modalità di azione dell'LCR, si è giunti alla conclusione che frammenti di HS, singoli o in combinazione, non riescono a raggiungere i livelli di attività dell'LCR integro. I primi modelli sperimentali, pur dimostrando specifiche interazioni e attivazioni tra singoli HS e geni globinici [11], sono stati superati e l'attuale modello in uso suggerisce che l'LCR funzioni come un olocomplexo di tutti gli HS, che interagisce con un solo gene globinico, durante lo sviluppo eritroide. Inoltre, l'LCR agisce sinergicamente con il promotore dei singoli geni β -globinici e con l'apparato dell'RNA polimerasi per controllare nel tempo il progressivo "spegnimento" di alcune catene a favore di altre [6,12].

Alcuni studi dimostrano che l'LCR è in grado di dirigere l'espressione globinica attraverso la formazione di *loop* di DNA, la cui formazione è dovuta all'interazione di complessi multipli DNA-proteina con i singoli promotori per le diverse globine del *cluster* β (Figura 2).

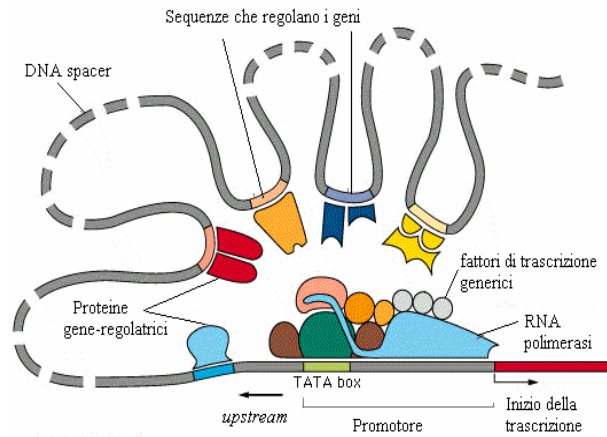


Figura 2. Esempio di una struttura a "loop". Nella figura sono schematizzate le regioni interessate dal legame di fattori trascrizionali, che a loro volta interagiscono grazie alla formazione di una struttura a *loop* con quelli riconosciuti in regioni più distali, regolando l'espressione genica a livello delle regioni promotrici dei singoli geni globinici.

L'efficacia e il tempo di interazione tra LCR e geni globinici sono influenzati dall'ordine dei geni [13], dalla distanza dall'LCR [14] e dai fattori di trascrizione specifici, che a turno intervengono a seconda dello stadio di sviluppo. Esistono diverse proteine e elementi di regolazione del DNA in cis coinvolte nella regolazione dell'espressione dei geni beta globinici.

Queste proteine funzionano come fattori trascrizionali leganti il DNA, regolatori coinvolti nella modulazione della struttura dell'eterocromatina e/o reclutano complessi trascrizionali. Sono anche coinvolti nel cambio conformazionale e nella relocalizzazione nucleare del locus globinico durante le diverse fasi dello sviluppo. Questi fattori sono proteine ematopoietiche specifiche, quali: GATA-1, EKLF (e le relative proteine), NF-E2 (e le relative proteine), e Tal-1. Inoltre in questo processo sono coinvolti anche dei fattori trascrizionali ubiquitari, come: Sp1 e USF (*Figura 3*).

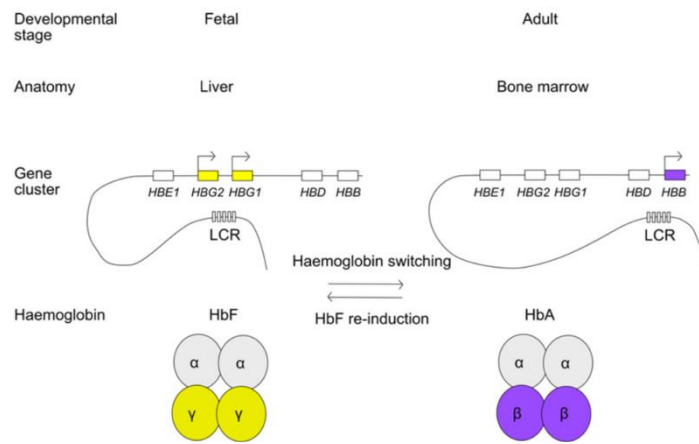


Figura 3. Schema delle proteine e dei co-regolatori che interagiscono con la porzione HS-2 del LCR e con il promotore del gene β -globinico nell'adulto. In alto è illustrata l'organizzazione del locus genico β -globinico: LCR (Locus control region), gene ϵ embrionale, gene $G\gamma$ e $A\gamma$ fetale, geni globinici nell'adulto (δ e β). All'interno della tabella troviamo tutti i fattori, proteine e co-regolatori, che interagiscono nell'espressione del gene β -globinico nell'adulto. I siti di legame per i fattori trascrizionali della porzione HS-2 del LCR, e del gene β -globinico sono indicati con dei quadrati. Gli attivatori sono indicati con cerchi colorati, mentre i repressori con cerchi neri e grigi. Infine i co-attivatori sono indicati con triangoli colorati, mentre i co-repressori da triangoli neri e grigi.

Il ruolo regolatore dell'LCR è anche fortemente supportato dal silenziamento di tutti i geni β -globinici, provocato da delezioni dell'LCR, che si verificano naturalmente nell' $\gamma\delta\beta$ -talassemia [15,16] e nell'espressione delle catene globiniche, dove è fondamentale l'interazione del promotore dei geni delle diverse catene con l'LCR. È stato dimostrato, usando un approccio moderno di reclutamento di fattori di trascrizione verso specifici elementi regolatori, che l'LCR interagisce con un solo promotore alla volta. Questo meccanismo di competizione tra i promotori è fondamentale nell'espressione delle diverse catene globiniche [17] (Figura 4).

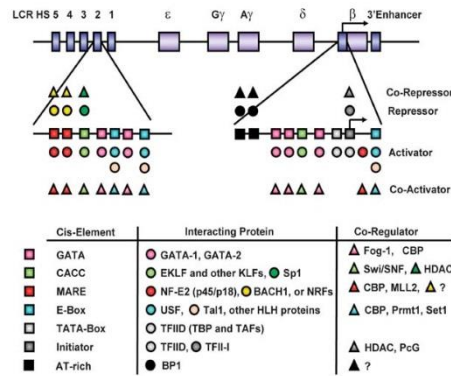


Figura 4. Produzione di due diversi tipi di emoglobina in base all'interazione con l'LCR. Nella figura viene mostrata la diversa produzione di catene emoglobiniche in base all'interazione con l'LCR. Quando l'LCR interagisce con il promotore dei geni *HBG1* e *HBG2* viene prodotta la catena γ -globinica; quando interagisce con il promotore del gene *HBB* è prodotta la catena β -globinica.

La struttura dell'emoglobina umana (Hb) cambia durante lo sviluppo (Figura 5) [18,19].

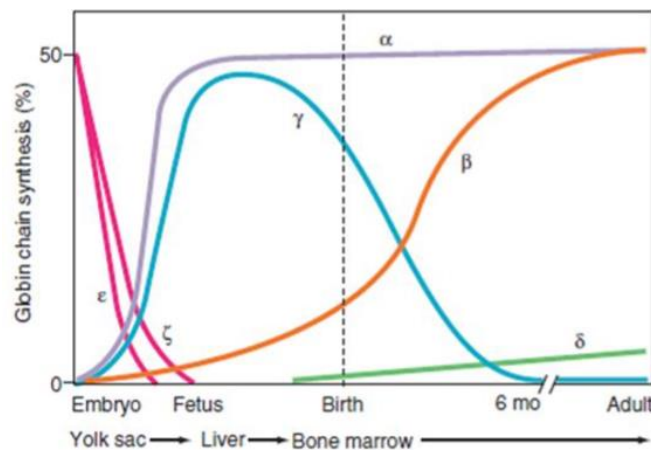


Figura 5. Rappresentazione dell'espressione nel tempo dei differenti tipi di catene globiniche. L'espressione dei geni alfa è attiva nelle fasi precoci dello sviluppo e continua durante tutta la vita. I membri γ e δ della famiglia beta sono espressi durante lo sviluppo fetale e sono disattivati appena prima la nascita. Il gene β -globina è attivato appena prima della nascita ed è espresso per tutta la vita.

Le emoglobine adulte (HbA) e quelle fetali (HbF), la cui produzione continua anche dopo la nascita, andando a costituire, per i primi mesi di vita, il 5% di tutta l'emoglobina, sono costituite da catene α che si combinano con le catene β (HbA, $\alpha_2\beta_2$), δ (HbA₂, $\alpha_2\delta_2$) o nel caso dell'emoglobina fetale, con le catene γ (HbF, $\alpha_2\gamma_2$); invece nell'embrione, le catene α -like sono rappresentate da catene ζ e si combinano con le catene ϵ (Hb Portland, $\zeta_2\epsilon_2$) o con le catene δ (Hb Gower1, $\zeta_2\delta_2$), mentre le catene α e ϵ formano l'Hb Gower2 ($\alpha_2\epsilon_2$) (Figura 6).

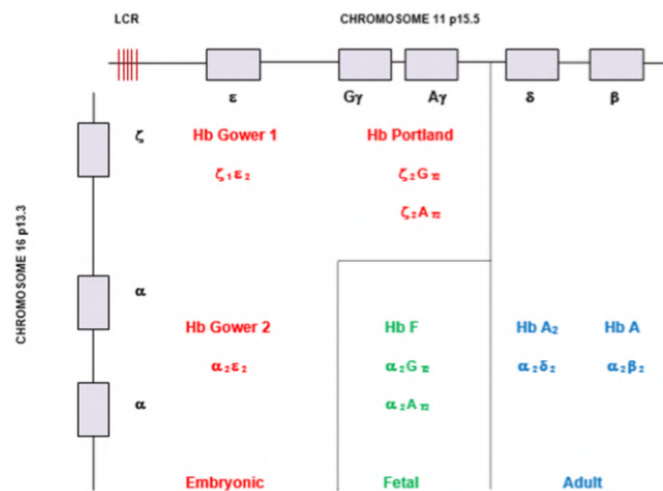


Figura 6. Differenti tipi di emoglobina nell'uomo. Le tre emoglobine embrionali vengono prodotte nei primi tre mesi di gravidanza, al termine dei quali vengono sostituite dall'emoglobina fetale, la quale a sua volta, viene sostituita dall'emoglobine adulte nei primi mesi di vita.

Rispetto all'emoglobina adulta, quella fetale ha una maggiore affinità per l'ossigeno: ciò permette un efficiente trasferimento dell'ossigeno dal sangue materno a quello fetale, attraverso la placenta.

La differente espressione delle diverse catene globiniche, dal concepimento alla vita adulta, può avvenire attraverso almeno due meccanismi: dalla competizione diretta tra un promotore del gene specifico per elementi regolatori condivisi, o dall'attivazione autonoma e/o silenziamento autonomo tramite attivatori e repressori specifici, nonché meccanismi epigenetici. Molto probabilmente, questi meccanismi non si escludono, ma si completano a vicenda rendendo completo lo switch dell'emoglobina [20-25].

Durante il periodo embrionale, sono attivi i geni responsabili della sintesi delle Hb Gower e Hb Portland, la cui espressione diminuisce progressivamente dopo le due prime settimane di gestazione. In particolare, l'espressione del gene ζ diminuisce con l'aumentare dell'espressione del gene per la globina α , invece, le globine ϵ sono sostituite dalle globine γ dopo sei settimane dal concepimento. Le globine α , dopo la ventiquattresima settimana dal concepimento, durante la fase fetale, raggiungono i livelli massimi, che rimarranno tali nell'adulto. Dopo la nascita, si ha una diminuzione della sintesi delle globine γ , tale da essere sostituite quasi completamente, al secondo anno d'età, dalle globine β . Durante la vita adulta, però, una piccola percentuale di HbF continua ad essere espressa, i cui livelli possono aumentare anche di dieci volte, e questo fenomeno può essere dovuto a diversi fattori: l'età, il sesso, e/o le peculiarità genomiche, come, ad esempio, mutazioni puntiformi nelle sequenze di DNA all'interno del *cluster* β o nei geni ad esso correlati.

L'evento più intrigante e allo stesso tempo più studiato, è lo *switching* dell'emoglobina, ossia la soppressione del gene γ -globinico, accompagnato dall'aumento complementare nell'espressione delle catene β -globiniche [15,25]. Si sa che questo processo ha un'implicazione terapeutica nella β -talassemia e nell'anemia falciforme, poiché pazienti con β -talassemia e anemia falciforme con livelli elevati di HbF mostrano anche un decorso della malattia meno grave. Dato che il gene γ -globinico può sostituire dal punto di vista funzionale il gene β difettivo [25,26], sono stati effettuati, molti studi che prevedono la riattivazione del gene γ e in particolar modo, un notevole passo avanti è stato compiuto grazie alla scoperta di nuovi fattori di trascrizione e regolatori epigenetici responsabili del silenziamento fisiologico dell'HbF [25].

Sono stati scoperti due repressori del gene γ -globinico: BCL11A e ZBTB7A. Il repressore trascrizionale BCL11A, codificato dal gene *BCL11A* che mappa nella regione 2p16.1 è una proteina con un motivo a dita di zinco espressa in numerosi tessuti ed il suo ruolo di repressore dell'espressione di HbF è stato convalidato da studi su knockdown e knockout in eritroblasti umani primari, linee cellulari eritroidi e modelli knockout di topo [17].

Studi di “Knockdown” hanno dimostrato che BCL11A è necessario per il mantenimento del silenziamento dell’espressione di HbF in progenitori eritroidi primari umani nell’adulto [27], mentre topi “Knockout” per *Bcl11a* generano un incompleto silenziamento dei geni gamma globinici [28]. BCL11A interagisce sia con fattori trascrizionali eritroidi quali GATA1, FOG1, SOX6, RUNX1, KLF1 che con il complesso NuRD (Nucleosome remodelling and deacetylase) coinvolto nel rimodellamento dell’eterocromatina [27,29]. Inoltre BCL11A occupa siti critici in mezzo al cluster genico beta globinico, incluse delle regioni che vengono delete in HPFH, promuovendo l’interazione fisica tra LCR e promotore del gene β globinico a spese del promotore del gene γ . Tutti questi risultati supportano il modello secondo il quale BCL11A coordina lo switch emoglobinico, partecipando ad un complesso multiproteico che occupa regioni specifiche del cluster genico beta globinico [29,30].

ZBTB7A, noto anche come fattore correlato alla leucemia / linfoma (LRF), è un fattore trascrizionale. Recenti studi su topi, cellule HUDEP-2 ed eritroblasti derivati da CD34 + HSPC hanno identificato ZBTB7A come un potente repressore dell’espressione genica embrionale e fetale della globina. I topi knockout *Zbtb7a* sono letali embrionali a causa di una grave anemia, mentre il knockout allo stadio adulto di *Zbtb7a* provoca una lieve anemia macrocitica. ZBTB7A lega e reprime direttamente l’accessibilità della cromatina dei geni γ -globinici nei precursori eritroidi umani [17].

Il grado di persistenza dell’HbF varia tra gli adulti ed è ampiamente controllato geneticamente. L’espressione del gene γ -globina in modo inappropriato negli adulti è associata all’HPFH, causata da delezioni di cluster HBB o mutazioni puntiformi nel promotore prossimale di HBG1 e HBG2. Mentre la persistenza di alti livelli di HbF non ha conseguenze cliniche in soggetti sani, la coinheritance dell’HPFH con uno dei due principali disturbi β -emoglobinici - la malattia da falciforme (SCD) o la β -talassemia - allevia la loro gravità clinica. Le mutazioni puntiformi associate all’HPFH portano un aumento significativo dell’HbF, che varia dal 10 al 40% dell’emoglobina totale negli eterozigoti.

Queste mutazioni sono raggruppate in due regioni, situate a 115 e 200 bp a monte del sito di inizio della trascrizione. Il denso raggruppamento delle mutazioni puntiformi associate all'HPFH suggerisce che questi sono potenziali siti di legame per i repressori trascrizionali che provocano il silenziamento autonomo del gene γ -globina.

Un esame approfondito dei fattori BCL11A e ZBTB7A associati ai promotori del gene della γ -globina ha dimostrato infatti che i due repressori si legano in regioni indipendenti a valle del sito di inizio della trascrizione dei geni della γ -globina. Nello specifico BCL11A a -115 bp e ZBTB7A a -200 bp. L'introduzione di mutazioni associate a HPFH nei siti a -200 e -115 bp ha comportato la perdita di attività di legame e un aumento dell'espressione di HbF rispetto alle cellule wild-type, imitando HPFH. I siti di legame indipendenti di BCL11A e ZFTB7A supportano uno studio precedente che ha dimostrato che le cellule a doppio knockout *ZBTB7A* e *BCL11A* avevano un aumento maggiore di HbF rispetto alle cellule a singolo knockout *ZBTB7A* o *BCL11A13*. Ciò che non è chiaro è se il legame dei due repressori sia successivo o quale ordine sia necessario per realizzare la piena riconfigurazione della cromatina [31] (Figura 7).

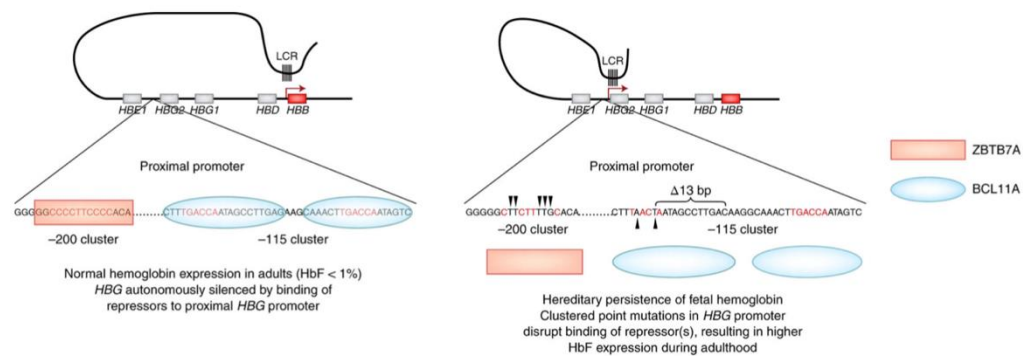


Figura 7. Le mutazioni puntiformi nei promotori del gene della γ -globina interrompono il legame dei repressori trascrizionali, con conseguente aumento dei livelli di emoglobina fetale in età adulta.

BCL11A e ZBTB7A si legano al promotore della γ -globina in due siti con mutazioni raggruppate in HPFH e inibiscono l'espressione. Tuttavia, gli individui con mutazioni associate a HPFH hanno una sequenza di consenso mutata per BCL11A e / o ZBTB7A che si legano, reprimendo l'espressione dell'emoglobina fetale durante l'età adulta. Le sequenze di consenso per BCL11A e ZBTB7A sono mostrate in rosso. Le mutazioni associate a HPFH nel cluster (frecche invertite) a -200 bp, da sinistra a destra, sono c.-202C> T / G, c. -201C> T, c. -197C> T, c. -196C> T e c. -195C> G; le mutazioni associate all'HPFH (frecche verso l'alto) nel cluster a -115 bp, da sinistra a destra, sono c.-117G> A e c. -114C> T / A / G.

KLF1 (conosciuto anche come EKLF) è un un fattore trascrizionale eritroide fondamentale per mantenere un alto livello d'espressione del gene beta globinico nell'adulto [26,32,33]. KLF1 lega la sequenza CACCC nel promotore del gene beta, con maggiore affinità rispetto al legame con la regione CACCC box nel promotore del gene γ [34,35]. Knockdown di *KLF1* in cellule progenitrici eritroidi umane genera un incremento dell'espressione di HbF mentre topi knockout per *KLF1* impedisce il silenziamento dei geni gamma globinici [36,37,38]. Inoltre la proteina KLF1 si lega al promotore di *BCL11A* attivandone l'espressione. KLF1 quindi sembra influenzare lo switching dell'emoglobina attraverso due meccanismi: attivando direttamente l'espressione del gene beta globinico nell'adulto e promovendo l'espressione di *BCL11A*, silenziatore del gene gamma globinico.

Il fattore trascrizionale ematopoietico c-Myb è un altro fattore coinvolto nella regolazione dell'HbF. Un aumento dell'espressione di c-Myb nella linea cellulare K562 inibisce l'espressione del gene γ -globinico [39]. Knockdown del gene *MYB* in progenitori eritroidi primari umani, è associato ad un incremento dell'HbF [40]. Inoltre c-Myb può legare il promotore di *KLF1* e quindi può influenzarne l'espressione attraverso un meccanismo di trans-attivazione [41].

Oltre a questi menzionati esistono altri fattori correlati con lo switch dell'emoglobina: COUP-TF, FOP (Friend of PRMT1), Ikaros, miR-15 e miR-16 (che influenzano negativamente l'espressione di c-Myb), MBD2, NF-E4, NRF2, TR2 e TR4. [40-48]. Ognuno di questi fattori, agisce in complessi multiproteici che hanno la capacità di modificare la struttura della cromatina.

L'osservazione che i promotori di *HBG1* e *HBG2* sono metilati con dinucleotidi CpG nelle cellule eritroidi adulte ha portato all'ipotesi che la metilazione abbia un ruolo nel silenziamento del gene γ -globinico nelle cellule eritroidi adulte.

La metilazione del DNA nel locus della γ -globina è riconosciuta dalla proteina MBD2 del dominio di legame alla metilcitosina (MBD), che aiuta a reclutare il complesso di silenziamento del NuRD nel locus. Questo complesso comprende due attività catalitiche, i rimodellatori della cromatina ATP-dipendenti CHD3 e CHD4, nonché le deacetilasi istoniche HDAC1 e HDAC2.

Sia BCL11A che ZBTB7A hanno dimostrato di interagire con i membri del complesso NuRD nelle cellule eritroidi, suggerendo che l'attività di NuRD può aiutare a eseguire la repressione genica da questi fattori.

Il passaggio dalla γ alla β -globina in concomitanza con le principali transizioni da feto ad adulto ha spinto la ricerca di fattori di sviluppo che potrebbero interagire con la regolazione genica della globina. I cosiddetti fattori eterocronici sono quelli che svolgono ruoli importanti nello stabilire le transizioni durante l'ontogenesi. Fattori fetali di spicco comprendono le famiglie LIN28 e IGF2BP di fattori che legano l'RNA, e importanti regolatori per adulti includono la famiglia di microRNA let-7. L'espressione dei fattori fetali LIN28B e IGF2BP1 sembra ridurre l'espressione di BCL11A, sebbene sia ancora da determinare sia il meccanismo di inibizione di BCL11A sia se ci sono bersagli eterocronici aggiuntivi oltre a BCL11A [17] (Figura 8).

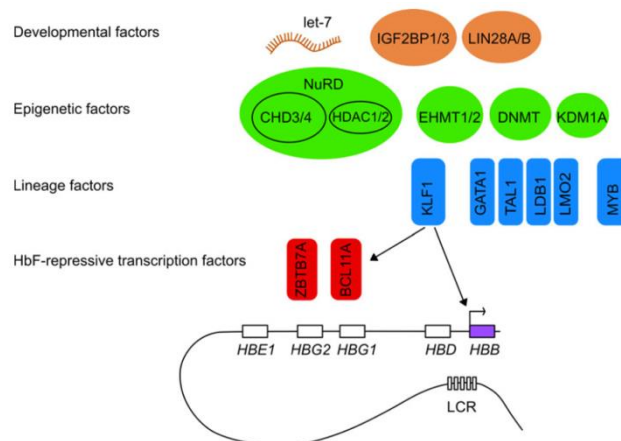


Figura 8. Macchinari molecolari coinvolti nello switch dell'emoglobina. Il silenziamento dei geni γ -globinici (*HBG1* / *HBG2*) è ottenuto dall'azione combinata di numerosi fattori, con i principali fattori rappresentati. I fattori di trascrizione BCL11A e ZBTB7A sono essenziali per il silenziamento della globina fetale. Entrambi questi fattori interagiscono con il complesso della cromatina NuRD. Altri fattori per la cromatina, come EHMT, DNMT e KDM1A, partecipano al silenziamento fetale della globina. Fattori che definiscono il lignaggio, come KLF1 e MYB, svolgono ruoli importanti, sebbene questi possano essere indiretti attraverso l'espressione coordinata di repressori diretti come BCL11A. Fattori di sviluppo, come let-7, LIN28 e IGF2BP promuovono le transizioni dal feto all'adulto in numerosi contesti cellulari, sebbene il modo in cui influenzano l'espressione genica della globina rimane incompletamente compreso.

2. MALATTIE GENETICHE DELL'EMOGLOBINA

2.1. Classificazione delle emoglobinopatie

La variabilità di mutazioni che interessano i geni globinici, con i loro differenti effetti sulla sintesi delle catene globiniche, si traduce in un ampio ventaglio di quadri clinici. Pertanto, i difetti della sintesi dell'emoglobina sono suddivisi classicamente in tre gruppi principali:

- a. Difetti qualitativi, vale a dire alterazioni in cui è sintetizzato in quantità regolare un prodotto proteico con caratteristiche diverse dal normale;
- b. Difetti quantitativi (talassemie), che generalmente interessano un solo gene, ma talora più di uno, con riduzione o assenza delle catene globiniche che però, quando sono presenti, non mostrano alterazioni della normale sequenza aminoacidica;
- c. Difetti del meccanismo che regola il calendario dell'attivazione/disattivazione genica nel corso dell'ontogenesi, definiti persistenza ereditaria di HbF (HPFH, Hereditary Persistence of Fetal Hemoglobin).

2.2. Emoglobinopatie qualitative

Sono causate prevalentemente da mutazioni missenso che portano alla sintesi di emoglobine con sequenza diversa dal normale. Queste mutazioni mostrano proprietà chimico-fisiche nuove rispetto a quelle del prodotto *wild-type* e, a seconda delle conseguenze della mutazione sulla funzione della proteina, danno origine a quadri patologici diversi fra loro.

2.2.a.1. Anemia falciforme

Il difetto genetico che dà luogo all'emoglobina alterata tipica della malattia drepanocitica (HbS) è una mutazione puntiforme del gene della catena beta dell'emoglobina a livello del locus 11p15.5. che comporta la sostituzione dell'adenina con la timina (GAG → GTG) con conseguente sostituzione del Glutammato in posizione 6 in Valina; ciò determina un cambiamento della forma della proteina [49]. La sostituzione di un amminoacido idrofilo con uno idrofobico riduce la solubilità di questa proteina, HbS, rispetto all'HbA, in configurazione

deossigenata con precipitazione dell'HbS e formazione di fibrille nelle emazie e deformazione a sickle o "falce" dell'eritrocita (*Figura 9*).

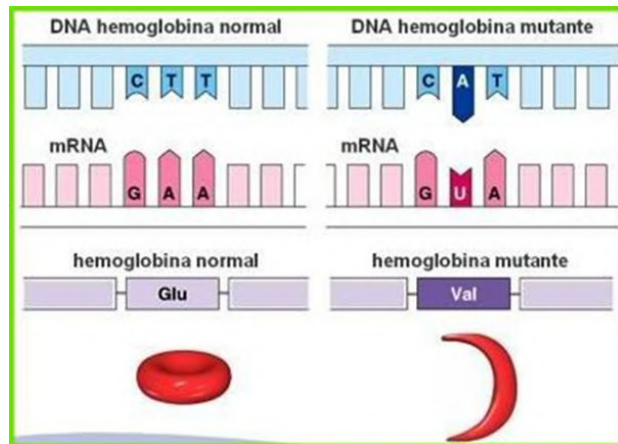


Figura 9. Sostituzione amminoacidica che provoca la falcizzazione.

Il processo di polimerizzazione dell'HbS, dal punto di vista cinetico, prevede una fase lenta e reversibile ed una fase di reazione rapida ed irreversibile. Il ripetersi di episodi di falcizzazione e di ritorno allo stato normale induce un danno di membrana a livello dei globuli rossi e comporta l'irreversibilità del fenomeno di falcizzazione. I globuli rossi a falce presentano membrane cellulari molto rigide ed indeformabili e vengono selezionati nei sinusoidi splenici per essere eliminati dal sistema reticolo-endoteliale; anche l'emolisi intravascolare è possibile, a causa dell'aumentata fragilità meccanica. La falcizzazione degli eritrociti avviene per lo più in presenza di ipossia, ma anche in altre particolari condizioni (acidemia, innalzamento della temperatura, presenza di un incremento dell'acido 2,3 bisfosfoglicerico). Sebbene numerosi studi abbiano tentato di individuare i fattori genetici e ambientali responsabili dell'ampia variabilità fenotipica della malattia, molti aspetti non sono ancora chiari [50].

Questa variazione strutturale modifica la carica superficiale dell'emoglobina provocando, in situazioni di deossigenazione, un'interazione idrofobica tra diversi tetrameri emoglobinici che portano alla formazione di polimeri che si ordinano in strutture parallele di fasci di fibre (gelificazione) [51]. La polimerizzazione dell'emoglobina si associa ad importanti alterazioni della membrana degli eritrociti che diventano meno deformabili e più fragili. Queste alterazioni di membrana, inizialmente reversibili, culminano nella formazione dei

classici eritrociti a forma di falce che danno il nome alla patologia associata alla presenza della HbS, l'anemia falciforme (*Figura 10*) [52].

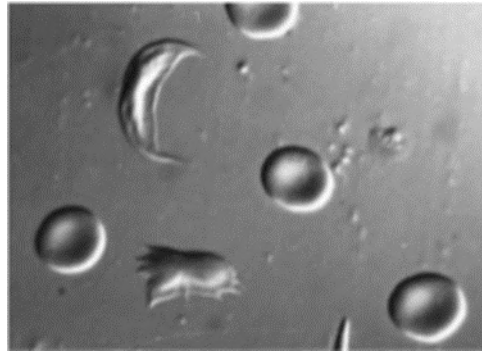


Figura 10. Esempio di falcizzazione su un campione di portatore eterozigote di emoglobina S e di β -talassemia. Fotografia della reazione con matabiosolfito. Si noti un'emazia a falce spontanea, una prodotta con reazione di matabiosolfito ed emazie ancora normali.

Nel 1978, Kan e Dozy, utilizzando un marcatore polimorfico del DNA hanno dimostrato che il gene della β -globina è contenuto nella maggior parte degli individui di origine africana in un frammento di restrizione di 7.6 kb. Allo stesso tempo hanno scoperto che l'allele della globina mutata si trova in un frammento di 13 kb sia in popolazioni africane, ad esempio del Ghana, e sia nel 70% degli afro-americani. Ciò ha permesso di utilizzare il frammento di 13 kb come marcatore dell'anemia falciforme in queste popolazioni. In altre popolazioni dell'Africa (per esempio in Kenya), la mutazione delle cellule falciformi è tipicamente associata al frammento di 7.6 kb.

Queste scoperte implicano che la mutazione si è originata nell'Africa Occidentale sul cromosoma che contiene il frammento di 13 kb e che poi è avvenuta altrove almeno una seconda volta. La protezione contro la malaria che il gene dell'anemia falciforme conferisce agli eterozigoti è responsabile della loro elevata frequenza nell'area della malaria stessa (*Figura 11*) [53].

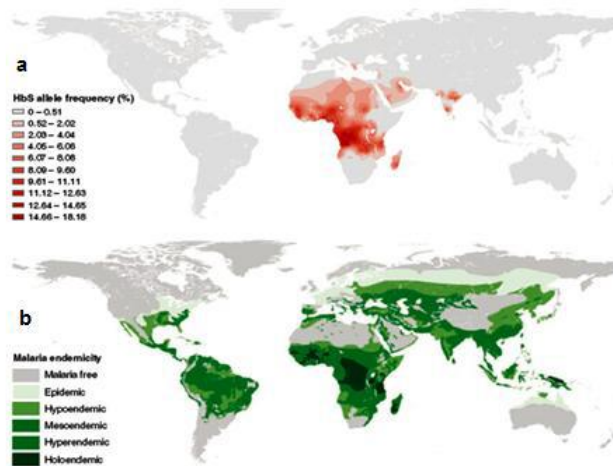


Figura 11. Distribuzione mondiale dell'anemia a cellule falciformi. (a) Mappa della frequenza allelica del HbS. (b) Mappa delle aree endemiche della malaria.

2.2.a.2. I diversi aplotipi e le differenze molecolari

A livello molecolare sono stati individuati dei diversi aplotipi HbS Beta6(A3) Glu>Val, (*HBB*: c.20A>T), con una specifica collocazione geografica, che sembrano correlare con la severità della malattia: aplotipo Sen (Senegal), il più lieve, l'aplotipo Car/Bant, il più severo, e l'aplotipo Ben (Benin), l'intermedio. Non è però pratica clinica definirli in ogni paziente [54, 55]. Inoltre anche i singoli differenti genotipi compresi nell'ambito delle sindromi drepanocitiche possono determinare differenze di fenotipo.

Gli individui eterozigoti per HbS (HbAS), possiedono uno solo dei due geni responsabili della produzione dell'HbS e sono portatori sani della malattia (Sickle Cell Trait-SCT); generalmente non presentano sintomi. Gli indici globulari sono normali, con contenuto di HbS di solito intorno al 40%. Tuttavia, è possibile l'insorgenza di quadri clinici SCD-like in situazioni di stress con sensibile riduzione nella concentrazione dell'ossigeno (esercizio fisico, disidratazione, altitudine) per la co-dominanza a livello molecolare delle due varianti alleliche; in alcuni casi sembrano soggetti a un maggior rischio di tromboembolismo venoso [56]. In caso di doppia eterozigosi HbS/ β -talassemia: (HbS/ β^0 -talassemia o HbS/ β^+ -talassemia), HbS/HbC, HbS/HbE, o per altre varianti emoglobiniche più rare (Hb D-Punjab e

HbO-Arab), sono possibili espressioni fenotipiche differenti, da gradi lievi pressoché asintomatici a quadri severi per la compromissione d'organo in termini di morbilità e mortalità [57].

2.2.b.1. HbC

L'emoglobina C si forma in seguito a una mutazione puntiforme sul nucleotide 6 (GAG AAG) localizzata esattamente sul primo codone del gene della beta globina [58]. È una sostituzione di guanina per adenina, che si traduce nella sostituzione dell'acido glutammico (sesto amminoacido della catena β) con una lisina: $\alpha_2\beta_2$ 6Glu-Lys [59]. La presenza di lisina (diamino monoacido) nella catena polipeptidica provoca la sostituzione di due cariche negative con due cariche positive. Pertanto, l'emoglobina C migra meno rapidamente di Hb S, a pH alcalino.

L'Hb C, originariamente descritta a livello dell'altopiano voltaico, ha conquistato il Golfo di Guinea e il Nord Africa (Marocco, Algeria, Tunisia). È particolarmente diffusa nelle regioni malariche dell'Africa occidentale (Ghana, Costa d'Avorio, Burkina Faso, Togo, Benin, Nigeria) dove il suo ruolo di protezione della malaria e quindi dal Plasmodium falciparum è stato ampiamente dimostrato.

La prevalenza di HbC è inferiore all'1% in Africa centrale e zero in Africa orientale [59,60]. Anche l'Europa meridionale, in particolare l'Italia e la Turchia. Hb C si trova anche nelle Antille francesi (circa il 3%) e tra le popolazioni di origine africana che vivono negli Stati Uniti [61] o nei Caraibi. Occasionalmente si incontra nella penisola arabica, nel vicino Oriente, nei Balcani, in Sicilia.

Gli eritrociti contenenti l'emoglobina C sono parzialmente disidratati (con perdita di acqua e efflusso di K^+), di piccole dimensioni ma con carico normale di emoglobina [60]. Quando la concentrazione di emoglobina aumenta, segue una formazione di cristalli e la rottura della membrana di scambio ionico. Ciò si traduce in una iperviscosità del globulo rosso. La forza meccanica viene così ridotta mentre la resistenza osmotica rimane buona, l'HbC ha infatti una tendenza trombotica più che una tendenza emolitica. La malattia viene trasmessa geneticamente con un meccanismo autosomico recessivo; la percentuale di HbC è quasi 100% nei soggetti omozigoti, mentre negli eterozigoti può arrivare al 35% in rapporto con la penetranza del gene anomalo. Talvolta l'HbC si può trovare combinata con l'anemia

falciforme, HbS (con peggioramento della falcizzazione delle emazie) o con beta talassemia [62]. Gli individui eterozigoti per Hb C sono asintomatici e possono presentare moderata microcitosi relativa al fenomeno della disidratazione [60,61]. Gli eterozigoti compositi C / β -talassemici sono caratterizzati dalla presenza di globuli rossi microcitici e ipocromici e dalla tendenza alla pseudo-poliglobulia. La combinazione di Hb C con β + -talassemia dà un quadro clinico paragonabile a quello di omozigotità C / C [63].

2.3.HbS/HbC

Per quanto riguarda invece l'eterozigosi composta HbS/HbC è la seconda variante emoglobinica nella popolazione africana, si riscontra infatti 1 in 1000 nasciate tra gli Afroamericani. È una condizione relativamente benigna in quanto si manifesta come una forma più lieve di anemia falciforme. I pazienti eterozigoti composti per la mutazione C/S presentano un grave quadro caratterizzato da cellule falciformi. Il quadro clinico è paragonabile a quello della malattia falciforme omozigote S / S, sebbene leggermente attenuato. Ha una cronica anemia emolitica moderata, crisi vaso-occlusive meno frequenti o assenti, e una diminuzione del rischio di sindrome toracica acuta. Complicazioni che possono essere certamente posteriori ma ugualmente gravi o più gravi, rimangono gli stessi e sono ancora più frequenti che nei pazienti con anemia falciforme. Forme gravi S / C sono caratterizzati da una elevata frequenza di crisi dolorose (66% dei pazienti S / C in questo studio) e / o sindrome acuta del petto [64].

2.4.Emoglobine M

Sono prodotte in seguito a mutazioni che interessano uno dei residui di istidina che prendono contatto con l'eme. Quattro di queste mutazioni, due nella catena α e due nella β , consistono nella sostituzione di questo aminoacido con un residuo di tirosina, mentre in un quinto tipo di mutazione si trova un acido glutamico. L'anello fenolico della tirosina crea un legame covalente con l'atomo di ferro dell'eme e lo stabilizza nella sua forma ossidata Fe⁺⁺⁺ dando così origine a metaemoglobina. Il difetto si manifesta negli eterozigoti e si trasmette come carattere autosomico dominante. La sintomatologia è rappresentata dalla cianosi, già presente alla

nascita nelle mutazioni delle catene α , mentre compare nei primi mesi di vita in quelle delle catene β [65].

2.5. Emoglobine chimeriche

L'elevata omologia di sequenza fra i geni del cluster β è responsabile di errori di allineamento alla meiosi. Il parziale scivolamento fra i due cromosomi può provocare un appaiamento fra le sequenze β di un cromosoma con la sequenza δ del cromosoma omologo o addirittura con quella del gene γ più proximale, l' γ^A : se interviene un crossing-over, si produrranno geni chimerici. Le emoglobine chimeriche formate da una sequenza al 5' del gene δ e una sequenza al 3' del gene β sono note come emoglobine Lepore, dal nome della famiglia in cui il fenomeno fu individuato per la prima volta (*Figura 12*).

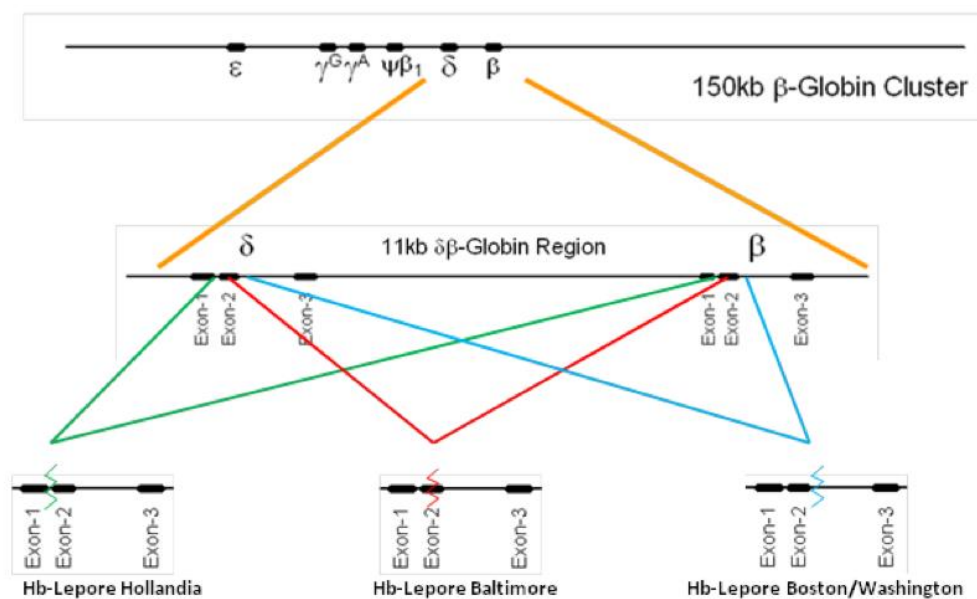


Figura 12. La figura mostra i tre possibili riarrangiamenti dell'alterazione Hb Lepore [109].

Il crossing-over ineguale sopra descritto dà origine anche a chimere speculari $\beta \delta$, dette anti-Lepore, le quali non sono associate a manifestazioni cliniche, dato che il cluster corrispondente contiene una copia funzionale sia del gene β sia del gene δ . Le chimere $\gamma \beta$ sono note come emoglobine Kenia. Non sono finora descritte emoglobine anti-Kenia. Questo tipo di proteina è altamente instabile e, nelle forme Lepore, il risultato in definitiva è rappresentato da un deficit di catene β con un

fenotipo clinico che ricalca quello della talassemia. Questo esempio dimostra anche che non è sempre agevole fare una distinzione netta tra emoglobinopatie qualitative e quantitative [65].

2.6. Talassemie

Le talassemie sono anomalie ereditarie, autosomiche recessive, della sintesi di emoglobina e caratterizzate da una quantitativa riduzione o abolizione delle catene globiniche α o β , che provocano rispettivamente l' α -talassemia o la β -talassemia, con la successiva riduzione della sopravvivenza dei globuli rossi [66, 67]. Esse derivano da alterazioni a carico dei geni globinici, in particolar modo da delezioni più o meno estese o da mutazioni puntiformi di diversa natura. Gli effetti di queste alterazioni si possono manifestare nei processi di trascrizione e traduzione, portando di conseguenza alla totale assenza della sintesi proteica, ad uno squilibrio nella velocità di sintesi, oppure alla produzione di catene altamente instabili [68]. La condizione di talassemia non è associabile ad un unico difetto genetico, ma è piuttosto un gruppo di alterazioni che producono effetti clinici simili. È quindi più opportuno distinguere tra classificazione genetica e classificazione clinica della patologia stessa. La classificazione clinica divide le talassemie in base alla loro gravità, mentre la classificazione genetica le definisce in funzione del tipo di catena globinica coinvolta. Infine le talassemie possono essere classificate anche a livello molecolare, sulla base del tipo di mutazione del gene globinico specifico. Sotto il profilo clinico nelle talassemie clinicamente rilevanti (α e β talassemia) si possono distinguere quattro diverse condizioni di gravità decrescente. La più grave è la talassemia major, trasfusione-dipendente, dovuta allo stato omozigote o al composto genetico per due mutazioni talasemiche gravi. La forma intermedia è caratterizzata da anemia di grado variabile e dalla non necessità di un continuo programma trasfusionale per la sopravvivenza. Tali forme sono molto eterogenee clinicamente mostrando una gravità che varia dal quadro asintomatico del portatore sano alla anemia grave della talassemia major. Anche sotto il profilo genetico le talassemie intermedie sono estremamente eterogenee. Il portatore sano, caratterizzato da manifestazioni esclusivamente ematologiche e da mancanza di sintomi clinici, è il prodotto dello stato di eterozigosi per α o β -talassemia. Infine vi

sono delle forme di eterozigosi sia per le α che per le β -talassemie che sono completamente silenti sia sul piano clinico che su quello ematologico e vengono indicate con il termine di “talassemie silenti”. La classificazione genetica distingue le talassemie in base alla catena globinica interessata in α , β , $\delta\beta$, γ , δ -talassemia, di cui solo le α , β , $\delta\beta$ -talassemie hanno una importante rilevanza clinica [16].

2.6.a. Distribuzione

I primi casi di talassemia sono stati descritti a partire dal 1925 in pazienti di origine mediterranea: infatti il Mar Mediterraneo veniva indicato con il termine greco *ταλασσα* (talassa) da cui ha origine la denominazione di questa classe di patologie [68,69]. Successivamente, è stata riscontrata una diffusione piuttosto elevata anche in Africa, nel Medio Oriente e nei paesi europei che si affacciano sul Mediterraneo [66,70,71].

L'incidenza maggiore è presente in Grecia (5-10%), Cipro (10-15%), Libano, Israele, Bulgaria, Romania, Georgia, Azerbaijan (7%) e fra gli Ebrei Curdi, dove raggiunge anche il 20% [72].

Attualmente, si può affermare che non esiste alcun luogo al mondo in cui le sindromi talasemiche non interessino almeno parte della popolazione; ciò è dovuto ai continui movimenti di individui da una regione all'altra [73].

L'Organizzazione Mondiale della Sanità ha stimato la presenza in Europa, Africa e Asia di circa 180 milioni di individui microcitemici, di cui il 40% è portatore sano di β -talassemia [72].

La talassemia è presente anche negli USA ed in Australia, dove, però, non è identificabile una distribuzione precisa e stabile nel tempo a causa del rilevante flusso migratorio di popolazioni eterogenee e comunque portatrici del tratto talasemico.

2.6.b. Fisiopatologia

2.6.b.1. Le α -talassemie

Nell'uomo esistono 4 geni, localizzati nel cromosoma 16 (16p13.3), che contengono l'informazione genetica per la produzione delle subunità alfa. Gli individui normali hanno 4 copie intatte del gene (2 copie situate sul cromosoma paterno e 2 sul

cromosoma materno) e si indicano con $\alpha\alpha/\alpha\alpha$; includendo un gene embrionale (ζ) e due geni fetali/adulti disposti lungo il cromosoma nell'ordine telomero- ζ - $\alpha 2$ - $\alpha 1$ -centromero, circondato da geni ampiamente espressi [74,75]. I portatori e i malati di talassemia alfa hanno invece un'alterazione in una o più copie del gene.

Le α -talassemie sono infatti solitamente causate da mutazioni del gene α -globinico che riducono (α^+) o aboliscono completamente (α^0) la produzione di catene α -globiniche da parte dell'allele affetto. Sono più frequentemente causate da delezioni che rimuovono parte o tutto il cluster del gene α . Meno comunemente, sono date da cambiamenti nucleotidici all'interno di entrambi i geni duplicati dell' α -globina (chiamate determinanti "non-delezione") e, molto raramente, da delezioni che includono la regione HS-40 ma lascia intatti i geni α [76].

La mancanza di catene α dà origine a catene β -globiniche in eccesso che formano:

- a. $\gamma 4$ tetramero, chiamata Hb Bart's (nella vita fetale);
- b. $\beta 4$ tetramero, chiamata HbH (nella vita adulta).

Finora sono state identificate oltre 100 forme genetiche di α -talassemia, con fenotipi che vanno da asintomatici a letali in quanto la gravità della malattia dipende dal numero dei geni colpiti.

Si distinguono quattro condizioni cliniche di diversa gravità, distinguibili a loro volta in due macrogruppi:

- a. Il primo macrogruppo è costituito dalle α -talassemie classificate sulla base della produzione totale della globina a cui appartengono le α^0 (nessuna formazione di α -globina dal cromosoma): quando entrambi i geni della α -globina su un cromosoma vengono eliminati o inattivati o le α^+ (viene prodotta una α -globina): quando un gene della α -globina su un cromosoma viene eliminato o inattivato da una variante non cancellante;
- b. Il secondo è costituito invece dalle α -talassemie classificate sulla base del numero dei geni globinici persi o anormali: alpha-talassemia 2 ($-\alpha/\alpha$, α -talassemia silent carrier), alpha-talassemia 1 ($-\alpha/-\alpha$ or $--/\alpha$, α -talassemia trait) [77].

I portatori silenti della talassemia alfa, α^+ -talassemia/ α -talassemia 2, hanno un'alterazione in una sola copia del gene alfa globinico, e si indicano con ($-\alpha/\alpha$). Questi individui non mostrano nessun sintomo da adulti, e solo alla nascita presentano una minima quota di emoglobina di Bart (1-2 %).

Il portatore classico, α^0 -talassemia/ α -talassemia 1, possiede due geni alfa globinici alterati e si indica con $-\alpha/-\alpha$ (omozigoti) oppure $--/\alpha\alpha$ (eterozigoti), a seconda che i geni alterati siano presenti sui due omologhi o nello stesso cromosoma. In questi individui alla nascita sono presenti quantità elevate di emoglobina di Bart (5-6 %). La malattia da emoglobina H (HbH) è dovuta all'alterazione di almeno 3 delle 4 copie del gene alfa globinico. Questi individui si indicano con $(--/-\alpha)$. Tale alterazione si riscontra nell'accumulo di catene gamma o beta inutilizzate in eccesso. La mancanza totale del gene alfa globinico $(--/--)$ causa una grave condizione chiamata Idrope Fetale, spesso letale per il feto durante la gravidanza, o subito dopo la nascita [78].

L' α -talassemia si trova prevalentemente nelle aree tropicali e sub-tropicali, dove la frequenza portante può raggiungere l'80-90% della popolazione. L'HbH si osserva prevalentemente nel sud-est asiatico, nel Medio Oriente e nel Mediterraneo. L'Idrope Fetale Hb Bart si osserva prevalentemente nel sud-est asiatico [77].

$-\alpha^+$ -talassemie causate da variazioni di sequenza (“non delezioni”) nei geni strutturali

Sebbene la maggior parte delle varianti di sequenza nel cluster dell' α -globina (compresa la variazione all'interno dell'HS-40) siano causate da SNP neutri, attualmente sappiamo di 69 mutazioni puntiformi o varianti oligonucleotidiche che alterano l'espressione genica, indicate come forme non-delezionali di α -talassemia (denotate $\alpha^T\alpha$ o $\alpha\alpha^T$ a seconda se è interessato il gene $\alpha 2$ o $\alpha 1$). In generale, le α^+ -talassemie di non-delezione possono determinare una riduzione più severa della sintesi della catena α rispetto alle precedenti.

Come per molte altre malattie genetiche umane, queste mutazioni possono influenzare le sequenze canoniche che controllano l'espressione genica, incluse le sequenze CCAAT e TATA box associate al promotore, il codone di inizio (ATG), i segnali di splicing (GT / AG), il codone di terminazione (TAA) e il segnale di poliadenilazione (polyA) (AATAAA).

Oltre a queste mutazioni, l' α -talassemia può anche essere causata da delezioni all'interno del frame, mutazioni frame-shift e mutazioni non-sense (che spesso portano al decadimento dell'RNA) e / o alla produzione di una proteina anormale [75].

- α^+ -talassemia causata da delezioni che rimuovono uno o più geni strutturali duplicati

L'eteroduplex e l'analisi della sequenza del DNA hanno mostrato che i geni dell' α -globina duplicati ($\alpha\alpha$) sono incorporati in due unità di duplicazione altamente omologhe, a 4-kb, la cui identità di sequenza sembra essere stata mantenuta durante l'evoluzione della malattia e della convergenza. Queste regioni sono divise in sottogruppi omologhi (X, Y e Z) da elementi non omologhi (I, II e III). La ricombinazione omologa reciproca tra segmenti Z, che distano 3,7 kb, produce cromosomi con un solo gene ($-\alpha^{3,7}$, delezione a destra) che causa α -talassemia e altri con tre geni alfa ($\alpha\alpha\alpha^{\text{anti}3,7}$). La ricombinazione tra scatole X omologhe, distanti tra loro 4,2 kb, dà origine a un determinante dell' α -talassemia ($-\alpha^{4,2}$, delezione a sinistra) e un cromosoma $\alpha\alpha\alpha^{\text{anti}4,2}$. Ulteriori eventi di ricombinazione tra i cromosomi risultanti (α , $\alpha\alpha$ e $\alpha\alpha\alpha$) possono dare origine a geni quadruplicati ($\alpha\alpha\alpha\alpha$), o quintuplicati ($\alpha\alpha\alpha\alpha\alpha$) o altri riarrangiamenti insoliti. Sebbene queste osservazioni di lunga data abbiano indicato il meccanismo mediante il quale si formano i cromosomi $-\alpha$ e $\alpha\alpha\alpha$, è stato osservato come possono verificarsi in vivo. Recenti studi hanno dimostrato che tale evento è generato dalla ricombinazione reciproca (e dello scambio ineguale) che si verifica nella mitosi (premeiotico) nella linea germinale [75].

- α^0 -talassemia causata da delezioni che rimuovono entrambi i geni strutturali duplicati

Ci sono attualmente circa 50 delezioni del *cluster* α -globinico, che eliminano completamente o parzialmente entrambi i geni dell' α -globina con conseguente perdita della capacità di sintetizzare catene globiniche da essi. Gli omozigoti per tali cromosomi ($--/--$) hanno l'Hb BHFS. Gli eterozigoti composti per queste delezioni che rimuovono un singolo gene alfa hanno una malattia chiamata HbH ($--/\alpha$). Queste delezioni possono essere raggruppate in quelle che si trovano interamente all'interno del *cluster* α -globinico e delezioni che si estendono fino a 800 kb oltre il cluster per includere i geni vicini. Sebbene queste delezioni rimuovano altri geni, gli eterozigoti colpiti appaiono fenotipicamente normali a parte la loro α -talassemia: in alcuni pazienti, il tratto talassemico ($--/\alpha\alpha$) e, in altri, la malattia HbH ($--/\alpha$). Nei

pazienti con delezioni più estese, con monosomia per un ampio segmento di 16p13.3, l' α -talassemia è associata a anomalie dello sviluppo e ritardo mentale (la cosiddetta sindrome ATR-16) [75].

2.6.b.2. Le β -talassemie

Le β -talassemie sono caratterizzate da una riduzione quantitativa o dalla totale assenza della sintesi delle catene β -globiniche, strutturalmente normali: il risultato di questo fenomeno è rispettivamente la diminuzione e la mancanza di molecole di HbA funzionali nel periodo postnatale e nell'età adulta [79]. L'eritropoiesi poco efficiente, tipica caratteristica della β -talassemia, è il risultato di numerosi effetti deleteri, parte di questi dovuti ad un accumulo delle catene α -globiniche [80]. Tale eccesso interferisce con molti stadi della normale maturazione eritroide: alcuni studi, a tal proposito, hanno dimostrato la morte intramidollare dei precursori dei globuli rossi, che si compie attraverso l'arresto nella fase G1 del ciclo cellulare [81] e l'aumentata apoptosi intramidollare degli eritroblasti formati per ultimi [82,83]. Inoltre, è stata anche dimostrata una correlazione tra l'accumulo delle catene α -globiniche e i loro prodotti di degradazione all'interno della membrana dei globuli rossi [84]. Tale anemia è dovuta anche al fatto che le catene α , che sono normalmente prodotte, non trovano un'equivalente concentrazione di catene β alle quali associarsi. Le catene α risultando così in eccesso, si legano tra loro per formare aggregati insolubili i quali precipitano all'interno dei precursori eritroidi intramidollari danneggiandoli; ne risulta così un'eritropoiesi inefficace. Non tutte le catene α in eccesso precipitano, infatti in certi casi possono associarsi alle catene γ , dando origine a molecole HbF, rendendo meno grave la malattia. Nonostante l'emoglobina fetale persista anche dopo la nascita, la sua produzione è insufficiente a compensare la riduzione della sintesi delle catene β globiniche e il relativo eccesso delle catene α -globiniche [85].

Nelle β -talassemie, il gene danneggiato non è sempre completamente inattivato, ma la sua funzione può essere semplicemente ridotta. Nel caso di portatori di un gene β -globinico normale e di un altro con mutazione che ne riduce o elimina la funzione, tali individui sono asintomatici, sebbene i loro globuli rossi siano di piccole dimensioni e presentino una lieve anemia. Quando entrambi i geni sono mutati, si

genera una situazione di anemia potenzialmente pericolosa; tuttavia, la gravità della malattia dipende dalla natura delle mutazioni e quindi dalla loro capacità di produrre catene β -globiniche normali. Le β -talassemie comprendono le tre sindromi cliniche accennate precedentemente:

- a. il trait talassemico (o talassemia *minor*) è clinicamente asintomatica e definita come microcitemia, in cui gli individui manifestano una leggera anemia e alti livelli di HbA₂ ($\alpha_2\delta_2$) associata ad un variabile incremento di HbF ($\alpha_2\gamma_2$). È causata dall'ereditarietà di un singolo allele β -talassemico;
- b. la talassemia intermedia, è caratteristica degli individui sintomatici, che presentano anemia significativa, ma non necessitano di trasfusioni. Questi individui sono caratterizzati da anomalie in entrambi i geni β , ma una o entrambe le mutazioni sono lievi e si ha ancora produzione di quantità sufficienti di β -globina. Pertanto, è una condizione intermedia tra la talassemia minor e quella major;
- c. la talassemia major, o morbo di Cooley, è rappresentata da una condizione in cui i pazienti hanno bisogno di periodiche trasfusioni; in alcuni casi, l'anemia è così grave che la mancanza di trasfusioni può portare al decesso. Tale forma, detta anche β^0 -talassemia e caratterizzata dalla totale assenza di HbA ($\alpha_2\beta_2$), è causata dall'ereditarietà di due alleli del tipo β^0 [82]. Il soggetto affetto da morbo di Cooley presenta, già dai primi mesi di vita, pallore, ittero ed epatosplenomegalia, a cui molto spesso si associano sintomi sistemici, come l'anoressia con decadimento delle condizioni generali e febbre. Inoltre, questi pazienti presentano alterazioni ossee, dovute all'espansione midollare, che si manifestano con la caratteristica "facies microcitemica", determinata da zigomi sporgenti, naso con radice piuttosto infossata e cranio rotondo e ingrossato.

Anche i fattori ambientali e/o acquisiti, come la progressiva splenomegalia, l'esposizione alle infezioni, fattori socio-economici, e la disponibilità delle cure mediche, contribuiscono a modificare l'entità della malattia; inoltre, anche diversi fattori genetici possono migliorare il decorso clinico nelle forme più gravi della β -talassemia, come un aumento dell'espressione dei geni per la γ -globina [86]. Inoltre, le relazioni tra le diverse varianti strutturali dell'emoglobina, descritte e responsabili

di altre patologie emopoietiche, possono produrre una serie complessa di fenotipi clinici [87]: ad esempio, le relazioni tra la β -talassemia e due di queste varianti, l'emoglobina S e l'emoglobina E, sono di particolare importanza.

In soggetti che manifestano β -talassemia sono state studiate ed individuate più di 200 differenti mutazioni geniche responsabili della patologia, sono note diverse delezioni più o meno estese del gene, ma le forme più frequenti sono rappresentate da mutazioni puntiformi, che possono interessare la sequenza codificante, la regione del promotore, il codone d'inizio o di stop e le sequenze di splicing (*Figura 13*) [88,89].

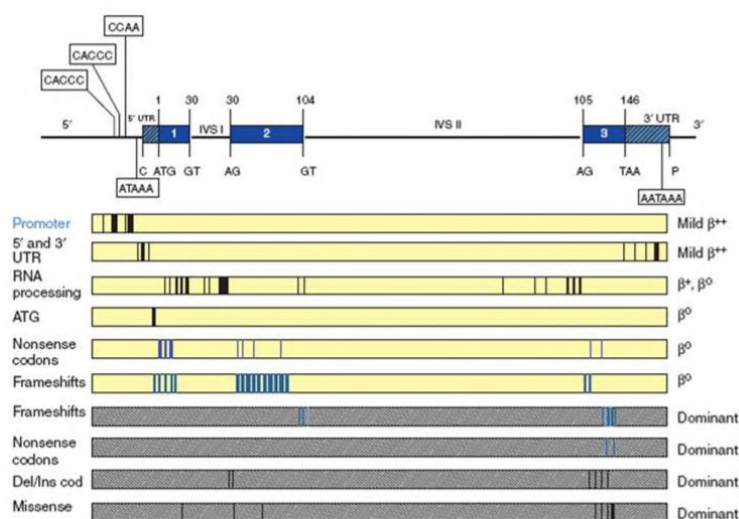


Figura 13. Le mutazioni puntiformi che causano β -talassemia. Pannello Superiore: Rappresentazione schematica del gene β -globinico, (Promotore, 3 esoni 5'/UTR, "invariant dinucleotides" delle giunzioni esone-introne). Pannello Inferiore: Distribuzione delle diverse mutazioni lungo il gene Beta Globinico. Le mutazioni dominanti si trovano all'interno delle caselle ombreggiate.

Le alterazioni del gene per le β -globine spesso presentano una distribuzione etno-geografica ben precisa, e comprendono:

- Mutazioni a livello della regione promotrice del gene: l'effetto di tali mutazioni è quello di ridurre l'efficienza del processo di trascrizione; in questo caso, le catene β -globiniche vengono prodotte anche se in quantità inferiore (β^+ -talassemia). Molte di queste alterazioni possono essere definite "lievi" [90];
- Mutazioni non-senso e frameshift: queste mutazioni portano ad una traduzione incompleta e quindi ad una proteina non funzionale. Sono da considerarsi mutazioni gravi (β^0 -talassemia) [91]. Interessano soprattutto gli esoni e causano

- una prematura terminazione della sintesi proteica; inoltre, possono interessare anche il codone d'inizio, quello di stop e la regione 3' UTR, generando nelle ultime due condizioni una β^+ -talassemia [74];
- c. Mutazioni che alterano il processo di maturazione dell'RNA: tale processo avviene attraverso:
- alterazione dei siti donatore e accettore, che impediscono il processo di escissione degli introni, processo di splicing, e provocano forme gravi di talassemia di tipo β^0 [44] e la degradazione dell'mRNA, dovuto all'alterazione dei dinucleotidi invarianti a livello delle giunzioni esone/introne, che impediscono il normale processo di separazione [92, 93];
 - alterazione delle regioni di consenso intorno ai siti accettore e donatore, che riducono l'efficienza con cui avviene l'escissione degli introni e danno luogo a forme di tipo β^+ di variabile gravità;
 - sostituzioni nucleotidiche che creano siti di splicing alternativo, sia all'interno degli introni, che degli esoni. Generalmente provocano un fenotipo di tipo β^+ ed hanno variabile gravità, che dipende dalla regione in cui il nuovo sito si attiva [90].
- d. Mutazioni nella regione di consenso, che funge da segnale di poliadenilazione;
- e. Mutazioni nel sito di inizio della trascrizione (CAP) e nella regione 3' non tradotta. Generano talassemie di tipo β^+ , alcune delle quali sono considerate tra le più lievi descritte [90];
- f. Mutazioni nella tripletta di inizio della traduzione. Queste forme sono le β^0 più gravi [90];
- g. Delezioni geniche comprendono anche l'Hb Lepore, che si originano da un crossing-over non omologo fra i geni β e δ adiacenti. Tale evento porta alla formazione di una subunità formata all'estremo N-terminale da un tratto di catena δ e all'estremo C-terminale da un tratto di catena β .

Di seguito vengono riportate delle rappresentazioni schematiche della distribuzione delle β -talassemie nel mondo (Figura 14), nella regione del Mediterraneo e dell'Asia (Figura 15) e in Italia (Figura 16).

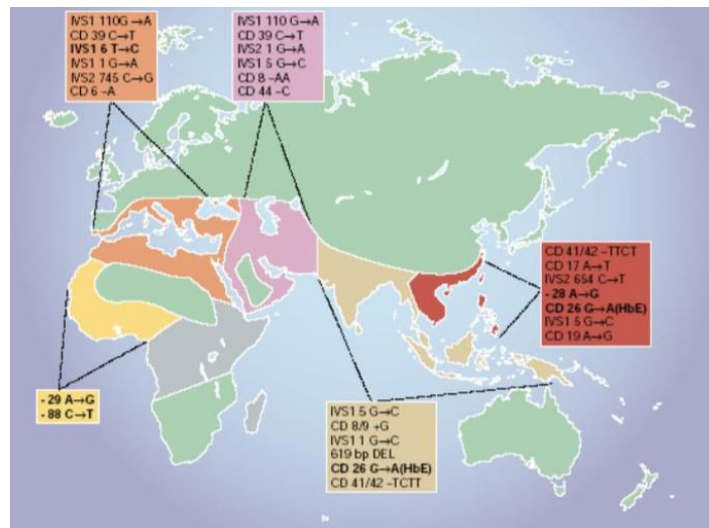


Figura 14. Distribuzione della β -talassemia nel mondo. In figura, nelle zone indicate dai diversi colori, sono riportate le mutazioni più frequenti.

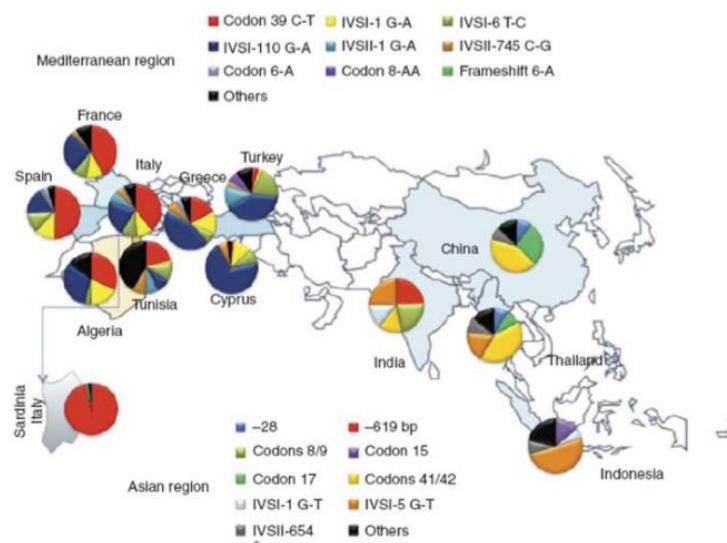


Figura 15. Distribuzione dettagliata delle mutazioni associate a β -talassemia nella regione del Mediterraneo e in Asia [71].

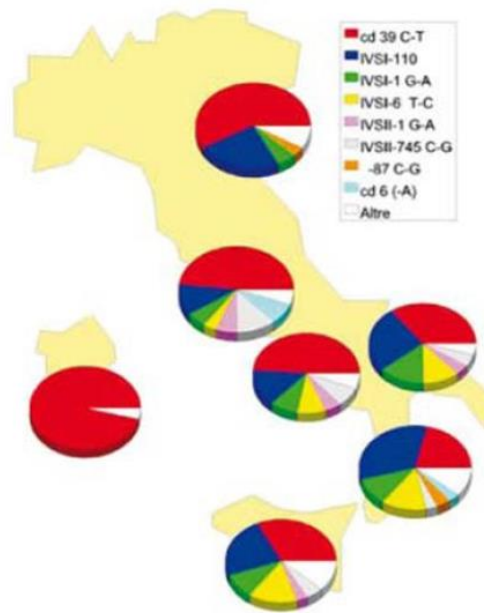


Figura 16. Distribuzione delle β -talassemie in Italia. Ogni colore rappresenta una mutazione diversa. Le alterazioni con incidenza maggiore al 2% sono indicate individualmente, mentre in "altre" sono comprese le incidenze inferiori a tale percentuale.

3. I principali fattori genetici che causano una marcata variabilità fenotipica

Diversi studi hanno suggerito che una varietà di determinanti genetici influenza i diversi fenotipi clinici tra pazienti che ereditano mutazioni identiche della β -globina. È stato osservato che livelli aumentati di emoglobina fetale (HbF, $\alpha_2\gamma_2$) e α -talassemia concomitante conferiscono importanti benefici clinici nei pazienti con β -talassemia e anemia falciforme, in termini di anemia, crescita, sviluppo, entità dell'epatosplenomegalia e necessità di trasfusioni [94]. I polimorfismi a singolo nucleotide (SNPs) in diversi modulatori genetici ed elementi di regolazione cis coinvolti nella regolazione dell'HbF può ritardare il passaggio dell'emoglobina fetale ad adulto (Hb), migliorando la gravità della β -talassemia e dell'anemia falciforme. Finora i meccanismi sottostanti che guidano la regolazione epigenetica della commutazione dell'Hb coinvolti nel miglioramento della gravità della β -talassemia sono in gran parte poco chiari [95]. Questi modificatori includono *HBG2*: -158C> T (rs7482144) nel cluster β -globina, [96, 97], e i geni principali coinvolti nella regolazione della commutazione dell'emoglobina fetale-adulta (Hb), compreso CLL a cellule B / linfoma 11A (*BCL11A*) [98, 99], fattore 1 simile a Kruppel (*KLF1*) [99] e

MYB [100-103]. I principali loci HbF includono: rs7482144 del gene della γ -globina (*HBG2*), rs4671393, rs1427407, rs10189857 di *BCL11A* (CLL a cellule B / linfoma 11A), e rs28384513, rs4895441 e rs9399137 della regione polimorfica *HBS1L-HMIPB* [95, 102-106].

Evidenze da pazienti afroamericani con anemia falciforme e portatori portoghesi di β -talassemia hanno suggerito che i pazienti con alti livelli di HbF presentano una frequenza allelica minore (A) significativamente più alta di rs7482144 rispetto ai pazienti con bassi livelli di HbF [107-108]. Il polimorfismo (A / G) nell'introne 2 di *BCL11A* (rs4671393, rappresentava il 7-12% della varianza HbF) era associato con HbF in nord europei sani [109], afroamericani con β -talassemia [110] e thailandesi con Talassemia HbE- β [111]. Gli SNP *HBS1L-MYB* rs28384513 (C / A) e rs4895441 (G / A) sono stati studiati principalmente in una coorte sana del Nord Europa e in pazienti con SCD con origini africane o europee [100].

Altri loci coinvolti nella regolazione dell'espressione di HbF sono *KCNK10*, *GPR65*, *RNase2*, *RNase3* e *C/EBPE*. Sembrano influenzare la gravità clinica in circa il 30% dei pazienti. *C/EBPE*, situato in 14q11.2, codifica per un fattore di trascrizione di tipo B / zip, coinvolto in diverse funzioni come differenziazione terminale e maturità funzionale delle cellule progenitrici dei granulociti e nell'apoptosi cellulare durante la differenziazione mieloide. I risultati del sequenziamento *C/EBPE* hanno rilevato timina invece di citosina a IVS-I-736, corrispondente al polimorfismo rs45496295 nello stato eterozigote. L'SNP rs45496295 (C> T) era associato a un fenotipo di β -talassemia moderato; alto tasso di emoglobina e regime trasfusionale ridotto, probabilmente dovuto alla conseguenza della riduzione dell'emolisi. Poiché la variazione dell'allele T non agisce direttamente sul livello di HbF, sono necessarie ulteriori indagini per chiarire il meccanismo coinvolto [112].

3.1.Co-ereditarietà dei geni alfa

Gli individui, normalmente, possiedono quattro geni alfa globinici situati sul cromosoma 16. Circa il 30% dei pazienti con SCD, di origine africana, però, presenta una o più delezioni considerate alfa-talasseemiche [29]. La variante più comune in questa popolazione è la delezione -3.7 del gene alpha globina (-3.7/ $\alpha\alpha$) [113]. La co-ereditarietà di una mutazione in omozigosi sui geni beta globinici e di una delezione o mutazione sui geni alfa globinici riduce lo squilibrio fra catene alfa

e non alfa; così non si hanno tutti gli effetti negativi dovuti alla produzione eccessiva di catene alfa globiniche libere e alla loro precipitazione all'interno dei globuli rossi, con gli effetti tossici correlati, e viene compensato lo sbilanciamento della sintesi globinica indotta dalla talassemia [110]. I globuli rossi dei pazienti con SCD, con delezioni del gene alfa subiscono meno emolisi rispetto a quelli senza delezioni alfa del gene e il grado di emolisi è proporzionale al numero di geni alfa-globina cancellati [113]. In pazienti affetti da anemia falciforme, la coesistenza dell' α -talassemia, riduce la concentrazione di emoglobina intracellulare, riducendo così la polimerizzazione dell'HbS e la falcizzazione delle emazie e diminuendo l'emolisi (*Tabella 1*) [110].

Studi dimostrano che, pazienti omozigoti per la β^0 -talassemia che co-ereditano la delezione di due geni alfa globinici ($-\alpha/-\alpha$) sviluppano più frequentemente il fenotipo clinico della talassemia intermedia piuttosto che quello della talassemia major. I pazienti invece che presentano il genotipo β^0 in omozigosi con una singola delezione α -globinica sviluppano di solito talassemia major anche se i segni clinici compaiono più tardivamente. Nel caso di omozigoti per la β^+ -talassemia anche una delezione su un solo gene alfa globinico sembra essere capace di produrre un fenotipo lieve [110].

Nei soggetti portatori delle più comuni varianti delle catene beta (Hb S, HbD-Punjab, HbE e HbC) la percentuale della variante dell'emoglobina dipende direttamente dal numero dei geni alfa globinici. In presenza di alfa talassemia è generalmente ridotta: a meno del 35% circa HbS, D-Punjab e Hb C e a meno del 25% circa HbE anche se, i valori misurati, possono variare a seconda delle tecnologie utilizzate. In assenza di un deficit di ferro, la percentuale della variante dell'Hb potrebbe quindi indicare la presenza di una co-ereditarietà di un'alfa talassemia e aiutare ad escludere una beta talassemia come la causa della microcitosi. Comunque, queste informazioni potrebbero non essere sufficienti per escludere con certezza la presenza di alfa talassemia a causa della possibilità di una co-ereditarietà addizionale di duplicazioni dei geni alfa globinici (ad esempio triplicazioni o quadruplicazioni).

È importante notare che per alcune varianti beta globiniche stabili, che sono cariche più negativamente dell'HbA, come l'HbJ, la situazione è inversa; quando il numero

dei geni alfa globinici diminuisce, la percentuale della variante incrementa in presenza di alfa talassemia [114].

	$\alpha\alpha/\alpha\alpha$	$-\alpha/\alpha\alpha$	$-\alpha/-\alpha$ or $--/\alpha\alpha$
Hb S (%)	35-40	29-34	24-28

Tabella 1. Percentuale del'HbS in presenza dell' α -talassemia.

3.2. Concentrazione di emoglobina fetale

L'importanza dei livelli di HbF è stata teorizzata da un pediatra, Janet Watson, nel 1948, quando notò che le complicanze cliniche della drepanocitosi erano rare prima dell'età di 1 anno, quando i livelli di HbF erano ancora elevati. L'effetto dell'HbF sulla fisiopatologia della polimerizzazione dell'HbS nella SCD è ben compreso. Quando presente nella cellula in concentrazioni sufficienti, l'HbF inibisce la falcizzazione, sia non partecipando alla formazione di polimeri, sia diluendo la concentrazione di HbS nella cellula. Studi in vitro suggeriscono che è necessario il 20% di HbF per prevenire la polimerizzazione di HbS; tuttavia, la quantità di HbF necessaria per migliorare i sintomi e le complicanze della SCD in vivo non è nota e può variare a seconda dell'individuo e dalla distribuzione di HbF nella cellula [113]. Sono stati esplorati diversi agenti terapeutici mirati ad aumentare i livelli di emoglobina fetale; il più usato è l'agente citotossico idrossiurea. L'idrossiurea inibisce il ribonucleotide riduttasi, l'enzima che catalizza la conversione di ribonucleotidi a desossiribonucleotidi. Questo riduce la produzione di globuli rossi contenenti alti livelli di HbS che derivano da cellule progenitrici a divisione rapida, favorendo le cellule con alti livelli di emoglobina F (Cellule F) che derivano da cellule progenitrici a divisione lenta. Clinicamente, l'idrossiurea riduce l'emolisi e il numero di cellule irreversibilmente a falce che verranno poi rimosse dalla circolazione dalla milza [12]. Di recente studi di associazione sull'intero genoma, condotti in differenti gruppi etnici, hanno segnalato che vi sono almeno tre principali loci che svolgono un importante ruolo nell'aumentare i livelli di HbF. Questi includono il -158 C>T nel promotore del gene *HBG2* (locus XmnI), il gene *BCL11A* sul cromosoma 2p16.1 e la regione intergenica *HBS1L-MYB* sul cromosoma 6q23.3. Insieme questi tre loci sono responsabili dal 20 al 50% [70] della variazione di HbF nei pazienti con beta talassemia o anemia falciforme e sono considerati i migliori predittori della gravità della malattia in β -emoglobinopatie [115-116, *Tabella 2*]. In entrambi i disordini le

manifestazioni cliniche sono estremamente variabili in gravità - si va da una forma quasi asintomatica sino alla totale dipendenza dalle trasfusioni - e dipende non solo dal tipo di difetto genico e dalla copresenza dell' α -talassemia ma anche dalla presenza di geni modificatori che portano alla produzione continua di catene γ -globiniche che nell'adulto svolgono un ruolo importante nel compensare in senso positivo i sintomi clinici. Tutte queste condizioni riducono, nella β -talassemia lo squilibrio del rapporto tra catene Beta e catene Alfa e l'eritropoiesi inefficacie, nella drepanocitosi la polimerizzazione dell'HbS e la falcizzazione delle emazie [94]. Sono stati descritti diversi polimorfismi a singolo nucleotide (SNPs) sui tre principali loci, *HBG2:g.-158C>T*, *BCL11A* e *HBS1L-MYB*, in grado di variare i livelli di HbF (Figura 17).

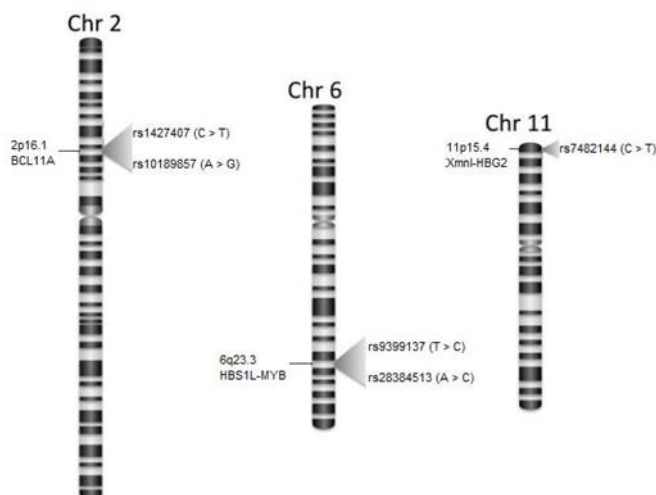


Figura 17. SNPs associati a variazioni di emoglobina fetale nei loci BCL11A (chr.2p16.1), HBS1L-MYB (chr.6q23.3) e XmnI-HBG2 (chr.11p15.4).

Il polimorfismo γ^G -158 (C>T), denominato anche rs7482144, svolge un ruolo importante nella gravità della β -talassemia e dell'anemia falciforme. È il sito di restrizione XmnI localizzato a 158 paia di basi a monte del gene HBG2 ed è associato ad un forte incremento della produzione di HbF. Questo polimorfismo è stato ben descritto tra gruppi di varie popolazioni ed è considerato un importante fattore genetico nel migliorare il fenotipo delle due malattie riducendo la presenza di catene α -globiniche libere [117-118].

Dalla letteratura emerge che gli SNPs rs1427407 (C>T) e rs10189857 (A>G) risultano essere significativamente associati ai livelli di HbF, entrambi situati all'interno dell'introne 2 sul gene *BCL11A*. Sono stati riscontrati più frequentemente nei pazienti con beta-talassemia omozigote con fenotipo lieve (talassemia intermedia) rispetto a quelli affetti dalla forma trasfusione-dipendente, ad indicare che le due varianti del *BCL11A*, incrementando la produzione di catene γ -globiniche, contribuiscono al miglioramento del fenotipo grazie alla parziale compensazione della sbilanciata produzione di emoglobina. Infatti, mentre i pazienti con talassemia major sono trasfusione-dipendenti, quelli con talassemia intermedia non ricevono, se non sporadicamente, trasfusioni e sono caratterizzati da alti livelli di emoglobina, quasi interamente composta da HbF [119-120,110, e 121]. Anche in pazienti affetti da anemia falciforme i due polimorfismi sono risultati fortemente associati all'aumento della quantità di HbF, modulando le caratteristiche cliniche ed ematologiche della malattia [122].

Studi su diverse popolazioni hanno dimostrato che anche polimorfismi sulla regione intergenica di *HBS1L-MYB* contribuiscono alla regolazione dell'emoglobina fetale, seppur in maniera minore rispetto alle varianti sul locus *BCL11A*. Il polimorfismo più significativo correlato ad alti livelli di HbF è l'allele C dello SNP rs9399137, presente maggiormente nei pazienti con talassemia intermedia rispetto a quelli con talassemia major e associato con la riduzione della frequenza degli episodi dolorosi attribuibili a crisi falcemiche [112, 103 e 122].

Diversamente dalla variante rs9399137 (T>C) e dagli altri polimorfismi sopra descritti sui loci *HBG2:g.-158C>T* e *BCL11A*, la presenza dello SNP rs28384513 sul gene *HBS1L-MYB* non aumenta, bensì abbassa i valori di HbF (*Tabella 2*). Questo polimorfismo non è associato ad un miglioramento del fenotipo della β -talassemia e dell'anemia falciforme, in quanto ridotti livelli di HbF non sono in grado di compensare lo squilibrio del rapporto delle catene α/β globiniche [123].

Gene	SNP	Effetti sul livello HbF
HBG2	rs7482144 [C>T] =XmnI -/+	+
BCL11A	rs1427407 [C>T]	+
	rs10189857 [A>G]	+
HBS1L-MYB	rs28384513 [A>C]	-
	rs9399137 [T>C]	+

Tabella 2. Polimorfismi associati alla gravità della β -talassemia e anemia falciforme.

4. Emoglobinopatie e ferro

Una delle principali complicanze della β -talassemia è il sovraccarico di ferro dovuto all'aumento della domanda di ferro associata agli sforzi per soddisfare un'attività eritropoietica aumentata. L'accumulo di ferro in diversi organi può portare a diverse complicazioni tra cui insufficienza cardiaca ed epatica, diabete mellito e ipopituitarismo. Per questo motivo, un focus principale della gestione clinica è stato sulla terapia di chelazione come mezzo per rallentare la progressione del sovraccarico di ferro. Nella β -talassemia major, dove non viene prodotta la β -globina, il sovraccarico di ferro è in gran parte dovuto alle trasfusioni di sangue necessarie per sostenere la vita. Tuttavia, molti studi hanno anche indicato che il sovraccarico di ferro nella β -talassemia è associato ad un aumento dell'assunzione di ferro dalla dieta. I pazienti che soffrono di β -talassemia intermedia, una forma più lieve della malattia, di solito non richiedono una terapia trasfusionale ma alla fine sviluppano un sovraccarico di ferro a causa di un maggiore assorbimento di ferro dal tratto gastrointestinale [124].

L'emocromatosi ereditaria (HH) è una condizione autosomica recessiva che coinvolge il metabolismo del ferro, portando in particolare a sovraccarico di ferro tossico multiorganico e insufficienza epatica. Il gene responsabile (*HFE*) è lungo 12kb, contiene sette esoni e codifica per una glicoproteina transmembrana di tipo I di 348 residui, che è espressa in molti tessuti, tranne il cervello [125]. Il gene *HFE* ha un ruolo importante nell'omeostasi del ferro regolando l'assorbimento del ferro. Le mutazioni del gene *HFE* sono attualmente conosciute come causa di aumento dell'assorbimento del ferro, sovraccarico di ferro e HH.

Tra le mutazioni più frequentemente riscontrate nell'HH vi sono tre mutazioni missenso nel gene *HFE* con un modello autosomico recessivo. La più comune è

una mutazione all'interno dell'esone 4 del gene *HFE*, con conseguente cambiamento di aminoacidi nella posizione 282 dalla cisteina alla tirosina (C282Y), questa mutazione è risultata responsabile del 60% dei casi di HH nelle popolazioni mediterranee. Un'altra mutazione è l'H63D che provoca una transizione CG al nucleotide 187 dell'esone 2 del gene *HFE* che porta a una sostituzione da istidina a acido aspartico [126]. La mutazione C282Y interrompe l'espressione della superficie cellulare della proteina HFE e la mutazione H63D influenza l'interazione dell'HFE con il recettore della transferrina (TfR) [97]. La sostituzione di una cisteina con tirosina nella mutazione C282Y interrompe un legame disolfuro e altera drasticamente la funzionalità della proteina. La mutazione H63D non disturba i legami disolfuro della proteina ma altera fortemente la carica della proteina a causa della sostituzione di un'istidina carica con un acido aspartico caricato. Apparentemente tali lievi menomazioni non sono sufficienti per causare una grave patologia nello stato omozigote, ma la combinazione del difetto H63D con la mutazione C282Y può causare patologia grave in alcuni individui [125].

Molti studi hanno dimostrato che i pazienti con β -talassemia che hanno mutazioni in *HFE* hanno maggiori probabilità di sviluppare un sovraccarico di ferro che può richiedere una chelazione precoce del ferro anche nello stato eterozigoti [126].

Sono presenti molti studi con risultati controversi riguardo il ruolo delle mutazioni a carico del gene *HFE* in presenza di β -talassemia. I dati pubblicati indicano sintomi di emocromatosi più gravi quando eterozigosi per la mutazione C282Y è associata a β -talassemia e quando livelli sierici più elevati nella talassemia co-ereditaria sono associati a eterozigosi e omozigosi per H63D. Questi risultati suggeriscono che il carico di ferro nel tratto β -talassemico tende ad aggravarsi con la coinheritance della mutazione H63D, anche se presente nello stato eterozigote [127].

In uno studio viene messo in evidenza che la mutazione H63D potrebbe avere un effetto modificante sull'assorbimento del ferro e che l'eredità della mutazione H63D potrebbe aggravare il tratto β -talassemico anche se presente in stato eterozigote; altri autori hanno concluso che i pazienti con talassemia intermedia con mutazione *HFE* coesistente hanno maggiori probabilità di sviluppare sovraccarico di ferro e potrebbero aver bisogno di una terapia di chelazione del ferro precoce.

Una scoperta diversa è stata osservata in studi che suggeriscono, sia che l'ereditarietà della mutazione H63D non influenza la gravità del sovraccarico di ferro

nei pazienti con β -talassemia, sia che la presenza della mutazione H63D non aumenta il ferro corporeo misurato dalla ferritina sierica nella β -talassemia.

La discrepanza tra i risultati può avere diverse spiegazioni in quanto l'espressione dell'emocromatosi da *HFE* è modificata da fattori genetici e acquisiti, ad esempio abitudini alimentari, perdita di sangue o donazione, gravidanza, menopausa, malassorbimento.

È importante lo screening delle mutazioni del gene *HFE* nei pazienti talassemici per la rilevazione di coloro che hanno una maggiore probabilità di sviluppare un sovraccarico di ferro anche quando la mutazione è presente nello stato eterozigote, poiché ciò può migliorare la qualità della vita in questi pazienti con un follow-up frequente per la rilevazione preventiva del sovraccarico di ferro e delle sue complicanze, applicando protocolli di chelazione del ferro più aggressivi [126].

Studi recenti indicano il ruolo chiave dell'epcidina nella modulazione dell'assorbimento del ferro nella β -talassemia. L'epcidina è un ormone peptidico sintetizzato nel fegato, che ha dimostrato di essere il principale regolatore del metabolismo del ferro.

È codificata dal gene *HAMP* e va a regolare negativamente l'assorbimento del ferro, degradando la ferroportina l'unico esportatore di ferro noto espresso da enterociti duodenali, epatociti e macrofagi. Bloccando l'esportazione di ferro, le cellule consentono a questo peptide di controllare il flusso di ferro nel flusso sanguigno, incluso l'assorbimento dal duodeno, il riciclaggio all'interno dei macrofagi e il rilascio dal compartimento di conservazione nel fegato. La regolazione dell'espressione dell'epcidina è un processo complesso che richiede l'integrazione di numerosi segnali tra cui ferro, ipossia, domanda eritropoietica e stato infiammatorio. L'accumulo e l'infiammazione di ferro spingono l'espressione dell'epcidina più in alto per limitare il rilascio di ferro nel flusso sanguigno. D'altra parte, l'anemia e l'ipossia riducono fortemente l'epcidina, aumentando la disponibilità di ferro. Nei pazienti con β -talassemia coesistono anemia e sovraccarico di ferro ed esercitano due segnali in competizione nella modulazione dell'espressione dell'epcidina. La prima evidenza di disregolazione dell'epcidina nella β -talassemia è stata riportata nel 2004, con bassa espressione dell'epcidina nei modelli murini di questa malattia. Utilizzando modelli murini di β -talassemia intermedia e major, ha è stato dimostrato

che il livello di IE determina la distribuzione di ferro degli organi e il livello di espressione dell'epcidina [124].

In diversi studi sono stati valutati i ruoli di *HAMP* e *HFE* in pazienti affetti da drepanocitosi. In uno è stata valutata la frequenza della mutazione G71D del gene *HAMP* e la mutazione H63D del gene *HFE* e la correlazione tra queste mutazioni, nonché la correlazione tra esse e il sovraccarico di ferro nei pazienti con anemia falciforme. Lo studio ha dimostrato che né la mutazione *HAMP*-G71D né la mutazione *HFE*-H63D da sole sono i principali determinanti dello stato di ferro corporeo totale nei pazienti con drepanocitosi, ma tuttavia, il numero di mutazioni ospitate può aumentare la probabilità di sovraccarico di ferro [99]. Altri autori invece hanno studiato le mutazioni del gene *HFE* H63D e C282Y in pazienti affetti da anemia falciforme e non è stato riscontrato che queste mutazioni vanno ad influenzare il grado di sovraccarico di ferro [128].

5. Flow chart per l'identificazione delle lesioni molecolari a carico dei geni Globinici

L'iter diagnostico per l'identificazione delle mutazioni a carico dei geni globinici prevede l'integrazione tra emocromo, elettroforesi delle proteine e analisi molecolari.

- a. esame emocromo-citometrico, che tiene conto del volume medio dei globuli rossi (MCV) e del contenuto medio di emoglobina nei globuli rossi (MCH);
- b. separazione delle frazioni emoglobiniche, con elettroforesi o HPLC, e quindi con la successiva definizione della percentuale di HbA2 e HbF (*Figura 18*) [71].

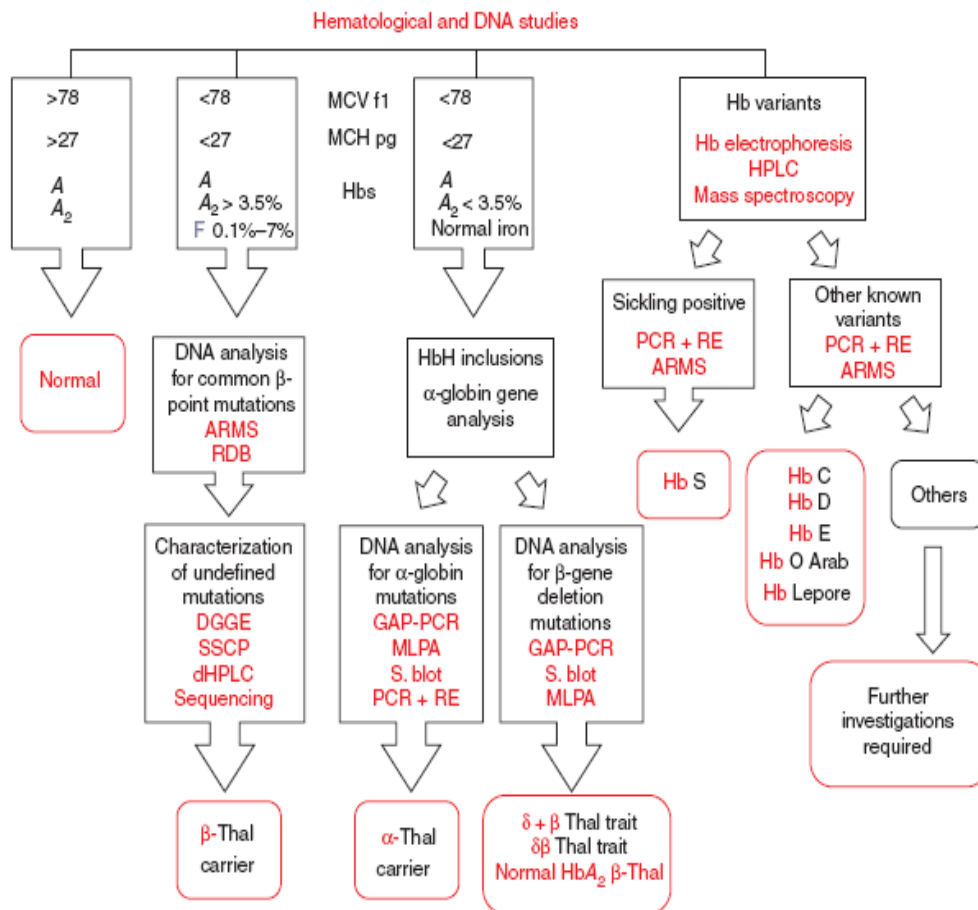


Figura 18. Flow chart per l'identificazione delle lesioni molecolari a carico dei geni Globinici. ARMS (Amplification refractory mutation system), RDB (Reverse dot blot), DGGE (Denaturing gradient gel electrophoresis), SSCP (Single-strand conformation polymorphism), dHPLC (Denaturing high-performance liquid chromatography), MLPA (Multiple ligation-dependent probe amplification), RE (Restriction enzyme).

In particolare differenti valori di MCV, MCH, HbA₂ e la sintesi delle catene globiniche, che rappresenta uno dei test funzionali per identificare un eventuale sbilanciamento delle catene globiniche, possono indirizzare l'analisi verso determinate mutazioni (Figura 19) [129].

GENOTYPE	MCV (fl)	MCH (pg)	HbA ₂ (%)	α/β ratio
β -101 (C→T)	88.5 ± 7.8	30.1 ± 1.0	3.1 ± 1.0	1.3 ± 0.04
β -92 (C→T)	83.0 ± 6.0	28.3 ± 2.0	3.5 ± 0.4	1.3 ± 0.08
β +33 (C→G)	82.0 ± 9.2	27.1 ± 3.4	2.5 ± 1.4	1.3 ± 0.06
β CAP+1 (A→C)	77.5 ± 2.5	24.5 ± 1.5	3.5 ± 0.3	-
β IVSI-6 (T→C)	71.0 ± 4.0	23.1 ± 2.2	3.4 ± 0.2	1.9 ± 0.05
β IVSII-844 (C→G)	96.0 ± 4.0	30.3 ± 1.8	3.2 ± 0.2	1.0 ± 0.06
β +1480 (C→G)	88.3 ± 9.5	27.9 ± 6.0	2.7 ± 0.8	1.6 ± 0.04
ααα/αα	85.5 ± 7.8	30.4 ± 5.0	2.8 ± 0.6	1.2 ± 0.04
δ- & β-thalassaemia	67.6 ± 7.6	21.8 ± 3.6	3.3 ± 0.4	1.7 ± 0.06
KLF1 mutations*	84.0 ± 6.6	28.2 ± 2.4	3.6 ± 0.3	0.9 ± 0.1

Figura 19. Valori di: MCV, MCH, HbA₂ e α/β ratio in alcune lesioni molecolari a carico dei geni globinici [76].

6. Tecniche biomolecolari per lo studio dei geni globinici

La diagnosi molecolare delle talassemie è effettuata utilizzando differenti tecnologie. Per l'identificazione delle mutazioni note dei geni globinici è importante l'utilizzo di tecniche, quali: RDB (reverse dot blot) e la tecnica ARMS (Amplification refractory mutation system). Entrambi i metodi, in particolare il RDB, possono identificare in un unico esperimento da 20-30 differenti mutazioni. Per identificare mutazioni sconosciute si possono utilizzare altre metodologie diagnostiche come ad esempio: dHPLC (denaturing high-performance liquid chromatography), DGGE (denaturing gradient gel electrophoresis), SSCP (single-strand conformation polymorphism). Il sequenziamento diretto mediante il metodo Sanger è inoltre utilizzato sia come conferma delle mutazioni identificate con i metodi sopra citati che come metodo diretto di analisi. GAP-PCR e MLPA (multiple ligation-dependent probe amplification) sono utilizzate per identificare delezioni dei geni globinici [71]. Sono attualmente in uso anche tecnologie che permettono l'analisi contemporanea di diversi geni globinici, e non, come NGS.

Negli ultimi dieci anni, il sequenziamento di nuova generazione (NGS), è diventato il metodo di riferimento in genetica, poiché è un metodo che produce dati di sequenziamento ad alta risoluzione con potenziale identificazione di variazioni rare (e nuove) che non possono essere identificate da altre piattaforme di genotipizzazione. Una delle principali proprietà di un sistema di NGS è la produttività, infatti è un sistema che consente di generare grandi quantità di dati

genomici al fine di rivelare la costituzione genetica delle persone e valutare i potenziali rischi per la salute [130].

La metodica prevede la preparazione di campioni di DNA genomico, la costruzione di librerie, il pooling, l'amplificazione tramite PCR e il sequenziamento.

La costruzione di librerie è uno dei passaggi fondamentali del sistema e può avvenire in diversi modi:

- a. Libreria basata su ampliconi: è un protocollo di preparazione di librerie che si basa sulla PCR, utilizzando oligonucleotidi per amplificare le regioni di interesse al fine di investigare le varianti a singolo nucleotide (Single Nucleotide Variant, SNV), le inserzioni/delezioni (Indel), le variazioni del numero di copie (Copy Number Variation, CNV) e le fusioni geniche. Quindi un flusso di lavoro in multiplex, altamente specifico e basato sulla PCR permette di ottenere librerie dotate di allineamento specifico per i target ed elevata uniformità di copertura. In un singolo saggio possono essere messi in multiplex fino a 24.000 ampliconi (AmpliSeq for Illumina), in questo modo la valutazione simultanea di più geni consente di risparmiare tempo.

La preparazione delle librerie inizia con l'amplificazione di regioni specifiche nell'input di DNA o cDNA iniziale. Le rimanenti sequenze dei primer sono sottoposte a digestione e gli adattatori per il sequenziamento aggiunti agli ampliconi ottenuti. La libreria viene amplificata ed è pronta per la quantificazione, normalizzazione e raggruppamento in pool prima del sequenziamento. Grazie a questo metodo, possono essere preparate più librerie in appena cinque ore (*Figura 20*);

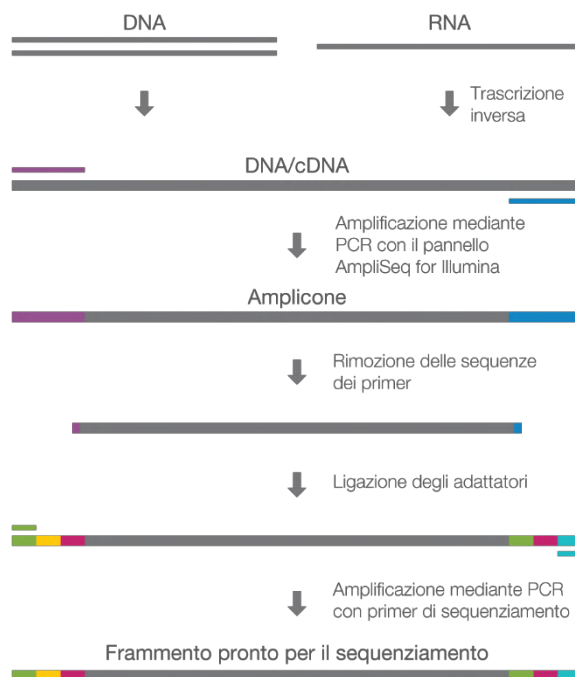


Figura 20. Preparazione delle librerie AmpliSeq for Illumina. Il flusso di lavoro è basato sulla PCR, dotato di elevato multiplex, amplifica fino a 24.000 ampliconi in un singolo saggio.

- b. Libreria basata su ibridazione a cattura: le regioni desiderate vengono “catturate” tramite ibridazione con oligonucleotidi esca e successivamente selezionate tramite purificazione. La frammentazione degli ampliconi (*Figura 21*), seguita da end-repair e A-tailing dei frammenti. Questo processo andrà a creare frammenti con una coda A, con una dimensione media di 400 bp. La frammentazione può avvenire per via enzimatica tramite enzimi che vanno a digerire il DNA.

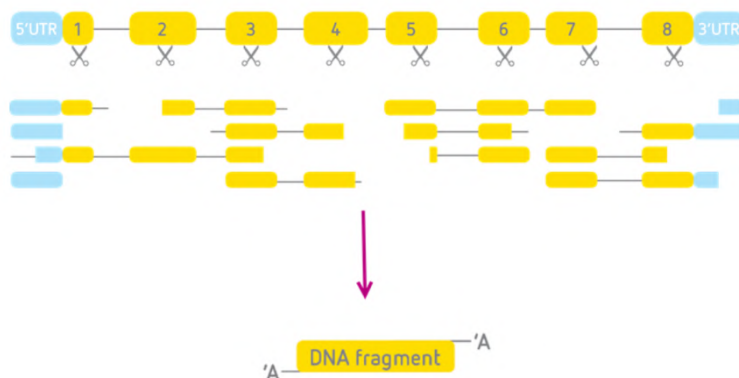


Figura 21. Rappresentazione della frammentazione di DNA.

La frammentazione provoca frammenti di DNA con brevi sporgenze di gruppi 5'-fosfati e 3'-idrossile. Le estremità dei frammenti di DNA vengono riparate per creare estremità smussate e quindi viene aggiunta una coda di A. Per creare siti di legame per gli adattatori compatibili sono necessari la riparazione finale (end-repair) del DNA e A-tailing (Figura 22). Gli adattatori vengono quindi legati al frammento di DNA. Durante questa fase si ha anche l'indicizzazione di ciascun frammento, tramite degli indici (i5 o i7) che sono direttamente legati all'adattatore. Il vantaggio di usare due indici è che si avrà bisogno di meno indici per etichettare un numero di campioni perché la combinazione di due indici crea un codice a barre univoco (Figura 23).



Figura 22. Rappresentazione della riparazione finale del DNA e A-tailing

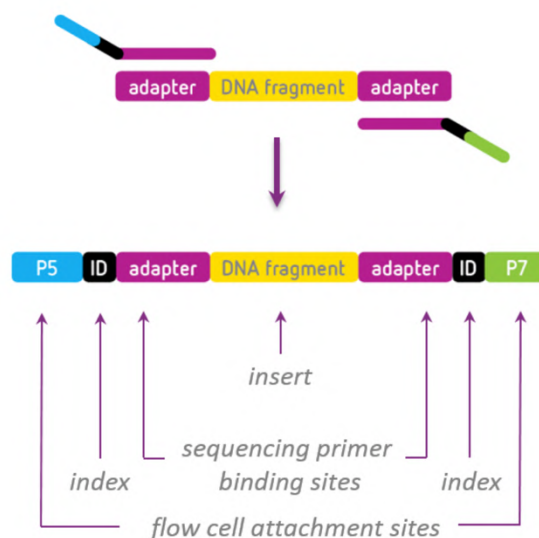


Figura 23. Rappresentazione generale del DNA frammentato legato agli adattatori. I siti P5 e P7 sono necessari per il sequenziamento.

Al fine di rimuovere i reagenti utilizzati e gli adattatori non legati, è necessaria una purificazione tramite l'utilizzo di biglie magnetiche, che vengono inoltre utilizzate per la selezione dei frammenti di nostro interesse (300 bp-700 bp). I frammenti

selezionati vengono amplificati tramite PCR. Successivamente a opportuni lavaggi, vengono eseguiti controlli qualitativi e quantitativi delle diverse librerie generate che poi vengono riunite in un unico pool per procedere poi alla sua liofilizzazione. Tale materiale è così pronto per la successiva cattura dei target tramite l'utilizzo di sonde specifiche biotinilate progettate per essere complementari alla regione di interesse. Le sonde non possono essere posizionate in regioni ripetitive poiché catturano sequenze dall'esterno della regione di interesse; pertanto, ci saranno delle lacune nel disegno della cattura. I frammenti vengono poi recuperati utilizzando biglie magnetiche con streptavidina; ciò consente una selezione della dimensione dei frammenti. I frammenti legati alle biglie sono amplificati tramite PCR e sottoposti a lavaggio. infine il pool viene sottoposto ad un controllo quantitativo e qualitativo, necessario per poi procedere al caricamento della flow cell (*Figura 24*).

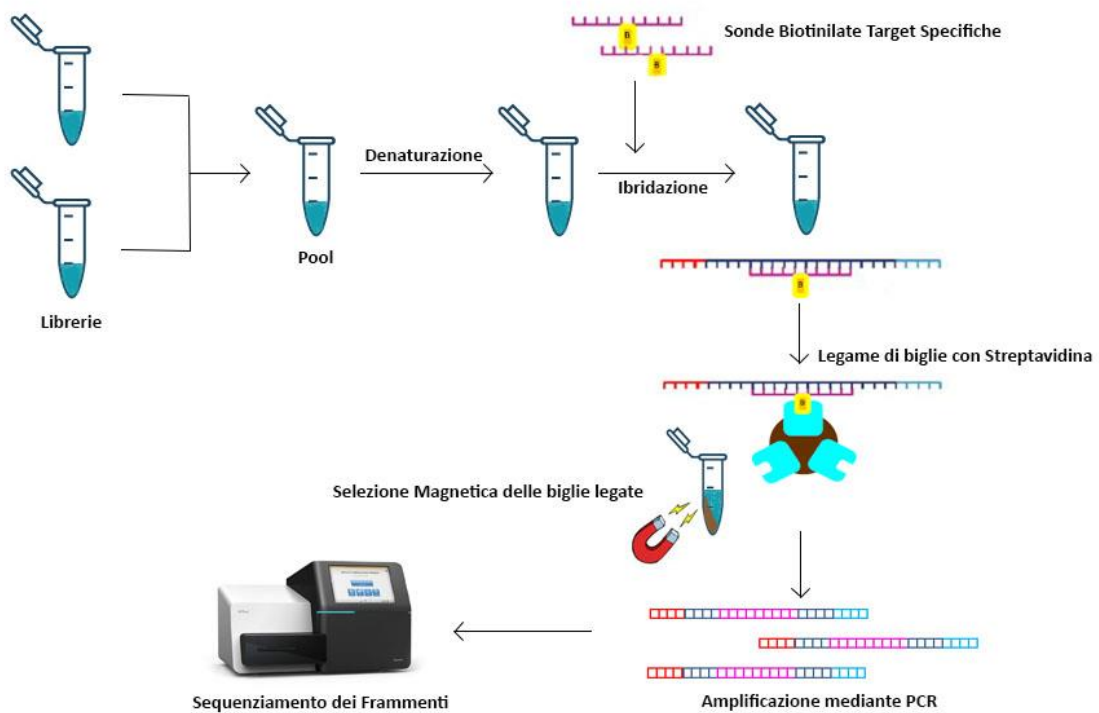


Figura 24. La figura rappresenta i passaggi dall'unione delle librerie al sequenziamento. Le librerie sono unite a formare un pool, nel quale vengono inserite delle sonde biotilate target specifiche. Le sonde permetteranno la selezione dei frammenti d'interesse tramite il legame con le biglie di streptavidina. Segue l'amplificazione dei frammenti e il sequenziamento.

La formazione di librerie basate su ibridazione cattura è un metodo più adatto per rilevare varianti strutturali in quanto vi è meno amplificazione basata su PCR ed i

frammenti quindi sono più fedeli alla rappresentazione della sequenza del genoma. Le varianti strutturali possono portare ad un inserimento o eliminazione di sequenze genomiche, ma anche ad un riarrangiamento del genoma stesso, come un'inversione o una traslocazione. Gli inserimenti o le delezioni di grandi dimensioni sono noti collettivamente come varianti del numero di copie (CNV) e possono essere > 1 Mb, mentre quelle più piccole hanno solitamente una dimensione <50 bp, esempio -3.7. NGS ha il vantaggio di essere in grado di identificare la presenza o l'assenza di un CNV e di sequenziare il punto di interruzione in una singola analisi. Ha anche il potenziale per identificare le inversioni e le traslocazioni rendendolo il metodo più completo fino ad oggi [131].

In linea con numerosi campi della biologia clinica, negli ultimi anni è emerso il sequenziamento di nuova generazione (NGS) dei geni dei gruppi sanguigni di pazienti affetti da emoglobinopatie, che sono sottoposti a trasfusioni. In uno studio è stato infatti dimostrato che NGS può in genere aiutare a identificare quelle variazioni che non sono state identificate con mezzi convenzionali e quindi, almeno parzialmente, a risolvere queste discrepanze [132]. Altri autori hanno messo in evidenza la difficile genotipizzazione ad esempio di sistemi Rh, data la complessa struttura di questi geni e l'elevata variabilità, ma allo stesso tempo è stata messa in luce la potenza dell'approccio basato su NGS in questo contesto, poiché sono state identificate varianti frequenti e rare, nonché nuove varianti geniche [133].

La tecnica di NGS è usata anche per lo studio di genotipi di pazienti affetti da talassemia; infatti nel 2015 è stata descritta la prima inversione / eliminazione che causava $\epsilon\gamma\delta\beta$ -talassemia usando un approccio NGS, che si basava sul rilevamento della presenza o dell'assenza delle variazioni di sequenza. L'elaborazione d'informatica dei dati di sequenziamento era manuale e richiedeva molto tempo. Con lo sviluppo di metodi analitici più automatizzabili è stata identificata la prima duplicazione dell'intero cluster di β -globina [131].

In uno studio è stato sviluppato un pannello di sequenziamento di nuova generazione (NGS) basato sonde a cattura. Questo pannello includeva i cluster dei geni globinici di tipo α e β , *BCL11A*, *KLF1*, *GATA1* e *HBS1L-MYB* ed ha permesso la caratterizzazione di una nuova mutazione della catena della β -globina [134].

Altri autori hanno sviluppato un pannello genetico mirato progettato, che includeva tutti gli otto geni globinici e modulatori validati (*KLF1*, *BCL11A* e *MYB*), applicandolo

allo screening molecolare e nella genotipizzazione clinica nella talassemia. Il risultato ha mostrato che l'NGS completo facilita notevolmente lo screening e la diagnosi di talassemia.

Tuttavia, ci sono ancora alcuni limiti delle tecniche NGS, poichè ad esempio in uno studio è stata identificata una nuova variante di epsilon-gamma-delta-beta talassemia usando MLPA e CGH, che è stata persa da NGS [135].

Rispetto alla caratterizzazione del gene β -globinico, il gene α -globinico è più difficile da caratterizzare, poichè che le sonde biotinilate non sono progettate per catturare sequenze ripetitive, caratteristica dei geni *HBA1* e *HBA2*. Un problema simile è presente con i due geni della γ -globina. Con il set di dati limitato generato fino ad oggi, la maggior parte dei punti di interruzione si trovano all'interno di ripetizioni di Alu. Le ripetizioni di Alu sono elementi genetici che hanno un'alta omologia e un alto numero di copie all'interno del genoma umano e quindi non catturate dalle sonde. È ben documentato che questi elementi genetici contribuiscono a eventi di ricombinazione non omologa che portano alla creazione di varianti strutturali. Le sequenze genetiche possono anche contribuire al DNA assumendo conformazioni del DNA non canonico, che non è Watson-Crick Helix [131].

La problematica principale della metodologia NGS è la limitazione imposta dalla chimica di sequenziamento che consente solo il sequenziamento di frammenti di DNA fino a 700 bp. Frammenti più grandi possono essere sequenziati ma la qualità della copertura in tutta la regione sarà ridotta a causa del numero di frammenti che possono essere sequenziati in parallelo. Se le letture potessero essere ancora più grandi, potrebbero essere analizzate correttamente anche regioni ripetitive.

Le tecnologie NGS applicate a tutto il genoma, all'esoma o ai pannelli genetici mirati, sono state efficacemente utilizzate in contesti di ricerca, nonché nei test clinici e nella diagnosi dei disturbi genetici. I risultati dell'applicazione di questa tecnologia hanno dimostrato che NGS è in grado di rilevare varianti patogene o probabilmente patogene in modo più preciso e generale, fornendo così una piattaforma efficace per lo screening molecolare e la genotipizzazione clinica in soggetti con emoglobinopatie [135].

Nonostante i numerosi punti a favore, questa tecnologia non può ancora essere sostitutiva delle altre metodiche di biologia molecolare, poichè quest'ultime sono

necessarie alla caratterizzazione sempre più completa di ogni paziente, come per coloro che presentano alterazioni a livello di regioni ripetitive.

7. L'ereditarietà e la variabilità

7.1. Ereditarietà materna

Nella progenie il DNA mitocondriale (mtDNA) viene ereditato per via materna a differenza di quello paterno è stato evidenziato da Studi effettuati sul mtDNA utilizzando gli RFLPs (*Restriction Fragment Length Polymorphisms*). Sono state formulate varie ipotesi, alcune mettono in dubbio la vitalità dei mitocondri degli spermatozoi, che anche riuscendo ad entrare nell'ovocita, dovrebbero essere fortemente danneggiati nelle loro funzioni a causa dell'eccessiva produzione di radicali liberi per la produzione di energia durante il movimento dello spermatozoo stesso. Inoltre è stata anche ipotizzata la presenza dell'ubiquitina all'interno degli ovociti, che degrada il materiale esterno alla cellula e quindi potrebbe attaccare anche il DNA mitocondriale paterno [136]. L'ipotesi più accreditata si basa su un calcolo probabilistico; infatti, nella maggior parte dei mammiferi, incluso l'uomo, al momento della fecondazione gli spermatozoi completi della coda, contenente i mitocondri, penetrano all'interno dell'ovocita. Ogni spermatozoo contiene al momento della penetrazione da 50 a 60 mitocondri: un numero molto piccolo se si considera che nell'ovocita maturo sono presenti da 100.000 a 200.000 copie di mtDNA. Sono stati eseguiti esperimenti che provano definitivamente che il DNA mitocondriale è ereditato solo per via materna da Nishimura, che ha eseguito una serie di osservazioni su alcuni spermatozoi viventi, dimostrando che sussiste una graduale diminuzione dei mitocondri durante la spermatogenesi e una rapida digestione dei pochi mtDNA rimanenti subito dopo la fecondazione, digestione che avviene ancor prima che le strutture mitocondriali siano completamente distrutte, forse per impedire la trasmissione di mtDNA spermatici dannosi alla progenie [137]. In merito alla ricombinazione nel mtDNA umano non sono mai stati osservati fenomeni ricombinativi a differenza di quanto succede per il DNA nucleare [138-139]: quindi salvo mutazioni casuali ciascuna molecola si trasmette quindi invariata. Altrimenti se il DNA mitocondriale ne fosse soggetto, tutte le ipotesi riferite alla nostra storia evolutiva finora effettuate con il modello dell'ereditarietà materna sarebbero prive di fondamento. In tale modello non sono previste né la

segregazione né la ricombinazione del mtDNA, pertanto l'ascendenza di un individuo anche dopo molte generazioni può essere ricondotta ad un solo antenato, analogamente a quanto avviene con il cromosoma Y (utilizzato per la storia evolutiva dell'uomo attraverso) [140].

7.2 Tasso evolutivo e mutazioni

È stato osservato che nonostante il DNA mitocondriale sia altamente conservato nel regno animale per struttura, dimensioni ed organizzazione, è presente una notevole variabilità a livello della sequenza nucleotidica. È stato stimato che il tasso evolutivo del mtDNA è di circa un ordine di grandezza superiore a quello dei geni nucleari a singola copia (in media circa 10-20 volte più veloce) [141]. Ciò è diretta conseguenza del tasso di mutazione, secondo la formula: $E = M \times F$, dove F rappresenta la frazione fissata di mutazioni ed M il tasso a cui avvengono le mutazioni. Il parametro M risulta maggiore nel DNA mitocondriale a causa dello sbilanciamento nella concentrazione dei quattro dNTPs [142] nonché della maggiore quantità dei cicli di replicazione, durante cui è possibile che si generino un maggior numero di errori. Nei studi evolutivi relativi all'origine della nostra specie e delle popolazioni umane l'elevata frequenza di mutazione di un gene mitocondriale rispetto ad un gene nucleare, risulta di notevole utilità ed è causa di diverse patologie. Nel mtDNA si rileva anche un aumento della frazione F; infatti poiché la dimensione effettiva della popolazione, basandosi sullo studio del solo DNA mitocondriale è un quarto di quella reale (una coppia di riproduttori diploidi contiene quattro genomi aploidi nucleari ma solo un tipo di mtDNA), l'effetto della deriva genetica (inversamente proporzionale alla dimensione della popolazione considerata) risulta quadruplicato rispetto alle osservazioni riguardanti il DNA nucleare. Quando insorge una mutazione in uno degli mtDNA presenti all'interno delle cellule, si viene a creare una situazione in cui coesistono molecole mutate e normali, situazione nota come eteroplasmia. I DNA mitocondriali di una cellula eteroplasmica che si divide, vanno incontro ad un processo noto come segregazione replicativa in cui la segregazione di molecole di tipo selvatico o di tipo mutato nelle cellule figlie, dipende esclusivamente da fenomeni casuali di deriva genetica. Il susseguirsi di tali eventi può portare nel tempo alla fissazione differenziale dei due tipi di mtDNA in cellule diverse dello stesso tessuto oppure in diversi tipi cellulari. Infine se c'è eteroplasmia anche nelle cellule della linea

germinale materna, si potrà riscontrare la fissazione differenziale dei diversi mtDNA in diversi discendenti della stessa madre eteroplasmica.

La condizione di eteroplasmia è relativamente comune nell'uomo in soggetti che presentano un quadro clinico di una certa gravità, dovuto ad esempio alla presenza di mutazioni altamente dannose nei geni dei tRNA; gli effetti di tali mutazioni sono così devastanti che possono essere il risultato solo di eventi mutazionali molto recenti; quando il livello dell'eteroplasmia è elevato per mutazioni dannose, infatti, queste ultime vengono perse in poche generazioni. Le patologie legate a mutazioni eteroplasmiche hanno gravità diversa a seconda della proporzione di mtDNA normali e mutati nei tessuti affetti di un individuo.

Non è ancora del tutto chiaro il meccanismo mediante il quale viene mantenuta l'omoplasmia, mentre risulta evidente un'estrema rapidità nel fissare o perdere le nuove mutazioni. Una possibile spiegazione tiene conto del fatto che le circa 100.000 molecole di mtDNA presenti in un oocita maturo di mammifero derivano da un numero molto ridotto di molecole capostipiti, forse addirittura solo cinque copie. Questa espansione clonale della popolazione di mtDNA nell'oocita durante l'oogenesi crea una situazione particolarmente idonea alla deriva genetica ed a fenomeni analoghi a quello definito "effetto del fondatore".

Le mutazioni del mtDNA possono essere di tre tipi: mutazione per sostituzione di una coppia di basi (mutazioni puntiformi), per inserzione, per delezione. Le sostituzioni nucleotidiche si possono suddividere in transizioni e transversioni. Al primo gruppo appartengono le sostituzioni adenina – guanina ($A \leftrightarrow G$) e citosina - timina ($C \leftrightarrow T$) ovvero purina – purina e pirimidina – pirimidina rispettivamente. Alle transversioni appartengono le sostituzioni di una purina con una pirimidina e viceversa (ad esempio $A \leftrightarrow C$ o $G \leftrightarrow T$). Complessivamente esistono due tipi di transizione e quattro tipi di transversione. Gli errori nella replicazione, che sono alla base delle mutazioni puntiformi sono dovuti principalmente alla proprietà chimica delle basi azotate di assumere due diverse forme con diversa stabilità (tautomerismo cheto – enolico o amino – iminico). Quando le basi si trovano nello stato più stabile, cheto – amminico, si appaiano secondo le regole di Watson – Crick; al contrario, il tautomerismo può generare sostituzioni puntiformi nella classe delle transizioni, consentendo appaiamenti non canonici. Tra le cause che danno origine a transversioni troviamo: a) la rotazione del legame glicosidico di una purina

in un orientamento energeticamente sfavorito; b) reazioni di ossidazione ad opera di radicali dell'ossigeno che trasformano la guanina in 8-idrossiguanina, causandone l'appaiamento preferenziale con l'adenina piuttosto che con la citosina. La stima delle probabilità che questi fenomeni avvengano è di $10^{-4} - 10^{-5}$ per le transizioni e di $10^{-5} - 10^{-6}$ per le trasversioni, ma nella realtà, sebbene il numero delle transizioni sia effettivamente superiore a quello delle trasversioni, accurati meccanismi di individuazione e riparo degli errori mantengono un tasso di mutazione di $10^{-8} - 10^{-10}$ nel DNA nucleare, tasso che risulta invece più elevato nel DNA mitocondriale. Questo, in effetti, possiede minori meccanismi di riparo rispetto al primo e la polimerasi mitocondriale è priva della capacità di correggere le bozze durante la replicazione.

Le conseguenze di ciascuna mutazione sono determinate dalla sua natura molecolare e del suo verificarsi nelle cellule somatiche o nella linea germinale [143-145], possono variare anche in funzione dello stadio del ciclo vitale dell'individuo portatore [146]. Per lo studio dell'evoluzione umana le mutazioni del mtDNA nella linea germinale sono uno strumento particolarmente idoneo e possono essere sia neutrali che dannose. Le mutazioni neutrali si sono accumulate sequenzialmente e indipendentemente sulle diverse linee di radiazione materne e permettono di distinguere individui o popolazioni umane che abbiano tempi di separazione ridotti da un punto di vista evolutivo. Invece le mutazioni dannose o potenzialmente letali vengono eliminate per effetto della selezione naturale tranne in alcuni casi in cui vengono mantenute in stato di eteroplasmia. Le mutazioni somatiche, non sono utili ai fini dello studio dell'evoluzione umana, ma possono giocare un ruolo notevole nel modulare l'espressione di patologie gravi e nei normali processi di invecchiamento. Queste possono essere dovute sia a riorganizzazioni genomiche, sia a singole sostituzioni nucleotidiche, entrambe comunque alterano la normale capacità di attuare la fosforilazione ossidativa in particolar modo nei tessuti post-mitotici.

7.3. Gli studi evoluzionistici

7.3.1 L'orologio molecolare

Il concetto di orologio molecolare, scoperto da Zuckerkandl e Pauling nel 1965, si basa sull'osservazione che, presa una qualunque proteina, il suo tasso intrinseco di evoluzione molecolare è costante nel tempo ed in tutte le linee. Secondo un

principio analogo è stato possibile stabilire un orologio molecolare associato al DNA mitocondriale, partendo dall'analisi di specie di cui erano disponibili dati sui tempi di divergenza. Il tasso medio di evoluzione risulta pari a 1-2% per milioni di anni (cioè 1-2 mutazioni su cento paia di basi ogni milione di anni). Questo orologio molecolare ha un ticchettio molto veloce utile nello studio di popolazioni che si sono differenziate in tempi recenti, come quelle umane [147-148].

Esistono due modi per esprimere le differenze evolutive: si può misurare il tasso di evoluzione (i cambiamenti avvenuti lungo una linea) contando le mutazioni che si sono accumulate a partire dal DNA ancestrale fino al suo attuale discendente, oppure si può calcolare il tasso di divergenza (numero di mutazioni tra due linee) tramite l'analisi delle mutazioni verificatesi tra due linee presenti in una popolazione rispetto all'antenato comune ad entrambe. Nel caso del DNA mitocondriale queste mutazioni si sono accumulate seguendo la linea materna ed i polimorfismi che si sono generati durante l'evoluzione possono essere analizzati per studiare il grado di divergenza tra gli individui e la loro distribuzione tra i vari gruppi umani [148].

7.3.2 La variabilità del mtDNA applicata agli studi evoluzionistici di popolazioni umane

In merito all'evoluzione umana, sono stati formulati due modelli principali: il modello multiregionale e quello dell'*Out of Africa*. Per il primo modello l'*Homo sapiens* si sarebbe evoluto indipendentemente e contemporaneamente nelle varie aree del mondo, a partire da un comune predecessore africano (probabilmente *Homo ergaster*) fuoriuscito dalla terra di origine circa 2 milioni di anni fa. L'ipotesi è compatibile con i dati archeologici delle varie aree geografiche, poiché sussiste una certa continuità anatomo – morfologica nelle sequenze di ritrovamenti fossili. Il punto debole di tale modello risiede nella difficoltà della realizzazione di un'evoluzione parallela, possibile unicamente grazie ad un flusso genico considerevole e continuo tra le popolazioni. Per il secondo modello, quello dell'*Out of Africa*, la suddivisione in realtà rispecchia un'origine africana unica e recente (circa 200000 anni fa) di tutte le altre popolazioni, poiché dei due rami principali in cui sarebbe diviso l'albero del DNA mitocondriale, uno porta esclusivamente tipi africani, mentre l'altro è composto da tutte le linee osservate nel resto del mondo e dai restanti tipi mitocondriali africani.

Gli studi effettuati prima del 1989 si basarono sull'analisi elettroforetica dei polimorfismi di lunghezza dei frammenti di restrizione (RFLP) mediante southern blotting o marcatura terminale mentre solo dai primi anni novanta fu introdotta la tecnica di sequenziamento. La prima ricerca fu effettuata da Brown nel 1980, il che analizzò 21 individui di varie regioni geografiche utilizzando 18 enzimi di restrizione. La quantità di variazione nucleotidica ottenuta, espressa come numero di sostituzioni per coppie di basi analizzate, fu 0.0018. Utilizzando quindi un tasso evolutivo dell'1% per milione di anni fu possibile stimare la data di speciazione dell'*Homo sapiens*, cioè 180000 anni fa. In realtà dato il piccolo numero di individui analizzati e il numero di enzimi, il dato non era conclusivo. Da quel momento fu chiaro che occorreva concentrare gli studi su singole aree geografiche o popolazioni per ottenere dati rilevanti [149].

Nel 1983 è stato analizzato da Johnson e collaboratori un campione di circa 200 individui (popolazioni bosciamani, bantu, nativi americani e asiatici) utilizzando 5 enzimi di restrizione ed evidenziando 35 aplotipi, di cui la maggior varietà era concentrata nelle popolazioni africane. Tale risultato portò a concludere che la nostra specie avesse un'origine africana. L'utilizzo dell'analisi di restrizione "a bassa risoluzione", condotta con 5 o 6 enzimi di restrizione, ha rivelato un albero filogenetico "star-like", cioè "simile ad una stella", con un tipo mitocondriale posto al centro, condiviso da diverse popolazioni del globo. Utilizzando l'analisi di restrizione "ad alta risoluzione", condotta con 12 enzimi di restrizione, su 147 individui provenienti da diverse aree geografiche, Cann et al ricostruirono un albero filogenetico dell'mtDNA molto più risolto [150]. Attraverso l'utilizzo del metodo della massima parsimonia (sistema che tende a ricostruire l'albero che minimizza il numero di cambiamenti necessari per ottenere l'albero osservato) i dati furono analizzati e ciò ha permesso di concludere che l'Africa è la fonte più verosimile di tutte le linee osservate nel resto del mondo [151]. In questo albero, si è osservato che un set di sequenze Africane risultava essere una radice di altre sequenze, africane e non. Questa ipotesi dell'"Out of Africa" (OOA) e della discendenza dall'africana "Eva mitocondriale" dell'uomo moderno, proposta è stata di fondamentale importanza nello stabilire il potenziale della molecola dell'mtDNA come marcatore genetico per gli studi sull'evoluzione umana [150].

Inoltre l'origine africana dell'uomo anatomicamente moderno fu datata a circa 200000 anni fa e si ricavò un modello genetico, definito "*Ipotesi del giardino dell'Eden*" (GOE), confermato da studi successivi [152], secondo cui l'espansione demografica e geografica delle popolazioni attuali avvenne contemporaneamente alla diversificazione degli aplotipi. Una rianalisi successiva [153] di tali dati riscontrò alberi filogenetici più brevi in lunghezza rispetto all'albero costruito con il metodo della massima parsimonia (MPT) del gruppo di Vigilant. Alcuni di questi alberi filogenetici non sostenevano la base africana dell'evoluzione umana. Questo portò ad un periodo di intenso dibattito, ma ulteriori dati e metodi perfezionati nella costruzione di alberi filogenetici, hanno alla fine reso solida l'ipotesi dell'OOA, proprio utilizzando le sequenze della regione di controllo [154-155].

Su larga scala fu effettuato sui nativi americani un importante studio che mise in evidenza la loro origine asiatica [156]. Proprio grazie a tale studio venne definita la nomenclatura base degli aplogruppi. Successivamente la caratterizzazione di altre popolazioni ha reso possibile la definizione di altri aplogruppi e sotto – aplogruppi. Nel 2000 l'emergenza della "mito genomica" è emersa in maniera incisiva con la pubblicazione di 53 genomi umani mitocondriali, completi, appartenenti a differenti Popolazioni [157-158]. Grazie al robusto albero filogenetico così ricavato sono stati così confermati gli studi precedenti in particolare l'osservazione che gli uomini sono migrati fuori dall'Africa per popolare il mondo. Man mano che più laboratori producevano dati di genomi mitocondriali completi, i frutti della mito genomica si sono prodotti in un gran numero di settori. Oltre alla conferma dell'ipotesi dell'"Out of Africa", vi è stata anche quella, rigorosa, dell'assenza di ricombinazione nell'mtDNA [157,159]. Interi genomi mitocondriali sono stati utilizzati per scoprire nuovi polimorfismi associati ad aplogruppi [160-161]: Torroni et al [162] hanno osservato differenze nel tasso di mutazione dei sottotipi dell'aplogruppo africano L2. È stato proposto come la selezione giochi un ruolo nella variazione del genoma mitocondriale tra differenti aplogruppi, e sono stati studiati i tassi di sostituzione sito-specifica nell'intero genoma mitocondriale [163]. È stato dimostrato come il DNA mitocondriale evolve in maniera complessa. La variabilità di ciascun sito rispetto ad un altro, relativamente ai tassi di mutazione nella regione di controllo, è una delle più evidenti peculiarità dell'mtDNA. La maggior parte dei siti nella regione di controllo ha un tasso di mutazione molto basso, o non varia assolutamente,

mentre altri siti mutano in maniera estremamente veloce. Questo enorme tasso di eterogeneità è stato riconosciuto da numerosi studi, utilizzando un'ampia varietà di metodi. Ciononostante, la complessità dell'evoluzione dell'mtDNA deve ancora essere caratterizzata e spiegata in maniera completa. L'eterogeneità nel tasso di mutazione può simulare l'espansione della popolazione ed influire sulle stime dei parametri genetici di popolazione, come quella relativa all'effettiva dimensione della popolazione [164] ed alla calibrazione dell'orologio molecolare [148].

7.3.3. Gli aplogruppi e la storia dell'uomo

L'uscita della nostra specie dal continente africano risale a circa 70000 anni, quando alcuni *sapiens* hanno abbandonato il Corno d'Africa per dirigersi verso le coste dell'Arabia, dell'Iran, dell'India, per arrivare in Asia orientale [158], espansione continuò fino all'Asia meridionale e circa 40000 anni fa raggiunse Australia e Nuova Guinea. La migrazione verso l'Europa sembra essere iniziata solo 35000 anni fa, probabilmente a causa delle condizioni climatiche sfavorevoli del nostro continente. L'ultima migrazione preistorica fu quella verso le Americhe, a partire dall'Asia settentrionale, attraverso il corridoio della Beringia, avvenuta in tre ondate successive (tra 30000 e 6000 anni fa, *Figura 25*).

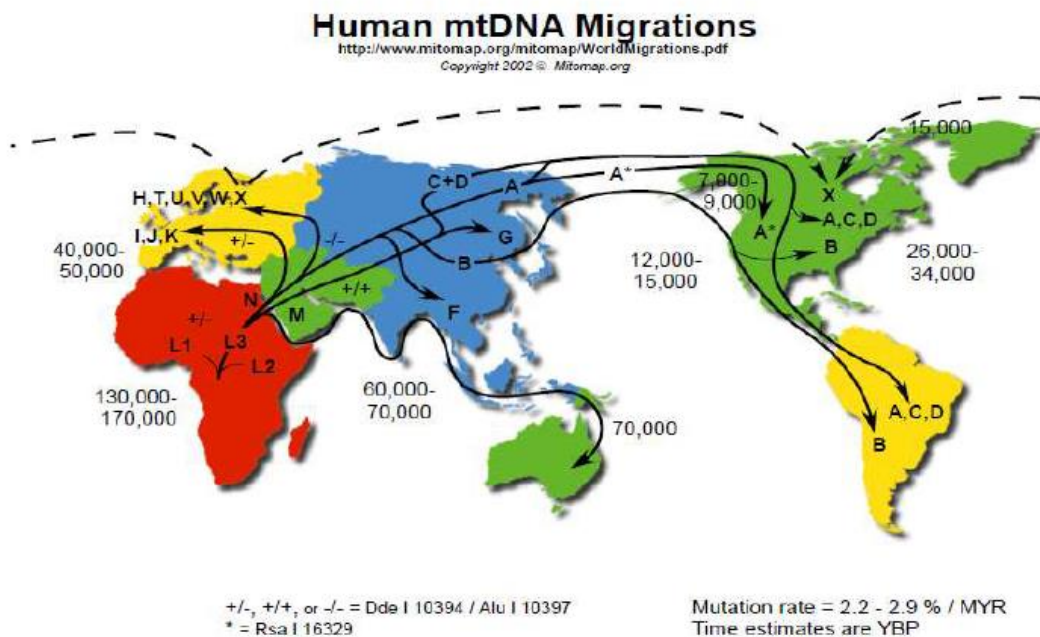


Figura 25. I flussi migratori della popolazione umana descritti dal DNA mitocondriale.

È stata dimostrata una migrazione a partire dall'Asia sud occidentale verso l'Europa e il Nord Africa, avvenuta circa 40000 anni fa, durante un periodo inter – glaciale, su rotte diverse da quelle seguite durante le prime espansioni verso l'oriente [165].

7.3.4. Gli aplogruppi principali

I polimorfismi del mtDNA sono associati a ed in particolari aplotipi, i quali possono essere raggruppati in aplogruppi (un aplogruppo, cioè, risulta costituito da un gruppo di aplotipi derivati da uno stesso mtDNA ancestrale e determinato da una mutazione specifica) [156]. Gli studi condotti su popolazioni di tutto il mondo portano alla conclusione che la variabilità del mtDNA è correlata all'origine etnica e geografica di un individuo. L'analisi della variazione degli RFLP nelle varie popolazioni, prese nella loro globalità, ha consentito l'identificazione di un certo numero di rami monofiletici, nei quali è stato possibile classificare tutti i DNA mitocondriali. Questi “clades”, o “aplogruppi”, sono stati differenziati grazie a mutazioni “antiche”, verificatesi migliaia di anni fa. Marcatori continente-specifici sono stati individuati per gli Africani [166], i Caucasiche di origine europea [167-168] e gli Asiatici [169]. Gli aplogruppi vengono designati da una lettera maiuscola seguita da un numero, che rappresenta il sottogruppo (“sub cluster”) dell'aplogruppo [168-170]. Quasi tutti gli mtDNA africani possono essere classificati in uno dei tre aplogruppi principali (L1, L2 o L3). Dal momento che lo si ritrova alla radice dell'albero filogenetico l'aplogruppo L1 è ritenuto essere quello ancestrale. I superaplogruppi Euroasiatici M ed N sono ritenuti aver lasciato l'Africa circa 50.000-60.000 anni fa, per poi dare origine agli aplogruppi attuali in Asia (A, B, C, D, F) ed in Europa (H, I, J, K, T, U, V, W, X) (Tabella 3). I motivi di sequenza nella regione di controllo del mtDNA correlano fortemente con gli aplogruppi definiti mediante l'analisi in RFLP [168].

In Africa l'aplogruppo principale è L [166] che, con i suoi sotto – aplogruppi (L1a, L1b, L2 ed L3), rappresenta quasi la totalità della popolazione sub sahariana. L1 è l'aplogruppo più diffuso (circa 50%), seguito da L2 e L3. Quest'ultimo rappresenta una piccola frazione della popolazione, ma sembra essere il progenitore di ben la metà delle linee di mtDNA in Europa, Asia e Nativi Americani. L'origine dell'aplogruppo L è datata tra 162000 e 123000 anni fa, considerando il tasso di evoluzione 2.2-2.9% per milione di anni [167].

In Asia l'aplogruppo M si riscontra in più della metà degli individui (circa 55%), ed è suddiviso nei sotto – aplogruppi C, D, G, E. La maggior parte dei restanti individui è compresa negli aplogruppi minori A, B e F.

Nei nativi americani sono stati trovati solo 4 aplogruppi asiatici: A, B, C e D. L'aplogruppo principale risulta essere l'A, uno dei più antichi [156]. Per l'Europa, l'analisi filogenetica è più complessa. L'aplogruppo principale è l'H, che raggruppa quasi il 40% della popolazione; sono stati definiti altri sottoaplogruppi, che mostrano una caratteristica distribuzione regionale. Tra questi, H1 ed H3 sono di comune riscontro nell'Europa occidentale, essendosi espansi dal rifugio Franco-cantabrico [171-172], mentre un sottotipo di H2, definito dalla transizione in 951, è tipico dell'Europa orientale e dell'Asia. Sono stati evidenziati altri 9 aplogruppi (tra cui anche M), di cui I, J e K sono i più rappresentati [168, *Tabella 3*].

CONTINENTE	APLOGRUPPI PRINCIPALE	APLOGRUPPI MINORI	% aplog. principali
Africa	L	M, U	76%
Europa	H, I, J, K	M, T, U, V, W, X	64%
Asia	M	A, B, F	55%
Nat. Americani	A	B, M	100%

Tabella 3. Schema sintetico degli aplogruppi presenti nel Mondo.

7.3.5. Il DNA mitocondriale europeo

Lo studio del DNA mitocondriale ha portato all'identificazione di aplogruppi specifici per l'Europa durante gli anni 90 [162, 167-168]. Il primo studio fu effettuato nel 1994 su 175 DNA caucasici provenienti dagli Stati Uniti e dal Canada, mediante PCR di amplificazione e successiva analisi dei polimorfismi di restrizione. Lo studio dimostrò che tutti i campioni erano ascrivibili per il 59% in 4 aplogruppi, H, I, J e K di cui l'H da solo costituiva il 40%. Un esame più completo di questi aplogruppi fu eseguito nel 1996 attraverso l'analisi di tre popolazioni europee (Svedese, Finlandese e Toscana, scelta sulla base di analisi linguistiche effettuate da Cavalli – Sforza come rappresentative di gruppi principali europei). Furono evidenziati 10 aplogruppi, tra cui l'aplogruppo M, tipico dell'Asia, e i quattro già identificati precedentemente. A questi si affiancarono gli aplogruppi T, U, V, X e W.

L'aplogruppo H è stato confermato come il più rappresentato (*Figura 26 e Tabella 4*).

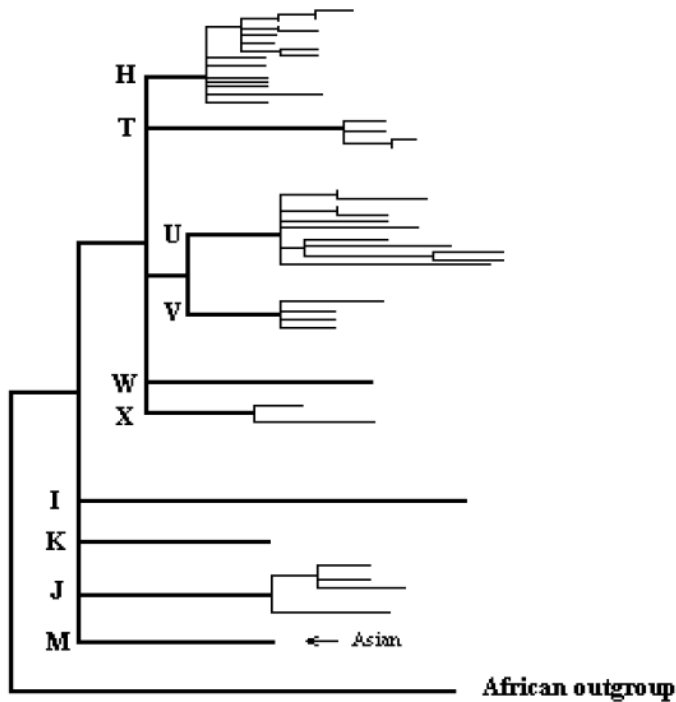


Figura 26. Albero filogenetico degli aplogruppi europei. Albero ottenuto col metodo della massima parsimonia rappresentante i risultati ottenuti nel lavoro di Torroni del 1996 relativamente alle popolazioni Finlandese e Svedese. L'outgroup utilizzato per la radice è l'aplotipo AF71, aplogruppo L, proveniente dal Senegal [166].

Tramite il paragone con i dati precedenti, fu dimostrato che gli aplogruppi H e K erano principalmente, ma non esclusivamente, europei, che l'aplogruppo J era esclusivo dell'Europa; che gli aplogruppi T, I e W non potevano essere collocati tra gli aplogruppi continente – specifico, poiché non c'erano ancora dati riguardo l'enzima di restrizione BstNI (identificativo dei suddetti aplogruppi) nei campioni non – europei analizzati; che l'aplogruppo X probabilmente era di derivazione asiatica, ma difficilmente analizzabile a causa del sito di restrizione al nucleotide 1715, perso varie volte in maniera indipendente nel corso della storia; che l'aplogruppo U era tipico dell'Europa e presente in alcuni individui dell'Africa Sub – Sahariana; che l'aplogruppo V non era classificabile con certezza per mancanza di informazioni. Si è concluso che gli aplogruppi H, I, J, K, T e W sembravano essere confinati essenzialmente in Europa ed essere il frutto di un'evoluzione avvenuta dopo l'arrivo in Europa dei primi *sapiens*, mentre per l'aplogruppo U era ipotizzabile un'origine africana più diretta. Ad oggi, dopo altri studi e con più risultati a livello mondiale, si possono ritenere ancora validi i risultati dei lavori appena citati, anche se l'albero

filogenetico risulta molto più accurato e le relazioni filogenetiche più complesse grazie al sequenziamento totale del mtDNA, che ha permesso la determinazione di numerosi marcatori polimorfici [160, 173]. Le frequenze per gli aplogruppi europei sono sintetizzate nella *Tabella 4*.

APLOGRUPPI	FREQUENZE
H	40.9%
T	12.7%
U	14.3%
V	3.2%
X	4.1%
W	2.1%
I	2.1%
J	10.5%
K	8.0%
Altri	2.1%

Tabella 4. Frequenze degli aplogruppi europei [168].

L'aplogruppo H sembra essersi originato nell'Europa dell'est circa 30000 – 25000 anni fa. Si sarebbe poi espanso nel resto del continente prima della seconda pleniglaciazione, contemporaneamente all'espansione della cultura Gravettiana (25000 – 20000 anni fa). La *Tabella 5* indica la frequenza dell'aplogruppo H in alcuni paesi e aree geografiche. Tra i sotto – aplogruppi in cui può essere suddiviso, almeno per quanto riguarda l'Europa dell'Ovest, i due principali sono H1 e H3 [171]. Infatti recenti studi di Roostelu e collaboratori hanno evidenziato la presenza di numerosi e frequenti nuovi sotto-aplogruppi nelle regioni dei balcani e del Medio Oriente.

Regione	Frequenza H (%)
Senegal	-
Marocco	36.8
Algeria	25.6
Egitto	1.4
Spagna	39.4
Francia	47.2
Olanda	38.2
Austria	44.8
Albania	48.6
Ungheria	42.3
Polonia	37.2
Turchia	26.0
Iraq	19.9
Penisola Arabica	10.6
Asia centrale	11.2

Tabella 5. Distribuzione dell'aplogruppo H nelle popolazioni.

8. SCOPO DELLA TESI

Lo scopo del primo progetto di ricerca è quello di eseguire uno studio pilota che permetta di correlare alcune alterazioni nucleari che mitocondriali in pazienti affetti da emoglobinopatie in modo da poter costruire una rappresentazione filogeografica degli stessi. Tale studio potrebbe così consentire una sempre più precisa stratificazione dei soggetti affetti da emoglobinopatie che rappresenta un requisito necessario sia ai fini diagnostici che per la gestione clinica del paziente.

Questo studio è stato condotto su 53 pazienti affetti da emoglobinopatie (15 beta-talassemia, 10 HbS/Beta tal, 20 HbS/HbS, 8 eterozigoti composti per varianti emoglobiniche qualitative) provenienti da sedici diversi paesi. Lo scopo è stato quello di investigare una possibile correlazione tra alterazioni dei geni globinici (*HBB*, *HBA1*, *HBA2*), polimorfismi nucleari di geni modulatori del fenotipo clinico (*HBG2*, rs7482144; *BCL11A*, rs1427407 e rs10189857; *HBS1L-MYB*, rs28384513 e rs939913, *C/EBPE*, rs45496295) e le informazioni derivanti dalla regione di controllo mitocondriale (tra i siti 15877 e 259).

Allo scopo di effettuare una sempre più accurata correlazione genotipo-fenotipo e poter identificare nuovi potenziali geni modificatori in soggetti affetti è stato condotto un secondo progetto di ricerca su due fratelli affetti da Beta Talassemia Major con identico genotipo beta -29 A>G /cod8 (-AA) e geni alfa normali con diversa espressività fenotipica, inclusi nella precedente analisi (pz 3 e 4). Ai fini di individuare alterazioni in tali soggetti che potrebbero spiegare tale diversità fenotipica sono stati analizzati mediante un pannello di Next Generation Sequencing, sviluppato e validato presso il Laboratorio di Genetica Molecolare dell'Università degli Studi di Perugia in collaborazione con la ditta SOPHiA GENETICS, che permette l'analisi di 75 geni, di cui otto geni associati ad Emoglobinopatie, undici coinvolti nel metabolismo del ferro ed i restanti legati ai disordini ematopoietici. Per avere maggiori informazioni in merito all'assetto genetico delle varianti nei due fratelli sono stati analizzati mediante la stessa metodica di sequenziamento anche i genitori.

9. MATERIALI E METODI DEL PRIMO STUDIO

9.1 Casistica

Lo studio è stato condotto su 53 pazienti (M/F: 38/15) affetti da emoglobinopatie che si sono rivolti al laboratorio di Genetica Molecolare dell'Università degli Studi di Perugia, Ospedale Santa Maria della Misericordia, tra il 1988 e il 2018 in Umbria una regione del Centro Italia. Ad eccezione del primo paziente di cui non si conosce l'origine, 16 pazienti (30,2%) erano europei, 28 (52,8%) africani, quattro (7,6%) asiatici e quattro (7,6%) americani. Informazioni complete sulle caratteristiche geografiche e sui genotipi sono state riportate nella Tabella 6, mentre i fenotipi clinici sono riportati nell'appendice A.

La coorte in esame presentava i seguenti soggetti omozigoti o eterozigoti composti: 15 pazienti con β -talassemia, 20 pazienti con drepanocitosi o anemia falciforme (HbS / HbS), dieci pazienti con microdrepanocitosi (HbS / β^+ o β^0) ed otto eterozigoti composti per varianti emoglobiniche di cui sette HbS / HbC e un HbO-Arab / HbC (Figura 27).

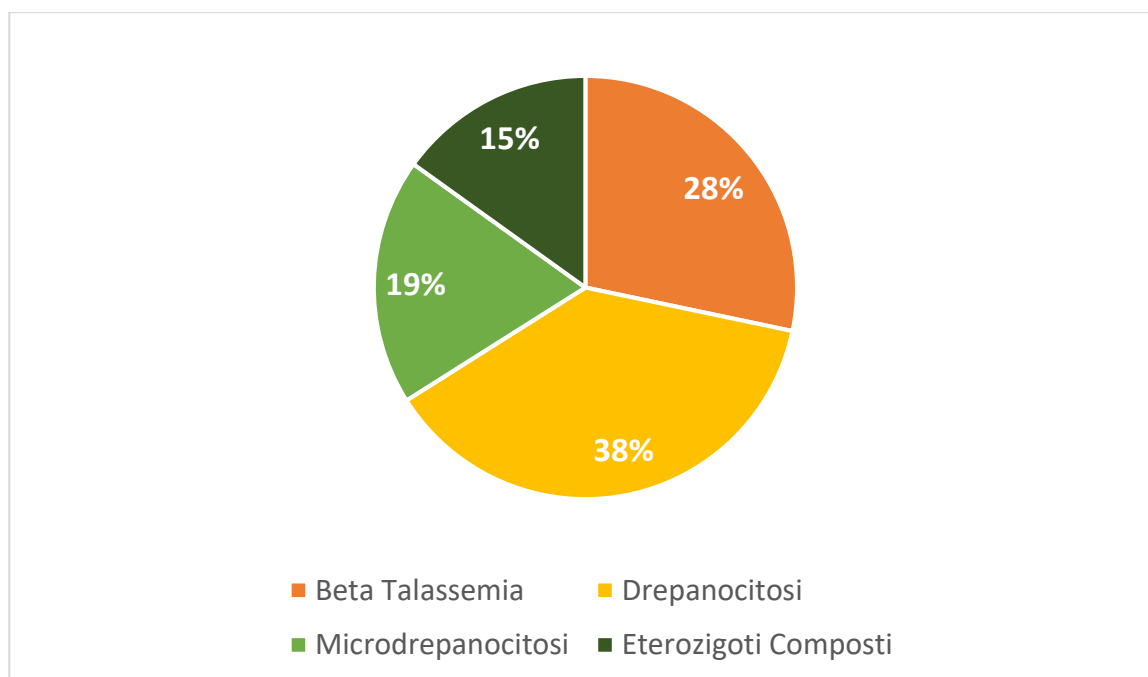


Figura 27. Genotipi dei 53 pazienti in analisi.

Dieci pazienti avevano mutazioni nei geni alfa globinici, in particolare: otto erano eterozigoti per la delezione $\alpha^{-3,7}$ e due per la triplicazione $\alpha\alpha^{\text{anti } 3,7}$. Sono state trovate diciassette diverse mutazioni del gene *HBB*: 12 difetti β -talassemici, due

varianti emoglobiniche talassemiche (Hb-Lepore e HbE) e tre varianti emoglobiniche (HbS, HbC e HbO- Arab).

Complessivamente, i pazienti con β -talassemia (pz 1-15) presentavano 17 mutazioni geniche differenti e 12 diversi genotipi. Sette pazienti su 15 (46,6%) erano europei, due (13,3%) africani, quattro (26,7%) asiatici e uno (6,7%) americano. Non erano disponibili informazioni relative all'origine etnica per un paziente (6,7%).

I pazienti con microdrepanocitosi (pz 16- 25) presentavano quattro diversi genotipi β mentre in due pazienti albanesi sono state riscontrate mutazioni α ($\alpha\alpha^{\text{anti } 3.7}$). Sette degli undici pazienti erano europei (63,6%), due (18,2%) africani e uno (9%) americano.

Tutti i pazienti con drepanocitosi (pz 25-45) portavano lo stesso genotipo β , mentre la presenza del genotipo $-\alpha^{3,7}/\alpha\alpha$ è stata dimostrata in cinque pazienti e il genotipo $-\alpha^{3,7}/-\alpha^{3,7}$ è stato evidenziato in un paziente. Sedici su 20 (80%) erano pazienti africani, due erano americani (10%) e due erano europei (10%).

Inoltre nel gruppo degli eterozigoti composti (pz 46- 53) sette pazienti su otto risultavano portatori del genotipo HbS / HbC (pazienti 46-51, 53), mentre l'altro paziente presentava genotipo HbC / HbO-Arab (paziente 52). Il genotipo $-\alpha^{3,7}/\alpha\alpha$ è stato riportato in due casi di HbS / HbC. Tutti e otto i pazienti erano africani (Tabella 6).

	Numero	Sesso-Età ^a	Origini etniche		Genotipo β				Genotipo α
			Continente	Paese	Allele 1		Allele 2		
					Nomenclatura Classica	HGVS nomenclature	Nomenclatura Classica	HGVS nomenclature	
β-Talassemia (β ⁺ /β ⁺ , β ⁰ /β ⁰ o β ⁺ /β ⁰)	1	M	Non Disponibile	Non Disponibile	IVS-I-1(G>A)(β ⁰)	HBB:c.92+1G>A	IVS-I-1(G>A)(β ⁰)	HBB:c.92+1G>A	αα/αα
	2	M-18	Europa	Albania	codon39(C>T)(β ⁰)	HBB:c.118C>T	codon 44(-C)(β ⁰)	HBB:c.135delC	αα/αα
	3 ^{ab}	M-3	Africa	Algeria	- 29(A>G)(β ⁺)	HBB:c.-79A>G	codon 8(-AA)(β ⁰)	HBB:c.25_26delAA	αα/αα
	4 ^{ab}	F-4	Africa	Algeria	- 29(A>G)(β ⁺)	HBB:c.-79A>G	codon 8(-AA)(β ⁰)	HBB:c.25_26delAA	αα/αα
	5	M	Europa	Italia	codon39(C>T)(β ⁰)	HBB:c.118C>T	IVS-I-6(T>C)(β ⁺⁺)	HBB:c.92+6T>C	αα/αα
	6	M-1	Europa	Albania	IVS-I-110(G>A)(β ⁺)	HBB:c.93-21G>A	IVS-I-110(G>A)(β ⁺)	HBB:c.93-21G>A	αα/αα
	7*	M-15	America	Venezuela	codon17(A>T)(β ⁰)	HBB:c.52A>T	- 29(A>G)(β ⁺)	HBB:c.-79A>G	αα/αα
	8	M-18	Europa	Albania	IVS-I-5(G>C)(β ⁺)	HBB:c.92+5G>C	IVS-I-6(T>C)(β ⁺⁺)	HBB:c.92+6T>C	αα/αα
	9	F-62	Europa	Italia	IVS-I-6(T>C)/(β ⁺⁺)	HBB:c.92+6T>C	IVS-I-6(T>C)(β ⁺⁺)	HBB:c.92+6T>C	αα/αα
	10	F-31	Europa	Italia	Codon39(C>T)(β ⁰)	HBB:c.118C>T	IVS-I-6(T>C)/(β ⁺⁺)	HBB:c.92+6T>C	αα/αα
	11*	M-3	Asia	Siria	codon 5 (-CT)(β ⁰)	HBB:c.17_18deICT	IVS-I-6(T>C)/(β ⁺⁺)	HBB:c.92+6T>C	αα/αα
	12*	M-4	Asia	Siria	codon 5 (-CT)(β ⁰)	HBB:c.17_18deICT	IVS-I-6(T>C)/(β ⁺⁺)	HBB:c.92+6T>C	αα/αα
	13	F-19	Asia	Cina	- 28(A>G)(β ⁺)	HBB:c.-78A>G	cd17(A>T); AAG(Lys)>TAG(Stop codon)(β ⁰)	HBB:c.52A>T	αα/αα
	14	M-29	Europa	Italia	IVS-II-745 (C>G) (β ⁺)	HBB:c.316-106C>G	Hb Lepore Boston-Washington	NG_000007.3:g.63632_71046del	αα/αα
	15	M-10	Asia	Cambogia	Hb E	HBB:c.79G>A	Hb E	HBB:c.79G>A	αα/αα
Microdrepanocitosi (HbS/β ⁺ o β ⁰)	16*	F-18	Europa	Albania	HbS	HBB:c.20A>T	codon 39(C>T)(β ⁰)	HBB:c.118C>T	αα/αα
	17*	M-11	Europa	Albania	HbS	HBB:c.20A>T	codon 39(C>T)(β ⁰)	HBB:c.118C>T	αα ^{anti3.7} /αα
	18	F-5	Europa	Albania	HbS	HBB:c.20A>T	codon 39(C>T)(β ⁰)	HBB:c.118C>T	αα/αα
	19	M-6	Africa	Costa d'Avorio	HbS	HBB:c.20A>T	-29(A>G)(β ⁺)	HBB:c.-79A>G	αα/αα
	20	M-46	Europa	Albania	HbS	HBB:c.20A>T	IVS-I-6(T>C)(β ⁺⁺)	HBB:c.92+6T>C	αα/αα
	21*	F-52	America	Venezuela	HbS	HBB:c.20A>T	-29(A>G)(β ⁺)	HBB:c.-79A>G	αα/αα
	22	M-51	Europa	Italia (Sicilia)	HbS	HBB:c.20A>T	IVS-I-110(G>A)(β ⁺)	HBB:c.93-21G>A	αα/αα
	23	M-2	Europa	Albania	HbS	HBB:c.20A>T	IVS-I-110(G>A)(β ⁺)	HBB:c.93-21G>A	αα/αα
	24	M-19	Africa	Guinea	HbS	HBB:c.20A>T	-29(A>G)(β ⁺)	HBB:c.-79A>G	αα/αα
	25	M-24	Europa	Albania	HbS	HBB:c.20A>T	codon 39(C>T)(β ⁰)	HBB:c.118C>T	αα ^{anti3.7} /αα
Drepanocitosi (HbS/HbS)	26	M-7	Europa	Albania	HbS	HBB:c.20A>T	HbS	HBB:c.20A>T	αα/αα
	27	M-32	America	Repubblica Dominicana	HbS	HBB:c.20A>T	HbS	HBB:c.20A>T	αα/αα
	28	M-4	Africa	Nigeria	HbS	HBB:c.20A>T	HbS	HBB:c.20A>T	(-) α ^{3.7} /αα

	29	M-1	Africa	Nigeria	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα
	30	M-4	Africa	Nigeria-Costa d'Avorio	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα
	31	M-2	Africa	Nigeria-Costa d'Avorio	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα
	32	M-5	Africa	Nigeria	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα
	33	M-2	Africa	Nigeria	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	(-) α ^{3.7} /αα
	34	M-1	Africa	Nigeria	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα
	35	M-49	Europa	Italia	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	(-) α ^{3.7} /αα
	36	F-5	Africa	Nigeria	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα
	37	F-1	Africa	Nigeria	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα
	38	M-14	Africa	Congo	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	(-) α ^{3.7} /αα
	39	M-1	Africa	Benin	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα
	40	M-12	Africa	Senegal	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	(-) α ^{3.7} /αα
	41	M-11	America	Venezuela	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα
	42	M-60	Africa	Nigeria	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	(-) α ^{3.7} -α ^{3.7}
	43	F-1	Africa	Nigeria	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα
44	M-10 days	Africa	Congo	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα	
45	M-2	Africa	Nigeria	HbS	<i>HBB:c.20A>T</i>	HbS	<i>HBB:c.20A>T</i>	αα/αα	
Eterozigoti Composti (HbS/HbC, HbO-Arab/HbC)	46	M-38	Africa	Costa d'Avorio	HbS	<i>HBB:c.20A>T</i>	HbC	<i>HBB:c.19G>A</i>	(-) α ^{3.7} /αα
	47*	M-2	Africa	Costa d'Avorio	HbS	<i>HBB:c.20A>T</i>	HbC	<i>HBB:c.19G>A</i>	αα/αα
	48	F-1	Africa	Ghana	HbS	<i>HBB:c.20A>T</i>	HbC	<i>HBB:c.19G>A</i>	αα/αα
	49*	F-14	Africa	Costa d'Avorio	HbS	<i>HBB:c.20A>T</i>	HbC	<i>HBB:c.19G>A</i>	(-) α ^{3.7} /αα
	50*	F-16	Africa	Costa d'Avorio	HbS	<i>HBB:c.20A>T</i>	HbC	<i>HBB:c.19G>A</i>	αα/αα
	51	F-24	Africa	Nigeria	HbS	<i>HBB:c.20A>T</i>	HbC	<i>HBB:c.19G>A</i>	αα/αα
	52	F-59	Africa	Marocco	HbO-Arab	<i>HBB:c.19G>A</i>	HbC	<i>HBB:c.364G>A</i>	αα/αα
	53	M-17	Africa	Marocco	HbS	<i>HBB:c.20A>T</i>	HbC	<i>HBB:c.19G>A</i>	αα/αα

Tabella 6. Informazioni geografiche e genotipi α/β dei 53 pazienti analizzati in questo studio. ^a Età al momento dei test genetici; * Pazienti imparentati; ^b Corrispondono rispettivamente ai pazienti 3b e 4b del secondo progetto di ricerca.

9.2. Pazienti ed estrazione del DNA

I campioni di sangue periferico sono stati raccolti durante gli esami medici di routine e genomici. Il DNA è stato isolato mediante procedura di salting-out [174]. Alla diagnosi e / o durante il follow-up tutti i pazienti o i loro genitori / tutori hanno fornito consenso informato per la raccolta e la conservazione dei campioni e per studi collaborativi di ricerca. I Protocolli sono stati approvati dal Comitato Etico per la Sperimentazione Clinica dell'Università di Perugia (protocollo n. 2017-01) e le analisi sono state eseguite in accordo con la dichiarazione di Helsinki.

9.3. Analisi del DNA genomico

È stata utilizzata sia la tecnica del reverse dot blot per l'identificazione delle mutazioni più frequenti dei geni β ed α globinici, basata sulla reazione a catena della polimerasi (PCR) e sull'ibridazione inversa (β -Globin StripAssay, α -Globin StripAssay, Vienna Lab; beta globina test, alfa globina test, Nuclear Laser Medicine), che il sequenziamento del gene beta secondo il metodo sanger.

Per il sequenziamento del gene beta i prodotti una volta amplificati sono stati purificati e quindi usati per la reazione di cycle sequencing utilizzando "BigDye terminator 1.1 cycle sequencing Kit" (Applied Biosystems) con i seguenti primers (Figura 28):

- *R95_F* (FORWARD) 5'-TAAGCCAGTGCCAGAAGAGCC-3'
- *pCO4_R* (REVERSE) 5'-CAACTTCATCCACGTTCCACC-3'
- *PCO5_F* (FORWARD) 5'-GGTGAACGTGGATGAAGTTG-3'
- *R94_R* (REVERSE) 5'-TCCTATGACATGAACTTAAC-3'
- *C_F* (FORWARD) 5'-GTGTACACATATTGACCAA-3'
- *AMD1_F* (FORWARD) 5'-CTGCATATAAATTGTAAGT-3'
- *R61_R* (REVERSE) 5'-CACTGACCTCCCACATTCCC-3'

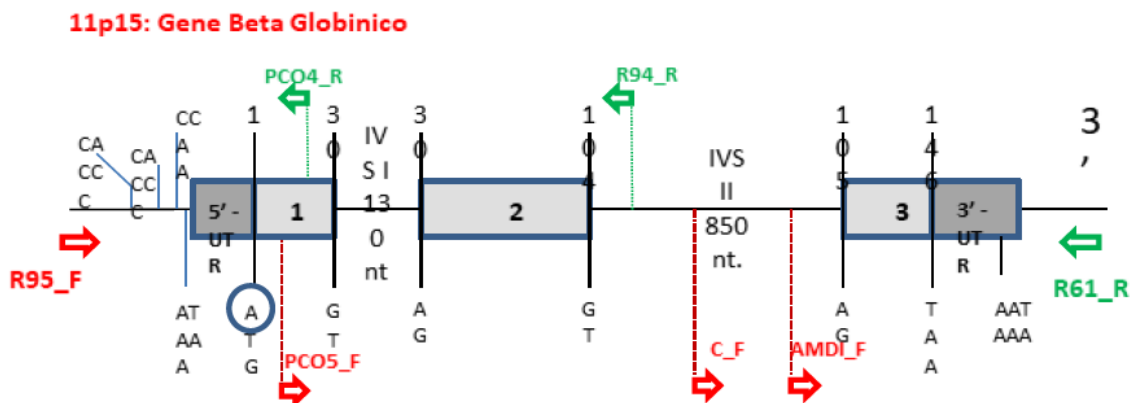


Figura 28. Schema del gene Beta globinico. In questa rappresentazione sono indicati in rosso i primers forward ed in verde quelli reverse utilizzati per il sequenziamento secondo metodo Sanger.

Sono stati studiati cinque SNPs associati alla gravità del fenotipo che modulano rispettivamente *HBG2* rs7482144 [C> T], *BCL11A* rs1427407 [G> T], *BCL11A* rs10189857 [A> G], *HBS1L-MYB* rs28384513 [A> C] e *HBS1L-MYB* rs9399137 [T>

C], utilizzando la reazione a catena della polimerasi (PCR), il reverse dot-blot basato su PCR (β -Thal Modifier StripAssay; ViennaLab Diagnostics GmbH, Vienna, Austria). C /EBPE, rs45496295 [C> T] è stato analizzato mediante PCR utilizzando i seguenti primers: CEBP_F: 5'-GACAGGAGCTGGCTCTGAGT-3' e CEBP_R: 5'-TCCGCAGAGTTAGGCCGT-3'. Il sequenziamento dei nucleotidi è stato eseguito utilizzando il kit BigDye™ Terminator Cyclor Sequencing e ABI PRISMTM 310 Genetic Analyzer (PE Applied Biosystems, Foster City, CA, USA) [175]. La frequenza degli SNPs nelle diverse popolazioni è stata ottenuta dal database GnomAD (Genome Aggregation Database, <https://gnomad.broadinstitute.org/>).

9.4. Analisi del DNA mitocondriale

La regione di controllo mitocondriale tra i siti 15877 e 259 è stata amplificata utilizzando due primers progettati sulla sequenza di riferimento revisionata, Cambridge Reference Sequence (rCRS), con numero di accesso NC_012920 [176]: forward 5'-CAAATGGGCCTGTCCTTGTA-3' e reverse 5'-TGTGCAGACATTCAATTGTT-3'. I frammenti di PCR sono stati prima purificati enzimaticamente (ExoSAP-IT1, USB Corporation, Cleveland, OH, USA), poi sottoposti a sequenziamento mediante l'applicazione di ABI Prism™ Big-Dye Terminator chimica sull'analizzatore genetico ABI PRISMTM 310 (PE Applied Biosystems, Foster City, CA, USA), con il forward primer 5' CCATTAGCACCCAAAGCTA-3'. Le sequenze sono state assemblate e allineate a rCRS utilizzando Sequencher™ 5.10 (Gene Codes Corporation). Ogni volta che gli elettroferogrammi mostravano ambiguità, sono state eseguite nuove amplificazioni PCR e reazioni di sequenziamento. L'assemblaggio finale comprendeva almeno l'intera sequenza ipervariabile I e parte della regione ipervariabile II (da np 16024 a np 210). Tutte le sequenze della regione di controllo mitocondriale sono state depositate in GenBank con i numeri di accesso MT176183-MT176235. Per ogni campione, gli aplotipi sono stati registrati e classificati in aplogruppi secondo i rispettivi motivi mutazionali facendo riferimento al software HaploGrep2 basato sul database PhyloTree build 17, www.phylotree.org [177]. Gli indici di variazione genetica sono stati stimati utilizzando il software DnaSP 5.1 [178]. Le relazioni evolutive tra aplotipi sono state visualizzate attraverso la costruzione di una rete di unione mediana utilizzando il software Network 10 (www.fluxus-engineering.com).

Come impostazione predefinita, tutti i polimorfismi relativi a rCRS sono stati pesati con un valore di dieci, ad eccezione del sito altamente variabile 16519 e delle variazioni nel poly-C si estendono intorno al np 16189 che sono stati ridotti a uno. Al contrario, ai motivi mutazionali sono stati assegnati pesi tripli 16223 e 73 al fine di rafforzare la topologia della rete. Le analisi dei componenti principali (PCA) sono state eseguite utilizzando il software Excel implementato da XLSTAT, come descritto altrove [179] per spiegare la varianza dei dati multivariati (variazione del DNA mitocondriale e nucleare, considerata separatamente in ogni campione).

10. MATERIALI E METODI DEL SECONDO STUDIO

10.1. Casistica

È stato sviluppato e validato, utilizzando 48 campioni, un pannello di Next generation Sequencing presso il laboratorio di Genetica Molecolare dell'Università degli Studi di Perugia in collaborazione con la ditta SOPHiA GENETICS (tecnologia Illumina Miseq) per lo studio di 75 geni, di cui otto geni associati ad Emoglobinopatie, undici coinvolti nel metabolismo del ferro ed i restanti implicati in disordini ematopoietici (appendice B). Dei 48 pazienti analizzati, vi erano 31 con patologie onco-ematologiche, mentre nove campioni affetti da emoglobinopatie. L'analisi dei dati nei nove campioni è stata effettuata con tre criteri di ammissione: 1) SNP >20% e INDEL >15%, 2) coverage>50X, 3) esclusione delle varianti off target (SOPHiA DDM®). In questo secondo studio sono stati analizzati due fratelli algerini, già inclusi ed analizzati nel primo progetto di ricerca, affetti da Beta Talassemia Major con identico genotipo beta e geni alfa normali, ma con diversa espressività fenotipica (pz 3 e 4). L'eterogeneità clinica si esprime in termini di diverso: supporto trasfusionale, profilo di sovraccarico marziale T2* e differente risposta alla ferrochelazione. In particolare:

- Paziente 3: regime trasfusionale più intenso (2 GRC/settimana) senza sovraccarico marziale Rmt2* (deferasirox)
- Paziente 4: regime trasfusionale meno intenso (1 GRC/settimana), moderato sovraccarico cardiaco ed epatico, ferritinemia elevata (deferasirox cpr

rivestite, pregressa idiosincrasia epatica a deferasirox dispersibile e agranulocitosi con deferiprone)

In seguito l'analisi è stata estesa anche ai genitori (pz 10b e 11b) allo scopo di avere maggiori informazioni in merito all'assetto genotipico (Figura 29).

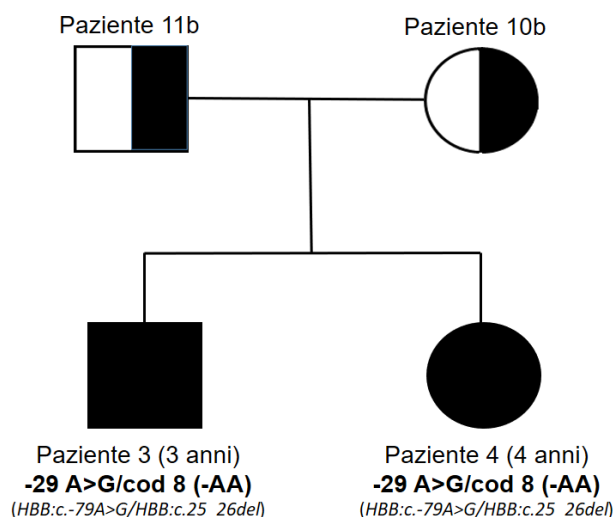


Figura 29. Albero genealogico della famiglia in studio.

I campioni dei soggetti affetti da emoglobinopatie sono stati analizzati presso il Laboratorio di Genetica Molecolare dell'Università degli Studi di Perugia dietro sospetto clinico già supportato da analisi ematobiochimiche di primo livello (Emocromo, Elettroforesi Emoglobine, HPLC emoglobina, metabolismo marziale). I genotipi, identificati tramite sequenziamento con metodo Sanger e Reverse Dot Blot sono riportati in *Tabella 7*.

N°	GENOTIPO β	GENOTIPO α
1b	HbS / cd 39 C>T	anti ^{-3.7} / anti ^{-3.7} - anti ^{-3.7} / $\alpha\alpha$
2b	HbS / HbS	α ^{-3.7} / $\alpha\alpha$
3b*	-29 A>G / cd8 (-AA);	$\alpha\alpha$ / $\alpha\alpha$
4b*	-29 A>G / cd8 (-AA);	$\alpha\alpha$ / $\alpha\alpha$
5b	HbS / HbS	$\alpha\alpha$ / $\alpha\alpha$
6b	HbS / HbC	$\alpha\alpha$ / $\alpha\alpha$
7b	Hb Lepore / IVS-II-745C>G	$\alpha\alpha$ / $\alpha\alpha$
8b	/	HBA2: c. 2T>C / $\alpha\alpha$
9b	IVS-II-745C>G / wt	/

Tabella 7. La tabella riporta il genotipo dei 9 pazienti affetti da emoglobinopatie. * Pazienti inclusi nel primo progetto di ricerca (rispettivamente pz 3 e 4).

10.2. Quantificazione del DNA genomico

Questo metodo di quantificazione prevede l'utilizzo di intercalanti fluorescenti in grado di legarsi in modo aspecifico al DNA, generando una fluorescenza rilevabile attraverso il fluorimetro Qubit®3.0 (Life technologies™, Invitrogen). Il saggio utilizzato è il dsDNA HS (high-sensitivity) Assay, rapido, altamente selettivo e accurato nella misurazione di concentrazioni di DNA comprese tra i 10 pg/μL e i 100 ng/μL e il dsDNA BR (broad-range) per la misurazione di concentrazioni di DNA comprese tra 100 pg/μL e i 1000 ng/μL.

Inizialmente è stata preparata la soluzione contenente l'intercalante fluorescente (Qubit® dsDNA HS reagent, 1 μl per ogni campione; Qubit® dsDNA BR reagent, 1 μl per ogni campione), il Qubit® dsDNA HS buffer (199 μl per ogni campione) e il Qubit® dsDNA BR buffer (199 μl per ogni campione). Da ogni campione è stato prelevato 1 μl ed è stato posto nelle rispettive Qubit® assay tubes, microtubi forniti con il kit in grado di far passare il fascio luminoso del fluorimetro e permettere in questo modo al fluoroforo legato al DNA di eccitarsi ed emettere la luminescenza ad ibridazione avvenuta. I campioni sono stati preparati nel modo seguente:

- a. 1 μl di campione + 199 μl di soluzione contenente il buffer e l'intercalante;
- b. 10 μl di Qubit® dsDNA HS standard #1 + 190 μl di soluzione contenente buffer e l'intercalante;
- c. 10 μl di Qubit® dsDNA BR standard #2 + 190 μl di soluzione contenente buffer e l'intercalante.

Prima della lettura dei campioni, il fluorimetro è stato tarato mediante i due standard precedentemente preparati e successivamente si è passati ad analizzare i campioni misurando l'ipotetica quantità di DNA presente in ciascuna libreria espressa in μg/mL.

10.3. NGS: custom bundle solution by Sophia Genetics, messa a punto del pannello

Un pannello custom con tecnologia “sonde a cattura” è stato sviluppato in collaborazione con la ditta Sophia Genetics, contenente 75 geni di cui 19 geni

GENI	NM
ATRX	NM_000489.5
BCL11A	NM_018014.3, NM_022893.3, NM_138559.1
BMP2	NM_001200.3
BMP6	NM_001718.5
ERFE	NM_001291832.1
FTL	NM_000146.3
HAMP	NM_021175.3
HBA1	NM_000558.4
HBA2	NM_000517.4
HBB	NM_000518.4
HBD	NM_000519.3
HFE	NM_000410.3, NM_139003.2, NM_139004.2, NM139006.2, NM_139007.2, NM139008.2, NM139009.2, NM_139010.2, NM139011.2, NM_001300749
HJV	NM_145277.4, NM_202004.3, NM_213652.3, NM_213653.3, NM_001316767.1
KLF1	NM_006563.4
PKLR	NM_000298.5, NM_181871.3
SCL11A	NM_000617.2, NM_001174125.1, NM_001174126.1, NM_001174127.1, NM_001174128.1, NM_001174129.1, NM_001174130.1
SLC40A1	NM_014585.5
TFR2	NM_003227.3, NM_001206855.1
TMPRSS6	NM_153609.3, NM_001289000.1, NM_001289001.1

(Tabella 8) coinvolti sia nelle emoglobinopatie che nei disordini del metabolismo del ferro e 56 legati a malattie onco-ematologiche. L'identificazione dei geni candidati è stata possibile grazie a un'attenta revisione della letteratura più recente, con lo scopo di avere uno strumento per una più completa caratterizzazione molecolare (pannello complessivo dei geni ed i rispettivi NM in appendice B).

Tabella 8. La tabella mostra i 19 geni coinvolti sia nelle emoglobinopatie che nei disordini del metabolismo del ferro con i rispettivi NM.

Il pannello è stato ideato considerando l'intera sequenza codificante di tutti i geni, le giunzioni classiche di splicing (+/- 5 bp fiancheggianti ciascun esone) ed entrambi

gli UTR; mentre per i geni globinici è stata richiesta la copertura dell'intero gene, comprendente quindi anche degli introni.

In breve, a partire da 200 ng di DNA genomico, vengono generate in 2 catture da 12 campioni l'una, 24 librerie dual-barcoded secondo il protocollo della ditta produttrice, adattato per pannelli custom, da sequenziare in modalità 2x300 bp paired-end su piattaforma Illumina MiSeq in modo tale da caricare 24 campioni su una flow cell di tipo v3.

La profondità di lettura media richiesta è di 200x.

Il processo di analisi dei dati generali segue un rigoroso flusso di lavoro bioinformatico.

- a. Analisi primaria: le reads prodotte vengono allineate ad un genoma di riferimento (il genoma attualmente utilizzato è il GCRh37 o hg19); in questa prima analisi, nella quale viene ricostruita la sequenza del genoma analizzato, le reads devono superare una serie di controlli di qualità per poter essere utilizzate;
- b. Analisi secondaria: il file generato verrà sottoposto ad un'analisi di sequenza per definire la presenza di eventuali varianti.

La sensibilità e la specificità del sequenziamento dipenderanno da parametri di analisi scelti precedentemente in base alla letteratura ed alla natura del sequenziamento.

Dal punto di vista operativo una volta terminata la corsa dello strumento, che dura più di 24 ore, si ottengono i files dei risultati in formato FASTQ. Tali files sono stati caricati e processati dal software Sophia DDM per mezzo di un algoritmo dedicato ILL1XG1G3 che è stato inoltre implementato anche per la rilevazione delle copy-number variations (CNVs). Tutte le variazioni di sequenza che superano i filtri qualità, con punteggio di qualità (Phred score) superiore a Q30, con sequencing read depth >20 per le single nucleotide variations (SNVs) e 15% per le indels and VAF (variant allele frequency) >5%, verranno utilizzate per l'elaborazione del risultato finale. In particolare l'analisi delle CNV viene eseguita analizzando i livelli di copertura di le regioni target tra i campioni all'interno dello stesso batch. Per ogni campione, l'algoritmo seleziona automaticamente una serie di campioni di riferimento della stessa corsa, in base alla somiglianza dei modelli di copertura. Inoltre, utilizzando i campioni di riferimento, la copertura è normalizzata per campione e per regione target e la chiamata delle CNVs viene eseguita utilizzando

un algoritmo hidden-Markov. Di conseguenza, il numero di copie più probabile per ciascuna regione target viene così determinata. Inoltre, viene determinato un livello di confidenza sia a livello di campione che a il livello della regione di destinazione. I campioni sono classificati in: rejected, medium-noise e low-noise, in base al rumore di copertura residuo dopo la normalizzazione e la chiamata delle CNVs. Nessun risultato CNV viene riportato per i campioni rifiutati, ma grafici della copertura del profilo per questi campioni è ancora incluso nel file prodotto dall'analisi a scopo illustrativo. Anche le singole regioni target (nei campioni non rifiutati) sono classificate in tre categorie: high-confidence, medium-confidence e undetermined. Le fasi principali dell'analisi bioinformatica comprendono filtraggi di qualità iniziali delle reads, rimozione dei duplicati di PCR, allineamento col genoma di riferimento e annotazione delle varianti. Al termine del processo di analisi, il software produce per ogni campione una lista di varianti, pre-classificate a seconda della "patogenicità" della variante stessa, corredata di informazioni sulla frequenza nella popolazione secondo i principali databases (1000Genomes, ExAc, ESP), sulla predizione *in silico* dell'effetto della mutazione sulla proteina (SIFT, Polyphen e Mutation Taster) e altri tipi di informazioni utili nell'interpretazione del dato. Tutte le variazioni con $MAF < 0.01$, quindi molto rare, vengono considerate come potenzialmente degne di nota e di sospetta patogenicità, visualizzate con il software Integrative Genomics Viewer (IGV <http://www.broadinstitute.org/igv/>) e poi validate con sequenziamento Sanger, con primers mutazione-specifici. La piattaforma oltre a fornire informazioni sulla frequenza della variante nella popolazione generale annotata nelle principali banche dati di riferimento (ESP, 1000 genomes, ExAC) verifica anche la frequenza della variante identificata in tutta la casistica delle analisi precedentemente caricate fino a fornire anche l'elenco dei pazienti in cui è stata identificata, e nella casistica di tutti gli utilizzatori della piattaforma. In funzione dall'analisi bioinformatica effettuata fornisce già una classificazione delle varianti in varianti di classe (category) A ovvero varianti ad alto significato patogeno, di classe B ovvero varianti a potenziale effetto patogeno, di classe C ovvero varianti a significato incerto infine varianti di classe D ovvero varianti verosimilmente benigne. La piattaforma permette di collegarsi direttamente sulla pagina di riferimento ExAC, IGV, ClinVar NCBI, Google, OMIM e ALAMUT.

10.4. NGS: custom bundle solution by Sophia Genetics

Per la validazione del pannello NGS sono stati analizzati 48 campioni, di cui 31 campioni con patologie onco-ematologiche e nove con alterazioni a carico dei geni globinici. In particolare 40 pazienti (indicati con UPN o con numerazione 1b-9b) sono stati analizzati una sola volta, mentre sette campioni (indicati come BIS) sono stati replicati, cioè sono stati analizzati più volte nel corso dell'esperimento al fine di validare la riproducibilità dei risultati, sia all'interno della stessa corsa (intra-run) sia in due corse distinte (inter-run); inoltre è stato inserito un l'ultimo campione costituito da un DNA di riferimento SG063, fornito dalla stessa ditta Sophia Genetics e costituito da DNA estratto da una linea cellulare avente varianti genomiche (SNPs) e variazioni del numero di copie (CNVs) note. Infine, ai 48 campioni totali, è stato aggiunto per ogni corsa, una libreria (PhiX control V3, Illumina), con la funzione di controllo di qualità del sequenziamento. L'obiettivo è stato quello di dimostrare la robustezza della metodica e, soprattutto, la riproducibilità dell'operatore, seguendo i vari step della metodica in modo estremamente preciso verificando di aver ottenuto il medesimo risultato da due stessi campioni preparati dallo stesso operatore in due momenti diversi.

La variabilità operatore-dipendente è stata valutata inserendo, nel test di validazione, due campioni inter-run, cioè il DNA dello stesso paziente replicato nella prima e nella seconda corsa.

L'algoritmo di analisi è stato elaborato rispettando alcuni valori soglia: Il limite di "low coverage", cioè di un numero insufficiente di letture per ciascuna sequenza, è stato fissato a 50x. Le regioni non analizzate in maniera idonea sono state segnalate dal software di analisi ed evidenziate con un simbolo di "warning". Il limite di bassa frequenza allelica è stato fissato nel 20% per le variazioni di singolo nucleotide (SNP) e nel 15% per le regioni di delezione/inserzione (indels). Ogni variante al di sotto di questa soglia è stata riportata come "low confidence".

Il limite di risoluzione per le inserzioni/delezioni in regioni omopolimeriche, cioè caratterizzate dalla presenza di ripetizioni della stessa base, è fissato in 10bp. Indels in regioni omopolimeriche della lunghezza superiore a 10bp potrebbero essere perse o comunque riportate dal software con un simbolo di "warning".

L'esperimento si è svolto in due settimane, nel corso delle quali abbiamo effettuato quattro diverse "catture" e due corse, alla fine delle quali abbiamo ottenuto risultati per 48 campioni.

La preparazione di ciascuna cattura avviene ogni due giorni, come da protocollo che segue (Figura 30, 31):

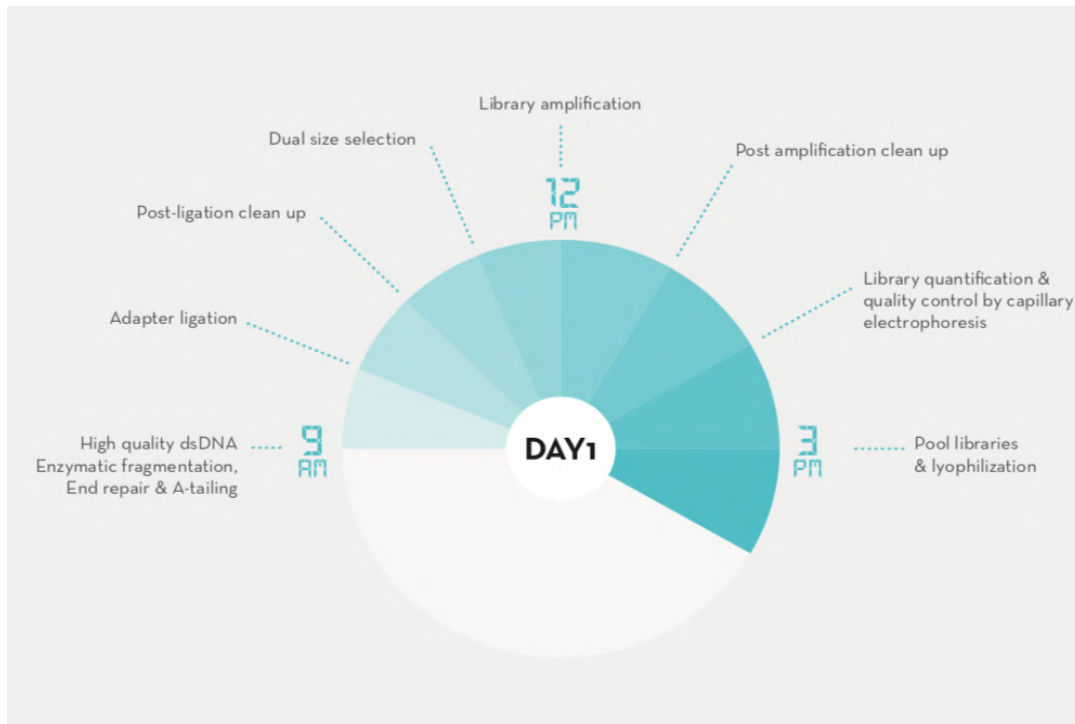


Figura 30. Suddivisione del lavoro durante il primo giorno.



Figura 31. Suddivisione del lavoro durante il secondo giorno.

a. Preparazione delle librerie:

- Preparazione dei campioni: la prima fase consiste nella preparazione dei DNA genomici (gDNA). In ciascuna cattura sono stati processati 12 campioni, per ciascuno dei quali è stato ottenuto un pool. La quantità di DNA raccomandata da protocollo per ciascun campione è compresa fra i 50 e i 100 ng/μl (la concentrazione del DNA deve essere confermata mediante una quantificazione fluorimetrica con strumento Qubit®, Thermo Fisher, e il valore ottenuto usato per calcolare la quantità finale). Un parametro molto importante è anche la purezza del DNA (lettura con spettrofotometro), i rapporti da rispettare sono rappresentati da:

- $A_{260}/A_{280} = 1.8 - 2.0$
- $A_{260}/A_{230} = 1.6 - 2.4$

Nel caso in cui le letture vengono eseguite il giorno precedente all'esperimento i gDNA devono essere posti a 4°C.

Una volta letti i campioni al Qubit, sono stati utilizzati 200 ng per ciascun DNA, eseguendo le opportune diluizioni in IDTE Low TE Buffer per portare ad un volume finale di 30 μl.

I gDNA diluiti sono stati trasferiti in delle 8-tube strips (RNase/DNase-free) e sono stati aggiunti con una pipetta multicanale 5 μl di FX Enhancer ai 30 μl dei rispettivi gDNA diluiti. Dopo aver posto i campioni in ghiaccio e si può procedere con la fase successiva di frammentazione;

- Allestimento e preparazione delle pre-mixes e dei reagenti: i componenti e i reagenti necessari, tranne gli enzimi che vanno solo invertiti e spinnati, sono stati vortexati e spinnati. Sono stati impiegati i seguenti reagenti:
 - FX Enzyme Mix;
 - FX Buffer 10x;
 - DNA Ligation Buffer 5x;
 - DNA Ligase;
 - HiFi PCR Master Mix 2x;
 - Primer Mix Illumina Library Amp;
 - Etanolo 80% fresco (40 ml or 70 ml per rispettivamente 24 o 48 reazioni);

- Acqua Nuclease-free;
- AMPure XP beads.

Sono stati preparate la FX reaction pre-mix, la Ligation pre-mix e la PCR pre-mix come riportato nelle seguenti tabelle (*Tabella 9, 10, 11*);

FX REACTION PRE-MIX	
Number of reaction	12
FX Buffer 10X	75µl
Fx Enzime Mix	150µl

Tabella 9. Schema preparazione pre-mix per FX REACTION.

LIGATION PRE MIX	
Number of reaction	12µl
DNA Ligation Buffer 5x	300µl
DNA Ligase	150µl
Water Nuclease-free	225µl

Tabella 10. Schema preparazione pre-mix per LIGATION.

PCR PRE-MIX	
Number of reaction	12µl
HiFi PCR Master Mix 2x	345µl
Primer Mix Illumina Library Amp	20.7µl
Water Nuclease-free	48.3µl

Tabella 11. Schema preparazione pre-mix per PCR.

- Frammentazione enzimatica e End Repair e A-tailing: durante questa fase è stato possibile la frammentazione enzimatica e il successivo blocco delle estremità dei frammenti precedentemente formati. Tale blocco viene garantito dall'aggiunta di un gruppo fosfato al 5' e l'aggiunta di una A al 3'. Sono stati aggiunti 15 µl di FX reaction pre-mix ai 35 µl di gDNA ad ogni campione, si passa poi a posizionare nel termociclature (Verity) ed è stato avviato il programma FXFrag Custom precedentemente creato (*Figura 32*);

	TEMPERATURE (°C)	TIME (MINUTES)
Lid	70	
Step 1	4	1
Step 2	32	5
Step 3	65	30
Step 4	4	Hold

Figura 32. Schema programma FXFrag Custom.

- Ligazione: ad ogni frammento generato si è legato un adattatore indicizzato (index), perciò occorre sempre associare adattatori-indici ai rispettivi campioni.

Utilizzando una pipetta multicanale (P10), sono stati aggiunti 5 µl di Dual Index Adapters al prodotto risultante della frammentazione secondo la strategia di indicizzazione. Utilizzando una pipetta multicanale (P100) sono stati inseriti 45 µl di Ligation pre-mix facendo attenzione che non ci siano bolle: occorre sempre miscelare bene ma lentamente.

Sono stati messi ad incubare nel termociclatore a 20 ° C per 15 minuti con coperchio aperto. Una volta terminato non occorre posizionare le 8-tube strips in ghiaccio alla fine della ligation, perché questo potrebbe diminuire il legame del DNA alle biglie;

- Clean Up Post Ligazione: questo primo step di lavaggi serve per eliminare i frammenti superiori alle 300 bp; questi, alla fine, insieme agli scarti di pcr rimangono adesi alle biglie.

Utilizzando una pipetta multicanale sono stati aggiunti 80 µl di AMPure XP beads ai 100 µl del prodotto di reazione della ligazione. Occorre mescolare accuratamente con un movimento dal basso verso l'alto in modo da non vedere più due fasi distinte. Dopo un'incubazione a temperatura ambiente di 5 minuti, le 8-tube strips sono state posizionate su un rack magnetico formato da 96 pozzetti per 5 minuti o fino a quando il liquido diventava limpido. Infine è stato scartato con cura il surnatante, usando una pipetta multicanale.

Mantenendo le provette sul rack magnetico, con una pipetta multicanale (P300), sono stati aggiunti 180 µl di etanolo 80% alle biglie e si è lasciato riposare per 1 minuto.

Dopo aver scartato con cura l'etanolo usando una pipetta multicanale; sono stati ripetuti i due passaggi precedenti ed è stato rimosso l'etanolo residuo utilizzando una pipetta multicanale P10.

Occorre poi far asciugare all'aria le biglie a temperatura ambiente per 5 minuti, non troppo perché questo potrebbe diminuire la quantità di DNA recuperato. In seguito le provette sono state rimosse dal rack magnetico e utilizzando una pipetta multicanale, sono stati aggiunti 105 µl di acqua nucleasi-free alle biglie. Si è provveduto a mescolare accuratamente ed incubare a temperatura ambiente per 5 minuti.

Dopo aver posizionato le 8-tube strips su un rack magnetico formato a 96 pozzetti per 3 minuti o fino a quando il liquido diventa nuovamente limpido, utilizzando una pipetta multicanale, sono stati trasferiti con cautela 100 µl del surnatante su una nuova 8-tube strips (occorre sempre fare attenzione a non aspirare le biglie);

- Dual Size Selection: in questa fase rimangono ancorati alle biglie i frammenti di 300-700 bp di nostro interesse; la procedura risulta simile a quella della Clean Up Post Ligazione, solo che all'inizio, utilizzando una pipetta multicanale, sono stati aggiunti 60 µl di AMPure beadsXP ai 100 µl prodotto di reazione della Ligazione e, alla fine, una volta rimosse le provette dal rack magnetico sono stati aggiunti 20 µl di IDTE alle biglie. Si può procedere alla fase dell'amplificazione;
- Amplificazione delle librerie: utilizzando una pipetta multicanale (P100), sono stati aggiunti 30 µl di PCR pre-mix al prodotto della dual size selection, eseguendo lo stesso movimento rotatorio di prima. Sono state inserite le provette nel termociclatore (Verity) ed è stato eseguito il programma LibAmp cust precedentemente impostato (*Figura 33*);

	TEMPERATURE (°C)	TIME (SECONDS)
Lid	99	
Initial Step	98	120
1st Step	98	20
2nd Step	60	30
3rd Step	72	30
Final Step	72	60

} 8 Cycles

Figura 33. Schema programma LibAmp cust.

- Clean Up Post Amplificazione: in questa fase, come nel Clean Up Post Ligazione si effettua una pulizia dai residui di reazione di pcr. All'inizio sono stati aggiunti 50 µl di AMPure XP beads ai prodotti di PCR, e alla fine, 20 µl di acqua nucleasi-free, e una volta rimosse le provette dal rack magnetico, sono stati trasferiti con cautela 18 µl del surnatante in un nuovo library storage tube precedentemente etichettato;
- Controllo quantitativo e qualitativo delle singole librerie: prima di procedere ad un controllo quantitativo delle librerie è stata eseguita una diluizione 1:4 di ciascuna di esse con acqua nucleasi-free. Si può così procedere con la quantificazione delle rispettive librerie con un metodo fluorimetrico (ad esempio quantificazione Qubit HS utilizzando 2 µl di campione diluito).

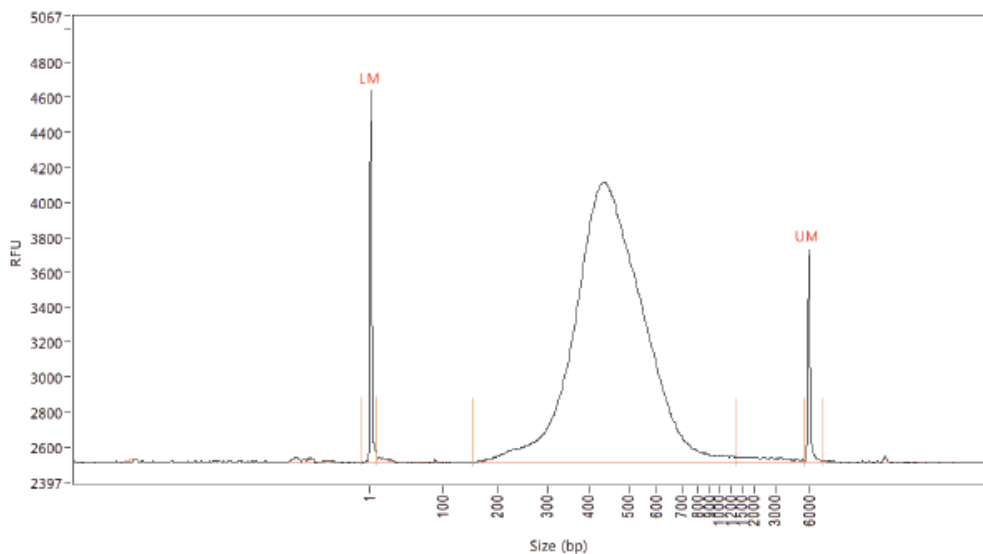
Per il controllo qualitativo delle librerie va analizzato il loro profilo tramite elettroforesi capillare (utilizzo della TapeStation, dove facciamo correre i campioni diluiti). I frammenti di DNA di ciascuna libreria dovrebbero avere una distribuzione delle dimensioni tra 300 / 700 bp (*Figura 34*).

Figura 34. Esempio di distribuzione della libreria DNA ottenuta con AATI Fragment Analyzer.

- b. Cattura (capture): pooling delle Librerie. Sono state unite le 12 librerie individuali utilizzando 150 ng di ciascuna (totale di 1800 ng) per la capture in una DNA low-binding tube. A questo pool sono stati aggiunti 2 μ l di Blocking oligos xGen Universal Blockers - TS Mix e 5 μ l di Human Cot DNA.

Si è provveduto ad essiccare il pool usando vacuum DNA concentrator a circa 43°C, finché la miscela non è completamente liofilizzata.

- Ibridazione: questo passaggio è cruciale, il pool viene ibridato con sonde biotinilate, che sfruttando il legame con la streptavidina garantiscono la specificità di “cattura”. È stato risospeso il pellet liofilizzato in 13 μ l della miscela di ibridazione (passaggio da fare sotto cappa), composta da 8.5 μ l



di 2x Hybridization Buffer, 3.4 μ l di Hybridization Buffer Enhancer e 1.1 μ l di acqua nuclease-free. Trasferire il tutto in un pcr tube da 0,2 ml. È stato programmato il termociclatore (Verity) con il programma Hib custom (10 minuti di denaturazione per passare direttamente a 65 °C all'infinito) ed è stato fatto partire un po' prima per preriscaldarlo. Occorre poi incubare la mix nel termociclatore a 95 °C per 10 minuti. Non appena la temperatura è scesa a 70°C sono stati aggiunti con una P10 4 μ l di sonde alla mix ed il tutto è

stato miscelato accuratamente. Occorre prestare attenzione che la temperatura del tubo non scenda al di sotto di 65 ° C, poiché ciò può comportare un non corretto annealing delle sonde. È stato lasciato overnight ad incubare nel termociclatore a 65 ° C (4 ore minimo). Prima di concludere il primo giorno di esperimento sono stati preparati i reagenti, per il giorno seguente, come indicato in *Tabella 12*;

BUFFER	b) Stock buffer	a) Water	Volume Finale
10x Wash Buffer I	33µl	297µl	330µl
10x Wash Buffer III	22µl	198µl	220µl
10x Wash Buffer II	22µl	198µl	220µl
10x Stringent Wash Buffer	44µl	396µl	440µl
2x Bead Wash Buffer	275µl	275µl	550µl

Tabella 12. Schema preparazione reagenti.

- Preparazione Biglie con Streptavidina: inizialmente bisogna procedere con una fase di equilibratura delle biglie, facendo dei lavaggi con opportuni buffer. Sono stati trasferiti 100 µl di biglie per cattura in una singola provetta da 1,5 ml, poi la provetta è stata posizionata su un rack magnetico ed è stata lasciata riposare fino a quando la soluzione diventava chiara. È stato rimosso e scartato con attenzione il sopranatante, per poi aggiungere 200 µl di 1x Bead Wash Buffer per cattura alla provetta, riposizionandola di nuovo sul rack magnetico per poi ripetere il precedente passaggio 2 volte. Sono stati trasferiti 100 µl delle biglie pulite in un nuovo pcr tube da 0,2 ml, e poi sono state posizionate le provette su un rack magnetico formato 96 pozzetti e lasciato riposare fino a che la soluzione diventava limpida. È stato rimosso con attenzione il sopranatante, occorre fare questo passaggio vicino al termociclatore in funzione. Non bisogna lasciare asciugare le biglie e procedere immediatamente al legame dei targets ibridati alle biglie con streptavidina;
- Legame dei targets ibridati con le biglie con streptavidina: in questo passaggio si lega in maniera specifica il DNA ibridato alle sonde. Importante è lavorare rapidamente per garantire che la temperatura rimanga vicina a 65°C; occorre trasferire 17 µl della soluzione di reazione di ibridazione nella

provetta dove vi sono le biglie pulite e sono state risospese le biglie fino a che la soluzione diventava omogenea.

Il Legame del DNA alle biglie avverrà inserendo la provetta nel termociclatore impostato a 65 °C e lasciandolo in incubazione per 45 minuti;

- Lavaggio delle biglie con streptavidina, per rimuovere DNA non legato: il Wash Streptavidin Beads rimuove il DNA non legato alle biglie. Quando mancavano 10 minuti al termine dei 45 minuti, sono stati spinnati i regenti riposti precedentemente a 65°C nel termo blocco; occorre sempre lavorare rapidamente per garantire che la temperatura rimanga il più vicina possibile a 65 ° C nelle fasi successive. Sono stati aggiunti 100 µl di 1x Wash Buffer I (a 65 ° C) al tubo con i target / biglie streptavidina ibridati, per poi trasferire il materiale in una nuova provetta che poi è stata collocata sul rack magnetico finché la soluzione non diventava limpida.

Dopo aver rimosso il sopranatante sono stati aggiunti 200 µl di 1x Stringent Wash Buffer (a 65 ° C) e si è risospeso delicatamente. Dopo un'incubazione a 65°C per 5 minuti, la provetta è stata riposizionata sul rack magnetico e si è ripetuto il precedente passaggio 2 volte. Nelle fasi successive si è dovuto lavorare a temperatura ambiente. Sono stati aggiunti 200 µl di 1x Wash Buffer I, di 1x Wash Buffer II, 1x Wash Buffer III e 1x IDTE. Questi vanno aggiunti uno alla volta intervallati dalla solita procedura di incubazione e scarto del sopranatante su rack magnetico. Terminati questi lavaggi è fondamentale rimuovere tutto l'IDTE rimanente e sono stati aggiunti infine 20 µl di acqua nucleasi-free per risospendere e poi trasferire la mix biglie e acqua in una nuova strip per PCR da 0.2 µl;

- Amplificazione Post Cattura: è stata preparata la pcr pre-mix come segue (*Tabella 13*), poi riposta in ghiaccio:

PCR PRE-MIX	
Number of reaction	1
2x HiFi HotStart Ready Mix	25µl
10x Library Amplification Primer Mix	2.5µl
Water. Nuclease-free	2.5µl

Tabella 13. Schema preparazione PCR pre-mix.

Sono stati aggiunti 30 µl di PCR pre-mix nella provetta Pool da 0.2 con la sospensione delle biglie e si è provveduto alla risospensione. Dopo aver posizionato le provette nel termociclatore (Verity), è stato eseguito il programma PostCapAmp Custom precedentemente creato (*Figura 35*);

	TEMPERATURE (°C)	TIME (SECONDS)
Lid	99	
Initial Step	98	45
1st Step	98	15
2nd Step	60	30
3rd Step	72	30
Final Step	72	60

} X Cycles (please see Appendix 4 for details)

Figura 35. Schema programma PostCapAmp Custom.

Al termine dei 14 cicli tenere le reazioni a 10 ° C.

- Post-capture Amplification Clean Up: questa fase, come nelle altre fasi di lavaggi post-amplificazione, si effettua una pulizia dai residui di reazione di pcr. All’inizio sono stati aggiunti 50 µl di AMPure XP beads ai prodotti di PCR, e alla fine, 20 µl di IDTE alle biglie, e una volta rimosse le provette dal rack magnetico, sono stati trasferiti con cautela 18 µl del supernatante in un nuovo library storage tube etichettato;
- Controllo qualitativo e quantitativo del pool delle librerie: è stata eseguita la quantificazione del pool di librerie con metodo fluorimetrico (Qubit HS) usando 2 µl della libreria; inoltre è stata eseguita un’analisi qualitativa del pool di librerie analizzandone il profilo tramite elettroforesi capillare (TapeStation).

I frammenti di DNA del pool della libreria dovrebbero avere una distribuzione delle dimensioni tra 300 e 700bp.

Nel caso il pool venisse congelato, una volta scongelato andrà riquantificato. Abbiamo ripetuto la metodica di questi primi due giorni nei seguenti due successivi con altri 12 campioni; ottenendo così 24 campioni finali per la corsa.

- c. Sequenziamento: preparazione delle librerie per il sequenziamento: è stata determinata la molarità di ciascun pool con la dimensione media della libreria (dimensione del picco in bp) e la concentrazione (ng/ul) ottenuta durante la fase precedente utilizzando la seguente formula:

$$\text{Library molarity (nM)} = \frac{\text{Library concentration (ng/}\mu\text{l)}}{\text{Average size in base pairs} \times 649.5} \times 10^6$$

Ciascun pool è stato diluito a 4 nM, per poi essere mescolati con cura in una quantità uguale (ad esempio 5 μl di ciascuno), e poter quindi utilizzare tale diluizione come riportato nelle linee guida dello strumento ILLUMINA. Eseguire opportuni passaggi per poi caricare infine una diluizione 10 pM delle librerie denaturate;

- d. Corsa delle librerie: per la corsa delle librerie è stata utilizzata la piattaforma Illumina Miseq, la quale supporta una vasta gamma di applicazioni, esegue la generazione di cluster, sequenziamento per sintesi chimica e l'analisi dei dati. Miseq è una piattaforma da bancone che può produrre fino a 15 gigabasi di dati in un'unica corsa, 25 milioni di letture (reads) con una lunghezza di 2x300 bp ciascuna, e parametri di qualità eccezionali (>Q30), garantendo un elevato grado di precisione a livello di sequenza.
- La seconda settimana abbiamo di fatto ripetuto il protocollo precedente ottenendo così, al termine del secondo e del quarto giorno altri 2 pool contenenti rispettivamente 12 campioni ciascuno. Il quarto giorno della seconda settimana abbiamo ripetuto una seconda corsa con Illumina MiSeq ottenendo così dati per 48 campioni.

11. RISULTATI DEL PRIMO STUDIO

11.1. Analisi dei SNPs modulatori del fenotipo clinico e distribuzione geografica

Al fine di caratterizzare e stratificare meglio i pazienti affetti da emoglobinopatie in esame sono stati analizzati cinque loci polimorfici legati alla modulazione dell'espressione dell'emoglobina fetale, ovvero *HBG2* rs7482144 [C> T], *BCL11A* rs1427407 [G> T], *BCL11A* rs10189857 [A> G], *HBS1L-MYB* rs28384513 [A> C], *HBS1L-MYB* rs939913 [T> C] e un altro modulatore, ovvero *C/EBPE* rs45496295 [C> T].

Da tale studio il sito polimorfico *BCL11A* rs10189857 [A> G] è risultato più frequente, poiché l'allele G è stato rilevato in 33/53 pazienti (21 eterozigoti, 12 omozigoti). Al contrario, *HBS1L-MYB* rs9399137 [T> C] è stato la variazione polimorfica meno frequente, in quanto l'allele C era presente solo in 6 pazienti su 53 (un omozigote e cinque eterozigoti). In riferimento agli altri SNPs analizzati: *HBG2* rs7482144 [C> T], *BCL11A* rs1427407 [G> T] e *HBS1L-MYB* rs28384513 [A> C] sono stati rilevati rispettivamente nei pazienti 9, 19 e 21, mostrando una frequenza intermedia (*Tabella 14*, *Tabella 15*). Non sono stati trovati pazienti portatori della variazione *C/EBPE* rs45496295 [C> T] (*Tabella 14*). Una descrizione dettagliata delle alterazioni trovate nei singoli sottogruppi è riportata nell'appendice C.

È stata eseguita una correlazione tra le frequenze dei sei SNPs analizzati e l'origine geografica dei pazienti (*Figura 36* e *Figura 37 A-C*). In particolare, *HBG2* rs7482144 [C> T] è stato trovato in nove pazienti, cinque africani, tre americani e uno asiatico (*Figura 37 A*, *Tabella 14*). Mentre, l'*HBS1L-MYB* rs9399137 [T> C], è stato rilevato solo in sei dei 53 pazienti analizzati, cinque europei e un asiatico (*Figure 37 A e B*, *Tabella 14*). Non sono stati trovati pazienti africani con *HBS1L-MYB* rs9399137 [T> C] nella nostra coorte sebbene questo rappresentasse il gruppo più numeroso. Le frequenze alleliche ottenute dal database GnomAD riportavano alleli C nelle popolazioni americane e africane, sebbene a frequenze inferiori (*Figure 37 B e C*).

	Numero	Sesso-Età ^a	Origine etnica		HBG2	BCL11A		HBS1L-MYB	HBS1L-MYB	C/EBPE
			Continente	Paese	rs7482144 [C>T]	rs1427407 [G>T]	rs10189857 [A>G]	rs28384513 [A>C] ^b	rs9399137 [T>C]	rs45496295 [C>T]
βTalassemia (β ⁺ /β ⁺ , β ⁰ /β ⁰ o β ⁺ /β ⁰)	1	M	Non disponibile	Non disponibile	CC	GG	GG	AA	TT	CC
	2	M-18	Europa	Albania	CC	GG	AA	AC	TC	CC
	3*	M-3	Africa	Algeria	CT	GG	AG	AA	TT	CC
	4*	F-4	Africa	Algeria	CT	GG	AG	AA	TT	CC
	5	M	Europa	Italia	CC	GG	AA	AA	TT	CC
	6	M-1	Europa	Albania	CC	GG	AG	AC	TC	CC
	7*	M-15	America	Venezuela	CT	GG	GG	AA	TT	CC
	8	M-18	Europa	Albania	CC	GG	AG	AC	TT	CC
	9	F-62	Europa	Italia	CC	GG	GG	AC	TT	CC
	10	F-31	Europa	Italia	CC	GG	AG	AC	TC	CC
	11*	M-3	Asia	Siria	CC	GT	AG	AA	TT	CC
	12*	M-4	Asia	Siria	CC	GG	GG	AA	TT	CC
	13	F-19	Asia	Cina	CC	GG	GG	AC	TT	CC
	14	M-29	Europa	Italia	CC	GG	AA	AA	TT	CC
	15	M-10	Asia	Cambogia	TT	GG	GG	CC	TC	CC
Microrepanocitosi (HbS/β ⁺ o β ⁰)	16*	F-18	Europa	Albania	CC	GG	GG	AC	TT	CC
	17*	M-11	Europa	Albania	CC	GG	GG	AA	TT	CC
	18	F-5	Europa	Albania	CC	GG	AG	AC	CC	CC

	19	M-6	Africa	Costa d'Avorio	CC	GG	AG	AC	TT	CC
	20	M-46	Europa	Albania	CC	GG	AG	AC	TT	CC
	21*	F-52	America	Venezuela	CT	GT	AG	AA	TT	CC
	22	M-51	Europa	Italia(Sicilia)	CC	GT	AA	AA	TT	CC
	23	M-2	Europa	Albania	CC	GG	AG	AA	TT	CC
	24	M-19	Africa	Guinea	TT	GT	AA	AA	TT	CC
	25	M-24	Europa	Albania	CC	GT	AA	AC	TC	CC
Drepanocitosi (HbS/HbS)	26	M-7	Europa	Albania	CC	GT	AG	AA	TT	CC
	27	M-32	America	Repubblica Dominicana	CC	GG	GG	AA	TT	CC
	28	M-4	Africa	Nigeria	CC	GT	AA	AA	TT	CC
	29	M-1	Africa	Nigeria	CC	GG	AA	AA	TT	CC
	30	M-4	Africa	Nigeria-Costa d'Avorio	CC	GT	AA	AA	TT	CC
	31	M-2	Africa	Nigeria-Costa d'Avorio	CC	GG	AG	AA	TT	CC
	32	M-5	Africa	Nigeria	CC	GT	AA	AA	TT	CC
	33	M-2	Africa	Nigeria	CC	GT	AA	AA	TT	CC
	34	M-1	Africa	Nigeria	CC	GT	AG	AC	TT	CC
	35	M-49	Europa	Italia	CC	GT	AA	AA	TT	CC
	36	F-5	Africa	Nigeria	CC	GG	AG	AC	TT	CC
	37	F-1	Africa	Nigeria	CC	TT	AA	AC	TT	CC
	38	M-14	Africa	Congo	CC	GG	AA	AC	TT	CC
	39	M-1	Africa	Benin	CT	GG	AA	AA	TT	CC
	40	M-12	Africa	Senegal	CT	GG	AA	AA	TT	CC
	41	M-11	America	Venezuela	CT	GG	GG	AA	TT	CC
	42	M-60	Africa	Nigeria	CC	GG	GG	AC	TT	CC
	43	F-1	Africa	Nigeria	CC	GG	AA	AC	TT	CC
	44	M-10 days	Africa	Congo	CC	GT	AG	AC	TT	CC

	45	M-2	Africa	Nigeria	CC	GG	AA	AA	TT	CC
Eterozigoti Composti (HbS/HbC o HbO-Arab/HbC)	46	M-38	Africa	Costa d'Avorio	CC	GT	AA	AA	TT	CC
	47*	M-2	Africa	Costa d'Avorio	CC	GT	AG	AA	TT	CC
	48	F-1	Africa	Ghana	CC	GG	GG	AC	TT	CC
	49*	F-14	Africa	Costa d'Avorio	CC	GT	AG	AA	TT	CC
	50*	F-16	Africa	Costa d'Avorio	CC	TT	AA	AA	TT	CC
	51	F-24	Africa	Nigeria	CC	GT	AG	AA	TT	CC
	52	F-59	Africa	Marocco	CC	GG	AG	AC	TT	CC
	53	M-17	Africa	Marocco	CC	GG	AG	AA	TT	CC

Tabella 14. Analisi molecolare degli SNP che modulano il fenotipo e distribuzione geografica dei 53 pazienti analizzati in questo studio. ^a Età al momento dei test genetici; ^b Solo SNP che riducono l'HbF; * Pazienti correlati

Modificatori genetici	N° pz	β -Talassemia	Microdrepanocitosi (HbS/ β^+ o β^0 tal)	Drepanocitosi (HbS/HbS)	Eterozigoti Composti
Pazienti, n (%)	53 (100.0)	15	10	20	8
Genotipo α , n (%)					
$\alpha\alpha/\alpha\alpha$	43 (81)	15	8	14	6
- $\alpha^{3.7}/\alpha\alpha$	7 (13)	0	0	5	2
- $\alpha^{3.7}/\alpha^{3.7}$	1 (2)	0	0	1	0
$\alpha\alpha^{\text{anti}3.7}$	2 (4)	0	2	0	0
<i>HBG2</i> , rs7482144 (T), n (%)	9 (17)	4	2	3	0
<i>BCL11A</i> rs10189857, (G), n (%)	33 (62)	12	7	8	6
rs1427407, (T), n (%)	19 (36)	1	4	9	5
<i>HBS1L-MYB</i> rs9399137, (C), n (%)	6 (11)	4	2	0	0
rs28384513, (C), n (%)	21 (40)	7	5	7	2

Tabella 15. Caratteristiche molecolari dei modulatori dei fenotipi clinici in pazienti omozigoti ed eterozigoti composti per difetti a carico del gene β .

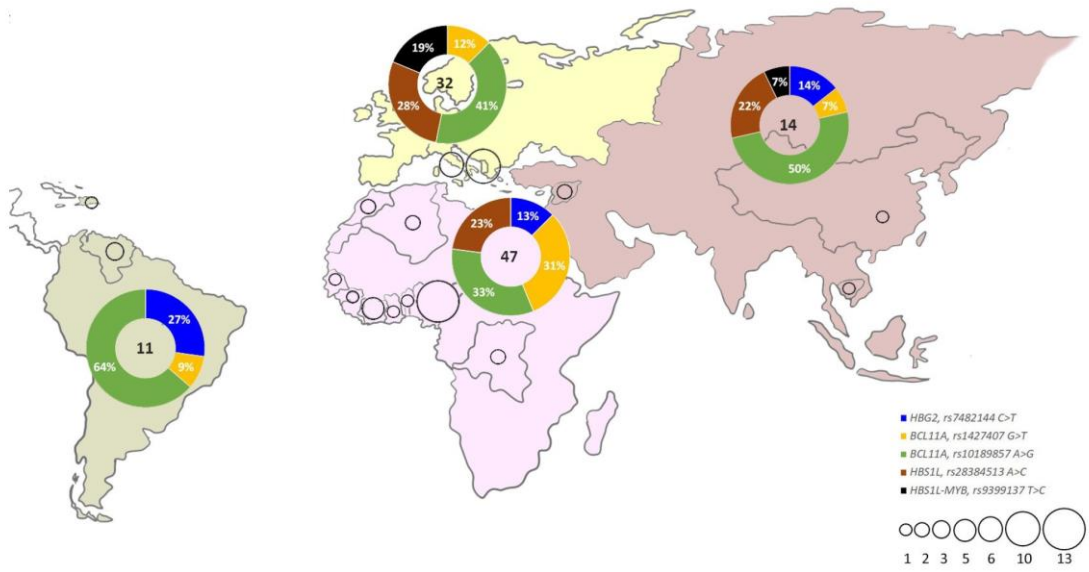
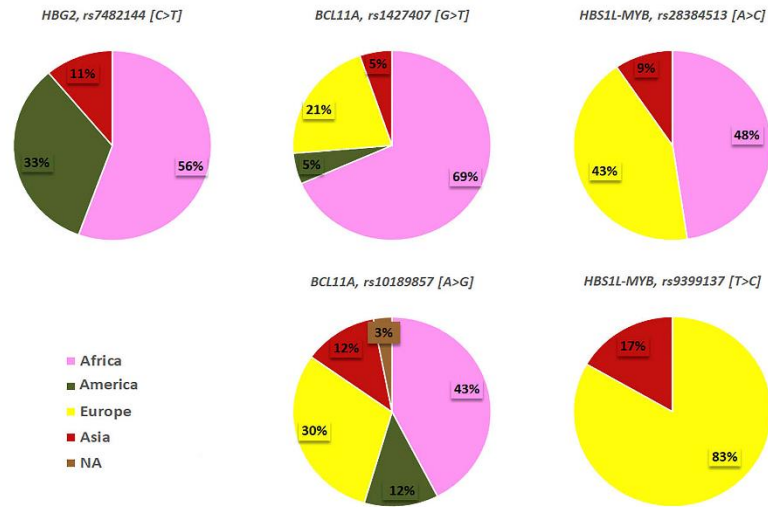
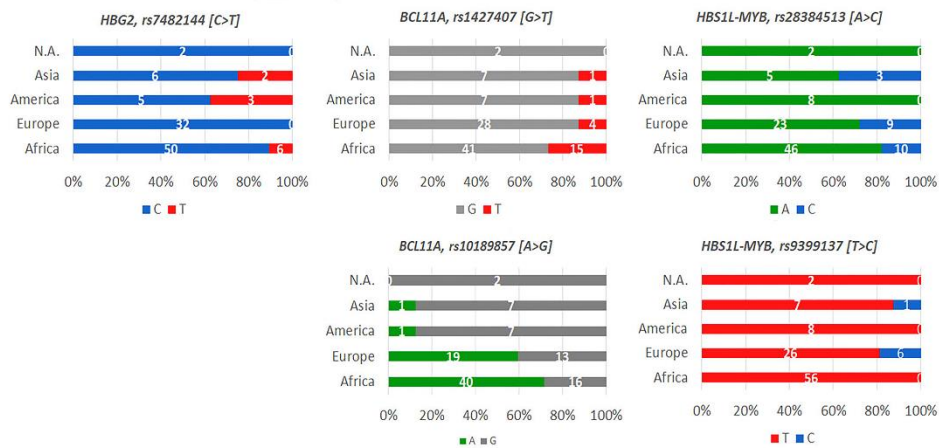


Figura 36. Frequenze e distribuzione geografica dei quattro SNPs nucleari nei 53 pazienti. Gli anelli neri rappresentano il numero di pazienti in diverse aree geografiche. I grafici a torta rappresentano le frequenze di ogni SNP. Il numero centrale indica gli alleli totali che trasportano l'alterazione [187].

A



B



C

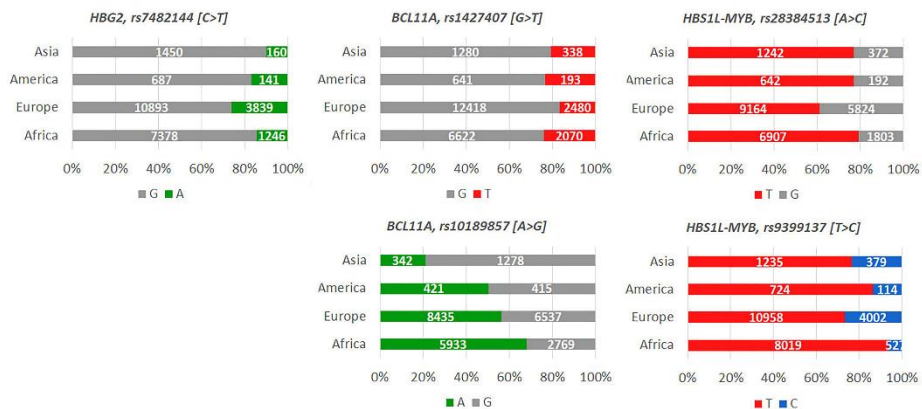


Figura 37. Frequenze degli 5 SNPs nucleari modulatori. (A) Frequenza degli SNPs dei pazienti analizzati in base all'origine geografica. (B) Frequenza per continente dell'allele ancestrale e di quello variante nella coorte di questo studio. (C) Frequenza per continente dell'allele ancestrale e di quello variante riportato dal database GnomAD. Si noti che i pannelli A e B si riferiscono alla sequenza di riferimento del DNA codificante mentre il pannello C si riferisce alla sequenza di riferimento genomica [187].

11.2. Analisi delle variazioni della sequenza di mtDNA e correlazione con le varianti dei geni modificatori

E' stata eseguita l'analisi delle sequenze della regione di controllo mitocondriale dal nucleotide 16024 al 210, comprendendo così la regione ipervariabile I e la maggior parte della regione ipervariabile II (*Tabella 14*). L' allineamento delle sequenze dei rispettivi campioni alla sequenza di riferimento Cambridge ha rivelato 99 siti polimorfici, rappresentati da dieci trasversioni e due indel (una inserzione e una delezione). È stata stimata la diversità dei nucleotidi (π) in tutti i campioni pari a 0,0173. La diversità degli aplotipi era molto alta ($H_d = 0,993$) così come la media numero di differenze nucleotidiche ($k = 13,052$).

Dall'allineamento alla sequenza di riferimento sono stati ottenuti un totale di 44 aplotipi diversi. I quali sono stati poi classificati attraverso i loro motivi mutazionali in 13 diversi macroaplogruppi: B, D, G, JT, L0, L1, L2, L3, M, N9, R0, UK e X (*Figura 38 A*). Sette aplotipi sono stati condivisi tra due o tre campioni, ma cinque coppie di questi individui hanno dichiarato di esserlo parenti materni (pz 3-4, 7 e 21, 11-12, 16-17, 47 e 49-50). Per i restanti due aplotipi, uno (HT02) è stato trovato in due fratelli dall'Algeria e un campione dal Marocco, l'altro (HT04) è stato rilevato in un paziente dalla Costa d'Avorio e uno dal Benin.

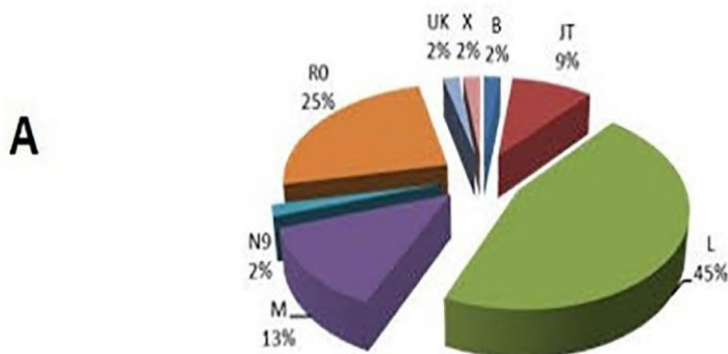


Figura 38 A. Grafico a torta delle frequenze dei macroaplogruppi [187].

I diversi aplogruppi sono stati correlati all'origine etnica dichiarata da ciascun paziente (*Figura 38 B*).

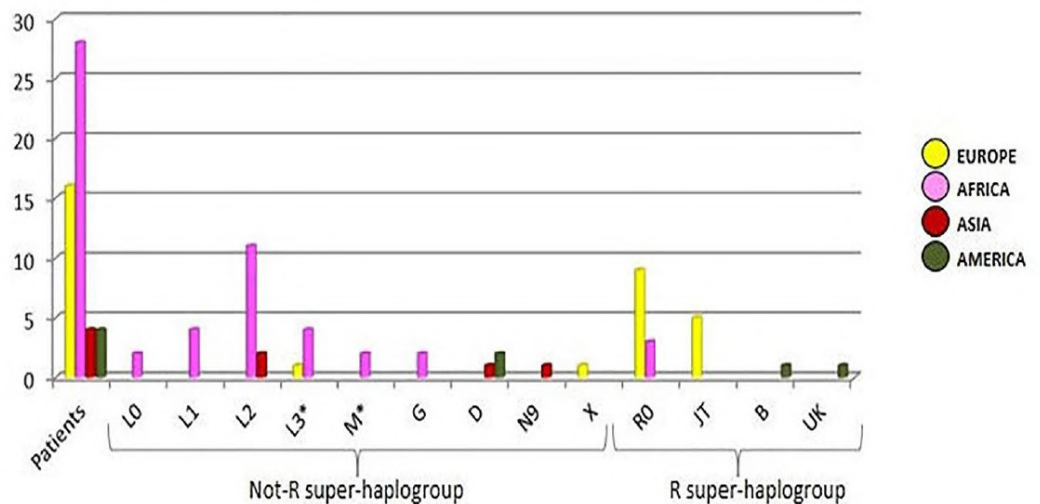
B

Figura 2B. Distribuzione geografica dei diversi macroaplogruppi [187].

Inoltre nel grafico riportato in *Figura 38 C* è possibile osservare la distribuzione dei diversi aplogruppi del mtDNA nei 4 sottogruppi di emoglobinopatie.

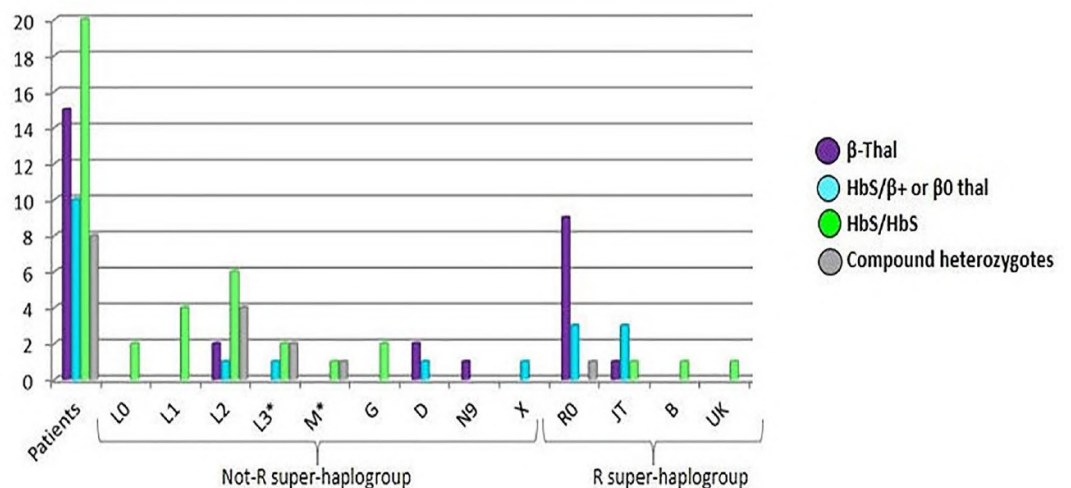
C

Figura 38 C. Distribuzione degli aplogruppi nei sottogruppi di emoglobinopatie [187].

Per meglio definire le relazioni filogenetiche tra i campioni, è stato costruito un network illustrato in *Figura 39*. Ciascun aplogruppo è stato riportato secondo la nomenclatura Phylotree [177]. Da tale analisi sono emersi due principali cluster di mtDNA, corrispondenti ai superaplogruppi R e non-R. All'interno di R, prevalevano gli aplogruppi JT e H. Quest'ultimo formava un ammasso stellare rappresentato principalmente dai sotto-aplogruppi H1, H2 e H7. Erano rappresentati anche R0a, HV, B e UK. Il cluster non-R includeva principalmente L1, L2 e L3 sub-lignaggi, seguiti da L0 e D; sono stati rivelati anche N9, M, X e G. Tra i campioni, il lignaggio più rappresentato apparteneva al super-aplogruppo L (45%), composto da L0 (8%),

L1 (17%), L2 (54%) e L3 (21%), seguito da R0 (25%) e M (13%) (Figura 38 B). L'aplogruppo JT è stato rilevato con il 9% di frequenza, mentre B, N9, UK e X i lignaggi erano rappresentati ciascuno da un solo campione.

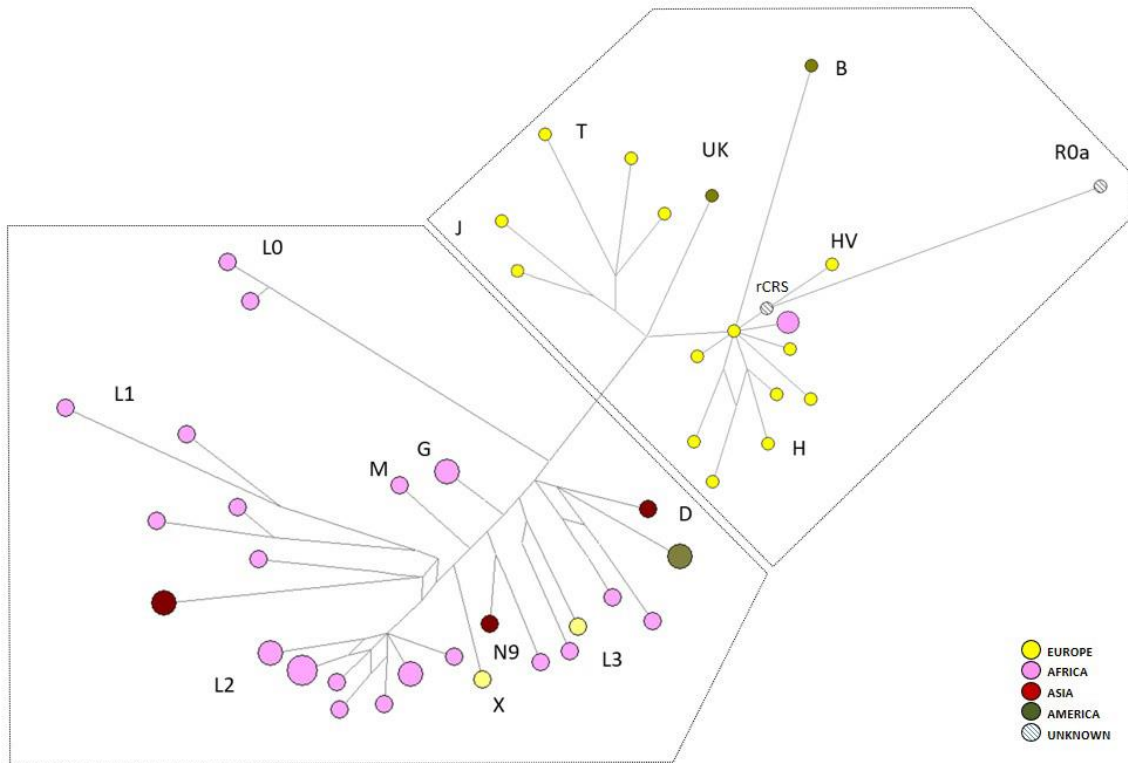


Figura 39. Network delle variazioni mitocondriali. Albero filogenetico dei 53 aplotipi di mtDNA osservati tra i pazienti, evidenziati con colori differenti a seconda della loro provenienza geografica. I cerchi sono proporzionali al numero di campioni che trasportano lo stesso aplotipo. Aplogruppi e sotto-aplogruppi sono indicati con lettere maiuscole. I super aplogruppi R e non-R sono definiti rispettivamente nella casella destra e sinistra [187].

Dopo aver valutato le relazioni tra la filogenesi mitocondriale e l'origine geografica degli individui, si è passati a un diverso livello di risoluzione, quindi sono stati inclusi in un nuovo network anche i dati degli SNPs nucleari (Figura 40, Tabella 14).

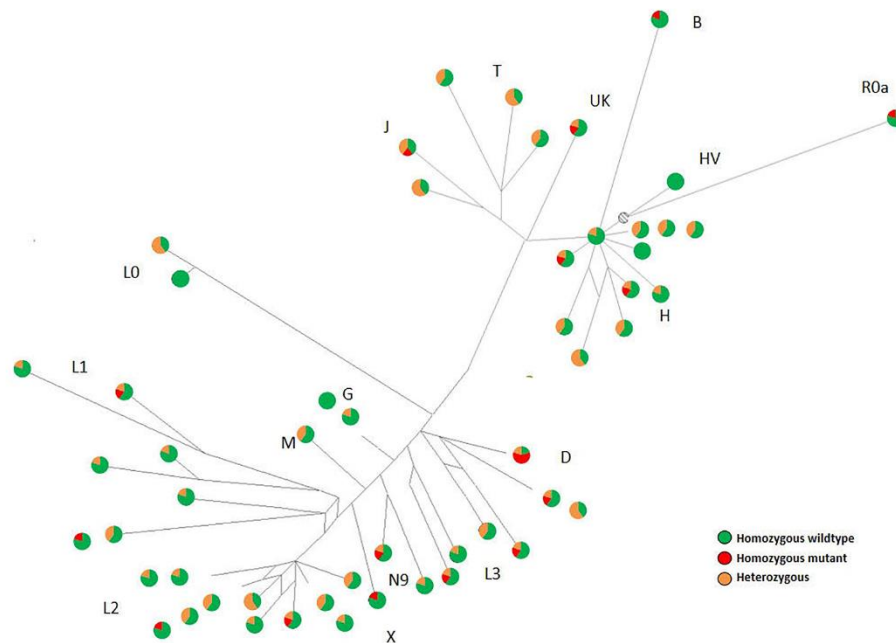


Figura 40. Network dei mtDNA e SNPs nucleari. Albero filogenetico dei 53 aplotipi di mtDNA osservati tra i pazienti in cui è stata evidenziata manualmente la zigosità di SNP (*rs45496295* è stato escluso perché i campioni erano tutti omozigoti wildtype per esso) [187].

In tale analisi sono stati considerati gli aplotipi della regione di controllo mitocondriale, le cui relazioni filogenetiche sono illustrate nella *Figura 39*, insieme alla zigosità dei cinque geni variabili, cercando di dettagliare il genotipo dei pazienti. I pazienti 5, 14, 29 e 45 erano i soli wildtype omozigote per tutte le variabili: due campioni provenienti dall'Italia (pz 5 e 14 entrambi appartenenti all'aplogruppo R0) e due dalla Nigeria (pz 29 e 45 che trasportano mitogenomi L0 e G). Un totale di 41/53 pazienti hanno mostrato la prevalenza dell'omozigosi wildtype, mentre solo sei individui (pz 6, 10, 21, 25, 34 e 44) erano eterozigoti per tre polimorfismi. Quattordici pazienti erano omozigoti mutati per un SNP, mentre solo il paziente 15 era omozigote mutato per tre variabili. Inoltre tutti gli individui all'interno del macroaplogruppo mitocondriale L erano wildtype omozigote per *HBS1L-MYB* rs9399137 [T > C], mentre i pazienti che presentavano il lignaggio mitocondriale JT erano omozigoti wildtype per *HBG2* rs7482144 [C > T]. Al contrario, *C/EBPE* rs45496295 [C > T] è sempre stato trovato come omozigote wildtype, quindi è stato escluso dalla rappresentazione.

È stata eseguita un'ulteriore analisi, cercando di ottenere una rappresentazione grafica della possibile integrazione tra la variabilità della regione di controllo del mtDNA e le mutazioni nucleari associati ai quattro sottogruppi di emoglobinopatie

(*Tabella 14 e Tabella 15*) mostrata nelle *Figure 41 e 42*. Il network ha rivelato una predominante distribuzione di determinati sottogruppi di emoglobinopatie all'interno delle diverse cladi (*Figure 41 e 42*). In particolare, tranne per tre aplotipi appartenenti a campioni provenienti da Albania (pz 26) e Sud America (pz 27 e 41) che si trovavano nel clade R, l'anemia falciforme sembrava essere comune tra gli aplogruppi non R, in particolare lignaggi africani (L0, L1, L2 e L3). È da notare che non sono presenti pazienti con drepanocitosi appartenenti al macro-aplogruppo R0 (inclusi gli aplogruppi R0a, H e HV) (*Figura 38 C*). Una distribuzione simile è stata osservata per i pazienti eterozigoti composti, tutti gli africani; appartenenti in clade non-R (aplogruppi M1, L2 e L3), ad eccezione del paziente 52, da Marocco, che apparteneva alla stirpe H1 ed era l'unico paziente caratterizzato dal genotipo HbC /HbO-Arab. Una distribuzione opposta è stata osservata per i pazienti talassemici, che si sono diffusi tra gli aplogruppi R, ad eccezione di quattro campioni dal Sud America (pz 7), Siria (pz 11 e 12, fratelli) e Cambogia (pz 15). In particolare, il clade R0 sembrava essere complessivamente caratterizzato da individui β -talassemici. D'altra parte, i pazienti con microdrepanocitosi (HbS/ β^+ o β^0) sembravano essere diffusi nel network senza particolari raggruppamenti.

Quando sono stati separati i dati in quattro network diversi, uno per ogni sottogruppo di emoglobinopatie e dopo aver evidenziato in diversi colori l'origine geografica di ciascun paziente, è risultato chiaro che i campioni asiatici fossero tutti talassemici, ma raggruppati lontani dai pazienti talassemici europei (*Figura 41E*). Nessun paziente africano era affetto da talassemia (ad eccezione di una coppia di fratelli algerini). Se esclusi due campioni con microdrepanocitosi (HbS / β^+ o β^0) (*Figura 41 F*) i soggetti africani prevalgono nel gruppo della drepanocitosi (*Figura 41 G*) e sono anche l'unica fonte degli eterozigoti composti (*Figura 41 H*). Non è emerso alcun modello di correlazione molecolare simile a una stella che potesse rappresentare qualsiasi evento di espansione locale. Quindi è stato osservato che i pazienti europei cadono prevalentemente nel gruppo dei talassemici o con microdrepanocitosi, mentre gli africani nei rimanenti due sottogruppi (ovvero, HbS/HbS e HbS/ HbC) (*Figura 42*).

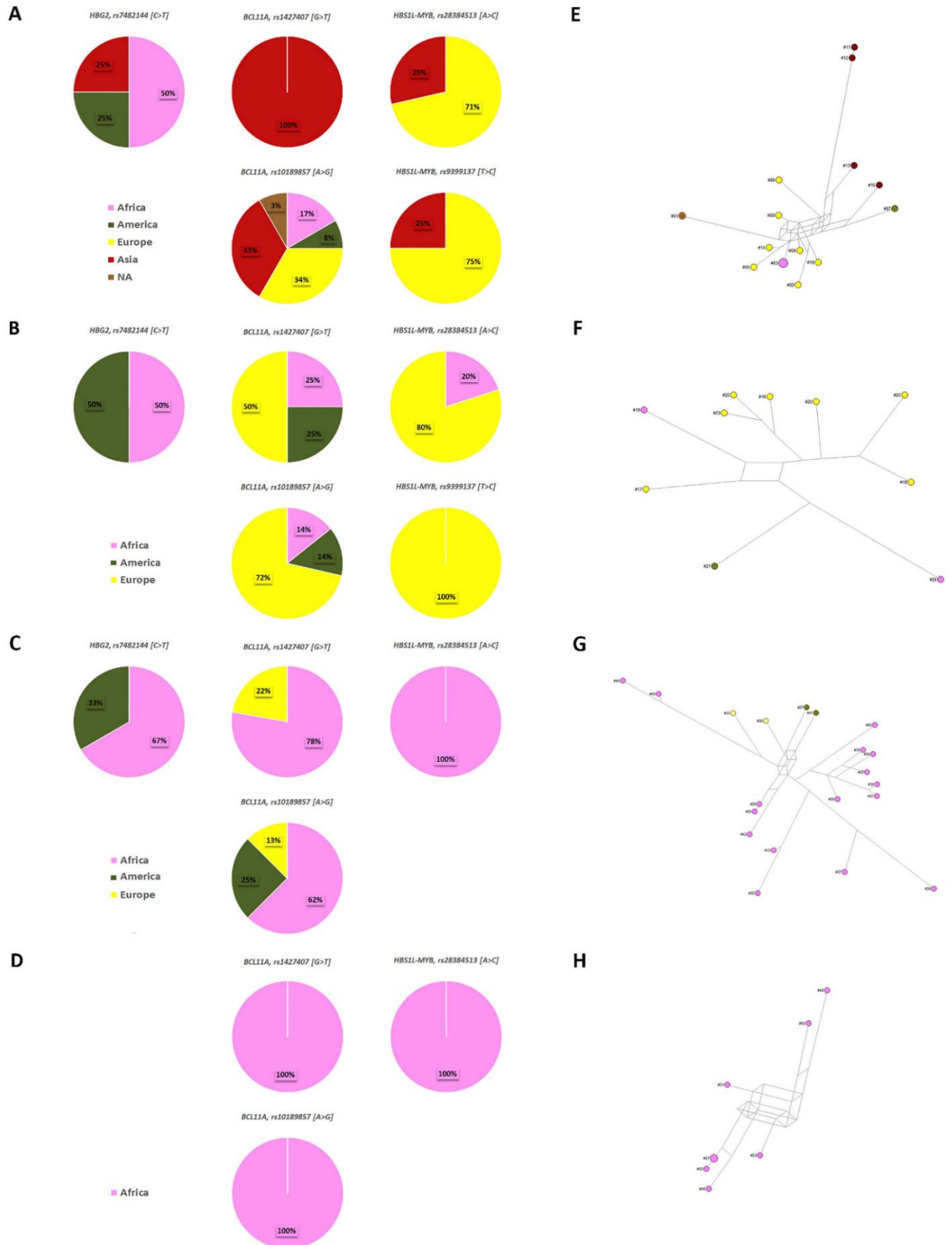


Figura 41. Frequenza degli SNPs nucleari, distribuzione geografica e network dei mtDNA nei diversi sottogruppi. (A) Frequenza degli SNPs nucleari e distribuzione geografica nei pazienti con β -talassemia (β^+/β^+ , β^0/β^0 e β^+/β^0), (B) pazienti microdrepanocitosi (HbS/β^+ o β^0), (C) pazienti con drepanocitosi (HbS/HbS) (D) eterozigoti composti (HbS/HbC e $HbO-Arab/HbC$) Pannello E-H: network ottenuti includendo tutte le variazioni del DNA mitocondriale e nucleare, considerate separatamente in ciascun gruppo patologico e colorate a seconda della loro origine geografica [187].

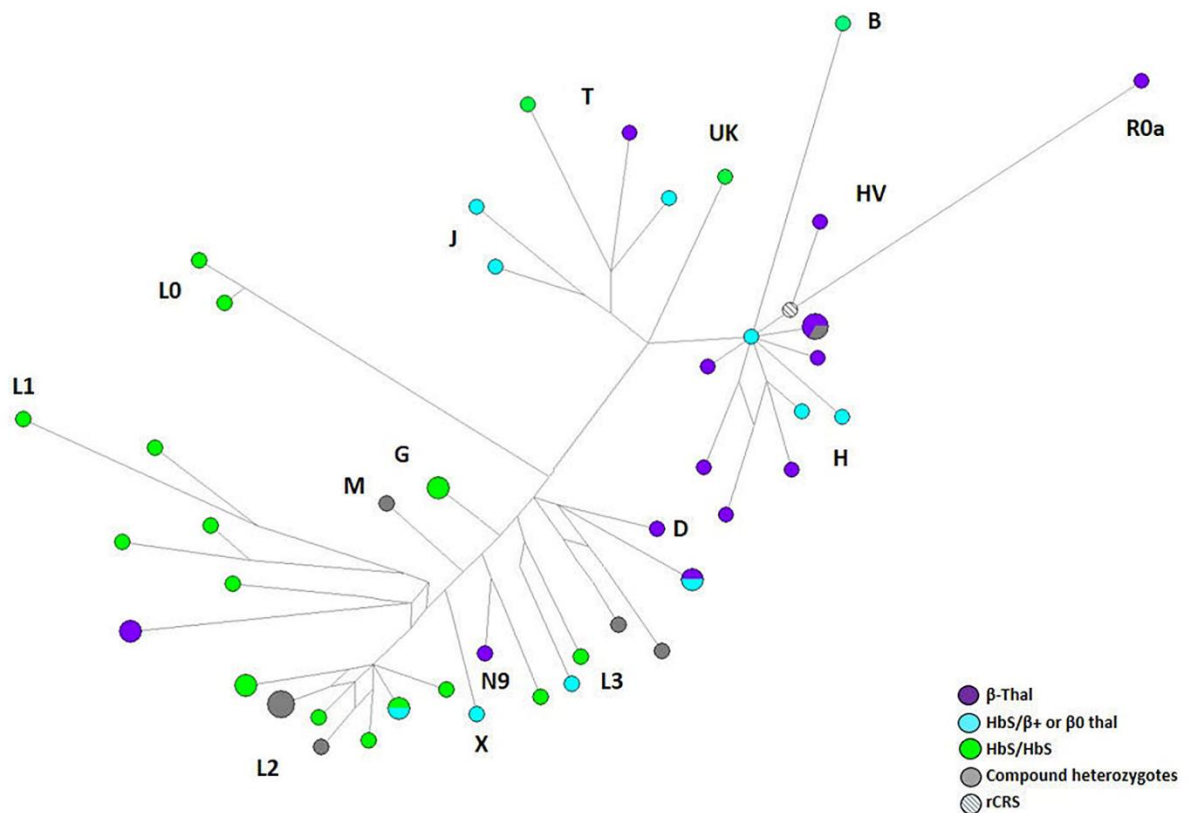


Figura 42. Network di variazione mitocondriale ed emoglobinopatie. Albero dei 53 aplotipi mtDNA osservati tra i pazienti in cui sono stati evidenziati diversi sottogruppi di emoglobinopatie [187].

È stata eseguita un'analisi delle componenti principali (PCA, *Figura 43*) considerando tutti i siti variabili sia mitocondriali che nucleari per i 53 campioni. È risultata una chiara suddivisione tra i due cladi africani, posti a sinistra, ed i due europei a destra. Inoltre, i campioni americani si sono distribuiti nei primi due quarti, vicino agli europei. Il grafico mostrava anche tre campioni europei vicini ai quarti africani, ovvero: i due pazienti albanesi il 17 (aplogruppo X) e il 26 (JT) e il paziente 35 (L3) dall'Italia. D'altro canto, cinque campioni africani si sono localizzati vicino alla linea verticale che separa i quarti di sinistra e destra, ed erano: il paziente 24 (L3) dalla Guinea; i pazienti 29 (G), 43 (L3) e 45 (L0) da Nigeria; e il paziente 44 (L0) dal Congo. È interessante notare che tutti questi pazienti insieme a tre europei (pz 17, 26, 35) avevano almeno un allele HbS, che è molto frequente nel Africa Sub-sahariana. Il quarto prevalentemente europeo (primo quarto a destra) mostra però anche quattro pazienti africani (pz 3 e 4 fratelli e pz 48 e 52). I due pazienti più

distanti erano africani provenienti dal Nord Africa (Marocco e Algeria) confermando la vicinanza alle cladi europee. È interessante notare che solo due campioni talassemici si trovano nella metà sinistra della rappresentazione della PCA: erano i pazienti 11 e 12, gli unici provenienti dall'Asia occidentale; mentre gli altri due campioni asiatici provenivano dalla Cina e dalla Cambogia e rientrano nella metà destra del grafico.

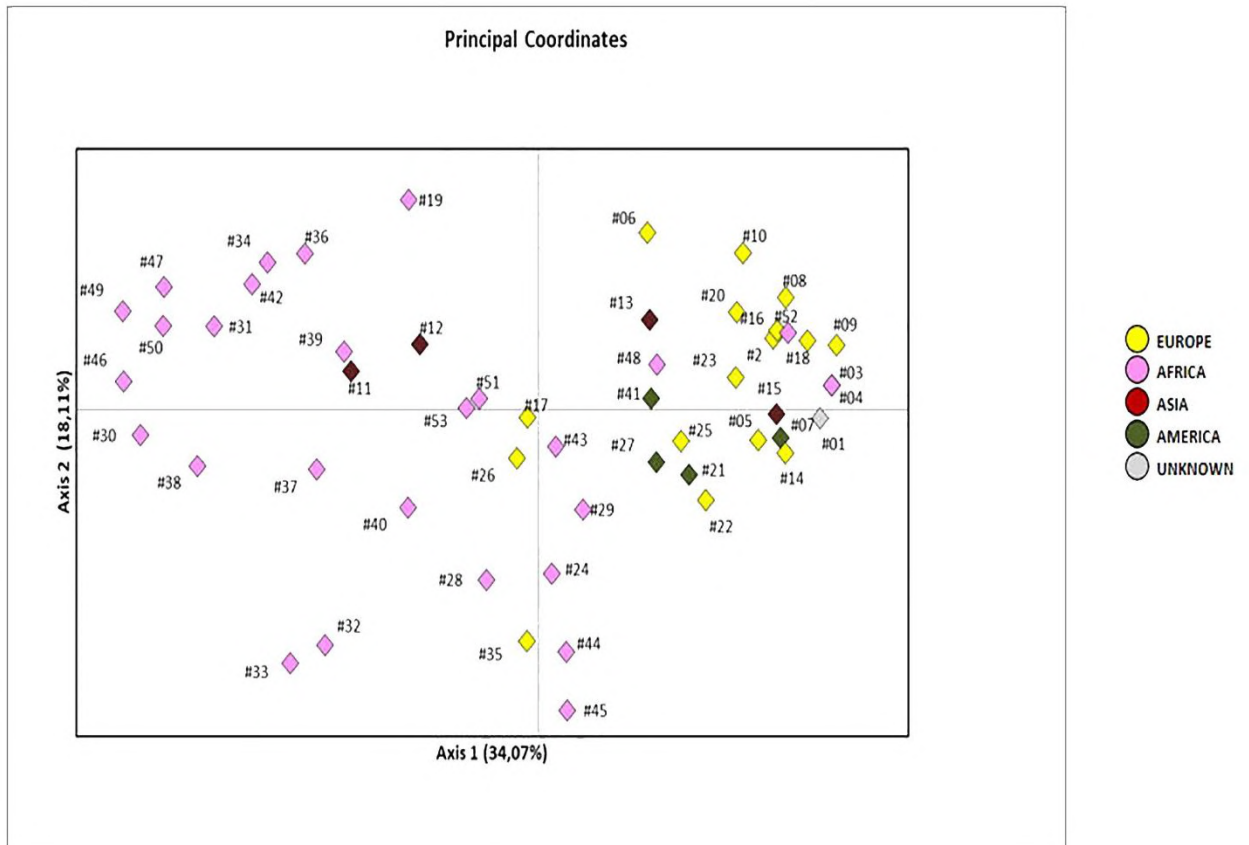


Figura 43. Analisi delle componenti principali (PCA). Grafico bidimensionale di analisi delle componenti principali ottenuto includendo tutte le variazioni del DNA mitocondriale e nucleare, considerate separatamente in ciascun campione e colorate in base alla loro origine geografica [187].

12. DISCUSSIONE DEL PRIMO STUDIO

12.1. Analisi dei SNPs modulatori del fenotipo clinico e distribuzione geografica

Dall'analisi dei cinque loci polimorfici legati alla modulazione dell'espressione dell'emoglobina fetale, ovvero *HBG2* rs7482144 [C> T], *BCL11A* rs1427407 [G> T], *BCL11A* rs10189857 [A> G], *HBS1L-MYB* rs28384513 [A> C], *HBS1L-MYB* rs939913 [T> C] e un altro modulatore, ovvero *C/EBPE* rs45496295 [C> T] è emerso che il sito polimorfico *BCL11A* rs10189857 [A> G] è risultato più frequente (allele G in 33/53 pazienti), invece *HBS1L-MYB* rs9399137 [T> C] è stato quello meno frequente (allele C in 6/53 pazienti). Mentre non sono stati trovati pazienti portatori della variazione *C/EBPE* rs45496295 [C> T] (*Tabella 14*, *Tabella 15*, appendice B). Dalla correlazione tra le frequenze dei sei SNPs analizzati e l'origine geografica dei pazienti (*Figura 36* e *Figura 37 A-C*) è stato riscontrato che *HBG2* rs7482144 [C> T] è stato identificato in nove pazienti, cinque africani, tre americani e uno asiatico (*Figura 37 A*, *Tabella 14*). L'assenza di questa variazione nei soggetti europei non si adattava completamente alle frequenze alleliche recuperate dal database GnomAD, che riportava alleli T nella popolazione europea, sebbene con una frequenza inferiore (alleli 3839 T e alleli C 10893) (*Figura 37 B* e *C*). Mentre l'*HBS1L-MYB* rs9399137 [T> C], è stato rilevato solo in sei pazienti, cinque europei e un asiatico (*Figura 37 A* e *B*, *Tabella 14*). Non sono stati trovati pazienti africani con *HBS1L-MYB* rs9399137 [T> C] nella nostra coorte sebbene questo rappresentasse il gruppo più numeroso. Le frequenze alleliche ottenute dal database GnomAD riportavano alleli C nelle popolazioni americane e africane, sebbene a frequenze inferiori (*Figura 37 B* e *C*). È interessante notare che è stato osservato come le frequenze dei cinque SNP analizzati fossero correlate all'origine geografica dei pazienti (*Figura 36* e *Figura 37 A-C*). Sebbene nella coorte ci fossero solo quattro casi asiatici, sono stati rilevati tutti gli SNPs indagati. Al contrario, negli altri pazienti (cioè europei, africani, americani), la frequenza di ogni SNPs era notevolmente diversa. *HBS1L-MYB* rs9399137 T> C non è stato rilevato nei pazienti africani o americani, mentre *HBG2* rs7482144 C> non è stato rilevato nei pazienti europei (*Figura 36*, *Tabella 14*, *Figura 37 A* e *B*).

12.2. Analisi delle variazioni della sequenza di mtDNA e correlazione con le varianti dei geni modificatori

Dall'allineamento delle sequenze dei rispettivi campioni alla sequenza di riferimento Cambridge ha rivelato 99 siti polimorfici (dieci trasversioni e due indel). È stata stimata la diversità dei nucleotidi (π) in tutti i campioni pari a 0,0173. La diversità degli aplotipi era molto alta ($H_d = 0,993$) così come la media numero di differenze nucleotidiche ($k = 13.052$).

Dall'analisi dei 13 diversi macroaplogruppi (B, D, G, JT, L0, L1, L2, L3, M, N9, R0, UK e X, mostrati in *Figura 38 A*), ottenuti dalla classificazione dei 44 diversi aplotipi, è emerso che sette aplotipi sono stati condivisi tra due o tre campioni, ma cinque coppie di questi individui hanno dichiarato di esserlo parenti materni (pz 3-4, 7 e 21, 11-12, 16-17, 47 e 49-50). Per i restanti due aplotipi, uno (HT02) è stato trovato in due fratelli dall'Algeria e un campione dal Marocco, l'altro (HT04) è stato rilevato in un paziente dalla Costa d'Avorio e uno dal Benin.

Il network rappresentante le relazioni filogenetiche tra i campioni, mostrato in *Figura 39*, ha permesso di individuare due principali cluster di mtDNA: superaplogruppi R (prevalevano gli aplogruppi JT e H) e non-R (includeva principalmente i sub-lignaggi L1, L2 e L3 sub-lignaggi). Tra i campioni, il lignaggio più rappresentato apparteneva al super-aplogruppo L (45%), composto da L0 (8%), L1 (17%), L2 (54%) e L3 (21%), seguito da R0 (25%) e M (13%) (*Figura 38 B*). L'aplogruppo JT è stato rilevato con il 9% di frequenza, mentre B, N9, UK e X i lignaggi erano rappresentati ciascuno da un solo campione.

È noto che per gli esseri umani il macro-aplogruppo L rappresenta il lignaggio mitocondriale più ancestrale, se si escludono i rami derivati L3, M e N, rappresenta la maggior parte della variabilità tipica del mtDNA subsahariana. All'interno di L, L2 è il più comune aplogruppo in Africa ed è stato precedentemente osservato in tutto il continente [180]. È stata evidenziata una concordanza tra la frequenza dell'aplogruppo e l'origine dei nostri campioni, specificamente quelli provenienti da Nigeria e Costa d'Avorio (ma anche Congo e Benin), fatta eccezione per i due fratelli siriani (PZ 11 E 12). L3 è stato trovato in: due campioni dalla Nigeria, uno dalla Guinea, uno da Ghana e uno dall'Italia. Anche se gli aplotipi L rappresentano <1% degli mtDNA in Europa, essendo L1b l'aplogruppo più comune, le frequenze

cambiano ampiamente, variando dal 3% in Europa meridionale allo 0,7% nell'Europa centrale e allo 0,5% nel Nord Europa. All'interno di R0, sono stati trovati solo un campione R0a e un HV1, mentre i restanti pazienti erano H, corrispondenti quindi a un totale di 11/53 (21%). L'aplogruppo H è il mitocondriale più frequente lignaggio in Europa (~40%) con un andamento in declino dall'Europa occidentale verso il Vicino Oriente e il Caucaso (~10-20%). È stato trovato anche in Africa, soprattutto nel nord regione, con un picco di H1 nel Maghreb e questo è in accordo con l'origine dei nostri campioni H. (Italia, Albania, Marocco e Algeria). Per quanto riguarda il macro-aplogruppo M, i campioni provenivano da Venezuela e Cambogia e apparteneva ai sottorami D4 (D1f1 e D4b2d), che rispecchiano esattamente il massimo geografico distribuzione di questi lignaggi (Sud America e Asia orientale, rispettivamente) [181]. Anche il paziente 53 del Marocco, appartenente all'aplogruppo mitocondriale M1a2a, era congruente con la distribuzione di frequenza del lignaggio, a differenza dei pazienti africani 28 e 29 (fratelli) e 40, che appartenevano rispettivamente a G1a e M30c, entrambe sotto-lignaggi tipici dell'Asia. Potrebbe risultare l'aplogruppo B4 dell'Asia orientale rivelato nel paziente 27 della Repubblica Dominicana strano, ma dovremmo tenere a mente che una sotto clade di B4b è uno dei cinque aplogruppi trovati tra gli indigeni delle Americhe, spiegando così una minore distribuzione anche in Centro America. Il paziente 13 dalla Cina appartenente a N9a10a riflette perfettamente l'origine dell'Asia orientale e il picco di distribuzione del lignaggio [182].

Dall'analisi degli aplotipi della regione di controllo mitocondriale, le cui relazioni filogenetiche sono illustrate nella *Figura 39*, e delle variazioni SNPs nucleari dei cinque geni variabili (*Figura 40, Tabella 14*) è emerso che tutti gli individui all'interno del macro-aplogruppo mitocondriale L erano wildtype omozigote per *HBS1L-MYB* rs9399137 [T> C], mentre i pazienti che presentavano il lignaggio mitocondriale JT erano omozigoti wildtype per *HBG2* rs7482144 [C> T]. Al contrario, *C/EBPE* rs45496295 [C> T] è sempre stato trovato come omozigote wildtype, quindi è stato escluso dalla rappresentazione.

Dall'analisi del network dell'albero filogenetico rappresentante la possibile integrazione tra la variabilità della regione di controllo del mtDNA e le mutazioni nucleari associati rispettivamente ai quattro sottogruppi di emoglobinopatie (*Tabella 14, Tabella 15, Figure 41 e 42*) è stato rilevato la predominanza di alcuni pazienti in

determinate cladi. In particolare, quasi la totalità dei pazienti con drepanocitosi si distribuivano tra gli aplogruppi non-R (lignaggi africani L0, L1, L2 e L3), tranne tre pazienti (pz 26, 27 e 41). Anche per i pazienti eterozigoti composti si distribuivano in calde non-R, ad eccezione del paziente (52, aplogruppo H1). Mentre i pazienti talassemici si riscontrano solo nel clade R, ad eccezione di quattro pazienti (7, 11, 12 e 15). Separando i diversi network per ciascun sottogruppo ed evidenziando l'origine geografica è emerso come tutti i campioni asiatici fossero tutti talassemici, ma raggruppati lontani dai pazienti talassemici europei (*Figura 41 E*). Mentre nessun paziente africano era affetto da talassemia (ad eccezione dei fratelli algerini). Inoltre se esclusi due campioni con microdrepanocitosi (HbS / β^+ o β^0) (*Figura 41 F*) i soggetti africani prevalgono nel gruppo della drepanocitosi (*Figura 41 G*) e sono anche l'unica fonte degli eterozigoti composti (*Figura 41 H*). Non è emerso alcun modello di correlazione molecolare simile a una stella che potesse rappresentare qualsiasi evento di espansione locale. Quindi è stato osservato che i pazienti europei cadono prevalentemente nel gruppo dei talassemici o con microdrepanocitosi, mentre gli africani nei rimanenti due sottogruppi (ovvero, HbS/HbS e HbS/ HbC) (*Figura 42*).

Anche dall'analisi delle componenti principali nei 53 campioni (*Figura 43*) è stata riscontrata una netta suddivisione tra i cladi africani (quarti di sinistra) e quelli europei (quarti a destra). Inoltre, i campioni americani si sono distribuiti nei primi due quarti, vicino agli europei. Il grafico mostrava anche tre campioni europei vicini ai quarti africani, ovvero: i due pazienti albanesi il 17 (aplogruppo X) e il 26 (JT) e il paziente 35 (L3) dall'Italia. D'altro canto, cinque campioni africani si sono localizzati vicino alla linea verticale che separa i quarti di sinistra e destra, ed erano: il paziente 24 (L3) dalla Guinea; i pazienti 29 (G), 43 (L3) e 45 (L0) da Nigeria; e il paziente 44 (L0) dal Congo. È interessante notare che tutti questi pazienti insieme a tre europei (pz 17, 26, 35) avevano almeno un allele HbS, che è molto frequente nel Africa Sub-sahariana. Il quarto prevalentemente europeo (primo quarto a destra) mostra però anche quattro pazienti africani (pz 3 e 4 fratelli e pz 48 e 52). I due pazienti più distanti erano africani provenienti dal Nord Africa (Marocco e Algeria) confermando la vicinanza alle cladi europee. È interessante notare che solo due campioni talassemici si trovano nella metà sinistra della rappresentazione della PCA: erano i pazienti 11 e 12, gli unici provenienti dall'Asia occidentale; mentre gli altri due

campioni asiatici provenivano dalla Cina e dalla Cambogia e rientrano nella metà destra del grafico. Anche se le migrazioni passate e recenti avrebbero potuto alterare parzialmente tale quadro, un alto livello di popolazione e la strutturazione geografica persiste ancora.

13.RISULTATI DEL SECONDO STUDIO

13.1. Risultati della validazione del pannello NGS

- a. Report di qualità: in entrambe le due corse di validazione sono state ottenute un elevato numero di reads in 45 dei 48 campioni, seppur con una maggiore variabilità nella seconda corsa: in tre di questi infatti, abbiamo riscontrato un numero di reads inferiore a 1 milione, considerato come valore limite. Questa variabilità può essere dovuta a molteplici fattori, fra cui, non ultimo, la qualità del DNA. Tuttavia, uno dei tre campioni con numero insufficiente di reads è un replicato inter-run, cioè analizzato in entrambe le corse, con un numero di reads nella prima corsa pienamente rispondente agli standards di qualità. Questo fenomeno può essere spiegato con una variazione fra i due esperimenti dovuta a variabilità dell'operatore.

Inoltre una parte delle reads generate viene scartata perché ritenuta di qualità insufficiente. In ogni caso, in tutti i campioni è stata raggiunta una quantità di reads correttamente mappate, con percentuali comprese fra il 96.5 e il 99.6% (valore soglia 95%). In particolare, anche se nella seconda corsa il numero delle reads generate risulta meno omogeneo, la percentuale delle reads che vengono mappate, in entrambe le corse, è in tutti i nostri campioni superiore al 99% (*Figura 44*).

VP_run1

A

Sample ID	Number of reads	Number of Mapped reads	Number of reads with problematic mate	Percentage of Mapped reads
Phix	1,210,214	1,168,377	49,193	96.54%
SG063	2,504,768	2,487,815	18,313	99.32%
SG063-BIS	1,691,482	1,676,550	18,168	99.12%
UPN61	1,934,442	1,922,716	15,958	99.39%
UPN62	1,875,818	1,862,283	17,925	99.28%
1	1,493,488	1,482,629	11,153	99.27%
1-BIS	1,595,084	1,584,118	13,724	99.31%
UPN64	1,763,248	1,750,754	15,254	99.29%
UPN65	1,658,806	1,647,311	14,269	99.31%
UPN66	1,862,270	1,851,213	11,745	99.41%
UPN67	2,234,478	2,219,394	22,506	99.32%
2	2,124,292	2,109,758	13,006	99.32%
3	2,056,294	2,043,591	14,447	99.38%
4	2,365,734	2,349,733	20,083	99.32%
UPN71	1,538,572	1,525,536	16,166	99.15%
UPN72	2,219,058	2,202,054	24,722	99.23%
UPN73	2,104,974	2,084,983	26,403	99.05%
UPN74	1,938,674	1,922,556	15,630	99.17%
UPN75	1,942,468	1,926,062	27,282	99.16%
UPN76	1,720,106	1,702,803	17,999	98.99%
5	1,514,166	1,501,661	16,277	99.17%
6	1,419,162	1,408,263	16,909	99.23%
7	1,334,282	1,321,003	24,381	99.00%
8	1,427,790	1,415,242	15,550	99.12%
UPN81	1,291,472	1,279,570	14,398	99.08%
Total	44,821,142	44,445,975	471,461	99.16%

B

VP_run2

Sample ID	Number of reads	Number of Mapped reads	Number of reads with problematic mate	Percentage of Mapped reads
Phix	1,484,788	1,436,958	61,656	96.78%
SGO63TRIS	1,801,834	1,790,119	26,371	99.35%
UPN62BIS	1,657,230	1,645,769	32,507	99.31%
2BIS	2,260,352	2,244,543	35,623	99.30%
6BIS	796,438	792,114	5,452	99.46%
UPN82	2,294,334	2,275,589	70,921	99.18%
UPN83	2,162,854	2,147,267	43,273	99.28%
UPN84	1,838,166	1,821,460	66,096	99.09%
UPN85	2,108,310	2,089,002	73,326	99.08%
UPN86	1,124,686	1,115,718	25,906	99.20%
9	2,886,878	2,871,505	18,051	99.47%
9BIS	1,424,904	1,416,123	17,307	99.38%
UPN88	2,499,126	2,483,867	19,989	99.39%
UPN89	2,382,542	2,367,219	31,619	99.36%
UPN90	2,279,046	2,267,508	18,314	99.49%
UPN91	2,212,228	2,203,314	17,960	99.60%
UPN92	1,790,102	1,778,967	27,229	99.38%
UPN93	1,665,906	1,656,020	23,558	99.41%
UPN94	4,778,382	4,754,380	60,310	99.50%
UPN94BIS	3,005,906	2,992,169	28,649	99.54%
UPN95	2,011,030	2,001,137	17,289	99.51%
UPN96	1,034,084	1,028,794	9,620	99.49%
UPN97	891,688	886,882	6,440	99.46%
UPN98	621,608	618,002	4,446	99.42%
Total	47,012,422	46,684,426	741,912	99.30%

Figura 44. Pannello A: Resoconto reads prima corsa. Pannello B: Resoconto film seconda corsa.

Come già accennato, lo standard di qualità delle reads è definito da un valore di Phred-score che deve essere maggiore di 30 per più della metà dei cicli di cattura. Nelle figure seguenti (*Figura 45*) è riportato un esempio di 5 campioni, per ogni corsa, pair-ended, in cui il valore di Phred-score si attesta ad oltre 30 fino ad almeno 200 cicli, per poi scendere al di sotto di questo valore negli ultimi 50-100 cicli, come atteso. Questi valori indicano un'ottima qualità delle reads in tutti i campioni studiati;

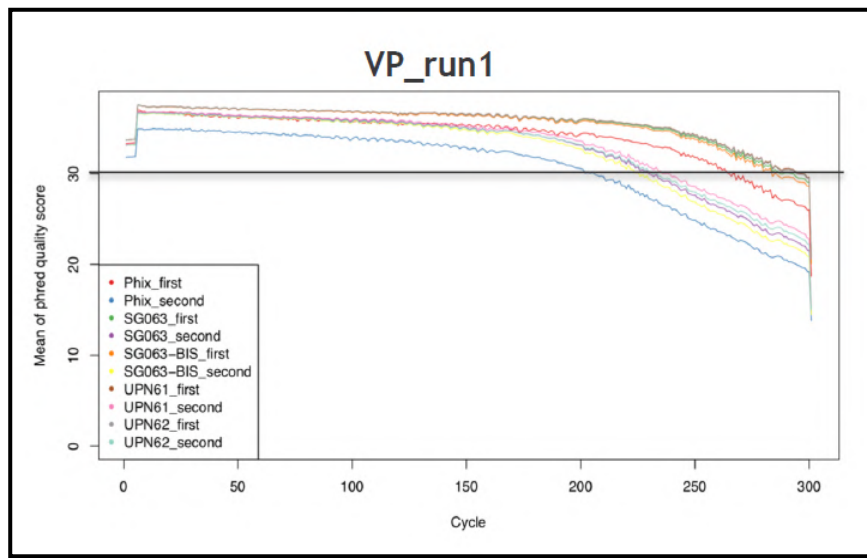
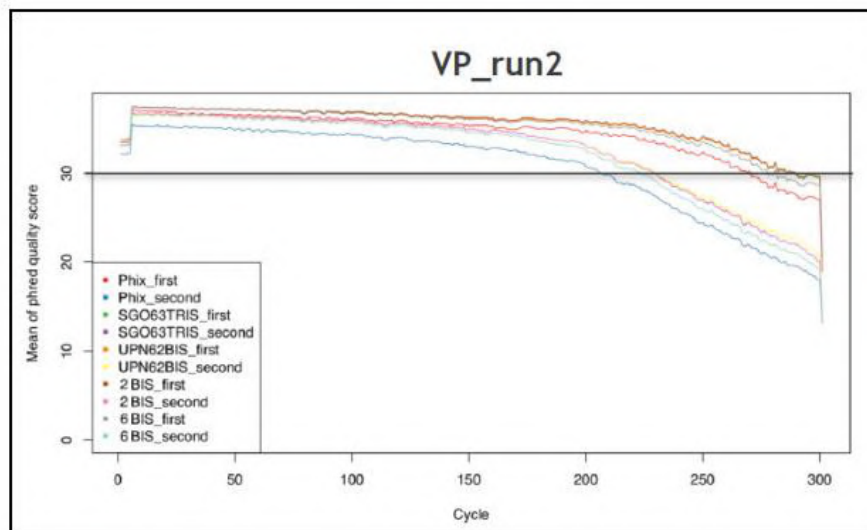
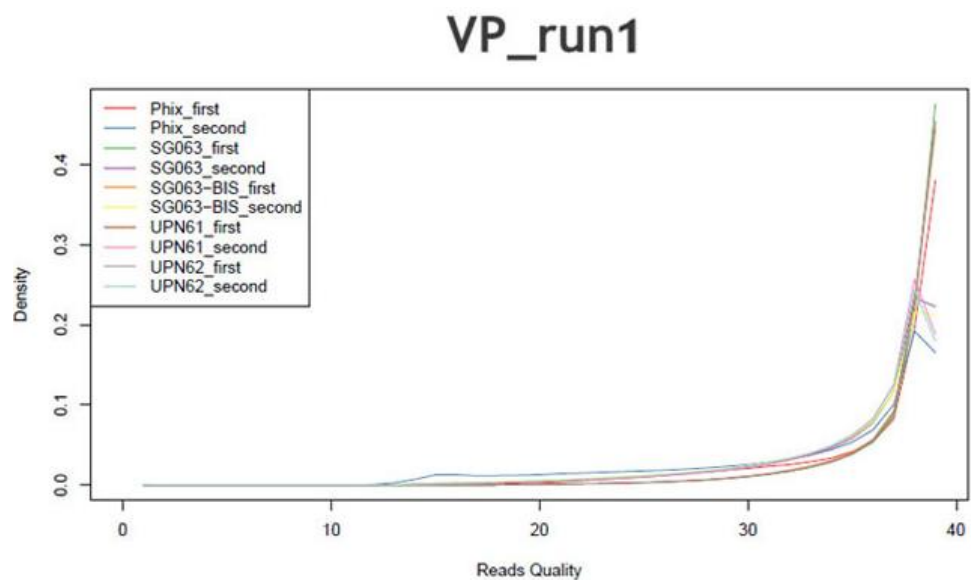
A**B**

Figura 45. Pannello A: Grafico con valore di Phred-score in 5 campioni della prima corsa. Pannello B: Grafico con valore di Phred-score in 5 campioni della seconda corsa.

Nella distribuzione media della qualità delle reads generate, deve essere presente un picco tra 35 e 40 per garantire la buona qualità delle stesse per la piattaforma Illumina (Figura 46).

A

B

VP_run2

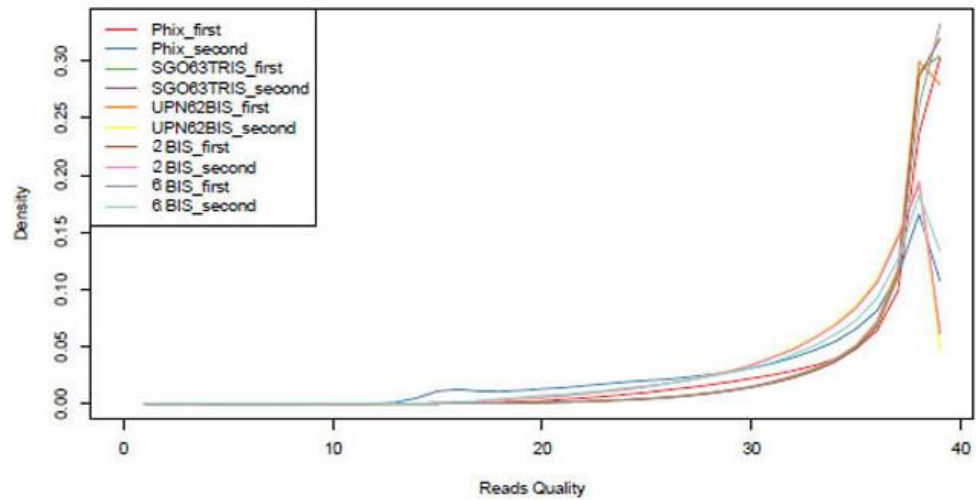


Figura 46. Pannello A: Distribuzione della media della qualità della reads per ognuno dei 5 campioni della prima corsa. Pannello B: Distribuzione della media della qualità della reads per ognuno dei 5 campioni della seconda corsa.

b. Analisi del Soft clipping: strumento bioinformatico utilizzato sia per ovviare a problematiche di lettura che per unire eventuali gaps presenti. Se una lettura non può essere allineata per tutta la sua lunghezza, alcune basi saranno "ritagliate" ai bordi della lettura. A seconda dei dati potrebbero esserci molti allineamenti di lettura con soft-clipping, tuttavia, il numero di basi che sono ritagliate dovrebbero essere piccoli. Se per molte letture il numero di basi tagliate è elevato, questo indica che si è verificato un problema e che molti dati vengono eliminati. Nelle due corse, la percentuale relativa a tale fenomeno risulta idonea (*Figura 47*) sottolineando così un'omogeneità dei dati prodotti.

A

Sample ID	No soft clip	Front soft clip	Soft clip both ends	End soft clip
Phix	1,336,123 (93%)	22,777 (1.6%)	18,867 (1.3%)	59,191 (4.1%)
SGO63TRIS	1,720,292 (96.1%)	17,465 (1%)	1,608 (0.1%)	50,754 (2.8%)
UPN62BIS	1,583,035 (96.2%)	14,243 (0.9%)	1,374 (0.1%)	47,117 (2.9%)
2BIS	2,151,902 (95.9%)	22,670 (1%)	2,047 (0.1%)	67,924 (3%)
6BIS	768,409 (97%)	5,753 (0.7%)	1,065 (0.1%)	16,887 (2.1%)
UPN82	2,160,570 (94.9%)	27,063 (1.2%)	2,080 (0.1%)	85,876 (3.8%)
UPN83	2,067,382 (96.3%)	17,779 (0.8%)	1,906 (0.1%)	60,200 (2.8%)
UPN84	1,728,438 (94.9%)	21,877 (1.2%)	2,423 (0.1%)	68,722 (3.8%)
UPN85	1,982,143 (94.9%)	26,294 (1.3%)	2,568 (0.1%)	77,997 (3.7%)
UPN86	1,073,961 (96.3%)	9,480 (0.8%)	1,233 (0.1%)	31,044 (2.8%)
9	2,783,494 (96.9%)	22,060 (0.8%)	1,819 (0.1%)	64,132 (2.2%)
9BIS	1,364,048 (96.3%)	12,894 (0.9%)	1,242 (0.1%)	37,939 (2.7%)
UPN88	2,399,631 (96.6%)	21,872 (0.9%)	1,541 (0.1%)	60,823 (2.4%)
UPN89	2,280,539 (96.3%)	20,301 (0.9%)	2,132 (0.1%)	64,247 (2.7%)
UPN90	2,189,161 (96.5%)	20,740 (0.9%)	1,973 (0.1%)	55,634 (2.5%)
UPN91	2,125,471 (96.5%)	21,318 (1%)	2,158 (0.1%)	54,367 (2.5%)
UPN92	1,701,623 (95.7%)	22,498 (1.3%)	1,952 (0.1%)	52,894 (3%)
UPN93	1,587,046 (95.8%)	19,812 (1.2%)	1,710 (0.1%)	47,452 (2.9%)
UPN94	4,518,968 (95%)	74,952 (1.6%)	5,949 (0.1%)	154,511 (3.2%)
UPN94BIS	2,866,972 (95.8%)	39,373 (1.3%)	2,209 (0.1%)	83,615 (2.8%)
UPN95	1,933,765 (96.6%)	16,816 (0.8%)	1,803 (0.1%)	48,753 (2.4%)
UPN96	994,514 (96.7%)	9,773 (0.9%)	1,094 (0.1%)	23,413 (2.3%)
UPN97	861,428 (97.1%)	6,069 (0.7%)	890 (0.1%)	18,495 (2.1%)
UPN98	600,412 (97.2%)	4,129 (0.7%)	691 (0.1%)	12,770 (2.1%)

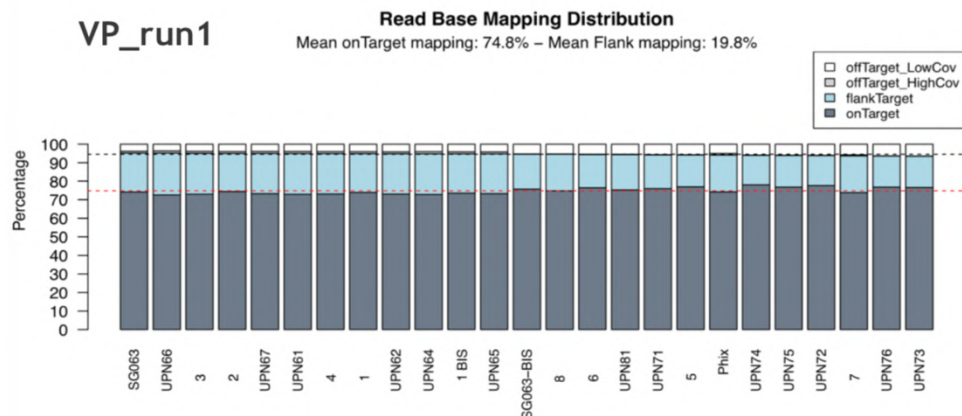
B

Sample ID	No soft clip	Front soft clip	Soft clip both ends	End soft clip
Phix	1,083,048 (92.7%)	27,169 (2.3%)	9,713 (0.8%)	48,447 (4.1%)
SG063	2,407,881 (96.8%)	22,760 (0.9%)	2,586 (0.1%)	54,588 (2.2%)
SG063-BIS	1,617,116 (96.5%)	17,227 (1%)	1,440 (0.1%)	40,767 (2.4%)
UPN61	1,859,413 (96.7%)	17,705 (0.9%)	2,021 (0.1%)	43,577 (2.3%)
UPN62	1,793,323 (96.3%)	20,514 (1.1%)	1,787 (0.1%)	46,659 (2.5%)
1	1,435,275 (96.8%)	13,669 (0.9%)	1,521 (0.1%)	32,164 (2.2%)
1 BIS	1,530,344 (96.6%)	15,640 (1%)	1,491 (0.1%)	36,643 (2.3%)
UPN64	1,694,896 (96.8%)	15,524 (0.9%)	1,886 (0.1%)	38,448 (2.2%)
UPN65	1,598,840 (97.1%)	13,366 (0.8%)	1,504 (0.1%)	33,601 (2%)
UPN66	1,797,288 (97.1%)	14,327 (0.8%)	1,602 (0.1%)	37,996 (2.1%)
UPN67	2,134,699 (96.2%)	24,621 (1.1%)	2,297 (0.1%)	57,777 (2.6%)
2	2,045,884 (97%)	17,861 (0.8%)	1,913 (0.1%)	44,100 (2.1%)
3	1,979,444 (96.9%)	17,730 (0.9%)	1,701 (0.1%)	44,716 (2.2%)
4	2,255,619 (96%)	29,424 (1.3%)	3,066 (0.1%)	61,624 (2.6%)
UPN71	1,470,451 (96.4%)	15,576 (1%)	1,315 (0.1%)	38,194 (2.5%)
UPN72	2,100,262 (95.4%)	31,905 (1.4%)	2,169 (0.1%)	67,718 (3.1%)
UPN73	2,000,276 (95.9%)	25,408 (1.2%)	1,733 (0.1%)	57,566 (2.8%)
UPN74	1,844,631 (95.9%)	24,940 (1.3%)	1,489 (0.1%)	51,496 (2.7%)
UPN75	1,821,592 (94.6%)	34,076 (1.8%)	2,258 (0.1%)	68,136 (3.5%)
UPN76	1,640,216 (96.3%)	18,503 (1.1%)	1,575 (0.1%)	42,509 (2.5%)
5	1,432,556 (95.4%)	23,201 (1.5%)	1,554 (0.1%)	44,350 (3%)
6	1,343,569 (95.4%)	20,272 (1.4%)	1,644 (0.1%)	42,778 (3%)
7	1,264,200 (95.7%)	16,150 (1.2%)	1,564 (0.1%)	39,089 (3%)
8	1,364,307 (96.4%)	15,046 (1.1%)	1,388 (0.1%)	34,501 (2.4%)
UPN81	1,232,312 (96.3%)	13,661 (1.1%)	1,328 (0.1%)	32,269 (2.5%)

Figura 47. Pannello A. Percentuale relativa al soft clipping della prima corsa. Pannello B. Percentuale relativa al soft clipping della seconda corsa.

c. Cattura delle regioni target: un parametro importante da considerare è quello della cattura delle regioni target (la sequenza codificante dei geni oggetto di studio), le regioni fiancheggianti 3' e 5' UTR, le regioni di splicing e le regioni regolatorie, ed in alcuni casi l'intero gene, così come richiesto in fase di elaborazione del pannello NGS. Questo parametro sta ad indicare la specificità della "card", cioè quali sequenze in percentuale mappano nelle regioni d'interesse. In entrambe le corse di validazione è stato ottenuto un valore medio di cattura del 94.6% nella prima corsa (74.8% on target e 19.8% flank region) ed un valore medio di 95.7% nella seconda corsa (72.7% on target e 23% flank region) (Figura 48).

A



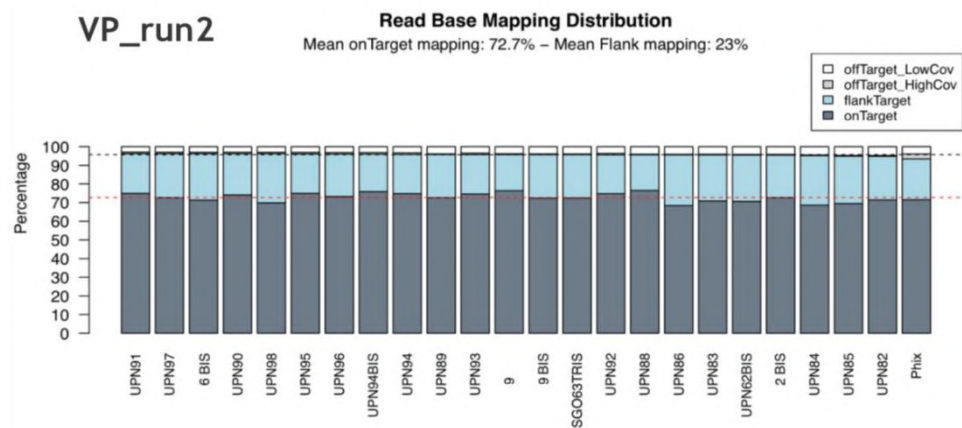
B

Figura 48. Pannello A: valore medio di cattura nella prima corsa. Pannello B: valore medio di cattura nella seconda corsa.

- d. Duplicati di PCR: un artefatto che è possibile riscontrare negli esperimenti NGS a cattura, è la presenza di duplicati di PCR; questo significa che ciascun filamento può funzionare da stampo ed essere risintetizzato (e quindi sequenziato) più volte. Questo fenomeno dovrebbe essere ridotto al minimo, in quanto può portare alla chiamata di varianti false positive, e la frazione di duplicati di PCR in ciascun campione dovrebbe essere compresa fra il 15 e il 25%. Nelle due corse di validazione tutti i campioni studiati hanno rispettato questi valori di riferimento, ad eccezione del campione UPN 80 che presenta una frazione di duplicati di PCR del 37.94%. Tuttavia, l'alto numero di reads del campione (circa 1.4 milioni) ha consentito di considerare affidabile anche il dato genetico emerso in tale paziente;
- e. Copertura: il termine copertura indica il numero di sequenze indipendenti che leggono la stessa posizione nel genoma sequenziato. Nel caso di pannelli di targeted-resequencing come quello elaborato per identificare mutazioni di tipo germinale nei pazienti, il livello di copertura richiesto è fissato a 200x. Questo valore è stato raggiunto in tutti i campioni in una percentuale variabile fra il 99.21% e il 99.95% nella prima corsa e il 91.43% e il 99.97% nella seconda corsa. La copertura osservata è risultata eterogenea fra i diversi campioni, ma tuttavia sempre superiore al valore di 200x. Il test di validazione ha perciò confermato un'ottima copertura, attribuendo al pannello di screening da noi proposto caratteristiche di alta sensibilità. L'analisi dei dati ha inoltre evidenziato una sostanziale omogeneità di copertura delle regioni target, mentre in nessuna regione genica specifica è stato osservato una copertura sistematicamente bassa (*Figura 49*).

A

VP_run1									
Sample ID	25x	50x	100x	200x	500x	1000x	Coverage 10% quantile	Coverage Heterogeneity	
Phix	99.98%	99.96%	99.95%	99.66%	75.26%	4.91%	398x	0.12%	
SG063	100.00%	99.99%	99.98%	99.95%	99.62%	84.04%	881x	0.08%	
SG063-BIS	99.98%	99.97%	99.93%	99.68%	91.48%	49.89%	524x	0.31%	
UPN61	99.99%	99.98%	99.96%	99.87%	98.57%	67.00%	744x	0.15%	
UPN62	99.99%	99.98%	99.96%	99.86%	98.05%	61.66%	672x	0.15%	
1	99.99%	99.98%	99.96%	99.82%	92.49%	28.87%	533x	0.10%	
1 BIS	99.99%	99.98%	99.96%	99.89%	94.17%	40.67%	560x	0.07%	
UPN64	100.00%	99.99%	99.96%	99.92%	97.07%	52.63%	644x	0.09%	
UPN65	99.99%	99.98%	99.96%	99.91%	94.13%	40.19%	566x	0.08%	
UPN66	99.99%	99.98%	99.95%	99.86%	98.04%	63.82%	716x	0.15%	
UPN67	99.99%	99.98%	99.96%	99.87%	99.06%	79.55%	846x	0.14%	
2	99.99%	99.98%	99.96%	99.94%	98.97%	72.45%	752x	0.08%	
3	99.99%	99.98%	99.96%	99.88%	98.73%	71.26%	737x	0.14%	
4	99.99%	99.98%	99.97%	99.93%	99.40%	83.05%	880x	0.14%	
UPN71	99.97%	99.94%	99.91%	99.40%	90.00%	46.47%	501x	0.56%	
UPN72	99.98%	99.97%	99.94%	99.86%	96.96%	73.63%	696x	0.38%	
UPN73	99.98%	99.97%	99.94%	99.86%	96.07%	67.10%	642x	0.27%	
UPN74	99.97%	99.95%	99.92%	99.80%	94.88%	61.81%	601x	0.29%	
UPN75	99.98%	99.95%	99.93%	99.73%	95.76%	65.74%	634x	0.38%	
UPN76	99.98%	99.98%	99.94%	99.74%	91.72%	47.98%	526x	0.25%	
5	99.97%	99.96%	99.93%	99.21%	84.47%	35.18%	420x	0.45%	
6	99.97%	99.96%	99.91%	99.47%	84.88%	31.45%	433x	0.28%	
7	99.97%	99.95%	99.90%	99.29%	83.61%	23.81%	427x	0.35%	
8	99.97%	99.96%	99.94%	99.54%	87.99%	26.77%	475x	0.27%	
UPN81	99.98%	99.94%	99.87%	99.23%	80.66%	18.26%	393x	0.40%	

B

VP_run2									
Sample ID	25x	50x	100x	200x	500x	1000x	Coverage 10% quantile	Coverage Heterogeneity	
Phix	99.99%	99.98%	99.88%	99.80%	89.02%	20.92%	490x	0.15%	
SG063TRIS	99.99%	99.96%	99.87%	99.61%	95.24%	61.15%	620x	0.47%	
UPN62BIS	99.98%	99.93%	99.83%	99.46%	92.97%	53.82%	561x	0.58%	
2 BIS	99.98%	99.96%	99.87%	99.73%	97.90%	76.36%	785x	0.44%	
6 BIS	99.99%	99.96%	99.90%	97.46%	39.33%	0.02%	263x	0.08%	
UPN82	100.00%	99.99%	99.94%	99.73%	96.82%	75.77%	729x	0.59%	
UPN83	99.99%	99.97%	99.87%	99.69%	96.51%	74.64%	735x	0.63%	
UPN84	100.00%	99.99%	99.91%	99.64%	94.12%	54.09%	590x	0.40%	
UPN85	100.00%	99.99%	99.94%	99.71%	95.31%	67.97%	632x	0.46%	
UPN86	99.99%	99.92%	99.70%	97.86%	74.25%	4.82%	365x	0.75%	
9	99.99%	99.98%	99.97%	99.94%	99.53%	91.79%	1.049x	0.23%	
9 BIS	99.98%	99.89%	99.85%	99.56%	91.05%	35.36%	515x	0.37%	
UPN88	99.99%	99.98%	99.96%	99.93%	98.86%	81.29%	852x	0.30%	
UPN89	99.99%	99.98%	99.88%	99.76%	98.07%	80.82%	822x	0.45%	
UPN90	99.99%	99.99%	99.97%	99.93%	99.22%	80.19%	820x	0.14%	
UPN91	99.99%	99.98%	99.97%	99.90%	98.64%	76.99%	804x	0.16%	
UPN92	99.99%	99.98%	99.96%	99.89%	96.42%	55.02%	610x	0.13%	
UPN93	99.99%	99.98%	99.86%	99.85%	95.36%	46.94%	594x	0.13%	
UPN94	100.00%	99.99%	99.99%	99.97%	99.87%	99.29%	1.826x	0.14%	
UPN94BIS	100.00%	99.99%	99.98%	99.96%	99.73%	94.11%	1.123x	0.07%	
UPN95	99.99%	99.98%	99.97%	99.92%	98.11%	70.10%	708x	0.15%	
UPN96	99.98%	99.96%	99.87%	99.22%	68.83%	0.32%	359x	0.15%	
UPN97	99.98%	99.95%	99.82%	98.92%	54.53%	0.02%	315x	0.18%	
UPN98	99.97%	99.86%	99.58%	91.43%	7.38%	0.00%	208x	0.21%	

Figura 49. Pannello A: Copertura osservata nella prima corsa; Pannello B: Copertura osservata nella seconda corsa.

Le regioni a bassa copertura (inferiore a 50x) di ciascun campione sono elencati nella *Figura 50*.

VP_run1		VP_run2	
Sample ID	Low coverage regions	Sample ID	Low coverage regions
Phix	1	Phix	2
SG063	1	SG063TRIS	2
SG063-BIS	1	UPN62BIS	3
UPN61	1	2 BIS	2
UPN62	1	6 BIS	1
1	1	UPN82	1
1 BIS	1	UPN83	3
UPN64	1	UPN84	1
UPN65	1	UPN85	3
UPN66	1	UPN86	4
UPN67	1	9	1
2	1	9 BIS	2
3	1	UPN88	1
4	1	UPN89	1
UPN71	2	UPN90	1
UPN72	1	UPN91	1
UPN73	1	UPN92	1
UPN74	2	UPN93	1
UPN75	3	UPN94	1
UPN76	1	UPN94BIS	1
5	1	UPN95	1
6	1	UPN96	2
7	4	UPN97	2
8	1	UPN98	3
UPN81	2		

Figura 50. Elenco di regioni a bassa copertura rilevate nella prima e seconda corsa.

Le rare regioni a bassa copertura, a causa di una relativa abbondanza di sequenze ripetute o di regioni ricche in GC e pertanto difficili da amplificare, andranno ulteriormente messe a punto in collaborazione con i tecnici bioinformatici, o indagate con metodologia standard di sequenziamento Sanger; Il 5' UTR dell'*HBA2* è in tutti i campioni sistematicamente caratterizzato da una bassa copertura (*Figura 51*), ad eccezione del campione UPN 84. Questa regione presenta un'elevata omologia con il 5' UTR del gene *HBA1*, tanto che il pannello non riesce a discriminare i due geni;

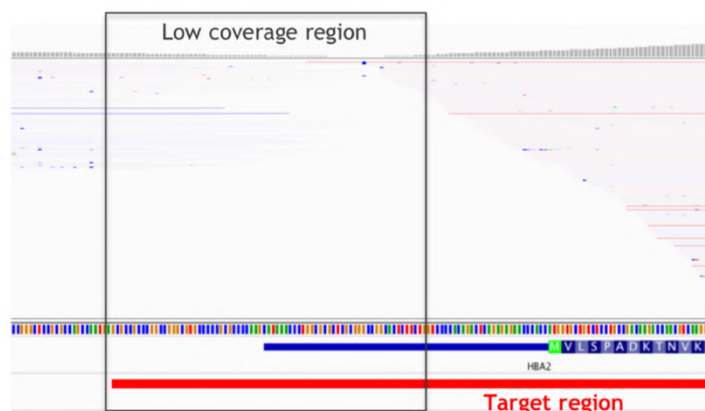


Figura 51. La figura rappresenta la regione 5' UTR del gene *HBA2* (16:222841-222882) caratterizzata da una bassa copertura.

- f. Falsi negativi e Falsi positivi: nel test di validazione è stato di grande importanza inserire campioni con varianti genomiche note a carico di geni da noi già esaminati tramite altre metodiche come il sequenziamento diretto e il Reverse Dot Blot, al fine di validare la robustezza e affidabilità della metodica. Tre varianti geniche a carico dei geni *BCL11A*, *TMPRSS6* e *TERF1* sono presenti nella totalità dei campioni, compresi i replicati di riferimento forniti dalla ditta stessa per la validazione. Queste variazioni sono state chiaramente identificata come falsi positivi ed il motivo di questo risultato è probabilmente imputabile al fatto che:
- la regione 5' UTR del gene *BCL11A* è una regione particolarmente ricca in ripetizioni GC, che possono andare incontro ad errori di allineamento delle reads, dando luogo ad un risultato interpretabile come delezione di tre nucleotidi (c.-19_-17delGCC) (*Tabella 16, Figura 52*);
 - la regione 3' UTR del gene *TMPRSS6* è una regione considerata rumorosa che, quindi, può portare ad errori di allineamento dando luogo ad un risultato interpretabile come delezione di quattordici nucleotidi (c.*73_*87delACAAGTATTCTGGCG) (*Tabella 16, Figura 53*);
 - l'esone 1 del gene *TERF1* è caratterizzato dalla presenza di numerose ripetizioni GAG, che possono portare ad un non corretto allineamento delle reads, tanto da dar luogo ad un risultato interpretabile come una delezione di tre nucleotidi (c.186_188delGGA) (*Tabella 16, Figura 54*).

ID	Gene	Chr	Start	%VF	Exon	cDNA	Tx	Sample
#1	BCL11A	2	60780421	15 - 17.6	1	c.-19_-17delGCC	NM_018014	VP_run1: SG063 VP_run2: SG063TRIS
#2	TMPRSS6	22	37462032	20.3	19	c.*73_*87delACAAGTATTCTGGCG	NM_001289000	VP_run1:SG063
#3	TERF1	8	73921283	15.8 - 22.5	1	c.186_188delGGA	NM_003218	VP_run1: SG063 VP_run1: SG063-BIS VP_run2: SG063TRIS

Tabella 16. Approfondimento dei geni *BCL11A*, *TMPRSS6* e *TERF1*.

g. Riproducibilità intra-run e inter-run: come già accennato, l'esperimento di validazione ha previsto l'analisi di 4 replicati intra-run, 2 per ciascuna corsa: 1 / 1BIS e SG063 / SG063bis, nella prima corsa; 9 / 9BIS e UPN 94 / UPN 94bis nella seconda corsa. La *Figura 55* mostra una perfetta corrispondenza della frazione allelica di ciascuna variante identificata nei replicati, indicando un'ottima ripetibilità di risultati all'interno della stessa corsa.

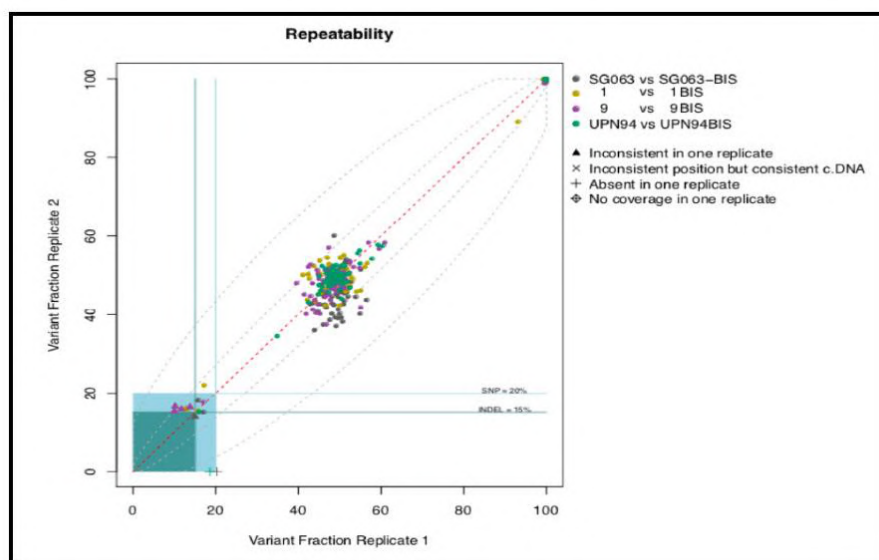


Figura 55. Grafico che mostra la corrispondenza tra le frazioni alleliche di ciascuna variante intra-run.

Inoltre, alcuni campioni sono stati replicati anche nelle due corse indipendenti: SG063 / SG063tris, UPN 62 / UPN 62bis, 2b / 2BIS e 6b / 6BIS (replicati inter-run). Anche in questo caso è stata confermata una ottimale riproducibilità, pur evidenziando un lieve discostamento delle frequenze alleliche in alcuni replicati (*Figura 56*). Fa eccezione una variante, identificate soltanto in uno dei due replicati, sia nel campione 2b e 6b.

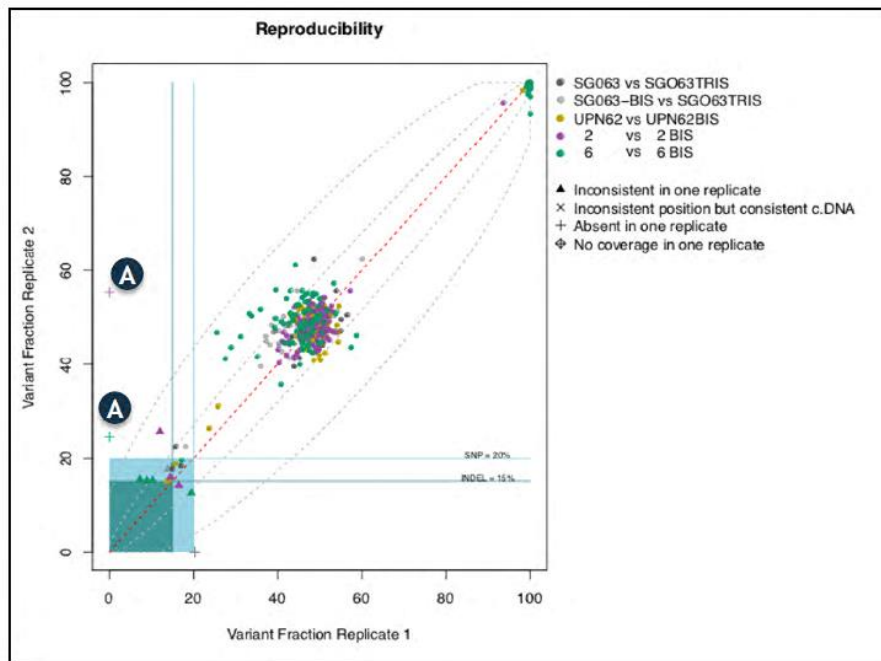


Figura 56. Grafico che mostra la corrispondenza tra le frazioni alleliche di ciascuna variante inter-run. Con il punto A è rappresentata la variante c.1842-6_1842-2dupCCCCA presente solo in uno dei replicati sia per il campione 2b e 6b.

Questa discrepanza si è verificata a livello di una regione del gene *TMPRSS6* ricca di sequenze ripetute TGGGG (*Figura 57*), che ha portato alla formazione di reverse reads provocando così un maggiore rumore di fondo.

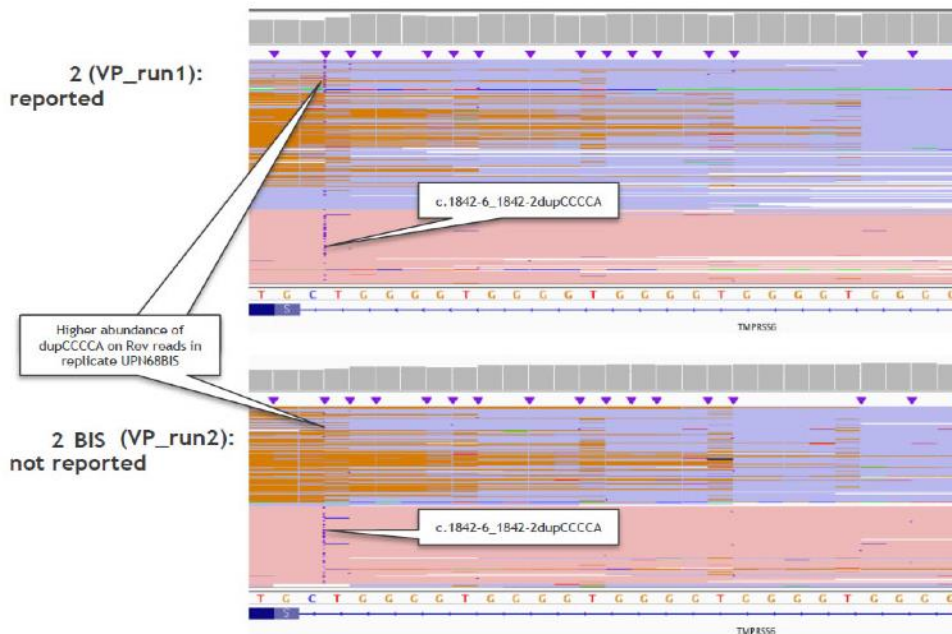


Figura 57. Gene *TMPRSS6*, regione ricca di sequenze ripetute nel 2b.

Complessivamente, i risultati sostanzialmente sovrapponibili fra i replicati intra- e inter-run hanno confermato una ottimale riproducibilità del dato genetico ottenuto, anche valutando la variabilità operatore-specifica che dovrà essere considerata nei futuri esperimenti.

13.2. Validazione dei genotipi dei campioni affetti da emoglobinopatie

La piattaforma Sophia DDM ha permesso di identificare ed annotare correttamente sei delle sette mutazioni a carico del gene β -globinico e l'identificazione di alterazioni a carico dei geni α -globinici, senza annotarle correttamente (*Tabella 17*). Nello specifico nei nove pazienti analizzati sono stati confermati sette degli otto genotipi β e quattro degli otto genotipi α (*Tabelle 7 e 17*).

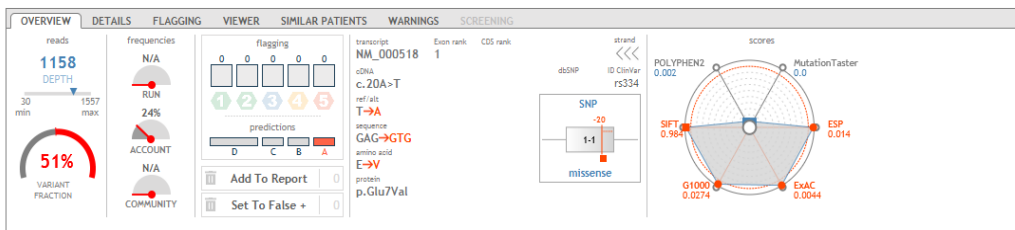
GENI	MUTAZIONI	IDENTIFICAZIONE
β	HbS (c.20 A>T)	✓
	cd 39 C>T (c.118 C>T)	✓
	-29 A>G (c.-79A>G)	✓
	cd 8 (-AA) (c.25_26delAA)	✓
	HbC (c.19 G>A)	✓
	IVS-II-745 C>G (c.316-106C>G)	✓
	Hb Lepore (NG_000007.3: g.63632_71046del)	*
α	anti ^{-3.7}	*
	α ^{-3.7} (NG_000006.1: g.34164_37967del3804)	*
	HBA2: c.2 T>C	*

Tabella 17. Identificazione mutazioni note dei geni globinici nei 9 campioni. La tabella mostra le mutazioni, presenti nei 9 pazienti a livello del gene α e β -globinico. Con il simbolo ✓ viene indicato che la mutazione è vista e riconosciuta dal pannello. Con il simbolo * viene indicato che il pannello vede un'alterazione genica che potrebbe essere ricondotta alla mutazione cercata, ma non riesce ad identificarla correttamente.

a. Nel campione 1b sono state confermate entrambe le mutazioni HbS e cd 39 C>T allo stato eterozigote. La profondità di lettura (depth) e la frazione allelica (variant fraction) sono riportate in *Figura 58*.

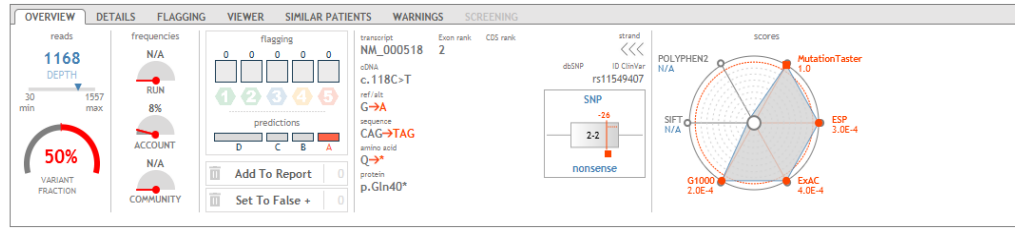
Il campione presenta anche la triplicazione del gene α, che è visibile nel pannello mediante l'utilizzo di un algoritmo specifico per l'identificazione delle copy number variation (CNV), ma non è possibile discriminare in quale, tra i due geni *HBA1-HBA2*, è avvenuta la triplicazione e nemmeno lo stato della triplicazione stessa. Come già dichiarato, questo campione è un replicato intra-run e in entrambi i replicati, 1b e 1BIS, sono stati ottenuti gli stessi risultati (*Figura 58*).

A

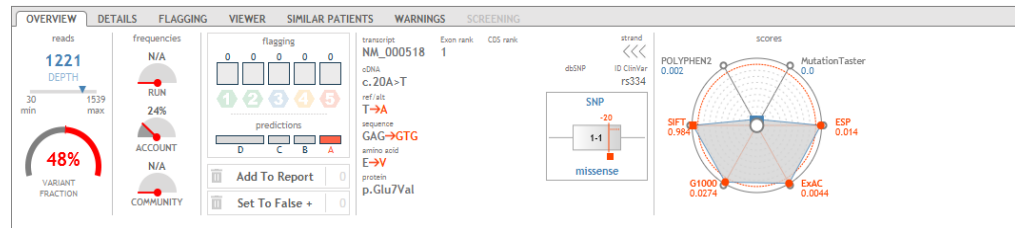


ExAC
ClinVar
COSMIC
IGV
NCBI
ALAMUT
Google
OMIM

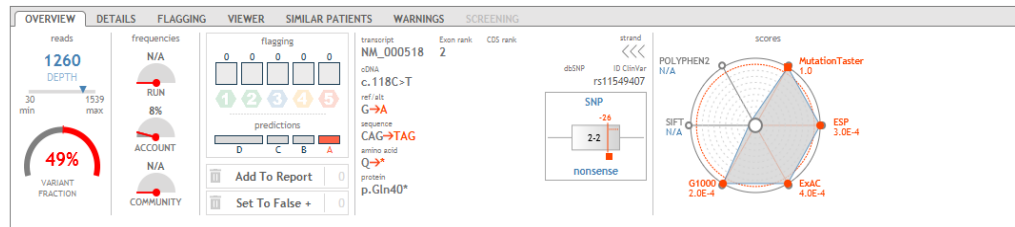
B



ExAC
ClinVar
COSMIC
IGV
NCBI
ALAMUT
Google
OMIM



ExAC
ClinVar
COSMIC
IGV
NCBI
ALAMUT
Google
OMIM

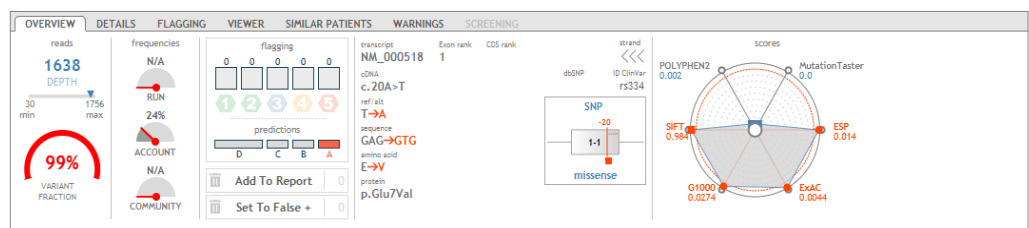


ExAC
ClinVar
COSMIC
IGV
NCBI
ALAMUT
Google
OMIM

Figura 58. Pannello A: il pannello mostra le mutazioni a livello del gene β -globinico per il campione 1b. **Pannello B:** il pannello mostra le mutazioni a livello del gene β -globinico per il campione 1BIS.

b. Nel campione 2b, è stata validata la presenza della mutazione HbS allo stato omozigote. La profondità di lettura (depth) e la frazione allelica (variant fraction) sono riportate in *Figura 59*. Mediante l'analisi delle CNV è stato possibile visualizzare una delezione a livello di un gene α , allo stato eterozigote. Anche in questo caso non è stato possibile discriminare in quale, tra i due geni *HBA1-HBA2*, sia presente tale delezione ed inoltre il sistema non permette di identificare la natura stessa della delezione. Come già dichiarato tale campione è un replicato inter run e in entrambi i replicati, 2 e 2BIS, sono stati ottenuti gli stessi risultati (*Figura 59*).

A



ExAC
ClinVar
COSMIC
IGV
NCBI
ALAMUT
Google
OMIM

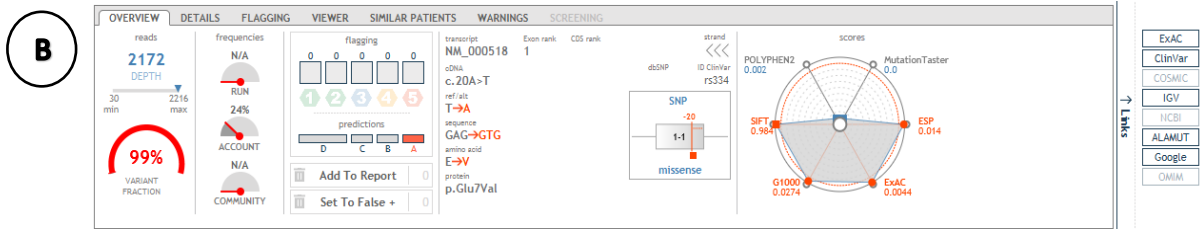


Figura 59. Pannello A: il pannello mostra la mutazione a livello del gene β -globinico per il campione 2b. **Pannello B:** il pannello mostra la mutazione a livello del gene β -globinico per il campione 2BIS.

- c. I campioni 3b e 4b che corrispondono ai campioni 3 e 4 del primo progetto di ricerca descritto in precedenza verranno riportati nel successivo paragrafo.
- d. Nel campione 5b è stata confermata la presenza della mutazione HbS allo stato omozigote (*Figura 60*). La profondità di lettura (depth) e la frazione allelica (variant fraction) sono riportate in *Figura 60*. Non è stato possibile analizzare le CNV per problemi derivanti dal DNA del campione, poiché nell'analisi è presente un elevato rumore di fondo; di conseguenza il genotipo α non può essere confermato.

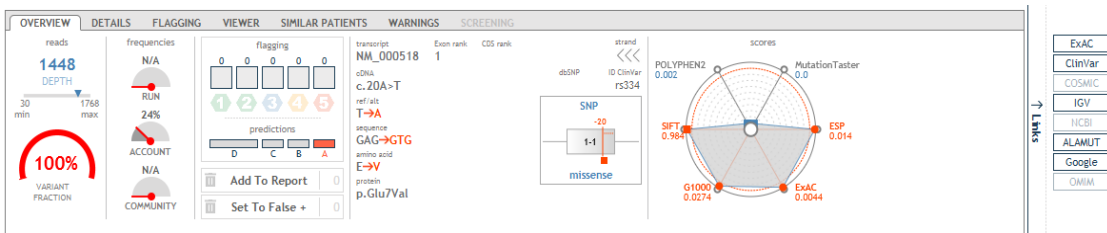
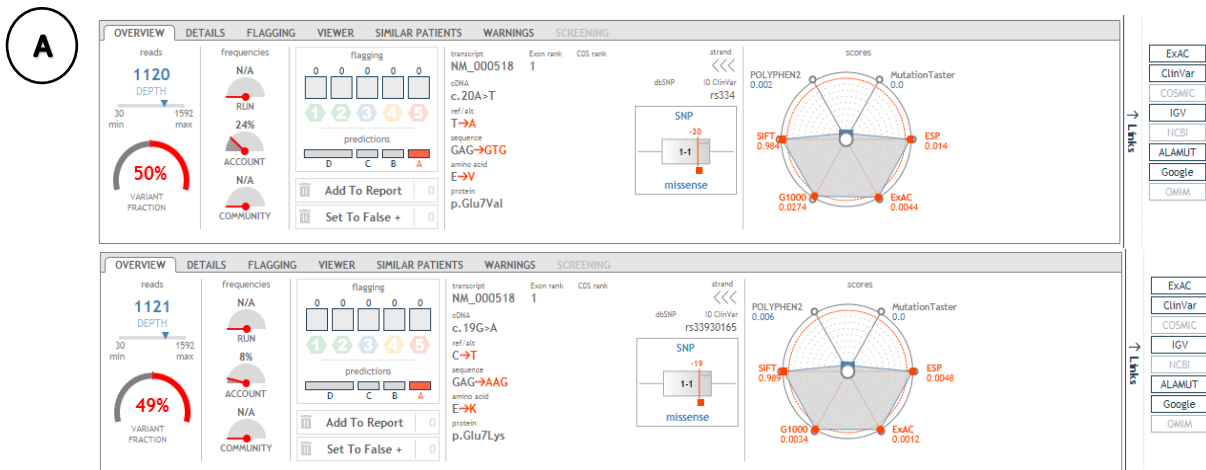


Figura 60. La figura mostra la mutazione a livello del gene β -globinico per il campione 5b.

- e. Per il campione 6b, il pannello ha permesso l'identificazione e quindi la validazione delle mutazioni HbS e HbC allo stato eterozigote, confermato il genotipo come quello di eterozigote composto. La profondità di lettura (depth) e la frazione allelica (variant fraction) sono riportate in *Figura 61*. Come già dichiarato, tale campione è un replicato inter run e in entrambi i replicati, 6b e 6BIS, sono stati ottenuti gli stessi risultati (*Figura 61*).



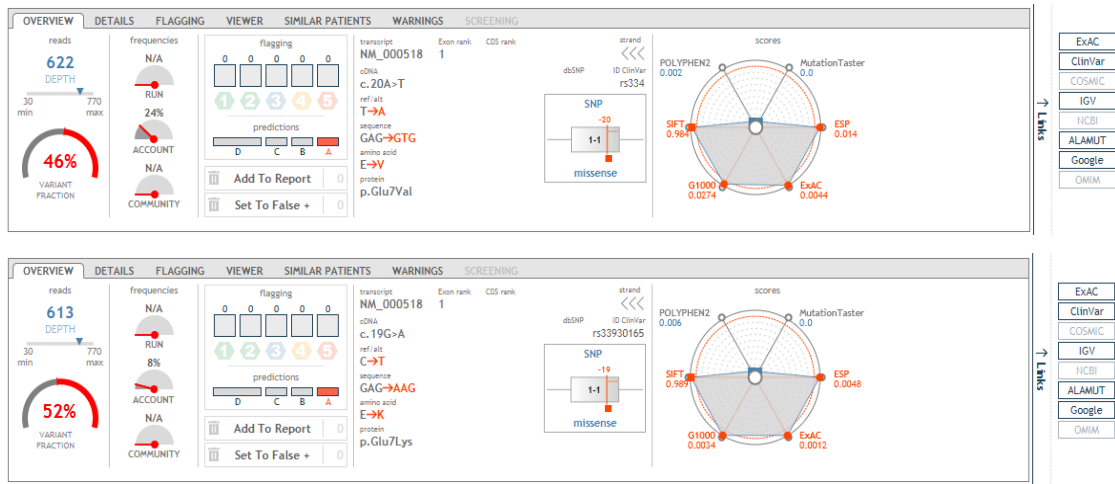
B

Figura 61. Pannello A: il pannello mostra le mutazioni a livello del gene β -globinico per il campione 6b. **Pannello B:** il pannello mostra le mutazioni a livello del gene β -globinico per il campione 6BIS.

- f. Nel campione 7b è stata confermata la presenza della mutazione IVS-II-745 C>G allo stato eterozigote (Figura 62). La profondità di lettura (depth) e la frazione allelica (variant fraction) sono riportate in *Figura 62*. Non è stato possibile validare la presenza del riarrangiamento Hb Lepore, poichè la piattaforma non ha identificato l'alterazione nello specifico, ma mostra soltanto la presenza di una delezione nel gene *HBB*, senza identificare la delezione nel gene *HBD*. Però, da un'analisi più approfondita mediante l'utilizzo del software integrative genomics viewer (IGV) è stato possibile visualizzare non solo la delezione di parte del gene *HBB* (esone 1-2), ma anche parte del gene *HBD* (esone 3); tale quadro risulta compatibile con l'Hb Lepore Boston-Washington (*Figura 63*). Un'analisi ulteriore a livello delle CNV non riporta alcuna alterazione a livello dei due geni.

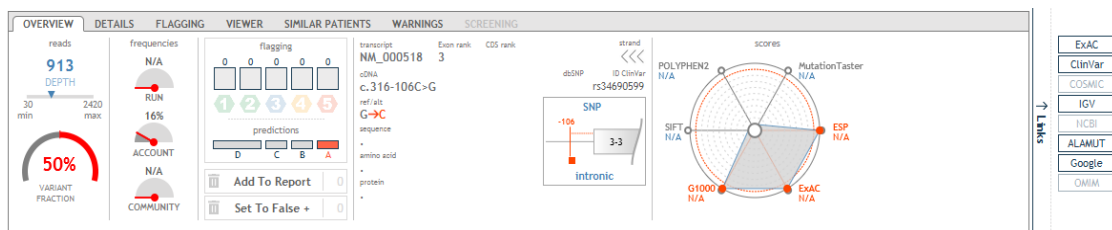


Figura 62. La figura mostra la mutazione a livello del gene β -globinico per il campione 7b.

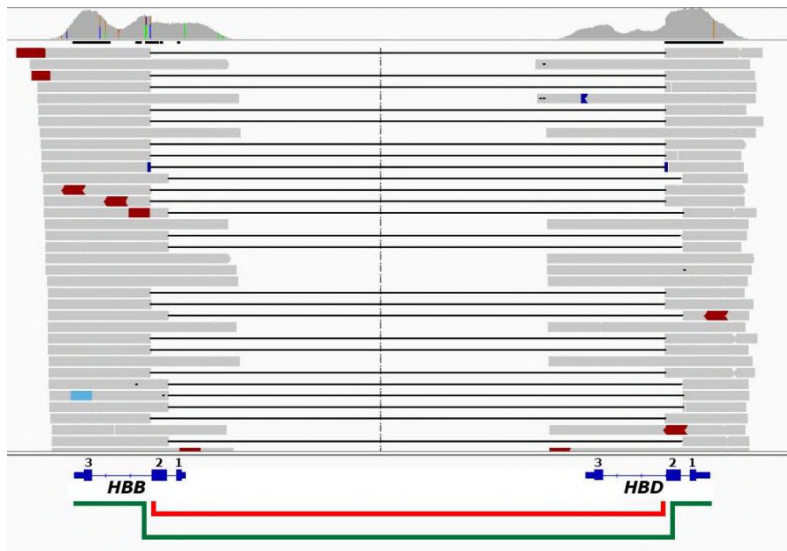


Figura 63. La figura mostra il riarrangiamento nel campione 7 a livello del gene *HBD* e *HBB*. In rosso è sottolineato il segmento che viene parzialmente perso; in verde è rappresentato il gene di fusione che presenta l'esone 1 e l'esone 2 del gene *HBD* e l'introne 2 e l'esone 3 del gene *HBB*.

- g. Nel campione 8b è stato possibile identificare la presenza della mutazione c.2T>C a livello del gene *HBA2*, (Figura 64). La profondità di lettura (depth) e la frazione allelica (variant fraction) sono riportate in Figura 64. A causa delle sequenze altamente ripetute dei geni α -globinici e ad alcune limitazioni legate al software di analisi, la mutazione potrebbe essere attribuita in maniera non specifica o al gene *HBA1* o al gene *HBA2*.

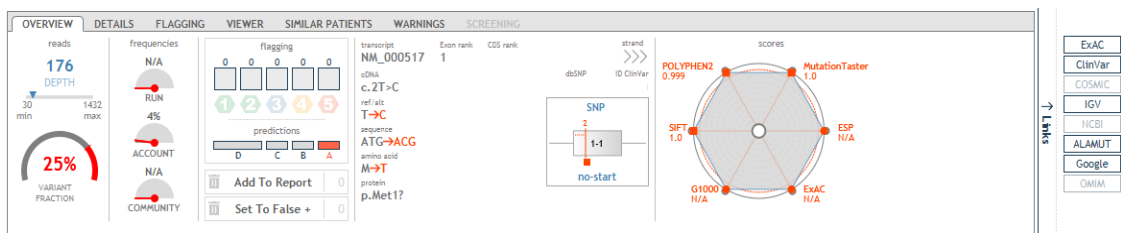
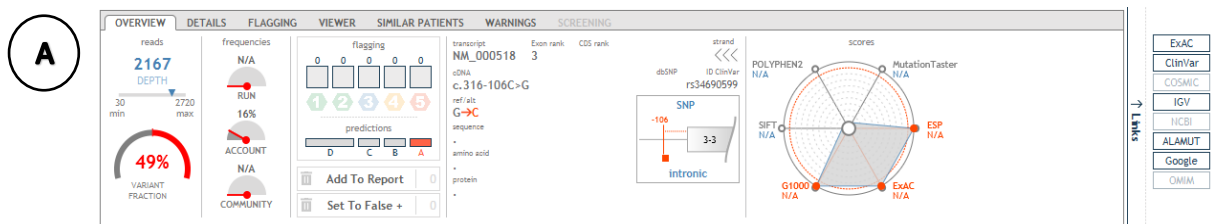


Figura 64. La figura mostra la mutazione a livello del gene α -globinico per il campione 8b.

- h. Nel campione 9b, presentante un trait β -talassemico, è stato possibile confermare la presenza della mutazione IVS-II-745C>G allo stato eterozigote. La profondità di lettura (depth) e la frazione allelica (variant fraction) sono riportate in Figura 65. Come già dichiarato, tale campione è un replicato intra run e in entrambi i replicati, 9 e 9 BIS, sono stati ottenuti gli stessi risultati (Figura 65).



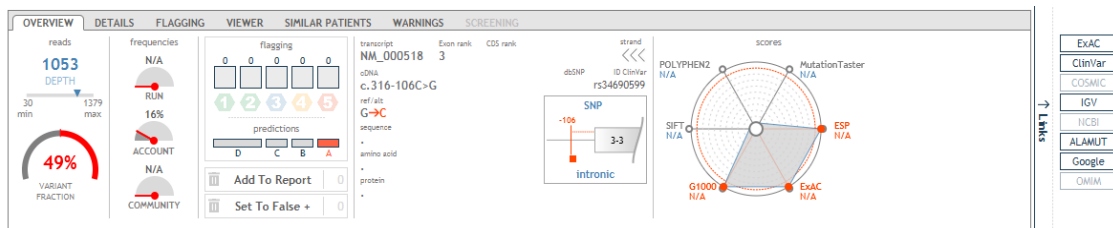
B

Figura 65. Pannello A: il pannello mostra la mutazione a livello del gene β -globinico per il campione 9b. **Pannello B:** il pannello mostra la mutazione a livello del gene β -globinico per il campione 9BIS.

13.3. Varianti identificate nei fratelli algerini e successivo studio familiare

Nel precedente progetto di ricerca sono stati descritti quali varianti legati ai geni modificatori del fenotipo clinico sono state riscontrate nei due fratelli algerini (pz 3-4, *Tabella 14*). In particolare è interessante notare come non sono presenti alterazioni nei geni alfa e come sono presenti le stesse alterazioni nei polimorfismi implicati nella modificazioni fenotipiche investigati. Inoltre è stato analizzato anche il promotore del gene *UGT1A*, in cui alterazioni delle ripetizioni TA presenti in numero superiore a 6 possono indurre la sindrome di Gilbert, che è nota essere un fattore modificatore delle complicanze secondarie in pazienti talassemici [183], i due fratelli sono risultati portatori di un allele con sette ripetizioni TA. Sono stati analizzati mediante reverse dot blot e sequenziamento del gene beta anche i genitori ed è stata identificata nella madre (pz 10b) la delezione cod.8 (-AA) in eterozigosi (*HBB:c.25_26del*/ wild type), mentre nel padre (pz 11b) la mutazione -29 A>G in eterozigosi (*HBB:c.-79A>G*/ wild type) (*Figura 66*).

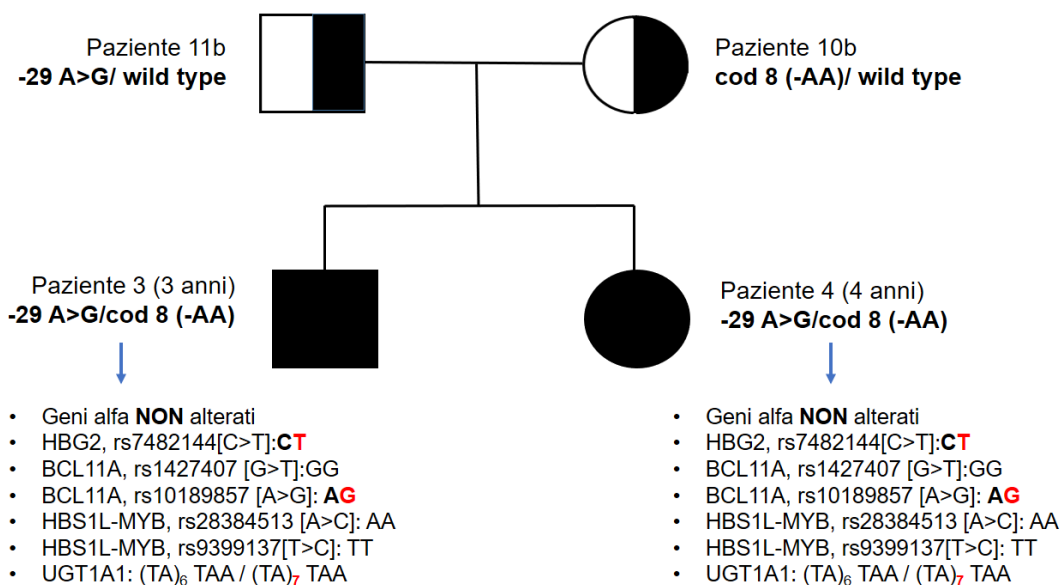


Figura 66. Caratterizzazione molecolare di alcuni modificatori genetici.

I due fratelli algerini (3b e 4b sono rispettivamente i campioni 3-4 del primo progetto di ricerca) affetti da talassemia major sono stati analizzati mediante il pannello NGS il quale ha permesso la conferma delle mutazioni note nel gene beta. In particolare:

- Nel campione 3 sono state identificate le mutazioni -29 A>G e cd 8 (-AA), entrambe allo stato eterozigote (*Figura 67*), confermando il genotipo talassemico. La profondità di lettura (depth) e la frazione allelica (variant fraction) sono riportate in *Figura 67*.

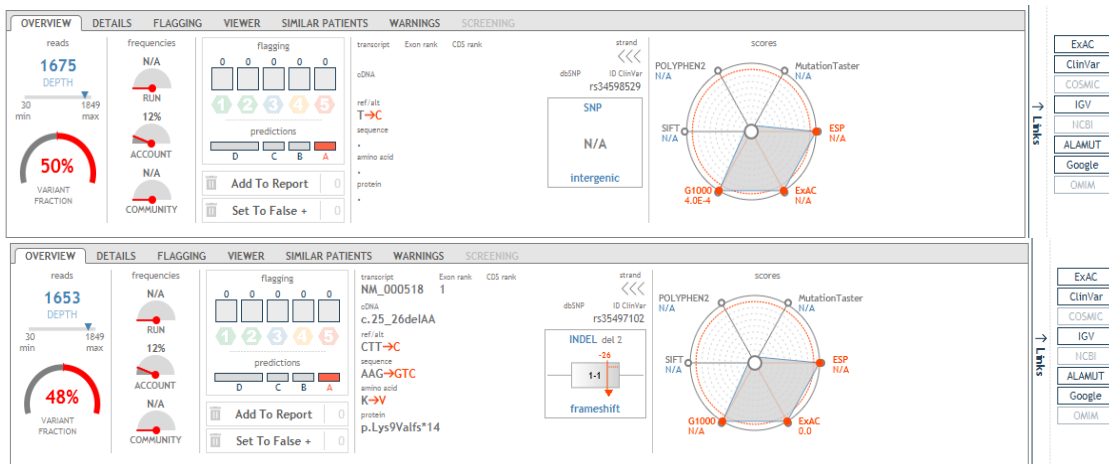


Figura 67. La figura mostra le mutazioni a livello del gene β -globinico per il campione 3.

- Il campione 4 presentava anchesso le mutazioni -29 A>G e cd 8 (-AA) allo stato eterozigote (*Figura 68*), confermando il genotipo talassemico. La profondità di lettura (depth) e la frazione allelica (variant fraction) sono riportate in *Figura 68*.

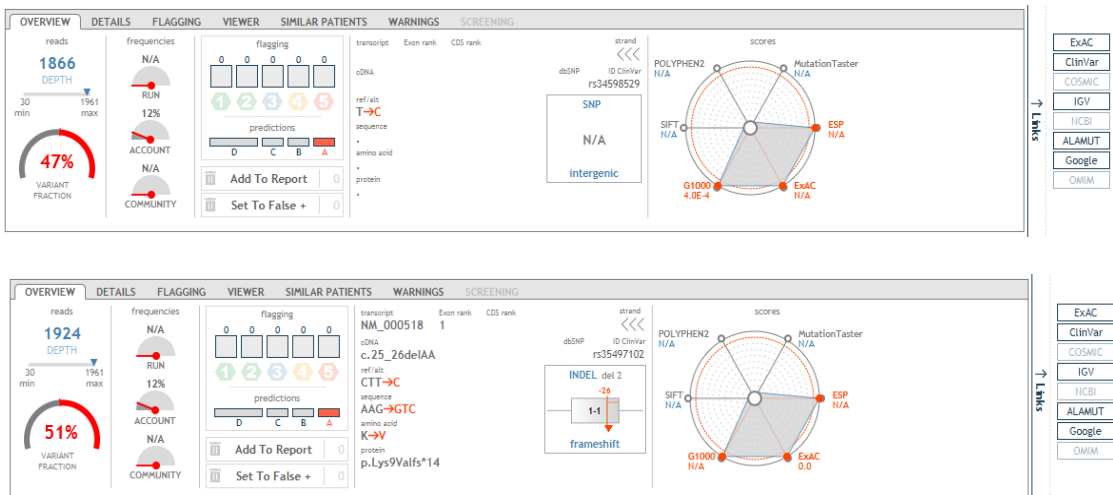


Figura 68. La figura mostra le mutazioni a livello del gene β -globinico per il campione 4.

L'analisi e l'interpretazione dei dati prodotti dagli esperimenti di NGS è stata eseguita utilizzando la piattaforma Sophia DDM, utilizzando l'algoritmo ILL1XG1G3, che ha permesso di individuare: 101 varianti retained, ovvero che soddisfanno i criteri di analisi posti, nel paziente 3 (appendice D) e 106 nel campione 4 (appendice E). In particolare nel campione 3 le varianti identificate sono state classificate come segue:

- Category A: 2 Altamente patogenetiche (*HBB:c.25_26del /HBB:c.-79A>G*)
- Category B: 1 Potenzialmente patogenetica (*HFE:c.187C>G* in eterozigosi)
- Category D: 93 Potenzialmente benigne
- Category C: 5 Incerto significato

Mentre nel campione 4 le varianti identificate sono state classificate come segue:

- Category A: 2 Altamente patogenetiche (*HBB:c.25_26del /HBB:c.-79A>G*)
- Category B: 2 Potenzialmente patogenetiche (*HFE:c.187C>G* in eterozigosi e *PINX1:c.63C>T* in eterozigosi)
- Category D: 98 Potenzialmente benigne
- Category C: 4 Incerto significato

È stata eseguita un'analisi incrociata tra i dati dei due fratelli per capire quali potessero essere le differenze sia in termini di differenti varianti individuate che in termini di diversità nell'assetto genotipico. Per eseguire tale analisi non solo sono stati consultati i dati delle varianti ritenute dal sistema, ma anche in alcuni casi le cosiddette varianti low confidence per capire se effettivamente una variante identificata solo in un paziente non fosse stata vista anche nell'altro ma non aveva passato tutti i criteri di qualità da noi introdotti con l'algoritmo di analisi scelto. Da tale analisi sono state identificate 95 varianti (2 varianti appartenenti alla category A: ovvero sono altamente patogenetiche; 1 variante B: Potenzialmente patogenetica, 3 varianti C: Incerto significato e 90 varianti D: Potenzialmente benigne), in entrambi i pazienti: di cui 76 varianti con lo stesso assetto genotipico mentre 19 varianti hanno un diverso assetto genotipico (tutte appartenenti alla category D, ovvero: Potenzialmente benigne) (*Tabella 18*).

Category	gene	type	codingConsequence	refGenome	chr	genome_position	exon_rank	c.DNA
A	<i>HBB</i>	INDEL	frameshift	GRCh37/hg19	11	5248225	1	c.25_26delAA
A	<i>HBB</i>	SNP	intergenic	GRCh37/hg19	11	5248330	0	
B	<i>HFE</i>	SNP	missense	GRCh37/hg19	6	26091179	2	c.187C>G
C	<i>DAXX</i>	SNP	missense	GRCh37/hg19	6	33287796	5	c.1457C>G
C	<i>WRAP53</i>	SNP	missense	GRCh37/hg19	17	7592543	3	c.433G>A
C	<i>BMP6</i>	SNP	synonymous	GRCh37/hg19	6	7862556	4	c.1029C>T
D	<i>HFE</i>	SNP	splice_donor_+4	GRCh37/hg19	6	26091336	2	c.340+4T>C
D	<i>BCL11A</i>	SNP	3'UTR	GRCh37/hg19	2	60679226	5	c.*484C>T
D	<i>BCL11A</i>	SNP	synonymous	GRCh37/hg19	2	60687959	4	c.2088T>C
D	<i>BCL11A</i>	INDEL	5'UTR	GRCh37/hg19	2	60780421	1	c.-19_-17delGCC (presente anche nel pz 3 ma nelle Low Confidence variants)
D	<i>BMP2</i>	SNP	synonymous	GRCh37/hg19	20	6751034	2	c.261A>G
D	<i>BMP2</i>	INDEL	3'UTR	GRCh37/hg19	20	6759758	3	c.*41_*42delAT (presente anche nel pz 3 ma nelle Low Confidence variants)
D	<i>BMP2</i>	SNP	missense	GRCh37/hg19	20	6759115	3	c.570A>T
D	<i>BMP2</i>	SNP	3'UTR	GRCh37/hg19	20	6760344	3	c.*608A>G
D	<i>BMP6</i>	SNP	synonymous	GRCh37/hg19	6	7862631	4	c.1104G>C
D	<i>BMP6</i>	INDEL	3'UTR	GRCh37/hg19	6	7881853	7	c.*1277_*1289del TTGTGTGTACAA GinsCC
D	<i>BMP6</i>	SNP	3'UTR	GRCh37/hg19	6	7881931	7	c.*1355T>C
D	<i>BMP6</i>	SNP	3'UTR	GRCh37/hg19	6	7881311	7	c.*735A>G
D	<i>ERFE</i>	INDEL	3'UTR	GRCh37/hg19	2	23907746 3	8	c.*1773delT

D	<i>ERFE</i>	SNP	3'UTR	GRCh3 7/hg19	2	23907593 5	8	c.*240A>G
D	<i>FTL</i>	SNP	synonymous	GRCh3 7/hg19	19	49469087	2	c.163T>C
D	<i>HBA2</i>	SNP	3'UTR	GRCh3 7/hg19	16	223706	3	c.*107A>G
D	<i>HBB</i>	SNP	intronic	GRCh3 7/hg19	11	5247733	2	c.315+74T>G
D	<i>HBB</i>	SNP	intronic	GRCh3 7/hg19	11	5247791	2	c.315+16G>C
D	<i>HBB</i>	SNP	intronic	GRCh3 7/hg19	11	5247141	3	c.316-185C>T
D	<i>HBB</i>	SNP	synonymous	GRCh3 7/hg19	11	5248243	1	c.9T>C
D	<i>HBB</i>	SNP	intronic	GRCh3 7/hg19	11	5247726	2	c.315+81C>T
D	<i>SLC11A2</i>	SNP	missense	GRCh3 7/hg19	12	51402905	1	c.17A>C
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh3 7/hg19	12	51374748	17	c.*761C>A
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh3 7/hg19	12	51375066	17	c.*443C>T
D	<i>SLC11A2</i>	SNP	intronic	GRCh3 7/hg19	12	51420164	1	c.49+1796G>T
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh3 7/hg19	12	51374250	17	c.*1259T>C
D	<i>SLC11A2</i>	SNP	intronic	GRCh3 7/hg19	12	51380232	16	c.1629+1933T>C
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh3 7/hg19	12	51373742	17	c.*1767G>A
D	<i>SLC11A2</i>	SNP	intronic	GRCh3 7/hg19	12	51381077	16	c.1629+1088G>A
D	<i>SLC11A2</i>	SNP	intronic	GRCh3 7/hg19	12	51419621	1	c.49+2339A>G
D	<i>SLC40A1</i>	INDEL	5'UTR	GRCh3 7/hg19	2	19044549 5	1	c.-312_- 310dupCGG (presente anche nel pz 3 ma nelle Low Confidence variants)
D	<i>TMPRSS 6</i>	SNP	5'UTR	GRCh3 7/hg19	22	37505552	2	c.-11G>A
D	<i>TMPRSS 6</i>	SNP	synonymous	GRCh3 7/hg19	22	37469591	13	c.1536C>T
D	<i>TMPRSS 6</i>	SNP	synonymous	GRCh3 7/hg19	22	37462926	18	c.2256C>T
D	<i>TMPRSS 6</i>	SNP	synonymous	GRCh3 7/hg19	22	37480797	9	c.1056G>A
D	<i>TMPRSS 6</i>	INDEL	splice_accep tor_indel	GRCh3 7/hg19	22	37465385	16	c.1842-6_1842- 2delCCCCA *
D	<i>ANKRD2 6</i>	SNP	missense	GRCh3 7/hg19	10	27389197	1	c.59A>G
D	<i>BMP6</i>	SNP	missense	GRCh3 7/hg19	6	7727271	1	c.83G>A

D	<i>COX4I1</i>	SNP	synonymous	GRCh37/hg19	16	85838653	3	c.184T>C
D	<i>CTC1</i>	SNP	splice_donor_+3	GRCh37/hg19	17	8141707	3	c.435+3A>G
D	<i>CTC1</i>	SNP	synonymous	GRCh37/hg19	17	8134785	15	c.2478A>G
D	<i>CTC1</i>	SNP	missense	GRCh37/hg19	17	8135061	14	c.2458A>G
D	<i>DAXX</i>	SNP	synonymous	GRCh37/hg19	6	33287183	6	c.1914A>G
D	<i>DCLRE1B</i>	SNP	missense	GRCh37/hg19	1	114448389	1	c.181C>T
D	<i>DCLRE1B</i>	SNP	synonymous	GRCh37/hg19	1	114449662	2	c.234T>C
D	<i>DDX41</i>	SNP	synonymous	GRCh37/hg19	5	176940384	10	c.822C>T
D	<i>DKC1</i>	SNP	synonymous	GRCh37/hg19	X	153994596	5	c.369G>T
D	<i>FUS</i>	SNP	synonymous	GRCh37/hg19	16	31195279	4	c.288C>T
D	<i>GAR1</i>	SNP	synonymous	GRCh37/hg19	4	110737389	2	c.69T>C
D	<i>GATA2</i>	SNP	synonymous	GRCh37/hg19	3	128205860	3	c.15C>G
D	<i>GATA2</i>	SNP	missense	GRCh37/hg19	3	128204951	4	c.490G>A
D	<i>GNAS</i>	SNP	synonymous	GRCh37/hg19	20	57478807	5	c.393C>T
D	<i>GNAS</i>	SNP	synonymous	GRCh37/hg19	20	57415812	1	c.651T>A
D	<i>LIG4</i>	SNP	synonymous	GRCh37/hg19	13	108861913	2	c.1704T>C
D	<i>NAF1</i>	SNP	missense	GRCh37/hg19	4	164048199	8	c.1102C>G
D	<i>NAF1</i>	SNP	missense	GRCh37/hg19	4	164085425	2	c.484A>G
D	<i>NBN</i>	SNP	missense	GRCh37/hg19	8	90990479	6	c.307G>C
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90958422	14	c.1770A>G
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90967711	11	c.951T>C
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90995019	2	c.102G>A
D	<i>NSD1</i>	SNP	synonymous	GRCh37/hg19	5	176721198	23	c.6829T>C
D	<i>PIF1</i>	SNP	synonymous	GRCh37/hg19	15	65116328	2	c.207T>C
D	<i>NAF1</i>	SNP	missense	GRCh37/hg19	4	164048199	8	c.1102C>G
D	<i>NAF1</i>	SNP	missense	GRCh37/hg19	4	164085425	2	c.484A>G
D	<i>NBN</i>	SNP	missense	GRCh37/hg19	8	90990479	6	c.307G>C
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90958422	14	c.1770A>G

D	<i>NBN</i>	SNP	synonymous	GRCh37/ hg19	8	90967711	11	c.951T>C
D	<i>NBN</i>	SNP	synonymous	GRCh37/ hg19	8	90995019	2	c.102G>A
D	<i>NSD1</i>	SNP	synonymous	GRCh37/ hg19	5	17672119 8	23	c.6829T>C
D	<i>PIF1</i>	SNP	synonymous	GRCh37/ hg19	15	65116328	2	c.207T>C
D	<i>RECQL4</i>	INDEL	intergenic	GRCh37/ hg19	8	14573876 7	15	c.2296+1del
D	<i>RECQL4</i>	SNP	synonymous	GRCh37/ hg19	8	14573763 6	19	c.3127T>C
D	<i>RECQL4</i>	SNP	missense	GRCh37/ hg19	8	14574251 4	4	c.274T>C
D	<i>RIF1</i>	SNP	missense	GRCh37/ hg19	2	15233141 8	35	c.7174C>G
D	<i>RIF1</i>	SNP	missense	GRCh37/ hg19	2	15232011 8	30	c.4084G>A
D	<i>RIF1</i>	SNP	missense	GRCh37/ hg19	2	15232209 5	30	c.6061A>T
D	<i>RIF1</i>	SNP	missense	GRCh37/ hg19	2	15231157 0	22	c.2506G>A
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/ hg19	20	62320968	23	c.1992T>C
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/ hg19	20	62322288	27	c.2544T>C
D	<i>RTEL1</i>	SNP	missense	GRCh37/ hg19	20	62326110	32	c.3126A>C
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/ hg19	20	62321655	26	c.2274G>A
D	<i>STN1</i>	SNP	missense	GRCh37/ hg19	10	10565731 6	7	c.743C>G
D	<i>STN1</i>	SNP	missense	GRCh37/ hg19	10	10565982 6	5	c.451A>G
D	<i>TERF1</i>	SNP	synonymous	GRCh37/ hg19	8	73958310	9	c.1198T>C
D	<i>TERF1</i>	INDEL	inframe_3	GRCh37/ hg19	8	73921283	1	c.186_188dup
D	<i>TERF1</i>	SNP	synonymous	GRCh37/ hg19	8	73958240	9	c.1128G>A
D	<i>TMPRSS6</i>	SNP	missense	GRCh37/ hg19	22	37462936	18	c.2246T>C
D	<i>TMPRSS6</i>	SNP	missense	GRCh37/ hg19	22	37485724	7	c.730A>G
D	<i>TPP1</i>	SNP	synonymous	GRCh37/ hg19	11	6638599	5	c.441A>G

Tabella 18. 95 varianti comuni ai due fratelli. In blu sono identificate quelle per le quali i due fratelli hanno un diverso assetto genotipico. *Variante identificata dal software di analisi come Falso Positivo.

A differenza solo nel paziente 3 sono state identificate 17 varianti (*Tabella 19*), in particolare:

- 2 classificate come Category C, ovvero varianti ad Incerto significato
- 15 classificate come Category D, ovvero Potenzialmente benigne

Category	gene	type	Coding Consequence	Ref Genome	chr	genome_position	depth	Var %	exon rank	c.DNA
C	<i>PINX1</i>	SNP	synonymous	GRCh37/hg19	8	10623289	1138	50,3	7	c.609T>C
C	<i>RIF1</i>	SNP	missense	GRCh37/hg19	2	1,52E+08	1493	49,7	30	c.5363T>C
D	<i>PINX1</i>	SNP	synonymous	GRCh37/hg19	8	10623214	1192	51,4	7	c.684C>G
D	<i>RIF1</i>	SNP	synonymous	GRCh37/hg19	2	1,52E+08	1017	46,5	27	c.3093G>A
D	<i>CBL</i>	INDEL	inframe_3	GRCh37/hg19	11	1,19E+08	907	46,6	1	c.125_127dup
D	<i>CEBPA</i>	INDEL	inframe_6	GRCh37/hg19	19	33792731	728	50,3	1	c.227_232dup
D	<i>ERFE</i>	SNP	3'UTR	GRCh37/hg19	2	2,39E+08	1481	52,3	8	c.*889G>C
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90955583	726	45,3	15	c.1836T>G
D	<i>NSD1</i>	SNP	synonymous	GRCh37/hg19	5	1,77E+08	1849	48,6	23	c.6903G>C
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/hg19	20	62321189	921	52,4	24	c.2112C>T
D	<i>SLC40A1</i>	SNP	3'UTR	GRCh37/hg19	2	1,9E+08	1630	48,5	8	c.*1130A>G
D	<i>SRP72</i>	SNP	synonymous	GRCh37/hg19	4	57356558	1541	45,8	12	c.1197T>C
D	<i>SRP72</i>	SNP	synonymous	GRCh37/hg19	4	57333822	1342	44	1	c.21G>T
D	<i>TEN1</i>	SNP	synonymous	GRCh37/hg19	17	73987652	729	99,9	3	c.198C>T
D	<i>TINF2</i>	SNP	synonymous	GRCh37/hg19	14	24709506	1274	48,7	7	c.1092G>A
D	<i>TMPRSS6</i>	SNP	intronic	GRCh37/hg19	22	37499692	1102	99,5	2	c.-1-234T>C
D	<i>TMPRSS6</i>	SNP	synonymous	GRCh37/hg19	22	37499565	1103	48,7	1	c.15C>T

Tabella 19. Varianti specifiche del paziente 3.

Mentre solo nel paziente 4 sono state identificate 19 varianti presenti (*Tabella 20*), in particolare:

- 1 classificata come Category B, ovvero Potenzialmente patogenetica
- 1 classificata come Category C, ovvero Incerto significato
- 17 classificate come Category D, ovvero Potenzialmente benigne

category	gene	type	Coding Consequence	Ref Genome	chr	genome_position	depth	Var %	Exon rank	c.DNA
B	<i>PINX1</i>	SNP	synonymous	GRCh37/hg19	8	10692242	794	47,2	2	c.63C>T
C	<i>MBD4</i>	SNP	missense	GRCh37/hg19	3	129155414	1360	47,6	3	c.1073T>C
D	<i>CBL</i>	SNP	synonymous	GRCh37/hg19	11	119170362	1547	50,6	16	c.2592C>T
D	<i>FUS</i>	SNP	synonymous	GRCh37/hg19	16	31193942	1270	47,7	3	c.147C>A
D	<i>GNAS</i>	SNP	missense	GRCh37/hg19	20	57428474	1858	48,4	1	c.154G>A

D	RECQL4	SNP	missense	GRCh37/hg19	8	145740375	1126	46,8	9	c.1565G>A
D	RECQL4	SNP	synonymous	GRCh37/hg19	8	145737086	1178	50,3	21	c.3480G>A
D	RECQL4	SNP	missense	GRCh37/hg19	8	145741702	1122	50,8	5	c.801G>C
D	RECQL4	SNP	synonymous	GRCh37/hg19	8	145742879	822	46,3	3	c.132A>G
D	RTEL1	SNP	synonymous	GRCh37/hg19	20	62319085	1253	46,6	17	c.1443G>A
D	SAMD9	SNP	synonymous	GRCh37/hg19	7	92734451	1357	51,8	2	c.960T>C
D	SAMD9	SNP	missense	GRCh37/hg19	7	92734983	1590	50,3	2	c.428T>C
D	SAMD9	SNP	missense	GRCh37/hg19	7	92734065	1475	46,2	2	c.1346A>G
D	SAMD9L	SNP	synonymous	GRCh37/hg19	7	92764982	1539	45,7	5	c.303G>A
D	SLC40A1	SNP	synonymous	GRCh37/hg19	2	190430177	1174	44,5	6	c.663T>C
D	STN1	SNP	synonymous	GRCh37/hg19	10	105657363	976	48,5	7	c.696G>C
D	TMPRSS6	SNP	synonymous	GRCh37/hg19	22	37499386	1317	43	2	c.72G>A
D	TERT	SNP	synonymous	GRCh37/hg19	5	1255520	1254	50,2	13	c.2850C>T
D	TERT	SNP	synonymous	GRCh37/hg19	5	1294086	1484	50,9	2	c.915G>A

Tabella 20. Varianti specifiche del paziente 4.

Al fine di poter analizzare le varianti rare identificate nei due fratelli sono stati applicati altri filtri alla analisi dei dati prodotti, ovvero frequenza sia in 1000 genomi che in ExAc inferiore a 0.01. Da tale analisi sono stati ottenuti i seguenti risultati:

- Nel paziente 3 sono presenti 13 alterazioni retained rare di cui: 2 varianti altamente patogeniche (category A), 3 varianti ad incerto significato (category C) e 9 potenzialmente benigne (category D) (Tabella 21)

Cat	gene	type	Coding Cons	depth	Var %	c.DNA	dbSNP	g1000	ExAC
A	HBB	INDEL	frameshift	1653	48,9	c.25_26del	rs35497102		0
A	HBB	SNP	intergenic	1675	50,6		rs34598529	0,0004	
C	DAXX	SNP	missense	1142	48,6	c.1457C>G	rs146304558	0,0086	0,0092
C	RIF1	SNP	missense	1493	49,7	<u>c.5363T>C</u>			
C	WRAP53	SNP	missense	1503	49,6	c.433G>A	rs147226406	0,0002	0
D	BMP6	INDEL	3'UTR	1360	42,4	c.*1277_*1289delinsCC	rs71548169		
D	CBL	INDEL	inframe_3	907	46,6	<u>c.125_127dup</u>	rs397507488	0,0056	0,0009
D	CTC1	SNP	synonymous	1764	51,6	c.2478A>G	rs200643329	0,0008	0,0011
D	GNAS	SNP	synonymous	1485	53,1	c.651T>A	rs75176432	0,0052	0,0056
D	SRP72	SNP	synonymous	1541	45,8	<u>c.1197T>C</u>	rs73818109	0,0072	0,0026
D	TERF1	SNP	synonymous	858	50,9	c.1128G>A	rs34194765	0,0074	0,0024
D	TERF1	SNP	synonymous	892	50,3	c.1198T>C	rs34368910	0,0074	0,0024
D	TINF2	SNP	synonymous	1274	48,7	<u>c.1092G>A</u>	rs184422577	0,0024	0,0049

Tabella 21. Varianti del paziente 3 con frequenza di popolazione in 1000 genomi e Exac <0.01. Le varianti sottolineate sono quelle specifiche del paziente 3.

- Nel paziente 4 sono presenti 11 alterazioni rare di cui: 2 altamente patogeniche (category A), 2 varianti ad incerto significato (category C) e 7 potenzialmente benigne (category D) (*Tabella 22*)

Cat	gene	type	Coding Cons	depth	Var %	c.DNA	dbSNP	g1000	ExAC
A	<i>HBB</i>	INDEL	frameshift	1924	51,6	c.25_26del	rs35497102		0
A	<i>HBB</i>	SNP	intergenic	1866	47,9		rs34598529	0,0004	
C	<i>DAXX</i>	SNP	missense	1280	45,6	c.1457C>G	rs146304558	0,0086	0,0092
C	<i>WRAP53</i>	SNP	missense	1777	49,7	c.433G>A	rs147226406	0,0002	0
D	<i>BMP6</i>	INDEL	3'UTR	1531	43,6	c.*1277_*1289delinsCC	rs71548169		
D	<i>CTC1</i>	SNP	synonymous	1916	47,5	c.2478A>G	rs200643329	0,0008	0,0011
D	<i>GNAS</i>	SNP	missense	1858	48,4	<u>c.154G>A</u>	rs527488103	0,0002	0,0001
D	<i>GNAS</i>	SNP	synonymous	1555	49	c.651T>A	rs75176432	0,0052	0,0056
D	<i>RTEL1</i>	SNP	synonymous	1253	46,6	<u>c.1443G>A</u>	rs762716598		0
D	<i>TERF1</i>	SNP	synonymous	916	47,2	c.1198T>C	rs34368910	0,0074	0,0024
D	<i>TERF1</i>	SNP	synonymous	914	48	c.1128G>A	rs34194765	0,0074	0,0024

Tabella 22. Varianti del paziente 4 con frequenza di popolazione in 1000 genomi e Exac <0.01. Le varianti sottolineate sono quelle specifiche del paziente 4.

Nei due soggetti sono stati analizzati anche i dati prodotti dall'algorithmo per l'identificazione di eventuali CNVs. Da tale analisi sono state escluse le seguenti regioni a causa del livello di rumore tipicamente elevato *ANKRD26_ex5*, *HBB_up* e *GNAS_in1_2*. Inoltre le CNVs frequenti in *HBA1_ex1-3* e *HBA2_ex1-3* possono limitare la precisione del rilevamento in queste regioni. In particolare entrambi i fratelli presentano un basso rumore di fondo che ha permesso una precisa analisi (*Figura 69*).

Table 1.2: Sample statistics

Sample	Average coverage per region	Residual noise	Noise status	Sex
3	1176	0.052	low-noise	male
4	1358	0.047	low-noise	female

Figura 69. Statiche rilevate dall' algoritmo per le CNVs.

Nell'appendice F è stato riportato il file prodotto dall'analisi per il rilevamento delle CNVs nei pazienti inclusi nella prima run di validazione di 24 campioni dove erano inclusi i due fratelli (UPN 69 ovvero pz 3 ed UPN70 ovvero pz 4); i grafici mostrano il riepilogo dei risultati del rilevamento CNVs per ciascun campione e per ciascuna regione target, ricordando che alcuni di essi sono stati utilizzati come campioni di riferimento per la necessaria normalizzazione. In tali grafici i quadrati blu corrispondono alle regioni target senza CNV, i quadrati rossi alle delezioni, i quadrati arancioni a duplicazioni. I punti neri indicano le regioni target con stato CNV indeterminato. I quadrati vuoti segnano le regioni con una confidenza ridotta del rilevamento CNV ("confidenza media"). Viene indicato il numero di copia dedotto all'interno di ogni quadrato (se può essere determinato in modo univoco). I campioni rifiutati sono contrassegnati da due asterischi. I campioni con rumore medio sono contrassegnati da un asterisco. A causa dell'elevato numero di campioni in questa corsa (24), sono stati suddivisi in 2 gruppi tracciati separatamente. In particolare tutte le regioni analizzate nei due fratelli non hanno riscontrato alterazioni in termini di CNVs (appendice F). Tutte le alterazioni identificate nei fratelli sono state identificate anche nei dati di sequenziamento prodotti da un esperimento indipendente di NGS con lo stesso pannello nei genitori (madre: pz 10 b e padre: pz 11b). Di seguito viene riportato l'assetto genotipico identificato nei genitori sia per le varianti comuni ai due fratelli nella *Tabella 23* che per le varianti specifiche di ogni patria nella *Tabella 24*.

Cat	gene	Ref Genome	chr	genome_position	c.DNA	Allele Madre	Allele Madre	Allele Padre	Allele Padre
A	<i>HBB</i>	GRCh37/hg19	11	5248225	c.25_26delAA				
A	<i>HBB</i>	GRCh37/hg19	11	5248330					
B	<i>HFE</i>	GRCh37/hg19	6	26091179	c.187C>G				
C	<i>DAXX</i>	GRCh37/hg19	6	33287796	c.1457C>G				
C	<i>WRAP53</i>	GRCh37/hg19	17	7592543	c.433G>A				

C	<i>BMP6</i>	GRCh37/ hg19	6	7862556	c.1029C>T				
D	<i>HFE</i>	GRCh37/ hg19	6	26091336	c.340+4T>C				
D	<i>BCL11A</i>	GRCh37/ hg19	2	60679226	c.*484C>T				
D	<i>BCL11A</i>	GRCh37/ hg19	2	60687959	c.2088T>C				
D	<i>BCL11A</i>	GRCh37/ hg19	2	60780421	c.-19_- 17delGCC (presente anche nel pz 3 ma nelle Low Confidence variants)	/	/	/	/
D	<i>BMP2</i>	GRCh37/ hg19	20	6751034	c.261A>G				
D	<i>BMP2</i>	GRCh37/ hg19	20	6759758	c.*41_*42delA T (presente anche nel pz 3 ma nelle Low Confidence variants)	/	/	/	/
D	<i>BMP2</i>	GRCh37/ hg19	20	6759115	c.570A>T				
D	<i>BMP2</i>	GRCh37/ hg19	20	6760344	c.*608A>G				
D	<i>BMP6</i>	GRCh37/ hg19	6	7862631	c.1104G>C				
D	<i>BMP6</i>	GRCh37/ hg19	6	7881853	c.*1277_*1289 delTTGTGTG TACAAGinsC C				
D	<i>BMP6</i>	GRCh37/ hg19	6	7881931	c.*1355T>C				
D	<i>BMP6</i>	GRCh37/ hg19	6	7881311	c.*735A>G				
D	<i>ERFE</i>	GRCh37/ hg19	2	239077463	c.*1773delT				
D	<i>ERFE</i>	GRCh37/ hg19	2	239075935	c.*240A>G				
D	<i>FTL</i>	GRCh37/ hg19	19	49469087	c.163T>C				
D	<i>HBA2</i>	GRCh37/ hg19	16	223706	c.*107A>G				
D	<i>HBB</i>	GRCh37/ hg19	11	5247733	c.315+74T>G				
D	<i>HBB</i>	GRCh37/ hg19	11	5247791	c.315+16G>C				
D	<i>HBB</i>	GRCh37/ hg19	11	5247141	c.316-185C>T				

D	<i>HBB</i>	GRCh37/ hg19	11	5248243	c.9T>C				
D	<i>HBB</i>	GRCh37/ hg19	11	5247726	c.315+81C>T				
D	<i>SLC11A2</i>	GRCh37/ hg19	12	51402905	c.17A>C				
D	<i>SLC11A2</i>	GRCh37/ hg19	12	51374748	c.*761C>A				
D	<i>SLC11A2</i>	GRCh37/ hg19	12	51375066	c.*443C>T				
D	<i>SLC11A2</i>	GRCh37/ hg19	12	51420164	c.49+1796G> T				
D	<i>SLC11A2</i>	GRCh37/ hg19	12	51374250	c.*1259T>C				
D	<i>SLC11A2</i>	GRCh37/ hg19	12	51380232	c.1629+1933T >C				
D	<i>SLC11A2</i>	GRCh37/ hg19	12	51373742	c.*1767G>A				
D	<i>SLC11A2</i>	GRCh37/ hg19	12	51381077	c.1629+1088G >A				
D	<i>SLC11A2</i>	GRCh37/ hg19	12	51419621	c.49+2339A> G				
D	<i>SLC40A1</i>	GRCh37/ hg19	2	190445495	c.-312_- 310dupCGG (presente anche nel pz 3 ma nelle Low Confidence variants)				
D	<i>TMPRSS6</i>	GRCh37/ hg19	22	37505552	c.-11G>A				
D	<i>TMPRSS6</i>	GRCh37/ hg19	22	37469591	c.1536C>T				
D	<i>TMPRSS6</i>	GRCh37/ hg19	22	37462926	c.2256C>T				
D	<i>TMPRSS6</i>	GRCh37/ hg19	22	37480797	c.1056G>A				
D	<i>TMPRSS6</i>	GRCh37/ hg19	22	37465385	c.1842- 6_1842- 2delCCCCA*				
D	<i>ANKRD26</i>	GRCh37/ hg19	10	27389197	c.59A>G				
D	<i>BMP6</i>	GRCh37/ hg19	6	7727271	c.83G>A				
D	<i>COX4I1</i>	GRCh37/ hg19	16	85838653	c.184T>C				
D	<i>CTC1</i>	GRCh37/ hg19	17	8141707	c.435+3A>G				
D	<i>CTC1</i>	GRCh37/ hg19	17	8134785	c.2478A>G				

D	<i>CTC1</i>	GRCh37/ hg19	17	8135061	c.2458A>G				
D	<i>DAXX</i>	GRCh37/ hg19	6	33287183	c.1914A>G				
D	<i>DCLRE1B</i>	GRCh37/ hg19	1	114448389	c.181C>T				
D	<i>DCLRE1B</i>	GRCh37/ hg19	1	114449662	c.234T>C				
D	<i>DDX41</i>	GRCh37/ hg19	5	176940384	c.822C>T				
D	<i>DKC1</i>	GRCh37/ hg19	X	153994596	c.369G>T				
D	<i>FUS</i>	GRCh37/ hg19	16	31195279	c.288C>T				
D	<i>GAR1</i>	GRCh37/ hg19	4	110737389	c.69T>C				
D	<i>GATA2</i>	GRCh37/ hg19	3	128205860	c.15C>G				
D	<i>GATA2</i>	GRCh37/ hg19	3	128204951	c.490G>A				
D	<i>GNAS</i>	GRCh37/ hg19	20	57478807	c.393C>T				
D	<i>GNAS</i>	GRCh37/ hg19	20	57415812	c.651T>A				
D	<i>LIG4</i>	GRCh37/ hg19	13	108861913	c.1704T>C				
D	<i>NAF1</i>	GRCh37/ hg19	4	164048199	c.1102C>G				
D	<i>NAF1</i>	GRCh37/ hg19	4	164085425	c.484A>G				
D	<i>NBN</i>	GRCh37/ hg19	8	90990479	c.307G>C				
D	<i>NBN</i>	GRCh37/ hg19	8	90958422	c.1770A>G				
D	<i>NBN</i>	GRCh37/ hg19	8	90967711	c.951T>C				
D	<i>NBN</i>	GRCh37/ hg19	8	90995019	c.102G>A				
D	<i>NSD1</i>	GRCh37/ hg19	5	176721198	c.6829T>C				
D	<i>PIF1</i>	GRCh37/ hg19	15	65116328	c.207T>C				
D	<i>NAF1</i>	GRCh37/ hg19	4	164048199	c.1102C>G				
D	<i>NAF1</i>	GRCh37/ hg19	4	164085425	c.484A>G				
D	<i>NBN</i>	GRCh37/ hg19	8	90990479	c.307G>C				
D	<i>NBN</i>	GRCh37/ hg19	8	90958422	c.1770A>G				

D	<i>NBN</i>	GRCh37/ hg19	8	90967711	c.951T>C				
D	<i>NBN</i>	GRCh37/ hg19	8	90995019	c.102G>A				
D	<i>NSD1</i>	GRCh37/ hg19	5	176721198	c.6829T>C				
D	<i>PIF1</i>	GRCh37/ hg19	15	65116328	c.207T>C				
D	<i>RECQL4</i>	GRCh37/ hg19	8	145738767	c.2296+1del				
D	<i>RECQL4</i>	GRCh37/ hg19	8	145737636	c.3127T>C				
D	<i>RECQL4</i>	GRCh37/ hg19	8	145742514	c.274T>C				
D	<i>RIF1</i>	GRCh37/ hg19	2	152331418	c.7174C>G				
D	<i>RIF1</i>	GRCh37/ hg19	2	152320118	c.4084G>A				
D	<i>RIF1</i>	GRCh37/ hg19	2	152322095	c.6061A>T				
D	<i>RIF1</i>	GRCh37/ hg19	2	152311570	c.2506G>A				
D	<i>RTEL1</i>	GRCh37/ hg19	20	62320968	c.1992T>C				
D	<i>RTEL1</i>	GRCh37/ hg19	20	62322288	c.2544T>C				
D	<i>RTEL1</i>	GRCh37/ hg19	20	62326110	c.3126A>C				
D	<i>RTEL1</i>	GRCh37/ hg19	20	62321655	c.2274G>A				
D	<i>STN1</i>	GRCh37/ hg19	10	105657316	c.743C>G				
D	<i>STN1</i>	GRCh37/ hg19	10	105659826	c.451A>G				
D	<i>TERF1</i>	GRCh37/ hg19	8	73958310	c.1198T>C				
D	<i>TERF1</i>	GRCh37/ hg19	8	73921283	c.186_188dup				
D	<i>TERF1</i>	GRCh37/ hg19	8	73958240	c.1128G>A				
D	<i>TMPRSS6</i>	GRCh37/ hg19	22	37462936	c.2246T>C				
D	<i>TMPRSS6</i>	GRCh37/ hg19	22	37485724	c.730A>G				
D	<i>TPP1</i>	GRCh37/ hg19	11	6638599	c.441A>G				

Tabella 23. Assetto genotipico dei genitori delle varianti comuni nei due fratelli. In rosso sono indicati gli alleli mutati mentre in verde gli alleli wild type. In blu sono identificate le varianti per le quali i due fratelli hanno un diverso assetto genotipico.* Variante identificata come Falso Positivo dal software di analisi.

Pz	Cat	gene	Ref Genome	chr	genome_position	depth	Var %	c.DNA	Allele Madre	Allele Padre	Allele Madre	Allele Padre
3	C	<i>PINX1</i>	GRCh37/hg19	8	10623289	1138	50,3	c.609T>C				
	C	<i>RIF1</i>	GRCh37/hg19	2	1,52E+08	1493	49,7	c.5363T>C				
	D	<i>PINX1</i>	GRCh37/hg19	8	10623214	1192	51,4	c.684C>G				
	D	<i>RIF1</i>	GRCh37/hg19	2	1,52E+08	1017	46,5	c.3093G>A				
	D	<i>CBL</i>	GRCh37/hg19	11	1,19E+08	907	46,6	c.125_127dup				
	D	<i>CEBPA</i>	GRCh37/hg19	19	33792731	728	50,3	c.227_232dup				
	D	<i>ERFE</i>	GRCh37/hg19	2	2,39E+08	1481	52,3	c.*889G>C				
	D	<i>NBN</i>	GRCh37/hg19	8	90955583	726	45,3	c.1836T>G				
	D	<i>NSD1</i>	GRCh37/hg19	5	1,77E+08	1849	48,6	c.6903G>C				
	D	<i>RTEL1</i>	GRCh37/hg19	20	62321189	921	52,4	c.2112C>T				
	D	<i>SLC40A1</i>	GRCh37/hg19	2	1,9E+08	1630	48,5	c.*1130A>G				
	D	<i>SRP72</i>	GRCh37/hg19	4	57356558	1541	45,8	c.1197T>C				
	D	<i>SRP72</i>	GRCh37/hg19	4	57333822	1342	44	c.21G>T				
	D	<i>TEN1</i>	GRCh37/hg19	17	73987652	729	99,9	c.198C>T				
	D	<i>TINF2</i>	GRCh37/hg19	14	24709506	1274	48,7	c.1092G>A				
	D	<i>TMPRSS6</i>	GRCh37/hg19	22	37499692	1102	99,5	c.-1-234T>C				
	D	<i>TMPRSS6</i>	GRCh37/hg19	22	37499565	1103	48,7	c.15C>T				
4	B	<i>PINX1</i>	GRCh37/hg19	8	10692242	794	47,2	c.63C>T				
	C	<i>MBD4</i>	GRCh37/hg19	3	129155414	1360	47,6	c.1073T>C				
	D	<i>CBL</i>	GRCh37/hg19	11	119170362	1547	50,6	c.2592C>T				
	D	<i>FUS</i>	GRCh37/hg19	16	31193942	1270	47,7	c.147C>A				
	D	<i>GNAS</i>	GRCh37/hg19	20	57428474	1858	48,4	c.154G>A				
	D	<i>RECQL4</i>	GRCh37/hg19	8	145740375	1126	46,8	c.1565G>A				
	D	<i>RECQL4</i>	GRCh37/hg19	8	145737086	1178	50,3	c.3480G>A				
	D	<i>RECQL4</i>	GRCh37/hg19	8	145741702	1122	50,8	c.801G>C				
	D	<i>RECQL4</i>	GRCh37/hg19	8	145742879	822	46,3	c.132A>G				
	D	<i>RTEL1</i>	GRCh37/hg19	20	62319085	1253	46,6	c.1443G>A				
	D	<i>SAMD9</i>	GRCh37/hg19	7	92734451	1357	51,8	c.960T>C				
	D	<i>SAMD9</i>	GRCh37/hg19	7	92734983	1590	50,3	c.428T>C				
	D	<i>SAMD9</i>	GRCh37/hg19	7	92734065	1475	46,2	c.1346A>G				
	D	<i>SAMD9L</i>	GRCh37/hg19	7	92764982	1539	45,7	c.303G>A				
	D	<i>SLC40A1</i>	GRCh37/hg19	2	190430177	1174	44,5	c.663T>C				
	D	<i>STN1</i>	GRCh37/hg19	10	105657363	976	48,5	c.696G>C				
	D	<i>TMPRSS6</i>	GRCh37/hg19	22	37499386	1317	43	c.72G>A				
D	<i>TERT</i>	GRCh37/hg19	5	1255520	1254	50,2	c.2850C>T					
D	<i>TERT</i>	GRCh37/hg19	5	1294086	1484	50,9	c.915G>A					

Tabella 24. Assetto genotipico dei genitori delle varianti specifiche di ogni fratello. In rosso sono indicati gli alleli mutati mentre in verde gli alleli wild type.

14. Discussione del Secondo studio

I risultati ottenuti dalla validazione del pannello NGS sono stati soddisfacenti, evidenziando un alto numero di letture di buona qualità, buona cattura delle regioni target e flanking, e bassa percentuale di duplicati di PCR. Mostrano inoltre ottima copertura, con una profondità di analisi di 200x in oltre il 99% delle regioni ed alta riproducibilità dei risultati sia nei replicati intra-run che in quelli inter-run. Nonostante ciò i risultati hanno mostrato problematiche nel discriminare regioni ripetute, ciò è stato evidenziato sia della presenza di tre delezioni nei geni *BCL11A*, *TMPRSS6* e *TERF1*, in tutti i campioni, che sono state classificate come falsi positivi che dal low coverage della regione 5' UTR del gene *HBA2* in tutti i campioni ad eccezione di uno. Per ciò che concerne il gene β , sette degli otto genotipi analizzati sono stati validati e correttamente annotati, evidenziando mutazioni puntiformi e piccole delezioni come cd8 (-AA), localizzate sia a livello esonico, intronico e del promotore. Il genotipo non validato è caratterizzato dalla mutazione Hb Lepore-Boston Washington, un riarrangiamento che porta alla formazione del gene di fusione *HBD-HBB*. Il sistema non ha permesso di annotare questo tipo mutazione, poiché è stata riportata soltanto una delezione a livello del gene *HBB*, senza mostrare alcuna alterazione nel gene *HBD*. Problematiche simili si sono verificate anche a livello dei geni α -globinici, dove le mutazioni più ricorrenti sono delezioni e duplicazioni. Infatti il sistema ha permesso di caratterizzare correttamente soltanto quattro degli otto genotipi α analizzati. L'analisi delle CNV ha evidenziato la presenza di una duplicazione nel campione con genotipo anti^{-3.7} e di una delezione nel campione con genotipo α ^{-3.7}, senza però riuscire ad annotare in maniera precisa il corretto riarrangiamento.

Per cercare di descrivere una sempre più accurata correlazione genotipo-fenotipo e poter identificare nuovi potenziali geni modificatori nei due fratelli (pz 3 e4) affetti da beta talassemia è risultata necessaria l'analisi delle varianti che soddisfavano i filtri imposti (SNP>20%, INDEL>15%, profondità di lettura>50X, esclusione delle varianti off target, Phred score>Q30) identificate mediante il pannello custom NGS, sviluppato dal laboratorio di Genetica Molecolare dell'Azienda Ospedaliera di Perugia in collaborazione con SOPHiA GENETICS, comprendente 75 geni (appendice B). Per avere maggiori informazioni in merito all'assetto genetico delle

varianti nei due fratelli sono stati analizzati mediante la stessa metodica di sequenziamento anche i genitori.

Interessante risulta analizzare in dettaglio i geni coinvolti nel metabolismo del ferro data l'eterogeneità clinica dei due fratelli che si esprime in termini di diverso: supporto trasfusionale, profilo di sovraccarico marziale T2* e differente risposta alla ferrochelazione. In particolare:

- Paziente 3: regime trasfusionale più intenso (2 GRC/settimana) senza sovraccarico marziale RMT2* (deferasirox)
- Paziente 4: regime trasfusionale meno intenso (1 GRC/settimana), moderato sovraccarico cardiaco ed epatico, ferritinemia elevata (deferasirox cpr rivestite, pregressa idiosincrasia epatica a deferasirox dispersibile e agranulocitosi con deferiprone).

Andando ad analizzare le varianti specifiche di ogni fratello nei geni coinvolti nel metabolismo nel paziente 3 sono state identificate: nel gene *TMPRSS6* le seguenti due varianti: la c.-1-234T>C (rs5756515) in omozigosi e la c.15C>T (rs115270691) in eterozigosi; nel 3'UTR del gene *ERFE* e del gene *SL40A1* sono state identificate rispettivamente la variante c.*889G>C (rs7595215) e la variante c.*1130A>G (rs11539983) entrambe in eterozigosi (Tabella 19). Mentre nel paziente 4 sono state identificate: nel gene *TMPRSS6* la variante c.72G>A (rs11704654) e nel gene *SLC40A1* la variante c.663T>C (rs2304704), entrambe in eterozigosi (Tabella 20). Nelle varianti in comune ai due fratelli ma con diverso assetto genotipico sono state identificate 6 varianti in *TMPRSS6* e 1 in *ERFE* (Tabella 18) tutte in omozigosi nel paziente 3 mentre presenti in eterozigosi nel paziente 4.

Da tale analisi emerge come il gene *TMPRSS6* risulta sede di dieci diverse variazioni (Tabella 25).

N	Type	Depth	Exon_rank	c.DNA	hpflags	g1000	ExAC
3	intronic	99,5	c.-1-234T>C		rs5756515		
	synonymous	48,7	c.15C>T	p.(Phe5=)	rs115270691	0,0298	0,021
	splice_acceptor_-1	98,6	c.1842-6_1842-2del*	p.(?)	rs200434923		0,157
	5'UTR	99,8	c.-6095G>A		rs228916		
	synonymous	100	c.1536C>T	p.(Asp512=)	rs4820268	0,6066	0,539
	synonymous	99,8	c.2256C>T	p.(Tyr752=)	rs2235321	0,3913	0,365
	synonymous	99,8	c.1056G>A	p.(Ser352=)	rs2111833	0,3475	0,252
	missense	99,6	c.2246T>C	p.(Val749Ala)	rs855791	0,6593	0,572

	missense	99,4	c.730A>G	p.(Lys244Glu)	rs2235324	0,3963	0,383
4	synonymous	43	c.72G>A	p.(Pro24=)	rs11704654	0,1756	0,164
	splice_acceptor_-1	55,3	c.1842-6_1842-2del*	p.(?)	rs200434923		0,157
	5'UTR	100	c.-6095G>A		rs228916		
	synonymous	44	c.1536C>T	p.(Asp512=)	rs4820268	0,6066	0,539
	synonymous	51,1	c.2256C>T	p.(Tyr752=)	rs2235321	0,3913	0,365
	synonymous	43,6	c.1056G>A	p.(Ser352=)	rs2111833	0,3475	0,252
	missense	52	c.2246T>C	p.(Val749Ala)	rs855791	0,6593	0,572
	missense	47,3	c.730A>G	p.(Lys244Glu)	rs2235324	0,3963	0,383

Tabella 25. Varianti identificate nel gene *TMPRSS6*. In rosso sono rappresentate le varianti specifiche per ogni paziente, in blu quelle in comune con diverso assetto genotipico e in nero quelle in comune con lo stesso assetto genotipico. *Falso positivo.

TMPRSS6, noto anche come matriptasi-2 (MT2), è una proteina di membrana espressa in abbondanza nel fegato, che inibisce l'espressione dell'epcidina, un ormone principale regolatore dell'omeostasi del ferro. L'epcidina è un inibitore diretto della ferroportina, una proteina di trasporto del ferro di membrana presente su enterociti e macrofagi, e quindi inibisce l'assorbimento del ferro intestinale e il rilascio di ferro dai depositi cellulari. Tale ormone risulta abbondantemente espresso in condizioni di eccesso di ferro e infiammazione e minimamente espresso in condizioni di carenza di ferro. Inibizione di *TMPRSS6*, che è inibitore dell'epcidina, stimola l'assorbimento e il rilascio di ferro dai depositi, aumentando così la disponibilità di ferro in tutto il corpo. Questo potente effetto di *TMPRSS6* sulla disponibilità di ferro sottolinea il ruolo fondamentale di questa proteina nella biologia del ferro. La soppressione dell'espressione di *Tmprss6* attenua la gravità della malattia nei modelli murini di β -talassemia ed emocromatosi ereditaria, mentre l'inibizione chimica dell'attività di *TMPRSS6* stimola l'espressione di epcidina nelle linee cellulari. La soppressione di *Tmprss6* nei modelli murini di malattia aumenta i livelli di epcidina e limita l'assorbimento del ferro e il rilascio di ferro dai depositi cellulari; si ritiene che la limitazione dei livelli di ferro nella β -talassemia migliori l'anemia limitando la tossicità mediata dal ferro nei precursori dei globuli rossi. La soppressione farmacologica dell'espressione di *TMPRSS6* è attualmente in fase di studio come potenziale trattamento per la β -talassemia negli studi clinici. I risultati di Enns et al suggeriscono che il targeting della funzione di *TMPRSS6* non dovrebbe concentrarsi sull'attività proteolitica di *TMPRSS6* ma piuttosto sulla sua capacità di legare le proteine come HJV. Ciò potrebbe portare alla generazione di un nuovo

tipo di agente farmacologico che gli scienziati potrebbero utilizzare per comprendere meglio le basi meccanicistiche della biologia del ferro e che i medici potrebbero utilizzare per trattare pazienti con anemie da carico di ferro, emocromatosi ereditaria e altre malattie legate al ferro [184].

In particolare nel paziente 3 sono state identificate queste due alterazioni:

- *TMPRSS6:c.-1-234T>C* in omozigosi, variante intronica, corrispondente al rs5756515, frequenza in 1000g: 0.2847 e in GnomAD:0.33621, in particolare negli africani 0.2191, riportata in Clin variation come benigna;
- *TMPRSS6:c.15C>T* in eterozigosi, variante sinonima, corrispondente al rs115270691, riportata in Clin variation come benigna, frequenza in 1000g:0,0186 e in ExAC:0.021192, in particolare negli africani 0.0167.

Mentre nel paziente 4 è stata identificata la variante:

- *TMPRSS6:c.72G>A* in eterozigosi, variante sinonima, privata, corrispondente al rs11704654, frequenza in 1000G:0.1516 ed in ExAC:0.163917, in particolare negli africani 0.1545, riportata in Clin variation come benigna.

Quest'ultimo SNP, a differenza degli altri due, è stata indentificato in una famiglia con severa anemia [185].

Interessante risulta discutere anche le varianti del gene *TMPRSS6* comuni nei due fratelli (Tabella 25). In un recente articolo in cui venivano descritte le differenze nella frequenza delle varianti associate allo squilibrio del ferro tra le popolazioni globali emergono undici SNPs in *TMPRSS6* che sono chiaramente associati a basso contenuto di ferro, anemia da carenza di ferro e / o IRIDA, tra cui riscontriamo anche le varianti identificate nei due fratelli, ovvero: rs855791, rs2235321, rs2235324, rs4820268, rs228916. La variante missenso A736V (*TMPRSS6* rs855791) è lo SNP più segnalato associato a carenza di ferro e ha un MAF del ~ 50% su tutte le popolazioni non africane, ma in Africa raggiunge solo il 10%. *TMPRSS6* rs855791 è associato ad anemia da carenza di ferro e IRIDA, con indici elevati di epcidina, ferro ridotto e emoglobina ridotta. Tale variante codifica per una sostituzione alanina-valina in posizione 736 all'interno del dominio della serina proteasi matriptasi-2 (p.Ala736Val). Questo polimorfismo a singolo nucleotide si trova nella parte funzionale di *TMPRSS6* e causa una sostituzione non sinonima che riduce la capacità dell'enzima di inibire la trascrizione dell'epcidina. Pertanto, *TMPRSS6*

A736V influenza l'omeostasi del ferro e l'eritropoiesi nei soggetti normali, sebbene in un altro studio condotto in pazienti con carenza di ferro, sulla base della presenza o assenza di anemia, TMPRSS6 non differisce nelle donne con carenza di ferro con o senza anemia. Rimane incerto se l'associazione sia mediata dal ferro o sia per effetto diretto delle varianti sull'eritropoiesi [186].

Infine all'interno delle 17 varianti specifiche del paziente 3 potrebbero essere interessanti alcune varianti associate con l'attività della telomerasi:

- *PINX:c.609T>C* in eterozigosi, sinonima, corrisponde ad rs17152329, potenzialmente benigna, frequenza in 1000G:0.0449 e in ExAC:0.014896, in particolare negli africani 0.0939;
- *RIF1:c.5363T>C* in eterozigosi, missense, per i criteri dell' ACMG benigna perché dominio non conservato ma molto rara e non descritta in letteratura, legata al complesso del telomero;
- *TEN1:c19C>T* in omozigosi, sinonima, complesso del telomero, corrisponde a rs10852767, frequenza in 1000G: 0.1891 in ExAC:0.09767, in particolare negli africani 0.4803.

Mentre all'interno delle 19 varianti specifiche del paziente 4 potrebbe essere interessante la seguente variante: *PINX:c.63C>T* in eterozigosi, per i criteri dell'ACMG benigna, sinonima, corrisponde a rs1052523, frequenza in 1000G: 0.0517 in GnomAD:0.04209, in particolare negli africani 0.0247.

15. CONCLUSIONI

Questa tesi di dottorato ha permesso in prima istanza di poter eseguire uno studio pilota su 53 pazienti affetti da emoglobinopatie (15 beta-talassemia, 10 HbS/Beta tal, 20 HbS/HbS, 8 eterozigoti composti per varianti emoglobiniche qualitative) provenienti da sedici diversi paesi andando a correlare alcune alterazioni sia nucleari che mitocondriali, giunti in una regione non endemica come l'Umbria. Mediante l'integrazione delle varianti dei geni globinici (*HBB*, *HBA1*, *HBA2*), dei polimorfismi nucleari dei geni modulatori del fenotipo clinico (*HBG2*, rs7482144; *BCL11A*, rs1427407 e rs10189857; *HBS1L-MYB*, rs28384513 e rs939913, *C/EBPE*, rs45496295) e le informazioni derivanti dalla regione di controllo mitocondriale (tra i siti 15877 e 259) è stato possibile costruire una rappresentazione filogeografica degli individui affetti che ha indentificato una netta separazione tra i pazienti africani e quelli europei [187]. In particolare, individuando la struttura filogeografica degli individui affetti, si potrebbe avere una sempre più accurata stratificazione genetica che potrebbe aiutare ad ottimizzare l'algoritmo diagnostico anche per i pazienti con origine incerta o sconosciuta. Inoltre questo studio pilota basato sulle varianti di marcatori uniparentali e dei geni dell'emoglobina potrebbe rappresentare uno strumento integrato e potente per delineare meglio le tante sfacciate tipiche delle emoglobinopatie.

Inoltre allo scopo di effettuare una sempre più accurata correlazione genotipo-fenotipo e poter identificare nuovi potenziali geni modificatori in soggetti affetti è stato sviluppato e validato un pannello di Next Generation Sequencing, presso il laboratorio di Genetica Molecolare dell'Università degli Studi di Perugia in collaborazione con la ditta SOPHiA GENETICS, che ha permesso l'analisi simultanea di 75 geni, di cui otto geni associati ad Emoglobinopatie, undici coinvolti nel metabolismo del ferro ed i restanti legati a malattie ematologiche. Tale strumento ha mostrato estrema accuratezza, sensibilità e specificità nell'identificare le mutazioni puntiformi e indel.

Indagini di varianti mitocondriali e nucleari condotte su una più ampia coorte di pazienti e su famiglie opportunamente selezionate, associate a studi di modellistica computazionale ed a studi funzionali in vitro ed in vivo, permetteranno di effettuare

una stratificazione sempre più dettagliata di tale patologie che rappresenta un requisito necessario sia ai fini diagnostici che per la gestione clinica del paziente.

BIBLIOGRAFIA

1. John Old, Prevention and diagnosis of haemoglobinopathies, a short guide for health professionals and laboratory scientists; Blishers thalassaemia international federation TIF publication NO. 21 Pub.
2. <http://www.summagallicana.it/Volume2/B.XVIII.03.htm>.
3. Higgs DR, Sharpe JA, et al. Understanding alpha globin gene expression: a step towards effective gene therapy. *Semin Hematol*, 35, 93-104; 1998.
4. Albitar M, Katsumata M, et al. Human alpha-globin genes demonstrate autonomous developmental regulation in transgenic mice. *Mol Cell Biol*, 11, 3786-3794; 1991.
5. Baralle FE, Blechl AE, et al. The structure and evolution of the human β -globin gene family. *Cell*, 21,653-668; 1980.
6. Cao A, Dalla Piccola B, et al. Malattie Genetiche molecole e geni, diagnosi, prevenzione e terapia. Piccin; 2004.
7. Cao A, Galanello R, et al. Beta-thalassemia. *Genetics in Medicine*. Volume 12, Number 2; February 2010.
8. Curtin P, Gobert-Jones JA, et al. A distant gene deletion affects β -globin gene function in an atypical $\gamma\delta\beta$ -thalassemia. *J. Clin. Invest*, 76, 1554; 1985.
9. Alter BP, Dobkin CS, et al. $\gamma\delta\beta$ Thalassemia due to a de novo mutation deleting the 5' β -globin gene activation-region hypersensitive sites. *Proc. Natl. Acad. Sci. USA*, 86 (19) 7470-7474; 1989.
10. Russel E Kaufman, Scott D Langdon. Gamma-Globin Gene Promoter Elements Required for Interaction With Globin Enhancers. *Blood*, 91 (1), 309-318; 1998.
11. Antoniou M, Grosveld F, et al. Each hypersensitive site of the human beta-globin locus region confers a different developmental pattern of expression on the globin genes. *Genes Dev*, 7, 106-113; 1993.
12. Thomas N. Williams, and Swee Lay Thein; Annual Review of Genomics and Human Genetics, Sickle Cell Anemia and Its Phenotypes, GG19CH04_Williams ARI 31 March 2018.
13. Fraser P, Grosveld F, et al. Importance of globin gene order for correct developmental expression. *Genes Dev*, 5, 1387-1394; 1991.

14. Dillon N, Fraser P, et al. The effect of distance on long-range chromatin interactions. *Mol Cell*, 1, 131-139; 1997.
15. Grosveld F, Stamatoyannopoulos G. Haemoglobin switching. In: Stamatoyannopoulos G, Perlmutter R, Majerus W, Varmus H (eds). *The Molecular Basis of Blood Diseases*. Saunders, Philadelphia, pp 135-165; 2001.
16. Cao A, Dalla Piccola B, et al. Malattie Genetiche molecole e geni, diagnosi, prevenzione e terapia. Piccin; 2004.
17. Divya S. Vinjamur, Daniel E. Bauer, et al. Recent progress in understanding and manipulating haemoglobin switching for the haemoglobinopathies. *British Journal of Haematology*, 2018, 180, 630–643.
18. Clegg J.B., Weatherall, et al. Thalassemia-a global public health problem. *Nature Med.* 2, 847-849; 1996.
19. Clegg J.B, Weatherall, D.J., et al. In *The Metabolic and Molecular Bases of Inherited Disease* (ed. Scriver, C. R.) 4571-4636 (McGraw-Hill, New York); 2001.
20. Cauchi MN, Loo LS. DNA methylation patterns of the gamma delta beta-globin genes in human fetal and adult erythroid tissues. *Am J Hematol*, 39(4), 289-93; 1992.
21. Enver T, Papayannopoulou T, et al. DNA methylation : a secondary event in globin gene switching? *Gene Dev*, 2(6), 698-706; 1988.
22. Cohen RA, Faller DV, et al. Butyrate infusion in the ovine fetus delay the biologic clock for globin gene switching. *Proceedings of the National Academy of Sciences*, 85, 8540-8542; 1988.
23. Wood HG. Then and now. *Annu Rev Biochem*, 54, 1-41; 1985.
24. Andrew W. De Vilbiss, Koiki R. Katsumura, et al. Transcriptional Mechanisms Underlying Hemoglobin Synthesis. *Cold Spring Harb Perspect Med*, 3:a015412;2013.
25. Daniel E. Bauer, Sophia C. Kamran, et al. Reawakening fetal hemoglobin: prospects for new therapies for the β -globin disorders. *Blood*, volume 120, number 15;11 October 2012.
26. Orkin SH, Perkins AC, et al. Lethal beta-talassemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, 375, 318-322; 1995.

27. Menne TF, Sankaran VG, et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science*, 322(5909): 1839-1842; 2008.
28. Ragoczy T, Sankaran VG, et al. Developmental and Species-divergent globin switching are driven by BCL11A. *Nature*, 460(7259):1093-1097; 2009.
29. Sankaran VG, Xu J, et al. Transcriptional silencing of (gamma)-globin by *BCL11A* involves long-range interaction and cooperation with SOX6. *Genes Dev*, 24(8):783-798;2010.
30. Byron R, Sankaran VG, et al. A functional element necessary for fetal hemoglobin silencing. *N Engl J Med*, 365(9):807-814; 2011.
31. Xunde Wang and Swee Lay Thein. Switching from fetal to adult hemoglobin. *Nat Genet*. Author manuscript; available in PMC 2019 March 15.
32. Bygrave A, Grosveld F, et al. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, 375(6529):316-318; 1995.
33. Guy LG, Orkins SH, et al. Erythroid kruppel-like factor is essential for beta-globin gene expression even in absence of gene competition, but is not sufficient to induce the switch from gamma-globin to beta beta-globin gene expression. *Blood*, 91(7):2259-2263; 1998.
34. Bieker JJ., Miller IJ, A Novel erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the kruppel family of nuclear proteins. *Mol Cell Biol*, 13(5):2776-2786; 1993.
35. Bieker JJ., Feng WC, et al. Analyses of beta-thalassemia mutant DNA interactions with erythroid kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem*, 269(2):1493-1500; 1994.
36. Borg J, Georgitsi M, et al. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet*, 42(9):801-805; 2010.
37. Liu K, Pawlik KM, et al. KLF1 regulates BCL11A expression and gamma-to-beta-globin gene switching. *Nat Genet*, 42(9):742-744; 2010.
38. Gaensler KM, Orkin SH, et al. Silencing of human fetal globin expression is impaired in the absence of the adult beta-globin gene activator protein EKLF1. *Proc Natl Acad Sci USA*, 93(22):12267-12271; 1996.

39. Best S, Menzel S, et al. cMYB is involved in the regulation of fetal hemoglobin production in adults. *Blood*,108(3):1077-1083; 2006.
40. Menne TF, Sankaran VG, et al. MicroRNA-15^o and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. *Proc Natl Acad Sci USA*, 108(4):1519-1524; 2011.
41. Bianchi E, Salati S, et al. C-mybsupports erythropoiesis through the transactivation of KLF1 and LMO2 expression. *Blood*, 116(22):e99-e110; 2010.
42. Aerbajinai W, Chin K, et al. SCF induces gamma-globin gene expression by regulating downstream transcription factor COUP-TFII. *Blood*, 114(1):187-194; 2009.
43. Gillemans N, Pourfarzad F, et al. Fetal globin expression is regulated by friend of Prmt1. *Blood*, 116(20):4349-4352; 2010.
44. Bank A, DeAngelis K, et al. Multiple hematopoietic defects and delayed globin switching in ikaros null mice. *Proc Natl Acad Sci U S A*, 99(2):602-607; 2002.
45. Gaensler K, Ginder GD, et al. Methyl binding domain protein 2 mediates gamma-globin gene silencing in adult human betaYAC transgenic mice. *Proc Natl Acad Sci U S A*, 103(17):6617-6622; 2006.
46. Sutton R, Zhao Q, et al. The role of p22 NF-E4 in human globin gene switching. *J Biol Chem*, 279(25):26227-26232; 2004.
47. Lowrey CH, Macari ER. Induction of human fetal hemoglobin via the NRF2 antioxidant response signaling pathway. *Blood*, 117(22):5987-5997; 2011.
48. Kobayashi S, McPhee D, et al. Embryonic and fetal beta-globin gene repression by the orphan nuclear receptors, TR2 and TR4. *EMBO J*, 26(9):2295-2306; 2007.
49. Ingram V. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature*, 1957 Aug 17. 180(4581): p. 326-8.
50. Steinberg, M. Predicting clinical severity in sickle cell anaemia. *Br J Haematol.*, 2005 May. 129(4): p. 465-81.
51. Lehmann H, Huntsman RG, et al. Man's haemoglobins. Oxford: North-Holland Publishing Company, 1974.
52. Eaton WA, Hofrichter S, et al. Hemoglobin S gelation and sickle cell disease. *Blood* 1987;70:1245-66.

53. Thompson & Thompson, Nussbaum, Mc innes, Willard. *Genetica in medicina*, Tr.it, Gasparini P, Jolascon A, Idelson-Gnocchi, 2005.
54. Serjeant GR, H.D., Hambleton IR. Geography and the clinical picture of sickle cell disease. An overview. *Ann N Y Acad Sci.*, 1989. 565: p. 109-19.
55. Powars D. Beta s-gene cluster haplotypes in sickle cell anemia. Clinical and hematologic features. . *Hematol Oncol Clin North Am*, 1991. 5: p. 475-93.
56. Austin H, K.N., et al. Sickle cell trait and the risk of venous thromboembolism among blacks. *Blood* 2007. 110: p. 908-12.
57. Weatherall DJ, C.J.. Inherited haemoglobin disorders: an increasing global health problem. *Bull World Health Organ.*, 2001 Oct 24. 79(8): p. 704-12.
58. Cotton F, Wolff F, et al. Automated capillary electrophoresis in the screening for hemoglobinopathies. *Methods Mol Biol* 2013;984:227-35. doi: 10.1007/978-1-62703-296-4_16. PubMed PMID: 23386347.
59. Steinberg MH, Chui DH. HbC disorders. *Blood* 2013 Nov 21;122(22):3698. doi: 10.1182/blood-2013-09-526764. PubMed PMID: 24263962.
60. Kohne E. Hemoglobinopathies: clinical manifestations, diagnosis, and treatment. *Dtsch Arztebl Int* 2011 Aug;108(31-32):532-40. doi: 10.3238/arztebl.2011.0532. Epub 2011 Aug 8. PubMed PMID: 21886666; PubMed Central PMCID: PMC3163784.
61. Bruyneel M, De Caluwé JP, et. Al. Hemoglobinopathy C and splenomegaly in an Ivory Coast patient. Value of splenectomy]. *Rev Med Brux* 2003 Apr;24(2):105-7. French. PubMed PMID: 12806878.
62. 2018©Medical Systems S.p.A. Società soggetta a direzione e coordinamento Finsystems; (sito web: <http://www.medicalsystems.it/didattica/emoglobinopatia-c/>).
63. Hafsia R, Marrakchi O, et al. Hemoglobin C disease: report of 16 Tunisian cases. *Tunis Med* 2007 Mar;85(3):209-11. French. PubMed PMID: 17668575.
64. Zhor Ouzzif, Aissam El Maataoui, et al. Hémoglobinoase C au Maroc : A propos de 111 cas; *La tunisie Medicale* - 2017 ; Vol 95 (n°012) : 229-233.
65. Genuardi M, Neri G. *Genetica umana e medica*. 2.ed.Elsevier Masson; 2010.
66. Clegg JB, Weatherall DJ. eds. *The Thalassaemia syndromes*, 3th ed. Oxford: *Blackwell Scientific*; 1981.

67. Bernini LF, Haeteveld CL. α -thalassaemia. In: Rodgers GP, ed. Sick cell disease and thalassaemia. *Baillière's clinical haematology*, 11, 91-126; 1998.
68. Myers RM, Maniatis T, et al. Fine structure genetic analysis of a beta-globin promoter. *Science*, 232, 613-618; 1986.
69. Olivieri NF. The β -Thalassaemias. *The New England Journal of Medicine*, 341(2),99-109; 1999.
70. Cao A, Galanello R, et al. Beta-thalassaemia. *Genetics in Medicine*. Volume 12, Number 2; February 2010.
71. Cao A, Wai Kan Y. The prevention of thalassaemia. *Cold Spring Harb Perspect Med* 2012;3:a011775.
72. Bianco Silvestroni I. Le talassemie. Un problema medico-sociale: ieri e oggi. Istituto Italiano di Medicina Sociale, Roma; 1998.
73. Cappellini N, Cohen A, et al. (Eds). Guidelines for the Clinical Management of Thalassaemia. Thalassaemia International Federation; 2000.
74. Bianco Silvestroni I. Le talassemie. Un problema medico-sociale: ieri e oggi. Istituto Italiano di Medicina Sociale, Roma; 1998.
75. Douglas R. Higgs, The molecular basis of α -Thalassaemia, *Cold Spring Harb Perspect Med* 2013; doi: 10.1101/cshperspect.a011718.
76. John Old, Michael Angastiniotis, et al. Prevention of thalassaemias and other haemoglobin disorders, volume 1: principles, 2013.
77. Frédéric B. Piel, Ph.D., et al. The α -Thalasseмии, From the Evolutionary Ecology of Infectious Disease Group, Department of Zoology (F.B.P.), and the Weatherall Institute of Molecular Medicine, John Radcliffe Hospital (D.J.W.), University of Oxford, Oxford, United Kingdom, 2014.
78. http://www.laboratorigenoma.it/prestazioni_sottocategoria.asp?IdCat=18&IdSubCat=89
79. Clegg, J. B., Weatherall, et al. The Thalassaemia Syndromes 4th edn, Blackwell Science, Oxford; 2001.
80. Centis F, Lucarelli G, et al. The importance of erythroid expansion in determining the extent of apoptosis in erythroid precursors in patients with β -thalassaemia major. *Blood*, 96, 3624-3629; 2000.
81. Cao A, Wai Kan Y. The prevention of thalassaemia. *Cold Spring Harb Perspect Med* 2012;3:a011775.

82. De Maria R, Lucchetti L, et al. Apoptotic role of Fas/Fas ligand system in the regulation of erythropoiesis. *Blood*, 91, 796-803; 1999.
83. Testa U. Apoptotic mechanism in the control of erythropoiesis. *Leukaemia*, 18, 1176-1199; 2004.
84. Angelucci E, Bai H, et al. Enhanced macrophagic attack on β -thalassaemia major erythroid precursor. *Haematologica*, 87, 578-583; 2002.
85. Cappellini MD, Comino A, et al. Membrane-bound iron contributes to oxidative damage of β -thalassaemia intermedia erythrocytes. *Br J Haematol*, 112, 48-50; 2001.
86. Weatherall DJ. Thalassaemia and malaria, revisited. *Ann Trop Med Parasitol*, 91(7),885-890; 1997b.
87. Franchini M, Veneri D. Iron-chelation therapy: an update. *Hematol J*, 5, 287- 292; 2004.
88. Genuardi M, Neri G. *Genetica umana e medica*. 2.ed.Elsevier Masson; 2010.
89. Swee Lay Thein. *The Molecular Basis of β -thalassemia*. Cold Spring Harb Perspect Med, doi:10.1101/cshperspect.a011700; 2013.
90. Huisman THJ. The β and δ -thalassemia repository. *Hemoglobin*, 17, 479-499; 1993.
91. Fisher TC, Mathias LA, et al. Ineffective erythropoiesis in β -thalassaemia major is due to apoptosis at the polychromatophilic normoblast stage. *Exp Hematol*, 28, 1343-1353; 2000.
92. Collins T, Cotran Ramzi S, et al. *Le basi patologiche delle malattie*. Piccin; 2000.
93. Clegg JB, Weatherall DJ. eds. *The Thalassaemia syndromes*, 3th ed. Oxford: Blackwell Scientific; 1981.
94. Dadheech S, Jain S et al. Association of Xmn1 -158 γ G variant with severity and HbF levels in β -thalassemia major and sickle cell anaemia. *Mol Biol Rep*. 2014;41(5):3331-7. Epub 2014/02/02. doi: 10.1007/s11033-014-3195-5. PubMed PMID: 24488321.
95. Lai Y, Zhou L, et al. The association between four SNPs (rs7482144, rs4671393, rs28384513 and rs4895441) and fetal hemoglobin levels in Chinese Zhuang β -thalassemia intermedia patients. *Blood Cells Mol Dis*. 2017;63:52-7. Epub 2017/01/25. doi: 10.1016/j.bcmd.2017.01.011. PubMed PMID: 28160732.

96. Gilman JG, Huisman TH. DNA sequence variation associated with elevated fetal G gamma globin production. *Blood*. 1985;66(4):783-7. PubMed PMID: 2412616.
97. Bianchi N, Cosenza LC, et al. Structural and Functional Insights on an Uncharacterized $\text{A}\gamma$ -Globin-Gene Polymorphism Present in Four β 0-Thalassemia Families with High Fetal Hemoglobin Levels. *Mol Diagn Ther*. 2016;20(2):161-73. doi: 10.1007/s40291-016-0187-2. PubMed PMID: 26897028.
98. Chaouch L, Moumni I, et al. rs11886868 and rs4671393 of *BCL11A* associated with HbF level variation and modulate clinical events among sickle cell anemia patients. *Hematology*. 2016;21(7):425-9. Epub 2016/01/22. doi: 10.1080/10245332.2015.1107275. PubMed PMID: 27077760.
99. Liu D, Zhang X, et al. *KLF1* mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of β -thalassemia. *Blood*. 2014;124(5):803-11. Epub 2014/05/14. doi: 10.1182/blood-2014-03-561779. PubMed PMID: 24829204; PubMed Central PMCID: PMC4118488.
100. Sedgewick AE, Timofeev N, et al. *BCL11A* is a major HbF quantitative trait locus in three different populations with beta-hemoglobinopathies. *Blood Cells Mol Dis*. 2008;41(3):255-8. Epub 2008/08/08. doi: 10.1016/j.bcmd.2008.06.007. PubMed PMID: 18691915; PubMed Central PMCID: PMC4100606.
101. Sankaran VG, Weiss MJ. Anemia: progress in molecular mechanisms and therapies. *Nat Med*. 2015;21(3):221-30. doi: 10.1038/nm.3814. PubMed PMID: 25742458; PubMed Central PMCID: PMC4452951.
102. Stadhouders R, Aktuna S, et al. *HBS1L-MYB* intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. *J Clin Invest*. 2014;124(4):1699-710. Epub 2014/03/10. doi: 10.1172/JCI71520. PubMed PMID: 24614105; PubMed Central PMCID: PMC3973089.
103. Lettre G, Sankaran VG, et al. DNA polymorphisms at the *BCL11A*, *HBS1L-MYB*, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc Natl Acad Sci U S A*. 2008;105(33):11869-74. Epub 2008/07/30. doi: 10.1073/pnas.0804799105. PubMed PMID: 18667698; PubMed Central PMCID: PMC2491485.

104. Thein SL, Menzel S, et al. Intergenic variants of *HBS1L-MYB* are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. *Proc Natl Acad Sci U S A*. 2007;104(27):11346-51. Epub 2007/06/25. doi: 10.1073/pnas.0611393104. PubMed PMID: 17592125; PubMed Central PMCID: PMCPMC2040901.
105. Al-Allawi N, Qadir SMA, et al. The association of *HBG2*, *BCL11A*, and *HMIP* polymorphisms with fetal hemoglobin and clinical phenotype in Iraqi Kurds with sickle cell disease. *Int J Lab Hematol*. 2019;41(1):87-93. Epub 2018/09/14. doi: 10.1111/ijlh.12927. PubMed PMID: 30216683.
106. Sebastiani P, Farrell JJ, et al. *BCL11A* enhancer haplotypes and fetal hemoglobin in sickle cell anemia. *Blood Cells Mol Dis*. 2015;54(3):224-30. Epub 2015/01/30. doi: 10.1016/j.bcmd.2015.01.001. PubMed PMID: 25703683; PubMed Central PMCID: PMCPMC4341902.
107. Akinsheye I, Solovieff N, et al. Fetal hemoglobin in sickle cell anemia: molecular characterization of the unusually high fetal hemoglobin phenotype in African Americans. *Am J Hematol*. 2012;87(2):217-9. Epub 2011/12/03. doi: 10.1002/ajh.22221. PubMed PMID: 22139998; PubMed Central PMCID: PMCPMC3302931.
108. Pereira C, Relvas L, et al. Polymorphic variations influencing fetal hemoglobin levels: association study in beta-thalassemia carriers and in normal individuals of Portuguese origin. *Blood Cells Mol Dis*. 2015;54(4):315-20. Epub 2015/02/21. doi: 10.1016/j.bcmd.2015.02.001. PubMed PMID: 25842369.
109. Sherva R, Sripichai O, et al. Genetic modifiers of Hb E/beta0 thalassemia identified by a two-stage genome-wide association study. *BMC Med Genet*. 2010;11:51. Epub 2010/03/30. doi: 10.1186/1471-2350-11-51. PubMed PMID: 20353593; PubMed Central PMCID: PMCPMC2853425.
110. Thein SL. Genetic association studies in β -hemoglobinopathies. *Hematology Am Soc Hematol Educ Program*. 2013;2013:354-61. doi: 10.1182/asheducation-2013.1.354. PubMed PMID: 24319204.
111. Pakdee N, Yamsri S, et al. Variability of hemoglobin F expression in hemoglobin EE disease: hematological and molecular analysis. *Blood Cells Mol Dis*. 2014;53(1-2):11-5. Epub 2014/02/26. doi: 10.1016/j.bcmd.2014.02.005. PubMed PMID: 24581976.

112. Mejri A, Mansri M, et al. First description of the rs45496295 polymorphism of the C/EBPE gene in β -thalassemia intermedia patients. *Hemoglobin*. 2016;40(6):411-6. Epub 2016/11/09. doi: 10.1080/03630269.2016.1252386. PubMed PMID: 27829304.
113. Alicia K. Chang, Carly C. Ginter Summarell, et al. Sheehan* Department of Pediatrics, Division of Hematology/Oncology, Baylor College of Medicine, Houston, TX, USA; Genetic modifiers of severity in sickle cell disease; *Clinical Hemorheology and Microcirculation* 68 (2018) 147–164 DOI 10.3233/CH-189004 IOS Press.
114. Joanne Traeger-Synodinos, Cornelis L Hartevelde, et al. on behalf of contributors to the EMQN haemoglobinopathies best practice meeting; EMQN Best Practice Guidelines for molecular and haematology methods for carrier identification and prenatal diagnosis of the haemoglobinopathies; *European Journal of Human Genetics* (2015) 23, 426–437 & 2015 Macmillan Publishers Limited All rights reserved 1018-4813/15.
115. Thein SL, Menzel S, et al. Control of fetal hemoglobin: new insights emerging from genomics and clinical implications. *Human molecular genetics*, 18(R2):R216-R223; 2009.
116. Cao A, Moi P, et al. Recent advances in β -thalassemias. *Pediatric reports*. 3(2):17; 2011.
117. Nguyen TKT, Joly P, et al. The *XmnI* G γ polymorphism influences hemoglobin F synthesis contrary to *BCL11A* and *HBS1L-MYB* SNPs in a cohort of 57 β -thalassemia intermedia patients. *Blood Cells, Molecules, and Diseases*. 45(2):124-127; 2010.
118. Bhagat S, Patra PK, et al. Association between *XmnI* polymorphism and HbF level in sickle cell disease patients from Chhattisgarh. *Int J Biomed Sci*. 1:36-39; 2012.
119. Danjou F, Anni F, et al. Genetic modifiers of β -thalassemia and clinical severity as assessed by age at first transfusion. *Haematologica*. 97(7):989-993; 2012.
120. Al-Allawi NA, Puehringer H, et al. Genetic Modifiers in β -Thalassemia Intermedia: A Study on 102 Iraqi Arab Patients. *Genetic testing and molecular biomarkers*. 19(5):242-247; 2015.

121. Fanis P, Kousiappa I, et al. Genotyping of *BCL11A* and *HBS1L-MYB* SNPs associated with fetal haemoglobin levels: a SNaPshot minisequencing approach. *BMC genomics*. 15(1):1; 2014.
122. Moi P, Campus S. Recenti progressi in ematologia pediatrica. *INDICE* numero 157-158 Gennaio-Giugno 2010:21; 2010.
123. Wonkam A, Bitoungui VJN, et al. Association of variants at *BCL11A* and *HBS1L-MYB* with hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. *PLoS one*. 9(3):e92506; 2014.
124. Sara Gardenghi, Pedro Ramos, et al. Hepcidin and Hfe in iron overload in β -thalassemia. *Ann N Y Acad Sci*. 2010 August ; 1202: 221–225. doi:10.1111/j.1749-6632.2010.05595.x.
125. Mehran Karimi, Majid Yavarian, et al. Spectrum and haplotypes of the HFE hemochromatosis gene in Iran: H63D in β -thalassemia major and the first E277K homozygous. *The Hematology Journal* (2004) 5, 524–527.
126. Azza Aboul Enein, Nermine A. El Dessouky, et al. Frequency of Hereditary Hemochromatosis (HFE) Gene Mutations in Egyptian Beta Thalassemia Patients and its Relation to Iron Overload. *Macedonian Journal of Medical Sciences*. 2016 Jun 15; 4(2):226-231.
127. Anita H. Nadkarni, Aradhana A. Singh, et al. Effect of the Hemochromatosis Mutations on Iron Overload among the Indian β Thalassemia Carriers. *Journal of Clinical Laboratory Analysis* 00: 1–5 (2016).
128. Yunus Kasım Terzi, Tuğçe Bulakbaşı Balcı, et al. Effect of Hereditary Hemochromatosis Gene H63D and C282Y Mutations on Iron Overload in Sickle Cell Disease Patients. *Turk J Hematol* 2016;33:320-325
129. Angastiniotis M, Galanello R, et al. Prevention of thalassaemias and other haemoglobin disorders. Vol. 2: Laboratory Protocols, Thalassemia International Federation TIF Publication 2012.
130. Angastiniotis M, Galanello R, et al. Prevention of thalassaemias and other haemoglobin disorders. Vol. 2: Laboratory Protocols, Thalassemia International Federation TIF Publication 2012.
131. Clark BE, Shooter C, et al. Next-generation sequencing as a tool for breakpoint analysis in rearrangements of the globin gene clusters. *Int J Lab Hem*. 2017;39(Suppl. 1):111–120. [https:// doi.org/10.1111/ijlh.12680](https://doi.org/10.1111/ijlh.12680).

132. Y. Fichou, M. Mariez, et al. The experience of extended blood group genotyping by next-generation sequencing (NGS): investigation of patients with sickle-cell disease. 2016 International Society of Blood Transfusion DOI: 10.1111/vox.12432.
133. Fichou Y, Férec C. NGS and blood group systems: State of the art and perspectives. *Transfusion Clinique et Biologique* (2017), <http://dx.doi.org/10.1016/j.tracli.2017.06.002>.
134. Jiajie Pu, Li Zhang, et al. Clinical Genotyping by Next Generation Sequencing Reveals a Novel, De Novo beta-Globin Gene Mutation Causing Hemolytic Anemia in a Chinese Individual, *Hemoglobin*, DOI: 10.1080/03630269.2018.1496928.
135. Zhang H, Li C, et al. Next-generation sequencing improves molecular epidemiological characterization of thalassemia in Chenzhou Region, P.R. China. *J Clin Lab Anal.* 2019;e22845. <https://doi.org/10.1002/jcla.22845>.
136. Sutovsky P., Moreno R.D., et al. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos; *Biology of reproduction* 63:582-90.
137. Nishimura Y., Yoshinari T., et al. Active digestion of sperm mitochondrial DNA in single living sperm revealed by optical tweezers; *PNAS* 103:1382-7, (2006).
138. Innan H & Nordborg M (2002): Recombination or mutational hot spots in human mtDNA? *Mol. Biol. Evol.* 19(7):1122–27.
139. Metspalu M., Kivisild T., et al. (2006). The pioneer settlement of modern humans in Asia. In: Bandelt HJ, Macaulay V, Richards M, editors. *Human mitochondrial DNA and the Evolution of Homo sapiens*. Berlin: Springer-Verlag. p 181–99.
140. Ke Y., Su B., et al. (2001) African origin of modern human in East Asia: a tale of 12000 Y chromosomes; *Science* 292:1151-3.
141. Wallace D.C., Ye J.H., et al. (1987) Sequence analysis of cDNAs for the human and bovine ATP synthase beta subunit: mitochondrial DNA genes sustain seventeen times more mutations; *Curr. Genet.* 12:81-90.
142. Kaguni L.S. (2004) DNA polymerase γ , the mitochondrial replicase; *Annu. Rev. Biochem.* 73:293-320.

143. Wallace D.C. (1994): Mitochondrial DNA sequence variation in human evolution and disease; PNAS 91:8739-46.
144. Ruiz-Pesini E., Mishmar D., et al. (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA; Science 303:223-6.
145. Lee H.C. & Wei Y.H. (2009): Mitochondrial DNA instability and metabolic shifts in human cancers; Int. J. Mol. Sci. 10: 674-701.
146. Jakupciak J.P., Maragh S., et al. (2008): Performance of mitochondrial DNA mutations detecting early stage cancer; BMC Cancer 8:285-96.
147. Wilson A.C., Cann R.L., et al. (1985) Mitochondrial DNA and two perspectives on evolutionary genetics; Biol. J. Linn. Soc. 26:375-400.
148. Howell N., Elson J.L., et al. (2004): African Haplogroup L mtDNA sequences show violations of clock-like evolution; Mol. Biol. Evol. 21(10):1843–54.
149. Brown W.M. (1980) Polymorphism in mitochondrial DNA of human as revealed by restriction endonuclease analysis; PNAS 77:3605-9.
150. Cann R.L., Stoneking M. and Wilson A.C. (1987) Mitochondrial DNA and human evolution; Nature 325:31-6.
151. Relethford J.H. (2008) Genetic evidence and the modern human origins debate; Heredity, 100:555–63.
152. Vigilant L., Stoneking M., Harpending H., Hawkes K. and Wilson A.C. (1991) African populations and the evolution of human mitochondrial DNA; Science 253:1503-7.
153. Templeton AR (1992) Human origins and analysis of mitochondrial DNA sequences; Science 255: 737.
154. Penny D., Steel M., et al. (1995) Improved analyses of human mtDNA sequences support a recent African origin for Homo sapiens; Mol. Biol. Evol. 12: 863-82.
155. Watson E., Forster P., et al. (1997) Mitochondrial footprints of human expansions in Africa; Am. J. Hum. Genet. 61: 691-704.
156. Torroni A., Schurr T.G., et al. (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs; Am J Hum Genet. 53:563-90.
157. Ingman M., Kaessmann H., et al. (2000) Mitochondrial genome variation and the origin of modern humans; Nature 408:708-13.

158. Maca-Meyer N., Gonzàlez A.M., et al. (2001) Major genomic mitochondrial lineages delineate early human expansions; *BMC Genetics* 2:13.
159. Elson J.L., Andrews R.M., et al. (2001) Analysis of European mtDNAs for recombination; *Am. J. Hum. Genet.* 68:145-53.
160. Finnilä S., Lehtonen M. and Majamaa K. (2001) Phylogenetic Network for European mtDNA; *Am. J. Hum. Genet.* 68:1475–84.
161. Herrnstadt C., Elson J.L., et al. (2002): Reduced-Median-Network Analysis of Complete Mitochondrial DNA Coding-Region Sequences for the Major African, Asian, and European Haplogroups; *Am. J. Hum. Genet.* 70:1152–71.
162. Torroni A., Bandelt H.J., et al. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe; *Am.J.Hum.Genet.* 69:844-52.
163. Meyer S. & Von Haeseler A. (2003) Identifying site-specific substitution rates; *Mol. Biol. Evol.* 20: 182-9.
164. Cavalli-Sforza L.L. & Feldman L.W. (2003) The application of molecular genetic approaches to the study of human evolution; *Nature Genetic Supplement* 33:266-75.
165. Olivieri A., Achilli A., et al. (2006) The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa; *Science* 314:1767-70.
166. Chen Y.S., Torroni A., et al. (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups; *Am. J. Hum.Genet.* 57:133-49.
167. Torroni A., Lott M.T., et al. (1994) mtDNA and the Origin of Caucasians: identification of ancient Caucasianspecific haplogroups, one of which is prone to recurrent somatic duplication in the D-loop region; *Am. J. Hum. Genet.* 55:760-76.
168. Torroni A., Huoponen K., et al. (1996) Classification of European mtDNAs from an analysis of three European populations; *Genetics* 144:1835-50.
169. Schurr T.G., Ballinger S.W., et al. (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages; *Am. J. Hum. Genet.* 46: 613-23.
170. Richards M.B., Macaulay V.A., et al. (1998) Phylogeography of mitochondrial DNA in western Europe; *Ann. Hum. Genet.* 62:241-60.

171. Achilli A., Rengo C., et al. (2004) The molecular dissection of mtDNA Haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool; *Am.J.Hum.Genet.* 75:910-18.
172. Loogväli E.L., Roostalu U., et al. (2004). Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia; *Mol Biol Evol* 21(11):2012-21.
173. Kivisild T., Shen P., et al. (2006) The role of selection in the evolution of human mitochondrial genomes; *Genetics* 172:373-87.
174. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988; 16(3):1215. <https://doi.org/10.1093/nar/16.3.1215> PMID: 3344216; PubMed Central PMCID: PMC334765.
175. Sanger F, Nicklen S, et al. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977; 74(12):5463–7. <https://doi.org/10.1073/pnas.74.12.5463> PMID: 271968; PubMed Central PMCID: PMC431765.
176. Andrews RM, Kubacka I, et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 1999; 23(2):147. <https://doi.org/10.1038/13779> PMID: 10508508.
177. Van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 2009; 30(2):E386–94. <https://doi.org/10.1002/humu.20921> PMID: 18853457.
178. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009; 25(11):1451–2. Epub 2009/04/03. <https://doi.org/10.1093/bioinformatics/btp187> PMID: 19346325.
179. Cardinali I, Lancioni H, Giontella A, Capodiferro MR, Capomaccio S, Buttazzoni L, et al. An Overview of Ten Italian Horse Breeds through Mitochondrial DNA. *PLoS One.* 2016; 11(4):e0153004. Epub 2016/ 04/07. <https://doi.org/10.1371/journal.pone.0153004> PMID: 27054850; PubMed Central PMCID: PMC4824442.
180. Cerezo M, Achilli A, et al. Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res.* 2012; 22(5):821–6. <https://doi.org/10.1101/gr.134452.111> PMID: 22454235; PubMed Central PMCID: PMC3337428.

181. Derenko M, Malyarchuk B, et al. Complete mitochondrial DNA analysis of eastern Eurasian haplogroups rarely found in populations of northern Asia and eastern Europe. *PLoS One*. 2012; 7(2):e32179. Epub 2012/02/21. <https://doi.org/10.1371/journal.pone.0032179> PMID: 22363811; PubMed Central PMCID: PMC3283723.
182. Behar DM, van Oven M, et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet*. 2012; 90(4):675–84. <https://doi.org/10.1016/j.ajhg.2012.03.002> PMID: 22482806; PubMed Central PMCID: PMC3322232.
183. Swee Lay Thein. Genetic association studies in β -hemoglobinopathies. American Society of Hematology, Hematology 2013.
184. Thomas Benedict B., et al. Cutting not the key to TMPRSS6 activity? *Blood*. 2020 Aug 20; 136(8): 922-923.
185. Kloss-Brandstätter A, Erhart G, et al. Candidate Gene Sequencing of SLC11A2 and TMPRSS6 in a Family with Severe Anaemia: Common SNPs, Rare Haplotypes, No Causative. *PLoS One*. 2012; 7(4): e35015.
186. Jallow MW, Cerami C, et al. (2020). Differences in the frequency of genetic variants associated with iron imbalance among global populations. *PLoS ONE* 15(7): e0235141. <https://doi.org/10.1371/journal.pone.0235141>
187. Barbanera Y, Arcioni F, et al. (2020) Comprehensive analysis of mitochondrial and nuclear DNA variations in patients affected by hemoglobinopathies: A pilot study. *PLoS ONE* 15(10): e0240632. <https://doi.org/10.1371/journal.pone.0240632>

RINGRAZIAMENTI

Il lavoro di tesi di Dottorato è stato uno dei più impegnativi, perché frutto di un percorso formativo nel quale si sceglie di mettersi personalmente in gioco.

Ringrazio il Dott. Paolo Gorello innanzitutto per avermi dato l'opportunità di entrare a contatto anche con il mondo della ricerca, per aver sempre creduto nelle mie possibilità, per avermi incoraggiato ed avermi insegnato a guardare le cose da una prospettiva nuova. Inoltre, esprimo la mia riconoscenza per avermi trasmesso parte della sua esperienza e per avermi guidato nel mio percorso di studio con preziosi consigli.

Devo altresì ringraziare la Prof.ssa Paola Grammatico la quale ha permesso che il mio lavoro di ricerca potesse continuare dove il mio percorso formativo e professionale era iniziato.

Grazie a tutte le persone che in questi anni mi hanno concesso l'opportunità di crescere insieme a loro mentre cercavo di insegnare tutto quello che avevo appreso del bellissimo mondo della Genetica Molecolare. In questo percorso ho conosciuto belle persone che porterò sempre nel mio cuore, sono state un sostegno prezioso e un aiuto insostituibile.

Grazie alla mia famiglia che è il senso del mio tutto.

Appendice A

A1. Pazienti β -talassemici (β^+/β^+ , β^0/β^0 e β^+/β^0)

L'analisi dei geni β globinici ha confermato la diagnosi di β -talassemia major in quattro dei 15 casi in esame mentre 11 presentavano beta-talassemia minor, mentre un paziente è risultato essere omozigote per la variante emoglobinica talassemica HbE (pz 15). In un paziente sebbene non erano disponibili informazioni cliniche il genotipo si rivolgeva a una β -talassemia major (pz 1). Dal punto di vista clinico 12 (80%) erano dipendenti dalla trasfusione (TD); la frequenza trasfusionale e il relativo genotipo sono stati riportati nella tabella A1. Un paziente TD (pz 11) ha avuto un trapianto di cellule staminali ematopoietiche di successo dal suo fratello identico HLA quando aveva 4 anni. Tutti i pazienti analizzati hanno mostrato un genotipo α normale (Tabella A1).

Tabella A1. Dati clinici di pazienti con β talassemia.

N°	Genotipo β	Frequenza trasfusionale (GR/settimane)	HbF%	Età
2	codon 39(C>T)(β^0) / codon 44(-C)(β^0)	2/2		
3	- 29(A>G)(β^+) / codon 8(-AA)(β^0)	2/2		
4	- 29(A>G)(β^+) / codon 8(-AA)(β^0)	1/2		
5	codon 39(C>T)(β^0) / IVS-I-6(T>C)(β^{++})	1/3	80	7 mesi
6	IVS-I-110(G>A)(β^+) / IVS-I-110(G>A)(β^+)	1/3		
7	codon 17(A>T)(β^0) / - 29(A>G)(β^+)	1/3		
8	IVS-I-5(G>C)(β^+) / IVS-I-6(T>C)(β^{++})	1/3		
10	codon 39(C>T)(β^0) / IVS-I-6(T>C)(β^{++})	1/2		
11	codon 5 (-CT)(β^0) / IVS-I-6(T>C)(β^{++})	1/2		
12	codon 5 (-CT)(β^0) / IVS-I-6(T>C)(β^{++})	1/3	50.5	4 anni
13	- 28(A>G)(β^+) /cd17(A>T);AAG(Lys)>TAG(Stop codon)(β^0)	1/3		
14	IVS-II-745 (C>G) (β^+) / Hb Lepore Boston-Washington	2 / 2		

A2. Pazienti con Microdrepanocitosi (HbS/ β^+ o β^0)

Nel gruppo dei dieci pazienti con microdrepanocitosi per un non erano disponibili informazioni cliniche (pz 21). Un paziente ha sviluppato un linfoma di Burkitt a 34

anni di età e ha ottenuto la remissione completa dopo la chemioterapia (follow-up 6 mesi) (pz 17). Solo due pazienti (pz 19 e 20) erano fuori trattamento (a causa di un fenotipo apparentemente lieve). Un paziente ha sviluppato una crisi dolorosa/anno tra il 2014 e il 2016 senza necessità di trasfondere globuli rossi (pz 19); uno è morto a 49 anni a causa di una sindrome toracica acuta (pz 20). Uno ha avuto supporto trasfusionale solo durante una sepsi e ha realizzato protesi d'anca a seguito di osteonecrosi (pz 24). Tre pazienti (18, 23 e 25) hanno risposto bene all'idrossiurea (HU), riducendo /sospendendo le trasfusioni (pz 18,23 e 25) (vedi Tabella A2). Uno (pz 18) è stato sottoposto a splenectomia all'età di undici anni. I pazienti 16 e 22 sono in trattamento con HU e scambi di globuli rossi (manuali il pz 16, automatici il pz 22) a causa della recidiva di crisi dolorose (Tabella A2). La genotipizzazione dei geni α globinici è stata riportata nella tabella A2.

Tabella A2. Dati clinici di pazienti con Microdrepanocitosi.

N°	Genotipo β	Genotipo α	Trattamento	HbF%	Età
16	Hb S / codon 39(C>T)(β^0)	$\alpha\alpha/\alpha\alpha$	Flebotomia/ trasfusioni + HU		
17	Hb S / codon 39(C>T)(β^0)	$\alpha\alpha$ anti- 3.7/ $\alpha\alpha$	Scambio eritrocitario/HU		
18	Hb S / codon 39(C>T)(β^0)	$\alpha\alpha/\alpha\alpha$	Trasfusioni/HU		
19	Hb S / -29(A>G)(β^+)	$\alpha\alpha/\alpha\alpha$	-	2.9	6 anni
20	Hb S / IVS-I-6(T>C)(β^{++})	$\alpha\alpha/\alpha\alpha$	-	0.8	46 anni
21	Hb S / -29(A>G)(β^+)	$\alpha\alpha/\alpha\alpha$	Non Disponibile		
22	Hb S / IVS-I- 110(G>A)(β^{++})	$\alpha\alpha/\alpha\alpha$	Scambio eritrocitario /HU		
23	Hb S / IVS-I- 110(G>A)(β^+)	$\alpha\alpha/\alpha\alpha$	Trasfusioni /HU	11.6	18 mesi
24	Hb S / -29(A>G)(β^+)	$\alpha\alpha/\alpha\alpha$	Trasfusioni Sporadiche	5.8	19 anni
25	Hb S / codon 39(C>T)(β^0)	$\alpha\alpha$ anti- 3.7/ $\alpha\alpha$	HU	25.5	24 anni

A3. Pazienti con Drepanocitosi (HbS / HbS)

Dal punto di vista clinico nel gruppo dei pazienti affetti da drepanocitosi è emerso che: tre pazienti sono morti (pz 27, 29 e 39); per il paziente 27 e il 29 non abbiamo dati clinici prima della morte; per il 29, che era il fratello del paziente 28, sappiamo

che ha richiesto un solo trasfusione prima della morte, a causa dell'infezione dal virus Varicella-zoster. Abbiamo perso al follow-up i pazienti 36, 43 e 44; il paziente 36 iniziava HU nel 2015 dopo una precedente profilassi eritrocitaria a causa di crisi toraciche ricorrenti, con solo una riduzione della frequenza. Cinque pazienti (pz 26, 28, 30, 32 e 38) hanno sospeso la profilassi eritrocitaria dopo l'inizio del trattamento con HU, con una marcata riduzione delle crisi dolorose. I pazienti 31, fratello del paziente 30, erano intolleranti a HU e sono entrati nella profilassi trasfusionale per la velocità del flusso sanguigno elevata al TCD (Transcranial Doppler Ultrasound). I pazienti 33, 34 e 41 hanno sospeso la profilassi trasfusionale dopo l'inizio del trattamento con HU. Il paziente 33 aveva bisogno di trasfusioni sporadiche per episodi ricorrenti di infarti splenici, mentre il paziente 34 aveva frequenti sindromi toraciche acute ed è stato sottoposto a trapianto di cellule staminali ematopoietiche nel giugno 2019 da un identico fratello donatore. Il paziente 41 necessita ancora di globuli rossi per l'anemia, senza una significativa ricorrenza di crisi dolorose. Il paziente 35 è stato sottoposto a scambio eritrocitario a causa di crisi dolorose ricorrenti nonostante il trattamento con HU. I pazienti 37, 42 e 45 hanno ricevuto trasfusioni sporadiche per anemia senza indicazione alla profilassi trasfusionale (TCD normale, bassa recidiva di crisi dolorose, nessuna sindrome toracica acuta). Il paziente 40 non ha avuto bisogno di cure perché finora ha subito solo una crisi dolorosa (il paziente ha ad oggi 14 anni). La genotipizzazione dell' α -globina è stata riportata nella tabella A3.

Tabella A3. Dati clinici di pazienti con Drepanocitosi.

N°	Genotipo α	Trattamento	HbF%	Età
26	$\alpha\alpha/\alpha\alpha$	HU	3.4	7 anni
27	$\alpha\alpha/\alpha\alpha$	Non Disponibile	4.7	32 anni
28	$-\alpha^{3.7}/\alpha\alpha$	HU	18.7	4 anni
29	$\alpha\alpha/\alpha\alpha$	ND	34.6	5 mesi
30	$\alpha\alpha/\alpha\alpha$	HU		
31	$\alpha\alpha/\alpha\alpha$	Profilassi trasfusionale	10.8	2 anni
32	$\alpha\alpha/\alpha\alpha$	HU	12.4	5 anni

33	- $\alpha^{3.7}/\alpha\alpha$	Trasfusioni Sporadiche+ HU	20	2 anni
34	$\alpha\alpha/\alpha\alpha$	Trasfusioni Sporadiche + HU	30	5 mesi
35	- $\alpha^{3.7}/\alpha\alpha$	Scambi eritrocitari	4	49 anni
36	$\alpha\alpha/\alpha\alpha$	Perso Follow Up	10.5	2 anni
37	$\alpha\alpha/\alpha\alpha$	Trasfusioni Sporadiche	45	1 anni
38	- $\alpha^{3.7}/\alpha\alpha$	HU	3	14 anni
39	$\alpha\alpha/\alpha\alpha$	Non Disponibili		
40	- $\alpha^{3.7}/\alpha\alpha$	Nessuna	16.3	12 anni
41	$\alpha\alpha/\alpha\alpha$	Trasfusioni Sporadiche + HU		
42	$\alpha^{-3.7}/\alpha^{-3.7}$	Trasfusioni Sporadiche	1.6	60 anni
43	$\alpha\alpha/\alpha\alpha$	Perso Follow-Up	24.5	9 mesi
44	$\alpha\alpha/\alpha\alpha$	Perso Follow-Up		
45	$\alpha\alpha/\alpha\alpha$	Trasfusioni Sporadiche	11.2	14 mesi

A4. Pazienti eterozigoti composti (HbS/HbC e HbC/O-Arab)

Nel gruppo dei otto pazienti eterozigoti composti due (pz 46 e 51) sono stati sottoposti a scambi eritrocitari. La paziente 46 è stata persa per il follow-up, mentre il 51 ha sospeso gli scambi eritrocitari e non ha avuto crisi dolorose (la paziente ha rifiutato il trattamento HU a causa del suo desiderio di rimanere incinta). I pazienti 47-50, 52 e 53 non sono stati sottoposti ad alcun trattamento. In particolare, la paziente 53 ha prodotto una sola trasfusione durante un episodio di crisi dolorosa ed è stata persa a follow-up dal 2017. La paziente 48 è stata persa a follow-up dal 2016, prima di questa data non ha sviluppato alcun problema clinico correlato all'emoglobinopatia. La genotipizzazione β e α è stata riportata nella tabella A4.

Tabella A4. Dati clinici di pazienti eterozigoti composti

N°	Genotipo β	Genotipo α	Trattamento	HbF%	Età
46	HbS / Hb C	- $\alpha^{3.7}/\alpha\alpha$	Perso Follow-Up	1.3	38 anni
47	HbS / Hb C	$\alpha\alpha/\alpha\alpha$	Nessuno	12.8	23 mesi

48	HbS / Hb C	$\alpha\alpha/\alpha\alpha$	Perso Follow-Up	24.1	1 anno
49	HbS / Hb C	$-\alpha^{3.7}/\alpha\alpha$	Nessuno	1.4	14 anni
50	HbS / Hb C	$\alpha\alpha/\alpha\alpha$	Nessuno	3.4	16 anni
51	HbS / Hb C	$\alpha\alpha/\alpha\alpha$	Nessuno	4.4	24 anni
52	Hb C / Hb O-Arab	$\alpha\alpha/\alpha\alpha$	Nessuno	0.7	59 anni
53	HbS / Hb C	$\alpha\alpha/\alpha\alpha$	Perso Follow-Up	3	17 anni

Appendice B

Gene list/target regions: the panel includes

all protein-coding exons for all transcripts of a given gene for the following genes with ± 5 bp flanking into the intron:

ANKRD26, ASF1A, ASF1B, ATRX, BCL11A, BLM, BMP2, BMP6, CBL, CEBPA, COX4I1, CTC1, CTCF, DAXX, DCLRE1B, DDX41, DICER1, DKC1, DNAJC21, ERF, ETV6, EZH2, FTL, FUS, GAR1, GATA2, GNAS, HAMP, HFE, HFE2, KLF1, LIG4, MBD4, MECOM, MRE11, NAF1, NBN, NHP2, NOP10, NSD1, NSD3, PABPN1, PARN, PIF1, PINX1, PKLR, POT1, RAD50, RECQL4, RIF1, RTEL1, RUNX1, SAMD9, SAMD9L, SLC11A2, SLC40A1, SRP72, STN1, TERC, TEN1, TERF1, TERF2, TERF2IP, TERT, TFR2, TIN2, TMPRSS6, TPP1, USB1, WRAP53, ZEB2

full UTR regions in addition to the exonic regions for the following genes (both 5' and 3' where present):

BCL11A (NM_018014.3, NM_022893.3, NM_138559.1), BMP2 (NM_001200.3), BMP6 (NM_001718.5), CEBPA (NM_001285829.1), COX4I1 (NM_001861.4, NM_001318786.1, NM_001318788.1, NM_001318797.1), ERF (NM_001291832.1), FTL (NM_000146.3), HAMP (NM_021175.3), HFE (NM_000410.3, NM_139003.2, NM_139004.2, NM_139006.2, NM_139007.2, NM_139008.2, NM_139009.2, NM_139010.2, NM_139011.2, NM_001300749.1), HFE2 (NM_145277.4, NM_202004.3, NM_213652.3, NM_213653.3, NM_001316767.1), KLF1 (NM_006563.4), PKLR (NM_000298.5, NM_181871.3), SLC11A2 (NM_000617.2, NM_001174125.1, NM_001174126.1, NM_001174127.1, NM_001174128.1, NM_001174129.1, NM_001174130.1), SLC40A1 (NM_014585.5), TFR2 (NM_003227.3, NM_001206855.1), TMPRSS6 (NM_153609.3, NM_001289000.1, NM_001289001.1)

5'UTR region for the following gene:

ANKRD26 (NM_014915.2, NM_001256053.1)

all exonic, intronic and UTR regions for the following genes:

HBA1 (NM_000558.4), HBA2 (NM_000517.4), HBB (NM_000518.4), HBD (NM_000519.3)

Sequencing recommendations:

	SAMPLES PER RUN	LIBRARY DILUTION	LOADING DILUTION
MiSeq V3	24	4 nM	10 pM

Appendice C

C1. Pazienti β -talassemici (β^+/β^+ , β^0/β^0 e β^+/β^0)

In 13 dei 15 pazienti analizzati è stata riscontrata almeno un'alterazione e solo due (pz 5 e 14) sono risultati negativi per tutti gli SNPs indagati (*Tabella 14*). Inoltre, i pazienti 3 e 4 (fratelli) hanno mostrato lo stesso genotipo SNP, mentre i pazienti 11-12 (fratelli) hanno mostrato genotipi diversi.

BCL11A, rs1427407 [G> T] era l'alterazione meno polimorfica, con un solo paziente mutato dall'Asia (pz 11). Diversamente, *BCL11A* rs10189857 [A> G] e *HBS1L-MYB* rs28384513 [A> C] erano le variazioni più polimorfiche. In particolare, l'allele G rs10189857, è stato rilevato in 12 dei 15 pazienti (sei omozigoti e sei eterozigoti). Questi pazienti provenivano da quattro continenti: quattro dall'Europa, due dall'Africa, quattro dall'Asia e uno dal Sud America.

HBS1L-MYB, rs28384513 A> C, è stato rilevato in sette dei 15 pazienti (un omozigote e sei eterozigoti). Cinque pazienti provenivano dall'Europa e due dall'Asia, mentre i pazienti americani e africani portatori di rs28384513 A> C non sono stati trovati. Allo stesso modo, *HBS1L-MYB* rs9399137 [T> C], è stato rilevato solo in quattro dei 15 pazienti, tre europei e uno asiatico.

Infine, *HBG2* rs7482144 [C> T] è stato rilevato in quattro dei 15 pazienti: due africani, uno sudamericano e uno asiatico. Non sono stati identificati pazienti europei portatori di questo SNP.

C2. Pazienti con Microdrepanocitosi (HbS/β^+ o β^0)

In tutti i dieci pazienti con microdrepanocitosi è stata rilevata almeno una variazione genica (*Tabella 14*). Le varianti meno polimorfiche erano *HBG2*, rs7482144 [C> T] e *HBS1L-MYB*, rs9399137 [T> C], entrambe rilevate solo in due pazienti. *HBG2*,

rs7482144 [C> T] è stato rilevato in un paziente africano e in un paziente americano (pz 21 e 24). Non è stato trovato nei pazienti europei. Al contrario, *HBS1L-MYB*, rs9399137 [T> C] è stato rilevato solo in pazienti europei (pz 18 e 25)

BCL11A rs10189857 [A> G] e *HBS1L-MYB* rs28384513 [A> C] sono state le variazioni più polimorfiche. In particolare, *BCL11A* rs10189857 [A> G] è stato identificato in sette dei 10 pazienti (due omozigoti e cinque eterozigoti). Cinque di questi pazienti erano europei, uno africano e uno dall'America. *HBS1L-MYB* rs28384513 [A> C] era presente in cinque dei 10 pazienti (pz eterozigoti: 16, 18-20, 25): quattro europei e un paziente africano. Non è stato trovato nei pazienti americani.

BCL11A rs1427407 [G> T] è stato rilevato in quattro dei 10 pazienti: due europei, uno africano e uno americano.

C3. Pazienti con Drepanocitosi (HbS / HbS)

In 18 pazienti con drepanocitosi su 20 è stata rilevata almeno una variante genomica mentre in due pazienti (pz 29 e 45) non sono state identificate varianti nei geni modulatori.

A differenza dei sottogruppi talassemia e microdrepanocitosi, non sono stati identificati pazienti con *HBS1L-MYB* rs9399137 [T> C]. *HBG2* rs7482144 [C> T] è stato rilevato solo in tre dei 20 pazienti (pz 39-41): due africani e un paziente americano mentre non è stato rilevato nei pazienti europei.

BCL11A rs1427407 [G> T] è stato rilevato in nove dei 20 pazienti (un omozigote e otto eterozigoti). Sette pazienti provenivano dall'Africa e due dall'Europa mentre non è stato identificato nei pazienti americani.

BCL11A rs10189857 [A> G], è stato identificato in otto dei 20 pazienti: cinque africani, due americani e un europeo. Tre pazienti erano omozigoti e cinque erano eterozigoti.

Inoltre, *HBS1L-MYB* rs28384513 [A> C], è stato identificato in sette dei 20 pazienti provenienti dall'Africa.

Sono state analizzate due coppie di fratelli (pz 28, 29 e 30, 31). I fratelli 28 e 29 hanno mostrato genotipi differenti: *BCL11A* rs1427407 [G> T] era presente nel paziente 28 mentre non sono state identificate alterazioni nel paziente 29. Allo

stesso modo, *BCL11A* rs1427407 [G> T] era presente nel paziente 30 mentre rs10189857 [A> G] era rilevato in suo fratello (pz 31).

C4. Pazienti eterozigoti composti (HbS/HbC e HbC/O-Arab)

In tutti gli otto pazienti eterozigoti composti è stata identificata almeno una variazione genetica nei geni modulatori. *BCL11A* rs1427407 [G> T] e rs10189857 [A> G] sono state le varianti più polimorfiche, rilevate rispettivamente in cinque dei otto pazienti (un omozigote: paziente 50 e quattro eterozigoti: 46-47, 49 e 51) e in 6 pazienti (un paziente omozigote 48 e cinque pazienti eterozigoti: 47, 49 e 51-53). *HBS1L-MYB* rs28384513 [A> C] è stato rilevato, in eterozigosi, in due pazienti (pz 48 e 52). Tutti i pazienti erano africani. Non sono stati trovati pazienti portatori di *HBS1L-MYB* rs9399137 [T> C] e *HBG2* rs7482144 [C> T].

Appendice D. Varianti retained identificate nel paziente 3.

Category	gene	type	Coding Consequence	refGenome	chr	Genome position	depth	Var %	Exon rank	c.DNA	protein
A	<i>HBB</i>	SNP	intergenic	GRCh37/hg19	11	5248330	1675	50,6	1		
A	<i>HBB</i>	INDEL	frameshift	GRCh37/hg19	11	5248225	1653	48,9	1	c.25_26del	p.(Lys9Valfs*14)
B	<i>HFE</i>	SNP	missense	GRCh37/hg19	6	26091179	1126	46,7	2	c.187C>G	p.(His63Asp)
C	<i>BMP6</i>	SNP	synonymous	GRCh37/hg19	6	7862556	1096	42,9	4	c.1029C>T	p.(Ala343=)
C	<i>DAXX</i>	SNP	missense	GRCh37/hg19	6	33287796	1142	48,6	5	c.1457C>G	p.(Ala486Gly)
C	<i>PINX1</i>	SNP	synonymous	GRCh37/hg19	8	10623289	1138	50,3	7	c.609T>C	p.(Ser203=)
C	<i>RIF1</i>	SNP	missense	GRCh37/hg19	2	1,52E+08	1493	49,7	30	c.5363T>C	p.(Leu1788Ser)
C	<i>WRAP53</i>	SNP	missense	GRCh37/hg19	17	7592543	1503	49,6	3	c.433G>A	p.(Ala145Thr)
D	<i>ANKRD26</i>	SNP	missense	GRCh37/hg19	10	27389197	1402	49,9	1	c.59A>G	p.(Gln20Arg)
D	<i>BCL11A</i>	SNP	3'UTR	GRCh37/hg19	2	60679226	1480	99,7	5	c.*484C>T	
D	<i>BCL11A</i>	SNP	synonymous	GRCh37/hg19	2	60687959	1411	99,4	4	c.2088T>C	p.(Ser696=)
D	<i>BMP2</i>	SNP	3'UTR	GRCh37/hg19	20	6760344	1523	99,6	3	c.*608A>G	
D	<i>BMP2</i>	SNP	synonymous	GRCh37/hg19	20	6751034	951	99,8	2	c.261A>G	p.(Ser87=)
D	<i>BMP2</i>	SNP	missense	GRCh37/hg19	20	6759115	1103	99,9	3	c.570A>T	p.(Arg190Ser)
D	<i>BMP6</i>	SNP	synonymous	GRCh37/hg19	6	7862631	1189	99,9	4	c.1104G>C	p.(Val368=)
D	<i>BMP6</i>	INDEL	3'UTR	GRCh37/hg19	6	7881853	1360	42,4	7	c.*1277_*1289delinsCC	
D	<i>BMP6</i>	SNP	3'UTR	GRCh37/hg19	6	7881931	1236	44,7	7	c.*1355T>C	
D	<i>BMP6</i>	SNP	missense	GRCh37/hg19	6	7727271	707	53,5	1	c.83G>A	p.(Arg28Gln)
D	<i>BMP6</i>	SNP	3'UTR	GRCh37/hg19	6	7881311	1681	46,1	7	c.*735A>G	
D	<i>CBL</i>	INDEL	inframe_3	GRCh37/hg19	11	1,19E+08	907	46,6	1	c.125_127dup	p.(His42dup)
D	<i>CEBPA</i>	INDEL	inframe_6	GRCh37/hg19	19	33792731	728	50,3	1	c.227_232dup	p.(His76_Pro77dup)
D	<i>COX4I1</i>	SNP	synonymous	GRCh37/hg19	16	85838653	950	100	3	c.184T>C	p.(Leu62=)
D	<i>CTC1</i>	SNP	splice_donor_+3	GRCh37/hg19	17	8141707	1647	100	3	c.435+3A>G	
D	<i>CTC1</i>	SNP	synonymous	GRCh37/hg19	17	8134785	1764	51,6	15	c.2478A>G	p.(Thr826=)
D	<i>CTC1</i>	SNP	missense	GRCh37/hg19	17	8135061	1634	100	14	c.2458A>G	p.(Ile820Val)
D	<i>DAXX</i>	SNP	synonymous	GRCh37/hg19	6	33287183	1134	48,5	6	c.1914A>G	p.(Gln638=)
D	<i>DCLRE1B</i>	SNP	missense	GRCh37/hg19	1	1,14E+08	773	50,3	1	c.181C>T	p.(His61Tyr)

D	<i>DCLRE1B</i>	SNP	synonymous	GRCh37/hg19	1	1,14E+08	735	46,7	2	c.234T>C	p.(His78=)
D	<i>DDX41</i>	SNP	synonymous	GRCh37/hg19	5	1,77E+08	1450	51,2	10	c.822C>T	p.(Arg274=)
D	<i>DKC1</i>	SNP	synonymous	GRCh37/hg19	X	1,54E+08	614	99,8	5	c.369G>T	p.(Thr123=)
D	<i>ERFE</i>	SNP	3'UTR	GRCh37/hg19	2	2,39E+08	1481	52,3	8	c.*889G>C	
D	<i>ERFE</i>	SNP	3'UTR	GRCh37/hg19	2	2,39E+08	1499	99,8	8	c.*240A>G	
D	<i>ERFE</i>	INDEL	3'UTR	GRCh37/hg19	2	2,39E+08	1216	99,8	8	c.*1773del	
D	<i>FTL</i>	SNP	synonymous	GRCh37/hg19	19	49469087	1131	50,2	2	c.163T>C	p.(Leu55=)
D	<i>FUS</i>	SNP	synonymous	GRCh37/hg19	16	31195279	984	99,8	4	c.288C>T	p.(Tyr96=)
D	<i>GAR1</i>	SNP	synonymous	GRCh37/hg19	4	1,11E+08	690	48,7	2	c.69T>C	p.(Gly23=)
D	<i>GATA2</i>	SNP	synonymous	GRCh37/hg19	3	1,28E+08	1261	99,9	3	c.15C>G	p.(Pro5=)
D	<i>GATA2</i>	SNP	missense	GRCh37/hg19	3	1,28E+08	1751	99,9	4	c.490G>A	p.(Ala164Thr)
D	<i>GNAS</i>	SNP	synonymous	GRCh37/hg19	20	57478807	1354	99,9	5	c.393C>T	p.(Ile131=)
D	<i>GNAS</i>	SNP	synonymous	GRCh37/hg19	20	57415812	1485	53,1	1	c.651T>A	p.(Arg217=)
D	<i>HBA2</i>	SNP	3'UTR	GRCh37/hg19	16	223706	1389	100	3	c.*107A>G	
D	<i>HBB</i>	SNP	intronic	GRCh37/hg19	11	5247733	1739	50,4	2	c.315+74T>G	
D	<i>HBB</i>	SNP	intronic	GRCh37/hg19	11	5247726	1762	48,9	2	c.315+81C>T	
D	<i>HBB</i>	SNP	intronic	GRCh37/hg19	11	5247791	1756	50,1	2	c.315+16G>C	
D	<i>HBB</i>	SNP	synonymous	GRCh37/hg19	11	5248243	1618	52,7	1	c.9T>C	p.(His3=)
D	<i>HBB</i>	SNP	intronic	GRCh37/hg19	11	5247141	1550	43,8	3	c.316-185C>T	
D	<i>HFE</i>	SNP	splice_donor_+4	GRCh37/hg19	6	26091336	940	99,8	2	c.340+4T>C	
D	<i>LIG4</i>	SNP	synonymous	GRCh37/hg19	13	1,09E+08	1401	48,9	2	c.1704T>C	p.(Asp568=)
D	<i>NAF1</i>	SNP	missense	GRCh37/hg19	4	1,64E+08	802	100	8	c.1102C>G	p.(Leu368Val)
D	<i>NAF1</i>	SNP	missense	GRCh37/hg19	4	1,64E+08	1032	99,9	2	c.484A>G	p.(Ile162Val)
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90955583	726	45,3	15	c.1836T>G	p.(Pro612=)
D	<i>NBN</i>	SNP	missense	GRCh37/hg19	8	90990479	888	44	6	c.307G>C	p.(Glu103Gln)
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90958422	770	47,1	14	c.1770A>G	p.(Pro590=)
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90967711	844	98,7	11	c.951T>C	p.(Asp317=)
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90995019	694	100	2	c.102G>A	p.(Leu34=)
D	<i>NSD1</i>	SNP	synonymous	GRCh37/hg19	5	1,77E+08	1849	48,6	23	c.6903G>C	p.(Gly2301=)
D	<i>NSD1</i>	SNP	synonymous	GRCh37/hg19	5	1,77E+08	1803	99,9	23	c.6829T>C	p.(Leu2277=)

D	<i>PIF1</i>	SNP	synonymous	GRCh37/hg19	15	65116328	892	100	2	c.207T>C	p.(Phe69=)
D	<i>PINX1</i>	SNP	synonymous	GRCh37/hg19	8	10623214	1192	51,4	7	c.684C>G	p.(Leu228=)
D	<i>RECQL4</i>	INDEL	intergenic	GRCh37/hg19	8	1,46E+08	850	99,5	15	c.2296+1del	
D	<i>RECQL4</i>	SNP	synonymous	GRCh37/hg19	8	1,46E+08	945	100	19	c.3127T>C	p.(Leu1043=)
D	<i>RECQL4</i>	SNP	missense	GRCh37/hg19	8	1,46E+08	883	99,9	4	c.274T>C	p.(Ser92Pro)
D	<i>RIF1</i>	SNP	synonymous	GRCh37/hg19	2	1,52E+08	1017	46,5	27	c.3093G>A	p.(Ser1031=)
D	<i>RIF1</i>	SNP	missense	GRCh37/hg19	2	1,52E+08	1072	99,8	35	c.7174C>G	p.(Leu2392Val)
D	<i>RIF1</i>	SNP	missense	GRCh37/hg19	2	1,52E+08	1371	99,9	30	c.4084G>A	p.(Val1362Met)
D	<i>RIF1</i>	SNP	missense	GRCh37/hg19	2	1,52E+08	1517	99,8	30	c.6061A>T	p.(Asn2021Tyr)
D	<i>RIF1</i>	SNP	missense	GRCh37/hg19	2	1,52E+08	1256	99,9	22	c.2506G>A	p.(Gly836Ser)
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/hg19	20	62320968	1039	100	23	c.1992T>C	p.(Asp664=)
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/hg19	20	62322288	988	99,8	27	c.2544T>C	p.(Pro848=)
D	<i>RTEL1</i>	SNP	missense	GRCh37/hg19	20	62326110	738	99,7	32	c.3126A>C	p.(Gln1042His)
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/hg19	20	62321655	1297	99,9	26	c.2274G>A	p.(Ala758=)
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/hg19	20	62321189	921	52,4	24	c.2112C>T	p.(Asp704=)
D	<i>SLC11A2</i>	SNP	intronic	GRCh37/hg19	12	51380232	1457	99,6	16	c.1629+1933T>C	
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh37/hg19	12	51375066	1603	99,8	17	c.*443C>T	
D	<i>SLC11A2</i>	SNP	intronic	GRCh37/hg19	12	51381077	1508	100	16	c.1629+1088G>A	
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh37/hg19	12	51373742	934	99,6	17	c.*1767G>A	
D	<i>SLC11A2</i>	SNP	missense	GRCh37/hg19	12	51402905	819	100	1	c.17A>C	p.(Tyr6Ser)
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh37/hg19	12	51374748	1602	99,4	17	c.*761C>A	
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh37/hg19	12	51374250	1089	100	17	c.*1259T>C	
D	<i>SLC11A2</i>	SNP	intronic	GRCh37/hg19	12	51419621	1219	99,4	1	c.-39+472A>G	
D	<i>SLC11A2</i>	SNP	intronic	GRCh37/hg19	12	51420164	1025	100	1	c.49+1796G>T	
D	<i>SLC40A1</i>	SNP	3'UTR	GRCh37/hg19	2	1,9E+08	1630	48,5	8	c.*1130A>G	
D	<i>SRP72</i>	SNP	synonymous	GRCh37/hg19	4	57356558	1541	45,8	12	c.1197T>C	p.(Tyr399=)
D	<i>SRP72</i>	SNP	synonymous	GRCh37/hg19	4	57333822	1342	44	1	c.21G>T	p.(Gly7=)
D	<i>STN1</i>	SNP	missense	GRCh37/hg19	10	1,06E+08	753	47,8	7	c.743C>G	p.(Ser248Cys)
D	<i>STN1</i>	SNP	missense	GRCh37/hg19	10	1,06E+08	770	48,6	5	c.451A>G	p.(Thr151Ala)
D	<i>TEN1</i>	SNP	synonymous	GRCh37/hg19	17	73987652	729	99,9	3	c.198C>T	p.(His66=)

D	<i>TERF1</i>	SNP	synonymous	GRCh37/hg19	8	73958310	892	50,3	9	c.1198T>C	p.(Leu400=)
D	<i>TERF1</i>	INDEL	inframe_3	GRCh37/hg19	8	73921283	755	38,8	1	c.186_188dup	p.(Glu62dup)
D	<i>TERF1</i>	SNP	synonymous	GRCh37/hg19	8	73958240	858	50,9	9	c.1128G>A	p.(Arg376=)
D	<i>TINF2</i>	SNP	synonymous	GRCh37/hg19	14	24709506	1274	48,7	7	c.1092G>A	p.(Leu364=)
D	<i>TMPRSS6</i>	SNP	intronic	GRCh37/hg19	22	37499692	1102	99,5	2	c.-1-234T>C	
D	<i>TMPRSS6</i>	SNP	synonymous	GRCh37/hg19	22	37462926	1046	99,8	18	c.2256C>T	p.(Tyr752=)
D	<i>TMPRSS6</i>	INDEL	splice_acceptor_-1	GRCh37/hg19	22	37465385	657	98,6	16	c.1842-6_1842-2del	p.(?)
D	<i>TMPRSS6</i>	SNP	synonymous	GRCh37/hg19	22	37480797	412	99,8	9	c.1056G>A	p.(Ser352=)
D	<i>TMPRSS6</i>	SNP	synonymous	GRCh37/hg19	22	37499565	1103	48,7	1	c.15C>T	p.(Phe5=)
D	<i>TMPRSS6</i>	SNP	missense	GRCh37/hg19	22	37485724	1041	99,4	7	c.730A>G	p.(Lys244Glu)
D	<i>TMPRSS6</i>	SNP	missense	GRCh37/hg19	22	37462936	1066	99,6	18	c.2246T>C	p.(Val749Ala)
D	<i>TMPRSS6</i>	SNP	5'UTR	GRCh37/hg19	22	37505552	1218	99,8	1	c.-11G>A	
D	<i>TMPRSS6</i>	SNP	synonymous	GRCh37/hg19	22	37469591	457	100	13	c.1536C>T	p.(Asp512=)
D	<i>TPP1</i>	SNP	synonymous	GRCh37/hg19	11	6638599	1007	51,8	5	c.441A>G	p.(Glu147=)

Appendice E. Varianti retained identificate nel paziente 4.

Category	gene	type	Coding Consequence	refGenome	chr	Genome position	depth	Var %	Exon rank	c.DNA	protein
A	<i>HBB</i>	SNP	intergenic	GRCh37/hg19	11	5248330	1866	47,9	1		
A	<i>HBB</i>	INDEL	frameshift	GRCh37/hg19	11	5248225	1924	51,6	1	c.25_26del	p.(Lys9Valfs*14)
B	<i>HFE</i>	SNP	missense	GRCh37/hg19	6	26091179	1262	50,2	2	c.187C>G	p.(His63Asp)
B	<i>PINX1</i>	SNP	synonymous	GRCh37/hg19	8	10692242	794	47,2	2	c.63C>T	p.(Ala21=)
C	<i>BMP6</i>	SNP	synonymous	GRCh37/hg19	6	7862556	1109	50,8	4	c.1029C>T	p.(Ala343=)
C	<i>DAXX</i>	SNP	missense	GRCh37/hg19	6	33287796	1280	45,6	5	c.1457C>G	p.(Ala486Gly)
C	<i>MBD4</i>	SNP	missense	GRCh37/hg19	3	1,29E+08	1360	47,6	3	c.1073T>C	p.(Ile358Thr)
C	<i>WRAP53</i>	SNP	missense	GRCh37/hg19	17	7592543	1777	49,7	3	c.433G>A	p.(Ala145Thr)
D	<i>ANKRD26</i>	SNP	missense	GRCh37/hg19	10	27389197	1499	99,8	1	c.59A>G	p.(Gln20Arg)
D	<i>BCL11A</i>	SNP	synonymous	GRCh37/hg19	2	60687959	1611	48,4	4	c.2088T>C	p.(Ser696=)
D	<i>BCL11A</i>	SNP	3'UTR	GRCh37/hg19	2	60679226	1560	50,7	5	c.*484C>T	
D	<i>BCL11A</i>	INDEL	5'UTR	GRCh37/hg19	2	60780421	736	17	1	c.-19_-17del	
D	<i>BMP2</i>	INDEL	3'UTR	GRCh37/hg19	20	6759758	702	18,1	3	c.*41_*42del	
D	<i>BMP2</i>	SNP	synonymous	GRCh37/hg19	20	6751034	1022	99,9	2	c.261A>G	p.(Ser87=)
D	<i>BMP2</i>	SNP	3'UTR	GRCh37/hg19	20	6760344	1575	99,7	3	c.*608A>G	
D	<i>BMP2</i>	SNP	missense	GRCh37/hg19	20	6759115	1093	100	3	c.570A>T	p.(Arg190Ser)
D	<i>BMP6</i>	SNP	synonymous	GRCh37/hg19	6	7862631	1224	99,8	4	c.1104G>C	p.(Val368=)
D	<i>BMP6</i>	INDEL	3'UTR	GRCh37/hg19	6	7881853	1531	43,6	7	c.*1277_*1289delinsCC	
D	<i>BMP6</i>	SNP	missense	GRCh37/hg19	6	7727271	869	60,8	1	c.83G>A	p.(Arg28Gln)
D	<i>BMP6</i>	SNP	3'UTR	GRCh37/hg19	6	7881931	1494	43,2	7	c.*1355T>C	
D	<i>BMP6</i>	SNP	3'UTR	GRCh37/hg19	6	7881311	1882	46,8	7	c.*735A>G	
D	<i>CBL</i>	SNP	synonymous	GRCh37/hg19	11	1,19E+08	1547	50,6	16	c.2592C>T	p.(Leu864=)
D	<i>COX4I1</i>	SNP	synonymous	GRCh37/hg19	16	85838653	1187	100	3	c.184T>C	p.(Leu62=)
D	<i>CTC1</i>	SNP	splice_donor_+3	GRCh37/hg19	17	8141707	1839	99,8	3	c.435+3A>G	
D	<i>CTC1</i>	SNP	synonymous	GRCh37/hg19	17	8134785	1916	47,5	15	c.2478A>G	p.(Thr826=)
D	<i>CTC1</i>	SNP	missense	GRCh37/hg19	17	8135061	1891	99,8	14	c.2458A>G	p.(Ile820Val)
D	<i>DAXX</i>	SNP	synonymous	GRCh37/hg19	6	33287183	1288	45,7	6	c.1914A>G	p.(Gln638=)
D	<i>DCLRE1B</i>	SNP	missense	GRCh37/hg19	1	1,14E+08	866	49	1	c.181C>T	p.(His61Tyr)
D	<i>DCLRE1B</i>	SNP	synonymous	GRCh37/hg19	1	1,14E+08	870	52,2	2	c.234T>C	p.(His78=)

D	<i>DDX41</i>	SNP	synonymous	GRCh37/hg19	5	1,77E+08	1803	49,3	10	c.822C>T	p.(Arg274=)
D	<i>DKC1</i>	SNP	synonymous	GRCh37/hg19	X	1,54E+08	1305	99,7	5	c.369G>T	p.(Thr123=)
D	<i>ERFE</i>	INDEL	3'UTR	GRCh37/hg19	2	2,39E+08	1440	99,9	8	c.*1773del	
D	<i>ERFE</i>	SNP	3'UTR	GRCh37/hg19	2	2,39E+08	1664	51,3	8	c.*240A>G	
D	<i>FTL</i>	SNP	synonymous	GRCh37/hg19	19	49469087	1299	49,2	2	c.163T>C	p.(Leu55=)
D	<i>FUS</i>	SNP	synonymous	GRCh37/hg19	16	31195279	1122	47,9	4	c.288C>T	p.(Tyr96=)
D	<i>FUS</i>	SNP	synonymous	GRCh37/hg19	16	31193942	1270	47,7	3	c.147C>A	p.(Gly49=)
D	<i>GAR1</i>	SNP	synonymous	GRCh37/hg19	4	1,11E+08	742	100	2	c.69T>C	p.(Gly23=)
D	<i>GATA2</i>	SNP	synonymous	GRCh37/hg19	3	1,28E+08	1439	99,8	3	c.15C>G	p.(Pro5=)
D	<i>GATA2</i>	SNP	missense	GRCh37/hg19	3	1,28E+08	1845	47,9	4	c.490G>A	p.(Ala164Thr)
D	<i>GNAS</i>	SNP	missense	GRCh37/hg19	20	57428474	1858	48,4	1	c.154G>A	p.(Glu52Lys)
D	<i>GNAS</i>	SNP	synonymous	GRCh37/hg19	20	57478807	1581	99,6	5	c.393C>T	p.(Ile131=)
D	<i>GNAS</i>	SNP	synonymous	GRCh37/hg19	20	57415812	1555	49	1	c.651T>A	p.(Arg217=)
D	<i>HBA2</i>	SNP	3'UTR	GRCh37/hg19	16	223706	1411	99,9	3	c.*107A>G	
D	<i>HBB</i>	SNP	intronic	GRCh37/hg19	11	5247726	1678	50,7	2	c.315+81C>T	
D	<i>HBB</i>	SNP	intronic	GRCh37/hg19	11	5247791	1728	47,3	2	c.315+16G>C	
D	<i>HBB</i>	SNP	intronic	GRCh37/hg19	11	5247733	1658	49,2	2	c.315+74T>G	
D	<i>HBB</i>	SNP	intronic	GRCh37/hg19	11	5247141	1596	47,8	3	c.316-185C>T	
D	<i>HBB</i>	SNP	synonymous	GRCh37/hg19	11	5248243	1961	47,3	1	c.9T>C	p.(His3=)
D	<i>HFE</i>	SNP	splice_donor_+4	GRCh37/hg19	6	26091336	1025	99,4	2	c.340+4T>C	
D	<i>LIG4</i>	SNP	synonymous	GRCh37/hg19	13	1,09E+08	1613	46,7	2	c.1704T>C	p.(Asp568=)
D	<i>NAF1</i>	SNP	missense	GRCh37/hg19	4	1,64E+08	1029	100	8	c.1102C>G	p.(Leu368Val)
D	<i>NAF1</i>	SNP	missense	GRCh37/hg19	4	1,64E+08	1080	99,9	2	c.484A>G	p.(Ile162Val)
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90995019	773	52,5	2	c.102G>A	p.(Leu34=)
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90967711	949	100	11	c.951T>C	p.(Asp317=)
D	<i>NBN</i>	SNP	synonymous	GRCh37/hg19	8	90958422	809	52,5	14	c.1770A>G	p.(Pro590=)
D	<i>NBN</i>	SNP	missense	GRCh37/hg19	8	90990479	990	51	6	c.307G>C	p.(Glu103Gln)
D	<i>NSD1</i>	SNP	synonymous	GRCh37/hg19	5	1,77E+08	1916	100	23	c.6829T>C	p.(Leu2277=)
D	<i>PIF1</i>	SNP	synonymous	GRCh37/hg19	15	65116328	1028	99,9	2	c.207T>C	p.(Phe69=)
D	<i>RECQL4</i>	SNP	missense	GRCh37/hg19	8	1,46E+08	1126	46,8	9	c.1565G>A	p.(Arg522His)
D	<i>RECQL4</i>	SNP	synonymous	GRCh37/hg19	8	1,46E+08	1178	50,3	21	c.3480G>A	p.(Val1160=)
D	<i>RECQL4</i>	SNP	missense	GRCh37/hg19	8	1,46E+08	1122	50,8	5	c.801G>C	p.(Glu267Asp)

D	<i>RECQL4</i>	SNP	synonymous	GRCh37/hg19	8	1,46E+08	822	46,3	3	c.132A>G	p.(Glu44=)
D	<i>RECQL4</i>	INDEL	intergenic	GRCh37/hg19	8	1,46E+08	870	99,9	15	c.2296+1del	
D	<i>RECQL4</i>	SNP	synonymous	GRCh37/hg19	8	1,46E+08	1083	100	19	c.3127T>C	p.(Leu1043=)
D	<i>RECQL4</i>	SNP	missense	GRCh37/hg19	8	1,46E+08	1001	100	4	c.274T>C	p.(Ser92Pro)
D	<i>RIF1</i>	SNP	missense	GRCh37/hg19	2	1,52E+08	1341	52	22	c.2506G>A	p.(Gly836Ser)
D	<i>RIF1</i>	SNP	missense	GRCh37/hg19	2	1,52E+08	1568	48,7	30	c.6061A>T	p.(Asn2021Tyr)
D	<i>RIF1</i>	SNP	missense	GRCh37/hg19	2	1,52E+08	1558	50,4	30	c.4084G>A	p.(Val1362Met)
D	<i>RIF1</i>	SNP	missense	GRCh37/hg19	2	1,52E+08	1113	51,3	35	c.7174C>G	p.(Leu2392Val)
D	<i>RTEL1</i>	SNP	missense	GRCh37/hg19	20	62326110	998	99,8	32	c.3126A>C	p.(Gln1042His)
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/hg19	20	62321655	1448	99,9	26	c.2274G>A	p.(Ala758=)
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/hg19	20	62320968	1111	99,6	23	c.1992T>C	p.(Asp664=)
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/hg19	20	62322288	1062	99,6	27	c.2544T>C	p.(Pro848=)
D	<i>RTEL1</i>	SNP	synonymous	GRCh37/hg19	20	62319085	1253	46,6	17	c.1443G>A	p.(Thr481=)
D	<i>SAMD9</i>	SNP	synonymous	GRCh37/hg19	7	92734451	1357	51,8	2	c.960T>C	p.(Tyr320=)
D	<i>SAMD9</i>	SNP	missense	GRCh37/hg19	7	92734983	1590	50,3	2	c.428T>C	p.(Ile143Thr)
D	<i>SAMD9</i>	SNP	missense	GRCh37/hg19	7	92734065	1475	46,2	2	c.1346A>G	p.(Asn449Ser)
D	<i>SAMD9L</i>	SNP	synonymous	GRCh37/hg19	7	92764982	1539	45,7	5	c.303G>A	p.(Gln101=)
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh37/hg19	12	51374250	1249	100	17	c.*1259T>C	
D	<i>SLC11A2</i>	SNP	intronic	GRCh37/hg19	12	51381077	1678	99,9	16	c.1629+1088G>A	
D	<i>SLC11A2</i>	SNP	intronic	GRCh37/hg19	12	51380232	1578	99,6	16	c.1629+1933T>C	
D	<i>SLC11A2</i>	SNP	intronic	GRCh37/hg19	12	51419621	1322	99,8	1	c.-39+472A>G	
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh37/hg19	12	51373742	1050	99,6	17	c.*1767G>A	
D	<i>SLC11A2</i>	SNP	missense	GRCh37/hg19	12	51402905	829	99,4	1	c.17A>C	p.(Tyr6Ser)
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh37/hg19	12	51375066	1814	99,9	17	c.*443C>T	
D	<i>SLC11A2</i>	SNP	intronic	GRCh37/hg19	12	51420164	1202	99,7	1	c.49+1796G>T	
D	<i>SLC11A2</i>	SNP	3'UTR	GRCh37/hg19	12	51374748	1666	100	17	c.*761C>A	
D	<i>SLC40A1</i>	SNP	synonymous	GRCh37/hg19	2	1,9E+08	1174	44,5	6	c.663T>C	p.(Val221=)
D	<i>SLC40A1</i>	INDEL	5'UTR	GRCh37/hg19	2	1,9E+08	1200	43,4	1	c.-312_-310dup	
D	<i>STN1</i>	SNP	synonymous	GRCh37/hg19	10	1,06E+08	976	48,5	7	c.696G>C	p.(Val232=)
D	<i>STN1</i>	SNP	missense	GRCh37/hg19	10	1,06E+08	923	50,7	7	c.743C>G	p.(Ser248Cys)
D	<i>STN1</i>	SNP	missense	GRCh37/hg19	10	1,06E+08	972	45,9	5	c.451A>G	p.(Thr151Ala)
D	<i>TERF1</i>	SNP	synonymous	GRCh37/hg19	8	73958310	916	47,2	9	c.1198T>C	p.(Leu400=)

D	<i>TERF1</i>	INDEL	inframe_3	GRCh37/hg19	8	73921283	860	42,1	1	c.186_188dup	p.(Glu62dup)
D	<i>TERF1</i>	SNP	synonymous	GRCh37/hg19	8	73958240	914	48	9	c.1128G>A	p.(Arg376=)
D	<i>TERT</i>	SNP	synonymous	GRCh37/hg19	5	1255520	1254	50,2	13	c.2850C>T	p.(His950=)
D	<i>TERT</i>	SNP	synonymous	GRCh37/hg19	5	1294086	1484	50,9	2	c.915G>A	p.(Ala305=)
D	<i>TMPRS S6</i>	SNP	missense	GRCh37/hg19	22	37485724	1187	47,3	7	c.730A>G	p.(Lys244Glu)
D	<i>TMPRS S6</i>	SNP	synonymous	GRCh37/hg19	22	37462926	1324	51,1	18	c.2256C>T	p.(Tyr752=)
D	<i>TMPRS S6</i>	SNP	synonymous	GRCh37/hg19	22	37469591	664	44	13	c.1536C>T	p.(Asp512=)
D	<i>TMPRS S6</i>	SNP	missense	GRCh37/hg19	22	37462936	1359	52	18	c.2246T>C	p.(Val749Ala)
D	<i>TMPRS S6</i>	INDEL	splice_acceptor_-1	GRCh37/hg19	22	37465385	669	55,3	16	c.1842-6_1842-2del	p.(?)
D	<i>TMPRS S6</i>	SNP	5'UTR	GRCh37/hg19	22	37505552	1264	100	1	c.-11G>A	
D	<i>TMPRS S6</i>	SNP	synonymous	GRCh37/hg19	22	37480797	546	43,6	9	c.1056G>A	p.(Ser352=)
D	<i>TMPRS S6</i>	SNP	synonymous	GRCh37/hg19	22	37499386	1317	43	2	c.72G>A	p.(Pro24=)
D	<i>TPP1</i>	SNP	synonymous	GRCh37/hg19	11	6638599	946	48,3	5	c.441A>G	p.(Glu147=)

1. CNV DETECTION



1.5.12 Plot for sample UPN69



[Back to top.](#)

1. CNV DETECTION

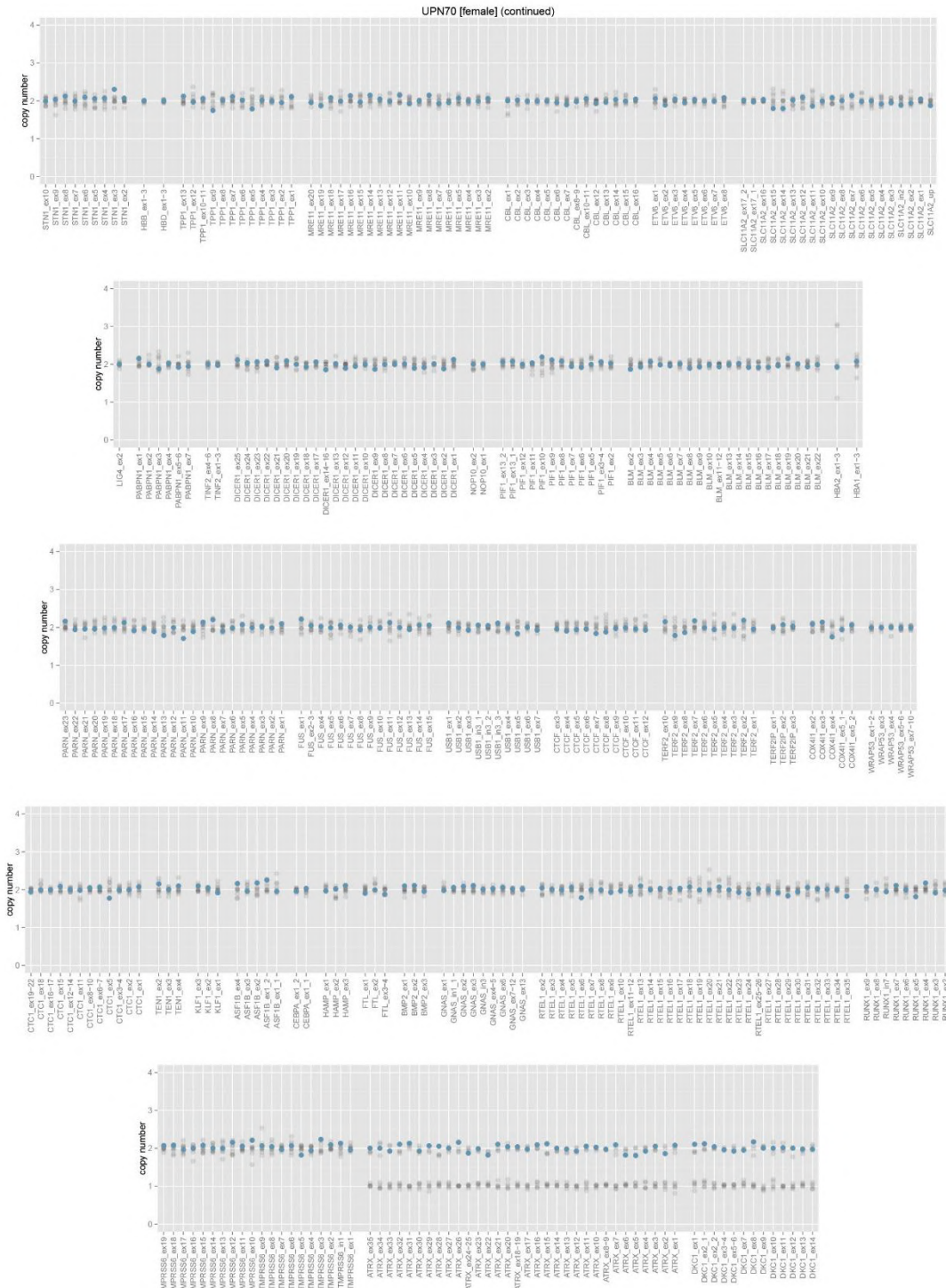


1.5.13 Plot for sample UPN70



Back to top.

1. CNV DETECTION



Back to top.