



UNIVERSITÀ DEGLI STUDI DI PADOVA

Scuola di Medicina e Chirurgia - Dipartimento di Medicina - DIMED

Corso di Laurea in

Tecniche di Laboratorio Biomedico

Presidente: Chiarissimo Prof. Matteo Fassan

TESI DI LAUREA

TIPIZZAZIONE HLA IN *NEXT GENERATION SEQUENCING* PER LA  
VALUTAZIONE DELLA COMPATIBILITÀ DONATORE-  
RICEVENTE NEL TRAPIANTO ALLOGENICO DI CELLULE  
STAMINALI EMOPOIETICHE

Relatore: Prof.ssa De Silvestro Giustina

Correlatore: Dr.ssa Balducci Elena

Laureanda: Falcon Alison

Matricola 1224347

---

Anno Accademico 2021 – 2022



## **ABSTRACT**

Il sistema HLA (Human Leukocyte Antigens) è un complesso di antigeni del sistema immunitario umano che gioca un ruolo fondamentale nei trapianti allogenici di cellule staminali emopoietiche (CSE); infatti, questi sono possibili solamente fra due individui che siano HLA-identici, o al più, se familiari, HLA-aploidentici. Il compito dei laboratori di immunogenetica è identificare e confermare la compatibilità fra donatore-ricevente con tecniche analitiche ad alta risoluzione, limitando così l'insorgenza di eventi avversi post-trapianto; il Laboratorio HLA dell'Ospedale dell'Angelo di Mestre ha introdotto l'NGS con tecnologia Illumina con questo scopo. L'NGS è una tecnica di sequenziamento che permette di processare un elevato numero di campioni in un'unica seduta offrendo risultati con massima risoluzione (allelica) e permettendo l'identificazione di nuovi alleli.

Lo scopo di questo elaborato è dimostrare come l'introduzione dell'NGS migliori il flusso di lavoro nel laboratorio, riducendo tempi di refertazione e costi, incrementando simultaneamente la qualità del risultato necessario alla cura del paziente.

Per raggiungere tale obiettivo, sarà illustrata la metodica di tipizzazione HLA in NGS dell'AULSS3 Serenissima, dimostrando come il sequenziamento di seconda generazione sia stato integrato nella routine laboratoristica, potenziandone le performance. Successivamente ad una breve descrizione dello strumento Illumina MiSeqDx, verranno descritti i software e le metriche di qualità adottate, oltre che gli algoritmi, che consentono una migliore caratterizzazione genetica dei geni HLA in maniera inequivocabile e senza ambiguità.

A dimostrazione di ciò, verranno illustrati due esempi di studio familiare per la compatibilità HLA per il trapianto di CSE, completi di test di conferma donatore-ricevente in NGS.

In conclusione, sarà deducibile l'enorme potenzialità dell'NGS applicata alla valutazione della compatibilità donatore-ricevente nell'ambito del trapianto allogenico di CSE.



# INDICE

<b>1. INTRODUZIONE</b> .....	<b>1</b>
<b>2. IMMUNOGENETICA E TRAPIANTI DI CELLULE STAMINALI EMOPOIETICHE</b> ....	<b>3</b>
2.1 IL SISTEMA HLA.....	3
2.1.1 HLA classe I.....	3
2.1.2 HLA classe II.....	4
2.1.3 HLA classe III.....	5
2.1.4 Polimorfismo, poligenia e codominanza .....	5
2.1.5 Nomenclatura dei geni HLA.....	6
2.2 TRAPIANTO DI CELLULE STAMINALI EMOPOIETICHE.....	8
2.2.1 Tipologie di trapianto allogenico di CSE .....	9
2.2.2 Protocollo dell'AULSS3 Serenissima per lo studio familiare.....	10
<b>3. NEXT GENERATION SEQUENCING</b> .....	<b>13</b>
3.1 TIPIZZAZIONE DEL SISTEMA HLA IN NGS.....	14
3.1.1 Workflow in NGS.....	15
<b>4. SCOPO DELLA TESI</b> .....	<b>17</b>
<b>5. MATERIALI E METODI</b> .....	<b>19</b>
5.1 GENERAZIONE DEL TARGET: ACCETTAZIONE ED ESTRAZIONE DEL DNA .....	19
5.1.1 Long PCR .....	21
5.2 GENERAZIONE DELLA LIBRERIA: FRAMMENTAZIONE E LIGAZIONE AGLI INDICI.....	24
5.2.1 Size selection della libreria.....	25
5.3 SEQUENZIAMENTO: AMPLIFICAZIONE PER CLUSTERS .....	26
5.3.1 Quantificazione con Qubit e caricamento della libreria .....	26
5.4 MiSeqDx™ Illumina® .....	28
5.4.1 Software MiSeqDx .....	31
5.5 ANALISI DEI DATI DI SEQUENZIAMENTO .....	31
5.5.1 Metriche di qualità in NGS.....	31

5.5.2 Algoritmi MIA FORA .....	33
<b>6. RISULTATI.....</b>	<b>39</b>
6.1 STUDIO FAMILIARE CON TEST DI CONFERMA IN NGS A 11 LOCI.....	39
6.2 STUDIO FAMILIARE CON TEST DI CONFERMA IN NGS A 6 LOCI.....	42
<b>7. DISCUSSIONE.....</b>	<b>45</b>
<b>BIBLIOGRAFIA E SITOGRAFIA.....</b>	<b>49</b>
<b>ICONOGRAFIA .....</b>	<b>53</b>
<b>ALLEGATO: APPROFONDIMENTO SUI GENI HLA DI I E II CLASSE.....</b>	<b>55</b>

## **ABBREVIAZIONI**

HLA	Human Leucocyte Antigen
MHC	Major Histocompatibility Complex
CSE	Cellule Staminali Emopoietiche
HSCT	Hematopoietic Stem Cell Transplant
ADR	Antigen Recognition Domain
IMGT/HLA	ImMunoGeneTics HLA database
PCR	Polymerase Chain Reaction
SSO	Sequence-Specific Oligonucleotids
SSP	Sequence-Specific Primers
SBT	Sequence Based Typing
NGS	Next Generation Sequencing
IBMDR	Italin Bone Marrow Donor Registry
TAT	Turn Around Time
S.I.T.	Settore Interdipartimentale Trasfusionale
PBSC	Peripheral Blood Stem Cell
GvHD	Graft versus Host Disease
GVL	Graft Versus Leukemia
EFI	European Federation for Immunogenetics
MUD	Matched Unrelated Donor





# 1. INTRODUZIONE

Il sistema HLA (Human Leukocyte Antigens) è un componente fondamentale del sistema immunitario umano e non, in quanto forma una delle prime linee di difesa attivate dal contatto fra l'organismo e un antigene estraneo; solo grazie a questo complesso, riusciamo ad attivare l'immunità acquisita, sia cellulo-mediata che umorale, riuscendo così a sopravvivere all'attacco da parte di molecole non-self. Ci sono delle situazioni però, in cui è necessario che, nell'organismo ospite, non si sviluppi un'immunoreattività, cioè che non vi sia il riconoscimento di un corpo estraneo come tale. È, ad esempio, il caso dei trapianti allogenici di cellule staminali emopoietiche (HSCT), procedura che prevede l'infusione in un paziente malato di cellule staminali emopoietiche (CSE) da donatore, dunque non self. È possibile un solo modo perché non si sviluppi un'alloreattività da trapianto: trovare un donatore HLA identico al paziente ricevente, cioè testare la compatibilità fra il sistema HLA del donatore e quello del ricevente.

Il problema di maggior rilevanza è che i geni MHC (Major Histocompatibility Complex) codificanti per questo sistema occupano solamente lo 0.5% del genoma umano [1], ma presentano una percentuale di variabilità nucleotidica altissima, tanto che il database ufficiale del sistema HLA IMGT/HLA conta 35 077 alleli con nomenclatura associata [2]. Tutto ciò rappresenta da anni una delle sfide tecnologiche più grandi nell'ambito della biologia, delle biotecnologie e della bioinformatica con il solo obiettivo di conoscere questo complesso sistema di antigeni in modo sempre più approfondito. Dallo studio degli antigeni HLA inizialmente basato sulla sierologia, con una specificità di analisi limitata al solo fenotipo, si è passati a tecniche molecolari come SSO (Sequence-Specific Oligonucleotide), SSP (Sequence-Specific Primers) e Real Time PCR, validate e certificate nei laboratori italiani per la determinazione degli aplotipi HLA in pazienti e donatori, fermandosi però ad un livello di risoluzione basso, in quanto risulta non essere sufficiente per evitare reazioni avverse post-trapianto [3]. Anche se un passo avanti è stato fatto con l'impiego della metodica di sequenziamento SBT (Sequence Based Typing), che ha rappresentato il gold standard per la definizione della compatibilità, tuttavia permane l'inadeguatezza del metodo per la selezione di donatori e pazienti a fini di trapianto; l'analisi con sequenziamento Sanger rimane infatti limitata ai soli esoni informativi (esoni 2 e 3 dei geni HLA di classe I ed esone 2 dei geni di classe II), risultando così un grosso limite, marcato ulteriormente dall'impossibilità di riconoscimento di nuovi alleli o della fase cromosomica (cis/trans) degli alleli HLA.

Con l'avvento del Next Generation Sequencing (NGS), si sono potuti risolvere gli errori e le ambiguità dei tradizionali metodi di genotipizzazione, grazie a due innovazioni tecnologiche quali l'amplificazione clonale e il sequenziamento massiccio in parallelo. L'NGS ha permesso così l'esecuzione di analisi complete dei geni HLA, di individuare alleli HLA non ancora ufficialmente riconosciuti, di stabilire la fase dei polimorfismi e, non da ultimo, di ottenere numeri molto alti di

sequenze in una singola corsa. Il sequenziamento di seconda generazione, infatti, si spinge, come vedremo in questo elaborato, fino alla risoluzione allelica a tre campi, talvolta arrivando al quarto campo, tipizzando gli esoni informativi o ARD (Antigen Recognition Domain), esoni al di fuori dell'ADR, regioni non tradotte e introni.

In questo elaborato, sono messi in risalto i principali vantaggi della metodica NGS nella sua tipizzazione, cercando di mettere in risalto i suoi vantaggi, quali la maggior velocità di produzione dei risultati, le ridotte tempistiche del TAT (Turn Around Time) e la minor presenza di ambiguità nei risultati, grazie all'alta risoluzione d'analisi [4]. Il tutto è dedotto dall'esperienza di tirocinio svolto presso il Laboratorio HLA del S.I.T. dell'Ospedale dell'Angelo di Mestre (AULSS3 Serenissima), durante il quale è stato possibile conoscere al meglio questa metodica, applicandola alla tipizzazione dei donatori di CSE iscritti al registro IBMDR (Italian Bone Marrow Donor Registry) e ai test di conferma donatore-ricevente per i pazienti candidati al trapianto allogenico di cellule staminali emopoietiche.

## 2. IMMUNOGENETICA E TRAPIANTI DI CELLULE STAMINALI EMOPOIETICHE

### 2.1 IL SISTEMA HLA

La base della sopravvivenza delle specie viventi è l'adattamento all'ambiente, la convivenza con esso e con le altre specie viventi, cioè la "sopravvivenza del più adatto" all'ecosistema. Meccanismo chiave dell'evoluzione e della selezione naturale darwiniana è la diversità genetica delle popolazioni che permette la diffusione dei geni cosiddetti "più forti" per la specie, con dunque una maggior probabilità di sopravvivenza. Allo stesso tempo però è necessario mantenere, almeno presso alcuni geni, un'elevata variabilità allelica, in quanto, al diffondersi fra la popolazione, permette una maggiore sopravvivenza globale e una maggior prontezza nel conformarsi con l'ecosistema. Uno dei più evidenti esempi di questi geni necessariamente ad elevato polimorfismo sono i geni appartenenti al sistema HLA, localizzato nel braccio corto del cromosoma 6 (6p 21.1-21.3, circa a 6000 Kb dal centromero), con un'estensione molto contenuta, ma dall'importante quantità di informazioni genetiche ad alta variabilità fra gli individui. Gli antigeni di questo complesso sono delle glicoproteine di membrana che svolgono un ruolo chiave nell'attivazione del sistema immunitario dell'individuo a diversi livelli, infatti possiamo dire che i prodotti dei geni di HLA sono classificabili in tre classi a seconda della loro distribuzione cellulare, ma anche delle loro caratteristiche morfologico-strutturali e funzionali [5] (Fig.1).

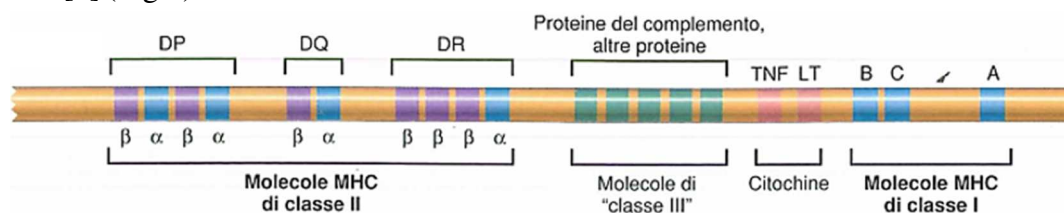


Figura 1. Localizzazione dei geni codificanti per gli antigeni del sistema HLA.

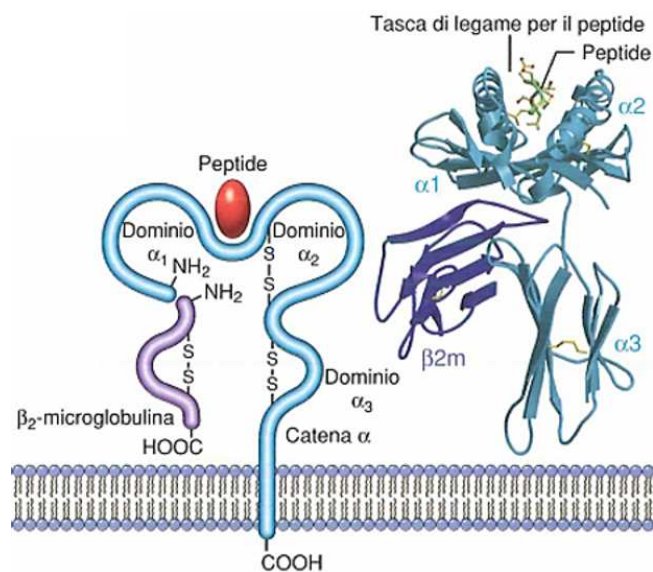
#### 2.1.1 HLA CLASSE I

Le molecole di classe I (HLA-I) sono espresse sulla membrana di tutte le cellule nucleate e delle piastrine, ma sono anche presenti allo stato solubile nel sangue. Sono codificate da tre loci strettamente correlati fra loro, HLA-A, HLA-B e HLA-C, che concorrono nel formare l'eterodimero sulla superficie cellulare. Questo è composto da una catena pesante, chiamata catena  $\alpha$ , con peso molecolare di circa 45 kDa, e da una catena leggera, o catena  $\beta$ , non polimorfica (12 kDa), conosciuta con il nome di  $\beta_2$ -microglobulina, il cui gene codificante è situato sul cromosoma 15 (15q21.1) (Fig.2); le due catene sono unite da legami non covalenti e la catena  $\beta$  non si inserisce all'interno della membrana, ma vi rimane slegata. È importante sottolineare anche che la molecola di HLA-I è suddivisibile in tre regioni a seconda del suo rapporto con la cellula:

- Regione intracitoplasmatica di 32 amminoacidi, porzione necessaria alla trasduzione del segnale esterno all'interno della cellula e codificata dagli esoni 6-7 del gene HLA-A
- Regione transmembrana idrofobica di 32 amminoacidi, codificata dall'esone 5 di HLA-A
- Regione extracellulare di 247 amminoacidi, essenziale per l'interazione con l'ambiente.

La regione extracellulare rappresenta quella di maggior interesse clinico e diagnostico, in quanto è determinante per la risposta immunitaria e per la compatibilità donatore-ricevente.

Essa è infatti costituita da tre domini di uguale lunghezza e con conformazione ad  $\alpha$ -elica;  $\alpha 1$  e  $\alpha 2$ , codificati dagli esoni informativi 2 e 3 rispettivamente, costituiscono la tasca di legame per gli antigeni da presentare ai linfociti T CD8+ dopo loro degradazione all'interno della cellula. Le facce laterali e il fondo del sito di legame sono delle regioni altamente polimorfiche, in modo da permettere il legame con la maggior quantità di antigeni possibili. Il terzo dominio è  $\alpha 3$ , codificato dall'esone 4, e caratterizzato da omologia con la regione costante delle immunoglobuline; è un dominio non polimorfico, necessario per il legame con CD8, co-recettore dei linfociti T citotossici, che determinano la risposta immunitaria cellulo-mediata.



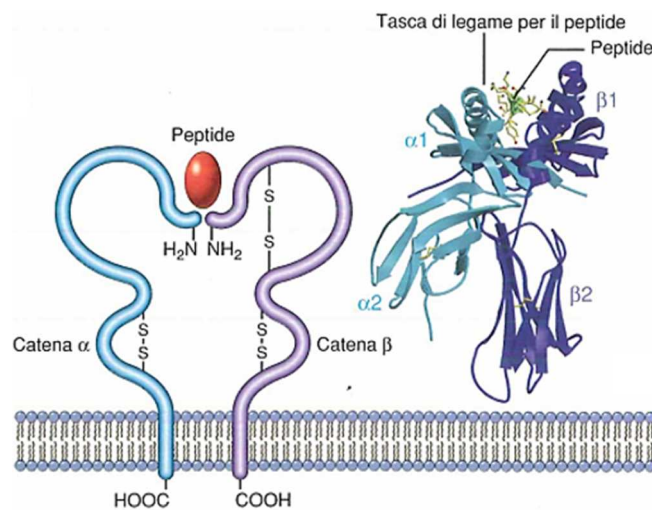
**Figura 2.** Animazione e struttura cristallografica della molecola MHC I.

### 2.1.2 HLA CLASSE II

Le molecole di classe II sono codificate dal gene HLA-D, composto da tre locus quali HLA-DR, HLA-DQ e HLA-DP che, come nel caso della classe I, si traducono in un eterodimero di superficie; questo differisce dal precedente per due caratteri principali: per prima cosa, non è ubiquitario sulla superficie cellulare, ma è situato solamente nelle cellule APC (Antigen Presenting Cells), cioè macrofagi, cellule dendritiche, cellule di Langerhans, alcune cellule dell'endotelio, linfociti B e linfociti T (quando attivati). Si riscontra una seconda differenza nella composizione stessa dell'eterodimero, generato anche in questo caso da due catene,  $\alpha$  e  $\beta$ , entrambe codificate dalla sequenza genica di questi loci: in particolar modo, la catena  $\alpha$  di 34 kDa è codificata dal gene A di

HLA-DR (HLA-DRA1) (solo in questo caso non è presente polimorfismo), HLA-DQ (HLA-DQA1) e HLA-DP (HLA-DPA1), mentre la catena  $\beta$  di 29 kDa è generata dal gene B di HLA-DQ (HLA-DQB1) e HLA-DP (HLA-DPB1) e da quattro geni B altamente polimorfici di HLA-DR, cioè DRB1, DRB3, DRB4 e DRB5 (Fig. 3).

Entrambe le catene sono ancorate all'interno della cellula, per cui sono costituite da una regione intracellulare, transmembrana ed extracellulare, la quale è composta da due domini per ciascuna catena:  $\alpha 1$ ,  $\alpha 2$ ,  $\beta 1$  e  $\beta 2$ . L'analisi cristallografica dimostra che i domini  $\alpha 1$  e  $\beta 1$ , codificati dagli esoni 2 di questi loci genici, formano la tasca di legame per i peptidi internalizzati all'interno della cellula tramite vescicole, i quali saranno esposti sulla membrana cellulare legati a MHC-II. Il dominio  $\beta 2$  invece è responsabile dell'interazione con il co-recettore CD4, caratterizzante la superficie dei linfociti T helper, responsabili del richiamo e dell'attivazione dei linfociti B e dunque della risposta immunitaria umorale.



*Figura 3.* Animazione e struttura cristallografica della molecola MHC II.

### 2.1.3 HLA CLASSE III

Le molecole di classe III sono infine molecole di varia natura, come componenti del complemento, in particolare le frazioni C2 e C4, citochine quali TNF $\alpha$  (fattore di necrosi tumorale) e linfo tossina, proteine implicate nel riconoscimento dell'antigene (trasportatori di peptidi, proteina DM e componenti del proteasoma), ma anche molecole che non fanno parte del sistema immunitario [6].

### 2.1.4 POLIMORFISMO, POLIGENIA E CODOMINANZA

Oltre a delineare la struttura e le componenti di questo articolato sistema di molecole è necessario descriverne le peculiarità, cioè i tratti caratteristici che lo rendono così interessante, ma soprattutto così importante per la sopravvivenza delle varie specie esistenti.

La prima caratteristica di questo complesso è il polimorfismo, cioè la presenza di un elevato numero di alleli che concorrono alla formazione delle proteine transmembrana, permettendo la presentazione dei frammenti peptidici derivati dalle proteine, soprattutto quelle riconosciute come antigeni non self,

potenzialmente pericolose per l'organismo; così facendo, rendono possibile l'attivazione del sistema immunitario, la nostra principale fonte di difesa dai patogeni e dagli agenti esterni.

Seconda caratteristica importante è la poligenia: a codificare per la proteina transmembrana, infatti, non è solamente un gene, ma sono tre geni che lavorano in sintonia, dunque i geni A, B e C per la prima classe e i geni DR, DQ e DP per la seconda classe. Ad essere più precisi, si è osservato che il locus comprende 253 geni, di cui il 57% sono attivi nella codifica delle proteine delle due classi, conferendo alta specificità per i diversi peptidi; a questa è da sommarsi anche la variabilità conferita dai vari alleli dei diversi geni, che differiscono l'uno dall'altro per circa il 5-10% della sequenza.

Ultima ed estremamente importante peculiarità del sistema HLA è la codominanza; abbiamo visto che il polimorfismo è essenziale per la nostra esistenza perché fa sì che all'interno di una popolazione ci sia sempre un buon numero di individui che è in grado di riconoscere determinati peptidi antigenici, garantendo così la sopravvivenza e la continuità della specie. Il maggior spettro di difesa dell'organismo è reso possibile anche dalla codominanza, cioè l'evento per cui sia l'allele ereditato dalla madre che quello ereditato dal padre vengano attivamente espressi sulla superficie delle cellule, aumentando così la resistenza ai patogeni esterni. La combinazione di questi alleli viene indicata come aplotipo HLA, che, per definizione, indica una specifica combinazione allelica fra diversi loci, posizionati su un unico cromosoma: si intende dunque la combinazione dei geni HLA-A, B, C, DR, DQ e DP presenti nei cromosomi 6 omologhi, cioè quello di origine paterna e quello di origine materna, che si coesprimono; se l'aplotipo del donatore non è identico o, al più, aploidentico, all'aplotipo del ricevente, non si può procedere al trapianto.

Stando a quanto si è detto fino a questo punto, il sistema HLA risulta essere il maggior vantaggio evolutivo che una specie possa possedere. Negli anni però è stato messo in risalto anche un suo limite: un così elevato polimorfismo indica una minor probabilità di trovare degli individui genotipicamente identici o aploidentici, cioè che condividono la metà degli alleli con un altro soggetto [7].

### **2.1.5 NOMENCLATURA DEI GENI HLA**

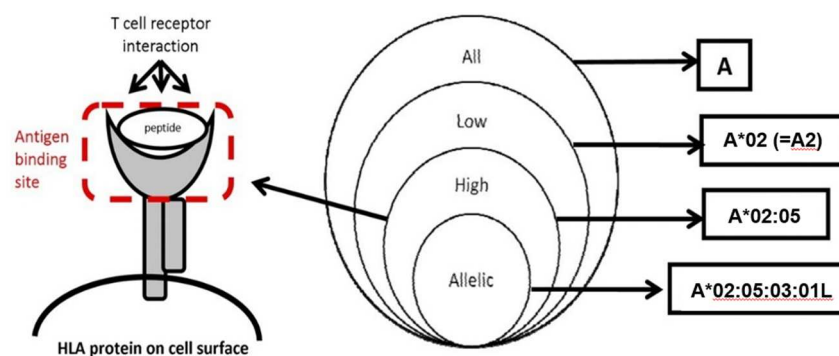
È ora necessario andare a delineare in modo chiaro come funziona la nomenclatura di questo sistema, cioè come è possibile andare a identificare in modo inequivocabile un allele (Fig. 4). Partiamo da un esempio:

HLA - A \* 02 : 05 : 03 : 01 L

La dicitura HLA identifica il complesso di geni, mentre la lettera A che lo affianca separata da un trattino indica il gene a cui si fa riferimento, che in questo caso è il gene del locus A, appartenente alla prima classe del sistema. Il simbolo \* è un separatore che fa capire che a seguire si incontreranno una serie di numeri che identificano l'allele preciso, in base a quelli che sono i polimorfismi riscontrati nella sequenza sia esonica che intronica. La posizione occupata dallo "02" viene definita "primo

campo" o "2 digits" e indica la famiglia allelica a cui si fa riferimento nel gene, della quale poi è necessario andare a definire l'allele preciso, dato dallo "05", chiamato "secondo campo" o "4 digits", che indica infatti la forma allelica che l'individuo presenta, cioè la proteina specifica che egli andrà ad esprimere sulle sue cellule. A livello genomico sono importanti anche le mutazioni silenti, cioè quelle variazioni di sequenza nucleotidica che non portano a cambiamenti a livello proteico grazie alla ridondanza del codice amminoacidico; queste sono indicate con lo "03", ovvero il "terzo campo" o "6 digits". Il "quarto campo" o "8 digits" è invece dato dallo "01" e serve ad identificare quelle variazioni della sequenza nucleotidica a livello degli introni, quindi dei cambiamenti nell'area non coding che non modificano la qualità della proteina, ma potrebbero influenzarne la quantità, cioè i livelli di espressione, i quali sono riportati dalla lettera finale (L in questo caso), che indica la quantità di produzione e di funzionalità della proteina. I differenti suffissi letterali possono essere:

- L (low): bassa espressione della proteina sulla superficie
- N (Null): sta ad indicare la presenza della sequenza di DNA che non viene tradotta in una proteina, per cui non viene espressa
- A (aberrante): proteina che non viene espressa oppure che presenta un'alterazione della struttura tale da non essere funzionale
- S (soluble): proteina che non si trova a livello della membrana, ma solamente in forma solubile
- C (cytoplasm): il prodotto allelico rimane nel citoplasma
- Q (questionable): la proteina prodotta da tale allele ha degli effetti che non sono ancora totalmente conosciuti o definibili.



**Figura 4.** Corrispondenza fra la nomenclatura HLA e il livello di risoluzione associato ad ogni campo.

Nel momento in cui si svolge la tipizzazione dei geni HLA di un individuo è importante chiarire che risoluzione si vuole ottenere, cioè a che campo ci si vuole fermare con l'analisi e, in base a questo, si può scegliere la metodica più adeguata per i propri scopi. Generalmente, per la ricerca di donatori compatibili si parte con una tecnica a bassa risoluzione, fermandosi al primo campo, in modo da focalizzarsi per bene solamente in quei pochi donatori che sono compatibili a campi più generici. Si

prosegue poi con una metodica più robusta, come la metodica SSP o il sequenziamento di Sanger, che identificano il secondo campo, o il sequenziamento NGS, che si spinge fino al terzo campo; obiettivo di quest'ultima tecnica è quello di non limitarsi a sequenziare solamente gli esoni informativi, ma l'intera sequenza di DNA della regione, diminuendo così i casi di mismatch e aumentando le possibilità di successo del trapianto.

## **2.2 TRAPIANTO DI CELLULE STAMINALI EMOPOIETICHE**

Il trapianto di cellule staminali emopoietiche è una delle terapie utilizzabili nella cura di neoplasie oncoematologiche (leucemie, linfomi, mieloma) o di altre malattie ematologiche importanti, quali immunodeficienza primitiva, aplasia midollare ed altre, elenco in progressivo aumento (attualmente sono circa 70 le malattie per cui è possibile l'accesso al trapianto). Lo scopo di questa procedura è andare a sostituire le cellule malate originanti dal midollo osseo con cellule staminali sane, in modo da eradicare la malattia. Si procede dunque in due step: prima di tutto, si eseguono dei trattamenti mieloablativi, quali chemioterapia, radioterapia o terapia bersaglio come farmaci inibitori o anticorpi monoclonali, con lo scopo di eliminare le cellule maligne; in secondo luogo, si sostituisce il midollo malato con uno sano, operando un trapianto autologo nel caso di autodonazione di CSE da parte del paziente stesso, oppure un trapianto allogenico, quando le cellule staminali sono prelevate da donatore familiare o non consanguineo [8]. Contrariamente a quanto accade per il trapianto autologo di CSE, lo scopo del trapianto allogenico nelle patologie neoplastiche non si basa solo sulla capacità della chemioterapia e/o radioterapia di condizionamento di eradicare la malattia, ma anche sull'effetto immunologico del trapianto stesso. Le nuove cellule del donatore, e in particolare linfociti T, hanno la capacità di eliminare, con un meccanismo noto come Graft Versus Leukemia (GVL), le cellule neoplastiche eventualmente ancora presenti nel ricevente nonostante la terapia di condizionamento. Nelle patologie ematologiche in cui le cellule del midollo non lavorano in modo corretto, lo scopo del trapianto è proprio quello di sostituire il midollo non funzionante con uno nuovo. Così facendo, nelle migliori delle ipotesi, le nuove cellule attecchiranno e fungeranno da precursori delle nuove cellule del sangue, ossia eritrociti, leucociti e trombociti.

Tutto questo è possibile poiché le cellule staminali sono cellule indifferenziate pluripotenti (positive al CD34 in citofluorimetria), cioè unità biologiche dotate delle capacità di automantenimento e di differenziamento in cellule figlie con caratteristiche ben definite a seconda dei fattori di crescita prodotti dall'organismo, dunque sulla base delle proprie necessità.

Per quanto riguarda le cellule del sangue, il processo si chiama emopoiesi, suddivisibile in due linee principali: mieloide, da cui originano eritrociti, trombociti, granulociti e macrofagi, e linfoide, che dà vita ai linfociti. Questo evento avviene nel midollo osseo, costituito da fibre reticolari che danno origine a due compartimenti: il compartimento emopoietico, in cui avviene la vera e propria



emopoiesi, e il compartimento vascolare, che contiene i sinusoidi necessari a veicolare le nuove cellule del sangue in circolo [9].

Sono possibili tre fonti per il prelievo di CSE da utilizzare in terapia: il midollo osseo, che presenta una quantità di CSE variabile fra 1-3%, prelevabili dalle creste iliache posteriori; il sangue di cordone ombelicale, che ospita una fonte di cellule staminali di 0.1-1%, per cui viene spesso utilizzata nella terapia infantile. L'ultima fonte è il prelievo di sangue periferico (PBSC), sede preferibile in quanto più facilmente tollerata dal donatore, sia per assenza di ricovero ospedaliero che per il rapido recupero della conta cellulare. In circolo però, è presente una concentrazione di CSE molto bassa (0.01-0.1%) per cui, per ottenere una raccolta di CD34+ efficace, è necessario procedere alla cosiddetta "mobilizzazione": questa prassi prevede due somministrazioni giornaliere peso-dipendenti di G-CSF, fattore di crescita granulocitario, fino alla staminoafèresi, la raccolta che avviene generalmente fra il quarto e quinto giorno di terapia. La procedura di raccolta richiede circa 4-5 ore e utilizza un separatore cellulare in circolazione extracorporea: in un circuito monouso, tramite un sistema di pompe peristaltiche, il sangue del donatore viene separato nei suoi componenti, vengono raccolte le cellule mononucleate, mentre tutti gli altri emocomponenti sono restituiti al donatore (in caso di trapianto allogenico) o al paziente/donatore (in caso di trapianto autologo). La quantità di cellule da raccogliere viene concordata con i sanitari del Centro Trapianti e, normalmente, è pari a  $5 \times 10^6$  cellule CD34+ / kg di peso del paziente. È possibile prevedere anche la quantità di raccolta grazie alla formula [10]

$$\frac{\text{conta CSE pre} \times V \text{ sangue processato (L)} \times \text{efficienza procedura}}{\text{peso del ricevente (kg)}}$$

### 2.2.1 TIPOLOGIE DI TRAPIANTO ALLOGENICO DI CSE

Per un determinato paziente, a discriminare la scelta di un donatore piuttosto che di un altro è la compatibilità del sistema HLA; in base a questa, il trapianto allogenico può essere:

- Fratello HLA-identico: è il donatore ideale, specialmente se gemello omozigote, poichè condivide con il paziente entrambi i loci HLA (compatibilità al 100%), usufruibile però solamente nel 25-30% dei casi;
- Fratello HLA-aploidentico: quando la compatibilità è del 50%, cioè è presente solamente un aplotipo HLA identico, come avviene fra genitori e figli, in quanto i figli ereditano metà sistema HLA dalla madre e metà dal padre; in questo caso, nei fratelli è concorde l'aplotipo ereditato da uno dei due genitori, mentre l'altro differisce;
- Donatore MUD (Matched Unrelated Donor): è il donatore volontario iscritto ai registri IBMDR, registro nazionale, o BMDWW (Bone Marrow Donor World Wide) a livello mondiale.

Per quanto riguarda l'Italia, nella pagina web del Ministero della Salute è possibile risalire ai dati ufficiali riguardanti i trapianti di CSE nell'anno 2021.

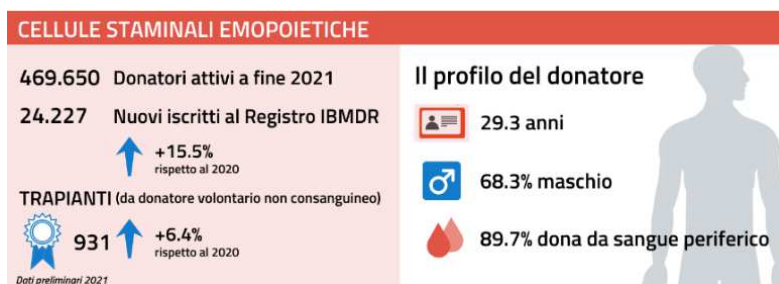


Figura 5. Report dei dati sui trapianti di CSE in Italia secondo il Ministero della Salute

Come mostra la Figura 5, in Italia meno dell'1% della popolazione è donatore attivo di cellule staminali (i dati IBMDR aggiornati al 30 giugno 2022 riporta un numero di donatori pari a 475.295); sembra comunque che, di anno in anno, stia aumentando la quota di italiani che compiono questa scelta, segnale che le campagne di sensibilizzazione su questa tematica stanno dando i loro frutti. Mobilitazioni di questo genere vengono spesso fatte presso le scuole superiori, ma anche fra i donatori di sangue periferico, infatti, dai dati riportati, si evince che quasi il 90% degli iscritti al registro IBMDR sono anche donatori di sangue regolari. Anche il numero di trapianti eseguiti è aumentato rispetto al 2020 [11].

### 2.2.2 PROTOCOLLO DELL'AULSS3 SERENISSIMA PER LO STUDIO FAMILIARE

Finora si è visto come il trapianto di cellule staminali possa essere una terapia valida per affrontare i tumori del sangue. Il primo step di questa strada è la ricerca di un donatore idoneo per il paziente, cioè la ricerca di una persona che, per prima cosa, presenti un sistema HLA compatibile con il ricevente, in modo da evitare eventi avversi di ogni tipo, come la Graft versus Host Disease (GvHD): questa complicanza si verifica quando le cellule immunocompetenti del donatore attivano una risposta immunologica nei confronti di cellule estranee, in questo caso le cellule MHC del ricevente, poiché queste presentano delle discordanze antigeniche che stimolano l'alloreattività dei linfociti T [12]. L'indicazione al trapianto è data dai clinici che hanno in cura il paziente, nel nostro caso dai medici del reparto di Ematologia e Oncoematologia dell'Ospedale dell'Angelo di Mestre, i quali collaborano con il Laboratorio HLA del settore trasfusionale, che si occupa di confermare l'idoneità tra ricevente e donatore. Per svolgere questa indagine il Laboratorio stila un proprio protocollo interno, basato sugli Standard Internazionali e sulle direttive dell'EFI (European Federation for Immunogenetics), il più importante ente europeo nel campo dell'immunogenetica, dell'istocompatibilità e dei trapianti; infatti, il laboratorio che intende eseguire ufficialmente tipizzazioni per trapianti di CSE deve possedere l'accreditamento presso questo ente, che ne certifica la qualità e l'affidabilità dell'attività e dei risultati prodotti e refertati. La certificazione è ottenuta con visita ispettiva da parte di ispettori appartenenti all'ente di accreditamento che hanno l'obbligo di

verificare che il laboratorio rispetti gli standard di qualità richiesti, necessari a garantire un risultato sicuro e ripetibile. L'accreditamento ha una validità di tre anni, al termine dei quali si ripeterà la visita per rinnovare la certificazione [13].

La ricerca di un donatore compatibile può essere attivata solamente dall'ematologo che ha in cura il paziente; egli è in stretto contatto con i medici del laboratorio di HLA dell'ospedale, per cui anticipa loro la necessità di avviare una nuova procedura di allotrapianto. La prima fase prevede l'apertura del cosiddetto "studio familiare" [14]: vengono inviati al laboratorio i prelievi di sangue del paziente e dei suoi familiari stretti, accompagnati da una richiesta dove sono riportati i gradi di parentela. Solitamente sono convocati i fratelli, poiché è maggiore la probabilità di trovare un donatore HLA identico o aploidentico, ma la richiesta viene ampliata anche ai figli o ai genitori se di età compatibile. È utile, in ogni caso, avere anche un campione dei genitori del paziente se disponibili, in quanto, conoscendo la loro tipizzazione, è possibile capire con certezza come è avvenuta la segregazione degli alplotipi all'interno della famiglia. Questo primo step, come indicato dagli standard IBMDR, prevede una tipizzazione a 5 loci in bassa risoluzione, per cui si va a studiare solamente la famiglia allelica di HLA-A, -B, -C, -DR, e -DQ, limitandosi cioè a 2 digits (ad esempio: A\*02, B\*01, C\*07). Questa viene svolta con la tecnica di SSO, un sistema che sfrutta l'ibridazione del DNA amplificato a microbglie eccitabili a lunghezze d'onda specifiche.

In caso si presenti una compatibilità, si procede al secondo step che prevede una tipizzazione in alta risoluzione su un nuovo campione della coppia paziente-donatore, in modo da aumentare i campi della ricerca: si effettua una tipizzazione a risoluzione allelica a 6 o 11 loci, includendo dunque i geni HLA-DQA1, -DPA1, -DPB1, -DRB3/4/5. Presso il laboratorio HLA di Mestre, il test di conferma si svolge in NGS, con la tecnologia Illumina, che permette di avere risultati attendibili fino a 6 digits, cioè al terzo campo (ad esempio: HLA-A\*02:01:02).

È inoltre raccomandato confermare la tipizzazione del locus DPB1 in alta definizione con la tecnica di SSP, per escludere una potenziale ricombinazione genetica frequente in questo gene a causa del basso linkage disequilibrium presente con i loci HLA-DR e -DQ. Clinicamente parlando infatti, è necessario individuare i mismatches permissivi e non permissivi di questo locus, in quanto discriminano una maggiore o minore capacità di scatenare i linfociti T in risposte alloreattive da incompatibilità donatore-ricevente.

Alla fine di questi step, il laboratorio invia le risposte all'ematologo che deciderà se andare a trapianto con il familiare selezionato o aprire una ricerca MUD da registro, cioè se ricercare nel registro IBMDR un donatore compatibile con il paziente. Discriminanti per questa scelta sono altre indagini effettuate sul paziente, come la ricerca di anticorpi anti-HLA, diagnostica ematologica in biologia molecolare e/o la citogenetica.



### 3. NEXT GENERATION SEQUENCING

A partire dalla scoperta del DNA da parte di Watson e Crick nel 1952, sono stati molti i progressi ottenuti in questo campo, proprio perché, essendo il DNA la base della vita, si è vista la necessità di doverlo conoscere sino alla sua forma più intima, con l'obbiettivo di ricavarne informazioni utili a sostegno della salute umana.

Prima di trovare delle metodiche che agissero in modo diretto sul DNA, la sierologia permetteva il riconoscimento dei suoi prodotti, le proteine: con un prelievo di sangue periferico, era possibile accedere ai linfociti e, sfruttando delle reazioni di agglutinazione con antisieri, si determinavano gli alleli del sistema HLA del paziente. Questa era una tecnica molto blanda, con una specificità di analisi limitata al primo campo e al solo fenotipo dell'individuo, non il genotipo in sé.

Un primo passo verso una maggior e facilitata accessibilità al DNA è stata l'invenzione della PCR, la reazione a catena della polimerasi, ideata dal premio Nobel per la chimica Kary Mullis (1944-2019) nel 1983; questa tecnica ha permesso di potenziare un'altra invenzione precedente di qualche anno, quando nel 1977 il premio Nobel Frederick Sanger (1918-2013) riporta un metodo per il sequenziamento del DNA, conosciuto come sequenziamento tramite terminatori di catena o, più semplicemente con metodo Sanger: per la prima volta, è possibile conoscere la sequenza del DNA base per base, accedendo così ad un'immensa quantità in informazioni che, con la giusta interpretazione e applicazione in medicina, contribuiscono in maniera determinante alla salute umana.

Una straordinaria evoluzione in questo campo si è verificata a partire dagli anni 2000, quando iniziano ad essere disponibili sul mercato le piattaforme di sequenziamento di nuova generazione, NGS, che aumentano in modo esponenziale la capacità di sequenziamento, abbattendo allo stesso tempo i costi e i tempi di attesa. Sono state introdotte diverse piattaforme NGS che operano secondo diverse modalità, ma tutte sono caratterizzate dalla creazione di una libreria di DNA target amplificato contenente migliaia di copie dei vari segmenti (reads), i quali permettono una sicura chiamata delle basi azotate presenti e, dunque, della sequenza nucleotidica del gene in esame. Esistono molteplici ambiti laboratoristici di applicazione, fra i quali si trova il laboratorio di Immunogenetica, all'interno del quale l'NGS ha introdotto enormi vantaggi, in quanto rende possibile il sequenziamento massiccio di un numero elevato di campioni in un'unica seduta con risultati certi ed inequivocabili, escludendo le ambiguità tipiche di altre tecniche di tipizzazione HLA. Questo è possibile grazie all'indicizzazione, cioè all'aggancio di "barcode" nucleotidici univoci a ciascun frammento derivante da uno stesso soggetto, in modo da poter ricondurre una determinata sequenza all'individuo di appartenenza. È inoltre possibile analizzare loci diversi in una singola corsa grazie all'amplificazione clonale dei frammenti, mantenendo sempre un alto livello di risoluzione (terzo campo, cioè risoluzione allelica), eliminando equivocità che possono nascere dalla presenza di polimorfismi in regioni introniche e associando mutazioni SNP ai giusti cromosomi (paterno o materno); in questo modo è possibile

ridurre la mortalità e l'insuccesso dei trapianti. Altri vantaggi degni di nota sono il costo di analisi, più contenuto per singolo campione, ma soprattutto la maggior velocità di refertazione, dato essenziale nella cura del paziente. Il risultato si ottiene in pochi giorni di lavoro e senza la necessità di indagini genetiche aggiuntive: in questo modo, il clinico saprà subito se fra i famigliari del paziente è presente una compatibilità oppure se sarà necessario aprire una ricerca in IBMDR, registro nel quale i donatori sono tipizzati in NGS, per cui la ricerca di un match è più immediata [15] [16].

Presso il laboratorio di HLA appartenente al S.I.T. dell'ospedale di Mestre, vengono eseguite tre metodiche di analisi del DNA, la cui combinazione permette di tipizzare correttamente gli alleli presenti nel sistema HLA di pazienti affetti da malattie oncoematologiche e di donatori di cellule staminali da sangue periferico o da midollo osseo. Le tecniche impiegate sono due di tipo molecolare, cioè SSO, applicata alla piattaforma Luminex, e SSP, e una terza di sequenziamento, cioè MIA FORA™ NGS MFlex, studiata per la piattaforma Illumina, supportata dallo strumento MiSeqDx (Fig. 16).

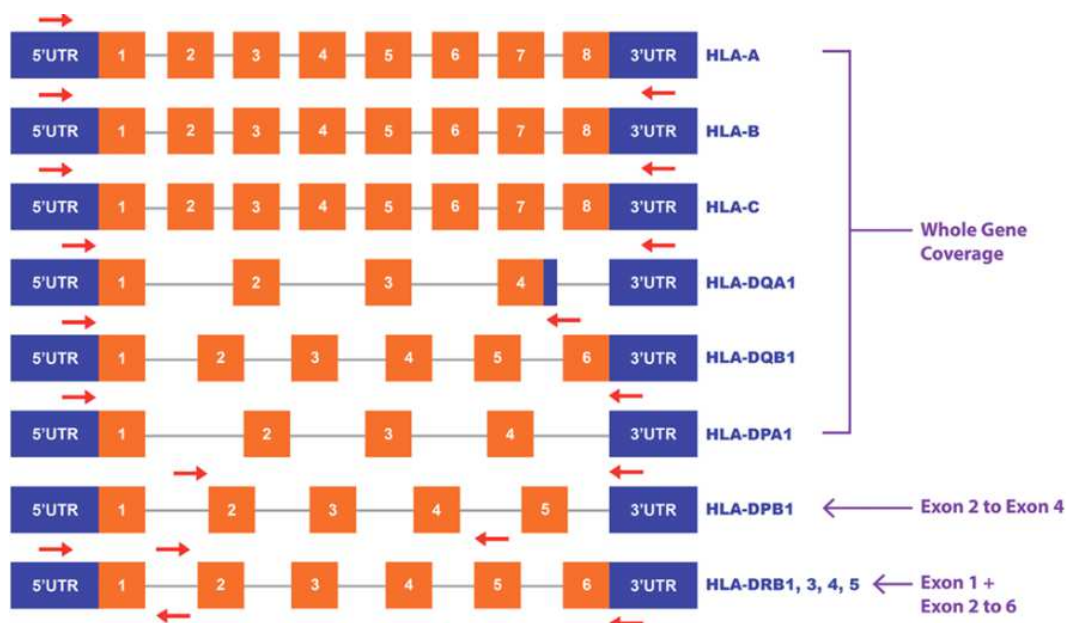
La trattazione si focalizza sulla metodica di sequenziamento NGS, introdotta recentemente nel laboratorio con grandi performance in termini di numero e qualità di tipizzazioni eseguite.

### **3.1 TIPIZZAZIONE DEL SISTEMA HLA IN NGS**

Per poter emettere dei referti, ogni laboratorio, come già riportato, deve possedere una certificazione che attesti la qualità del lavoro svolto. Uno dei requisiti necessari ad ottenere questo riconoscimento è stilare un proprio protocollo che descriva le proprietà di ciascuno strumento e il flusso di lavoro richiesto per arrivare al referto [17]. Anche il laboratorio HLA di Mestre ha redatto proprie procedure di tipizzazione in NGS: sono riportati i materiali, gli strumenti, le loro specifiche, ma anche il workflow, descritto in seguito. A tal proposito, sono necessari reagenti e strumentazione dedicati affinché la seduta di NGS abbia la massima performance.

MIA FORA™ NGS Mflex è un kit di tipizzazione del sistema HLA progettato dalla ditta Immucor, specializzata in medicina trasfusionale e trapianti con sede principale a Norcross, GA (USA). Viene utilizzato all'interno del laboratorio di HLA del S.I.T. di Mestre con l'obiettivo di ottenere una tipizzazione HLA in alta risoluzione di donatori di cellule staminali iscritti all'IBMDR, pazienti candidati a trapianto di CSE e per test di conferma fra paziente ricevente e donatore.

I geni che vengono sequenziati sono quelli appartenenti alla classe I, HLA-A, HLA-B e HLA-C, e alla seconda classe, HLA-DRB, -DQB1, -DPB, -DPA1 e -DQA1. È una tipizzazione a 6 o 11 loci ad alta risoluzione poiché si propone di analizzare i loci selezionati nella loro interezza, cioè sia esoni che introni, con l'eccezione dei geni HLA-DPB1 e HLA-DRB1/3/4/5, di cui non si ha una sequenza completa, ma che comunque copre la maggior parte di esoni ed introni, non limitandosi dunque ai soli esoni informativi (esoni 2 e 3) (Fig.6) [18].



**Figura 6.** Illustrazione grafica dei geni HLA sequenziabili dal kit MIA FORA, con i rispettivi esoni (riquadri arancioni). le frecce rosse indicano il punto di attacco e stacco della Taq polimerasi durante la Long PCR.

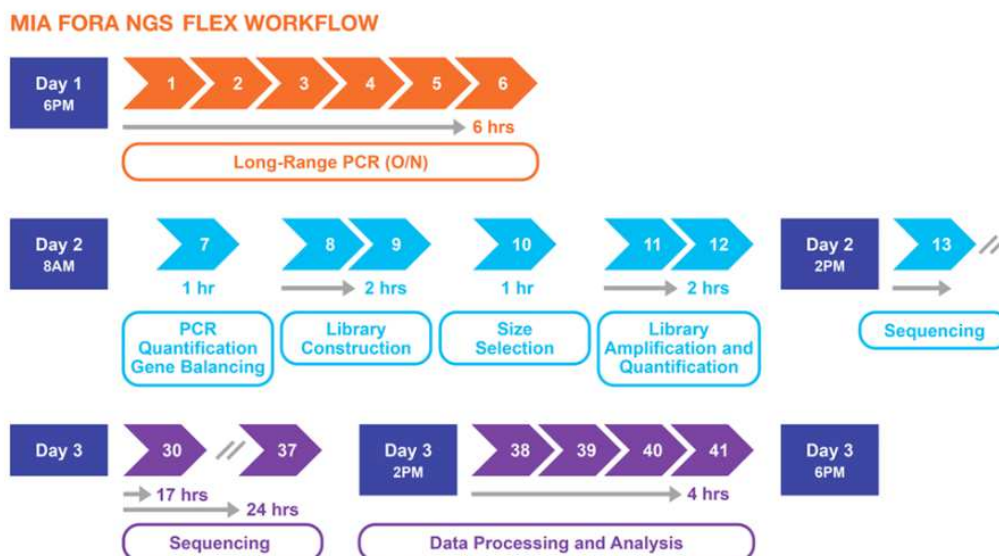
MIA FORA™ NGS Mflex sfrutta la tecnologia Illumina, una metodica di sequenziamento di seconda generazione che opera con una bridge PCR, cioè una PCR "a ponte" che permette il cosiddetto "Sequencing By Synthesis" (SBS), ovvero una procedura di sequenziamento per la quale la sequenza nucleotidica è letta grazie alla sintesi degli ampliconi con dNTPs (nucleotidi) terminatori di catena marcati con fluorescenza.

Come riportato nel capitolo 5 “Materiali e Metodi”, il kit di reagenti che si utilizza per il sequenziamento NGS Illumina è MIA FORA™ NGS MFlex HLA Typing Kits, prodotto dalla ditta Immucor. In base al kit utilizzato, è possibile la tipizzazione a 6 o 11 loci di 24 campioni: la tipizzazione a 6 loci include HLA-A, B, C, DRB1, DQB1 e DPB1, mentre il sequenziamento a 11 loci comprende HLA-A, B, C, DRB1, DRB3/4/5, DQA1, DQB1, DPA1 e DPB1.

### 3.1.1 WORKFLOW IN NGS

La tipizzazione con tecnologia Illumina, come ogni metodica di sequenziamento di seconda generazione, può essere suddivisa in tre macrofasi essenziali per un buon sequenziamento (Fig. 7):

1. Generazione del target: prevede una long PCR (PCR a lungo raggio) che amplifica la quantità di campione che sarà poi disponibile per la libreria
2. Generazione della libreria: composta da diversi sottopassaggi necessari per eseguire il sequenziamento in formato multiplex
3. Sequenziamento: è l'effettiva tipizzazione del DNA, cioè la lettura della successione di basi che determinano l'aplotipo del soggetto in esame.



*Figura 7.* Diagramma riassuntivo del workflow in NGS, suddivisibile in tre giornate lavorative.

Ovviamente, il tutto è una procedura complessa, che richiede un lavoro manuale che si protrae per alcune ore; nel laboratorio di Mestre, l'intero flusso si svolge in tre giornate lavorative, che corrispondono alle tre macrofasi delineate precedentemente:

1. Prima giornata: è dedicata alla preparazione dei campioni per la reazione di long PCR;
2. Seconda giornata: si opera la frammentazione e riparazione delle estremità dei segmenti di DNA target amplificati, avviene la ligazione agli indici per l'univoca identificazione dei frammenti e la size selection della libreria per il target richiesto dal metodo di sequenziamento;
3. Terza giornata: inizia con l'amplificazione della libreria tramite primers specifici per la suddivisione dei frammenti in clusters, sino ad arrivare ad una libreria correttamente quantificata e pronta per essere caricata nello strumento per il sequenziamento con tecnologia Illumina.



## **4. SCOPO DELLA TESI**

Nell'ambito dei trapianti è fondamentale conoscere minuziosamente l'aplotipo HLA del paziente ricevente e del donatore di CSE. Ad oggi, sono state messe a punto numerose tecnologie e metodiche che permettono di identificare gli antigeni HLA con differente sensibilità e risoluzione. Fra le tecniche più affidabili, troviamo SSO, SBT e, più recentemente, l'NGS. Fra queste, il sequenziamento di seconda generazione ha introdotto enormi vantaggi all'interno dei laboratori di Immunogenetica, primo fra tutti la risoluzione dei risultati spinta fino al terzo/quarto campo, quindi una risoluzione allelica, che evidenzia anche le variazioni delle regioni esoniche al di fuori dell'ARD e introniche, che possono essere responsabili di cambiamenti a livello di espressione della proteina MHC. Oltre a ciò, come già riportato, altri vantaggi sono il ridotto tempo di refertazione, il deciso abbattimento dei costi di analisi, grazie alla possibilità di sequenziamento massiccio di più loci su più campioni in un'unica corsa, e l'alta affidabilità dei dati ottenuti, che sono certi, inequivocabili e privi di ambiguità, in merito all'introduzione di algoritmi informatici all'avanguardia.

Lo scopo della tesi è dimostrare come l'introduzione dell'NGS abbia migliorato il flusso di lavoro all'interno del Laboratorio HLA del S.I.T. dell'Ospedale dell'Angelo di Mestre, evidenziando come i vantaggi sopraelencati siano riscontrabili anche in questa realtà. Si descriverà come è stata standardizzata e validata la procedura operativa interna al laboratorio a partire dalla metodica indicata nel kit MIA FORA NGS MFlex della ditta Immucor [19], analizzando nel dettaglio ogni passaggio e cercando di evidenziare l'importanza che ricopre ciascuna fase per la buona riuscita della tipizzazione HLA.

Infine, sulla base del protocollo steso dai dirigenti in merito al percorso di studio familiare (vedi capitolo 2.2.2), si porteranno due esempi di tale studio, i cui risultati sono stati confermati applicando la genotipizzazione a 11 loci e a 6 loci, a dimostrazione di quanto sia fondamentale per l'identificazione del donatore compatibile per il paziente candidato a trapianto allogenico di cellule staminali emopoietiche.



## 5. MATERIALI E METODI

### 5.1 GENERAZIONE DEL TARGET: ACCETTAZIONE ED ESTRAZIONE DEL DNA

Il campione da analizzare deve essere raccolto in una provetta Vacutainer contenente EDTA come anticoagulante; solamente con l'EDTA, infatti, è possibile conservare correttamente un prelievo per biologia molecolare, in quanto, con la sua azione, sottrae dal sangue  $\text{Ca}^{2+}$  e  $\text{Mg}^{2+}$ , sostanze che attivano le nucleasi (DNAsi e RNAsi), responsabili della degradazione degli acidi nucleici.

Si procede poi alla fase di accettazione del campione, che consiste nell'assegnare alla provetta un codice numerico, in modo da ottenere un'identificazione univoca che evita errori derivanti da omonimia e che rispetta la privacy del paziente. La richiesta dell'esame va inserita nel sistema gestionale del laboratorio, in questo caso di Emonet, per la completa tracciabilità dei dati. [20]; in caso di donatori di CSE, si consulta IBMDR per verificare che i dati anagrafici del donatore siano inseriti correttamente [21].

Si procede quindi all'estrazione del DNA, procedura da svolgere in una stanza dedicata o pre-PCR, cioè un'area in cui si utilizza solamente DNA non amplificato allo scopo di evitare contaminazioni che potrebbero compromettere il risultato.

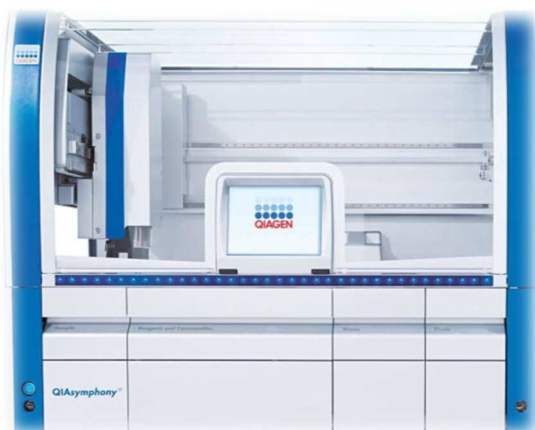
Nel laboratorio HLA di Mestre, l'estrazione del DNA si esegue con lo strumento QiaSymphony SP della ditta QIAGEN (Fig.8), che permette di ottenere estrazioni automatizzate di acidi nucleici da differenti matrici, con bassi livelli di frammentazione del DNA e di sostanze contaminanti, sia chimiche che biologiche, per cui l'eluato si presta anche a metodiche di tipizzazione ad alta risoluzione [22]. La procedura di estrazione e purificazione sfrutta quattro tipologie di reagente principali, ognuno dei quali corrisponde ad una fase diversa, come illustra la Tabella 1.

Tampone	Azione
Tampone di lisi	Lisi delle cellule nucleate contenute nel campione di sangue intero (leucociti) in modo da recuperare il DNA contenuto al loro interno
Particelle magnetiche + tampone di legame	Cattura del DNA liberato dalle cellule, che si lega alle biglie magnetiche grazie al tampone di legame; le biglie sono attratte e catturate da un magnete così da eliminare eventuali residui e impurità derivati dall'estrazione del DNA
Tampone di lavaggio	Rimozione di eventuali sostanze inquinanti presenti nel campione, lasciando una soluzione di DNA puro

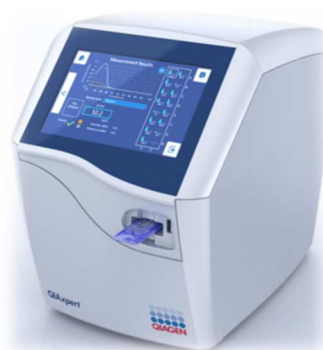
Tampone di eluizione	Prodotto dell'estrazione dispensato in Micro Tube da 2 MI (200 $\mu$ L); è un DNA di alta qualità, utilizzabile con tecniche molecolari altamente sensibili.
----------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------

*Tabella 1.* Reagenti usati durante l'estrazione con lo strumento QiaSymphony

Il DNA estratto può essere conservato in frigorifero, ad una temperatura di 2-8°C per un massimo di 5 giorni oppure, per una conservazione a lungo termine, lo si congela a temperature di -20°C o -80°C.



*Figura 8.* Estrattore QiaSymphony SP della ditta Qiagen.



*Figura 9.* Spettrofotometro QIAxpert della ditta Qiagen

Prima di procedere al congelamento, è opportuno andare a misurare la concentrazione di DNA ottenuta con l'eluato. La concentrazione di DNA da un prelievo di sangue periferico è dipendente dalla quantità di cellule nucleate presenti in esso, cioè dalla conta dei leucociti. Generalmente, pazienti ematologici con leucemie producono una grande quantità di globuli bianchi e di conseguenza presentano un'elevata concentrazione di DNA in una singola estrazione da 200  $\mu$ L di eluato; al contrario, pazienti in aplasia midollare possiedono pochi leucociti per cui si otterrà un'estrazione con DNA molto diluito, anche inferiore a 10 ng/ $\mu$ L, tanto che, in molti casi, è necessario procedere con un'ulteriore estrazione da 100  $\mu$ L, in modo da ottenere un eluato più concentrato e meglio applicabile alle varie tecniche di biologia molecolare, ciascuna delle quali necessita di una determinata quantità di acido nucleico per poter dare risultati affidabili.

Lo strumento utilizzato per la quantificazione è il QIAxpert, prodotto sempre dalla ditta QIAGEN (Fig. 9) [23]. Questo strumento è uno spettrofotometro che sfrutta una combinazione di raggi UV e di microfluidica ad alta velocità per la quantificazione degli acidi nucleici e per la valutazione della purezza degli eluati. Il protocollo di lavoro prevede l'utilizzo di 2  $\mu$ L di eluato, prelevati dopo aver vortexato e centrifugato il campione in modo opportuno, in modo che siano rappresentativi dell'intero campione. Si utilizza una cartuccia con sedici sistemi di microfluidica, in grado quindi di processare sedici campioni in un'unica seduta d'analisi. Una volta inserita la slide nello strumento, si imposta la misurazione dell'assorbanza a 260 nm, lunghezza d'onda a cui le basi azotate degli acidi nucleici

assorbono la luce ultravioletta. Dopo pochi secondi, lo strumento riporta i dati misurati sotto forma di numero e di grafico; di nostro interesse sono principalmente i tre parametri elencati in Tabella 2 [24].

Parametro	Valore	Descrizione
Concentrazione di DNA	Superiore a 1 ng/ $\mu$ L	Concentrazione del DNA dipendente dalla quantità di leucociti del paziente
A <sub>260/280</sub>	Compreso fra 1,7 e 2,0	Rapporto fra le assorbanze alle lunghezze d'onda di 260 nm (valido per gli acidi nucleici) e 280 nm, assorbanza delle proteine. Indica il livello di pulizia del campione, quindi la purezza dell'eluato. Una diminuzione indica una contaminazione da proteine, mentre un aumento è indicatore di contaminazione da RNA
A <sub>260/230</sub>	Compreso fra 1,8 e 2,2	230 nm è la lunghezza d'onda di assorbanza dei carboidrati, dei fenoli e degli agenti caotropici; rapporti inferiori ai valori di riferimento sono indice di contaminazione da solventi

**Tabella 2.** Parametri misurati dallo strumento QiaExpert

A questo punto si è pronti per iniziare una seduta di NGS.

### 5.1.1 LONG PCR

All'interno del laboratorio di HLA di Mestre, si effettuano di norma sedute di NGS da 24 campioni per l'utilizzo ottimale del kit di reagenti MIA FORA™ NGS MFlex HLA Typing Kits (Fig. 10); il kit è composto da differenti mix, ognuna con un ruolo specifico, come riassunto in Tabella 3.



**Figura 10.** Reagenti del kit di tipizzazione HLA MIA FORA, piastra degli indici/adattatori e biglie magnetiche per la pulizia del campione.

Nome	Contenuto	Descrizione
MFlex 6/Mflex 11	All-in-one PCR Mix	Provetta contenente tutte le componenti necessarie per lo svolgimento della long PCR, compresa la Taq polimerasi
L1	Mix di enzimi primari	Enzimi di frammentazione degli ampliconi della PCR Enzimi di riparazione delle estremità frammentate Tail A per l'estremità '3
L2	Mix di buffer primario	Buffer associato alla reazione di amplificazione di L1
L3	Enzima ligasi	Ligasi per mediare il legame degli ampliconi con gli indici
L4	Buffer ligasi	Buffer associato alla reazione di amplificazione di L3
L5	Buffer mix	Enzimi e mix necessari alla reazione di PCR post selezione della dimensione degli ampliconi
L6	Primer di amplificazione	25 µL di primer di amplificazione F e 25 µL di primer di amplificazione R per la creazione dei cluster Illumina
L7	Acqua esente da nucleasi	

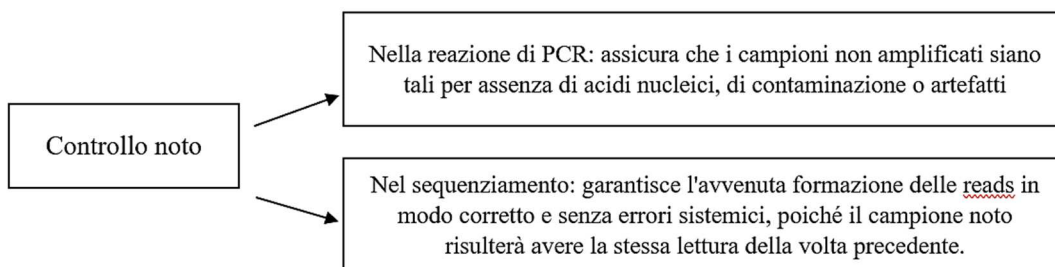
**Tabella 3.** Elenco dei reagenti contenuti nel kit MIA FORA NGS MFlex HLA Typing Kit

Oltre a questi reagenti, è presente anche una piastra di adattatori/indici. Si tratta di sequenze a codice a barre univoche che hanno lo scopo di identificare ogni filamento di DNA durante il sequenziamento, così da poter sempre risalire a che campione appartiene una determinata sequenza. Gli indici sono forniti in una piastra da 96 che, dopo essere stata centrifugata, viene assemblata al contenuto della piastra "Primary prep clean": gli indici si dispensano con una pipetta multicanale fra colonne corrispondenti delle due piastre, per cui, per una seduta da 24 campioni, si vengono ad utilizzare solamente le colonne 1-3.

La fase della generazione del target prevede la realizzazione di una long PCR, che ha lo scopo di amplificare le sequenze di interesse dei campioni di hgDNA (human genomic DNA) per ottenere una grande quantità di target da cui iniziare la procedura di sequenziamento.

Per prima cosa, è essenziale stilare una lista dei campioni da analizzare, selezionandone 22 fra donatori di CSE e pazienti ematologici che necessitano una tipizzazione in alta risoluzione; questi andranno ad occupare i pozzetti 3-24 della piastra da 96 pozzetti. I primi due pozzetti sono invece il luogo in cui si verranno ad inserire rispettivamente il NTC (Non-Template Control) e il controllo noto. L'NTC è il controllo negativo, costituito da un pozzetto (nel nostro caso il pozzetto 1) riempito con acqua libera da nucleasi (UltraPure™ DNase/RNase-Free Distilled Water) in modo da poter verificare la presenza di eventuali contaminazioni della piastra di lavoro con DNA estraneo. Il

controllo noto è un controllo positivo, cioè un campione che è stato precedentemente eseguito con la medesima metodica e di cui si è certi del risultato, che funge dunque da garante del successo della seduta. Nel caso in cui si operi un sequenziamento, il controllo noto viene scelto fra i campioni che sono già stati tipizzati in precedenza, generando una sequenza nucleotidica certa, chiara e dunque ripetibile; solo in questo modo si è sicuri che il sequenziamento sia avvenuto senza errori tecnici. Lo Schema 1 riassume gli obiettivi del controllo noto.



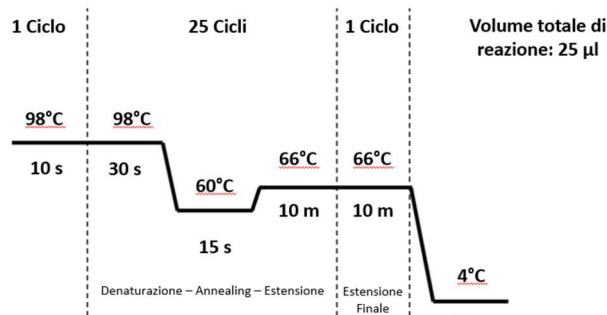
Schema 1. Riassunto della funzione del controllo noto in NGS.

La lista dei campioni viene riportata all'interno di un foglio di calcolo Microsoft Excel (Fig. 11), prodotto dal fornitore MIA FORA, assieme alle corrispondenti concentrazioni. Questa tabella si basa su un algoritmo che riporta il volume di diluizione del DNA da provetta madre per arrivare ad ottenere una concentrazione ottimale di DNA pari a 15 ng/μL in 50 μL totali; come si vede in Figura 11, la quinta colonna riporta la quantità di acqua nucleasi free, mentre la quarta il volume di eluato madre da diluirvi in base alla concentrazione di DNA di partenza.

IMMUCOR		Volume finale di diluizione (uL)		50	
MIA FORA™ NGS					
Pos. Piastra	Inserisci ID del campione	Inserisci la concentrazione iniziale campione(ng/uL)	Volume DNA da prelevare (uL)	Acqua per diluizione(uL)	
A1	NTC	0	#DIV/0!	#DIV/0!	
B1	Controllo noto 211153	42.3	20.1	29.9	
C1	211160	34.9	24.4	25.6	
D1	211161	57.6	14.8	35.2	
E1	211162	34.5	24.6	25.4	
F1	211175	55.1	15.4	34.6	
G1	211180	51.7	16.4	33.6	
H1	211181	35	24.3	25.7	
A2	211182	23.4	36.3	13.7	
B2	211189	58.4	14.6	35.4	
C2	211190	43.5	19.5	30.5	
D2	211191	64.8	13.1	36.9	
E2	211192	36.3	23.4	26.6	
F2	211195	42.4	20.0	30.0	
G2	211196	48	17.7	32.3	
H2	211197	41.8	20.3	29.7	
A3	211198	58.5	14.5	35.5	
B3	211199	69.9	12.2	37.8	
C3	211200	36.1	23.5	26.5	
D3	211201	50.8	16.7	33.3	
E3	211202	60.5	14.0	36.0	
F3	211203	57.1	14.9	35.1	
G3	211204	44.6	19.1	30.9	
H3	211206	62.4	13.6	36.4	

Figura 7. Esempio di foglio di calcolo Excel utilizzato nelle diluizioni con il kit MIA FORA NGS. La prima colonna riporta l'identificativo del campione, mentre la seconda la corrispondente concentrazione. In terza e quarta colonna sono invece indicate rispettivamente le quantità di DNA e acqua necessarie. Il volume totale è di 50 μL.

Dopo la diluizione dell'eluato madre e la dispensazione in piastra, si procede all'aggiunta al DNA di 15 μL di una miscela di reazione (all-in-one PCR mix). La reazione di long PCR dura all'incirca quattro ore, durante le quali si alternano cicli con durata e temperatura di reazione differenti (Fig. 12).



**Figura 8.** Schema della reazione di long PCR svolta dal termociclatore Veriti.

Terminata questa fase, è possibile sospendere il lavoro e conservare la piastra amplificata a  $-20^{\circ}\text{C}$  fino a quattro giorni senza alterare la qualità del prodotto.

## 5.2 GENERAZIONE DELLA LIBRERIA: FRAMMENTAZIONE E LIGAZIONE AGLI INDICI

La fase successiva è dedicata alla preparazione della libreria NGS, cioè un insieme di ampliconi dei loci genici in esame univocamente identificati che abbiano tutti una lunghezza di basi specifica. Da questa fase in poi, l'intero lavoro è svolto in stanza post-PCR in quanto si opera con sequenze di DNA amplificate.

La costruzione della libreria si presenta in due fasi principali: la preparazione dei campioni per il legame con gli adattatori e la size selection degli ampliconi.

Per prima cosa viene preparata una mix con i reagenti L1 ( $49.5\ \mu\text{L}$ ) e L2 ( $313.5\ \mu\text{L}$ ) per una reazione di amplificazione della durata di circa 40 minuti, che si svolge in una nuova piastra denominata "Primary Prep". In questo momento, i prodotti della long PCR subiscono una frammentazione enzimatica, la riparazione delle estremità e l'aggiunta di una coda terminante con un A in estremità 3' per il successivo legame con l'adattatore che presenta una sporgenza 5'-T.

Prima di procedere alle fasi successive, è necessario purificare il prodotto amplificato tramite una pulizia con biglie magnetiche, processo che sarà svolto in diversi momenti della preparazione della libreria in modo da arrivare al sequenziamento con il prodotto più puro possibile. Per la pulizia sono necessari:

- Biglie magnetiche: fornite dalla ditta Immucor assieme al kit di reagenti NGS
- Etanolo 80%: da preparare al momento
- Tris HCl 10 mM pH 8: una soluzione tamponante da preparare al momento, che permette di mantenere stabile il pH della soluzione senza inibire l'azione degli enzimi presenti; allo stesso tempo, permette di staccare i filamenti di DNA agganciati alle biglie magnetiche per il recupero dell'eluato dopo la purificazione.

A ciascuna delle tre soluzioni corrisponde una fase della purificazione del campione; inizialmente vengono aggiunte le biglie magnetiche, miscelandole bene in ogni pozzetto contenente il DNA target



e incubandole per cinque minuti, in modo da permettere il legame fra biglie e DNA. Si procede poi ponendo la piastra "Primary prep" nella Magnum FLX, una piastra magnetica: così facendo, le biglie con legato il DNA si dispongono ad anello attorno alla parete del pozzetto, lasciandovi un surnatante contenente impurità e aspecificità, che saranno eliminate dopo tre minuti di incubazione. Il secondo step prevede tre lavaggi con l'etanolo (30 secondi ciascuno), durante i quali la piastra permane adesa al magnete per fissare per bene il DNA target alle biglie ed eliminare l'eccesso. Dopo aver fatto evaporare l'eccesso di etanolo allontanando la piastra dal magnete, si procede con il terzo passaggio che consiste nell'aggiunta di 25  $\mu\text{L}$  di Tris HCl che si va a miscelare con le biglie, facendole staccare dalla parete; si lascia incubare per cinque minuti e poi si riposiziona la piastra sul magnete per tre minuti. In tal modo, si consente al Tris HCl di sciogliere il legame che tiene unito il DNA alle biglie magnetiche, liberandolo nella soluzione (eluato); alla fine, vengono prelevati 23  $\mu\text{L}$  di eluato da ciascun pozzetto e pipettati nei pozzetti corrispondenti di una nuova piastra chiamata "Primary prep clean".

La fase successiva prevede il legame dell'adattatore/indice ai campioni di DNA. Questo passaggio consiste nell'allestimento di una nuova PCR, che sfrutta come master mix la miscela fra i reagenti L3 (30  $\mu\text{L}$ ) e L4 (750  $\mu\text{L}$ ) da miscelare in ogni pozzetto contenente DNA assieme a 8  $\mu\text{L}$  di indici; dopo l'assemblaggio agli adattatori, la piastra viene denominata "Ligasi". Anche questa reazione dura circa quaranta minuti.

### **5.2.1 SIZE SELECTION DELLA LIBRERIA**

Dopo la ligazione degli indici, ogni frammento di DNA è sempre identificabile per cui è possibile unirli tutti in un'unica sede di lavoro; questo processo si chiama consolidamento della libreria o pooling, e prevede il raggruppamento del contenuto di ogni pozzetto della piastra in una provetta Eppendorf da microcentrifuga denominata "Consolidated". Prendendo la piastra "Ligasi", si sceglie una colonna centrale, ad esempio la colonna 7, nella quale, con una pipetta multicanale, si depositano 27  $\mu\text{L}$  del contenuto delle colonne 1-3 contenenti i campioni. Da questa colonna, con una pipetta singola impostata a 72  $\mu\text{L}$ , si porta il contenuto dei pozzetti A-H all'interno della Eppendorf, ottenendo un volume totale di campioni del pool finale pari a 576  $\mu\text{L}$ .

Segue poi la pulizia del campione con le biglie magnetiche, con il prelievo di 60  $\mu\text{L}$  di eluato che viene depositato in una nuova provetta da microcentrifuga denominata "Consolidated clean". Da qui, si prelevano 30  $\mu\text{L}$  di eluato, depositandoli in una nuova Eppendorf etichettata come "Pippin Size Selection", conservando il campione originale a  $-20^{\circ}\text{C}$ .

Vengono quindi selezionati i prodotti ligati agli indici con una lunghezza compresa fra 500-900 bp, eliminando quindi le aspecificità. Questo passaggio viene compiuto con lo strumento Pippin Prep, prodotto dalla ditta Sage Science (Fig. 13), e ha lo scopo di consentire la produzione di librerie NGS

con solo i frammenti con lunghezza compresa nelle dimensioni target di 500-900 bp [25]. La corsa del DNA si svolge in gel di agarosio 1.5% (Fig. 14), con una durata di circa quaranta minuti. Al termine, si prelevano circa 40  $\mu$ L di eluato dall'apposita camera, che saranno utilizzati per l'ultima parte del flusso di lavoro NGS.

### 5.3 SEQUENZIAMENTO: AMPLIFICAZIONE PER CLUSTERS

Il sequenziamento vero e proprio ha luogo in terza giornata, dopo aver amplificato l'eluato, al fine di aumentare la quantità di copie di target che possono generare reads appropriate e di permettere l'amplificazione delle sequenze con i primers Illumina forward (F) e reverse (R), che contengono le sequenze necessarie alla generazione di cluster nella flow cell di sequenziamento. Questa reazione avviene grazie alla master mix preparata con L5 (25  $\mu$ L), L6 (1  $\mu$ L) e L7 (9  $\mu$ L), unite a 15  $\mu$ L di libreria selezionata per dimensione.

Anche alla fine di questa amplificazione si procede con la pulizia dell'amplificato con le biglie magnetiche, ottenendo un eluato finale di 17  $\mu$ L da depositare in una nuova Eppendorf denominata "Clean library".



Figura 10. Strumento Pippin Prep della ditta Sage Science.

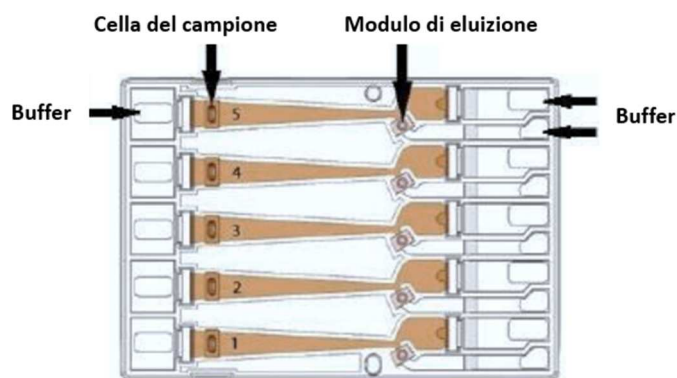


Figura 9. Rappresentazione grafica di una Gel Cassette per la DNA size selection.

#### 5.3.1 QUANTIFICAZIONE CON QUBIT E CARICAMENTO DELLA LIBRERIA

Per procedere al sequenziamento con la tecnologia Illumina è necessario che la libreria sia portata ad una concentrazione prestabilita di 10 pM, tale per cui possa essere amplificata in cluster con la bridge PCR e poi sequenziata nel modo più pulito possibile. Per ottenere tale diluizione, si deve procedere con la quantificazione della libreria, cioè si deve determinare quanto dsDNA sia ancora presente a seguito delle precedenti manipolazioni. Questa tappa avviene con uno strumento compatto della ditta ThermoFisher Scientific chiamato Qubit 4 Fluorometer (Fig. 15), che rende possibile una quantificazione di RNA, DNA e proteine molto più accurata rispetto alle metodiche che sfruttano l'assorbanza dei raggi UV, rilevando i coloranti fluorescenti legati allo specifico target di interesse, dsDNA nel nostro caso, anche in bassissime quantità [26]. Per fare ciò, lo si associa ad un kit chiamato Qubit™ dsDNA HS and BR Assay Kits, che contiene una soluzione di buffer, un colorante

fluorescente e due standard di misurazione: questi ultimi sono denominati Standard 1 e Standard 2 e rendono lo strumento capace di leggere selettivamente e velocemente una quantità molto bassa di DNA e una molto alta, andando a creare la curva di calibrazione necessaria per la corretta quantificazione del DNA presente nel campione.

Prima di procedere alla determinazione della concentrazione del DNA, si procede preparando quattro microprovette che corrispondono a standard 1 (S1), standard 2 (S2), controllo negativo (N) e libreria (L), come riportato in Tabella 4.

S1	190 $\mu$ L di buffer con colorante fluorescente + 10 $\mu$ L di standard 1
S2	190 $\mu$ L di buffer con colorante fluorescente + 10 $\mu$ L di standard 2
N	198 $\mu$ L di buffer con colorante fluorescente + 2 $\mu$ L di acqua distillata nuclease-free
L	198 $\mu$ L di buffer con colorante fluorescente + 2 $\mu$ L di libreria

**Tabella 4.** Descrizione delle provette necessarie alla quantificazione con il Qubit

Prima di inserire le provette nel Qubit si devono lasciar incubare per due minuti al buio e poi si può proseguire con la quantificazione. Il valore ottimale di concentrazione è compreso fra 4 e 10 ng/ $\mu$ L, anche se abbiamo avuto esperienza di sedute di sequenziamento ben riuscite anche con valori di libreria inferiori a 4 ng/ $\mu$ L, come 1,89 ng/ $\mu$ L, o superiori a 10 ng/ $\mu$ L, come 17,3 ng/ $\mu$ L.



**Figura 11.** Fluorometro Qubit, prodotto dalla ditta Thermo Fischer.

A questo punto, si ritorna sul foglio di calcolo di Microsoft Excel utilizzato inizialmente per il calcolo delle diluizioni dei vari campioni (Fig. 8), posizionandosi su una differente tabella denominata "Qubit Calculator"; si inserisce il valore di libreria trovato nell'apposita casella, ottenendo così i valori di libreria e di Tris HCl da cui partire per portare la libreria alla concentrazione di 10 pM richiesta dal sequenziamento Illumina.

A questo punto inizia il cosiddetto "caricamento della libreria in cartridge", cioè tutta una serie di diluizioni che permettono di raggiungere la quantità ottimale. Per prima cosa è necessario avere a portata di mano 100  $\mu$ L di NaOH 0,2 N, preparati al momento aggiungendo 20  $\mu$ L di NaOH 1 N a 80  $\mu$ L di acqua, e il tampone di ibridazione HT1, un buffer specifico per le diluizioni di librerie denaturate e librerie di controllo che viene fornito congelato a -20°C assieme alla cartridge Illumina (Fig. 19). Nella Tabella 5 sono illustrati i vari passaggi necessari ad ottenere una libreria 10 pM.

Step	Libreria	PhiX
1°	4 nM: Tris HCl + libreria secondo le quantità indicata dal foglio di calcolo	2 nM: 8 µL Tris HCl + 2 µL PhiX 10 nM
2°	40 pM: 10 µL libreria 4 nM + 10 µL NaOH 0,2 N + 980 µL HT1	20 pM: 10 µL PhiX 2 nM + 10 µL NaOH 0,2 N + 980 µL HT1
3°	10 pM: 250 µL libreria 40 pM + 750 µL HT1 + 25 µL PhiX 20 pM	

**Tabella 5.** Diluizioni pre-caricamento della libreria in MiSeqDx.

Il PhiX è una libreria pronta all'uso che funge da libreria di controllo per le corse di sequenziamento Illumina. Ottenuta per la prima volta nel 1977 dal sequenziamento di un batteriofago denominato PhiX-174, il PhiX è una libreria concentrata costituita da DNA 10 nM in 10 µL con sequenze di circa 500 bp, caratterizzata da una composizione bilanciata di basi (45% CG e 55% AT), con l'obiettivo di fungere da ottimo controllo di calibrazione, di qualità e di sequenziamento. Se il Phix non fosse già alla giusta concentrazione, è possibile raggiungere i 20 pM con passaggi paralleli a quelli necessari per la diluizione della libreria (Tab. 5).

A questo punto basta solamente inserire 600 µL della libreria 10 pM nel pozzetto 17 della cartridge di reazione Illumina (Fig. 19) e inserirla nell'apposito vano dello strumento MiSeqDx. Al contempo, sono da inserire la flow cell e il tampone di lavaggio PR2; a questo punto si è pronti per iniziare il sequenziamento.

#### **5.4 MiSeqDx™ Illumina®**

Parallelamente al flusso di lavoro manuale per la costruzione delle librerie, è necessario anche preparare lo strumento di lavoro. Per capire come impostarlo, è giusto chiarire il funzionamento generale dello strumento e delle sue componenti, descrivendo anche il software in uso, in quanto ci troviamo di fronte ad una tecnologia che racchiude in un unico corpo la generazione di cluster, il sequenziamento vero e proprio e l'analisi dei dati [27].

Lo strumento impiegato nel laboratorio HLA di Mestre è il MiSeqDx™ Illumina®, prodotto dalla ditta Illumina, Inc [28]. È un sequenziatore da banco molto semplice che sfrutta il Sequencing By Synthesis (SBS), il sequenziamento per sintesi, realizzando l'analisi genomica in sole otto ore e con automazione completa a seguito dell'inserimento del campione e dei reagenti. È composto dai seguenti elementi (Fig. 16):

- Scomparto della cella a flusso (flow cell): area di inserimento della flow cell di reazione, sulla quale avviene la generazione dei cluster e l'analisi dei campioni
- Scomparto ottica incorporato: componenti ottici che consentono l'imaging della flow cell
- Scomparto dei reagenti: area di caricamento del tampone di lavaggio PR2, del flacone dello scarico; contiene anche un vano refrigerato, detto chiller, che mantiene la cartuccia dei reagenti ad una temperatura costante di 4° C
- Monitor touch screen: sistema di interfaccia dell'operatore con lo strumento, tramite il quale è possibile controllare la configurazione del sistema, l'impostazione della corsa, ma anche creare il collegamento fra lo strumento e i software di analisi.



*Figura 12.* Strumento MiSeqDx della ditta Illumina.

Andiamo ora a vedere nel dettaglio ogni componente necessaria allo strumento MiSeqDx per una sessione di NGS.

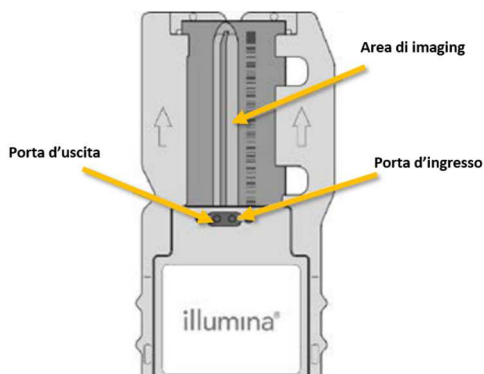
Il primo elemento che lo strumento chiede di inserire è la flow cell o cella a flusso, uno substrato su vetro monouso nel quale si viene a depositare la libreria per la creazione dei cluster (Fig. 17) [29]. È costituita da una porta d'ingresso, un'area di imaging a singola corsia e una porta d'uscita, dalla quale gli scarti defluiscono nell'apposito flacone (Fig. 18). Questa flow cell è "patterned", cioè contiene dei nanopozzetti che permettono di avere un elevato livello di throughput, cioè di rendimento sia in qualità e quantità che come riduzione dei tempi e dei costi d'analisi, poiché vi è un miglior utilizzo di tutta la superficie della cella. I clusters, infatti, possono formarsi solamente in corrispondenza dei pozzetti predefiniti, evitando così il sovraccarico della superficie; in questo modo si ottengono una miglior risoluzione dell'imaging delle coordinate X e Y dei cluster e una miglior densità distributiva, fattori che garantiscono delle buone metriche qualitative e un'analisi più sicura. Questo accade poiché ogni pozzetto è coattato con dei probe di DNA specifici, che occupano una posizione ben precisa e mappata dal software, rendendo così possibile il tracciamento di ogni amplificazione e sequenziamento. I clusters si formano all'avvio della corsa, quando la soluzione contenente la libreria viene fatta scorrere sulla superficie della flow cell; i frammenti di libreria complementari ai probe si

legano nella posizione corrispondente e in contemporanea avviene la bridge PCR, l'amplificazione a ponte peculiare della tecnologia Illumina. Operando semina e amplificazione allo stesso tempo viene ottimizzato il tempo di lavoro e si riduce il rischio di aspecificità, cioè l'attacco di un frammento di libreria presso nanocelle con primers non complementari.

Dopo la flow cell, avviene il caricamento del tampone di lavaggio PR2 e del flacone degli scarti,



**Figura 13.** Flow cell Illumina



**Figura 14.** Rappresentazione grafica dei componenti della flow cell Illumina.

seguito dalla cartuccia dei reagenti, chiamata cartridge (Fig. 19). È una cartuccia monouso, costituita da 22 serbatoi numerati e pre-riempiti con differenti reagenti utili allo strumento per dirigere la generazione dei cluster della libreria e per il sequenziamento; il pozzetto 17, identificato come “Load Samples”, è dedicato al caricamento dei 600  $\mu\text{L}$  di libreria 10 pM denaturata. La cartuccia viene conservata congelata ad una temperatura di circa  $-20^{\circ}\text{C}$  e scongelata solamente all'utilizzo con due modalità: ponendola in un frigorifero alla fine della seconda giornata di lavoro oppure scongelandola in terza giornata immergendola in un bagno d'acqua a temperatura ambiente; in base all'esperienza maturata all'interno del laboratorio dell'ospedale di Mestre, si può affermare che il metodo migliore è lo scongelamento a  $4^{\circ}\text{C}$  overnight poiché i reagenti, al mattino seguente, risultano perfettamente scongelati e privi di cristalli di ghiaccio. Al contrario, la flow cell viene conservata assieme il tampone di lavaggio PR2 ad una temperatura di  $4-8^{\circ}\text{C}$ .

Assieme alla cartridge viene fornito anche il tampone di ibridazione HT1, utilizzato poco prima dell'avvio del sequenziamento per le diluizioni della libreria fino alla concentrazione di 10 pM.



**Figura 15.** Componenti necessari al sequenziamento Illumina nel MiSeqDx.

### **5.4.1 SOFTWARE MiSeqDx**

Prima di iniziare una nuova corsa, è necessario preparare virtualmente lo strumento al tipo di analisi che si vuole svolgere e su quali campioni; il MiSeqDx viene gestito da quattro applicazioni integrate che gestiscono aree differenti [30]. La prima è il MiSeqDx Control Software (MCS) che dirige lo strumento facendosi da tramite con l'operatore, indirizzandolo durante il caricamento di flow cell, tampone di lavaggio e cartridge; durante la corsa, permette la visione di una panoramica sulle statistiche di qualità.

Il secondo software è il Real Time Analysis (RTA) che si occupa dell'analisi dell'imaging della flow cell, infatti identifica le basi che compongono la libreria amplificata assegnando anche un punteggio qualitativo a ciascuna base per ogni ciclo di lettura.

Il software più importante per la gestione di un sequenziamento è invece il Local Run Manager, accessibile direttamente dalla schermata dello strumento. Oltre che a monitorare il sequenziamento dei campioni per tutta la durata della corsa, permette di impostare e creare la corsa stessa, registrando le informazioni riguardo i campioni presenti e la posizione che essi occupano nella piastra, associandovi dunque il corretto numero di indici per un'identificazione univoca; alla fine, gestisce l'analisi secondaria, che associa la chiamata delle basi ai campioni in esame.

Esiste infine un quarto software molto importante per la valutazione delle metriche di qualità del sequenziamento; si tratta di un programma Illumina chiamato Sequencing Analysis Viewer (SAV), che compie analisi della qualità della corsa sia durante che al suo termine, integrando i dati ricavati da RTA e generando diagrammi e grafici per da avere una visualizzazione globale qualitativa più immediata [31].

## **5.5 ANALISI DEI DATI DI SEQUENZIAMENTO**

L'analisi dei dati è uno step di routine fondamentale all'interno di un laboratorio sia diagnostico che di ricerca poiché serve a garantire dei risultati corretti e affidabili ai pazienti; figure come il tecnico di laboratorio, il dirigente medico o biologo devono dedicarsi al controllo di qualità dei dati a seconda della strumentazione con cui lavorano. Nei paragrafi successivi, si metterà in evidenza come la combinazione fra le informazioni ricavate dalle metriche di qualità dello strumento MiSeqDx e dagli algoritmi MIA FORA, permettano di ottenere dei risultati di sequenziamento certi.

### **5.5.1 METRICHE DI QUALITÀ IN NGS**

Ogni strumento laboratoristico deve avere dei sistemi di controllo di qualità per poter emettere dei risultati che rispecchino la realtà, cioè che siano affidabili. Normalmente, si valuta che lo strumento agisca con precisione e con accuratezza, fornendo dei risultati ripetibili nel tempo e nello spazio. Prendendo come esempio il Laboratorio Analisi, accade che gli strumenti svolgono quotidianamente i cosiddetti CQI (Controlli di Qualità Interni) prima di iniziare con la routine degli esami da eseguire.

I dati vengono poi raccolti in un grafico di Levy-Jennings, il quale viene interpretato con le regole di Westgard, cioè delle regole di validazione che permettono al tecnico di determinare se uno strumento sta lavorando in linea con gli standard di qualità richiesti.

Allo stesso modo, anche le tecnologie di NGS devono dimostrare di lavorare in modo sicuro, affidabile e riproducibile, ma non sfruttano dei CQI quotidiani. Le prime forme di controllo di qualità sono, come visto in precedenza, il controllo negativo e il controllo noto, che garantiscono rispettivamente l'assenza di contaminazione e l'avvenuta reazione; la loro azione è però limitata alla seduta stessa, per cui si può affermare che i loro valori non possono essere garanzia di un corretto funzionamento dello strumento nel tempo.

Quando si parla di sequenziamento, questi controlli non bastano; infatti, è necessario andare ad istituire tutta una serie di parametri che garantiscano la fiducia del risultato, controllando le varie fasi dell'analisi durante ogni seduta. Queste si chiamano metriche di qualità e variano in base alle capacità e alla performance dello strumento utilizzato [32]. Nel nostro caso, lo strumento è il MiSeqDx e i principali parametri sono:

- Intensità: valuta l'intensità di ogni cluster per il 90° percentile, vagliando quindi la qualità dell'imaging
- Resa prevista: indica il numero di basi che vengono generate in ciascuna corsa, valutabile solo dopo il ciclo 25; solitamente questo è un parametro garantito dal provider, per cui non è motivo di preoccupazione da parte dell'utente
- Error rate: identifica la percentuale di basi chiamate in modo errato in un ciclo, valore che aumenta all'aumentare della lunghezza dei frammenti; per calcolarlo, lo strumento allinea le reads alla libreria di controllo PhiX
- %Q30: questo parametro indica il punteggio qualitativo che viene assegnato dal software per ciascuna read di ciascun ciclo, fungendo quindi da indice di predizione della probabilità di un'errata identificazione della sequenza nucleotidica. Il quality score necessario per lavorare in modo affidabile deve essere pari o superiore a Q30, con il quale si intende una probabilità errata di identificazione di una base di 1 su 1000; il MiSeqDx Illumina garantisce che ogni corsa avvenga sempre con una percentuale di Q30 maggiore del 70-80%.
- Densità dei cluster: misurata in  $K/mm^2$ , indica la densità dei cluster formatisi sulla flow cell, permettendo di verificare se i dati ottenuti sono di bassa qualità
- Cluster Passing Filter (PF): parametro che vaglia la percentuale di cluster che superano il filtro Chastity Illumina, dato dal rapporto fra l'intensità del segnale maggiore e la somma delle due intensità maggiori, che deve essere superiore a 0,6; in caso di valori inferiori, non



si può essere sicuri del fatto che un singolo template abbia generato un singolo cluster con un chiaro segnale, dunque si procede tralasciando le informazioni ottenute da questo.

- Phas/Prephas: il phasing è un passaggio molto importante nel ricostruire le sequenze di DNA a partire dai frammenti della libreria, infatti permette di ottenere in modo chiaro la tipizzazione dell'allele materno e di quello paterno. Rileva che tutti i frammenti della libreria si allineino correttamente rispetto alla sequenza consenso, generando così una lettura univoca e certa dell'allele in esame.

Le metriche di pre-phasing indicano quando le singole basi nucleotidiche vengono a sfasarsi le une con le altre nei vari cicli di lettura, causando delle chiamate errate; ad ogni ciclo, le sequenze si allungano di una base per volta, ma può accadere che alcuni filamenti vadano fuori fase rispetto al ciclo di incorporazione corrente, anticipandone o ritardandone una. Un'accortezza per evitare questa situazione è valutare la composizione di basi del campione in esame: se le quattro basi azotate non sono bilanciate fra di loro, è consigliabile aumentare la quantità della libreria di controllo PhiX, che permette di oltrepassare questo problema.

Un'altra forma di controllo di qualità sono le VEQ (Valutazione Esterna di Qualità), sistema con il quale i laboratori clinici accreditati, compresi i laboratori HLA, confrontano i risultati ottenuti sullo stesso campione e ne traggono informazioni utili specialmente per valutare l'accuratezza e la riproducibilità dei risultati; sono indispensabili per armonizzare le procedure fra i laboratori che utilizzano lo stesso sistema diagnostico, quindi per valutare le performance in conformità ai risultati ottenuti.

### **5.5.2 ALGORITMI MIA FORA**

L'analisi che porta alla identificazione degli alleli con estrema accuratezza e, di conseguenza, alla determinazione dell'aplotipo del soggetto in esame, seppur facilitato dalla presenza di un software fornito, è comunque una procedura da porre attenzione. Per garantire un'identificazione delle basi, e quindi per l'interpretazione dei risultati, il software utilizza tre algoritmi: uno per il mapping, e due per il phasing. L'insieme di questo processo viene detto analisi secondaria.

Il primo step dell'analisi secondaria viene chiamato demultiplex e consiste nell'esame del sample sheet, cioè del file .csv nel quale è stata riportata la lista dei campioni in esame collegati al valore degli indici di legame delle corrispondenti sequenze, confrontando che i dati teorici corrispondano a quelli della libreria voluta. Questa procedura permette la generazione di file intermedi chiamati FASTQ, cioè dei file di testo utilizzati per la rappresentazione delle sequenze con il rispettivo punteggio qualitativo, ovviamente solo in caso di cluster che abbiano oltrepassato il filtro di qualità. I file FASTQ sono l'input fondamentale per la fase di allineamento, procedura tramite cui è possibile

confrontare le reads con le sequenze di riferimento e assegnare loro un punteggio, basandosi sul grado di similarità delle regioni presenti e sulle metriche di qualità del sequenziamento.

Il FASTQ file è composto da quattro righe di testo che riportano rispettivamente le seguenti informazioni (Fig. 20):

- Prima riga: “@” seguita da informazioni identificative della sequenza, della corsa e del cluster
- Seconda riga: sequenze nucleotidiche ottenute dal sequenziamento
- Terza riga: indice di separazione indicato con il simbolo (+)
- Quarta riga: parametri del controllo di qualità



Figura 16. Esempio di FASTQ file.

Il secondo step prevede l’elaborazione delle informazioni di genotipizzazione ricavate dall’analisi secondaria i tre algoritmi che, nell’insieme, forniscono la sequenza nucleotidica corretta. Il primo algoritmo si definisce consenso, poiché si occupa della costruzione dei filamenti corretti grazie al paragone con le cosiddette “sequenze consenso”, cioè le sequenze presenti nei database genetici riportanti tutte le sequenze di basi azotate del genoma conosciute fino a quel momento (Fig. 22-A); nel nostro caso utilizziamo il database IMGT/HLA Database che raggruppa tutte le informazioni riguardanti i geni del sistema HLA. A seguire, il software si occupa della costruzione del contig, cioè una sorta di mappa genetica di basi azotate derivanti dai frammenti di libreria (Fig. 22-B). Accade che i vari frammenti amplificati durante la formazione della libreria presentano delle regioni sovrapponibili alle estremità di modo che, in base all’uguaglianza, si possa ricostruire la sequenza originale per intero. Questa tecnica è stata adottata in quanto è molto più facile e sicuro lavorare con frammenti di DNA di lunghezza contenuta (inferiore a 1000 bp) piuttosto che lavorare con lunghe sequenze che potrebbero portare a risultati falsati a causa di una maggior facilità di errore della DNA polimerasi che lavora per tratti estesi. Il contig, dunque, si genera dalle reads dei vari frammenti, ma funge allo stesso tempo da controllo della corrispondenza con la sequenza dei frammenti; contig e singole letture devono infatti essere identiche per fornire un’interpretazione univoca e certa degli alleli presenti (Fig. 22-C).

L’ultimo algoritmo è il phasing, essenziale per le tecniche di NGS in quanto, combinandosi con la costruzione del contig, permette l’identificazione e la separazione fra due alleli di uno stesso gene, cioè fra i geni di eredità materna e quelli derivanti dal padre (Fig. 22-D).

Nell'esempio riportato nella Figura 21, la presenza di una sola sequenza consenso non permette di distinguere fra i due cromosomi per cui le due basi che differiscono vengono interpretate come mutazione SNP rispetto alla sequenza originale, senza prendere in considerazione l'ipotesi che si possa trattare di due alleli diversi; adottando una tecnica di phasing invece, si riesce a stabilire che le basi A e G non sono delle variazioni della sequenza, ma rappresentano la realtà in quanto appartenenti ai due cromosomi omologhi dell'individuo [33].

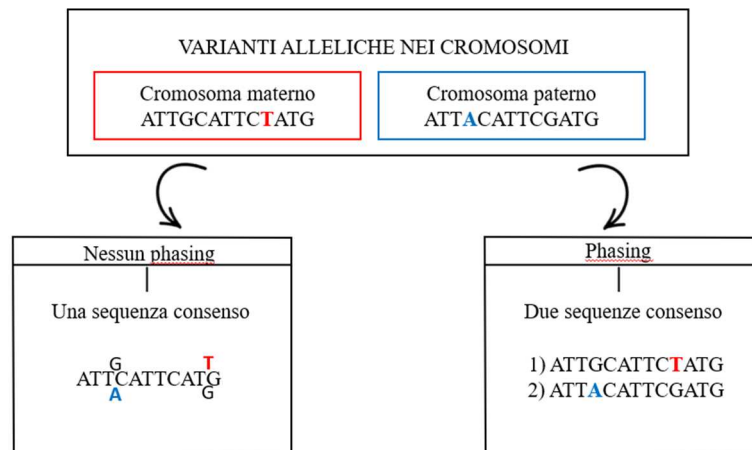


Figura 17. Rappresentazione di come funziona il phasing Illumina.

Esiste anche il phasing indiretto, cioè un algoritmo che non si basa solamente sui dati ricavati dal sequenziamento, ma vaglia database di sequenze geniche. Il suo obiettivo è quello di creare delle statistiche che suggeriscano dei pattern di ereditarietà fra i geni del sistema HLA; esistono infatti dei particolari alleli di determinati loci che tendono a segregare assieme a livello di crossing over meiotico, aumentando quindi la probabilità di essere ereditati assieme. Si parla di linkage disequilibrium, termine ad uso comune nel linguaggio genetico che indica l'associazione statistica fra alleli occupanti due o più loci a presentarsi sempre assieme fra le generazioni, andando così a costituire un particolare aplotipo (Fig. 22-E). Un esempio di linkage disequilibrium lo si trova con il DRB1\*03:01:01, che tende a trascinare con sé il DRB3\*01:01:02 e il DQB1\*02:01:01.

Il phasing viene sfruttato al suo massimo quando si applica allo studio familiare, quindi quando si va ad analizzare la segregazione dei geni HLA nei genitori e nei figli, specialmente se si vuole applicare nella scoperta e comprensione di malattie trasmissibili geneticamente.

In sintesi, la combinazione di sequenza consensus, costruzione del contig e phasing permettono la corretta e univoca identificazione degli alleli costituenti i loci HLA, ma è sempre da tenere in conto che il buon funzionamento di questi algoritmi è imprescindibile da una buona qualità del DNA di partenza e da una buona libreria di sequenziamento.

```

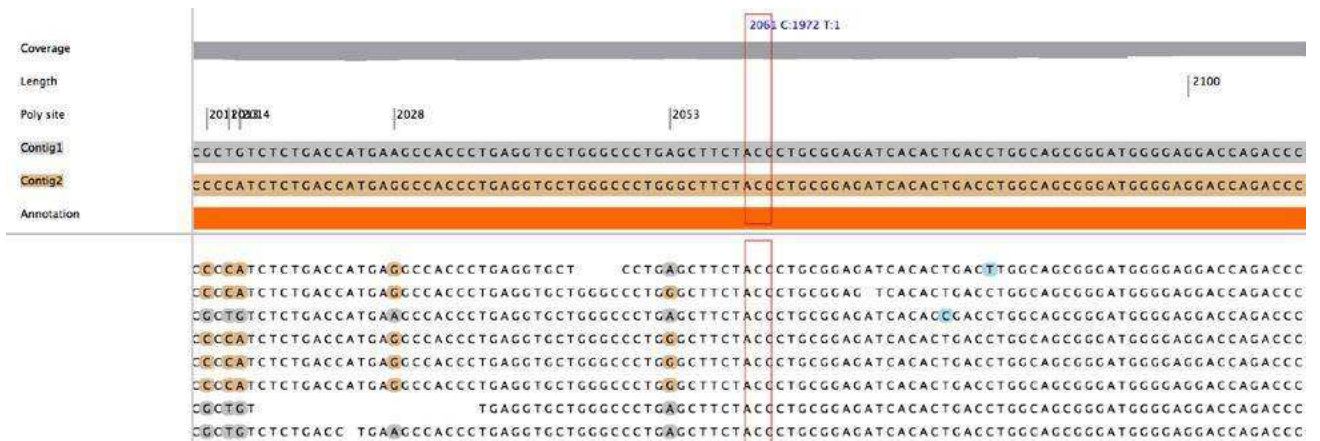
AA pos:          140          150          160
AA:             F L R G Y E Q H A Y D G K D Y I A L N E D L R S W T A A D M A A Q I
A*30:02:01:03 CCTCCGCGGGTAACTAAAGCAAGCCTAAGAAGGCAAGGATTACATCGCCCTGAAAGAGGACCTGGCGCTCTTGGACCGGGGGACATGGCGCTCAGATC
A*68:01:02:02 CCTCCGCGGGTAACTAAAGCAAGCCTAAGAAGGCAAGGATTACATCGCCCTGAAAGAGGACCTGGCGCTCTTGGACCGGGGGACATGGCGCTCAGATC

Nuc pos:        00          520          540          560          580          6
AA pos:          170          180          190          200
AA:             T Q R K W E A A R R A E Q L R A Y L E G T C V E W L R R Y L E N G
A*30:02:01:03 ACCGAGC CAAGTGGGAGGGCGGCCCTGGGGGAGCAGTTGAGAGCCTACCTGGAGGGCACGTGGTGGAGTGGCTCGCAGATACCTGGAGAAAGGGA
A*68:01:02:02 ACCGAGC CAAGTGGGAGGGCGGCCCTGGGGGAGCAGTTGAGAGCCTACCTGGAGGGCACGTGGTGGAGTGGCTCGCAGATACCTGGAGAAAGGGA

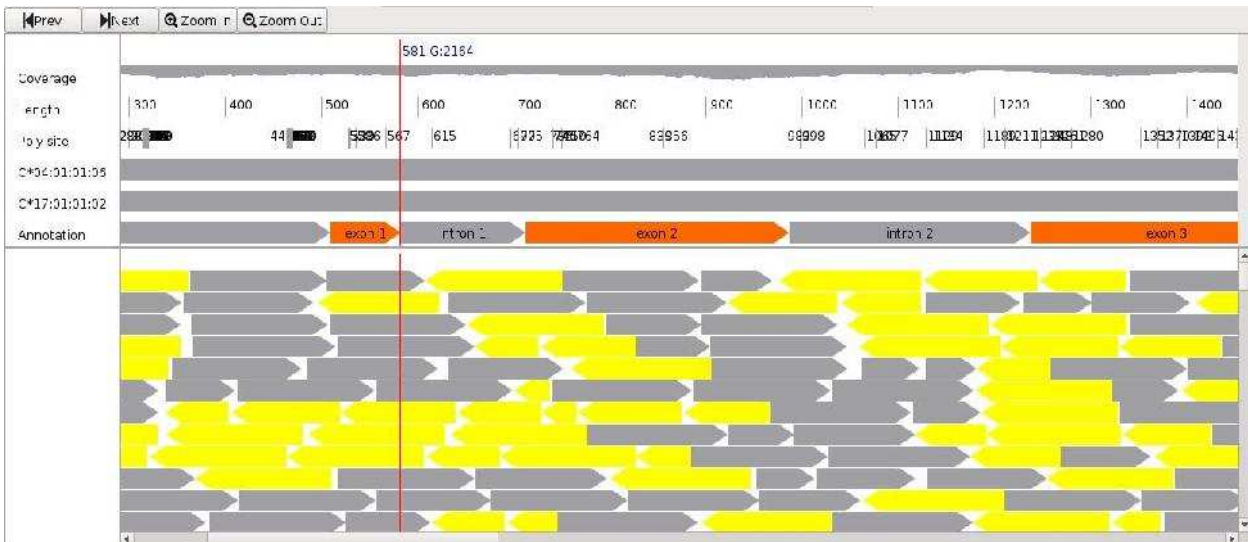
Nuc pos:        00          620          640          660          680          230
AA pos:          210          220          230
AA:             K E T L Q R T D P P K T H M T H H P I S D H E A T L R C W A L G F
A*30:02:01:03 AGGAGAGCTGCAGCGCAAGGACCCCCAAACCATATGACCACCACCTCTCTGACCATGAGCCACCTGAGGTGCTGGCCCTGCTTCT
A*68:01:02:02 AGGAGAGCTGCAGCGCAAGGACCCCCAAACCATATGACCACCACCTCTCTGACCATGAGCCACCTGAGGTGCTGGCCCTGCTTCT

```

A.



B.



C.

Approve		Confirm	
PC	Allele 1	Allele 2	
2	A*02:01:01:01	A*68:01:02:02	
2	B*07:02:01	B*39:06:02	
1	C*07:02:01:01	C*07:02:01:03	
1	DPA1*01:03:01:04	DPA1*01:03:01:05	
2	DPB1*04:02:01:01	DPB1*04:01:01:01	
2	DQA1*01:02:01:03	DQA1*01:01:01:01	
2	DQB1*06:02:01	DQB1*05:01:01:01	
2	DRB1*15:01:01:03	DRB1*01:01:01	
1	DRB5*01:01:01		

D.

Variants	LD Info	Smart Guide
<b>LD Suggestion</b>		
Candidates are highlighted in yellow and lines are ordered based on Chi-Square P-values (lowest[strongest LD] to highest [weakest LD])		
DRB1*03:01:01~DRB3*02:02:01~DQB1*02:01:01		
DRB1*13:02:01~DRB3*03:01:01~DQB1*06:09		
DRB1*03:01~DQA1*05:01~DQB1*02:01		
DRB1*13:02~DQA1*01:02~DQB1*06:09		
DRB1*03:01:01~DRB3*03:01:01~DQB1*02:01:01		
DRB1*03:01:01~DRB3*01:01:02~DQB1*02:01:01		
DRB1*13:02:01~DRB3*03:01:01~DQB1*06:04:01		

E.

Figura 18. Algoritmi Illumina:

- Allineamento al cDNA di riferimento derivante dal database IMGT-HLA
- Browser di allineamento alla sequenza consensus e contig
- Browser di allineamento al contig che mostra le letture mappate sul contig de novo; il grigio rappresenta le letture paired-end, mentre il giallo le non-paired-end.
- Schema riassuntivo della tipizzazione dei due alleli del candidato, distinguibili grazie al phasing.
- Esempio di alleli in linkage disequilibrium fra loro suggeriti dal software



## 6. RISULTATI

A dimostrazione di quanto detto precedentemente sull'NGS, porterò ora due casistiche di applicazione pratica di questa metodica nell'ambito dello studio familiare di pazienti candidati a trapianto di cellule staminali emopoietiche. Come spiegato nella sessione 2.2.2 "Protocollo dell'AULSS3 Serenissima per lo studio familiare", il primo step di questa procedura prevede lo studio della compatibilità fra il paziente e i suoi famigliari con la tecnica di SSO in bassa risoluzione; si ottiene quindi la tipizzazione dei loci HLA-A, -B, -C, -DR e -DQ limitata al primo campo o 2 digits, indicante solamente la famiglia allelica di appartenenza del gene. Utilizzare una tecnica di tipizzazione a bassa risoluzione ha un suo vantaggio: l'SSO ha un TAT di una sola mattinata di lavoro per cui permette di visualizzare i risultati in breve tempo e con costi limitati; così facendo, si escludono subito i famigliari non compatibili con il paziente e, dopo comunicazione al clinico, è possibile proseguire con il test di conferma in NGS ad alta risoluzione fra il paziente e il donatore compatibile prescelto, ottimizzando dunque le risorse del laboratorio.

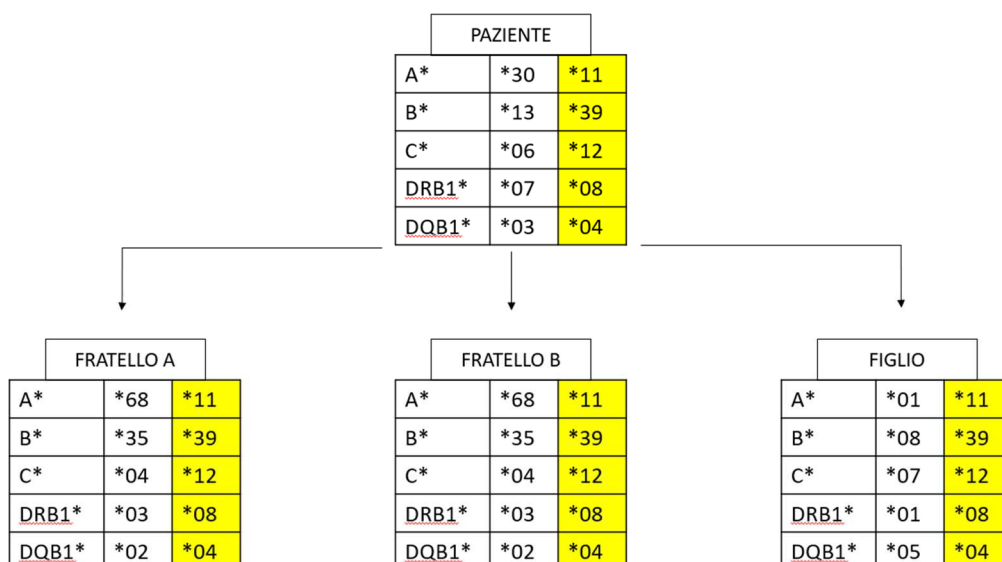
Durante il mio periodo di tirocinio (aprile - settembre 2022), nel Laboratorio di HLA sono stati aperti 16 studi famigliari, 8 dei quali sono arrivati al test di conferma donatore-ricevente in NGS; oltre a questi, sono stati eseguiti dei test di conferma di studi famigliari aperti precedentemente, per un totale di 15 conferme donatore-ricevente. Fra questi ho preso in esame due studi famigliari completi di test di conferma in NGS, il primo svolto con un kit MIA FORA a 11 loci e il secondo con un kit a 6 loci.

### 6.1 STUDIO FAMIGLIARE CON TEST DI CONFERMA IN NGS A 11 LOCI

Per il paziente (P) in esame è stato aperto uno studio familiare in quanto presenta una diagnosi di sindrome mielodisplastica con lesioni di alto grado, cioè una malattia del sangue eterogenea che causa un'incapacità del midollo osseo di rilasciare cellule funzionanti nel circolo sanguigno. In accordo con le linee guida IBMDR [34], sono giunti al laboratorio i campioni del paziente, dei due fratelli (identificati come A e B) e del figlio (F), per verificare la presenza di una compatibilità HLA all'interno dei famigliari stretti. Come primo step, è stata eseguita la tipizzazione a 5 loci in SSO a bassa risoluzione, che ha portato ai risultati riassunti nello Schema 2.

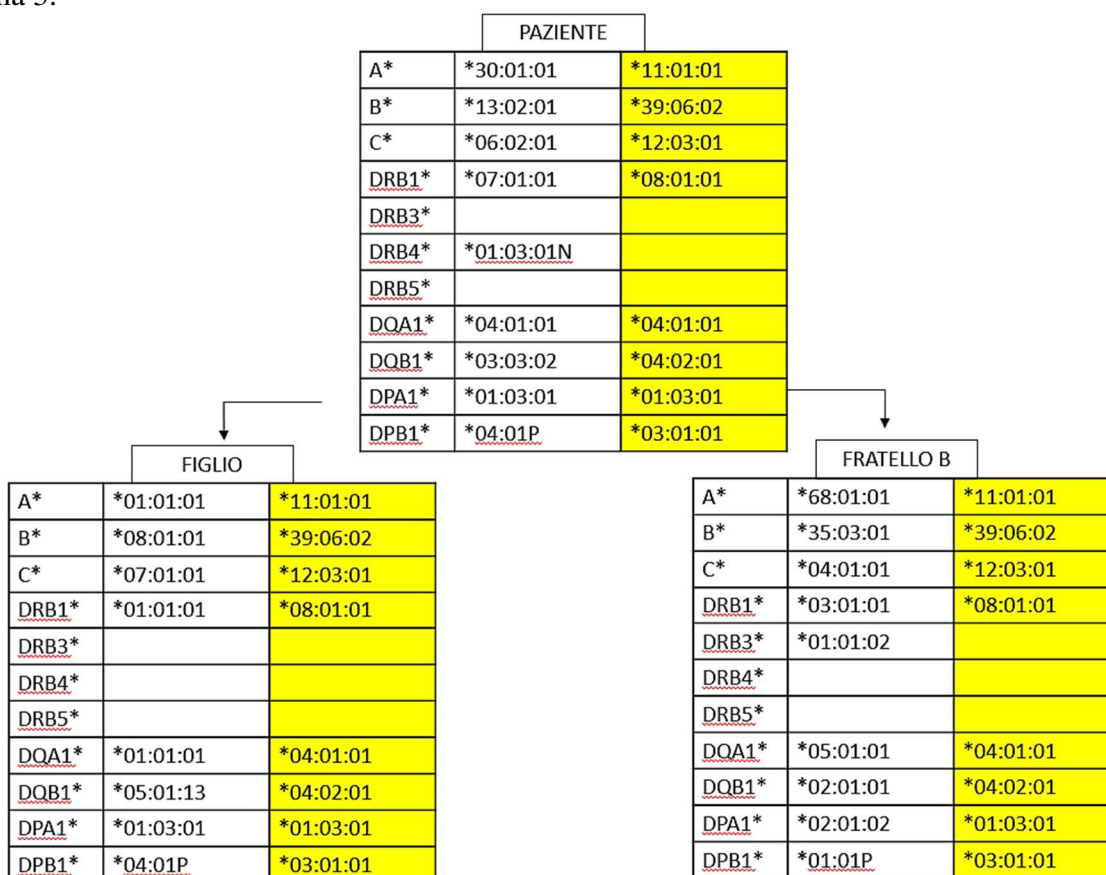
Come si può vedere dalla colonna evidenziata in giallo, il paziente risulta essere aploidentico sia ai fratelli A e B che al figlio; ciò significa che tutti condividono metà dell'aplotipo HLA: i tre fratelli lo hanno ereditato da uno dei due genitori, anche se è impossibile stabilire la corretta segregazione in quanto i genitori non sono risultati disponibili al momento dello studio, mentre il figlio lo ha ereditato dal paziente.





**Schema 2.** Riassunto della tipizzazione del paziente e dei suoi famigliari con la tecnica di SSO in bassa risoluzione.

Successivamente, il clinico ha avviato il secondo step dello studio familiare, inviando al laboratorio un nuovo campione del paziente accompagnato dalla richiesta di test di tipizzazione HLA di conferma donatore-ricevente; assieme alle sue provette, in tempi diversi, sono giunti anche i campioni del figlio e del fratello B. A questo punto, per un maggior approfondimento e per guidare il clinico nella scelta migliore fra i donatori nell'ambito familiare, si è passati all'analisi in NGS estesa ad un numero maggiore di loci, ottenendo dei risultati con risoluzione al terzo campo. I risultati sono riportati nello Schema 3.



**Schema 3.** Riassunto del test di conferma del paziente e dei suoi famigliari aploidentici eseguito in NGS in alta risoluzione.



In questo caso, la richiesta di test di conferma è stata accompagnata da una nota dell'ematologo che, per ragioni cliniche, chiedeva espressamente l'estensione dell'analisi al locus HLA-DRB3, per cui si è utilizzato il protocollo di tipizzazione a 11 loci, che, oltre ai geni analizzati precedentemente, estende la tipizzazione ai loci, sfruttando sempre l'amplificazione clonale dei frammenti di DNA per ottenere risultati certi in un'unica corsa.

Come si vede dallo Schema 3, la risoluzione allelica a 3 campi è mantenuta per tutti i risultati, ad eccezione del locus DPB1, che è accompagnato dalla lettera "P" (P group secondo il database IMGT/HLA), una dicitura peculiare che sta ad indicare che l'allele identificato appartiene ad un gruppo di alleli che codificano per lo stesso ARD, cioè proteine MHC costituite da una tasca di legame composta dagli stessi amminoacidi; gli alleli identificati, quindi, non portano a modificazioni del fenotipo del sito di legame della proteina MHC, ma, a livello di genotipo, si parla di alleli differenti, poiché presentano comunque delle mutazioni legate ad esoni non-ARD o ad introni. Tale dicitura nasce per sopperire alle ambiguità che potrebbero generarsi dalla metodica NGS in loci quali DPB1, DRB1/3/4/5, per i quali manca una copertura dell'intera sequenza del gene nella reazione di PCR (Fig. 6): l'amplificazione e il sequenziamento avvengono correttamente a livello degli esoni informativi, determinando correttamente l'ARD, ma, visto le problematiche di phasing e determinazione della fase gametica, non è possibile determinare in modo assoluto il terzo campo. Nel caso del DPB1\*04:01P del paziente, l'allele DPB1\*939:01 era ugualmente possibile e non escludibile, per cui si è optato per questa scrittura.

Un altro risultato di rilievo è il DQA1 del paziente, che risulta essere in omozigosi, presente quindi in entrambi i cromosomi. Secondo le linee guida, i casi di omozigosi devono essere sempre confermati con una diversa metodica che permetta lo stesso livello di risoluzione per scongiurare casi di "allelic drop out", cioè di mancata amplificazione o mancato riconoscimento di un allele a favore di un altro; in questo caso, il test in SSP ha confermato trattarsi di una vera omozigosi.

Si può ora focalizzare l'attenzione sui loci DRB3/4/5; osservando le tipizzazioni ottenute, si può evidenziare che il paziente possiede il gene DRB4, il fratello il DRB3, mentre il figlio risulta essere negativo per DRB3/4 e 5. Anche in questo caso, visto la specifica richiesta del clinico, si è voluto confermare la tipizzazione di questi loci mutualmente esclusivi con l'SSP ad alta risoluzione, che ha dimostrato l'affidabilità e inequivocità dei risultati NGS.

Il paziente è stato confermato un HLA-DRB4\*01:03:01:02N e HLA-DRB5 assente, come suggerito dall'algoritmo MIA FORA del linkage disequilibrium, in quanto la presenza del gene HLA-DRB1\*07:01:01 fa presumere una forte probabilità di essere HLA-DRB4 positivi. È però anche da evidenziare la lettera "N" che accompagna la nomenclatura del gene, che suggerisce la non traduzione del gene in proteina. Il figlio si è dimostrato essere negativo per HLA-DRB3/4/5, risultato

concordante con la tipizzazione in NGS e con il linkage disequilibrium, poiché la presenza di HLA-DRB1\*01:01:01 indica la molto probabile assenza di HLA-DRB5, mentre HLA-DRB1\*08:01 è indice di un HLA-DRB3 assente. L'affidabilità dell'NGS è avvalorata anche dal riscontro ottenuto sul DRB3 del fratello in SSP, che in entrambi i casi conferma la presenza dell'allele DRB3\*01:01:02:01, in forte linkage disequilibrium con l'allele DRB1\*03:01:01:01, presente nel soggetto [35].

Al termine dell'indagine diagnostica, la refertazione emessa dai dirigenti secondo lo Schema 3 è stata consegnata agli ematologi per la definizione clinica.

Inoltre, possiamo affermare che i tempi di analisi sono stati molto veloci, infatti la refertazione completa dei due test di conferma si è conclusa a meno di 15 giorni dall'arrivo dell'ultimo prelievo dei soggetti dello studio, il tutto grazie ad un flusso di lavoro ottimizzato in tutti i passaggi e ad una metodica di sequenziamento certosina.

## 6.2 STUDIO FAMILIARE CON TEST DI CONFERMA IN NGS A 6 LOCI

Questo studio familiare si è aperto a seguito della diagnosi di neoplasia a cellule dendritiche plasmacitoidi blastiche (BPDCN), malattia maligna molto aggressiva che coinvolge il midollo osseo, pelle, linfonodi e siti extra-nodali, per la quale l'unica cura che permette la sopravvivenza e la remissione è il trapianto di CSE. Anche in questo caso, il primo step della ricerca del donatore compatibile è partito dal familiare di primo grado, la sorella; l'analisi a 5 loci in SSO in bassa risoluzione ha identificato gli aplotipi riassunti nello Schema 4, dal quale si può dedurre che il paziente è verosimilmente aploidentico alla sorella, in quanto non è possibile determinare con certezza come sia avvenuta la segregazione degli alleli fra figli e genitori, a causa dell'assenza di quest'ultimi.

PAZIENTE			SORELLA		
A*	*02	*31	A*	*24	*31
B*	*35	*55	B*	*18	*55
C*	*04	*03	C*	*12	*03
<u>DRB1*</u>	*13	*16	<u>DRB1*</u>	*11	*16
<u>DQB1*</u>	*06	*05	<u>DQB1*</u>	*03	*05

**Schema 4.** Riassunto della tipizzazione del paziente e del suo familiare con la tecnica di SSO in bassa risoluzione.

L'utilizzo di una metodica più sensibile è l'unico modo per determinare la corretta compatibilità fra due individui, anche in assenza dei genitori. A questo punto, si è proceduto al secondo passaggio, con l'arrivo al laboratorio di una nuova richiesta di test di conferma della compatibilità HLA donatore-ricevente.

I risultati del test in NGS sono riassunti nello Schema 5.

PAZIENTE			SORELLA		
A*	*02:01:01	*31:01:02	A*	*24:03:01	*31:01:02
B*	*35:08:01	*55:01:01	B*	*18:01:01	*55:01:01
C*	*04:01:01	*03:04:01	C*	*12:03:01	*03:04:01
<u>DRB1*</u>	*13:02:01	*16:01:01	<u>DRB1*</u>	*11:04:01	*16:01:01
<u>DQB1*</u>	*06:04:01	*05:02:01	<u>DQB1*</u>	*03:01P	*05:02:01
<u>DPB1*</u>	*02:01:02	*14:01P	<u>DPB1*</u>	*04:02P	*14:01P

**Schema 5.** Riassunto del test di conferma del paziente e del suo familiare in NGS in alta risoluzione.

In questa seduta è subito visibile l'alta risoluzione dei risultati, ottenuti con grazie al kit di sequenziamento MIA FORA MFlex6, che esamina i 5 principali loci del sistema HLA già analizzati in precedenza, aggiungendo anche il DPB1, allele che gioca un ruolo importante riguardo la permissività del trapianto [12].

Il phasing opera la distinzione fra l'aplotipo del cromosoma di origine materna e quello di origine paterna, attribuendo le giuste sequenze ai due e discriminando fra mutazioni in cis (sullo stesso cromosoma) o in trans (su due cromosomi diversi). Questo algoritmo, unito alla risoluzione a 6 digits, permette di confermare che fra i due familiari non sono presenti variazioni genetiche, per cui è possibile affermare che il paziente è aploidentico alla sorella senza alcuna ambiguità.

Al termine dell'indagine diagnostica, la refertazione riportata nello Schema 5 è stata consegnata agli ematologi per la definizione clinica.

Anche per questo test di conferma è stato possibile operare un sequenziamento massiccio di più campioni in un'unica corsa, portando ad ottenere dei risultati inequivocabili con un costo di analisi contenuto e con tempi rapidi di refertazione: il referto è infatti stato chiuso in meno di 10 giorni dalla ricezione del campione del paziente e del familiare, assicurando così un più rapido intervento per la cura della grave condizione che ha colpito il soggetto.



## 7. DISCUSSIONE

L'obiettivo del presente elaborato è dimostrare come il sequenziamento di seconda generazione possa essere introdotto all'interno della routine di un laboratorio ospedaliero; in particolar modo, grazie alla mia esperienza di tirocinio svolta presso il Laboratorio di HLA dell'ospedale di Mestre, ho potuto accertare come la standardizzazione di una metodica NGS in un laboratorio di istocompatibilità Hub provinciale migliori il flusso di lavoro e l'affidabilità dei risultati. Questa metodica offre infatti la possibilità di un sequenziamento massiccio di un numero elevato di campioni in un'unica seduta, con alti livelli di standardizzazione del lavoro. I tempi e, di conseguenza, i costi di lavoro sono contenuti se paragonati ad altre tecniche di analisi, quali SSP, SSO e SBT, grazie alla possibilità di indicizzazione degli ampliconi generati dall'amplificazione clonale dei frammenti di DNA target di partenza. A sostegno di ciò, vi è l'esperienza di NGS svolta nel laboratorio di Mestre, che ha ottimizzato il protocollo di tipizzazione, suddividendo in tre giornate lavorative la seduta, durante la quale vengono sequenziati 6 o 11 loci HLA di 24 campioni (22 campioni, NTC e controllo noto); al termine, si ottengono dei risultati affidabili e certi, che non necessitano di indagini aggiuntive per il chiarimento di ambiguità sui dati rilevati. Algoritmi quali la sequenza consenso, il contig e il phasing permettono l'identificazione di polimorfismi presenti in regioni esoniche, esoniche non-ARD e introniche, che aumentano il livello di risoluzione dell'analisi, arrivando ai 6-8 digits, e stabiliscono la fase cromosomica (cis/trans) degli alleli HLA, riconoscendo eventuali nuovi alleli non ancora definiti, grazie al continuo confronto con il database internazionale IMGT/HLA.

Dal punto di vista clinico, l'NGS rappresenta un'enorme vantaggio per pazienti in attesa di trapianto di CSE in quanto è possibile velocizzare il processo che porta all'identificazione di un donatore sano di CSE compatibile: il test di conferma di compatibilità giunge all'ematologo dopo pochi giorni dalla ricezione dei campioni da testare da parte del laboratorio; in seguito, in base al risultato, egli può prontamente decidere se andare a trapianto con un donatore familiare oppure se aprire una ricerca MUD da registro, con la speranza di trovare un candidato migliore per il paziente. La stessa ricerca MUD è agevolata dall'introduzione dell'NGS, poiché gli stessi donatori volontari vengono tipizzati direttamente in alta risoluzione, velocizzando il processo di matching e conferma. Il tutto si traduce con una maggior percentuale di successo nella cura dei pazienti ematologici, soggetti per i quali l'accelerazione dei tempi di intervento e l'incremento della qualità delle indagini sono imprescindibili per il buon esito della terapia e per annullare i rischi di insorgenza di eventi avversi, migliorando e salvando così quante più vite possibili.









## BIBLIOGRAFIA E SITOGRAFIA

- [1] J. L. S. E. T. W. C. Cornaby, «Next-generation sequencing and clinical histocompatibility testing,» *Elsevier*, vol. Human Immunology, n. 82, pp. 829-837, 2021.
- [2] Anthony Nolan Research Institute , «HLA nomenclature,» Giugno 2022. [Online]. Available:  
<http://hla.alleles.org/alleles/index.html#:~:text=There%20are%20currently%2035%2C077%20HLA,IPD%2DIMGT%2FHLA%20Database.>
- [3] J. D. H. T. R. T. e. a. N. P. Mayor, «Recipients Receiving Better HLA-Matched Hematopoietic Cell Transplantation Grafts, Uncovered by a Novel HLA Typing Method, Have Superior Survival: A Retrospective Study,» *Journal of the American Society for Blood and Marrow Transplantation* , vol. Biology of Blood and Marrow Transplantation, n. 25, pp. 443-450, Marzo 2019.
- [4] O. V. R. Y. A. E. O. S. Glotov, «Comparative analysis of NGS and Sanger sequencing methods for HLA typing at a Russian University Clinic,» *Cellular Therapy and Transplantation*, pp. 72-82, 2018.
- [5] A. K. A. J. C. A. Vinay Kumar, Robbins Basic Pathology, Elsevier Saunders, IX edizione, 2013.
- [6] S. Boegel, HLA Typing. Methods and protocols, Humana Press, 2018.
- [7] F. Fantolini, «Correlazione e suscettibilità alle malattie degli alleli del sistema HLA,» Università degli studi di Pisa , 2013/2014.
- [8] M. Hertl, «Trapianto di cellule staminali ematopoietiche,» giugno 2020. [Online]. Available: [https://www.msmanuals.com/it-it/professionale/immunologia-malattie-allergiche/trapianto/trapianto-di-cuore.](https://www.msmanuals.com/it-it/professionale/immunologia-malattie-allergiche/trapianto/trapianto-di-cuore)
- [9] P. C. M. M. G. S. M. S. E. Z. S. Adamo, Istologia per i corsi di laurea in professioni sanitarie, Piccin, V ristampa, 2017.
- [10] P. Marson, «Raccolta di cellule staminali emopoietiche da sangue periferico,» 2021.
- [11] M. d. Salute, «Trapianti - Sito ufficiale dei Centro Nazionale Trapianti,» [Online]. Available:  
[https://www.trapianti.salute.gov.it/trapianti/dettaglioContenutiCnt.jsp?lingua=italiano&area=cnt&menu=cittadini&id=252.](https://www.trapianti.salute.gov.it/trapianti/dettaglioContenutiCnt.jsp?lingua=italiano&area=cnt&menu=cittadini&id=252)

- [12] A. (. I. d. I. e. B. d. Trapianti), *Raccomandazione AIBT per a valutazione della istocompatibilità nel trapianto di cellule staminali emopoietiche*, Versione 1.0 del 2 dicembre 2021.
- [13] M. d. S. C. N. S. e. C. N. T. Collaborazione fra ISS, *Linee di indirizzo CNT/CNS in tema di requisiti minimi organizzativi, strutturali e tecnologici del Programma Trapianto (PT) di CSE e delle Unità ad esso afferenti*, Prot. 155/CNT 2021 e Prot. 0138/CNS 2021.
- [14] O. d. ULSS3, *Protocollo per lo studio familiare*.
- [15] «Nuove tecnologie al servizio dei trapianti,» *Galliera News 1*, 2015.
- [16] M. C. e. A. Moschetti, «Il ruolo del laboratorio di Immunogenetica - DM 2/2022,» *Anaa.it*, 2022.
- [17] O. d. ULSS3, *Tipizzazione HLA in NGS*, Seconda versione in vigore dal 29 giugno 2021.
- [18] Immucor, «MIA FORA NGS MFlex HLA Typing Kit».
- [19] Immucor, «MIA FORA NGS Mflex - Guida per l'utente - Protocollo manuale».
- [20] Insiel spa, «EMONET, l'informazione in rete per il buon uso del sangue,» Aprile 2004. [Online]. Available: [http://archive.forumpa.it/archivio/1000/1700/1790/1792/finsiel\\_Emonet.pdf](http://archive.forumpa.it/archivio/1000/1700/1790/1792/finsiel_Emonet.pdf).
- [21] Ente Ospedaliero Ospedali Galliera, «IBMDR - Registro Italiano Donatori di Midollo Osseo,» [Online]. Available: <https://www.ibmdr.galliera.it/ibmdr>.
- [22] O. d. ULSS3, «Estrazione automatizzata di DNA con strumento QiaSymphony,» Prima versione.
- [23] QIAGEN, «QIAxpert System,» [Online]. Available: <https://www.qiagen.com/us/products/instruments-and-automation/quality-control-fragment-analysis/qiaxpert-system/>.
- [24] «Molecular,» [Online]. Available: <https://www.molecularlab.it/principi/calcoli-laboratorio/letture-spettrofotometro.asp>.
- [25] Sage Science , «Pippin Prep - Sistema di DNA size selection,» Manuale d'uso.
- [26] ThermoFisher Scientific , «Qubit Fluorometric Quantification,» [Online]. Available: [https://www.thermofisher.com/it/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit.html?ef\\_id=Cj0KCQjwntCVBhDdARIsAMEwACkqWc1nEE36-](https://www.thermofisher.com/it/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit.html?ef_id=Cj0KCQjwntCVBhDdARIsAMEwACkqWc1nEE36-)

jKPw2Mz9HdZYzWzb2ET1\_xOT0z0T8Xt3mxP0gB0BLQaApnPEALw\_wcB:G:s&s\_kwc  
id=AL!365.

- [27] K. W. M. K. R. K. Ravi, «MiSeq: a Next Generation Sequencing Platform for genomic Analysis,» *Methods in Molecular Biology* , pp. 223-232, 2018.
- [28] Immucor, «MiSeq - Guida del sistema».
- [29] Illumina , «Sequencing By Synthesis: Patterned Flow Cell Technology,» [Online]. Available: <https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/patterned-flow-cells.html>.
- [30] Immucor, «Guida per l'utente del software v.4.5 - Linux Server».
- [31] Illumina, «Sequencing Analysis Viewer Software - User Guide,» 2014.
- [32] «Our top 5 quality control (QC) metrics every NGS user should know,» [Online]. Available: <https://blog.horizondiscovery.com/diagnostic/the-5-ngs-qc-metrics-you-should-know>.
- [33] Illumina , «What is phasing?,» [Online]. Available: [https://www.illumina.com/company/video-hub/15NPZCGP\\_e4.html](https://www.illumina.com/company/video-hub/15NPZCGP_e4.html).
- [34] IBMDR, *Standard di funzionamento del programma nazionale italiano di donazione di cellule staminali emopoietiche da non familiare*, Versione XXIV, Agosto 2021.
- [35] S. D. M. B. A. Sanchez-Mazas, «A linkage disequilibrium map of the MHC region based on the analysis of 14 loci haplotypes in 50 French families,» *European Journal of Human Genetics*, vol. 8, pp. 33-41, 2000.
- [36] «National Library of Medicine - National Center for Biotechnology Information,» Luglio 2022. [Online]. Available: <https://www.ncbi.nlm.nih.gov/gene/3105>.
- [37] «GeneCards - The human gene database,» [Online]. Available: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=HLA-DRA#summaries>.
- [38] C. C. Lionetti Elena, «Co-localization of gluten consumption and HLA-DQ2 and -DQ8 genotypes, a clue to the history of celiac disease,» *Editrice Gastroenterologica Italiana* , pp. PubMed, PMID: 25200477, 2014.



## ICONOGRAFIA

- Figura 1-2-3. Robbins e Cotran, Volume 1, “Le basi patologiche delle malattie – Patologia generale”, VIII Edizione, Elsevir, edizione italiana a cura di Vincenzo Eusebi
- Figura 4. 2° Convention Galliera – 2016 “La fabbrica delle idee... realizzabili”, NGS: immunogenetica e trapianti, Michela Mazzocco, S.C. Laboratorio di istocompatibilità e IBMDR
- Figura 5.  
<https://www.trapianti.salute.gov.it/trapianti/dettaglioContenutiCnt.jsp?lingua=italiano&area=cnt&menu=cittadini&id=252>
- Figura 6-7. <https://immdocs.immucor.com/en-us/Products/Pages/MIA-FORA-NGS.aspx>
- Figura 8. <https://www.jantdx.com/product/qiasymphony-sp/>
- Figura 9. <https://www.qiagen.com/media/product-tools/qiexpert-web/index.html>
- Figura 10. <https://www.immucor.com/it/product/mia-fora-ngs-mfex-hla-typing-kits/>
- Figura 11. Esempio di foglio di calcolo utilizzato nel laboratorio HLA dell’Ospedale di Mestre
- Figura 12. Schema riassuntivo tempo/temperatura della reazione di long PCR.
- Figura 13. <https://www.labtech.com/product/pippin-prep>
- Figura 14. "Pippin Prep - Sistema DNA Size Selection", Manuale d'uso, Sage Science
- Figura 15. <https://www.thermofisher.com/order/catalog/product/cn/en/Q33216>
- Figura 16. <https://www.medicalexpo.it/prod/illumina-inc/product-83632-604889.html>
- Figura 17. <https://verogen.com/products/miseq-fgx-sequencing-reagents/>
- Figura 18. "MiSeq - Guida del sistema", Immucor
- Figura 19. <https://www.illumina.com/systems/sequencing-platforms/miseq/products-services.html>
- Figura 20. <https://slidetodoc.com/introduction-to-next-generation-sequencing-ngs-data-analysis/>
- Figura 21. Immagine ispirata a [https://www.illumina.com/company/video-hub/15NPZCGP\\_e4.html](https://www.illumina.com/company/video-hub/15NPZCGP_e4.html)
- Figura 22. "Guida per l'utente del software v.4.5 - Linux Server", MIA FORA™ NGS, Immucor



## **ALLEGATO: APPROFONDIMENTO SUI GENI HLA DI I E II CLASSE**

Negli ospedali di riferimento provinciale (HUB) sono presenti i laboratori di "HLA" o di "immunogenetica", ai quali il paziente si rivolge quando il medico di base sospetta particolari condizioni che possono essere condizionate dalla presenza di particolari alleli del complesso MHC. In questi contesti, oltre al sequenziamento ampiamente esplorato lungo questo elaborato, si eseguono anche altre indagini di tipizzazione HLA, come quelle necessarie per determinare possibili associazioni con malattie come celiachia, varie forme di artrite e diabete di tipo I oppure quelle associate agli eventi avversi che si verificano a seguito dell'impiego di alcuni farmaci a bersaglio molecolare. Anche il laboratorio HLA di Mestre non si focalizza solamente sull'NGS e sulla tipizzazione di pazienti ematologici e donatori di CSE, ma offre tutta una serie di altre prestazioni di routine a servizio di pazienti ospedalieri o di esterni; ad esempio, fra le analisi svolte, troviamo il test di biologia molecolare per la predisposizione della celiachia, che evidenzia la presenza dell'aplotipo DQ2 o DQ8. Oltre a ciò, si utilizzano le tecniche di SSO e SSP per la tipizzazione HLA dei soggetti richiedenti, in quanto il loro medico sospetta una particolare malattia o un'alterata risposta terapeutica; si è infatti visto che determinati alleli del sistema HLA sono indice di predisposizione a particolari malattie o alterazioni cliniche, per cui la loro diagnosi può essere di aiuto al medico curante.

In seguito, sono riportati alcuni esempi di cui sono venuta a conoscenza durante il mio tirocinio, suddividendoli fra HLA di prima classe e seconda classe. I tre geni costituenti MHC I sono:

- HLA-A: questo gene si trova in 6p22.1 [36] ed è costituito da 8 esoni codificanti per le diverse regioni della catena pesante  $\alpha$ ; sono stati descritti più di 6000 alleli di questo gene, con diversa funzionalità e attività associata, come nel caso dell'infezione da virus dell'HIV, in cui troviamo che HLA-A\*30 è predittore di una maggior capacità di ridurre la carica virale da parte dell'organismo, mentre HLA-A\*02, quando associato con HLA-B\*45, è indice di una carica virale di molto maggiore;
- HLA-B: localizzato in 6p21.33, formato anch'esso da 8 esoni e assieme a HLA-A e HLA-DR gioca un ruolo di fondamentale importanza nel rigetto dei trapianti, per cui è necessario avere un match fra donatore e ricevente quanto più simile e preciso. Sono note centinaia di forme alleliche differenti, fra cui sono da menzionare:
  - HLA-B\*27, associato ad un aumentato rischio di sviluppare spondilite anchilosante, spondiloartropatie, artrite reattiva, malattie infiammatorie croniche dell'intestino e psoriasi;

- HLA-B\*57 è un altro gene molto richiesto dai clinici al laboratorio di HLA in quanto è responsabile dell'inibizione della risposta citotossica delle cellule natural killer e di ipersensibilità al trattamento del virus dell'HIV con Abacavir;
- HLA-B\*51, gene indice di una maggior predisposizione allo sviluppo della malattia di Behçet, che causa infiammazione della parete dei vasi sanguigni, oltre che ad altre malattie.
- HLA-C: localizzato in 6p21.33, costituito anch'esso da 8 esoni; anche di questo gene esistono vari alleli che mostrano generalmente un linkage disequilibrium con alleli appartenenti ad HLA-B, creando così degli aplotipi comuni etnia-correlati come Cw4-B35 (sierotipo Cw4 che comprende gli alleli HLA-C\*0401, \*0407, e \*0410) nelle popolazioni dell'Africa occidentale o nei nativi americani, oppure Cw7-B7 (sierotipo Cw7 con gli alleli HLA-C\*07:01-\*07:06, \*07:12, \*07:14, \*07:16) nell'Eurasia occidentale e nel Sud Africa.

Per quanto riguarda invece i tre geni che caratterizzano il sistema HLA di seconda classe, abbiamo:

- HLA-DR: il suo locus è 6p21.32 ed è costituito da cinque geni diversi che sono, come visto precedentemente HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4 e HLA-DRB5. Il gene A [37] codifica per la catena  $\alpha$ , è composto da cinque esoni e non è polimorfico, mentre i quattro geni B codificano per la catena  $\beta$ , sono composti da 6 esoni e sono altamente polimorfici, tanto che sono conosciuti centinaia di alleli, la cui presenza è associabile a particolari condizioni o malattie, come il DRB1\*13:02 che è associato a persistenza del virus dell'epatite B acuta e cronica; inoltre, presso i geni B sono presenti molti pseudogeni, cioè sequenze nucleotidiche simili a quelle dei geni ma che non portano alla codifica di proteine, evento attribuibile ad un'ampia gamma di mutazioni che hanno fatto perdere la funzionalità del gene, mantenendone comunque la sequenza all'interno del genoma.
- HLA-DQ: il suo locus è 6p21.32, composto da due geni, DQA1 (sei esoni) e DQB1 (sei esoni), entrambi altamente polimorfici, che codificano rispettivamente per le due catene. Sono dei geni che presentano frequentemente isoforme, cioè degli SNP (Single Nucleotide Variant) o delle forme di splicing alternativo, che possono o meno far variare la funzionalità della proteina prodotta, arrivando anche a legare dei peptidi differenti, per cui sono da osservare attentamente per l'assegnazione di un sierotipo ad un paziente. Questo gene gioca spesso un importante ruolo per quanto riguarda lo sviluppo di malattie autoimmuni, in particolar modo la malattia celiaca e il diabete di tipo 1. L'importante per queste condizioni non è possedere un singolo allele preciso, ma un aplotipo, cioè una specifica combinazione di varianti alleliche lungo una sequenza genica (o lungo un cromosoma) che tendono ad essere ereditate assieme: gli aplotipi HLA che espongono a maggior suscettibilità verso la



celiachia sono l'aplotipo DQ2, che indica la presenza contemporanea della famiglia allelica HLA-DQA1\*05 e HLA-DQB1\*02, e l'aplotipo DQ8, che indica la combinazione HLA-DQA1\*0301 e HLA-DQB1\*0302, con un rischio di malattia aumentato in caso di omozigosi. Questi aplotipi sono in forte linkage disequilibrium con alleli di HLA-DR: in particolare, DQ2 lo è con DR3 e DR5/7, mentre DQ8 con DR4, utili quindi in caso sia necessaria un'analisi genetica di II livello per la conferma di predisposizione alla malattia celiaca [38].

- HLA-DP: è situato nel locus 6p21.32 ed è costituito da due geni HLA-DPA1 e HLA-DPB1 che, con i loro polimorfismi, permettono di creare fino a quattro differenti molecole specifiche per peptidi. Molto importante è il gene DPB1, che gioca un ruolo fondamentale nel trapianto di midollo osseo: in particolar modo, è da evidenziare se i mismatch che si vengono a riscontrare fra donatore e ricevente siano o meno permissivi, cioè se abbiano un'alta o una minima probabilità di sviluppare un'allorreattività post-trapianto da parte dei linfociti T, che poi sfocia in GvHD o rigetto del trapianto.

