



**UNIVERSITÀ POLITECNICA DELLE MARCHE
DIPARTIMENTO SCIENZE DELLA VITA E DELL'AMBIENTE**

Corso di Laurea Magistrale

BIOLOGIA MOLECOLARE E APPLICATA

**POTENZIALITA' E LIMITI DEI MICROAPLOTIPI DEL DNA
NELL'ANALISI DI RELAZIONI PARENTALI**

**PROS AND CONS OF MICROHAPLOTYPE MARKERS IN KINSHIP
ANALYSIS**

Tesi di Laurea Magistrale
di:

Moreno Brenda Denise

Relatore
Chia.ma Prof.ssa

Chiara Turchi

**Sessione Luglio 2024
Anno Accademico 2023-2024**

INDICE	
<i>Capitolo 1</i>	4
INTRODUZIONE	4
1.1 La genetica forense	4
1.2 Polimorfismi del DNA	5
1.2.1 Polimorfismi di sequenza: SNP	7
1.3 Microaplotipi	10
1.3.1 Nomenclatura dei microaplotipi	12
1.3.2 Criteri di selezione dei microaplotipi	15
1.4 Vantaggi e svantaggi dei microaplotipi	17
1.5 Applicazioni dei microaplotipi	19
Microaplotipi nell'identificazione personale	20
Microaplotipi nella deconvoluzione di campioni misti	21
Microaplotipi nelle relazioni parentali	22
Microaplotipi nell'ascendenza biogeografica	24
<i>Capitolo 2</i>	26
<i>SCOPO DELLA TESI</i>	26
<i>Capitolo 3</i>	28
MATERIALI E METODI	28
3.1 Descrizione del pannello sviluppato	28
3.2 Campioni analizzati per lo studio popolazionistico	32
3.3 Campioni analizzati per le relazioni parentali	32
3.4 Estrazione del DNA	35
3.4.1 Estrazione DNA da tamponi buccali	36
3.4.2. Estrazione DNA da resti scheletrici	37
3.4.3 Quantizzazione del DNA estratto	39
3.4.3.1 Quantizzazione estratto con E-Gel™ Agarose Gels with SYBR Safe™	39
3.4.3.2 Quantizzazione estratto con Qubit™ 1X dsDNA High Sensitivity (HS)	40

3.5 Preparazione manuale delle librerie	41
3.5.1 Amplificazione delle regioni target.....	41
3.5.2 Parziale digestione degli ampliconi	43
3.5.3 Ligazione parziale dei barcode.....	43
3.5.4 Purificazione delle librerie	45
3.5.5 Eluizione delle librerie	47
3.5.6 Quantizzazione delle librerie.....	47
3.6 Diluizione, raggruppamento delle librerie e posizionamento sullo ION CHEF	49
3.7 Preparazione del template sullo HID ION CHEF	49
3.8 Sequenziamento con HID Ion GeneStudio S5 System	53
3.9 Analisi dei dati di sequenziamento	54
3.10 Software per il calcolo biostatistico delle relazioni parentali	56
3.10.1 Familias (v.3.3.1).....	56
<i>Capitolo 4</i>	58
RISULTATI	58
4.1 Risultati quantizzazione estratto	58
4.2 Risultati quantizzazione librerie	59
4.3 Analisi risultati del sequenziamento	60
4.6 Analisi con Familias	67
<i>Capitolo 5</i>	76
DISCUSSIONI.....	76
<i>Capitolo 6</i>	83
CONCLUSIONI.....	83
BIBLIOGRAFIA.....	85

INTRODUZIONE

1.1 La genetica forense

La genetica forense è una branca della genetica, che permette di risalire all'identità di un individuo a partire da diversi materiali di partenza, che possono essere: sangue, capelli, resti scheletrici, sperma o altri liquidi biologici. Questa disciplina impiega marcatori genetici del DNA per poter compiere una serie di applicazioni quali:

- Identificazione personale
- Indagini di parentela o test di paternità
- Analisi di cadaveri non identificati

Per far ciò, vengono prese in considerazione regioni del DNA definite polimorfiche, ossia differenti tra un individuo e l'altro e questo è necessario per poter discriminare soggetti da un punto di vista genetico: differenze che si osservano nel fenotipo sono causate da differenze nel genotipo. A queste si sommano interferenze ambientali o eventi stocastici durante il processamento dei campioni.

L'origine dell'identificazione mediante analisi del DNA viene attribuita ad Alec Jeffreys, un genetista britannico che ha messo a punto il fingerprinting genetico, metodo che ha rivoluzionato le tecniche di identificazione. In particolare, è stato sfruttato per la prima volta per la risoluzione di un omicidio negli anni 80'.

La genetica forense, quindi, combina principi di genetica e tecniche innovative di biologia molecolare, basandosi su variabilità e probabilità, per raggiungere diversi scopi.

1.2 Polimorfismi del DNA

In genetica forense, è molto importante avere dei marcatori del DNA caratterizzati da un'elevata variabilità o marcatori meno polimorfici ma che, combinati insieme, ci rendano possibile una discriminazione di individui diversi ⁽¹⁾. Analizzando il genoma, si stima che la maggior parte della sequenza di bp del nostro DNA (oltre il 99,7%) sia la stessa tra le persone. Soltanto una piccola frazione del DNA (circa lo 0,3% o 10 milioni di nucleotidi) ci rende individui unici e quindi ci permette di discriminarci gli uni dagli altri. ⁽²⁾

In genetica si parla di polimorfismo quando nella popolazione esistono almeno due forme alleliche e l'allele più raro è presente con una frequenza uguale o

superiore all'1%; con il termine variante, invece, si indica un allele con frequenza al di sotto dell'1%. Considerando che le frequenze alleliche spesso variano tra le popolazioni, una variante per una popolazione potrebbe essere un polimorfismo per un'altra. ⁽¹⁾

Esistono due tipologie di polimorfismi che si verificano a livello della sequenza del DNA: polimorfismi di lunghezza (STR, short tandem repeat) e polimorfismi di sequenza (SNP, single nucleotide polymorphisms).

Polimorfismi di lunghezza: consistono in cambiamenti nel numero di sequenze di DNA ripetute disposte in tandem. Questi cambiamenti sono classificati in minisatelliti, noti anche come variable number of tandem repeat (VNTR), microsatelliti e satelliti, in base alla taglia delle loro unità ripetitive.

I microsatelliti, noti anche con il nome di STRs (short tandem repeats), sono i marcatori genetici più conosciuti e studiati in ambito forense. La caratteristica fondamentale che li rende i candidati migliori per l'analisi forense è data dalla ridotta lunghezza della sequenza di DNA che va a formare l'unità ripetuta in tandem, definita unità ripetitiva, che varia da 2 a 6 paia di basi. Questi polimorfismi danno origine a diverse forme multi-alleleliche, come conseguenza della lunghezza variabile della sequenza che è data dal numero di ripetizioni. ⁽¹⁾

In tutti i casi, si presume che questi polimorfismi derivino da scambi ineguali mitotici/meiotici o da slittamenti del DNA durante la replicazione.

Polimorfismi di sequenza: conosciuti come SNP, o single nucleotide polymorphism, consistono nella variazione di una singola base lungo la sequenza del DNA.

Sono modificazioni stabili, ereditarie e bi-alleleliche, ossia si distinguono due varianti alleliche che segregano nella popolazione. Dalla combinazione di queste forme alleliche sono possibili solo 3 genotipi. A questa categoria di modifiche, si aggiunge anche l'inserimento e/o la delezione di una singola base (InDel), che provoca quindi una variazione nella lunghezza della sequenza del DNA ^(4,5)

1.2.1 Polimorfismi di sequenza: SNP

I polimorfismi a singolo nucleotide rappresentano la forma più comune di variazione genetica in cui la mutazione avviene a livello di una singola base nucleotidica. Questa modifica è la conseguenza di un'errata incorporazione di nucleotidi durante il processo replicativo del DNA oppure una mutagenesi causata da una modificazione chimica delle basi nucleotidiche. ⁽¹⁾

L'attenzione nei confronti degli SNP nasce dal loro possibile uso come marcatori molecolari per poter identificare i geni coinvolti nelle malattie geniche. A tal proposito, il Progetto HapMap è nato con l'obiettivo di determinare le variazioni più comuni del genoma umano, andando a caratterizzare le varianti genetiche, le loro frequenze e le possibili correlazioni tra di esse. È stato stimato che, all'interno della popolazione mondiale umana, circa 10 milioni di siti variano in modo che si possano osservare entrambi gli alleli con una frequenza dell'1%.⁽⁶⁾

Considerando la probabilità di trovare una posizione nucleotidica in eterozigosi confrontando due cromosomi presi a caso (media della diversità nucleotidica), ci si aspetta di trovare in media 1 SNP ogni 1.331 bp circa. Sono già stati identificati 1,42 milioni di questi polimorfismi, ma se si considera la frequenza minima dell'1% per allele meno frequente, una stima di SNP nel genoma umano si spinge oltre 11 milioni di polimorfismi a singolo nucleotide.⁽¹⁾

Sachidanandam et al.⁽⁵⁾ hanno descritto una mappa di 1,42 milioni di SNP e hanno evidenziato che questi polimorfismi a singolo nucleotide hanno una densità relativamente costante tra i cromosomi autosomici. Analizzando la densità degli SNP in finestre cromosomiche definite (considerate come coordinate cromosomiche), si evince che la maggior parte del genoma umano possiede SNP ad alta densità.

I polimorfismi a singolo nucleotide sono dotati di una serie di caratteristiche che li rendono ideali come marcatori genetici a scopi forensi. Come prima cosa, presentano tassi di mutazione più bassi rispetto agli STR, si parla di tassi circa 100.000 volte inferiori (10^{-8} contro 10^{-3}), per cui essendo più stabili in termini di ereditarietà sono utili nell'analisi delle relazioni parentali. A tal proposito, lo studio di Amorim e Pereira ⁽⁹⁾ ha evidenziato che l'utilizzo esclusivamente di SNP con un potere informativo paragonabile a quello degli STR fornisce risultati statistici inconclusivi in analisi parentali di routine.

In secondo luogo, gli SNP sono particolarmente adatti per l'analisi di campioni di DNA altamente degradati in quanto vengono sfruttate le piccole dimensioni degli ampliconi. In terzo luogo, essendo marcatori capaci di originare solo due forme alleliche sono dotati di uno scarso potere informativo, svantaggio significativo nel caso di deconvoluzione di miscele. Diversi studi hanno dimostrato che per poter eguagliare il potere discriminatorio di 10-15 STR si stima che siano necessari circa 50 SNP polimorfici. ⁽⁸⁾

Il confronto diretto tra SNP e STR suggerisce che per ora gli SNP non sono pronti a sostituire le STR nell'identificazione del DNA forense, nonostante la loro semplicità e la loro abbondanza nel genoma umano. Grazie a tutte queste caratteristiche positive e negative, gli SNP possono ricoprire un ruolo utile in alcune applicazioni specifiche come i test sul DNA mitocondriale (mtDNA),

gli Y-SNP per scoprire discendenze dirette dagli antenati paterni, i marcatori informativi dell'ascendenza (AIM), la previsione di tratti fenotipici, ma anche la presenza di due o più SNP collegati insieme a formare i microaplotipi. ⁽²⁾

Characteristics	Short Tandem Repeats (STRs)	Single Nucleotide Polymorphisms (SNPs)
Occurrence in human genome	~1 in every 15kb	~1 in every 1 kb
General informativeness	High	Low; only 20% to 30% as informative as STRs
Mutation rate	~1 in 1000	~1 in 100,000,000
Marker type	Di-, tri-, tetra-, and pentanucleotide repeat markers with many alleles	Mostly biallelic markers with six possibilities: A/G, C/T, A/T, C/G, T/G, A/C
Number of alleles per marker	Usually 5 to 20	Typically 2 (some triallelic SNPs exist)
Detection methods	Gel/capillary electrophoresis	Sequence analysis; microchip hybridization
Multiplex capability	> 10 markers with multiple fluorescent dyes	Difficult to amplify more than 50 SNPs (detection of 1000s with microchips)
Amplicon size	~75 to 400bp	Can be less than 100 bp
Ability to predict ethnicity (biogeographical ancestry)	Limited	Some SNPs associated with ethnicity
Phenotypic information	No	Possible to predict some hair colors, etc.
Major advantage for forensic application	Many alleles enabling higher success rates for detecting and deciphering mixtures	PCR products can be made small, potentially enabling higher success rates with degraded DNA samples

Note: SNPs are more common in the human genome than STRs but are not as polymorphic.

Fig.1. STR vs SNP: caratteristiche e applicazioni a confronto⁽²⁾

1.3 Microaplotipi

L'introduzione del concetto di microaplotipo nelle analisi forensi risale al 2013, grazie al prof. Kidd, il quale prese in considerazione la possibilità di utilizzare i microaplotipi come marcatori genetici in diverse applicazioni forensi. ⁽¹⁰⁾

Il termine microaplotipo (microhaps o Mh) fa riferimento ad una breve regione di DNA, di circa 200-300bp, caratterizzata dalla presenza di due o più SNP

strettamente collegati tra di loro che segregano insieme presentano tre o più combinazioni alleliche e considerando che si tratta di multi-SNP, il potere informativo che si ottiene nel complesso, per ogni singolo locus, è molto più ampio rispetto ad uno SNP se considerato singolarmente.

L'interesse nei confronti di questi nuovi marcatori genetici è dato dalla presenza di un potere discriminatorio paragonabile a quello degli STR, che rappresentano il "gold standard" per l'identificazione personale. Al tempo stesso, la breve dimensione dei microaplotipi li rende adatti per ricostruire profili genetici a partire da campioni di DNA degradato. Infine, la breve distanza che contraddistingue i singoli SNP implica un tasso di ricombinazione estremamente basso, per cui possono essere sfruttati al meglio nella determinazione delle parentele e nella deconvoluzione delle tracce miste.

Per analizzare i microaplotipi, la tecnologia migliore è rappresentata dalla Massive Parallel Sequencing (Fig.2) che rispetto ad altre tecnologie come Sanger, Real Time PCR, permette di avere informazioni sulla fase cis/trans dei singoli SNP che compongono il microaplotipo. In questo modo, si può risalire agli alleli che compongono i microhaps grazie ad una singola lettura della sequenza. senza dover ricorrere all'utilizzo di software, come PHASE, per compiere una ricostruzione statistica degli aplotipi in modo da dedurre i modelli di microaplotipi osservati nelle popolazioni. ⁽¹¹⁾

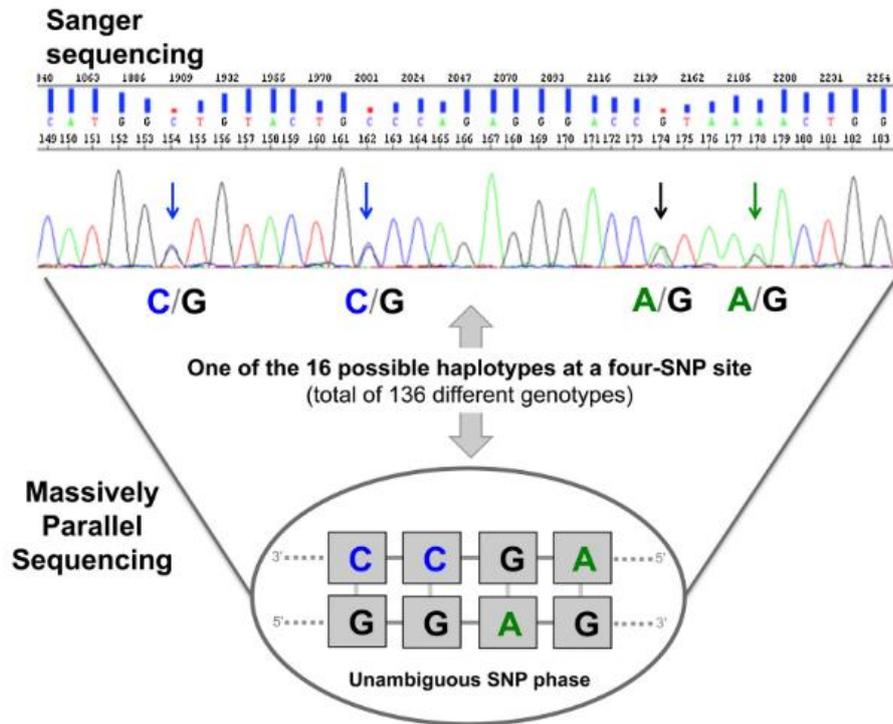


Fig.2 Tecniche di analisi dei microplotipi⁽¹¹⁾

1.3.1 Nomenclatura dei microaplotipi

Per praticità durante la manipolazione dei microaplotipi, all'inizio degli anni 80' la Gene Mapping Community ha istituito un primo sistema di denominazione per poter distinguere i diversi polimorfismi del DNA. In particolare, si inizia dalla lettera "D" che sta per DNA, seguita dal numero del cromosoma, poi dalla lettera "S" che sta per sito e infine un numero assegnato per il sito polimorfico. Ad esempio, la sigla del polimorfismo D3S1358 sta a indicare il 1358° sito che è stata catalogato sul cromosoma 3.

Per permettere una facile consultazione di tutti i MH identificati nei diversi laboratori, Kidd et al. ⁽¹²⁾ ha proposto una nomenclatura seguendo le linee guida dell'HUGO Gene Nomenclature Committee (HGNC), per ottenere una denominazione efficace e informativa. In particolare, viene proposto che il nome dei microaplotipi sia organizzato come segue (Fig.)

- Una radice rappresentata dalla sigla “mh” o “MH”,
- Due cifre che indicano il numero del cromosoma,
- Un simbolo di due lettere che identifichi il laboratorio che ha pubblicato per la prima volta il microaplotipo in questione,
- Uno specifico numero univoco associato al laboratorio ⁽¹²⁾

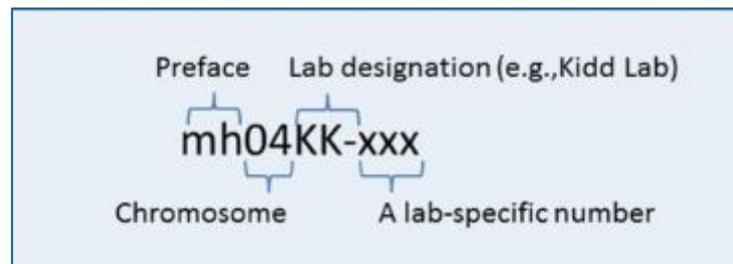


Fig.3 Nomenclatura proposta da Kidd

Successivamente, Zhu et al. ⁽¹³⁾, ha preso in considerazione il fatto che gli SNP che compongono ogni microaplotipo, possono essere modificati con studi successivi per cui si potrebbero aggiungere nuovi SNP oppure quelli già preesistenti potrebbero non essere più considerati. Per tale motivo, è stata messa a punto una nomenclatura, che a differenza di quella proposta da Kidd,

tenesse in considerazione i singoli SNP che vanno a costituire i microhaps. Il nome è composto da 5 parti:

- Una radice data dalla sigla “mh” o “MH”,
- Il numero del cromosoma,
- Un codice univoco stabilito dai primi autori che hanno studiato il microaplotipo,
- Un numero specifico del laboratorio,
- Un numero che definisce i diversi cluster di SNP ⁽¹³⁾

(N.B: le sigle mh e zl vengono proposte da Kidd per separare i diversi numeri nella denominazione).

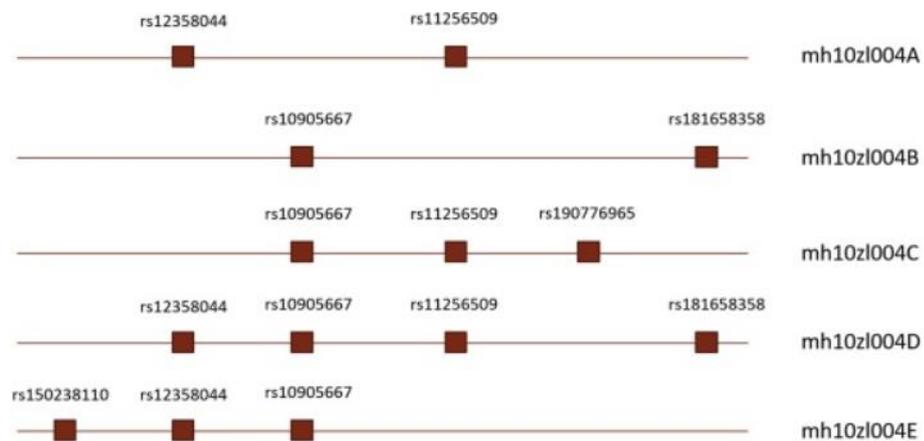


Fig.4 Esempio di versioni di SNP che costituiscono il microaplotipo nel chr.10⁽¹³⁾

La Fig 4 permette di vedere nel dettaglio come vengono considerati i diversi SNP in questo tipo di nomenclatura. Le lettere A, B, C, D ed E indicano, infatti,

le diverse versioni di SNP presenti nel cromosoma 10 che formano il microplotipo, caratterizzato dalla condivisione di SNP o dalla loro localizzazione ravvicinata.

1.3.2 Criteri di selezione dei microplotipi

Al fine di compiere una scelta ponderata sui microplotipi ottimali per i diversi scopi forensi, si possono considerare due parametri statistici: Effective Number of Alleles (A_e) e l'Informativeness (In). La prima variabile viene indicata come “il numero di alleli neutri uniformemente frequenti, basati sulla dinamica di popolazione, che sono in grado di produrre la stessa eterozigosi del locus, con differenti alleli a frequenze diverse”. Ciò si riferisce al livello di variazione di una determinata popolazione, alla misura dell'eterozigotità di un microplotipo e alla capacità di distinguere gli individui, ossia la random match probability (RPM).

Viene calcolata come il reciproco dell'omozigosi, data dalla seguente formula:

$$1 / \sum p_i^2$$

dove p_i è uguale alla frequenza dell'allele i e la somma è su tutti gli alleli del locus.

Maggiore è il valore dell' A_e , maggiore sarà la capacità di un locus del MH di rilevare e distinguere i campioni che compongono una traccia mista, avendo a disposizione gli alleli più ugualmente frequenti. ^(14,15)

I microhaps meglio studiati sono caratterizzati da un A_e che va da circa 3 fino a 10; Kidd et al ⁽¹⁵⁾ ha evidenziato che un piccolo gruppo di 24 MH con valori di A_e compresi tra 5 e 10 è più informativo rispetto ai 24 STR più utilizzati in applicazioni forensi.

Il secondo parametro è un indicatore della variazione delle frequenze alleliche osservate tra le popolazioni che vengono analizzate; ciò vuol dire che un MH caratterizzato da valori elevati di I_n è in grado di distinguere alcune delle popolazioni in studio per una corretta assegnazione dell'ascendenza biogeografica di un individuo.

Un microaplotipi ottimale, quindi, dovrebbe avere valori elevati di A_e ed I_n , ad indicare che ognuna delle popolazioni considerate presenta un gran numero di alleli che variano in frequenza tra le diverse popolazioni. ⁽¹¹⁾

I primi dati che riguardano le frequenze alleliche di popolazione dei microaplotipi sono stati annotati nel database ALFRED (ALLele FREquency) che nasce come un database di varianti e polimorfismi, ma che contiene anche una buona percentuale di microaplotipi, anche se ad oggi non viene più

aggiornato. Recentemente è stato introdotto MicroHapDB ⁽²⁸⁾ (<https://github.com/bioforensics/MicroHapDB>), un ampio database delle varianti microaplotipiche umane. Il database integra informazioni sui microaplotipi e dati di frequenze nelle diverse popolazioni che provengono da numerosi studi pubblicati. Nel database sono altresì presenti i dati di Ae e delle frequenze alleliche di 26 popolazioni, consentendo quindi ai ricercatori di selezionare i marcatori più appropriati per le specifiche applicazioni.

1.4 Vantaggi e svantaggi dei microaplotipi

Nella comunità forense, gli STR vengono considerati il gold standard per l'identificazione personale, ma in alcune applicazioni, come la deconvoluzione di tracce miste e l'analisi di campioni di DNA degradato, presentano delle limitazioni.

Mettendoli a confronti con i microaplotipi, si possono estrapolare tre importanti vantaggi:

- 1- Rispetto alle STR, l'assenza di un unità ripetuta in tandem evita lo slittamento della polimerasi durante l'amplificazione; di conseguenza non vengono generati gli artefatti noti come stutter, che potrebbero complicare l'analisi di tracce miste nel momento in cui il picco stutter

ha un'altezza paragonabile all'allele del contribuente minore della traccia;

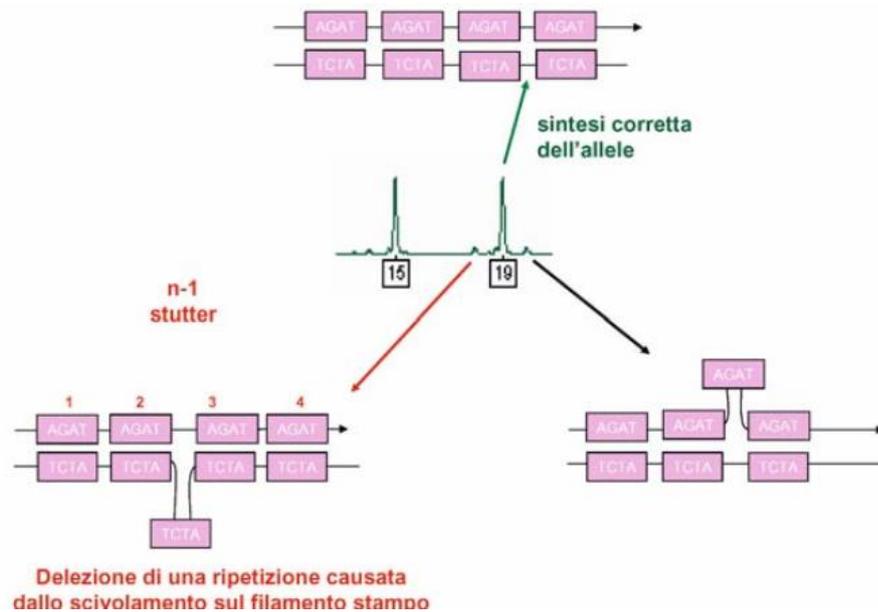


Fig. 5 Meccanismo di formazione delle stutter ^[1]

- 2- Il tasso di mutazione dei microaplotipi è molto più basso rispetto a quello delle STR di circa 5-6 ordini di grandezza e ciò risulta importante nell'analisi delle relazioni parentali, dove una variazione minore degli alleli permette di compiere un'interpretazione adeguata dei risultati;
- 3- Gli ampliconi che si generano dai diversi alleli Mh sono tutti della stessa lunghezza, cosa diversa nelle STR dove gli alleli possono variare in lunghezza. Questo risulta essere problematico perché in campioni degradati si può verificare un'amplificazione preferenziale

dell'allele più corto rispetto agli altri e di conseguenza si attribuisce in maniera errata il profilo genetico.

In sintesi, il basso tasso di mutazione, l'assenza di stutter e la lunghezza più corta degli ampliconi sono tutte caratteristiche che rendono i microaplotipi dei potenziali marcatori per essere applicati in diversi scopi forensi, tra cui anche previsione dell'ascendenza geografica e la previsione delle caratteristiche visibili esterne. ^(17, 24)

1.5 Applicazioni dei microaplotipi

I microaplotipi rappresentano un nuovo strumento versatile che può essere utilizzato per estrarre simultaneamente molteplici e rilevanti informazioni forensi da uno o più individui che contribuiscono a un campione di DNA.

L'avvento della Massive Parallel Sequencing (o MPS) ha reso possibile la genotipizzazione di tutti i loci forensi in un unico esperimento permettendo il sequenziamento efficiente dell'intero amplicone. Sfruttando questa tecnologia, i set di MH possono essere ottimizzati e rendersi utili per applicazioni forensi quali: identificazione individuale, relazioni tra individui, informazioni sull'origine biogeografica, deconvoluzione di campioni misti, identificazione di persone scomparse e identificazione DNA non umano. ⁽¹¹⁾

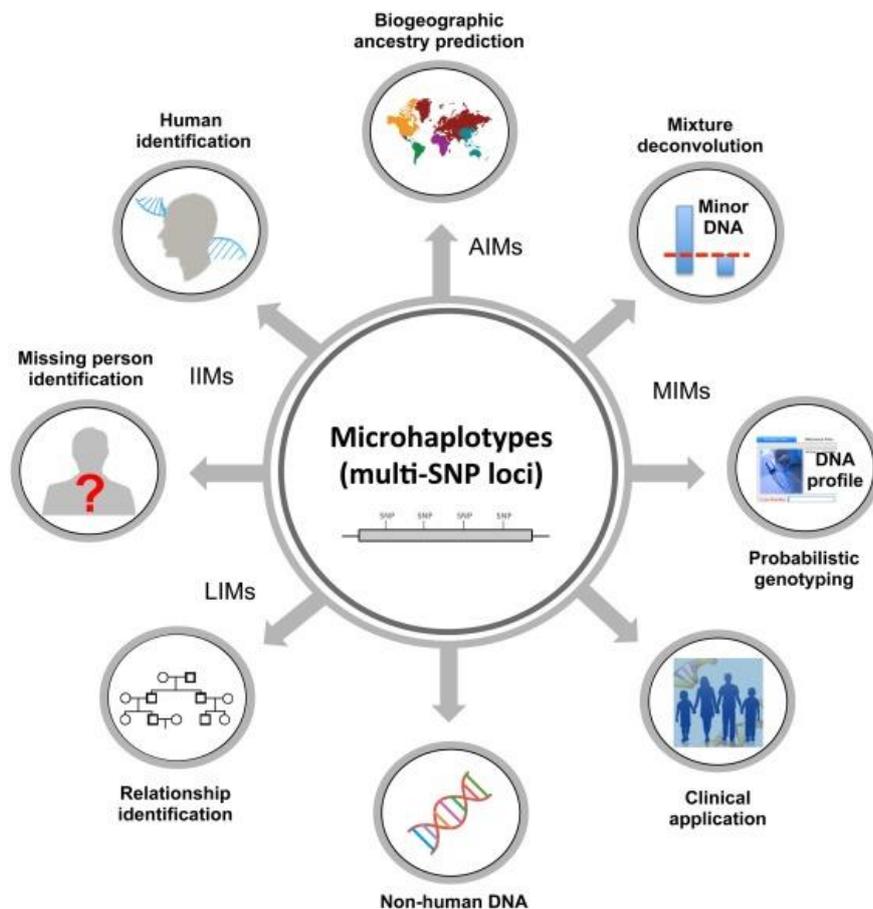


Fig. 6 Possibili applicazioni dei microaplotipi in campo forense⁽¹¹⁾

Microaplotipi nell'identificazione personale

Essendo costituiti dalla presenza di più SNP, i microaplotipi sono dotati di caratteristiche ottimali per l'identificazione individuale,

A tal proposito, si sfruttano i microaplotipi per determinare se un profilo in esame corrisponde al profilo di un sospettato; si fa riferimento alla random match probability (RMP), rappresenta la probabilità che due individui presi casualmente nella popolazione condividano lo stesso profilo genetico.

Di conseguenza, più è alto il valore di questa probabilità, più è probabile che la corrispondenza evidenziata non sia casuale. Nel suo studio, Kidd ⁽¹⁵⁾ ha selezionato 24 MH con un valore Ae più alto tra i 90 MH di un pannello che aveva presentato in precedenza ⁽²⁹⁾ per confrontarli con i 24 loci STR più utilizzati. Questo piccolo set di microplotipi ha mostrato dei valori di RPM più piccoli rispetto agli STR usati abitualmente; si è dedotto che alti livelli di individualizzazione calcolati mediante la RPM derivano da alti valori di Ae.

Microplotipi nella deconvoluzione di campioni misti

Per testare la capacità dei microplotipi di risalire ai diversi contributori che costituiscono una traccia mista, il gruppo di lavoro di Oldoni ha sviluppato un pannello di 74 microplotipi basato sul sistema Ion Torrent S5 partendo dai 130 microplotipi caratterizzati dal laboratorio di Kidd. ⁽³⁰⁾ Questo pannello si è dimostrato efficiente per l'analisi di miscele caratterizzate da diversi rapporti di contributori e quantità di DNA; è stato messo a confronto con gli STR ed è risultato uno strumento che può andare ad integrare e migliorare la deconvoluzione delle miscele mediante STR.

Risultati confermati anche da uno studio effettuato considerando solo 36 MH per l'analisi di miscele ⁽¹⁸⁾, partendo sempre dai 130 MH indicati da Kidd.

Si evince quindi che l'analisi complessiva dei microplotipi mediante MPS mostra una sensibilità maggiore rispetto a quella che si osserva con gli STR.

Un vantaggio importante nell'utilizzo del sequenziamento massivo è la possibilità di sequenziare contemporaneamente più loci. Questo aumento di informatività permette una maggiore copertura di tutti gli alleli di una miscela, in modo da poter distinguere gli alleli dei diversi contributori.

Microplotipi nelle relazioni parentali

L'identificazione delle relazioni parentali ha come obiettivo la ricostruzione della relazione biologiche all'interno delle famiglie, ma anche di rintracciare persone scomparse mediante l'analisi di parentela. I marcatori abitualmente utilizzati per queste applicazioni sono gli STR, ma un vantaggio importante nell'utilizzo dei microplotipi è un basso tasso di ricombinazione e il fatto che ogni microplotipo viene trasmesso in blocco di generazione in generazione.

In diversi studi è stata testata la capacità dei microplotipi nella ricostruzione delle relazioni parentali. Wu et al. ⁽¹⁹⁾ ha dimostrato l'efficienza di un pannello di 54 MH altamente polimorfici, studiando due famiglie complesse e ottenendo un'elevata affidabilità dei risultati. Allo stesso modo, Tomas, insieme al suo gruppo di lavoro, ha testato un set di 74 microplotipi definiti in due popolazioni, i danesi e i groenlandesi, andando a definire diversi gradi di

parentela. I risultati ottenuti sfruttando questo pannello di MH ha superato quello ottenuti con set di STR o SNP attualmente utilizzati. ⁽²⁴⁾

Nel loro studio, Turchi et al ⁽³¹⁾, hanno valutato la performance di due pannelli, uno da 87 MH e uno da 29MH, in diversi scenari complessi di relazioni parentali mediante il software Familias ⁽²⁵⁾ Questi vengono messi a confronto con gli STR inclusi nel PowerPlex Fusion 6C System (Promega), considerando rapporti di mezzi fratelli, fratelli pieni e cugini di primo grado. Viene considerato il valore del LR nelle diverse relazioni considerando tali relazioni come reali o prendendo i soggetti come non imparentati: dall'analisi delle simulazioni effettuate, è stato dimostrato che il pannello da 87 MH risulta essere più informativo nella risoluzione delle relazioni di mezzi fratelli e fratelli pieni rispetto agli STR.

Questo dimostra che i microplotipi possono essere più utilizzabili nella ricostruzione delle relazioni parentali.

A conferma di questo risultato, è presente anche il lavoro di De la Puente ⁽²²⁾, il quale ha deciso di compiere delle simulazioni di differenti pedigree analizzando il valore del LR ricavato dal software Familias ⁽²⁵⁾. In particolare, hanno dimostrato che il pannello da 118 MH analizzato permette di ottenere una probabilità di parentela sufficientemente elevata per quanto riguarda fratelli e fratellastri, osservando che non c'è una sovrapposizione della distribuzione del

rapporto di verosomiglianza in questi casi, rispetto ai cugini di primo e di secondo grado.

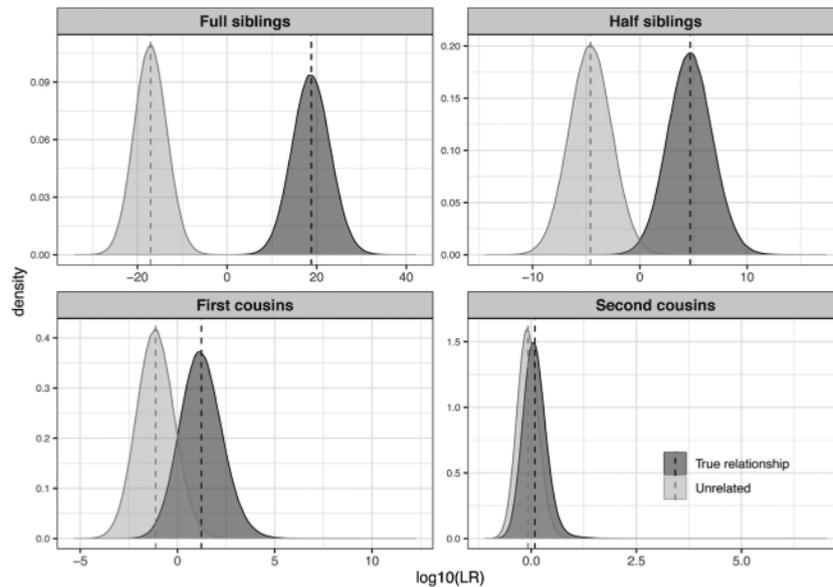


Fig.7 Distribuzione del valore di LR in tre scenari parentali, distinguendo soggetti imparentati da quelli non imparentati

Questi studi dimostrano che i microplotipi rappresentano dei marcatori preziosi per la verifica delle relazioni parentali.

Microplotipi nell'ascendenza biogeografica

A partire dal 2006 molti pannelli di SNP sono stati utilizzati con lo scopo di risalire all'ascendenza degli individui, evidenziando le popolazioni che possono essere raggruppate insieme. A tal proposito, sono stati pubblicati diversi pannelli di polimorfismi a singolo nucleotide informativi

dell'ascendenza, indicati come *AISNP*, che permettono di distinguere le origini biogeografiche degli individui.

Con l'emergere dei microaplotipi, ci si è focalizzati su questi nuovi marcatori considerando che il potere informativo di più SNP collegati tra loro è maggiore di un singolo SNP di per sé. ⁽¹⁵⁾

Il parametro statistico più importante per questo tipo di applicazione è l'informativeness (*In*), che misura la variazione della frequenza allelica nelle diverse popolazioni; ciò indica che loci con un valore elevato di *In* presentano una maggiore variazione nella popolazione. Sono stati condotti diversi studi sull'ascendenza biogeografica, che hanno dimostrato un ottimo potere dei microhaps nel discriminare le differenti popolazioni. ⁽²⁰⁾

Capitolo 2

SCOPO DELLA TESI

I microaplotipi rappresentano una classe di marcatori genetici emergenti nel campo della genetica forense. Consistono in brevi regioni del DNA caratterizzate da uno o più SNP che possono essere facilmente sequenziate mediante le nuove tecnologie di sequenziamento (Next Generation Sequencing). Nell'identificazione delle relazioni parentali i microaplotipi hanno attirato l'attenzione grazie ai diversi vantaggi che mostrano rispetto agli STR in questo tipo di analisi, tra cui il tasso di mutazione più basso, l'assenza di artefatti stutter.

Lo scopo principale di questo studio è quello di esplorare l'applicabilità dei microaplotipi in analisi parentali. Per perseguire questo obiettivo è stato progettato un pannello da 77MH con cui sono stati tipizzati soggetti italiani, al fine di stimare le frequenze aplotipiche nella popolazione italiana, e famiglie con relazioni parentali note, per verificare se la relazione è supportata con l'analisi di questi marcatori.

Per questo tipo di analisi è possibile utilizzare il software Familias ⁽²⁵⁾ il quale permette di definire le relazioni parentali restituendo un valore di LR, ossia un rapporto tra le probabilità di due ipotesi mutualmente esclusive che vengono considerate. In particolare, in questa applicazione viene definito come *paternity index* (PI).

MATERIALI E METODI

3.1 Descrizione del pannello sviluppato

I microaplotipi che sono stati utilizzati per la costruzione del pannello da 77MH sono stati scelti dalla letteratura, considerando i dati ottenuti da diversi studi: 23 MH sono stati selezionati dal progetto Global Ae Average 1KG, 48 MH dal lavoro di De la Puente ⁽²²⁾ e 6 MH dal pannello di 89MH sviluppato nel 2017 da Turchi ⁽³²⁾.

La selezione dei microaplotipi viene svolta seguendo 3 criteri:

- Microaplotipi composti da 2 a 6 SNP,
- Microaplotipi caratterizzati da un valore di $A_e \geq 3$,
- Microaplotipi amplificabili con taglia < 140 bp

Per tre dei sei microaplotipi selezionati dallo studio (Turchi, 2019) è stato necessario eliminare uno dei due SNP esterni così da ridurre la grandezza dell'amplicone. Nella tabella sottostante sono riportate le informazioni relative ai sei microaplotipi di interesse.

Microhaps pannello 89					Microhaps pannello 77			SNP eliminato
Chr	Microhap	SNP	Amplicon size	Ae	SNP	Amplicon size	Ae	
2	mh02KK-134	4	374	6.818	3	39	6,129	ultimo SNP: rs72623112
	mh02KK-136	3	344	4.538	3	70	4,538	
4	mh04KK-010	2	374	3.117	2	34	3,117	
5	mh05KK-022	2	374	2.988	2	31	2,988	
	mh13KK-218	4	298	8.344	3	65	5,089	ultimo SNP: rs9536430
	mh21KK-324	4	374	5.949	3	66	4,081	primo SNP: rs6518223

Tab.1 Microaplotipi modificati nel pannello da 77 MH

Il pannello finale è stato disegnato con *Ion AmpliSeq Designer Tool* (TFS), definendo la grandezza dell'amplicone in un range di 125-140bp.

Va specificato che la dimensione dell'amplicone è una stima a partire da ciò che viene fornito dal tool, ossia l'insert, che rappresenta la regione che viene effettivamente sequenziata senza considerare i primer. A questo dato vengono aggiunte 40 bp ipotizzando che ciascun primer sia lungo 20 bp.

Nella tabella sottostante vengono riportati tutti i microaplotipi che compongono questo pannello, indicando il nome relativo, la posizione cromosomica e la dimensione dell'amplicone stimata.

Nome	Posizione cromosomica	Dimensione amplicone (bp)
FNDC10	chr1:1529937-1530018 --GRCh37	121
mh01KK-160	chr1:153170258-153170323 --GRCh37	105
mh01usc-1qB	chr1:187622410-187622491 --GRCh37	121
mh02KK-140	chr2:4822383-4822455 --GRCh37	128
mh02KK-141	chr2:41188288-41188373 --GRCh37	123
D2S259	chr2:76368295-76368370 --GRCh37	102
mh02USC-2pC	chr2:79252938-79253030 --GRCh37	126
mh02USC-2qA	chr2:103092496-103092584 --GRCh37	132
mh02KK-134*	chr2:161079365-161079459 --GRCh37	120
mh02USC-2qD	chr2:206822599-206822677 --GRCh37	124
mh02KK-136	chr2:228092383-228092465 --GRCh37	124
D3S2426	chr3:1717833-1717911 --GRCh37	128
mh03USC-3pA	chr3:5534414-5534492 --GRCh37	132
mh03KK-217	chr3:10823050-10823139 --GRCh37	132
ZNF385D	chr3:22057765-22057856 --GRCh37	127
3pB	chr3:42924607-42924695 --GRCh37	126
LINC02000	chr3:133806138-133806228 --GRCh37	120
LOC100505920	chr3:193946498-193946582 --GRCh37	122
mh04KK-010	chr4:1986908-1986996 --GRCh37	123
SORCS2	chr4:7499107-7499207 --GRCh37	134
mh04USC-4pA	chr4:8122448-8122551 --GRCh37	132
4pB	chr4:36639739-36639813 --GRCh37	112
mh04USC-4qA	chr4:59436032-59436105 --GRCh37	125
mh04USC-4qC	chr4:155794578-155794665 --GRCh37	115
mh04KK-217	chr4:179715535-179715603 --GRCh37	132
mh04KK-218	chr4:182795865-182795945 --GRCh37	128
D4S2390	chr4:189794162-189794256 --GRCh37	134
D4S3214	chr4:189891668-189891754 --GRCh37	118
mh05KK-125	chr5:3082770-3082849 --GRCh37	122
mh05KK-022	chr5:9619877-9619963 --GRCh37	130
mh0USC-5qB	chr5:106366209-106366296 --GRCh37	122
5qC	chr5:132916318-132916416 --GRCh37	141
mh05USC-5qD	chr5:178267354-178267452 --GRCh37	119

mh06USC-6pA	chr6:23765315-23765400 --GRCh37	130
mh06KK-103	chr6:31145388-31145478 --GRCh37	118
D6S2662-B	chr6:32700783-32700875 --GRCh37	118
LOC105375075	chr6:44595088-44595183 --GRCh37	129
6pB	chr6:53700970-53701061 --GRCh37	131
mh06USC-6qA	chr6:77942995-77943075 --GRCh37	128
mh06USC-6qB	chr6:96372294-96372382 --GRCh37	130
mh06USC-6qC	chr6:120130310-120130392 --GRCh37	124
mh07USC-7pA	chr7:11613136-11613224 --GRCh37	128
mh07USC-7pB	chr7:25447583-25447668 --GRCh37	140
mh07USC-7pC	chr7:57218093-57218198 --GRCh37	143
mh07USC-7qA	chr7:68052846-68052940 --GRCh37	114
mh07USC-7qC	chr7:151518731-151518818 --GRCh37	113
8pA	chr8:3306392-3306481 --GRCh37	127
GATA-B-GATA4-A	chr8:11607303-11607402 --GRCh37	108
mh08USC-8pB	chr8:40664178-40664261 --GRCh37	120
9pA	chr9:2288627-2288718 --GRCh37	134
mh09USC-9pB	chr9:31196668-31196747 --GRCh37	126
mh09USC-9qA	chr9:73362598-73362689 --GRCh37	119
mh09USC-9qB	chr9:104903823-104903911 --GRCh37	126
mh09USC-9qC	chr9:136768832-136768923 --GRCh37	127
10pA	chr10:10412198-10412286 --GRCh37	138
11pB	chr11:34437342-34437425 --GRCh37	138
12qA	chr12:40786705-40786767 --GRCh37	125
mh12USC-12qB	chr12:68508276-68508362 --GRCh37	130
mh13USC-13qA	chr13:33553572-33553664 --GRCh37	132
mh13KK-218*	chr13:54060822-54060902 --GRCh37	135
mh13USC-13qC	chr13:85907630-85907714 --GRCh37	131
mh14USC-14qA	chr14:25724211-25724295 --GRCh37	120
D16S2835	chr16:9675226-9675314 --GRCh37	128
mh17USC-17qB	chr17:49449766-49449858 --GRCh37	122
mh17USC-17qC	chr17:78919831-78919923 --GRCh37	128
mh18USC-18pA	chr18:5280008-5280095 --GRCh37	125
mh18USC-18qB	chr18:50547455-50547541 --GRCh37	145
mh18USC-18qC	chr18:63842490-63842570 --GRCh37	134
mh19USC-19pB	chr19:16151666-16151748 --GRCh37	127
LINC00906	chr19:29376017-29376100 --GRCh37	129

19qA	chr19:33764632-33764726 --GRCh37	139
19qB	chr19:54217617-54217709 --GRCh37	123
20qB	chr20:61091310-61091400 --GRCh37	131
21qA	chr21:15899204-15899286 --GRCh37	119
mh21KK-324*	chr21:46714618-46714719 --GRCh37	131
22qA	chr22:30800301-30800380 --GRCh37	128
22qB-mh22KK-070 (+rs77899570)	chr22:49060958-49061048 --GRCh37	131

Tab.2 Caratteristiche dei microplotipi del pannello

I microplotipi con * sono stati modificati eliminando uno SNP al 5' o al 3' rispetto all'originale come precedentemente descritto.

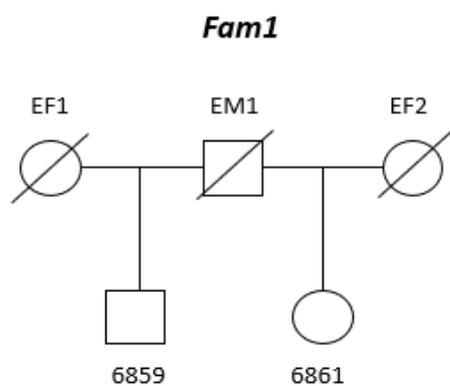
3.2 Campioni analizzati per lo studio popolazionistico

Per la validazione del pannello da 77 MH sono stati tipizzati 102 soggetti italiani, con lo scopo di stimare le frequenze alleliche nella popolazione italiana da utilizzare nel software Familias v.3.3.1 ⁽²⁵⁾ per il calcolo biostatistico.

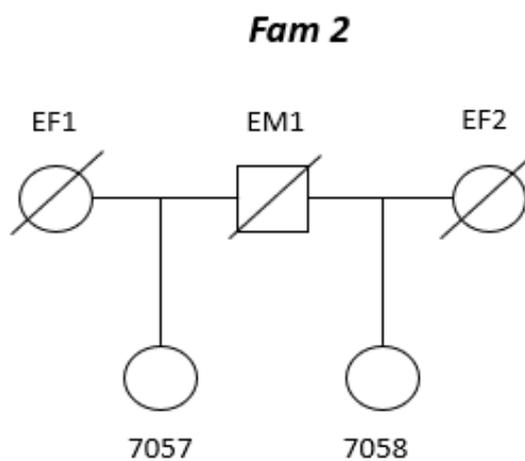
3.3 Campioni analizzati per le relazioni parentali

Per poter valutare le potenzialità del pannello nelle analisi parentali, vengono prese in considerazione 5 casi di relazioni parentali complesse o deficitarie, che vengono di seguito descritte nei 5 pedigree (FAM1, FAM2, FAM3, FAM4, FAM5).

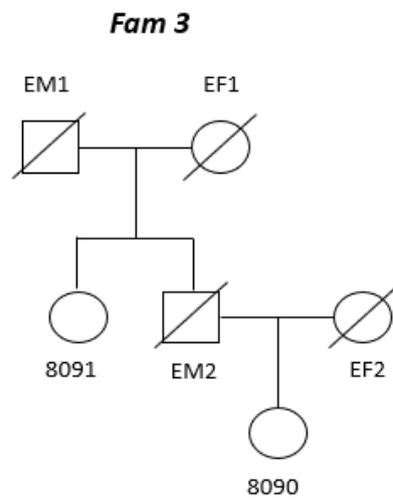
FAM 1 – fratello/sorella con madri diverse ma stesso padre (fratelli uniparentali o mezzi fratelli)



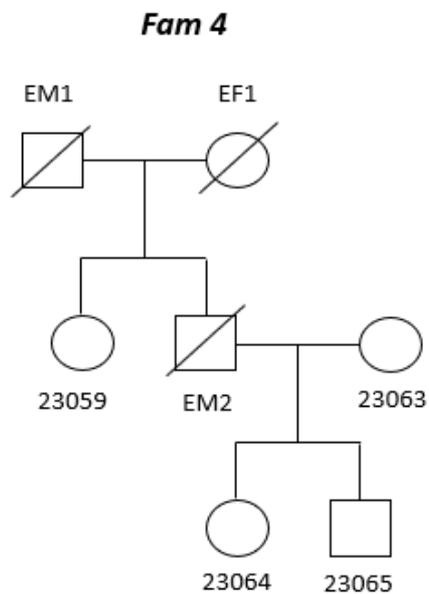
FAM 2 – sorelle con madri diverse ma stesso padre (sorelle uniparentali o mezze sorelle)



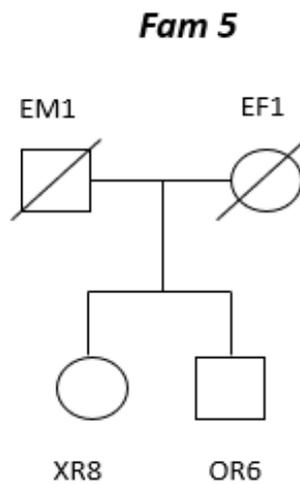
FAM 3 – zia/nipote per via paterna, in assenza di tipizzazione materna



FAM 4 – fratello/sorella dello stesso padre, con la tipizzazione della madre e di una zia paterna



FAM 5 – fratello/sorella, in assenza di genitori



Le matrici biologiche utilizzate sono tutte tamponi buccali, ad eccezione di un reperto osseo rappresentato dal campione XR8.

EF e EM sono soggetti “extra” di sesso maschile (EM) e di sesso femminile (EF) di cui sono ignote le tipizzazioni genetiche, ma che devono essere considerati per la correttezza del calcolo biostatistico.

3.4 Estrazione del DNA

Vengono applicate diverse procedure a seconda della matrice biologica da cui si isola il DNA, secondo protocolli interni stabiliti dal laboratorio, seguendo tutte le procedure necessaria per ridurre il rischio di contaminazione.

3.4.1 Estrazione DNA da tamponi buccali

L'estrazione del DNA da tamponi buccali viene eseguita in maniera automatizzata utilizzando lo strumento Maxwell 16 (Promega) mediante il kit “*DNA IQ Reference Kit for Maxwell*” (Promega, Revised 7/12-TB347).

La procedura prevede diversi step:

1. Preparare per ogni campione 500 µl di buffer di lisi combinando 495 µl di lysis buffer e 5 µl di DTT 1M;
2. Posizionare il tampone buccale in una provetta da 2 ml codificata;
3. Aggiungere il buffer di lisi preparato nel punto 1;
4. Incubare a $70^{\circ} \pm 1^{\circ}$ in agitazione per 30 minuti;
5. Trasferire il tampone buccale in uno spin basket che verrà posizionato dentro la provetta contenente il lisato;
6. Centrifugare alla massima velocità per 2 minuti;
7. Rimuovere lo spin basket e chiudere il tappo;
8. Posizionare le cartucce del kit sul supporto, rimuovendo il tappo;
9. Posizionare un plunger nel pozzetto 7 di ogni cartuccia;
10. Posizionare gli elution tube nell'apposito supporto e aggiungere 300 microlitri di elution buffer;
11. Aggiungere il campione lisato nel pozzetto 1 delle cartucce;

12. Posizionare le elution tubes e le cartucce nello strumento e impostare la modalità “SEV” che sta per Standard Elution Volume;
13. Al termine della corsa, trasferire l’eluito in una nuova provetta codificata di tipo eppendorf da 1.5 ml.

3.4.2. Estrazione DNA da resti scheletrici

Per l’estrazione del DNA da resti scheletrici viene usato lo strumento Maxwell 16 (Promega) con il “*Bone Extraction Kit*” e il “*DNA IQ Casework Pro Kit for Maxwell 16*” (Promega, Revised 12/23-TM332). Per questo tipo di matrici è necessaria una fase preliminare di demineralizzazione del campione.

1. Viene prelevato un piccolo frammento non più grande di circa un cm cubo, eventualmente per mezzo di seghetto, mortaio e martello precedentemente decontaminati con soluzione alcolica e UV;
2. La superficie della porzione del dente viene ripulita da residui di tessuti molli o altre incrostazioni tramite bisturi, eventualmente tamponandola con ipoclorito di sodio decontaminato; lasciare nella cappa sotto flusso d’aria finché il campione non è completamente asciutto;
3. Il campione prelevato viene posto all’interno del molino in acciaio e questo all’interno della busta ermetica;

4. Inserire il molino con il campione e quello senza campione (contrappeso) nel Tissue Lyser, selezionare il tempo (30 secondi) e la frequenza (30Hz) e poi avviare;
5. Aggiungere 400 μ L di demineralization buffer, 40 μ L di Proteinasi K e 10 μ L di tioglicerolo;
6. Vortexare al massimo e incubare a $56^{\circ} \pm 1^{\circ}$ in agitazione overnight;
7. Centrifugare per 5 minuti e trasferire il sovranatante in una provetta nuova senza toccare il pellet;
8. Aggiungere 72 μ L di lysis buffer e 8 μ L di tioglicerolo e centrifugare brevemente;
9. Posizionare le cartucce del kit sul supporto, rimuovendo il tappo;
10. Posizionare un elution tube davanti ad ogni cartuccia e aggiungere 300 μ L di elution buffer;
11. Posizionare un plunger nel pozzetto 8 di ogni cartuccia;
12. Aggiungere il campione lisato nel pozzetto 1 delle cartucce;
13. Posizionare le elution tubes e le cartucce nello strumento e impostare la modalità "LEV" ossia Low Elution Volume;
14. Al termine della corsa, trasferire l'eluito in una nuova provetta codificata di tipo eppendorf da 1.5 ml.

3.4.3 Quantizzazione del DNA estratto

La quantizzazione del DNA estratto viene eseguita con procedure diverse per i campioni analizzati: i campioni derivati dallo studio popolazionistico vengono quantizzati mediante E-Gel™ Agarose Gels with SYBR Safe™, mentre i campioni di DNA relativi alle famiglie sono stati quantizzati utilizzando il Qubit™ 1X dsDNA High Sensitivity (HS) (Thermo Fisher Scientific, TFS).

3.4.3.1 Quantizzazione estratto con E-Gel™ Agarose Gels with SYBR Safe™

La quantizzazione con E-Gel™ Agarose Gels with SYBR Safe™ prevede l'allestimento di un'elettroforesi su gel d'agarosio al 2%, in modo da determinare la taglia e quantizzare le regioni amplificate.

Al fine di ottenere buoni risultati, viene suggerito di caricare 20-100 ng di DNA per banda. È possibile preparare delle eventuali diluzioni del campione in acqua deionizzata o tampone. Il volume totale del campione per il caricamento è di 20 µL.

Per allestire la corsa è necessario avere a disposizione anche il DNA molecular weight marker, il quale a differenza dei campioni, viene diluito 1:20 e vengono caricati sul gel solo 10 µL. Il programma da impostare sulla cameretta

elettroforetica (E-Gel™ iBase™ device) è “E-Gel 0.8-2%” e la corsa dura 26 minuti. Al termine della corsa è possibile visualizzare le bande accendendo il transilluminatore (E-Gel® Safe Imager™ Real-Time Transilluminator), che va ad evidenziare gli ampliconi.

3.4.3.2 Quantizzazione estratto con Qubit™ 1X dsDNA High Sensitivity (HS)

La quantificazione dei campioni relative alla famiglie viene effettuata attraverso lo strumento Qubit 2.0 (TFS), seguendo il protocollo “*Qubit™ 1X dsDNA HS Assay Kits*” (MAN0017455). Questo strumento permette di quantizzare DNA a doppio filamento, DNA a singolo filamento, RNA e miRNA, sfruttando dei fluorofori che si legano ai diversi acidi nucleici analizzati. Questo kit prevede l'utilizzo di due standard che permettono di costruire una retta di calibratura, usata dallo strumento per quantificare i campioni.

Per eseguire il test, è sufficiente diluire il campione nella soluzione di lavoro 1X fornita; quindi, leggere la concentrazione utilizzando il fluorimetro Qubit™: il volume finale di ogni provetta deve essere di 200 µL. Ogni provetta standard richiede 190 µL di soluzione di lavoro Qubit™, mentre ogni provetta di campione ne richiede 180-199 µL.

Il test è raccomandato per concentrazioni iniziali di campione comprese tra 10 pg/ μ L e 100 ng/ μ L, con un intervallo di rilevamento di base compreso tra 0,2 ng e 100 ng di DNA.

3.5 Preparazione manuale delle librerie

La preparazione delle librerie viene svolta manualmente con il kit Precision ID Library seguendo il protocollo *Precision ID SNP Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System Application* (ThermoFisher Scientific, MAN0017767). In questo studio, si opta per l'utilizzo di metà del volume rispetto a quello che viene riportato nel kit.

3.5.1 Amplificazione delle regioni target

La quantità di DNA suggerita dal kit per ogni reazione di amplificazione è di 1 ng, ma il volume può variare in base alla concentrazione del DNA di partenza.

Ogni reazione di PCR viene allestita come mostrato nella tabella sottostante.

Componenti	Volume
5X Ion AmpliSeq™ HiFi Mix	2 μ L
Custom Ion AmpliSeq™ SNP Panel (2x)	5 μ L
DNA (1 ng)	X μ L
Acqua Nucleasi-Free	3-X μ L
Totale	10 μ L

Tab. 3 Mix di PCR per l'amplificazione della regione di interesse

Lo step di amplificazione viene svolto utilizzando il termociclatore *Veriti™ 96-Well Thermal Cycler* (TFS) impostando una serie di condizioni ad ogni step.

Step	Temperatura	Tempo
Active the enzyme	99°C	2 minuti
Denature	99°C	15 secondi
Anneal and extend	60°C	4 minuti *
–	10°C	Hold

Tab. 4 Condizioni di amplificazione

*Il tempo necessario per la fase di annealing e di extension viene stabilito in base alla coppia di primer per pool; nel nostro caso si hanno 49-96 coppie di primer quindi il tempo della fase di anneal/extend è di 4 minuti.

Viene anche raccomandato un numero di cicli di amplificazione in base alla qualità del DNA del campione:

- 22 cicli per “high-quality DNA”
- 25 cicli per “low-quality DNA”

Nel nostro studio, tutti i campioni analizzati vengono amplificati in 22 cicli, ad eccezione del campione XR8 per il quale si è proceduto anche ad una amplificazione a 25 cicli PCR data la bassa concentrazione di DNA input.

3.5.2 Parziale digestione degli ampliconi

A questo punto, ottenuto il prodotto di amplificazione, lo step successivo prevede l'aggiunta di 1 μ L di un reagente chiamato *Fupa Reagent* che permette la digestione parziale dei campioni. Le provette vengono caricate sul *Veriti™ 96-Well Thermal Cycler* con le condizioni riportate in Tab. 5.

Temperatura	Tempo
50°C	10 minuti
55°C	10 minuti
60°C	20 minuti
10°C	Hold (max 1 ora)

Tab. 5 Condizioni per la digestione parziale

3.5.3 Ligazione parziale dei barcode

A seguito della digestione parziale degli ampliconi è possibile il legame all'estremità di un Barcode, ossia un codice a barre specifico costituito da una sequenza oligonucleotidica, che permette di identificare in maniera univoca i frammenti di ciascun campione.

Prima del loro utilizzo, i barcode devono essere diluiti come mostrato in Tab 6.

Componenti	Volume
Ion P1 Adapter	1 μ L
Ion Xpress TM Barcode	1 μ L
Nuclease-free Water	2 μ L
Totale	4 μ L

Tab. 6 Diluizione Barcode

I barcode diluiti possono essere conservati a -20°C e riutilizzati entro massimo 20 giorni. Dopo essere diluiti, i barcode vengono ligati ad un' estremità degli ampliconi sfruttando l'azione dell'enzima Ligasi che catalizza la reazione di ligazione. In questo step è importante tenere l'enzima DNA ligasi in ghiaccio.

La reazione viene portata avanti utilizzando i componenti descritti della Tab 7.

Componenti	Volume
Switch Solution	2 μ L
Barcode diluito	1 μ L
DNA Ligasi	1 μ L
Totale	~15 μ L

Tab. 7 Mix di PCR per la reazione di ligazione dei barcode

A questo punto, le provette vengono caricate su VeritiTM 96-Well Thermal Cycler con le condizioni mostrate nella Tab 8.

Temperature	Tempo
22°C	30 minuti
72°C	10 minuti
10°C	Hold (max 24 ore)

Tab.8 Condizioni reazione di ligazione

3.5.4 Purificazione delle librerie

Dopo la ligazione dei barcode, viene svolta la purificazione delle librerie ottenute utilizzando Agencourt™ AMPure™ XP Reagent, una resina con biglie magnetiche a cui aderiscono gli ampliconi. In questo modo è possibile eliminare tutti i componenti in eccesso delle precedenti fasi che vanno ad interferire con le librerie ottenute. Gli step da seguire sono i seguenti:

- Trasferire le librerie in una provetta 1,5 mL e posizionarle in un rack magnetico;
- Vortexare bene la resina in modo da far risospendere le biglie al suo interno;
- Aggiungere 22,5 µL di resina ad ogni campione;
- Incubare a temperatura ambiente per 5 minuti;

Nel frattempo, preparare l'etanolo al 70% utilizzando 230 µL di etanolo e 100 µL di acqua nucleasi-free;

- Trasferire le provette con campione e resina sul rack magnetico;
- Incubare a temperatura ambiente per 2 minuti (o finché la soluzione non risulta limpida);
- Rimuovere il surnatante limpido, evitando di toccare il precipitato;
- Aggiungere 150 μ L di etanolo al 70% preparato in precedenza;
- Ruotare le provette sul rack magnetico per 5-6 volte, controllando che il precipitato migri correttamente ad ogni rotazione;
- Rimuovere il surnatante, evitando di toccare il precipitato;
- Ripetere nuovamente i due passaggi precedenti:
 - Aggiungere 150 μ L di etanolo al 70%;
 - Ruotare le provette sul rack magnetico per 5-6 volte controllando che il precipitato migri in modo corretto;
- Rimuovere il surnatante, evitando di toccare il precipitato;
- Effettuare una veloce centrifuga, facendo attenzione a rivolgere il precipitato verso la parte esterna della centrifuga;
- Riposizionare le provette sul rack magnetico;
- Eliminare l'etanolo in eccesso che si è depositato sul fondo della provetta;
- Incubare a temperatura ambiente con il tappo della provetta aperto per 2 minuti (o finché l'etanolo non è completamente evaporato).

3.5.5 Eluizione delle librerie

- Aggiungere 50 μL di LOW TE facendo in modo di sciogliere il precipitato;
- Vortexare e incubare le provette a temperatura ambiente per 5 minuti;
- Centrifugare velocemente le provette;
- Posizionare le provette su rack magnetico;
- Incubare per 2 minuti;
- Prelevare il surnatante e trasferirlo su nuove provette, già codificate con il nome del campione.

3.5.6 Quantizzazione delle librerie

Una volta ottenute le librerie purificate, si procede con la quantizzazione delle librerie mediante una PCR quantitativa, utilizzando il kit *Ion Library TaqMan™ Quantitation Kit* (MAN0015802, TFS).

Come prima cosa è necessario preparare una diluizione 1:100 combinando 2 μL di surnatante e 198 μL di acqua nucleasi.-free. Successivamente è necessario preparare delle diluizioni seriale dello standard, rappresentato da *E.Coli DH10B Control Library*, come riportato dalla Tab 9.

Standard	Control Library volume	Nucleasi-free Water volume	Concentration
1	5 μ L (undiluted)	45 μ L	6,8 pM
2	5 μ L di Std 1	45 μ L	0,68 pM
3	5 μ L Std 2	45 μ L	0,068 pM

Tab. 9 Concentrazioni diluizioni seriali librerie

Preparare la mix di PCR per il numero totale delle librerie, a cui si aggiunge anche il controllo negativo e gli standard.

Componenti	Volume (1 reaction)
Ion Library TaqMan™ qPCR Mix, 2X	10 μ L
Ion Library TaqMan™ Quantitation Assay, 20X	1 μ L

Tab.10 Mix reazione di PCR

Aliquotare 11 μ L della mix nelle provette per la quantizzazione e aggiungere 9 μ L di campione (standard, controllo negativo o librerie diluite). Il totale per ciascuna provetta è 20 μ L. Per la quantizzazione delle librerie dei campioni di DNA si utilizza lo strumento *Rotor Gene 6000 PCR System (Qiagen)*, impostando le opportune condizioni.

Da questa procedura di quantizzazione, vengono esclusi un paio di campioni che, invece, vengono quantificati mediante la *Tape Station 2200 (Agilent)*, una strumentazione che attraverso un'elettroforesi automatizzata su gel d'agarosio,

permette di avere una rapida quantizzazione e verificare se è avvenuta o meno l'amplificazione.

3.6 Diluizione, raggruppamento delle librerie e posizionamento sullo ION CHEF

Dopo aver quantizzato le librerie, queste vengono diluite per poter essere caricate sullo strumento *Ion Chef Instrument* (TFS), come mostrato in Tab.11.

Diluite to	Minimum volume	Templating size in Planned Run setup
30 pM	25 μ L	200 bp

Tab. 11 Diluizione librerie

Dopo aver diluito la libreria del campione alla concentrazione target (pM), combinare volumi uguali di più librerie diluite in una provetta.

3.7 Preparazione del template sullo HID ION CHEF

La preparazione del template sullo *Ion Chef Instrument* (TFS) avviene attraverso un'amplificazione clonale che permette di ottenere delle biglie ciascuna delle quali presenta sulla superficie tante copie dello stesso frammento di DNA. Si parla di Ion Sphere Particles o semplicemente ISP.

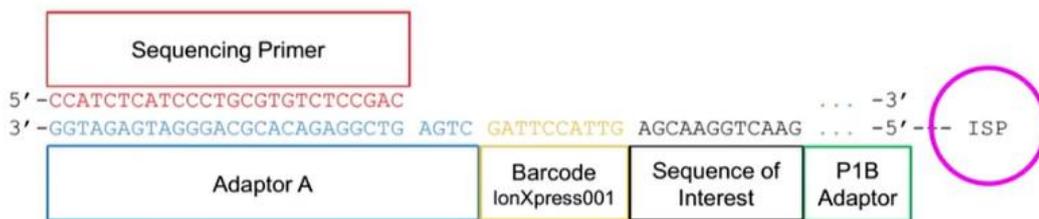


Fig. 8 Frammento di DNA all'interno della libreria

La condizione ideale prevede, quindi, di ottenere un'amplificazione monoclonale, tale per cui si ottiene l'amplificazione di un unico frammento di DNA piuttosto che una sfera policlonale in cui si ha un mix di frammenti di DNA che non sono sequenziabili.

Le ISP che non presentano l'amplificato possono essere eliminate nella fase di arricchimento, attraverso una purificazione magnetica automatizzata sfruttando la capacità della streptavidina di formare un legame con la biotina presente nel primer, grazie all'utilizzo di un magnete che può separare le particelle senza DNA da quelle che contengono il DNA amplificato.

Queste sono caricate dallo strumento su un chip di sequenziamento definito Ion Chip e, seguendo il protocollo, viene prima pianificata una corsa utilizzando il *Torrent Suite software* (TFS).

Vengono effettuate le seguenti fasi:

- È necessario disimballare le cartucce Ion S5™ Precision ID Chef Reagents 45 minuti prima dell'uso, in modo da stabilizzarle a temperatura ambiente;
- Successivamente pipettare 25 µL della libreria diluita nell'apposito Ion Chef™ Library Sample Tube, come mostrato in figura.

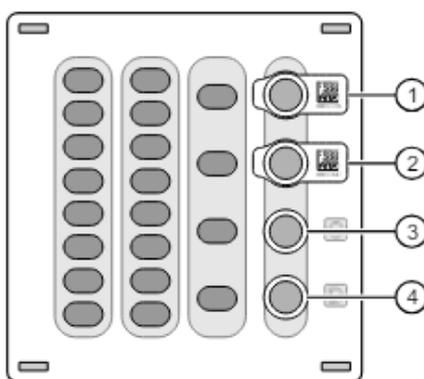


Fig. 9 Organizzazione delle cartucce: posizione 1-2 per il pool di primer, posizione 3 per le tube vuote, posizione 4 per tube output

- Coprire e conservare le due librerie di DNA diluite in ghiaccio fino a quando non si è pronti per caricarle sullo strumento;
- Rimuovere tutte le cartucce e i materiali da utilizzare dalla confezione;
- Caricare un rack di puntali vuoto nella posizione dei puntali usati, di scarto;
- Scartare una nuova cartuccia di puntali v2 e rimuovere il coperchio per esporre i puntali delle pipette;

- Caricare una nuova piastra PCR nel blocco campioni del termociclatore;
- Caricare la cartuccia dei reagenti e le soluzioni nella posizione corretta;
- Caricare il Chip-loading sulla centrifuga;
- È importante verificare se i materiali di consumo siano stati installati correttamente, come mostrato in figura;

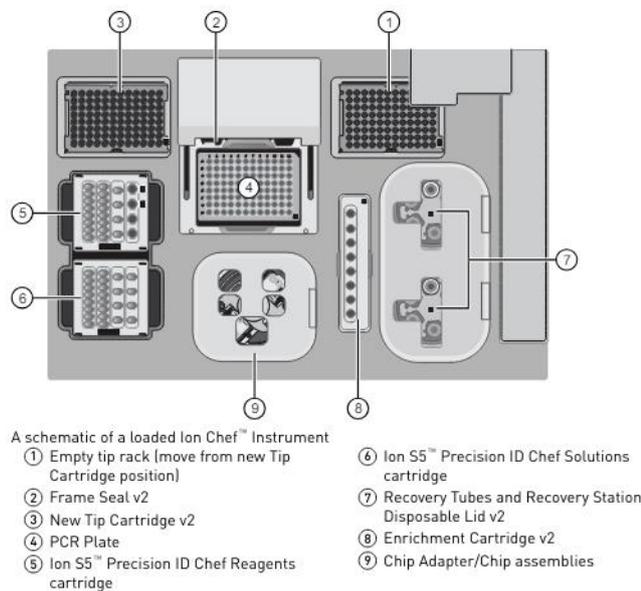


Fig.10. Illustrazione del caricamento dello strumento Ion Chef

- Selezionare Set up Run dallo schermo touchscreen dello strumento e seguire step by step le istruzioni per raggiungere il setup dello strumento;
- Selezionare Start check per iniziare Deck Scan e una volta completato cliccare next per visualizzare la Data Destination, che mostra il tipo di kit e il chip utilizzato;

- Selezionare nelle Run options le condizioni appropriate per la una corretta corsa e poi premere Start Run;
- Al termine della corsa, scaricare lo strumento e sequenziare subito il chip.

3.8 Sequenziamento con HID Ion GeneStudio S5 System

Il sequenziatore HID Ion S5™ (TFS) è in grado di verificare la compatibilità di ogni chip e il consumo dei materiali che vengono caricati durante l'inizializzazione e il sequenziamento. Prima di avviare il sequenziamento è necessario aprire le cartucce 45 minuti dell'uso, aprire il flacone della soluzione di lavaggio, capovolgerla 5 volte e farla ruotare per mescolare bene il contenuto.

Per avviare il sequenziamento seguire i diversi step:

- Inizializzare premendo Initialize nel menu principale dello schermo;
- Quando richiesto, rimuovere il flacone della soluzione di lavaggio Precision ID e svuotare i rifiuti, poi reinserire il serbato dei rifiuti vuoto.
- Sostituire la cartuccia Ion S5 Precision ID Sequencing Reagents esaurita con una cartuccia nuova equilibrata a temperatura ambiente;

- Installare il nuovo flacone della soluzione di lavaggio Precision ID, chiudere lo sportello e avviare l'inizializzazione. Al termine, toccare Run nel menu principale;
- Rimuovere il chip di sequenziamento usato e inserire il nuovo chip;
- Verificare che sia selezionata la corsa pianificata corretta, premere Review e confermare le impostazioni;
- Chiudere lo sportello dello strumento e premere Start-Run per iniziare la corsa di sequenziamento.

3.9 Analisi dei dati di sequenziamento

I dati ottenuti dal sequenziamento sono stati processati con il *Torrent Suite Software* (TFS) e le reads sono state allineate con la sequenza di riferimento del genoma umano (hg19), settando una serie di parametri, tra cui il coverage minimo totale di read da considerare per la chiamata allelica e il valore minimo della frequenza allelica per distinguere eterozigote/omozigote.

In particolare, viene utilizzata una soglia minima di copertura di base di 20X per cui le basi che non superano questo threshold non vengono assegnate.

Un'analisi primaria viene effettuata lanciando il *Coverage Analysis V5.12.00* (TFS) che fornisce un report in cui vengono riportati i valori medi del coverage per campione e per singolo amplicone.

Oltre al plug-in relativo al coverage viene lanciato il Plugin *HID_Microhaplotype_Research_PluginV1.5* (TFS), che è in grado di risolvere le fasi aplotipiche restituendo le frequenze genotipiche ed il coverage di ogni aplotipo.

Successivamente, viene svolta un'analisi secondaria utilizzando IGV, *Investigative Genomics Viewer* ⁽³³⁾, che consente di ispezionare le chiamate alleliche ambigue per verificare se si tratta di artefatti dovuti a tratti omopolimerici o a regioni ripetute.

L'analisi dei dati per lo studio popolazionistico è stata condotta mediante il software *PowerMarker* ⁽³⁴⁾, il quale permette di calcolare diversi parametri statistici per ogni locus del marcatore, tra cui il numero di alleli, la frequenza genotipica e aplotipica e l'eterozigosità.

Da questa analisi vengono ottenute le frequenze alleliche della popolazione italiana, necessarie per eseguire il calcolo biostatistico delle relazioni parentali.

3.10 Software per il calcolo biostatistico delle relazioni parentali

3.10.1 Familias (v.3.3.1)

Il software Familias v.3.3.1 ⁽²⁵⁾ viene utilizzato per calcolare probabilità e likelihood in tutti quei casi in cui è necessario determinare le relazioni parentali di uno o più soggetti, per poterne ricostruire il pedigree.

Inserendo le frequenze alleliche della popolazione esaminata e il genotipo di ogni soggetto in esame è possibile determinare il caso più semplice di relazione, ossia padre-figlio, ma anche legami più complessi.

Il software Familias v.3.3.1 ⁽²⁵⁾ è in grado di generare una moltitudine di pedigree, considerando sia gli individui disponibili per l'analisi genetica sia individui non genotipizzati ma che fanno parte delle relazioni parentali ipotizzate.

Nel caso in cui si osserva una compatibilità genetica tra presunto padre e figlio, questa deve essere valutata statisticamente per capire se è casuale oppure se il presunto padre è effettivamente il padre biologico.

Per far questo, il software consente di calcolare una stima della Likelihood Ratio (LR) che è data dal rapporto tra la probabilità di ottenere quel profilo genetico se è vera l'ipotesi di paternità e la probabilità di ottenere quel profilo se è vera l'ipotesi di non paternità. Il valore di LR che si ottiene indica quante

volte è più probabile osservare una compatibilità se il padre presunto è il padre biologico piuttosto che il presunto padre e il figlio condividano casualmente gli alleli analizzati.

Capitolo 4

RISULTATI

4.1 Risultati quantizzazione estratto

I campioni relativi alle analisi parentali sono stati quantizzati mediante lo strumento Qubit 2.0 (TSF), basato sull'utilizzo di fluorofori specifici e una retta di calibratura per quantificare i campioni. Per il campione XR8 sono state preparate due repliche di libreria NGS (XR8-1 e XR8-2).

<i>Famiglie</i>	<i>Campioni</i>	<i>[DNA] ng/ul quantizzazione estratto</i>
FAM1	6859	34,6
	6861	0,075
FAM2	7057	0,5
	7058	0,078
FAM3	8090	0,14
	8091	0,22
FAM4	23059	30,00
	23063	56,00
	23064	56,00
	23065	90,00
FAM5	OR6	0,22
	XR8	0,05

Tab.12 Quantizzazione DNA estratto

4.2 Risultati quantizzazione librerie

Le librerie genomiche preparate per il sequenziamento sono state quantizzate mediante una PCR quantitativa utilizzando il kit *Ion Library TaqMan™ Quantitation Kit* (MAN0015802). La quantizzazione in Real Time di ogni campione è stata eseguita in doppio per poter ottenere una media delle due prove. Considerando che le librerie vengono diluite 1:100, si moltiplica per 100 il valore medio ottenuto, il quale servirà nella fase di sequenziamento.

Famiglie	Campioni	Ion TaqMan library			
		qPCR1	qPCR2	media	*100
FAM3	8090	6,63	6,97	6,80	680,11
	8091	5,50	5,45	5,48	547,60
FAM4	23059	0,40	0,42	0,41	41,14
	23063	0,84	0,83	0,84	83,78
	23064	0,40	0,40	0,40	38,55
	23065	2,60	2,60	2,60	258,73
FAM5	OR6	3,40	3,4	3,40	342,13
	XR8-1	7,10	5,90	6,50	650,21
	XR8-2	26,50	-	26,50	2647,58

Tab. 13 Quantizzazione con Real Time

Quattro librerie sono state quantizzate sfruttando lo strumento Tape Station come riportato in Tab.14.

Famiglie	Campioni	Tape Station (D1000)
FAM1	6859	2700pM
	6861	214 pM
FAM2	7057	7460 pM
	7058	2600 pM

Tab.14. Quantizzazione con TapeStation

4.3 Analisi risultati del sequenziamento

Al termine del sequenziamento, i dati sono stati analizzati lanciando il Plugin *Coverage Analysis V5.12.00* (TFS) con il quale è possibile risalire ad una serie di informazioni che permettono di valutare l'efficienza del sequenziamento, riportate in tab.15.

<i>Famiglie</i>	<i>Campioni</i>	<i>Mapped Reads</i>	<i>On Target</i>	<i>Mean Depth</i>	<i>Uniformità</i>	<i>Risultati</i>
FAM1	6859	20007	97,21%	249	94,12%	Completo
	6861	55435	97,29%	676,4	95,44%	Completo
FAM2	7057	663958	96,48%	8194	95,28%	Completo
	7058	21631	95,95%	264,7	97,49%	Completo
FAM3	8090	65815	95,49%	794,5	95,72%	Completo
	8091	85406	96,15%	1045	94,24%	Completo
FAM4	23059	120463	94,49%	1425	91,97%	Completo
	23063	107405	84,73%	1148	90,73%	Completo
	23064	117730	87,67%	1285	91,86%	Completo
	23065	114500	94,00%	1367	91,87%	Completo
FAM5	OR6	125924	96,51%	1553	90,02%	Completo
	XR8-1	28103	38,78%	134,9	93,29%	Parziale, mancano 5/77 mh
	XR8-2	35366	36,78%	160,2	93,53%	Parziale, mancano 3/77 mh

Tab.15 Risultati del sequenziamento

Vengono analizzati:

- *Mapped read*: numero totale di letture mappate all'interno del genoma di riferimento (hg19);
- *On target*: numero di reads che ricadono nel target scelto;
- *Mean Depth*: la media del numero di volte che ogni nucleotide è stato letto per ogni singolo campione;

- *Uniformità*: indica quanto uniformemente siano distribuite le letture del target.

Un altro importante parametro da considerare è il *relative depth of coverage*, che rappresenta la somma di tutte le read per campione per ogni microarray. Per ottenere questo valore si calcola la media di ogni amplicone relativo al numero totale di campioni facendo un rapporto tra il numero di reads per amplicone diviso il numero totale di reads per campione. Questo valore medio permette di valutare l'efficienza di amplificazione dei singoli ampliconi durante la reazione di preparazione delle librerie e di valutare l'uniformità di amplificazione tra i vari ampliconi che compongono il pannello.

Osservando il grafico, è possibile notare che il marcatore, 6pB mostra un coverage normalizzato più basso rispetto agli altri microarray, seguito poi da D2S259; per questo motivo vengono definiti underperforming.

Inoltre, il grafico viene sviluppato per grandezza dell'amplicone dal più piccolo al più grande per poter mostrare che l'underperforming che si osserva non dipende dall'ampiezza.

Successivamente viene lanciato il Plugin HID_Microhaplotype_Research_PluginV1.5 (TFS), il quale fornisce una descrizione dei microaplotipi che costituiscono il pannello, per ciascun campione. Da questa analisi è disponibile un report riepilogativo che mostra la genotipizzazione di ogni singolo campione relativi alle famiglie, come riportato in Tab.16-17.

Marcatore	6859	6861	7057	7058	8090	8091
FNDC10	ACGC/GCCG	ACGC/ACGC	GCCG/GCCG	GCCG/ACGC	ACGC/ACGC	ACGC/GCCG
mh01KK-160	CGAA/TGAA	TGAA/TTAA	TGAA/TGAA	CGAA/TGAA	TGAA/TTAA	CGAA/TGAA
1qB	TCGC/CGGC	CGGC/CGGC	CGGC/CGAC	TCGC/CGGC	CGAC/TCGT	CGGC/CGGC
mh02KK-140	GGGG/CGGA	GGGG/CGGA	GGGG/GGAG	CGGG/GAGG	CGGG/GGGG	CGGG/GGGG
mh02KK-141	GCTA/GCTA	GCTA/GGGG	GCTG/GGTG	GCTA/GGGG	GCTG/GGGG	GGGG/GCTA
D2S259		ACGC/ATGG	ATGG/ATGG	ATGG/ATGG	ACGC/ACGC	ACAG/ACGC
2pC	ACGT/GCCG	ACGT/GCCG	GCTT/GCTT	GCTT/GCCG	GCTT/GCCG	GCCG/GCCG
2qA	CCAC/CTGC	CCAG/CCAG	TTAC/CTGC	TTAC/CTGC	CCAC/CCAG	CCAC/CTGC
mh02KK-134	TTC/TTT	TTT/TTT	ACT/TTT	ACC/ACT	ATC/TCT	ATT/TCT
2qD	GCA/GGG	GCG/GCA	AGG/GGG	GGG/GCA	GCG/GCA	GCA/AGG
mh02KK-136	GTC/GTC	GTC/GTC	GTA/TTC	GTC/TTC	GTC/GTC	TCA/GTA
D3S2426	ATC/ATA	ATC/ATC	GTC/GTC	ATC/ATA	GTC/ATC	GTC/ATC
3pA	CTAC/CCGT	CCGT/CCGT	CTAC/ACGT	CTGC/ACGT	CTGC/CTAC	CTGC/CTAT
mh03KK-217	CdGCC/CTGCC	CdGCC/ATGCT	ATGCT/CdGCC	CdGCC/ATGCT	AdGCC/ATGCT	ATGCC/CdGCC
ZNF385D	CATG/AATG	AATG/CGTG	AATG/CACG	AATG/CACG	CATG/CATG	AATG/CATG
3pB	AACCG/CGCGA	AACCG/AACCG	AACCG/AACCG	AACCG/AACCG	AACCA/AACCA	CGCGA/AACCG
LINC02000	CGTC/CGTC	CGTC/CGTC	CGTC/CCTC	CGTC/CCTC	CGTC/GGTC	CGTC/CGTC
LOC100505920	GCGT/GCGT	GCGT/GCGT	GCGA/CTGA	GCGT/CTGA	CTGA/GCGA	GCGA/GCGT
mh04KK-010	AA/AA	GA/AA	GA/GA	GA/GA	AG/GA	GA/GA
SORCS2	CCT/CCT	CCT/CCT	CCT/CCT	CCT/CCT	CCT/CCG	CCG/CCT
4pA	GTAA/GTAG	GTAG/CTCA	GGCA/GGCA	GTAG/GGCA	CTCA/GGCA	GGCA/GGCA
4pB	ACGC/GCAG	GCAC/GCAG	ACGC/GCAC	ACGC/GCAC	GCAC/ATGC	ACGC/GCAG

4qA	CCT/CTT	CCT/CCT	CCG/CCG	CCT/CCT	GCG/CTT	CCT/CTT
4qC	TGAG/TGAG	TAGG/TGAG	TAGG/GGGG	GGGG/GGGG	TGAG/TAGG	TAGG/GGGG
mh04KK-217	ATT/ACT	ATT/ACC	ATC/ACT	ATC/ATC	ACT/ATC	ATC/ACT
mh04KK-218	GGTC/GGTC	GGTC/GGTC	AACC/AACC	AACC/GGTC	GGTC/AACC	AACC/AACC
D4S2390	AGAGT/AGAGC	AGAGT/AGAGC	AGAGC/AGGCT	AGAGC/GGAGC	AAAGT/AGGGC	GGAGC/AGGCT
D4S214	AAAG/AAAG	AGGG/AGGG	AAAG/AAAA	AGGG/AAAG	GGAG/AGGG	AGGG/AGGG
mh05KK-125	AATG/GATG	AATG/GATG	AATG/GGTG	AATG/GATG	AATG/AATG	GATC/AATG
mh05KK-022	CA/TC	CA/CC	TC/CC	CA/CC	CA/CC	CA/TC
5qB	CCAT/CCAC	TTAC/CCAC	CCAT/TTAC	TTAC/TTAC	TTAC/CAAC	TTAC/CCAT
5qC	CTGT/TCAC	CCGC/CCGC	TCAC/TCAC	TCGC/CCGC	CCGC/CCGT	CCGT/TCAC
5qD	GAA/GAA	GAA/GAA	GAG/AAG	AAG/AAG	GGG/GAA	GGG/GAA
6pA	CCA/GCT	CCT/CCT	CCT/CCT	CCA/CTT	CCT/CCT	CCT/CCA
mh06KK-103	CCTA/CCTG	CCCA/CCCA	CCCA/CCCA	CCCA/CCCA	CCTG/CCTA	CCTG/CCCA
D6S2662-B	AGAA/AGGT	AGGA/TGGT	AGGA/TGGT	TGGT/TGGT	AGGT/TGGT	AGGT/TGGT
LOC105375075	TTCC/GTTC	GTTC/GTTC	TCTC/GTTC	TCTC/GTTC	GTTC/TCTC	TCTC/TTCC
6pB						GTAG/GTAA
6qA	GCT/GTT	GCT/GTT	GCC/GTT	GCC/GTT	CCT/GTT	CCT/GTT
6qB	GTGG/GCGG	GCGG/GCGG	GCGG/GCGG	ACAG/GCGG	ACAG/GTGG	GCGG/GTGG
6qC	CTC/CTC	ATC/CTC	ATC/ATC	ATC/ATC	ATC/ATC	ATT/ATC
7pA	AGTC/TATC	AATA/TATC	AGTC/TATC	AGTC/TATC	AAGA/TATC	AGTC/TATC
7pB	AAGC/TAAT	AGGC/AAGC	TAGT/AGGC	TAGT/AGGC	TAAC/AAGC	AAAC/AAGT
7pC	CTCGT/CTCGT	CCTAC/TCCGC	CCTAC/CCTAC	CCTAC/TCCGC	CTCGT/TCCGC	CTCGT/CTCGT
7qA	CAT/GAT	CAT/CAT	CGC/CGC	CGC/CGC	CGC/CAC	CAC/CAC
7qC	AGCAC/TACGC	AGCAC/TACGC	AGCAC/TACGC	AGTGC/AGTGC	TACGC/AGTGC	AGTGC/AGTGC
8pA	CAGTA/CAGTG	CAGTG/CAGTG	CGGTG/CAGTG	CGGTA/CAGTG	CAGTA/CAGTG	CGGTG/CAGTA
GATA-B-GATA4-A	CCTAG/CTCAA	CCTAG/CTCAA	CCTAG/TTTAA	CCTAG/CCCAG	CCTAG/CCTAG	CCTAG/CCCAG
8pB	TTCTA/CATCA	CATCA/TTCTA	CATCA/CATCA	CATCA/CATCA	CATCA/CATCA	CTCTG/CATCA
9pA	GGTCGC/GGCCA C	AGTCAC/GGCCA C	GATCAA/AGTCA C	AATTGC/GGCCA C	AGTCAC/AGTCA C	AGTCAC/AGTCA C
9pB	TCGT/TCGT	TCGC/GTAC	TCGC/GTAC	GTAC/GTAC	TCGT/TCGT	GTAC/TCGT
9qA	GGC/TGC	GGC/TGC	TGC/GGG	GGC/TCC	GGG/TGC	GGC/GGG
9qB	TGCT/TGCT	TGCT/TGCT	TGCT/TGCT	TGCT/TGCT	TGCT/CATT	TGCT/CATT
9qC	ATCGC/ATCGC	ACCGC/ATCGC	GCCGC/ATCGC	GCCGC/ATCGC	GCTGA/GCTCA	GCTCA/ATCGC
10pA	CTGG/CTGG	CTGG/CTGG	CTAG/CTAG	CTAG/CTAG	CTGG/CTGA	CTAG/CTGG
11pB	TCGCG/TCGCA	TCGCG/CCATG	TTGCG/TTGCG	TTGCG/TCGCG	TCGCA/TTGCG	TCGCA/CCGTG
12qA	GTGG/GTGG	GTGG/GTGG	GGGG/GTGG	GGGG/GTGG	GGGG/GGGG	GGGG/GGGG
12qB	GAAC/AAAT	GAAC/GCAC	AAAC/GACC	GAAC/AAAT	AAAT/AAAT	AAAT/AAAT
13qA	GCAA/GCAA	GCAA/ATAA	ATAA/GTAA	GTAA/ATGG	ATAA/ATAA	ATAA/ATAA
mh13KK-218	TTC/TTC	TTT/CCC	TTC/TTT	TTT/TTT	TTT/CTC	TTT/CCC
13qC	TGGC/TGGC	TGTG/TGGC	CGTC/TGGC	CGTC/TGGC	TGTG/CGTC	CCTC/CCTC
14qA	TCC/GCG	TCC/GTG	TCC/GCG	TCC/GTG	TCC/TCC	GCG/TCC
D16S2835	TCGTA/TCGTA	TCGTA/TCGTA	GTGCG/TCGTA	TCGTA/GTGCG	TCGTA/TCGTG	TCGTA/GTGCG
17qB	AGT/AGC	AGC/GGC	GGC/GGT	AGC/GGC	GGC/GGC	AGT/GGC
17qC	GGAC/GCGT	GCGT/CCGC	GGAC/CCGC	GCAT/CCGC	GCGT/CCGC	CCGC/CCGC

18pA	TCGTG/TCGTG	CAATG/TCGTG	TCGTG/CAATG	TCATA/TCATG	TCGTG/TCATG	TCATA/TCGTG
18qB	TAC/TAC	GGC/TAC	GGC/TAC	TAC/GAT	GGC/GAT	TAC/GAT
18qC	TGAT/TGAT	TGAT/TGAT	TAGG/TGAT	TGAT/TAGG	TGAT/TAGG	TAGG/TAGG
19pB	AGTCG/GGTTG	AGTCG/GGTTG	AACTG/AGTCT	GGTTG/AGTCT	AGTCT/AACTG	AGTCT/AACTG
LINC00906	ATCA/ATCA	ATCG/ATCA	ATCA/ATCG	ATCG/ATCA	ATCG/ATCA	ATCG/ATCG
19qA	TCGCC/CCATC	TCGCC/TCGCC	TCGCC/CCACC	TCGCC/CCACC	CCACC/TCGCC	CCATC/TCGCC
19qB	AGCG/AGGG	GAGA/AGCG	AGGG/AGCG	AGCG/AGGA	AGCA/AGCA	AGCA/AGGG
20qB	CCAAC/CCAAC	TTCGC/CCAAC	CCCAT/TTCGC	CCCAT/CCAAC	TTCGC/CCAAC	CCCGT/TTCGC
21qA	GTGG/CAGG	GTGG/GTGG	GAGA/GAGA	GAGA/GAGA	GAGA/GAAG	GAGA/GAGA
mh21KK-324	CAG/TAA	CAG/CTA	CAG/CAG	CAG/TAA	CAG/TAA	TAA/CAG
22qA	TCGG/TCGG	TCGG/ACTA	TCGG/TCGG	ACTG/ATGG	TCGG/ACTG	ATGG/ACTA
22qB-mh22KK-070	AATATG/GCGTA	GCGTA/AATATG	AATATG/AGTGA	AGCATG/AATATG	GGCATG/AGTGA	GGCATG/AATATG

Tab. 16 Tipizzazione relativa a FAM1, FAM2 e FAM3

Marcatore	OR6	XR8	23059	23063	23064	23065
FNDC10	ACGC/GCGG	GCGG/GCGG	ACGC/ACGC	GCGG/GCGG	ACGC/GCGG	GCGG/ACGC
mh01KK-160	CGAA/TGAA	CGAA/TGAA	TTAA/CGAA	CGAA/TTAA	CGAA/TTAA	CGAA/CGAA
1qB	CGGC/TCGT	CGGC/TCGT	CGGC/TCGC	CGGC/CGGC	CGGC/TCGC	CGGC/TCGC
mh02KK-140	CGGA/CGGG	GGAG/GGGG	GGGG/GGAG	CGGG/GAGG	GGAG/CGGG	GAGG/GGAG
mh02KK-141	GCTA/GCTA		GCTA/GGTG	GCTA/GCTA	GCTA/GCTA	GCTA/GCTA
D2S259	ACAG/ACGC	ACAG/ACGC	ATGG/ACAG	ACGC/ACAG	ACAG/ACAG	ACGC/ACAG
2pC	GCGT/GCTT	ACGT/GCGG	GTGT/GCGG	ACGT/ACGT	GCGT/ACGT	ACGT/GCGT
2qA	CTGC/CCAC	CTGC/TTAC	CCAC/TTAC	TTAC/CCAG	TTAC/CCAG	TTAC/CCAG
mh02KK-134	ATC/ACT	ATC/ACT	TCT/TTC	ACT/ATC	TCT/ACT	ATC/ATC
2qD	AGG/GCG	AGG/GGG	GCA/GGG	AGG/GGG	GGG/GCA	GCA/AGG
mh02KK-136	TCC/GTC	TCC/GTA	GTC/TTC	GTC/GTC	GTC/GTC	GTC/GTC
D3S2426	GTC/ATC	ATC/GTC	ATC/ATC	ATA/ATC	ATC/ATC	ATC/ATA
3pA	CTAT/CTAC	CTGC/CTAC	CCGT/CTAT	CCGT/CTAT	CTAT/ACGT	ACGT/CCGT
mh03KK-217	CdGCC/CTGCC	ATGCT/ATTCT	ATGCT/ATGCT	AdGCT/ATGCT	CdGCC/AdGCT	ATGCT/CdGCC
ZNF385D	CATG/AATG	CATG/AATG	CACG/CACG	CGTG/CATG	CGTG/CGTG	CGTG/CGTG
3pB	AACCA/AACCA	AACCA/AACCA	AACCG/CGCGA	AACCA/AACCG	CGCGA/AACCG	CGCGA/AACCG
LINC02000	CGTC/CCTC	CCTC/CGTC	GGTC/CGTC	CGTC/GGTC	CGTC/CGTC	CGTC/GGTC
LOC100505920	GCGT/GCGT	GCGT/GCGT	GCAT/GCGT	GCGT/GCGT	GCGT/GCGT	GCGT/GCGT
mh04KK-010	AG/GG	GG/GA	AA/AA	AG/GA	GA/GA	AG/GA
SORCS2	CCT/CCG	CCT/CCT	CCT/CCT	CCT/CCG	CCG/CCG	CCT/CCT
4pA	GGCG/CTCA	GTAG/GGCA	CTCA/GTAA	GGCA/GGCA	GGCA/GTAG	GGCA/GGCA
4pB	GCAC/ACGC	GCAC/ACGC	GCAC/GCAG	ACGC/ACGC	ACGC/GCAG	GCAG/ACGC
4qA	CCG/CTT	CTT/CCG	CCG/GCG	CTT/CCT	CTT/CCG	CCT/CTT
4qC	TAGG/TGAG	TAGG/TGAG	TGAG/TGAG	TGAG/TGAG	TGAG/TGAG	TGAG/TGAG

mh04KK-217	GCC/ACC	ACC/GCC	ACC/ACC	ACC/ATC	ACC/ACC	ATC/ACC
mh04KK-218	AACC/AACT	AACT/GGCC	AACC/GGCC	GGCC/AACT	AACT/AACC	GGCC/GGCC
D4S2390	AAAGT/AGAGC	AAAGT/AAAGT	GGAGC/GGAGC	AGAGC/AGGCT	GGAGC/AGAGC	GGAGC/AGGCT
D4S3214	GGAG/GGAG	GGAG/GGAG	AAAG/GGAG	AGGG/AAAG	AGGG/AAAG	AAAG/AAAG
mh05KK-125	AATG/GATC	GATC/AATG	GGTG/AATG	AATG/AATG	AATG/GGTG	AATG/AATG
mh05KK-022	CC/TC	TC/TC	TC/TC	CC/CC	TC/CC	CC/TC
5qB	TTAC/TTAC	TTAC/CCAT	CCAC/CCAT	CCAC/TTAC	CCAC/CCAC	CCAC/TTAC
5qC	CTGT/CCGT	TCAC/CCGC	CCGC/TCAC	CCGT/CTGT	CCGC/CCGT	CCGT/CCGC
5qD	GAG/GAG	GAA/GAG	GAA/GAA	GGG/GGG	GGG/GGG	GGG/GAA
6pA	CCT/CCA	CCA/CCA	CCT/CCT	CCA/CCA	CCA/GCT	CCA/GCT
mh06KK-103	CCTG/CCTG	CCTG/CCTA	CCTG/CCTG	CTCA/CCTA	CCCA/CTCA	CCTA/CCCA
D6S2662-B	TGGT/AGGA	TGGT/TGGT	AGAT/AGGA	TGGT/AGAA	TGGT/TGGT	TGGT/AGAA
LOC105375075	TCTC/TCTC	TCTC/TCTC	TTTT/TTTC	TTCC/TTTT	TTCC/TTTT	TTTT/TTTT
6pB	GCAA/GCAA		GCAA/GTAA	GCAA/GTAG		GCAA/GCAA
6qA	GCC/GCT	GCT/GCT	GCT/GCT	CCT/CCT	CCT/GCC	CCT/GCT
6qB	GCAC/ACAG	GCAC/ACAG	GTGG/GTGG	GCAC/GCGG	GCAC/GCAC	GTGG/GCGG
6qC	ATC/ATC	ATC/ATC	CTC/CTC	CTC/ATC	CTC/CTC	CTC/CTC
7pA	AATA/TAGC	AGTC/TAGC	AATA/TATC	AGTC/AAGA	TATC/AAGA	AGTC/AATA
7pB	TAGC/TAAC	TAGC/TAAC	AAGC/AAGC	TAGC/TAGT	AAGC/TAGT	AGGC/TAGT
7pC	CTCGT/CTCGT	CTCGT/CTCGT	CTCGT/CTCGT	CCTAC/CCTAC	CTCGT/CCTAC	CTCGT/CCTAC
7qA	CAT/CGC	CAT/CAC	CGC/CGC	CAC/CGC	CAC/CAC	CAC/CAC
7qC	AGTGC/AGTGC	AGTGC/AGTGC	AGTGC/AGTGC	AGCAC/AGTGT	AGTGC/AGCAC	AGTGC/AGTGT
8pA	CGGTA/CGGTA	CGGTA/CGGTA	CAGTA/CGGTA	CGGTG/CGGTA	CGGTA/CGGTG	CAGTA/CGGTA
GATA-B-GATA4-A	CCCAG/CTCAA	CTCAA/CCCAG	CTTAG/CCTAG	CTCAA/TTAA	CTCAA/CTCAA	CTCAA/CTCAA
8pB	TTCTA/TTCTA	TTCTA/TTCTA	CTCTG/CATCA	CTCTG/CATCA	TTCTA/CATCA	CTCTG/TTCTA
9pA	AGTCAC/GGTCC _C	GGTCGC/GATCA _A	GATTGC/AATTG _C	GGTCGC/GATCA _A	GGTCGC/AATTG _C	AATTGC/GATCA _A
9pB	TCGT/TCGC	TCGT/GTAC	TCGC/TCGT	TCGT/GTAC	GTAC/TCGC	TCGC/TCGT
9qA	GGC/GGG	GGG/GGC	GGG/TCC	GGC/TCC	TCC/TCC	GGC/TCC
9qB	TGCT/CGTT	CGTT/TGCT	CATT/TGCT	TGCT/TGCT	TGCT/CATT	TGCT/TGCT
9qC	ATCGC/ATCGC	ATCGC/ATCGC	ATCGC/ATCGC	ATCGA/ACCGC	ATCGC/ACCGC	ACCGC/ATCGC
10pA	CTGA/CTGG	CTGA/CTGG	CTAG/CTGG	TAGG/TAGG	CTGG/TAGG	TAGG/CTGG
11pB	CCATG/TCGCA	TTGCG/CCATG	TCGCA/TTGCG	CCATG/CCATG	TTGCG/CCATG	TTGCG/CCATG
12qA	GTGG/GGGG	GTGG/ATAA	GGGG/ATAA	GTGG/ATAA	GTGG/ATAA	ATAA/ATAA
12qB	GACC/GCAC	GCAC/GACC	GAAC/GAAC	AAAT/GAAC	GAAC/GAAC	GAAC/AAAT
13qA	ATGG/ACAG	ATGG/ACAG	ACAG/ATGG	GCAA/ACAG	ACAG/ACAG	GCAA/ATGG
mh13KK-218	CTC/TTT	TTT/TTC	CCC/TTT	CTC/CCC	CTC/TTT	TTT/CCC
13qC	CCTC/TGGC	TGGC/CGTC	CCTC/TGGC	TGTG/TGGC	TGGC/TGGC	CCTC/TGGC
14qA	TCG/GCG	TCC/GCG	TCC/TCC	TCC/TCC	TCC/TCC	TCC/TCC
D16S2835	GTGCG/TCGTG	TCGTG/GTGCG	TCGTA/TCATG	TCATG/TCATG	TCATG/TCGTA	TCGTA/TCATG
17qB	AGT/AGT	AGT/GGC	GGC/GGC	AGT/GGC	AGT/GGC	AGT/AGT
17qC	GGAC/CCGC	GGAC/CCGC	CCGC/GGAC	GCGT/CCGC	CCGC/GCGT	GCGT/GCGT
18pA	TCATA/TCGTG	TCGTG/TCATA	TCGTG/TCATA	TCGTG/TCGTG	TCGTG/TCATA	TCATA/TCGTG
18qB	GGC/TAC	TAC/TAC	TAC/GAT	TAC/TAC	TAC/TAC	TAC/GAT

18qC	TGAT/TAGG		TGAT/TGAT	GAAG/TAGG	TGAT/TAGG	TGAG/GAAG
19pB	AGTCT/GGTTG	AGTCG/GGTTG	AGTCG/AGTCT	AGTCT/AACTG	AGTCG/AGTCT	AGTCT/AGTCG
LINC00906	ATCA/ATCA	ATCA/ATCA	ATTG/ATCG	ATCG/ATCG	ATCG/ATTG	ATCG/ATTG
19qA	CCATC/CCACC	CCATC/CCATC	CCACC/CCACC	CCATC/CCATC	CCATC/CCACC	CCATC/CCACC
19qB	AGCG/GAGA	GAGA/AGCG	AGGG/AGCG	AGCA/GAGA	GAGA/AGCA	GAGA/AGCA
20qB	CCCAT/TTCGC	CCCAT/TTCGC	TTCGC/CCAAC	CCCAT/TTCGC	TTCGC/CCAAC	TTCGC/CCAAC
21qA	GAGA/GTGG	GAGA/GTGG	GTGG/GAGA	GAGG/GAGA	GAGA/GAGG	GAGA/GAGG
mh21KK-324	CAG/TAA	TAA/CAG	CAG/CTA	CTG/CAG	CTA/CTG	CTA/CTG
22qA	ACTG/ATGG	ACTA/TCGG	ATGG/TCGG	ATGG/ATGG	ATGG/ATGG	ATGG/TCGG
22qB-mh22KK-070	GGCATG/AATATG	AATATG/GGCATG	AATATG/AATATG	AATATG/AATATG	AATATG/AATATG	AATATG/GGCATG

Tab. 17 Tipizzazione relativa a FAM4 e FAM5

A partire da questo report, viene svolta un'ulteriore ispezione manuale tramite il software IGV ⁽³³⁾ per analizzare le varianti che non superano il filtraggio del plug-in dei microplotipi.

Per verificare se i 6 microplotipi selezionati dal pannello di 89 MH e poi modificati nel pannello da 77 MH restituivano risultati concordanti per quanto riguarda i genotipi, è stato fatto un confronto di tutti i campioni analizzati.

Dal confronto è emersa una discrepanza relativa ad un solo campione per il microplotipo mh02KK-134 ed è dovuta ad un minor coverage nel pannello da 89MH per quel campione; di conseguenza non viene assegnata l'eterozigosità, in quanto il valore è inferiore al threshold stabilito.

4.6 Analisi con Familias

Sulla base delle frequenze alleliche ottenute dai campioni dello studio popolazionistico, sono state effettuate delle simulazioni con il software

Familias v.3.3.1 ⁽²⁵⁾ al fine di valutare il potere informativo del pannello sviluppato.

Le simulazioni sono state effettuate con diversi scenari di parentela, quali fratelli pieni, mezzi fratelli e cugini di primo grado, considerando sia i marcatori STR sia i microplotipi, in modo da poter confrontare il potere informativo di questi marcatori. Il software non è stato in grado di calcolare LR dei cugini di primo grado per gli STR.

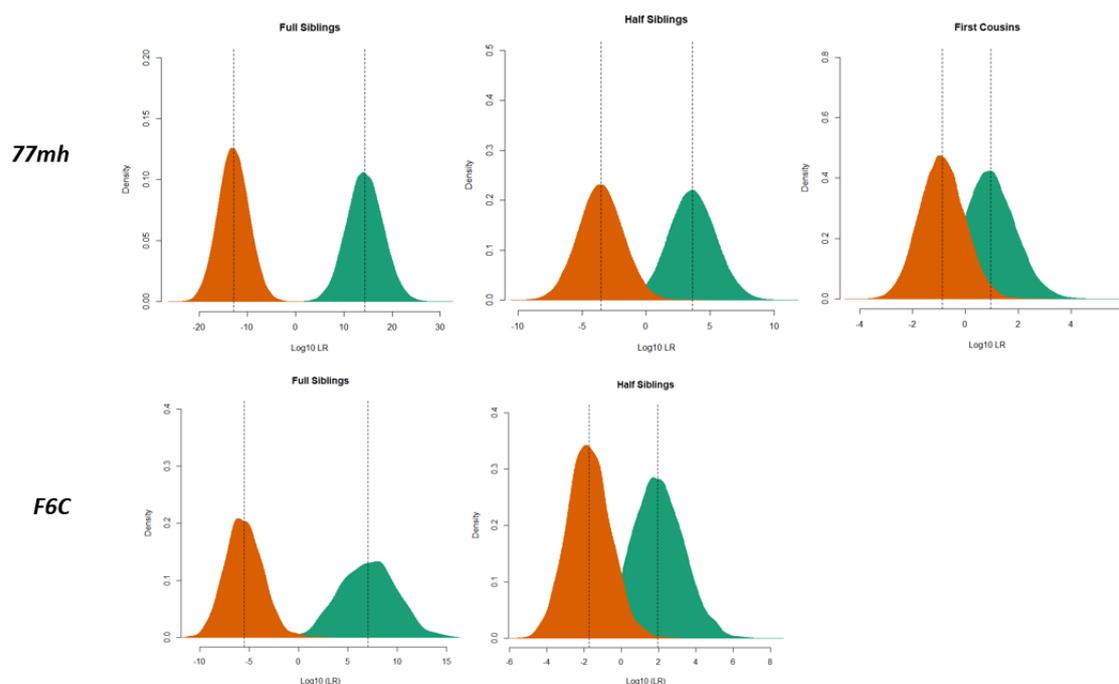


Figura 12 Confronto della distribuzione di LR tra MH e STR in tre scenari parentali

Nel grafico viene mostrata la distribuzione dei rapporti di verosomiglianza su base logaritmica di individui imparenti, indicata in arancione, e quella di individui non imparenti, indicata in verde, per tutti e tre gli scenari di parentela.

Ciò che si va a valutare è la possibile sovrapposizione delle distribuzioni della LR nei casi di parentela analizzati: distribuzioni sovrapposte indicano che il pannello presenta dei limiti nel distinguere soggetti imparentati da quelli non imparentati.

Dopo aver stabilito che il pannello da 77 MH ha un potere informativo tale da poter risolvere casi di parentela, sono state effettuate le analisi su 5 famiglie per ricostruire le relazioni parentali in base al valore di LR fornito dal software.

Per l'interpretazione dei risultati si fa riferimento alle linee guida SIGU ⁽²⁷⁾ sulle analisi genetiche di accertamento parentale che consentono di stabilire tre possibili risultati: ⁽²⁶⁾

- Esclusione: per escludere una compatibilità biologica il valore del rapporto di verosomiglianza deve essere inferiore a 1:10.000;
- Attribuzione: per definire l'attribuzione il valore del rapporto di verosomiglianza deve essere superiore a 10.000:1;
- Inconclusività: viene definito inconclusivo qualsiasi risultato che generi un rapporto di verosomiglianza compreso 1:10.000 e 10.000:1.

Seguendo queste linee guida sono stati interpretati i risultati ottenuti dall'analisi di Familias v.3.3.1 ⁽²⁵⁾ per le cinque famiglie analizzate.

Nella tabella sottostante sono riportati i valori di LR ottenuti sia dall'analisi degli STR che dall'analisi di due diversi set di frequenze aplotipiche dei MH. In dettaglio l'analisi è stata condotta sia con le frequenze aplotipiche ottenute dallo studio popolazionistico descritto nel presente studio che con un secondo database costituito da frequenze riportate: nel catalogo 1000Genome Project (<https://internationalgenome.org/>), nel lavoro ⁽²³⁾ e nel database MicroHapDB ⁽²⁸⁾ (<https://github.com/bioforensics/MicroHapDB>).

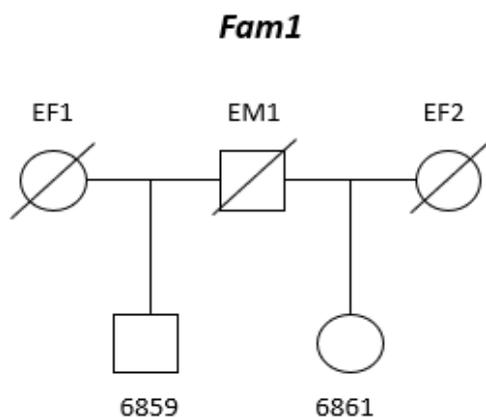
		STR	MH (Freq. Studio popolazionistico)	MH (Freq. 1000K, Turchi 2019, MicrohapDB)
FAM1	6859	5,89E+03	1,60E+07	9,46E+07
	6861			
FAM2	7057	16,80	2,08E+06	6,21E+06
	7058			
FAM3	8090	74,27	4,95E+03	1,93E+04
	8091			
FAM4	23059	-	5,26E+20	2,80E+23
	23063			
	23064			
	23065			
FAM5	OR6	3,06E+03	4,94E+21	1,56E+25
	XR8			

Tab. 18 Valori di LR ottenuti dall'analisi degli STR e dall'analisi di frequenze aplotipiche ricavate da due differenti studi

Vengono descritti nel dettaglio i risultati ottenuti famiglia per famiglia.

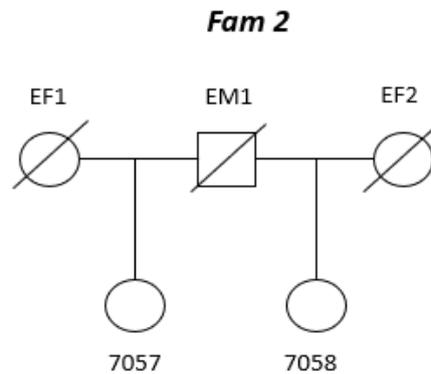
FAM1: il valore di LR ottenuto analizzando gli STR è di $5,89E+03$ e rappresenta un risultato inconclusivo in quanto è inferiore alla soglia stabilita, Al contrario, il valore di LR ottenuto dall'analisi delle frequenze aplotipiche ricavate dallo studio popolazionistico è pari a $1,60E+07$ e quello ottenuto dall'analisi del secondo database di microplotipi è di $9,46E+07$. Questi valori di LR, molto simili tra loro, permettono di supportare l'ipotesi di parentela seguendo le linee guida stabilite ⁽²⁷⁾.

Questo rappresenta un caso familiare in cui i MH hanno portato ad una attribuzione della parentela, rispetto ad un risultato inconclusivo dato dall'analisi degli STR.

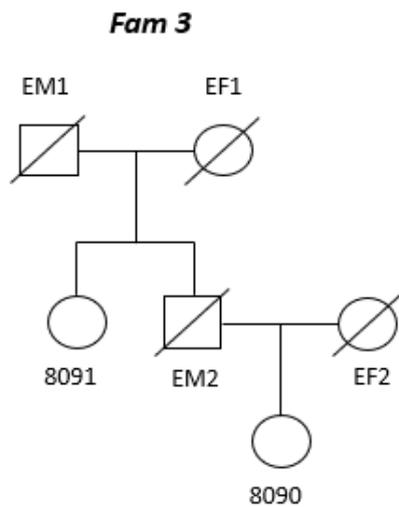


FAM2: il valore di LR ottenuto con gli STR è pari a 16,80, un valore che rientra nella definizione di “risultato inconclusivo”. Al contrario, il valore di LR ottenuto dall'analisi con i MH per entrambe le frequenze utilizzate, rientra nella definizione di attribuzione della paternità. Rispettivamente, il valore di LR è

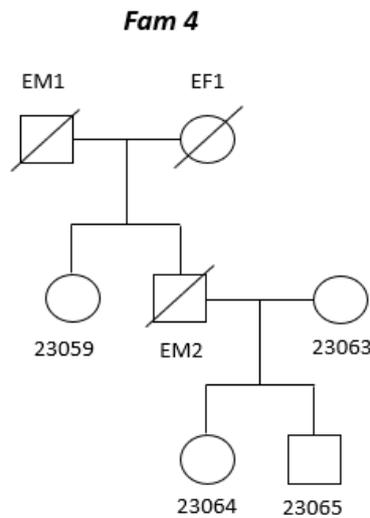
pari a $2,08E+06$ per le frequenze dello studio popolazionistico, e $6,21E+06$ per le frequenze presenti nel secondo database.



FAM3: il valore di LR ottenuto dall'analisi degli STR è pari a 74,27, quindi fornisce un risultato inconclusivo. Per quanto riguarda il valore di LR ottenuto analizzando le frequenze microaplotiche derivate dallo studio popolazionistico, questo valore si trova al di sotto della soglia indicata dalle linee guida ⁽²⁷⁾. Di conseguenza, non è possibile attribuire una paternità, come accade per gli STR. Al contrario, il valore di LR che si ricava dall'analisi delle frequenze aplotipiche derivate dal secondo database, è pari a $1,93E+04$, quindi rientra nel range di attribuzione di paternità.

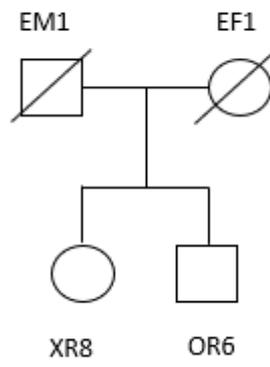


FAM4: per questa famiglia non si dispone di valori di LR derivati dall'analisi con gli STR; diversamente i valori di LR ottenuti dall'analisi dei due set di frequenze aplotipiche ha permesso di attribuire una paternità. In particolare, l'analisi con le frequenze aplotipiche, derivate dai due studi citati in precedenza, fornisce un valore di LR maggiore di quattro ordini di grandezza rispetto a quello ottenuto esaminando le frequenze aplotipiche ricavate dallo studio popolazionistico. Rispettivamente, i valori di LR sono pari a $2,80E+23$ e $5,26E+20$.



FAM5: anche in questo caso l'analisi degli STR fornisce un valore che ricade del range di inconclusività. Contrariamente, il valore di LR ottenuto dall'analisi delle frequenze aplotipiche, ottenute dallo studio popolazionistico, supportata in maniera forte l'ipotesi di paternità indicando che è $4,94E+21$ volte più probabile osservare una compatibilità se il presunto padre è il padre biologico piuttosto che il presunto padre e il figlio condividano casualmente gli alleli analizzati. Lo stesso discorso vale nel caso delle frequenze aplotipiche derivate dal secondo database e in particolare si osserva un supporto ancora maggiore per l'ipotesi di paternità considerando che il valore di LR ottenuto è pari a $1,56E+25$.

Fam 5



DISCUSSIONI

Negli ultimi anni l'utilizzo dei microplotipi nelle indagini forensi è risultato essere molto promettente. Confrontando questi marcatori con gli STR, che rappresentano il "gold standard" per l'identificazione personale, è stato dimostrato in diversi studi che la tipizzazione dei microplotipi permette di aumentare notevolmente il potere discriminativo nelle analisi forensi. ^(10,11)

Un vantaggio importante nell'utilizzo dei microplotipi è la possibilità di studiare un numero elevato di marcatori in un unico saggio mediante il Massive Parallel Sequencing. Al contrario, gli STR vengono analizzati mediante l'elettroforesi capillare, una tecnica che presenta dei limiti di detection in quanto è in grado di rilevare soltanto otto marcatori fluorescenti. A questo si aggiunge l'assenza di artefatti stutter e un basso tasso di ricombinazione, importante nella ricostruzione di relazioni parentali.

Il presente studio si è concentrato sull'applicabilità dei microplotipi nell'analisi di relazioni parentali utilizzando un pannello di 77 microplotipi con cui sono stati tipizzati soggetti italiani di uno studio popolazionistico e famiglie con relazioni parentali note, con lo scopo di valutare se la relazione

già verificata con altri marcatori, quali STR, era supportata anche con i microaplotipi. Nei casi in cui l'analisi con gli STR era risultata inconclusiva, l'obiettivo era quello di verificare se la tipizzazione con i microaplotipi potesse essere risolutiva.

I campioni dello studio popolazionistico sono stati tipizzati al fine di ottenere le frequenze alleliche della popolazione italiana da utilizzare nel calcolo biostatistico delle relazioni parentali.

I campioni relativi alle famiglie sono stati sequenziati e le reads sono state allineate con il genoma di riferimento hg19. I livelli di reads on target osservati sono buoni e il valore di uniformità indica che le letture sono distribuite in maniera uniforme sul target per tutti i campioni analizzati.

Un altro parametro calcolato è il *relative depth of coverage* che ha evidenziato la presenza di microaplotipi con bassa efficienza di amplificazione, come 6pb, seguito da D2S259, 12qA, 9qB. Ciò indica che questi marcatori possono essere soggetti a locus drop-out o perdita di alleli.

Questi risultati sono stati confermati anche da uno studio di De la Puente ⁽²²⁾, il quale fornisce come possibile soluzione il disegno di un nuovo primer per i microaplotipi underperforming.

L'analisi con il plug-in `HID_Microhaplotype_Research_PluginV1.5` ha consentito l'estrapolazione degli assetti genotipici di tutti i campioni analizzati nell'ambito dello studio familiare. I genotipi ottenuti sono stati successivamente utilizzati per il calcolo biostatistico mediante il software `Familias v.3.3.1` ⁽²⁵⁾

Prima di eseguire il calcolo sulle singole famiglie, è però necessario valutare il potere informativo dell'intero pannello. A tal fine sono state effettuate delle simulazioni considerando diversi pedigree: fratelli pieni, mezzi fratelli (o fratelli uniparentali) o cugini di primo grado. Al termine delle simulazioni viene generato un report da caricare sul software *R* in modo da costruire un grafico con la distribuzione del valore di LR su base logaritmica per i diversi pedigree, considerando soggetti imparentati e soggetti non imparentati.

Analizzando il grafico, è stato dedotto che il pannello dei microaplotipi è in grado di distinguere correttamente fratelli pieni e mezzi fratelli, rispetto a soggetti non imparentati. Per lo scenario dei cugini di primo grado è stata riscontrata una sovrapposizione delle due distribuzioni e ciò indica che in questo tipo di relazione parentale il pannello potrebbe presentare delle criticità e dei limiti che devono essere accuratamente valutati caso per caso.

Per quanto riguarda gli STR, si osserva una buona separazione delle distribuzioni di LR nel caso di fratelli pieni, ma nello scenario dei mezzi fratelli

le distribuzioni non sono separate quindi questi marcatori hanno un potere inferiore rispetto ai microplotipi nel distinguere mezzi fratelli da individui non imparentati. Da questa analisi il pannello dei microplotipi risulta essere più informativo rispetto agli STR per quanto riguarda i fratelli pieni e i mezzi fratelli.

Sono stati condotti diversi studi su pannelli di microplotipi con l'obiettivo di valutare il potere informativo del pannello sviluppato. Nel suo lavoro, Tomas ⁽²⁴⁾ ha eseguito delle simulazioni di parentela tra fratelli, mezzi fratelli, cugini di primo grado e individui non imparentati utilizzando le frequenze alleliche dei MH, STR e SNP di popolazioni danesi e groenlandesi. Ha dimostrato che il suo pannello da 74 microplotipi è informativo per quanto riguarda la distinzione tra fratelli rispetto a soggetti non imparentati nella popolazione danese, mentre con i pannelli STR e di SNP si osservano alcune sovrapposizioni. Nel caso dei groenlandesi si riscontra già una piccola sovrapposizione nelle distribuzioni LR di fratelli e soggetti non imparentati utilizzando un numero inferiore di loci MH, ma si ottiene comunque un maggiore potere informativo rispetto agli STR e agli SNP.

Nel caso dei primi cugini, in entrambe le popolazioni analizzate, si osserva che nessuno dei tre marcatori utilizzati è in grado di risolvere queste relazioni più

complesse, risultato concordante anche con il pannello da 77MH del nostro studio.

Anche De la Puente ⁽²²⁾ ha condotto uno studio riguardante il potere informativo del pannello di 118 microaplotipi (107 autosomici e 11 del cromosoma X) effettuando delle simulazioni con il software Familias nei tre scenari di parentela descritti finora: fratelli, mezzi fratelli e cugini di primo grado, usando frequenze europee. Da questo studio viene dimostrato che la distribuzione dei rapporti di verosimiglianza (LR) di individui imparentati e non imparentati non si sovrappone per fratellastri e fratelli, ma nel caso dei cugini di primo grado potrebbe essere necessario utilizzare marcatori aggiuntivi.

Questi studi forniscono dei risultati in linea con quelli ottenuti per quanto riguarda il potere informativo del nostro pannello.

Una volta validata la capacità del pannello di distinguere soggetti imparentati da quelli non imparentati, sono state effettuate le analisi di relazioni parentali nelle cinque famiglie in esame, ottenendo un valore di LR dal software Familias, v3.3.1 nel caso di microaplotipi e STR.

In base a questo valore, è stato dimostrato che il pannello dei microaplotipi è in grado di attribuire una relazione parentale nei casi in cui l'utilizzo degli STR fornisce risultati inconclusivi. In particolare, in FAM2 e FAM4, il valore di LR

ottenuto attraverso la tipizzazione dei microaplotipi supera la soglia stabilita di 10^4 e consente di attribuire la paternità, con valori di LR molto simili sia per le frequenze aplotipiche ricavate dallo studio popolazionistico sia per le frequenze recuperate da 1000K, dal lavoro (Turchi 2019) e dal database MicroHapDB ⁽²⁸⁾

Con l'analisi degli STR si ottiene un valore di LR inferiore rispetto ai microaplotipi e comunque risulta essere molto inferiore alla soglia stabilita per l'attribuzione di una parentela secondo le linee guida.

Nella FAM1 e FAM5, il valore di LR che si ottiene con gli STR è inferiore ma molto vicino alla soglia stabilita. Diversamente, il valore di LR che si ottiene dall'analisi dei due set di frequenze aplotipiche è nettamente superiore alla soglia, dimostrando la loro potenzialità nella risoluzione delle relazioni parentali.

Infine, per FAM3, il valore di LR ottenuto con gli STR porta ad una inconclusività nei risultati, mentre il valore di LR relativo all'analisi delle frequenze aplotipiche ricavate dallo studio popolazionistico è sotto la soglia di attribuzione della parentela. L'analisi condotta con le frequenze aplotipiche ricavate dagli altri due studi precedentemente nominati, supporta l'ipotesi di paternità dimostrando di essere in grado di fornire un accertamento parentale. Questo può essere considerato uno scenario di cugini di primo grado, per cui i

risultati ottenuti sono conformi a ciò che si osserva nel grafico delle distribuzioni del LR per questo tipo di relazione.

CONCLUSIONI

Grazie all'avvento del Massive Parallel Sequencing, è stato possibile esplorare dei nuovi tipi di marcatori genetici, noti come microaplotipi. Per definizione un microaplotipo corrisponde ad una piccola regione del DNA (<300 bp) costituito da 2 o più SNPs geneticamente vicini. Questi loci rappresentano uno strumento emergente che può essere applicato in diverse analisi forensi, come l'identificazione personale, la deconvoluzione di tracce miste, la ricostruzione di relazioni parentali e la stima dell'origine biogeografica.

In questo studio è stato esplorato un pannello di 77 microaplotipi per valutare la sua applicabilità in analisi di relazioni parentali. È stato prima valutato il potere informativo del pannello effettuando delle simulazioni con il software Familias, per verificare se il pannello è in grado di risolvere le relazioni di fratelli, mezzi fratelli e cugini di primo grado.

In base ai valori di LR forniti dal software, è possibile confermare l'utilità dei microaplotipi nella ricostruzione delle relazioni parentali, soprattutto in casi in cui l'utilizzo degli STR fornisce risultati inconclusivi.

Studi come questo, sono essenziali per poter validare l'utilizzo di questi marcatori in ambito forense, con l'obiettivo di ampliare ulteriormente le conoscenze di cui si dispone nell'applicabilità dei microplotipi nella ricostruzione delle relazioni parentali.

Nonostante la comunità scientifica ed i numerosi studi pubblicati negli ultimi dieci anni confermano l'applicabilità dei microplotipi nelle analisi di relazioni parentali, evidenziando i vantaggi rispetto ai classici marcatori STR, l'utilizzo dei microplotipi nella pratica forense e nelle indagini giudiziale di accertamento o disconoscimento di paternità non è ancora accettato in quanto mancano linee guida e raccomandazioni universalmente accettate.

BIBLIOGRAFIA

1. Tagliabracci A., Alessandrini F., Mazzarini L., Onofri V., Onori N., Turchi C., *Introduzione alla genetica forense. Indagini di identificazione personale e di paternità* (2010), Ancona, Springer Biomed.;
2. Butler, J. M. (2009). *Fundamentals of forensic DNA typing*. Academic press;
3. Siegel, J.; Saukko, P. *Encyclopedia of Forensic Sciences*, 2nd ed.; Academic Press: Cambridge, MA, USA, 2014
4. Lee HY, Park MJ, Yoo JE, Chung U, Han GR, Shin KJ. *Selection of twenty-four highly informative SNP markers for human identification and paternity analysis in Koreans*. *Forensic Sci Int*. 2005 Mar 10;148(2-3):107-12.
5. The International Project HapMap (2003); *Nature*, Volume 426
6. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D;

- International SNP Map Working Group. *A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms*. Nature. 2001 Feb 15;409(6822):928-33
7. John M. Butler, Michael D. Coble, Peter M. Vallone *STRs vs. SNPs: thoughts on the future of forensic DNA testing*, Forensic Sci Med Pathol (2007) 3:200–205
 8. Juan J. Sanchez; Chris Phillips; Claus Børsting¹; Kinga Balogh; Magdalena Bogus; Manuel Fondevila; Cheryl D. Harrison; Esther Musgrave-Brown, Antonio Salas; Denise Syndercombe-Court; Peter M. Schneider; Angel Carracedo; Niels Morling; *A multiplex assay with 52 single nucleotide polymorphisms for human identification*, Electrophoresis 2006, 27;
 9. Amorim A, Pereira L. *Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs*. Forensic Sci Int. 2005 May 28;150(1):17-21
 10. K.K. Kidd, A.J. Pakstis, W.C. Speed, R. Lagace, J. Chang, S. Wootton, N. Ihuegbu, *Microhaplotype loci are a powerful new type of forensic marker*, Forensic Science International: Genetics Supplement Series, Volume 4, Issue 1, 2013, e123-e124, ISSN 1875-1768,

11. Oldoni F, Kidd KK, Podini D. *Microhaplotypes in forensic genetics* 2019. *Forensic Sci Int Genet.* Jan;38:54-69
12. K. Kidd; *Proposed nomenclature for microhaplotypes* (2016); *Human Genomic*
13. Jing Zhu, Bing Long, Shengqiu Qu, Lu Yin, Wang He, Jiong Mao, Hui Wang, Lin Zhang, Bo Jin, Dan Chen, Weibo Liang; *A new proposed nomenclature for microhaplotypes*, *Forensic Science International: Genetics Supplement Series* 2019, Volume 7, Issue 1, 813-815,
14. Kidd et al *Criteria for selecting microhaplotype mixture detection and deconvolution* 2015; *Invest. Genet.*; 6: 1,
15. Kidd KK, Pakstis AJ. *State of the Art for Microhaplotypes* 2022, *Genes* (Basel). 24;13(8):1322,
16. Kidd KK, Pakstis AJ, Gandotra N, Scharfe C, Podini D. *A multipurpose panel of microhaplotypes for use with STR markers in casework*; *Forensic Sci Int Genet* 2022.; 60:102729,
17. Xue et al; *An overview of SNP-SNP microhaplotypes in the 26 populations of the 1000 Genomes Project* 2022, *International Journal of Legal Medicine* 136:1211–1226
18. Bennett L, Oldoni F, Long K, Cisana S, Madella K, Wootton S, Chang J, Hasegawa R, Lagacé R, Kidd KK, Podini D. *Mixture deconvolution by*

- massively parallel sequencing of microhaplotypes*. Int J Legal Med. 2019 May;133(3):719-729
19. Wu R, Chen H, Li R, Zang Y, Shen X, Hao B, Wang Q, Sun H. *Pairwise kinship testing with microhaplotypes: Can advancements be made in kinship inference with these markers?* Forensic Sci Int. 2021;325:110875
20. Zou X, He G, Liu J, Jiang L, Wang M, Chen P, Hou Y, Wang Z. *Screening and selection of 21 novel microhaplotype markers for ancestry inference in ten Chinese subpopulations* 2022, Forensic Sci Int Genet. May;58:102687
21. Brissenden JE, Roscoe JM, Evsanaa B, Togtokh A, Paschou P, Grigorenko EL, Gurwitz D, Wootton S, Lagace R, Chang J, Speed WC, Kidd KK. *Ancestry inference of 96 population samples using microhaplotypes* 2018. Int J Legal Med.;132(3):703-711
22. M. de la Puente, C. Phillips, C. Xavier, J. Amigo, A. Carracedo, W. Parson, M.V. Lareu, *Building a custom large-scale panel of novel microhaplotypes for forensic identification using MiSeq and Ion S5 massively parallel sequencing systems*, Forensic Science International: Genetics 45 (2020) 102213
23. Turchi C., Melchionda F., Pesaresi M., Tagliabracci A., *Evaluation of a microhaplotypes panel for forensic genetics using massive parallel*

sequencing technology, (2019) Forensic Science International: Genetics.

41

24. Tomas, C.; Rodrigues, P.; Jønck, C.G.; Barekzay, Z.; Simayijiang, H.; Pereira, V.; Børsting, C. *Performance of a 74-Microhaplotype Assay in Kinship Analyses*. Genes 2024, 15, 224
25. Kling, D., Tillmar, A. O., & Egeland, T. (2014). *Familias 3-Extensions and new functionality*. Forensic Science International: Genetics, 13, 121-127
26. Turchi C, (2023) - Genetica Forense, Università Politecnica delle Marche, inedito
27. Società Italiana di Genetica Forense (SIGU) - *Analisi di accertamento parentale, aggiornamento 2022*, https://sigu.net/wp-content/uploads/2022/11/2023_05_02_AccertamentoParentale_2022_Rev03.pdf)
28. Standage, D.S.; Mitchell, R.N. *MicroHapDB: A Portable and Extensible Database of All published Microhaplotype Marker and Frequency Data*. Front. Genet. 2020, 11, 781.
29. Pakstis A.J., Gandotra N., Speed W.C., Murtha M., Scharfe C., Kidd K.K. *The population genetics characteristics of a 90 locus panel of microhaplotypes*. Hum. Genet. 2021;140:1753–1773

30. Oldoni F, Bader D, Fantinato C, Wootton SC, Lagacé R, Kidd KK, Podini D. *A sequence-based 74plex microhaplotype assay for analysis of forensic DNA mixtures*. *Forensic Sci Int Genet*. 2020 Nov;49:102367
31. Turchi C, Melchionda F, Pesaresi M, Ciarimboli E, Bini C, Fattorini P, et al. Exploring the usefulness of microhaplotypes in forensic identification using massive parallel sequencing technology (2020), *Minerva Medicoleg*; 140:26-36
32. C. Turchi, M. Pesaresi, A. Tagliabracci A microhaplotypes panel for forensic genetics using massive parallel sequencing *Forensic Sci. Int.: Genet. Suppl. Ser.*, 6 (2017), pp. e117-e118
33. Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, *Briefings in Bioinformatics*, Volume 14, Issue 2, March 2013, Pages 178–192
34. Kejun Liu, Spencer V. Muse, PowerMarker: an integrated analysis environment for genetic marker analysis, *Bioinformatics*, Volume 21, Issue 9, May 2005, Pages 2128–2129