



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

**Corso di Laurea Magistrale**

**Biologia Molecolare e Applicata**

**ANALISI DI NUOVI MARCATORI GENETICI PER  
L'IDENTIFICAZIONE INDIVIDUALE: SVILUPPO E VALIDAZIONE DI  
UN SET DI MICROAPLOTIPI DEL DNA MEDIANTE TECNOLOGIA  
MPS ED APPLICAZIONI IN GENETICA FORENSE**

**ANALYSIS OF NEW GENETIC MARKERS FOR INDIVIDUAL  
IDENTIFICATION: DEVELOPMENT AND VALIDATION OF A SET OF  
DNA MICROHAPLOTYPES USING MPS TECHNOLOGY AND  
APPLICATIONS IN FORENSIC GENETICS**

Tesi di Laurea Magistrale  
di: Riccardo Lamonica Maria

Relatore:  
Chiar.ma Prof.ssa  
Chiara Turchi

Correlatore:  
Magg. Inv. Sc. Fabiano Gentile

**Sessione Febbraio 2024  
Anno Accademico 2022/2023**



## INDICE

INTRODUZIONE.....	6
1.1 La genetica forense .....	6
1.2 Polimorfismi del Dna.....	7
1.2.1 Polimorfismi di sequenza: SNP .....	9
1.3 I Microaplotipi.....	12
1.3.1 Nomenclatura dei microaplotipi.....	13
1.4 Vantaggi e Svantaggi dei microaplotipi .....	15
1.6 Applicazione dei microaplotipi in genetica forense .....	19
Microaplotipi nell'identificazione personale .....	21
Microaplotipi nella deconvoluzione di campioni misti.....	22
Microaplotipi nell'identificazione parentale .....	23
Microaplotipi nell'ascendenza biogeografica .....	24
1.7 Tecniche di rilevamento dei microaplotipi: Massive Parallel Sequencing .....	24
1.7.1 Tecnologia Ion Torrent .....	26
1.8 Precision ID SNP Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System .....	36
SCOPO DELLO STUDIO .....	37
MATERIALI E METODI.....	38
3.1 Selezione dei loci d'interesse .....	38
3.3 Campioni utilizzati per la validazione del pannello .....	41
3.4 Estrazione e quantificazione del DNA .....	43
3.4.1 Estrazione del DNA con QIAamp® DNA Micro .....	43

3.4.2 Estrazione del DNA con EZ1 ® DNA Investigator KIT .....	46
3.4.3 Quantificazione con E-Gel™ Agarose Gels with SYBR Safe™ ....	49
3.4.4 Quantificazione con PowerQuant® System.....	50
3.5 Massive Parallel Sequencing.....	53
Amplificazione delle regioni target e preparazione delle librerie .....	54
Parziale digestione degli ampliconi .....	57
Ligazione dei barcode .....	57
Purificazione delle librerie .....	58
Eluizione delle librerie .....	60
Quantizzazione delle librerie.....	60
3.6 Diluizione, raggruppamento delle librerie e posizionamento sullo ION CHEF .....	62
3.7 Preparazione del template su HID ION CHEF.....	64
3.8 Sequenziamento con HID Ion GeneStudio S5™ System.....	67
3.9 Analisi dei dati.....	69
RISULTATI.....	70
4.1 Selezione dei marker .....	70
4.2 Risultati della quantificazione delle librerie .....	71
4.3 Risultati analisi delle prestazioni del sequenziamento .....	73
4.4 Analisi del Coverage .....	79
4.5 Risultati della genotipizzazione.....	82
DISCUSSIONI.....	92
CONCLUSIONI.....	98



BIBLIOGRAFIA .....100

## Capitolo 1

### **INTRODUZIONE**

#### ***1.1 La genetica forense***

La genetica forense è un ramo della scienza forense che ha il compito di caratterizzare tracce biologiche per assegnare una corrispondente identità, conducendo analisi su diversi tessuti biologici, come ad esempio il sangue, la saliva, frammenti ossei, sperma, frammenti di unghie, denti, capelli. Dai campioni viene estratto il DNA che sarà analizzato al fine di ottenere un profilo genotipico. Il laboratorio di genetica forense fornisce un valido supporto tecnologico e scientifico al sistema giudiziario per molteplici ragioni:

- Individuare l'identità di autori di reati;
- Mettendo in correlazione i dati genetici di un sospettato con il materiale biologico;
- L'attribuzione individuale di resti umani sconosciuti (per esempio in disastri di massa);
- Verifica dei rapporti di parentela (indagini di paternità);
- L'attribuzione individuale di resti umani sconosciuti;
- Esonerare o identificare un sospettato.

A oggi la genetica forense utilizza le moderne tecniche di biologia molecolare, anche se il primo timido inizio delle analisi forensi in campo biologico si ebbe nei primi anni del secolo scorso, in cui l'analisi forense si basava sulla caratterizzazione del polimorfismo dell'AB0 che definiscono i gruppi sanguigni studiati da Karl Landsteiner, i quali consentivano di classificare un individuo sulla base delle diverse caratteristiche antigeniche <sup>(1)</sup>.

### ***1.2 Polimorfismi del Dna***

Due individui presi casualmente nella popolazione, condividono il 99,7% del materiale genetico. Paradossalmente solo uno 0.3% differisce un individuo dall'altro, ed è infatti in tali regioni che sono presenti i marcatori d'elezione per lo studio della variabilità <sup>(2)</sup>.

Le regioni studiate in genetica forense vengono definite polimorfismi. 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, è chiamato un allele con frequenza al di sotto dell'1%. Chiaramente, poiché le frequenze alleliche spesso variano tra le popolazioni, una variante per una popolazione potrebbe essere un polimorfismo per un'altra <sup>(3)</sup>.

Esistono due tipologie di polimorfismi del DNA:

- polimorfismi di lunghezza: consistono in cambiamenti nel numero di sequenze del DNA disposte in tandem che possono trovarsi in posizioni cromosomiche differenti. Questi sono classificati, in base alla taglia delle loro unità ripetitive in: *microsatelliti*, conosciuti anche con il nome di STRs (*short tandem repeats*), sono costituiti da sequenze di DNA lunghe 2-6 bp e ripetute in tandem numerose volte. Al giorno d'oggi, gli STR sono ancora i marcatori genetici forensi predominanti per i test di identità e l'analisi di parentela e i kit STR convalidati sono facilmente disponibili e utilizzati di routine nella maggior parte dei laboratori forensi di tutto il mondo <sup>(2)</sup>. Seguono i *minisatelliti*, costituiti da unità di 8-100 bp ripetute dalle 5 alle 1.000 volte e infine abbiamo i *satelliti* sono larghe regioni ripetute in tandem che vanno da centinaia di kilobasi a megabasi e sono composte da unità ripetitive di diverse dimensioni <sup>(2)</sup>.
- polimorfismi di sequenza: comunemente chiamati SNPs (*single nucleotide polymorphisms*), comprendono le sostituzioni di una base, in cui una base azotata viene sostituita da un'altra, e l'inserimento e/o la delezione di una base singola (InDel), in cui una base viene aggiunta o rimossa determinando così una variazione della lunghezza del DNA <sup>(41)</sup>.

(a) Sequence polymorphism

-----AGACTAGACATT-----  
-----AGATTAGGCATT-----

(b) Length polymorphism

-----**(AATG)**(AATG)(AATG)-----  
                    3 repeats  
-----  (AATG)(AATG)-----  
                    2 repeats

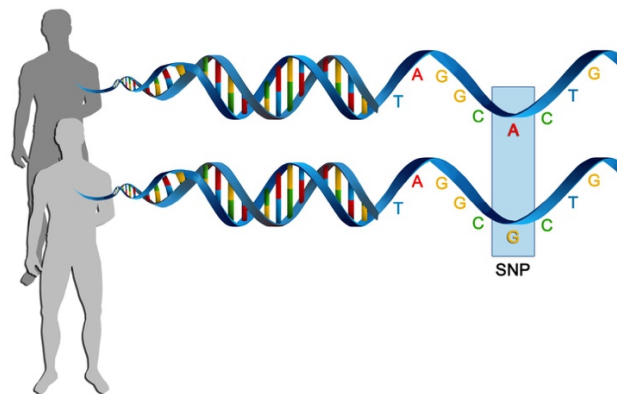
*Fig. 1 Rappresentazione di un polimorfismo di sequenza (a) e di un polimorfismo di lunghezza(b). (2)*

### ***1.2.1 Polimorfismi di sequenza: SNP***

I Single Nucleotide Polymorphisms (SNP) sono polimorfismi di sequenza in cui l'evento di mutazione ha creato una variazione a carico di una singola posizione nucleotidica. Si tratta di polimorfismi derivati da una errata incorporazione di basi di nucleotidi durante la replicazione, oppure generati da modificazioni chimiche e danni fisici a cui è sottoposta la molecola di DNA <sup>(3)</sup>. Esistono milioni di SNP per individuo, questo potrebbe essere molto vantaggioso per poter differenziare gli individui<sup>(2)</sup>. Sono considerati dalla comunità forense come potenziali marcatori genetici, in quanto particolarmente adatti per l'analisi di campioni di DNA altamente degradati, grazie alle dimensioni complessivamente ridotte degli ampliconi. Si tratta di marcatori bi-allelici, ciò implica l'origine di sole 2 forme alleliche che combinandosi tra loro, portano alla formazione di 3 possibili genotipi, costituendo di per se uno scarso potere

informativo. Vari studi hanno dimostrato che in media per poter arrivare a identificare un soggetto bisognerebbe analizzare un minimo di 50 marcatori SNPs rispetto ai soli 13 marcatori STR <sup>(4)</sup>.

Al contrario invece sono caratterizzati da un tasso di mutazione più basso <sup>(3)</sup>, di circa  $10^{-8}$  - $10^{-9}$ , rispetto agli STR rendendoli molto utili in vari ambiti d'analisi distanti dall'identificazione personale, ma comunque altrettanto importanti per spunti investigativi come ad esempio la previsione dell'ascendenza biogeografica o dei caratteri fenotipici. Proprio in merito a quest'ultima applicazione è stato messo a punto un pannello di SNP <sup>(5)</sup> che consente di ottenere una previsione sul colore degli occhi di un soggetto ignoto a partire dal DNA estratto da una traccia.



*Fig.1 Illustrazione di un esempio di SNP <sup>(37)</sup>*

Nel 2017 è stato fondato il consorzio VISible Attributes trough GENomics (VISAGE) con l'obiettivo principale di sviluppare e validare strumenti che

permettano di ottenere una più completa ricostruzione delle caratteristiche fenotipiche di un soggetto <sup>(6)</sup>.

A tal proposito è stato messo a punto un pannello contenete all'incirca 524 SNPs in grado di fornire informazioni fenotipiche aggiuntive quali la presenza/assenza di lentiggini, colore delle sopracciglia, la predisposizione alla calvizie, la tipologia dei capelli, ascendenza biogeografica e l'età <sup>(7)</sup>.

Le principali categorie di SNPs oggetto di studio sono:

- LISNPs (*Lineage Informative SNPs*): SNPs in linkage disequilibrium, che formano un blocco aplo tipico e funzionano come marcatori multiallelici in grado di fornire informazioni in analisi di parentela, in quanto vengono trasmessi in modo uniparentale <sup>(8)</sup>.
- AISNPs (*Ancestry Informative SNPs*): gruppo di SNPs che fornisce una previsione su basi probabilistiche dell'ascendenza geografica di un individuo; grazie al tasso di mutazione più basso rispetto agli STR, gli SNPs vengono tramandati inalterati da una generazione a quella successiva, sedimentandosi in modo specifico e caratteristico per ogni determinata popolazione <sup>(9)</sup>.
- PISNPs (*Phenotype Informative SNPs*): forniscono una predizione di fenotipo su un individuo ignoto, ad esempio legata al colore di pelle, occhi e capelli. Oggi sono disponibili sistemi per identificare

caratteristiche fenotipiche quali addirittura la presenza/assenza di lentiggini, la tipologia di capello liscio/riccio, la predisposizione alla calvizie e tante altre <sup>(10)</sup>.

### ***1.3 I Microaplotipi***

Nel 2013 il professore Kennet Kidd ha introdotto il concetto di *microaplotipo* (MH) e lo ha applicato per la prima volta in campo della genetica forense <sup>(11)</sup>.

I microaplotipi, proposti come nuovi marcatori genetici, corrispondono a una piccola regione del DNA (<300 bp) caratterizzati dalla presenza di due o più SNP.

Questi nuovi marcatori genetici hanno suscitato un crescente interesse nella genetica forense per svariati motivi tra cui il possedere un'alta variabilità e informatività, caratteristica dovuta proprio alla loro struttura composta da più SNP e ciò si traduce come un'alta variabilità genetica all'interno di una popolazione. Questo li rende molto informativi e utili nell'identificazione individuale, uguagliabile a quello degli STR, marcatori d'elezione in genetica forense.

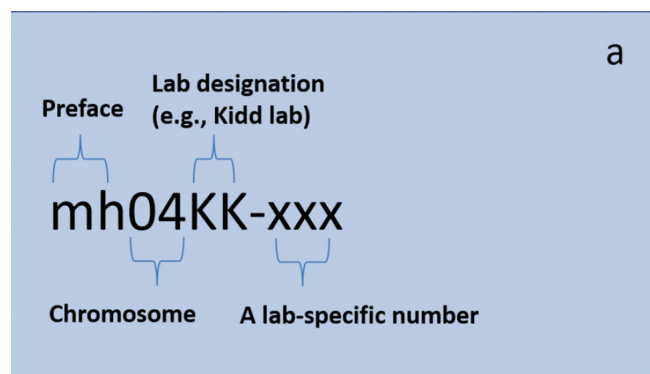
Inoltre le sequenze dei microaplotipi sono relativamente corte, il che facilita la loro amplificazione anche da frammenti di DNA degradato o danneggiato, situazione comune nei campioni forensi.



### ***1.3.1 Nomenclatura dei microaplotipi***

In passato i microaplotipi venivano classificati utilizzando la codifica “rs” relativo a ciascuno degli SNP che lo componevano, definiti e catalogati in dbSNP. Con il tempo però questo metodo risultò scomodo da applicare, motivo per il quale ciascun laboratorio di ricerca iniziò ad utilizzare metodi propri per classificare i microaplotipi.

Fu Kennet Kidd a proporre un metodo standardizzato per la nomenclatura dei microaplotipi <sup>(11)</sup> che seguisse le raccomandazioni della HUGO Gene *Nomenclature Committee* <sup>(12)</sup>. La nomenclatura proposta è riportata in *Fig.2*



*Fig. 2 Illustrazione grafica delle regole di nomenclatura proposte da Kenneth Kidd <sup>(11)</sup>*

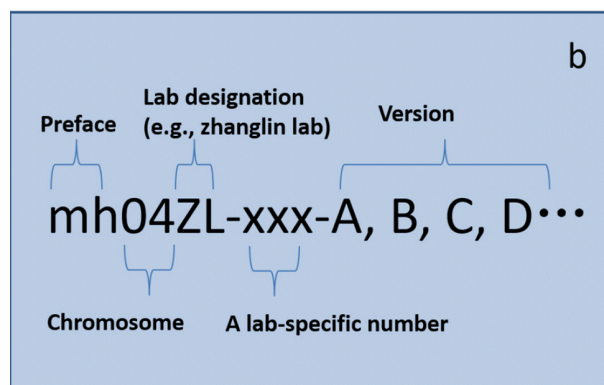
Questa nomenclatura implica:

- Una radice “mh” (scritta in maiuscolo o minuscolo) ad identificare che si tratta di un microaplotipo;

- Seguono due numeri, stanno a identificare il numero del cromosoma su cui si trova il microplotipo (da 01 a 22 per gli autosomici e 0X o 0Y per i cromosomi sessuali);
- Abbiamo poi un codice da 2 a 4 lettere che segue (es. KK) identifica il laboratorio in cui il microplotipo è stato scoperto e caratterizzato;
- Il codice termina con un numero a 3 cifre che rappresenta la combinazione di due o più SNP. Quel codice è specifico per quel determinato microplotipo.

Ad esempio, il simbolo del microplotipo mh01KK-001 si riferisce a un locus sul cromosoma 1 pubblicato dal Kidd Lab (KK-) come #001.

Recentemente, un gruppo di ricercatori asiatici, ha proposto una modifica alla nomenclatura di Kidd (*Fig.3*) <sup>(14)</sup>.



*Fig.3* Illustrazione grafica delle regole di nomenclatura proposte da J. Zhu, et al.<sup>(14)</sup>

La modifica consiste nell'aggiunta di una lettera alla fine del numero indicante gli SNP del microplotipo. In questo caso la lettera "A" indica che questo gruppo di SNPs è stato definito per primo, a cui potrebbero essere aggiunti dei nuovi, il che si traduce in un polimorfismo più elevato.

I primi dati che riguardano le frequenze alleliche di popolazione dei microplotipi sono stati annotati nel database ALFRED ( ALlele FREquency) <sup>(15-16)</sup> che nasce come un database di varianti e polismorfismi, ma che contiene anche una buona percentuale di microplotipi, anche se ad oggi non viene più aggiornato. Recentemente è stato introdotto un nuovo database, il MicroHapDB contenete le informazioni sui marcatori microhap e sulla frequenza della popolazione, provenienti da numerosi articoli di ricerca pubblicati. <sup>(47)</sup>

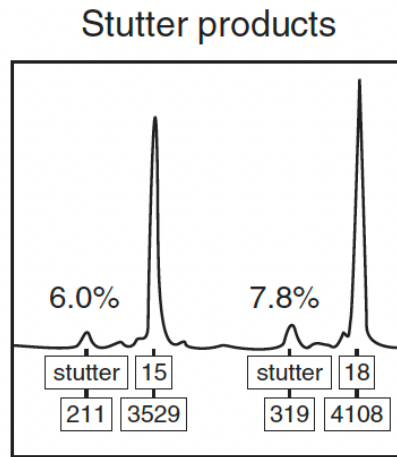
#### ***1.4 Vantaggi e Svantaggi dei microplotipi***

Nella pratica forense l'analisi dei marcatori genetici del DNA è attualmente lo strumento più prezioso per l'identificazione individuale. Gli STR in quanto marcatori genetici del DNA più comunemente utilizzati, svolgono un ruolo importante nell'identificazione forense dell'individuo e della parentela; tuttavia, quando si incontrano campioni biologici altamente degradati, gli STR spesso non riescono a fornire informazioni sufficienti.

Gli SNP possiedono diverse caratteristiche che li rendono marcatori utili, non solo nell'analisi di campioni di DNA degradato, ma anche per la previsione di caratteristiche fenotipiche umane <sup>(30)</sup>, più efficaci degli STR per svariati motivi: tasso di mutazione più basso ( $10^{-8}$  - $10^{-9}$  gli SNP e  $10^{-3}$  per gli STR)<sup>(18)</sup> ; dimensione dell' amplicone più piccola (50–150 bp); maggiore presenza nel genoma umano (circa 1 su 1000 bp, milioni per individuo) <sup>(40)</sup>.

Un vantaggio dei microaplotipi è che essendo costituiti da più SNP, possono fornire informazioni più dettagliate rispetto al singolo SNP.

L'introduzione dei microaplotipi in campo forense ha evidenziato molti aspetti vantaggiosi: prima fra tutti i microaplotipi, non essendo costituiti da ripetizioni in tandem, non producono *stutter*, ovvero frammenti di DNA in cui manca una ripetizione rispetto al frammento originale a causa di un errore prodotto dalla DNA polimerasi, cosa invece molto frequente con gli STR. Questo aspetto risulta essere di grande utilità soprattutto quando si ha a che fare con tracce miste <sup>(17)</sup>



*Fig.4 Rappresentazione di un esempio di stutter <sup>(3)</sup>*

Gli alleli dei microplotipi presentano la stessa lunghezza, cosa che non osserviamo negli STR che invece possono anche essere costituiti da lunghezze differenti che possono variare fino a 100 nucleotidi <sup>(11)</sup>, portando ad un'amplificazione preferenziale dell'allele corto rispetto all'allele lungo dello stesso contribuente, complicando l'attribuzione allelica nelle miscele.

Altro importante vantaggio è che i microplotipi presentano un tasso di ricombinazione molto basso, dovuto al fatto che gli SNP sono molto vicini tra loro. Importante per l'identificazione individuale è anche il livello di eterozigotità, tutti gli alleli hanno la stessa frequenza.

Tra gli svantaggi dei microplotipi troviamo però che essendo brevi sequenze di DNA, presentano meno alleli per cui è necessario un pannello più grande di marcatori per ottenere un potere di discriminazione comparabile agli STR <sup>(11)</sup>.

Inoltre, per quanto riguarda i microplotipi, bisogna considerare che le frequenze alleliche variano significativamente tra le popolazioni; per cui conoscere le frequenze alleliche delle popolazioni è necessario per il calcolo *Random Match Probability* (RMP) o del *Likelihood Ratio* (LR).

Un ulteriore inconveniente riguarda le operazioni di laboratorio, infatti per l'analisi dei microplotipi viene utilizzata la MPS che risulta essere però una tecnica costosa e laboriosa, rispetto all'elettroforesi capillare usata per la rilevazione degli STR.

### ***1.5 Criteri di selezione dei microplotipi***

Esistono diverse fonti sul web per lo screening dei loci candidati tra cui HapMap e HGDP, ma anche i dati derivati dallo studio 1000 genomi umani.<sup>(11)</sup>

I criteri per selezionare i microplotipi ottimali sono due: *Effective Number of Alleles* (Ae) e l'*Informativeness* (In)<sup>(30)</sup>.

Ae è definita come “il numero di alleli neutri frequenti, basato sulle dinamiche di popolazione che produce la stessa eterozigosità del locus con numerosi alleli a frequenze diverse”<sup>(30)</sup>.

Si calcola come il reciproco dell'omozigosi (*Fig.5*):

$$1 / \sum p_i^2$$

Fig. 5 Formula per il calcolo del valore  $A_e$ ; dove  $p_i$  è uguale alla frequenza dell'allele  $i$  e la somma è su tutti gli alleli nel locus. <sup>(30)</sup>

Maggiore sarà questo parametro e più potente sarà l'informazione ottenuta da quell'aplotipo. Inoltre valori di  $A_e$  più elevati sono correlati ad una maggiore capacità di un locus di rilevare un campione di origine mista.

Altro criterio importante è l'informativennes, denominata anche *Informatività di Rosenberg*, secondo il quale fornisce una stima delle variazioni delle frequenze alleliche osservate tra le diverse popolazioni. Di conseguenza è importante per la biogeografic ancestry: elevati valori di questo parametro indicano che l'allele è più frequente in certe popolazioni rispetto che in altre.

Il locus ideale avrà valori alti sia di  $A_e$  che di  $In$ , il che significa che ogni popolazione presenta molti alleli e alcuni variano in frequenza tra le popolazioni.

### ***1.6 Applicazione dei microaplotipi in genetica forense***

I microaplotipi sono un nuovo strumento promettente e versatile che può essere utilizzato per estrarre simultaneamente molti tipi di informazioni forensi

rilevanti da uno o più individui che contribuiscono a un campione di DNA. I set di microaplotipi possono essere ottimizzati in modo da risultare utili per: identificazione individuale, relazioni tra individui, informazioni sul fenotipo, informazioni sull'origine biogeografica, deconvoluzione di campioni misti, riconoscimento di cadaveri nei disastri di massa, identificazione di persone scomparse e identificazione DNA non umano <sup>(11)</sup>.

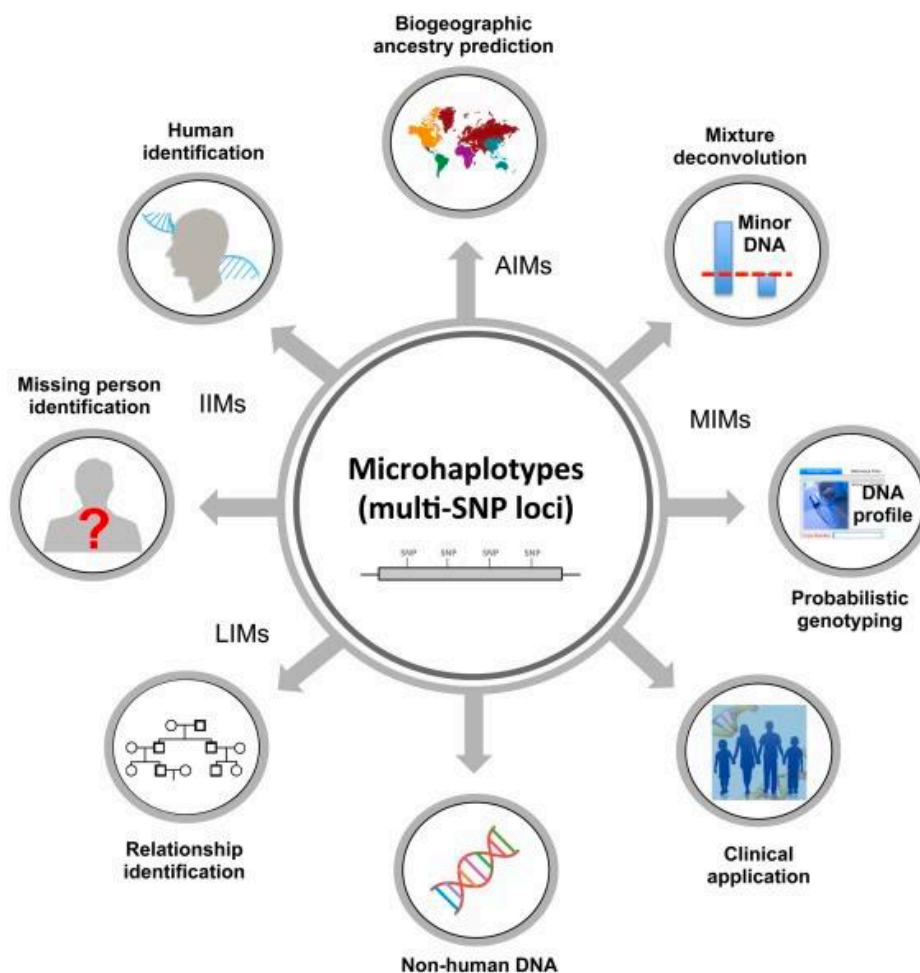


Fig.6 Panoramica delle potenziali applicazioni dei microaplotipi in campo forense <sup>(11)</sup>.



### Microaplotipi nell'identificazione personale

Il primo comune utilizzo dei microaplotipi è stato il voler determinare se un profilo di prova corrispondesse al profilo di un sospettato. Una misura forense, la probabilità di corrispondenza casuale (RMP), (nota anche come probabilità di corrispondenza cumulativa, CMP), rappresenta la probabilità che due individui presi casualmente nella popolazione condividano lo stesso profilo genetico. Il calcolo probabilità di corrispondenza casuale (RMP) viene spesso utilizzata nei casi penali per sottolineare quanto sia improbabile che qualcun altro abbia lo stesso genotipo dell'imputato <sup>(21)</sup>, infatti più basso è l'RMP, meno probabilità c'è che la corrispondenza sia un evento casuale ed è relativamente più probabile che la corrispondenza indichi che le prove provengono dal sospettato. <sup>(19)</sup>

Un approccio è stato presentato da Kidd et al. <sup>(20)</sup>, presentando un pannello multiuso di 90 MH. I 90 loci selezionati, per 79 popolazioni avevano numero effettivo in media  $A_e > 5,08$ . Da tale studio si arrivò alla conclusione che alti livelli di individualizzazione misurati dalla probabilità di corrispondenza casuale (RMP) sono una conseguenza degli elevati valori  $A_e$  dei loci. Successivamente, sempre partendo dal pannello di 90MH, Kidd et al. <sup>(20)</sup> selezionò i 24 MH con il più alto numero effettivo di  $A_e$ . Questi furono confrontati con un pannello STR aumentato di 24 loci comunemente usati nelle

pratiche forensi. Vengono confrontate, in particolare, le distribuzioni Ae e l'informatività di ascendenza (In). È stato dimostrato che il pannello MH ha una migliore individualizzazione e distinzione della popolazione rispetto ai CODIS STR, risultando essere migliori anche per le analisi delle miscele rispetto agli STR. <sup>(21)</sup>

#### Microplotipi nella deconvoluzione di campioni misti

Mediante un recente studio sono state valutate le prestazioni di un pannello di 74 MH tramite la piattaforma *Ion S5*. I dati emersi dallo studio hanno confermato la possibilità di identificare e distinguere i due profili, di due persone in campioni misti <sup>(29)</sup>. Inoltre questo pannello ha dimostrato di essere efficace non solo per l'analisi delle miscele, ma anche per l'individualizzazione e per l'ascendenza biogeografica. Un aspetto importante che è emerso nei vari studi condotti dai vari ricercatori, è stato che un valore Ae elevato è particolarmente rilevante per una migliore risoluzione delle miscele. A dare conferma di tale aspetto c'è il lavoro condotto da Kidd et al. <sup>(21)</sup>, nel quale è stato proposto un pannello di 24 MH con il più alto valore di Ae. Dimostrando che maggiore è l'intervallo Ae e più alta sarà la probabilità di vedere 4 alleli in un locus in una miscela di due persone.

### Microaplotipi nell'identificazione parentale

L'identificazione della parentela ricopre un ruolo importante nell'analisi di relazioni biologiche all'interno della famiglia, ma anche nell'identificazione delle vittime di disastri o delle persone scomparse. Attualmente l'identificazione di relazioni familiari viene eseguita utilizzando gli STR standard, ma anche utilizzando dei marcatori classici, come ad esempio i gruppi sanguigni <sup>(19)</sup>. Nei test di relazione che coinvolgono DNA degradato, come l'identificazione di persone scomparse <sup>(22)</sup>, vittime di disastri o resti riesumati, gli ampliconi STR più lunghi generalmente falliscono, mentre i microaplotipi essendo costituiti strutturalmente da ampliconi più corti possono risultare più risolutivi. In più i microaplotipi, rispetto al singolo SNP, risultano essere più polimorfici e hanno un potere discriminante più elevato, il che riduce sostanzialmente il numero di loci richiesto in un pannello. Diversi studi hanno dimostrato l'utilità dei microaplotipi nella misurazione della parentela, come il pannello proposto da un gruppo di ricercatori asiatici <sup>(23)</sup>, in grado di separare le coppie genitore-figlio fino ai parenti di secondo grado.

Recentemente è stato sviluppato e convalidato un pannello di 74 MH <sup>(24)</sup> per l'analisi di parentela, in quattro gruppi di una popolazione statunitense e i dati ottenuti confrontati con i dati ottenuti utilizzando gli STR.

I risultati ottenuti <sup>(23-24)</sup> supportano ulteriormente l'uso promettente dei microaplotipi nelle applicazioni forensi.

### *Microaplotipi nell'ascendenza biogeografica*

Dal 2006 molti pannelli di polimorfismi sono stati proposti per permettere l'individualizzazione dell'ascendenza biogeografica<sup>(19-9)</sup>. Essendo quindi i microaplotipi costituiti da più SNP, sono stati introdotti diversi pannelli di MH che abbiano come caratteristica principale un diverso Ae tra le popolazioni.

I primi studi che utilizzavano MH mostravano gli stessi modelli di relazione mostrati con i pannelli SNP <sup>(19-25)</sup>.

Tuttavia i pannelli più recenti consentono di ottenere raggruppamenti più dettagliati delle popolazioni come nel caso del pannello da 90 MH <sup>(20)</sup>. Tra i numerosi recenti studi sulle relazioni tra le popolazioni, c'è quello proposto da Zou et al. <sup>(26)</sup> i quali hanno sviluppato un nuovo pannello che ha fornito un elevato potere discriminatorio nella differenziazione tra le popolazioni continentali.

### ***1.7 Tecniche di rilevamento dei microaplotipi: Massive Parallel Sequencing***

Il *sequenziamento Sanger* rappresenta tutt'ora la metodologia “gold standard” per il sequenziamento del DNA <sup>(2)</sup> e per fornire risultati utilizzabili in ambito

giudiziario, tuttavia il limite principale di questa tecnologia è l'incapacità di determinare le relazioni cis/trans tra i singoli SNP, cioè la FASE, ostacolo superato con l'introduzione nelle pratiche forensi della *Massive Parallel Sequencing (MPS)*. Il nuovo millennio ha portato a un grande avanzamento tecnologico per quanto riguarda il sequenziamento del DNA: la *Next Generation Sequencing*, anche nota come *Massive Parallel Sequencing (MPS)*. Questa tecnologia innovativa ha consentito di poter effettuare ricerche approfondite, tramite il sequenziamento rapido di milioni di frammenti di DNA simultaneamente, a differenza del sequenziamento Sanger con il quale si analizza una reazione di sequenziamento alla volta (*Fig.7*), fornendo informazioni complete sulla struttura del genoma, sulle variazioni genetiche e sulle modifiche epigenetiche. Questo ha portato a una riduzione dei costi, ma anche a una riduzione dei tempi di attesa per il sequenziamento di singoli geni, pannelli di geni, esoma ed intero genoma.

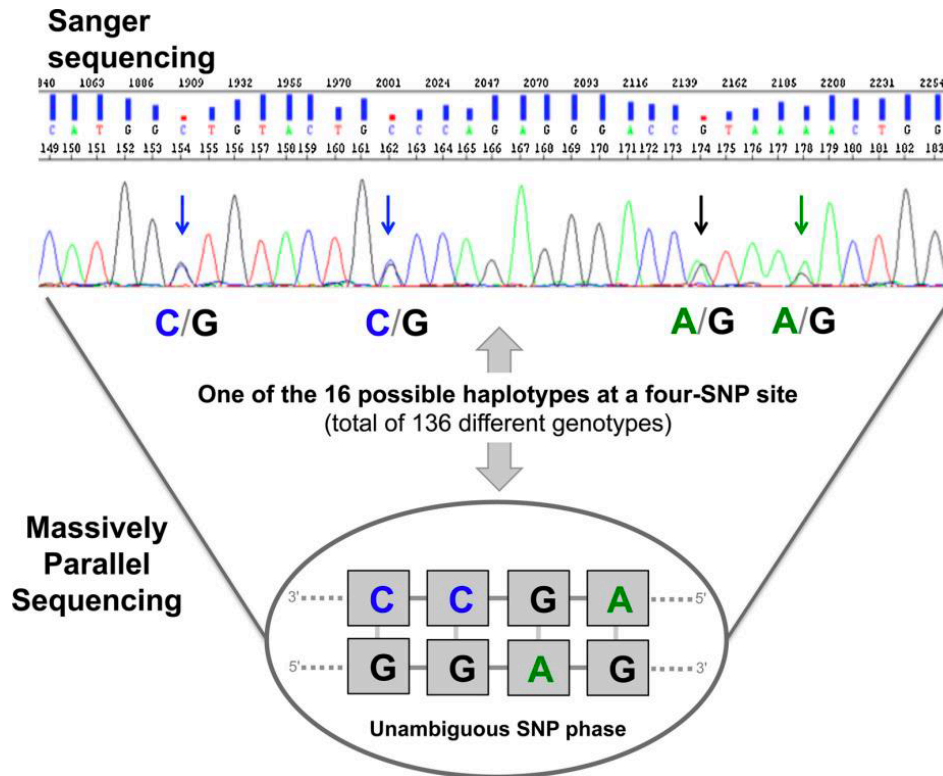


Fig. 7 Sequenziamento di Sanger e Massive Parallel Sequencing <sup>(11)</sup>.

Esistono in commercio diverse tipologie di piattaforme NGS, tuttavia quelle maggiormente utilizzate, definite anche approcci di seconda generazione, sono: *Illumina* e *Ion Torrent*. Entrambe le metodiche generalmente prevedono 4 fasi di lavoro: costruzione delle librerie, preparazione del template, sequenziamento e analisi dei dati <sup>(42)</sup> che verranno approfonditi più avanti.

### 1.7.1 Tecnologia Ion Torrent

La tecnologia Ion Torrent è stata la prima piattaforma NGS senza rilevamento ottico, per cui invece di utilizzare una cascata enzimatica per generare un

segnale, la piattaforma Ion Torrent rileva gli ioni  $H^+$  che vengono rilasciati quando ciascun dNTP viene incorporato. <sup>(44)</sup>

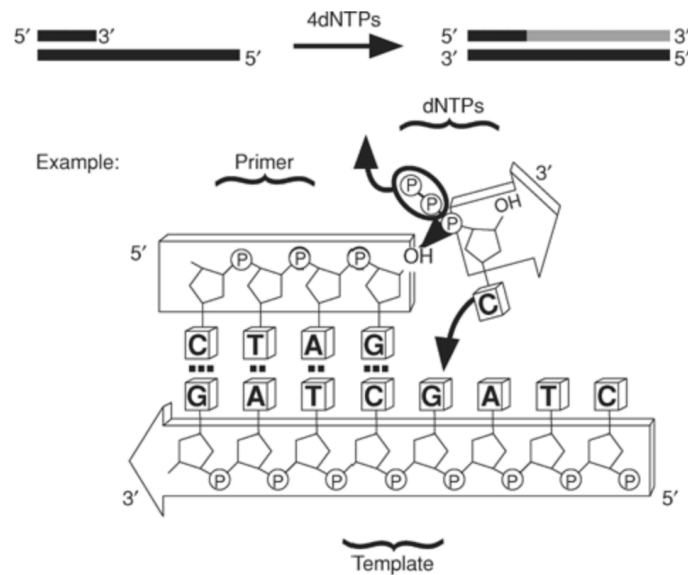


Fig.8 Principio di funzionamento della tecnologia Ion Torrent. <sup>(28)</sup>

In particolare sfrutta la tecnologia di sequenziamento tramite chip semiconduttore (Ion S5 chip) <sup>(28)</sup>. La dimensione del chip è determinata dal numero di basi che devono essere sequenziate.

L' *Ion Chip* è costituito da uno strato superiore, in cui sono presenti dei pozzetti che consentono di poter attirare solo una biglia e i reagenti per il sequenziamento. Questo primo strato superficiale poggia su un secondo costituito da semiconduttori, il quale permette la trasmissione dei segnali allo strato sottostante, strutturato come una piastra di sensori. Ogni sensore corrispondente a ciascun pozzetto, in grado di registrare le piccolissime

variazioni di pH che avvengono all' interno del pozzetto durante il sequenziamento (Fig.9). Ciascun pozzetto può essere considerato come un pHmetro, questo perché ogni qual volta che un nucleotide viene incorporato sul filamento in crescita viene rilasciato un protone: le variazioni di pH che avvengono all'interno dei singoli pozzetti che costituiscono il Chip sono rilevate dallo strumento che associa ad ogni variazione l'incorporazione di un nucleotide trasformando il segnale chimico in segnale digitale. Ogni nucleotide viene irrogato singolarmente (vale a dire che il filamento in crescita viene esposto ad un nucleotide per volta) per cui si avrà un segnale solo se il nucleotide viene incorporato.

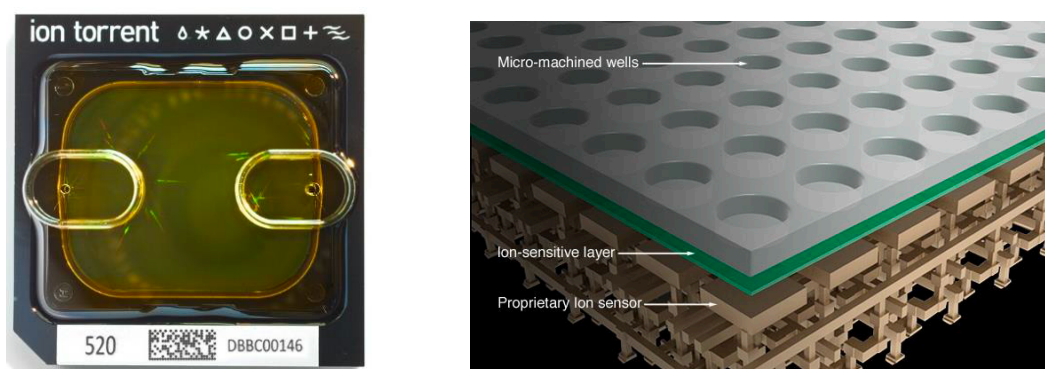


Fig.9: A destra è rappresentato un Ion Chip<sup>(28)</sup> a sinistra invece lo schema della struttura a tre strati dell' Ion Chip.



### **Flusso di lavoro Ion Torrent**

Il flusso di lavoro dell' NGS, costa di 4 fasi: costruzione delle librerie, preparazione del template, sequenziamento e analisi dei dati.

Prima di arrivare alla costruzione delle librerie è necessario estrarre il DNA da una traccia. Con il tempo sono state messe appunto differenti tecniche estrattive da differenti matrici biologiche, ma tutte si basano essenzialmente su tre punti principali: una prima fase di frammentazione e lisi delle membrane cellulari che consente il rilascio degli acidi nucleici, una seconda fase di denaturazione delle proteine e una terza di separazione del DNA dalle proteine e di rimozione di tutti quei contaminanti che potrebbero interferire con le successive fasi di analisi del campione <sup>(3)</sup>. Per poter ottenere risultati ottimali in un'indagine genetica occorre quindi una buona purificazione del DNA estratto del quale andranno valutate, prima delle successive fasi analitiche, qualità e quantità. La fase di quantificazione viene effettuata mediante una Real-Time PCR.

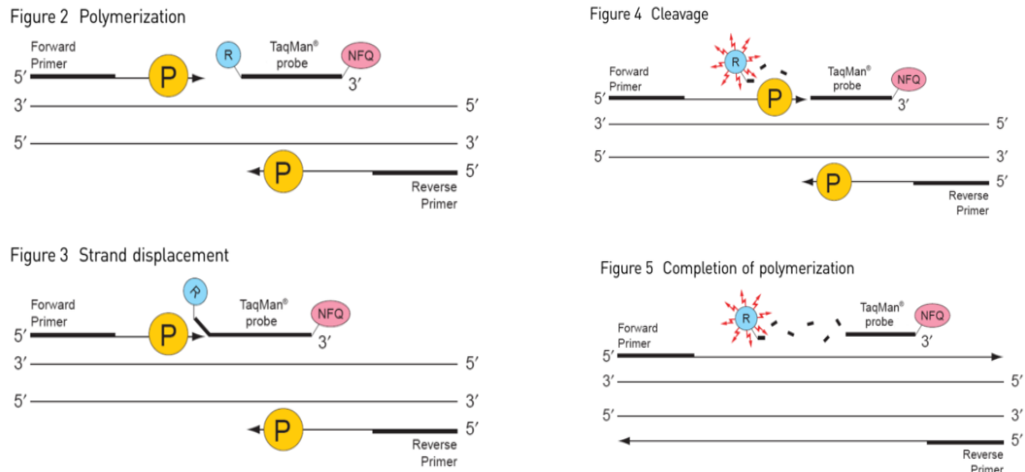
La Real-time PCR è un test in grado di stimare accuratamente sia la quantità che la qualità di DNA presente in un campione. E' una tecnica che combina l'azione di un termociclatore, in cui l'alternanza di temperatura e il numero di cicli variabile determina l'amplificazione dei frammenti di DNA, con quella di un fluorimetro che invece grazie alla presenza di un detector consente la rilevazione di fluorescenza con intensità differenti. Monitorare l'accumulo di

fluorescenza durante la PCR, consente di risalire alla quantità di DNA presente nei campioni mediante la rilevazione in tempo reale di frammenti di DNA amplificati ad ogni ciclo.

Tale analisi prevede l'utilizzo di una sonda marcata con due differenti coloranti che emettono fluorescenza a diverse lunghezze d'onda, chiamata TaqMan.

La sonda TaqMan è costituita da una specifica sequenza di DNA alle cui estremità 5' e 3' sono legati un colorante Reporter (R) e un Quencher (Q), definiti rispettivamente donatore (R) e accettore (Q).

Se lo spettro di emissione del donatore si sovrappone a quello di assorbimento dell'accettore e contemporaneamente le due molecole si trovano in stretta vicinanza, accade che il donatore quando eccitato non emette luce, ma trasferisce l'energia all'accettore per risonanza. Ne consegue che, finché la sonda è intatta e il Reporter è in prossimità del Quencher, il trasferimento di energia tra i due coloranti risulta in un annullamento della fluorescenza del Reporter. Durante la polimerizzazione, la sintesi del filamento complementare degrada le sonde TaqMan. ibridate alla sequenza bersaglio, grazie all'attività 5'-esonucleasica della polimerasi, e il Reporter così rilasciato può emettere fluorescenza rilevabile. <sup>(3)</sup>



*Fig.10 Illustrazione dell'attività della sonda Taq Man*

La quantificazione del DNA viene definita dal Cycle threshold (Ct, ciclo soglia) che è il ciclo di amplificazione nel quale la fluorescenza supera un valore soglia che rappresenta il rumore di fondo osservabile anche nei primi cicli di amplificazione. Il ripetersi di numerosi cicli di amplificazione provoca l'aumento della quantità di DNA, consentendo per ciascun campione il raggiungimento del ciclo soglia. <sup>(3)</sup>

Utilizzando degli standard a concentrazione nota crescente, si andrà a valutare dove questi intersecano nel threshold (Ct) e si andrà a costruire una retta di taratura, sulle ordinate troveremo i valori di Ct degli standard mentre in ascisse le concentrazioni di questi. Per valutare l'efficienza della qPCR vengono presi in considerazione alcuni parametri:

- $R^2$  (coefficiente di correlazione), il quale stabilisce lo scostamento della retta di regressione lineare della curva standard rispetto ai  $C_t$  osservati sperimentalmente. Il valore ideale dell' $R^2$  è vicino ad 1<sup>(45)</sup>;
- La pendenza, definita anche slope, è una misura dell'efficienza della reazione di amplificazione.

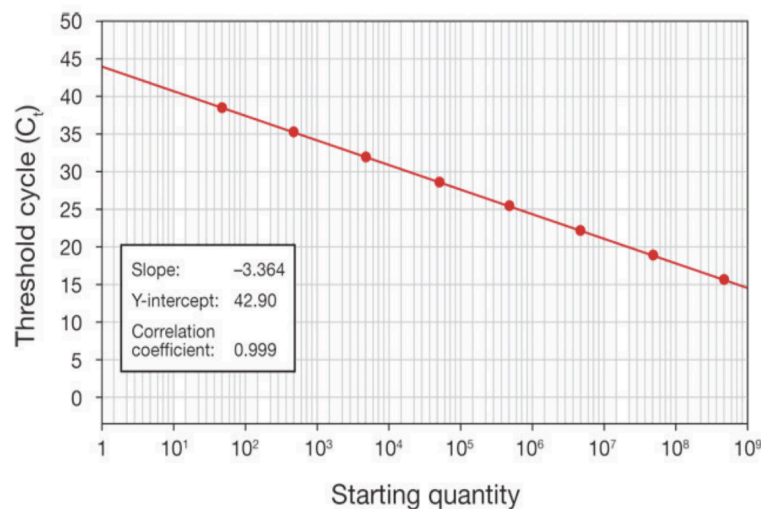


Fig.11 Esempio di retta di taratura (45)

Dopo aver quantificato i campioni si può iniziare con il flusso NGS.

1. *Costruzione delle librerie*: per libreria genomica si intende una collezione completa di tutti i segmenti provenienti dal medesimo organismo. Per ottenerle, ciascun campione viene frammentato, gli ampliconi ottenuti vengono trattati con il reagente *FuPa* che permetterà il legame alle estremità con i *Barcode Adapters*, ovvero un codice a barre specifico costituito da una sequenza oligonucleotidica, in grado di

contrassegnare in modo univoco tutti i frammenti del medesimo campione e poter eseguire un'analisi in multiplexing. Le librerie così formate vengono quindi purificate, quantificate e normalizzate a formare una library-pool dove ciascun campione è identificabile grazie al barcode assegnatogli.

2. *Preparazione del Templato*: in questa fase viene eseguita la clonazione delle librerie tramite PCR in emulsione (ePCR). Per effettuare ciò vengono associati gli ampliconi ad una nanosfera chiamata *Ion Sphere Particle* (ISP). Si allestisce un'amplificazione in emulsione, fatta in acqua e olio, dei frammenti di DNA a cui sono legati i barcode. Questa emulsione favorirà la formazione di micelle in cui possono essere facilmente incorporati una biglia e un singolo amplicone. Nella micella avviene l'amplificazione della molecola di DNA incorporata.

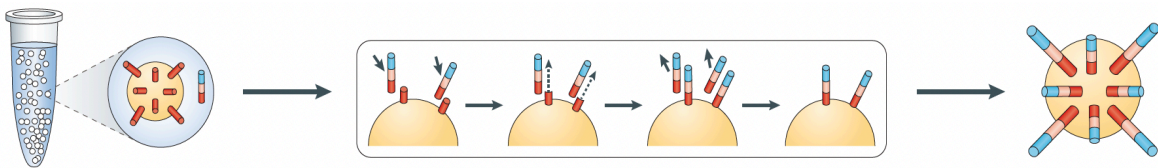
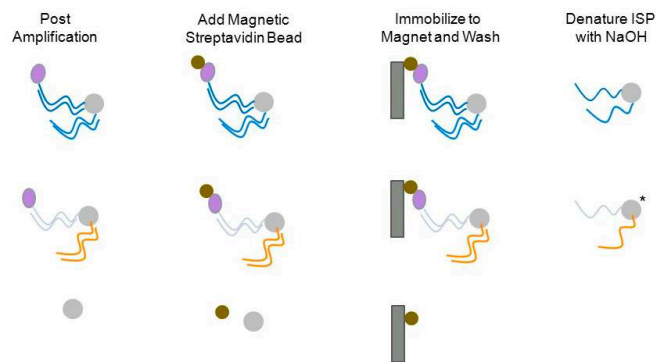


Fig.12 Emulsion PCR <sup>(44)</sup>

Successivamente all'ePCR viene effettuata la rottura dell'emulsione così da rilasciare le ISPs con il template legato. Viene effettuata quindi la fase di

arricchimento che sfrutta l'affinità tra biotina e streptavidina. In particolare, viene utilizzato un secondo tipo di biglia, a cui è adesa la streptavidina, in grado di selezionare solo le biglie ricoperte da molecole di dsDNA, questo perché dopo l'amplificazione clonale, ciascuna particella sferica da un lato è biotinilato. Tutte le ISPs vuote, che non contengono alcun frammento di DNA, vengono eliminate.



*Fig.13 Illustrazione grafica della fase di arricchimento*

Prima del sequenziamento viene effettuato il caricamento del template sul Chip, il quale è costituito da pozzetti all'interno dei quali entra una sola ISP (Fig. 14). Tale processo è svolto dall'*Ion Chef™ Instrument*.

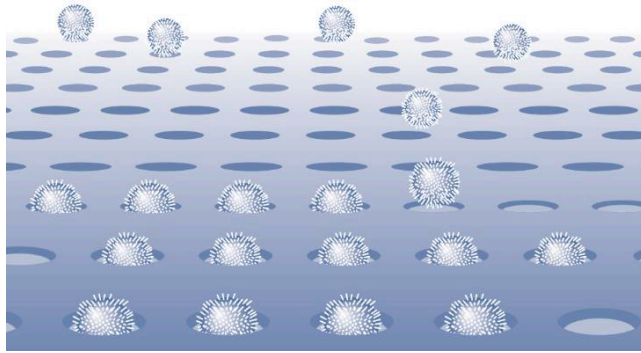


Fig.14 Rappresentazione dei pozzi del Chip in cui alloggia una sola biglia <sup>(43)</sup>

3. Sequenziamento Massivo Parallelo: Il sequenziamento con la tecnologia *Ion Torrent* viene effettuato con l'*Ion S5 Semiconductor Sequencing System*. Come già detto in precedenza, la determinazione della sequenza nucleotidica è resa possibile da una variazione di pH a seguito del rilascio di ioni  $H^+$ .
4. Analisi dei dati: Software appropriati consentono di effettuare il processo di demultiplexing, riconducendo ogni frammento sequenziato ad uno specifico campione e restituendo i risultati del processo di sequenziamento.

Questa tecnologia offre tempi di esecuzione del sequenziamento rapidi (tra 2,5 e 4 ore) con letture comprese tra 200 e 600 bp <sup>(42-43)</sup>.

## ***1.8 Precision ID SNP Panels with the HID Ion S5™/HID Ion***

### ***GeneStudio™ S5 System***

Il pannello utilizzato per lo svolgimento di questo progetto di tesi sperimentale è il Precision ID Library Kit utilizzando il protocollo *Precision ID SNP Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System*.

Per l'amplificazione dei microaplotipi oggetto dello studio, è stato progettato online un pannello di tipo *custom* tramite il sito web Ion AmpliSeq Designer sul sito della ThermoFisher® Scientific <sup>(33)</sup> e allestito per la metodica Ion Torrent. In particolare, nella sezione Ion AmpliSeq designer è possibile creare un pannello fornendo tutti i parametri richiesti, come ad esempio le regioni cromosomiche in cui mappano i loci da sequenziare, il nome da assegnare al pannello, il genoma di riferimento e il file contenente tutti i microaplotipi selezionati. Le analisi dei risultati sono state affidate al plugin *HID\_Microhaplotype\_Research\_PluginV1.5*, il quale ha fornito informazioni relative alla presenza dei microaplotipi che costituiscono il pannello, per ciascun campione. Tra le informazioni ottenute abbiamo: la lettura degli alleli, la relativa frequenza allelica, la percentuale di lettura delle singole basi, le reads ed il coverage di ciascuna mutazione. Questi dati, sono stati organizzati dal plug-in in FileExport, restituiti in formato Excel, utilizzati successivamente per l'interpretazione dei dati.



## *Capitolo 2*

### **SCOPO DELLO STUDIO**

L'obiettivo dello studio è stato quello di valutare le performance di un pannello di 78 microaplotipi su campioni di DNA degradato.

La scelta dei microaplotipi, a partire dal pannello da 89 MH precedentemente testato <sup>(17)</sup> e dagli studi condotti da De La Puente et al. <sup>(31)</sup>, è ricaduta su coloro che avessero un valore di Ae elevato.

Il presente studio si è concentrato esclusivamente sui microaplotipi come nuovi markers per le applicazioni forensi.

Inoltre, un altro parametro fondamentale per la scelta dei microaplotipi, è stato valutare l'ampiezza dell'amplicone, infatti minore sarà l'ampiezza e maggiore sarà la possibilità di poter usare il set di microaplotipi su campioni di DNA moderatamente o fortemente degradati.

La valutazione del pannello è stata effettuata su campioni di DNA diverso, ovvero su campioni di DNA in buone condizioni e su campioni in cui non è stato possibile risalire ad un profilo genotipico derivanti da scene del crimine.

## *Capitolo 3*

### *MATERIALI E METODI*

#### *3.1 Selezione dei loci d'interesse*

I microaplotipi sono stati scelti a partire dal pannello di 89 MH <sup>(19)</sup> precedenti studiati e dagli studi condotti da De La Puente et al. <sup>(31)</sup>. Tutti i loci studiati sono presenti nel database ALFRED (in cui troviamo il numero di SNP per quel determinato MH, la FMFM, ovvero la frequenza dell'allele più elevato per quell'aplotipo e il parametro HET ovvero l'eterozigosità attesa). Inoltre per alcuni dei microaplotipi selezionati è stato necessario effettuare delle modifiche per permettere la riduzione dell'amplicone, ovvero eliminando o aggiungendo degli SNP. Questo è stato fatto mediante visualizzazione manuale della sequenza su *Ensembl* <sup>(32)</sup>. La scelta degli SNP da aggiungere si è basata sul valore della MAF (*Minor Allele Frequency*), ovvero la frequenza con cui si presenta il secondo allele più comune in una data popolazione. In particolare, tale valore deve essere maggiore dello 0,1%; questo perché uno SNP per essere polimorfico, deve avere almeno una frequenza dell'1% in una particolare popolazione, per essere considerata polimorfica.

La scelta è stata fatta seguendo determinati criteri:

- Microaplotipi con valore di  $A_e > 3$ ;
- Microaplotipi composti con un massimo di 6 SNP;

- Microaplotipi amplificabili con ampiezza compresa tra 125-140 bp;

Per il presente studio i microaplotipi selezionati sono 78, in *Tab. I* sono riportate le caratteristiche dei microaplotipi come: il nome relativo ad ogni microhap, la posizione cromosomica, l'ampiezza dell'amplicone e il numero di SNP per locus.

Nome	Posizione cromosomica	Dimensione dell'amplicone (pb)	Numero di SNP per locus
FNDC10	chr1:1529937-1530018 --GRCh37	121	4
mh01KK-160	chr1:153170258-153170323 --GRCh37	105	4
1qB	chr1:187622410-187622491 --GRCh37	121	4
LMX1A	chr1:165270325-165270411 -- GRCh37	126	4
mh02KK-140	chr2:4822383-4822455 --GRCh37	112	4
mh02KK-141	chr2:41188288-41188373 --GRCh37	125	4
D2S259	chr2:76368295-76368370 --GRCh37	115	4
2pC	chr2:79252938-79253030 --GRCh37	132	4
2qA	chr2:103092496-103092584 --GRCh37	128	4
mh02KK-134	chr2:161079365-161079459 --GRCh37	134	3
2qD	chr2:206822599-206822677 --GRCh37	118	3
mh02KK-136	chr2:228092383-228092465 --GRCh37	122	3
D3S2426	chr3:1717833-1717911 --GRCh37	118	3
3pA	chr3:5534414-5534492 --GRCh37	118	4
mh03KK-217	chr3:10823050-10823139 --GRCh37	129	5
ZNF385D	chr3:22057765-22057856 --GRCh37	131	4
3pB	chr3:42924607-42924695 --GRCh37	128	5
LINC02000	chr3:133806138-133806228 --GRCh37	130	4
LOC100505920	chr3:193946498-193946582 --GRCh37	124	4
mh04KK-010	chr4:1986908-1986996 --GRCh37	128	2
SORCS2	chr4:7499107-7499207 --GRCh37	140	4
4pA	chr4:8122448-8122551 --GRCh37	143	4
4pB	chr4:36639739-36639813 --GRCh37	114	4
4qA	chr4:59436032-59436105 --GRCh37	113	3
4qC	chr4:155794578-155794665 --GRCh37	127	4
mh04KK-217	chr4:179715535-179715603 --GRCh37	108	3
mh04KK-218	chr4:182795865-182795945 --GRCh37	120	4
D4S2390	chr4:189794162-189794256 --GRCh37	134	5
D4S3214	chr4:189891668-189891754 --GRCh37	126	4
mh05KK-125	chr5:3082770-3082849 --GRCh37	119	4
mh05KK-022	chr5:9619877-9619963 --GRCh37	126	2
5qB	chr5:106366209-106366296 --GRCh37	127	4
5qC	chr5:132916318-132916416 --GRCh37	138	4

5qD	chr5:178267354-178267452 --GRCh37	138	3
6pA	chr6:23765315-23765400 --GRCh37	125	3
mh06KK-103	chr6:31145388-31145478 --GRCh37	130	4
D6S2662-B	chr6:32700783-32700875 --GRCh37	132	4
LOC105375075	chr6:44595088-44595183 --GRCh37	135	4
6pB	chr6:53700970-53701061 --GRCh37	131	4
6qA	chr6:77942995-77943075 --GRCh37	120	3
6qB	chr6:96372294-96372382 --GRCh37	128	4
6qC	chr6:120130310-120130392 --GRCh37	122	3
7pA	chr7:11613136-11613224 --GRCh37	128	4
7pB	chr7:25447583-25447668 --GRCh37	125	4
7pC	chr7:57218093-57218198 --GRCh37	145	5
7qA	chr7:68052846-68052940 --GRCh37	134	3
7qC	chr7:151518731-151518818 --GRCh37	127	5
8pA	chr8:3306392-3306481 --GRCh37	129	5
GATA-B-GATA4-A	chr8:11607303-11607402 --GRCh37	139	5
8pB	chr8:40664178-40664261 --GRCh37	123	5
9pA	chr9:2288627-2288718 --GRCh37	131	6
9pB	chr9:31196668-31196747 --GRCh37	119	4
9qA	chr9:73362598-73362689 --GRCh37	131	3
9qB	chr9:104903823-104903911 --GRCh37	128	4
9qC	chr9:136768832-136768923 --GRCh37	131	5
10pA	chr10:10412198-10412286 --GRCh37	128	4
11pB	chr11:34437342-34437425 --GRCh37	123	5
12qA	chr12:40786705-40786767 --GRCh37	102	4
12qB	chr12:68508276-68508362 --GRCh37	126	4
13qA	chr13:33553572-33553664 --GRCh37	132	4
mh13KK-218	chr13:54060822-54060902 --GRCh37	120	3
13qC	chr13:85907630-85907714 --GRCh37	124	4
14qA	chr14:25724211-25724295 --GRCh37	124	3
D16S2835	chr16:9675226-9675314 --GRCh37	128	5
17qB	chr17:49449766-49449858 --GRCh37	132	3
17qC	chr17:78919831-78919923 --GRCh37	132	4
18pA	chr18:5280008-5280095 --GRCh37	127	5
18qB	chr18:50547455-50547541 --GRCh37	126	3
18qC	chr18:63842490-63842570 --GRCh37	120	4
19pB	chr19:16151666-16151748 --GRCh37	122	5
LINC00906	chr19:29376017-29376100 --GRCh37	123	5
19qA	chr19:33764632-33764726 --GRCh37	134	5
19qB	chr19:54217617-54217709 --GRCh37	132	4
20qB	chr20:61091310-61091400 --GRCh37	130	5
21qA	chr21:15899204-15899286 --GRCh37	122	4
mh21KK-324	chr21:46714618-46714719 --GRCh37	141	3
22qA	chr22:30800301-30800380 --GRCh37	119	4
22qB-mh22KK-070	chr22:49060958-49061048 --GRCh37	130	6

*Tab.1 Caratteristiche dei 78 microaplotipi*

Di seguito vengono elencate le caratteristiche tecniche del pannello:

- Range di lunghezza degli ampliconi tra 125-140 bp;
- Un solo pool di primer;
- 78 ampliconi;
- 58 bp perse;
- Coverage del 99.07%;
- Target: Genoma Umano (hg19).

### ***3.3 Campioni utilizzati per la validazione del pannello***

Per valutare la funzionalità del pool di primer disegnato, sono stati selezionati 34 campioni di DNA (*Tab.2*), estratti da sangue fresco, testati inizialmente nel pannello originario da 89 MH <sup>(19)</sup>.

SAMPLE	TIPO DI CAMPIONE
ANCS022	SANGUE FRESCO
ANCS031	SANGUE FRESCO
ANCS032	SANGUE FRESCO
ANCS033	SANGUE FRESCO
ANCS034	SANGUE FRESCO
ANAS060	SANGUE FRESCO
ANAS061	SANGUE FRESCO
ANAS064	SANGUE FRESCO
ANAS092	SANGUE FRESCO
ANAS093	SANGUE FRESCO
ANAS103	SANGUE FRESCO
ANCS017	SANGUE FRESCO
ANCS018	SANGUE FRESCO
ANCS019	SANGUE FRESCO
ANCS020	SANGUE FRESCO
ANCS002	SANGUE FRESCO
ANCS005	SANGUE FRESCO
ANCS023	SANGUE FRESCO
ANCS001	SANGUE FRESCO

ANCS011	SANGUE FRESCO
ANCS021	SANGUE FRESCO
ANCS027	SANGUE FRESCO
ANCS029	SANGUE FRESCO
ANAS001	SANGUE FRESCO
ANAS014	SANGUE FRESCO
ANAS015	SANGUE FRESCO
ANAS018	SANGUE FRESCO
ANAS039	SANGUE FRESCO
ANAS040	SANGUE FRESCO
ANAS046	SANGUE FRESCO
ANAS047	SANGUE FRESCO
ANAS054	SANGUE FRESCO
ANAS062	SANGUE FRESCO
ANAS070	SANGUE FRESCO
ANAS071	SANGUE FRESCO

*Tab.2 Informazioni sui campioni per studio popolazionistico.*

I test sono stati effettuati anche su campioni forensi reali, l'elenco dei campioni è riportato in *Tab.3*

SAMPLE	TIPO DI CAMPIONE
CAMPIONE 1	SUPPORTO SOLIDO
CAMPIONE 2	SUPPORTO SOLIDO
CAMPIONE 3	SUPPORTO SOLIDO
CAMPIONE 4	SUPPORTO SOLIDO
CAMPIONE 5	SUPPORTO SOLIDO
CAMPIONE 6	SUPPORTO SOLIDO
CAMPIONE 7	SUPPORTO SOLIDO
CAMPIONE 8	SUPPORTO SOLIDO
CAMPIONE 9	SUPPORTO SOLIDO
CAMPIONE 10	SUPPORTO SOLIDO
CAMPIONE 11	SUPPORTO SOLIDO
CAMPIONE 12	SUPPORTO SOLIDO
CAMPIONE 13	SUPPORTO SOLIDO
CAMPIONE 14	SUPPORTO SOLIDO
CAMPIONE 15	SUPPORTO SOLIDO
CAMPIONE 16	SUPPORTO SOLIDO
CAMPIONE 17	SUPPORTO SOLIDO
CAMPIONE 18	SUPPORTO SOLIDO
CAMPIONE 19	SUPPORTO SOLIDO
CAMPIONE 20	SUPPORTO SOLIDO

*Tab.3 Informazioni sui campioni reali forensi.*

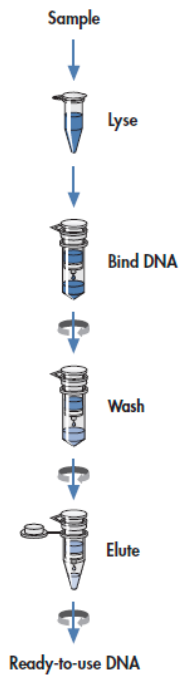
### ***3.4 Estrazione e quantificazione del DNA***

Per questo studio sono stati analizzati campioni di DNA estratti e quantificati con metodiche diverse in accordo alle procedure interne dei laboratori. In particolare per i campioni della *Tab.1* il DNA è stato estratto con QIAamp® DNA Micro e successivamente quantificato con la metodica *E-Gel™ Agarose Gels with SYBR Safe™*. Mentre il DNA per i campioni della *Tab.2* è stato estratto e quantificato rispettivamente con *EZI ® DNA Investigator KIT* e *PowerQuant® System*.

#### ***3.4.1 Estrazione del DNA con QIAamp® DNA Micro***

L'estrazione del DNA genomico tramite QIAamp® DNA Micro della QIAGEN® (*QIAamp® DNA Micro Handbook, Third Edition, December 2014, QIAGEN®*) <sup>(34)</sup> prevede una fase iniziale di lisi cellulare, necessaria per il rilascio dell'acido nucleico, che avviene in condizioni denaturanti ad alta temperatura; il passaggio del lisato attraverso una colonnina e il legame del DNA alla matrice silicea della colonnina. Dopodiché si effettuano diversi lavaggi e infine si eluisce il DNA con un buffer o semplice acqua distillata.

#### QIAamp DNA Micro Procedure



*Fig.15 Protocollo d'estrazione del DNA con QIAamp® DNA Micro*

La procedura prevede l'utilizzo di sangue e anticoagulante (EDTA, citrato o eparina) come materiale di partenza.

- Aggiungere 100 µl di buffer ATL a 2 gocce di sangue;
- Aggiungere 10 µl di Proteinasi K;
- Aggiungere 100 µl di Buffer AL;
- Vortexare per 15 secondi;
- Incubare a 56°C per 10 minuti in agitazione su Eppendorf Thermomixer;
- Breve centrifugazione;
- Aggiungere 50 µl di etanolo (96-100%);
- Vortexare per 15 secondi;



- Incubazione a temperatura ambiente per 3 minuti;
- Breve centrifugazione;
- Preparare una QIAmp MinElute column per ogni campione (ogni colonnina è già dotata di una collection tube per la raccolta del liquido in eccesso);
- Trasferire l'intero lisato (dopo breve centrifugazione) all'interno della colonnina, senza toccarne la membrana;
- Centrifugazione a 8000 rpm per 1 minuto;
- Trasferire la colonnina all'interno di una nuova collection tube e gettare quella che contiene il liquido di scarto;
- Aggiungere 500 µl di Buffer AW1, senza toccare la membrana della colonnina;
- Centrifugazione a 8000 rpm per un minuto;
- Trasferire la colonnina all'interno di una nuova collection tube e gettare quella che contiene il liquido di scarto;
- Aggiungere 500 µl di Buffer AW2, senza toccare la membrana della colonnina;
- Centrifugazione a 8000 rpm per 1 minuto;
- Trasferire la colonnina all'interno di una nuova collection tube e gettare quella che contiene il liquido di scarto;

- Centrifugazione a 14000 rpm per 3 minuti, per asciugare la membrana;
- Trasferire la colonnina in una provetta da 1,5 mL (in cui viene eluito il DNA);
- Aggiungere 50 µl di Buffer AE, al centro della membrana della colonnina;
- Incubare a temperatura ambiente per 1 minuto;
- Centrifugare a 14000 rpm per 1 minuto;
- Gettare la colonnina, il DNA eluito si trova nella provetta da 1,5 mL.

### ***3.4.2 Estrazione del DNA con EZ1 ® DNA Investigator KIT***

A seguito di prelievo tramite tampone cotonato, campioni relativi a diverse matrici biologiche, contenuti in provette dotate di Spin Basket, sono stati incubati con reagenti e rispettive quantità di seguito riportati:

- 190uL di Buffer di lisi G2, un agente caotropico che consente la denaturazione delle proteine e la rottura delle strutture tipiche delle cellule;
- 10uL di Proteinasi K, un enzima in grado di determinare la liberazione del DNA dalle strutture nucleari e dalle proteine istoniche nonché inibitore di tutti gli enzimi presenti nelle cellule - RNAsi e DNAsi - capaci di determinare la degradazione degli acidi nucleici.

- Una volta terminata la lisi, che può essere condotta (opzionale) direttamente in uno spin basket nel tubo di lisi, si procede a centrifugazione per 2' a 14000 rpm;
- Lo spin basket eventualmente utilizzato viene rimosso e smaltito;
- I campioni sono pronti per la fase estrattiva vera e propria;

La procedura di estrazione del materiale genetico è avvenuta mediante strumento EZ1® Advanced XI (Qiagen, 2017), una piattaforma automatizzata che consente di isolare e purificare il DNA umano tramite l'utilizzo di particelle paramagnetiche totalmente ricoperte di silice in grado di legare il DNA in presenza di specifiche condizioni di reazione. I reagenti necessari per il protocollo di estrazione sono interamente contenuti nell'apposito EZI® DNA Investigator Kit <sup>(35)</sup> sotto forma di cartucce con pozzetti; ciascun pozzetto contiene, in quantità stabilite e sufficienti biglie paramagnetiche ricoperte di silice, buffer di lisi, buffer per i lavaggi e buffer di eluizione. Una volta ultimato il caricamento delle diverse componenti dello strumento, l'operatore, a seconda delle esigenze e della tipologia di accertamento che deve effettuare, può selezionare protocolli differenti e variare alcuni parametri, quali il volume finale ed il buffer di eluizione.

Lo strumento è in grado di processare contemporaneamente un numero massimo di 14 campioni, incluso un controllo negativo di estrazione cioè un

campione costituito unicamente dalla mix di incubazione, quindi privo di materiale genetico, la cui funzione principale è quella di dimostrare l'assenza di contaminazioni nei reagenti riportati in precedenza. Per l'elaborato in analisi, lo strumento è stato settato sul protocollo "Tracce", indicato il buffer TE come buffer di eluizione e 50ul come volume di eluizione.

Una volta ultimata la procedura di estrazione, gli estratti di DNA sono stati trasferiti in appositi tubini matrix contrassegnati da un univoco QR-code per garantire l'identificazione di ciascun campione.

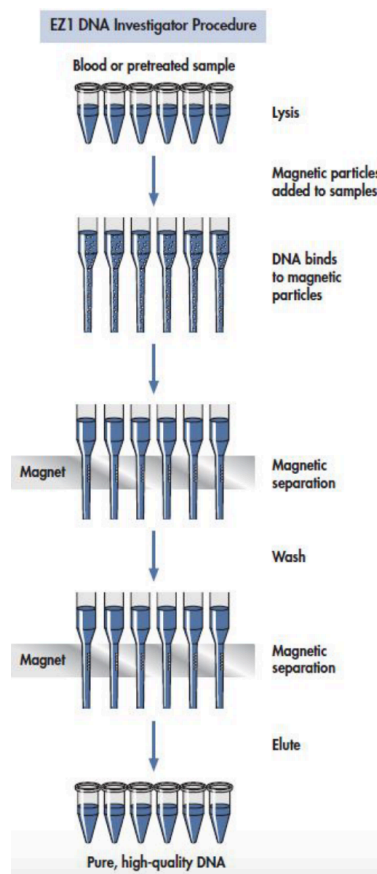


Fig.16 Protocollo d'estrazione del DNA EZI® DNA Investigator Kit.

### **3.4.3 Quantificazione con E-Gel™ Agarose Gels with SYBR Safe™**

La quantificazione con E-Gel™ Agarose Gels with SYBR Safe™ (39) prevede l'allestimento di un'elettroforesi su gel d'agarosio al 2%, per la determinazione della taglia e per la quantizzazione delle regioni amplificate.

Per ottenere buoni risultati, viene suggerito di caricare 20-100 ng di DNA per banda. Eventuali diluizioni del campione possono essere preparate in acqua deionizzata o tampone.

Il volume totale del campione per il caricamento è di 20 µL.

Per allestire la corsa è necessario anche il DNA molecular weight marker. A differenza dei campioni, per questo viene fatta una diluizione 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%". La corsa dura 26 minuti. Al termine della corsa si possono visualizzare le bande accendendo il transilluminatore (E-Gel® Safe Imager™ Real-Time Transilluminator), che mette in evidenza gli ampliconi.

SAMPLE	[DNA] ng/ul
ANCS022	30
ANCS031	50
ANCS032	50
ANCS033	50
ANCS034	40
ANAS060	40
ANAS061	40
ANAS064	40
ANAS092	40
ANAS093	40
ANAS103	40

ANCS017	40
ANCS018	40
ANCS019	40
ANCS020	40
ANCS002	41
ANCS005	41
ANCS023	40
ANCS001	50
ANCS011	50
ANCS021	50
ANCS027	50
ANCS029	50
ANAS001	30
ANAS014	30
ANAS015	30
ANAS018	30
ANAS039	30
ANAS040	30
ANAS046	30
ANAS047	30
ANAS054	30
ANAS062	30
ANAS070	30
ANAS071	30

*Tab.4 Risultati della quantificazione con E-Gel™ Agarose Gels with SYBR Safe™ per i campioni per studio popolazionistico.*

#### **3.4.4 Quantificazione con PowerQuant® System**

La procedura di quantificazione PowerQuant<sup>(40)</sup> prevede una doppia serie di diluizioni scalari del DNA standard [50 ng//uL stock], con concentrazioni comprese tra 50 ng/uL (Std.1) e 3,2 pg/uL (Std.4), oppure è possibile utilizzare la doppia curva standard prevista per il kit Plexor HY su 4 standard, con concentrazioni comprese tra 50 ng/uL (Std.1) e 3,2 pg uL.

Standard	Concentrazione (ng/mL)	Come diluire	Fattore di diluizione
Std. 1	50	Usare lo standard del kit non diluito.	0x
Std. 2	2	4 pL [Std. 1] + 96 uL PowerQuant Dilution Buffer	25 X
Std. 3	0.08	4 uL [Std. 2] + 96 uL PowerQuant Dilution Buffer	25 X

Per preparare le diluizioni del DNA standard:

- Procedere ad eseguire il protocollo sulla piattaforma automatizzata Hamilton Starlet che in automatico procederà ad effettuare le diluizioni della curva standard e a caricarle sulla piastra di reazione. Altrimenti eseguire manualmente la curva standard come segue.
- Contrassegnare ciascuna provetta con la scritta Std.1, Std.2, e così via e procedere al loro smaltimento al termine della prova;
- Dispensare la quantità richiesta di DNA e di acqua distillata autoclavata o DNA free o di dilution buffer dove previsto come indicato in tabella, manualmente, come descritto nei successivi punti 3 e 4 con le micropipette opportune, avendo cura di cambiare i puntali ad ogni passaggio;
- Per la preparazione dello Std.1 vortexare il DNA Standard da 3 a 5 secondi, riporre la quantità calcolata di DNA Standard nel tubo indicato come Std. 1, poi mescolare vigorosamente.
- preparare poi gli standard: da 2 a 4 (Powerquant)

ponendo le quantità calcolate di DNA nei tubi dei passaggi successivi come indicato in tabella, mescolando vigorosamente ad ogni passaggio.

#### Preparazione della reazione

- Nel calcolare il numero di campioni necessari occorre aggiungere anche un ulteriore CN (cosiddetto NTC, "no template control"), da porre nelle posizioni successive agli standard più diluiti o a seguire i campioni da quantificare.
- Calcolare il volume richiesto per ogni componente sulla base della numerosità dei campioni da quantificare, tenendo conto che si deve dispensare una doppia serie di standard, utilizzando i seguenti riferimenti:

*Mix per singolo campione  $x[(n^\circ \text{ campioni}) + (n^\circ \text{ campioni}/10)]$*

MIX PER SINGOLO CAMPIONE	PRIMER MIX	PCR REACTION MIX	ACQUA PENTA AUTOCLAVATA O DNA FREE
9,0 $\mu$	0,5 $\mu$	5,0 $\mu$	3,5 $\mu$

Se la quantificazione è effettuata con l'ausilio del robot Hamilton Starlet occorre preparare una Mix maggiorata del 30%.

- Far sciogliere completamente la Primer mix, poi vortexare per 3-5 secondi. Agitare delicatamente la Reaction mix prima dell'uso ma senza vortexare.



SAMPLE	[DNA] pg/ul
CAMPIONE 1	88
CAMPIONE 2	127
CAMPIONE 3	35
CAMPIONE 4	8
CAMPIONE 5	11
CAMPIONE 6	6
CAMPIONE 7	36
CAMPIONE 8	7300
CAMPIONE 9	6606
CAMPIONE 10	17678
CAMPIONE 11	36
CAMPIONE 12	0,6
CAMPIONE 13	887
CAMPIONE 14	87
CAMPIONE 15	905
CAMPIONE 16	1662
CAMPIONE 17	5452
CAMPIONE 18	5730
CAMPIONE 19	15438
CAMPIONE 20	4502

*Tab.5 Risultati della quantificazione con PowerQuant® System per i campioni reali forensi.*

### ***3.5 Massive Parallel Sequencing***

Il sequenziamento è stato eseguito seguendo il protocollo *Precision ID SNP Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System* <sup>(36)</sup>. Nel caso dei campioni di DNA di buona qualità (*Tab.2*), abbiamo lavorato utilizzando i volumi dimezzati, mentre per i campioni di DNA degradato (*Tab.3*) sono stati rispettati i volumi proposti dal manuale.

### Amplificazione delle regioni target e preparazione delle librerie

La prima fase prevede l'amplificazione della regione di nostro interesse, questo avviene grazie all'utilizzo di un pool di primer. Prima di procedere con la reazione di PCR, i campioni di DNA sono stati diluiti alla concentrazione finale di 1ng/ $\mu$ L, a seconda delle concentrazioni ottenute dalla quantificazione.

(Tab.6 e Tab.7)

CAMPIONE	QUANTITA' DI CAMPIONE	QUANTITA' DI ACQUA
ANCS022	2 $\mu$ L	58 $\mu$ L
ANCS031	2 $\mu$ L	48 $\mu$ L
ANCS032	2 $\mu$ L	48 $\mu$ L
ANCS033	2 $\mu$ L	48 $\mu$ L
ANCS034	2 $\mu$ L	78 $\mu$ L
ANAS060	2 $\mu$ L	78 $\mu$ L
ANAS064	2 $\mu$ L	78 $\mu$ L
ANAS092	2 $\mu$ L	78 $\mu$ L
ANAS093	2 $\mu$ L	78 $\mu$ L
ANAS103	2 $\mu$ L	78 $\mu$ L
ANCS017	2 $\mu$ L	78 $\mu$ L
ANCS018	2 $\mu$ L	78 $\mu$ L
ANCS019	2 $\mu$ L	78 $\mu$ L
ANCS020	2 $\mu$ L	78 $\mu$ L
ANCS002	2 $\mu$ L	78 $\mu$ L
ANCS005	2 $\mu$ L	78 $\mu$ L
ANCS023	2 $\mu$ L	78 $\mu$ L
ANCS001	2 $\mu$ L	48 $\mu$ L
ANCS011	2 $\mu$ L	48 $\mu$ L
ANCS021	2 $\mu$ L	48 $\mu$ L
ANCS027	2 $\mu$ L	48 $\mu$ L
ANCS029	2 $\mu$ L	48 $\mu$ L
ANAS001	2 $\mu$ L	58 $\mu$ L
ANAS014	2 $\mu$ L	58 $\mu$ L
ANAS015	2 $\mu$ L	58 $\mu$ L
ANAS018	2 $\mu$ L	58 $\mu$ L
ANAS039	2 $\mu$ L	58 $\mu$ L
ANAS040	2 $\mu$ L	58 $\mu$ L
ANAS046	2 $\mu$ L	58 $\mu$ L
ANAS047	2 $\mu$ L	58 $\mu$ L

ANAS054	2 $\mu$ L	58 $\mu$ L
ANAS062	2 $\mu$ L	58 $\mu$ L
ANAS070	2 $\mu$ L	58 $\mu$ L
ANAS071	2 $\mu$ L	58 $\mu$ L

*Tab.6 Quantità di acqua e campioni utilizzate per la diluizione campioni di DNA per studio  
popolazionistico*

CAMPIONE	QUANTITA' DI CAMPIONE	QUANTITA' DI ACQUA
CAMPIONE 1	Campione non diluito	
CAMPIONE 2	Campione non diluito	
CAMPIONE 3	Campione non diluito	
CAMPIONE 4	Campione non diluito	
CAMPIONE 5	Campione non diluito	
CAMPIONE 6	Campione non diluito	
CAMPIONE 7	Campione non diluito	
CAMPIONE 8	1 $\mu$ L	44 $\mu$ L
CAMPIONE 9	1 $\mu$ L	40 $\mu$ L
CAMPIONE 10	1 $\mu$ L	106 $\mu$ L
CAMPIONE 11	Campione non diluito	
CAMPIONE 12	Campione non diluito	
CAMPIONE 13	2 $\mu$ L	10 $\mu$ L
CAMPIONE 14	Campione non diluito	
CAMPIONE 15	2 $\mu$ L	10 $\mu$ L
CAMPIONE 16	1 $\mu$ L	10 $\mu$ L
CAMPIONE 17	1 $\mu$ L	33 $\mu$ L
CAMPIONE 18	1 $\mu$ L	34 $\mu$ L
CAMPIONE 19	1 $\mu$ L	93 $\mu$ L
CAMPIONE 20	1 $\mu$ L	27 $\mu$ L

*Tab.7 Quantità di acqua e campioni utilizzate per la diluizione dei campioni reali forensi.*

La reazione di PCR è stata allestita come in *Tab.8*, mantenendo il ghiaccio 5 X  
*Ion AmpliSeq™ HiFi Mix*.

COMPONENTI	VOLUME	
	<i>High-quality DNA</i>	<i>Low-quality DNA</i>
<b>5X Ion AmpliSeq™ HiFi Mix</b>	2µL	4µL
<b>Costum Ion AmpliSeq™ SNP Panel</b>	5µL	10µL
<b>DNA (1ng)</b>	X µL	X µL
<b>Acqua nucleasi-free</b>	3 – X µL	6 - XµL
<b>VOLUME</b>	10 µL	20 µL

*Tab.8 Mix di PCR per l'amplificazione della regione target.*

Il volume di DNA può variare in base alla concentrazione del DNA di partenza.

La quantità ottimale suggerita dal kit è di 1 ng.

Le condizioni di amplificazione sono riportate in *Tab.9* e impostate su *Veriti™ 96-Well Thermal Cycler*.

Step	Temperature	Time
Active the enzyme	99°C	2 minutes
Denature	99°C	15 minutes
Anneal and exted	60°C	4 minutes
	10°C	Hold

*Tab.9 Condizioni di amplificazione.*

I campioni sono stati amplificati utilizzando il medesimo protocollo, ottimizzando i cicli in base alla tipologia dei campioni, precisamente 22 cicli per i campioni “high-quality” (*Tab.2*) e a 25 cicli per i campioni “low-quality” (*Tab.3*)

### Parziale digestione degli ampliconi

A questo punto si procede con una parziale digestione dei campioni amplificati che viene effettuata con un reagente chiamato *FuPa Reagent*. Anche in questo caso 2  $\mu\text{L}$  o 1  $\mu\text{L}$  di *FuPa Reagent* vengono aggiunti direttamente al prodotto di amplificazione, a seconda del tipo di campione raggiungendo un volume totale di  $\sim 22 \mu\text{L}$  o di  $\sim 11 \mu\text{L}$ .

Le provette vengono caricate su *Veriti™ 96-Well Thermal Cycler* con le seguenti condizioni (*Tab.10*).

Temperature	Time
50°C	10 minutes
55°C	10 minutes
60°C	20 minutes
10°C	Hold (per massimo 1 ora)

*Tab.10 Condizioni della reazione di digestione*

### Ligazione dei barcode

Come prima cosa è necessario diluire i barcode prima di poterli usare. Ogni barcode è stato diluito come in *Tab.11*

Component	Volume
Ion P1 Adapter	1 $\mu\text{L}$
Ion Xpress™ Barcode X	1 $\mu\text{L}$
Nuclease-free Water	2 $\mu\text{L}$
TOTALE	4 $\mu\text{L}$

*Tab.11 Diluizione dei barcode. La X rappresenta il codice del barcode scelto.*

Una volta diluiti i barcode possono essere conservati a  $-20^{\circ}$  e riutilizzati fino a un massimo di 20 giorni. Successivamente i barcode diluiti vengono ligati ad una estremità degli ampliconi. La preparazione della reazione di ligazione è riportata in *Tab.12*. E' importante ricordare che durante questo passaggio la DNA ligasi va mantenuta in un porta provette ghiacciato.

Component	Volume	
	High-quality DNA	Low-quality DNA
Switch Solution	2 $\mu$ L	4 $\mu$ L
Barcode X diluito	1 $\mu$ L	2 $\mu$ L
DNA ligase	1 $\mu$ L	2 $\mu$ L
Totale	$\sim 15$ $\mu$ L	$\sim 20$ $\mu$ L

*Tab.12 Mix di PCR per la ligazione dei barcode.*

Il caricamento delle provette su *Veriti™ 96-Well Thermal Cycler* segue le seguenti condizioni (*Tab.13*).

Temperature	Time
22°C	30 minutes
72°C	10 minutes
10°C	Hold (for up to 24 hours)

*Tab.13 Condizioni della reazione di ligazione.*

### Purificazione delle librerie

A questo punto si procede con la purificazione delle librerie, mediante l'utilizzo di una resina, *Agencourt™ AMPure™ XP Ragent* che permetterà

l'eliminazione di tutte le componenti in eccesso, usate durante gli step precedenti.

- Trasferire le provette con campione e resina su 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  $\mu\text{L}$  di etanolo al 70% appena preparato; La preparazione dell'etanolo si ottiene combinando 230  $\mu\text{L}$  di etanolo con 100  $\mu\text{L}$  di acqua nuclease-free.
- Ruotare le provette sul rack magnetico per 5-6 volte (controllare che il precipitato migri correttamente);
- Rimuovere il surnatante, evitando di toccare il precipitato;
- Ripetere nuovamente i due passaggi precedenti:
  - Aggiungere 150  $\mu\text{L}$  di etanolo al 70%
  - Ruotare le provette sul rack magnetico per 5-6 volte (controllare che precipitato migri in modo corretto);
- Rimuovere il surnatante, evitando di toccare il precipitato;
- Fare una centrifuga veloce, facendo attenzione a posizionare il precipitato sulla 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);

#### Eluizione delle librerie

- Aggiungere 50  $\mu\text{L}$  di LOW TE facendo in modo di sciogliere il precipitato;
- Vortexare
- 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 provette nuove, già codificate con il nome del campione.

#### Quantizzazione delle librerie

Dopo aver purificato ed eluito le librerie è necessario quantizzarle tramite PCR quantitativa (*qPCR*), utilizzando il *kit Ion Library Taq Man® Quantitation Kit*.

Prima di procedere con l'allestimento della mix per la qPCR è necessario fare una diluizione 1:100 della libreria, combinando in una provetta 198  $\mu\text{L}$  di acqua nucleasi free e 2  $\mu\text{L}$  di libreria purificata.



Dopo di che si procede con la preparazione delle diluizioni seriali dello standard, *E. coli DH10B Ion Control Library*, contenuto nel kit seguendo le concentrazioni elencate in *Tab.14*:

Standard	Control Library volume	Nuclease-free Water volume	Concentration
1	5 $\mu$ L (undiluted)	45 $\mu$ L	6.8 pM
2	5 $\mu$ L Std 1	45 $\mu$ L	0.68 pM
3	5 $\mu$ L Std 2	45 $\mu$ L	0.068 pM

*Tab.14 Concentrazioni per la diluizione seriale della libreria*

Preparare la mix di PCR per il numero totale di campioni, per il controllo negativo e per gli standard appena diluiti come da *Tab.15*.

Component	Volume (1 reaction)
Ion Library TaqMan <sup>®</sup> qPCR Mix	10 $\mu$ L
Ion Library TaqMan <sup>®</sup> Quantitation Assay, 20X	1 $\mu$ L

*Tab.15 Mix per PCR*

Aliquotare 11  $\mu$ L della mix nelle provette per la quantizzazione e aggiungere 9  $\mu$ L campione (standard, controllo negativo o librerie diluite). Il totale per ogni provetta è 20  $\mu$ L. Impostare le condizioni opportune sul *Rotor Gene 6000 PCR System*, per la quantizzazione delle librerie dei campioni di DNA non degradato; mentre la fase di quantizzazione delle librerie dei campioni di DNA degradato la fase è stata affidata allo strumento *Applied Biosystem<sup>®</sup> 7500 Real-Time PCR System*. I risultati della quantificazione sono riportati nelle *Tabelle 18 e 19*.

### 3.6 Diluizione, raggruppamento delle librerie e posizionamento sullo ION

#### CHEF

- Al completamento della qPCR, calcolare la concentrazione media di ciascuna libreria non diluita utilizzando la seguente equazione:

$$\text{Concentrazione media della libreria non diluita} = (\text{quantità media della qPCR}) \times (\text{diluizione della libreria})$$

- Diluire le librerie come descritto in *Tab.16*:

Dilute to	Minimum volume	Templating size in Planned Run setup
30 pM	25 µL	200 bp

*Tab16. Diluizione della libreria*

Ogni libreria è stata diluita a 30pM in Low TE, in *Tab.17* e in *Tab.18* vengono riportate le opportune quantità.

SAMPLE	VOLUME CAMPIONE	VOLUME TE
ANCS031	2 µL	128,0 µL
ANCS032	4 µL	172,0 µL
ANCS033	5 µL	21,9 µL
ANCS034	5 µL	47,4 µL
ANAS060	5 µL	13,3 µL
ANAS061	5 µL	37,9 µL
ANAS064	5 µL	41,4 µL
ANAS092	5 µL	27,9 µL
ANAS093	5 µL	31,8 µL
ANAS103	5 µL	22,5 µL
ANCS017	5 µL	8,2 µL
ANCS018	5 µL	13,7 µL
ANCS019	5 µL	15,6 µL
ANCS020	5 µL	5,2 µL
ANCS002	5 µL	10,0 µL
ANCS005	5 µL	9,2 µL
ANCS023	5 µL	10,9 µL
ANCS001	5 µL	18,9 µL

ANCS011	5 µL	28,8 µL
ANCS021	5 µL	7,8 µL
ANCS027	5 µL	38,8 µL
ANCS029	5 µL	34,1 µL
ANAS001	5 µL	36,5 µL
ANAS014	5 µL	77,5 µL
ANAS015	5 µL	17,0 µL
ANAS018	5 µL	30,2 µL
ANAS039	5 µL	18,9 µL
ANAS040	5 µL	61,4 µL
ANAS046	4 µL	99,3 µL
ANAS047	5 µL	60,4 µL
ANAS054	5 µL	33,5 µL
ANAS062	5 µL	25,8 µL
ANAS070	5 µL	15,8 µL
ANAS071	5 µL	39,4 µL

*Tab.17 Diluizioni con le opportune quantità di TE, campioni per studio popolazionistico.*

SAMPLE	VOLUME CAMPIONE	VOLUME TE
CAMPIONE 1	5 µL	109 µL
CAMPIONE 2	5 µL	100 µL
CAMPIONE 3	5 µL	165 µL
CAMPIONE 4	5 µL	214 µL
CAMPIONE 5	4 µL	274 µL
CAMPIONE 6	5 µL	39 µL
CAMPIONE 7	5 µL	48 µL
CAMPIONE 8	2 µL	272 µL
CAMPIONE 9	3 µL	221 µL
CAMPIONE 10	4 µL	208 µL
CAMPIONE 11	2 µL	144 µL
CAMPIONE 12	1 µL	301 µL
CAMPIONE 13	5 µL	189 µL
CAMPIONE 14	4 µL	177 µL
CAMPIONE 15	4 µL	227 µL
CAMPIONE 16	5 µL	121 µL
CAMPIONE 17	5 µL	46 µL
CAMPIONE 18	5 µL	121 µL
CAMPIONE 19	5 µL	35 µL
CAMPIONE 20	5 µL	46 µL
CAMPIONE 21	5 µL	3 µL

*Tab.18 Diluizioni con le opportune quantità di TE, campioni reali forensi.*

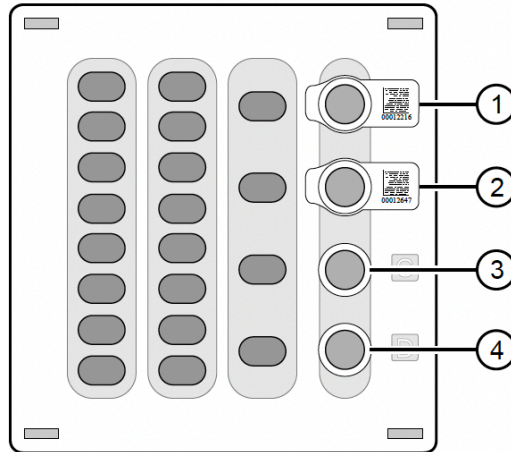
Dopo aver diluito la libreria di campioni alla sua concentrazione target (pM), raggruppare volumi uguali di più librerie in una provetta.

### ***3.7 Preparazione del template su HID ION CHEF***

La preparazione del template sullo strumento Ion Chef consiste nell'amplificazione clonale di sequenze target in una libreria su *Ion Sphere Particles* (ISP). Queste sono caricate dallo strumento su un chip di sequenziamento: *Ion Chip*. Per ogni chip Ion possono essere sequenziati contemporaneamente fino a un massimo di 34 campioni ed è necessario creare una corsa pianificata seguendo le istruzioni (*Precision ID SNP Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System*) utilizzando il software *Torrent Suite*. Sono state eseguite due corse poiché non tutti i campioni potevano essere caricati su un unico chip.

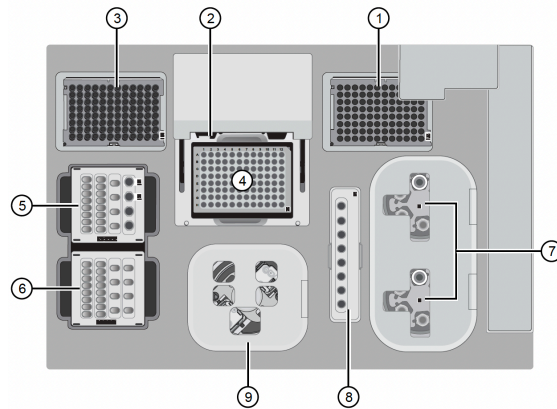
La configurazione dello strumento per la corsa richiede le seguenti procedure:

- Disimballare le cartucce dei reagenti Ion S5 Precision ID Chef 45 minuti prima dell'uso, per consentire la stabilizzazione a temperatura ambiente, pipettare 25 µl di pool di librerie diluite sul fondo dell'appropriata Ion Chef Library Sample Tube



*Fig.17 Cartucce per i reagenti dello Ion S5 Precision ID Chef. Posizione 1 (pool di librerie di DNA), Posizione 2 (pool di librerie di DNA), Posizione 3 (NaOH) e Posizione 4 (provetta vuota)*

- 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, poi metterli sulla postazione accanto allo Ion Chef.
- Caricare i puntali delle pipette e la piastra PCR.
- Caricare le cartucce dei reagenti e le soluzioni.
- Caricare le provette di ripristino e le cartucce di arricchimento v2.
- Caricare il Chip-loading sulla centrifuga.
- Verificare che i materiali di consumo siano installati correttamente (Fig.18).



A schematic of a loaded Ion Chef™ Instrument

- |   |   |
|---|---|
| ① Empty tip rack (move from new Tip Cartridge position) | ⑥ Ion S5™ Precision ID Chef Solutions cartridge         |
| ② Frame Seal v2   | ⑦ Recovery Tubes and Recovery Station Disposable Lid v2 |
| ③ New Tip Cartridge v2                                  | ⑧ Enrichment Cartridge v2                               |
| ④ PCR Plate   | ⑨ Chip Adapter/Chip assemblies                          |
| ⑤ Ion S5™ Precision ID Chef Reagents cartridge          |   |

*Fig.18 Strumento Ion Chef caricato di tutti i materiali di reazione*

- Avviare la corsa Ion Chef selezionando l'opzione sullo schermo touch nella home.
- Seguire passo dopo passo per fare in modo che lo strumento vada verso la configurazione seguendo le istruzioni sullo schermo.
- Quando richiesto, chiudere lo sportello dello strumento. Il sistema si attiverà.
- Toccare avvia controlli per avviare Deck Scan, in cui lo strumento esegue la scansione dei codici a barre di tutti materiali di utilizzo e dei reagenti.
- Quando la Deck Scan sarà completata, toccare sul display Data Destination, durante il quale lo strumento visualizza il tipo di kit, il

tipo di chip, i codici a barre del chip e le corse pianificate. Se tutte le opzioni sono corrette, tocca Next.

- Il menu Run Option consente di impostare l'ora di completamento dell'esecuzione, se necessario. Successivamente premere Start Run.
- Inizializzare lo Ion S5™ almeno 40 minuti prima che Ion Chef™ System termini il caricamento del chip.
- Al termine della corsa, scaricare lo strumento e sequenziare il chip.

### ***3.8 Sequenziamento con HID Ion GeneStudio S5™ System***

Il sequenziatore *HID Ion S5™* è equipaggiato per verificare la compatibilità di ogni chip e il consumo dei materiali che vengono caricati durante l'inizializzazione e il sequenziamento. Prima dell'inizio della reazione di sequenziamento:

- Disimballare le cartucce *Ion S5™ Precision ID Sequencing Reagents* 45 minuti prima dell'utilizzo e lasciarli equilibrare a temperatura ambiente.
- Disimballare il flacone della soluzione di lavaggio *Precision ID*. Capovolgere la bottiglia 5 volte, per mescolare bene. Togliere dal suo sacchetto sottovuoto, togliere il tappo rosso dalla soluzione di

lavaggio e dai detergenti immediatamente prima di installare lo strumento.

- Inizializzare toccando Initialize nel menu principale della schermata touch.
- Rimuovere il flacone della soluzione di lavaggio Precision ID per accedere e svuotare i rifiuti serbatoio, e 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.
- Lasciare il chip che è stato sequenziamento nel cumulo dei chip, 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 di sequenziamento
- Verificare che sia selezionata la corsa pianificata corretta, premere Review, confermare le impostazioni sono corretti o apportare modifiche.
- Chiudere lo sportello dello strumento, premere Start-Run per iniziare la corsa di sequenziamento.



### 3.9 Analisi dei dati

Al termine dell'MPS, tutti i sequenziamenti sono stati processati con il *Torrent Suite Software* e le reads allineate con la sequenza di riferimento del genoma umano (hg19). L'analisi degli aplotipi è stata eseguita con i plug-in, *HID\_Microhaplotype\_Research\_PluginV1.5*, proposto da Oldoni et. al per la validazione del pannello da 74 MH<sup>(29-49)</sup>. Di seguito i parametri relativi al plug-in:

Global Parameters	
Min total read coverage per position	<input type="text" value="20"/>
Min # of allele count to include in report	<input type="text" value="5"/>
Min allele frequency (heterozygous)	<input type="text" value="10"/>
Min allele frequency (homozygous)	<input type="text" value="90"/>

Fig.19. Parametri del plug-in *HID\_Microhaplotype\_Research\_PluginV1.5*<sup>(49)</sup>

Oltre al plugin relativo ai microaplotipi abbiamo lanciato anche i plug-in del *Coverage Analysis V5.12.00*, il quale restituisce informazioni per singolo amplicone di ciascun campione e del *Variant Caller V5.12.04*. Successivamente per confermare e revisionare le chiamate alleliche ottenute è stata effettuata un'analisi secondaria con il software Integrative Genomics Viewer IGV.

## Capitolo 4

### RISULTATI

#### *4.1 Selezione dei marker*

Il pannello utilizzato in questo studio permette di amplificare un set di 78 microplotipi. Il presente pannello è costituito da 17 MH formati da 3 SNP, 41 MH da 4 SNP, 19 da SNP e 2 MH da 6 SNP.

Come già anticipato precedentemente (*3.1 Selezione dei loci d'interesse*), in alcuni casi sono stati aggiunti degli SNP presenti nelle vicinanze, mediante visualizzazione manuale su Esembl.

Mediante un controllo su dbSNP (Single Nucleotide Polymorphism Database) <sup>(38)</sup>, è stato possibile notare che lo SNP rs149217113 presente sul cromosoma 1, è costituito da un InDel (abbreviazione che indica un evento di mutazione/ricombinazione che può far parte di due classi: una inserzione [in] e una delezione [del]), per cui è stato eliminato durante l'analisi.

Il totale di microplotipi analizzati sarà quindi 77 e il totale degli SNP è 298 sparsi sui 22 cromosomi umani. Il range di lunghezza dell'amplicone è compreso tra le 125-140 bp.

#### 4.2 Risultati della quantificazione delle librerie

Al termine dell'amplificazione, i campioni sono stati quantificati utilizzando il *Ion Library TaqMan Quantitation Kit*. La quantificazione delle librerie dei 34 campioni di DNA per lo studio popolazionistico è stata effettuata in *Real-Time PCR* con *Rotor-Gene 6000 PCR System®* e i dati in tempo reale sono stati analizzati con il software *Rotor Gene Qv.2.1.0* della *Qiagen*.

In Tab.18 sono riportate le concentrazioni ottenute effettuate in duplicato, il valore medio delle due concentrazioni viene moltiplicato per 100 poiché prima della quantificazione è stata eseguita una diluizione delle librerie 1:100.

Sample	qPCR1	qPCR2	Media	Media*100
ANCS020	0,55	0,67	0,61	61,03
ANCS021	0,77	0,77	0,77	76,73
ANCS017	0,75	0,83	0,79	79,13
ANCS005	0,74	0,97	0,85	85,31
ANCS002	0,81	0,99	0,90	90,11
ANCS023	0,96	0,95	0,96	95,69
ANAS060	1,12	1,07	1,10	109,84
ANCS018	1,08	1,17	1,12	112,18
ANCS019	1,21	1,26	1,24	123,54
ANAS070	1,19	1,31	1,25	125,06
ANAS015	1,37	1,27	1,32	131,97
ANCS001	1,46	1,40	1,43	143,20
ANAS039	1,40	1,47	1,43	143,30
ANCS033	1,59	1,64	1,62	161,62
ANAS103	1,67	1,63	1,65	165,03
ANAS062	1,81	1,88	1,85	184,67
ANAS092	2,12	1,82	1,97	197,10
ANCS011	1,94	2,12	2,03	202,95
ANAS018	2,10	2,12	2,11	210,95
ANAS093	2,30	2,12	2,21	220,88
ANAS054	2,14	2,48	2,31	230,90
ANCS029	2,21	2,49	2,35	234,87
ANAS001	2,59	2,39	2,49	249,12
ANAS061	2,70	2,45	2,57	257,34

ANCS027	2,67	2,58	2,63	262,87
ANAS071	2,61	2,72	2,66	266,48
ANAS064	2,69	2,88	2,78	278,46
ANCS034	3,08	3,21	3,14	314,41
ANAS047	3,89	3,96	3,92	392,24
ANAS040	4,03	3,94	3,98	398,50
ANAS014	5,00	4,90	4,95	495,04
ANAS046	7,78	7,72	7,75	775,12

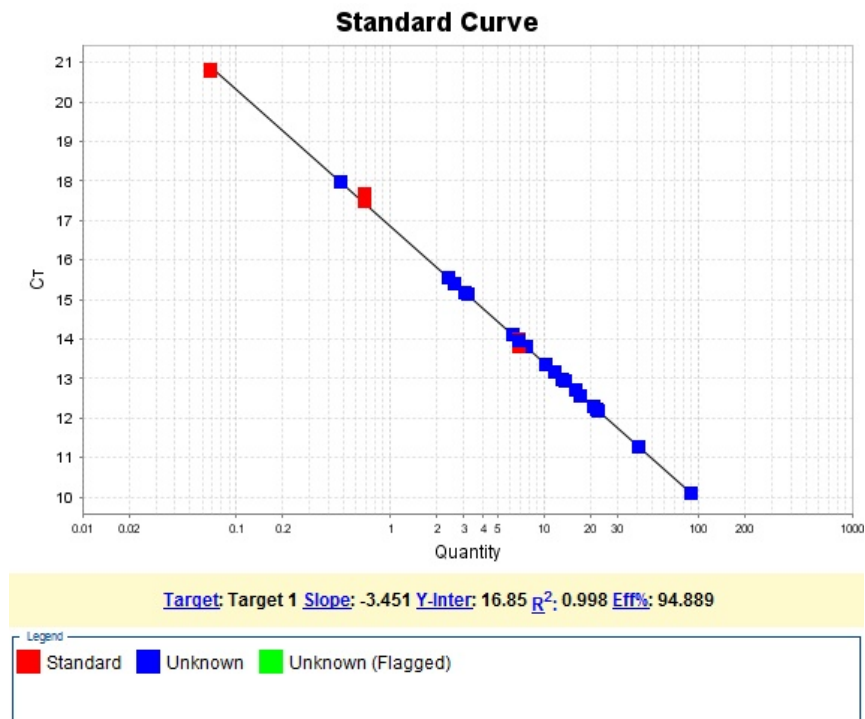
*Tab.18 Risultati della quantizzazione delle librerie in qPCR*

La quantificazione dei campioni forensi reali è stata effettuata con *Applied Biosystem* ®7500 *Real-Time PCR System* associato al software *HID Real-Time PCR Analysis*. I risultati sono visualizzati in *Tab.19*.

Sample	qPCR	Concentrazione *100
CAMPIONE 1	6,85	685
CAMPIONE 2	6,27	627
CAMPIONE 3	10,22	1021
CAMPIONE 4	13,16	1316
CAMPIONE 5	20,82	2082
CAMPIONE 6	2,62	262
CAMPIONE 7	3,16	316
CAMPIONE 8	41,12	4112
CAMPIONE 9	22,43	2243
CAMPIONE 10	15,91	1591
CAMPIONE 11	21,88	2188
CAMPIONE 12	90,62	9062
CAMPIONE 13	11,62	1162
CAMPIONE 14	13,60	1360
CAMPIONE 15	17,33	1733
CAMPIONE 16	7,58	758
CAMPIONE 17	3,05	305
CAMPIONE 18	7,56	756
CAMPIONE 19	2,38	238
CAMPIONE 20	3,05	305
CAMPIONE 21	0,47	47

*Tab.19 Risultati della quantizzazione delle librerie in qPCR*

Nella *Fig.20* viene riportata la retta di taratura relativa alla quantificazione delle librerie dei campioni con DNA degradato. I parametri di  $R^2$  (0,998) e della slope (-3,451), sono considerati idonei per valutare efficace la qPCR.



*Fig.20* Retta di taratura delle librerie relative ai campioni forensi reali.

### 4.3 Risultati analisi delle prestazioni del sequenziamento

Sono state eseguite due corse di sequenziamento utilizzando due Ion 520™ Chip. Le metriche della corsa del primo chip sono riportate nelle *Fig.21* e *Fig.22*:

## Run Summary

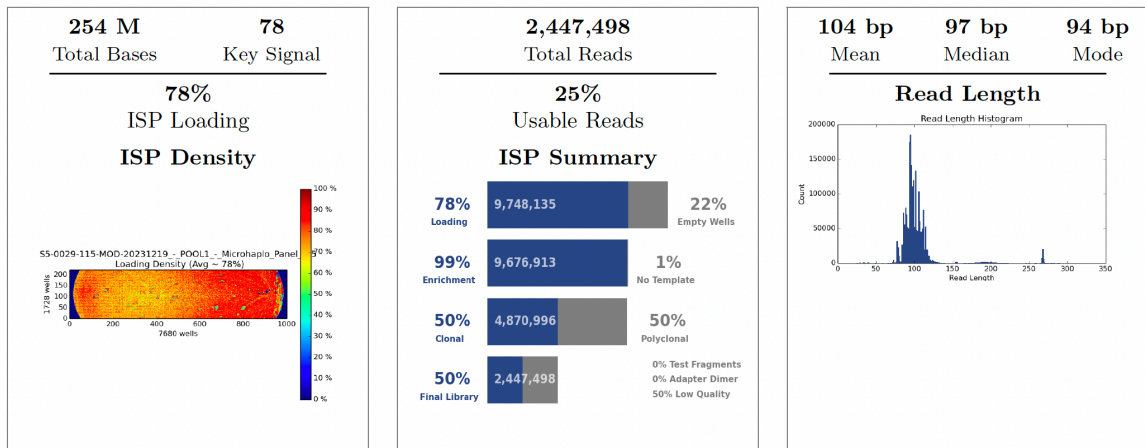


Fig.21 Run summary della corsa di sequenziamento

Nel primo riquadro della Fig.21 è possibile osservare il totale di basi sequenziate (254M) ed è inoltre rappresentato il carico effettivo delle *Ion Sphere Particles* (ISPs) sul chip con un valore del 78%, le aree raffigurate in rosso sono quelle con un caricamento ottimale, le aree con densità di caricamento inferiori variano dal giallo al verde, mentre le aree blu indicano un caricamento non riuscito. Il numero di letture totali è illustrato nel secondo riquadro ed è di 2,447,498, di queste secondo lo strumento *Ion S5*, solo il 25% sono letture utilizzabili poiché superano determinati parametri, tra cui, l'avvenuto caricamento, l'arricchimento e fanno parte delle ISP clonali. È stato osservato un livello relativamente elevato di policlonalità, una percentuale alta di ISPs policlonali è indice di un numero elevato di campioni in analisi o di una non corretta quantificazione delle librerie. Per poter considerare questa percentuale in accordo con i parametri di riferimento, il suo valore dovrebbe

essere compreso tra 25-50% <sup>(45)</sup>. Nel presente studio, è stato ottenuto un intervallo del 50%.

Nel terzo riquadro è riportato l'istogramma che restituisce le informazioni in merito alla lunghezza di tutte reads sequenziate, troviamo infatti i valori della Media, Mediana e Moda.

**Alignment Summary (aligned to *Homo sapiens*)**

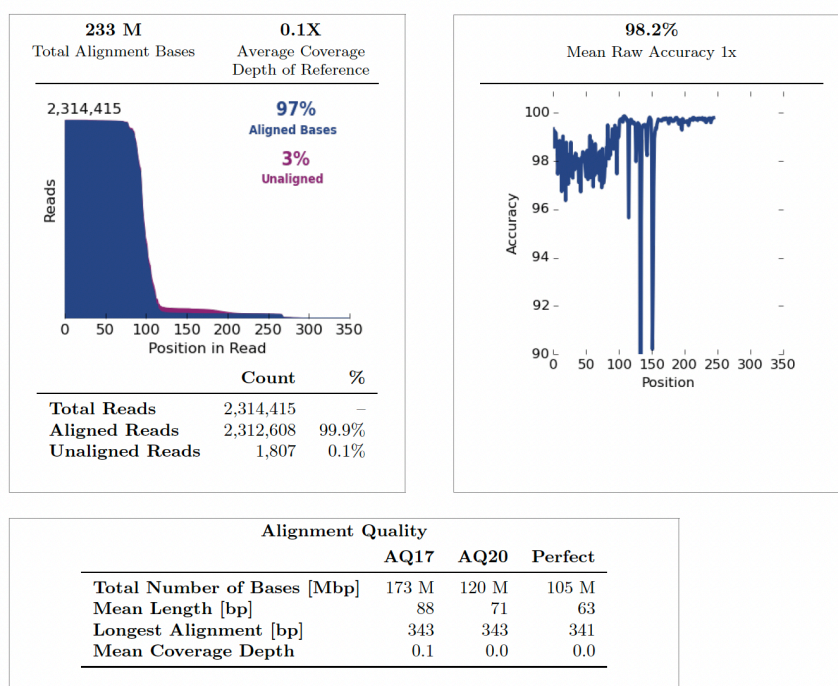


Fig.22 Metriche post-allineamento incluse nel report di esecuzione

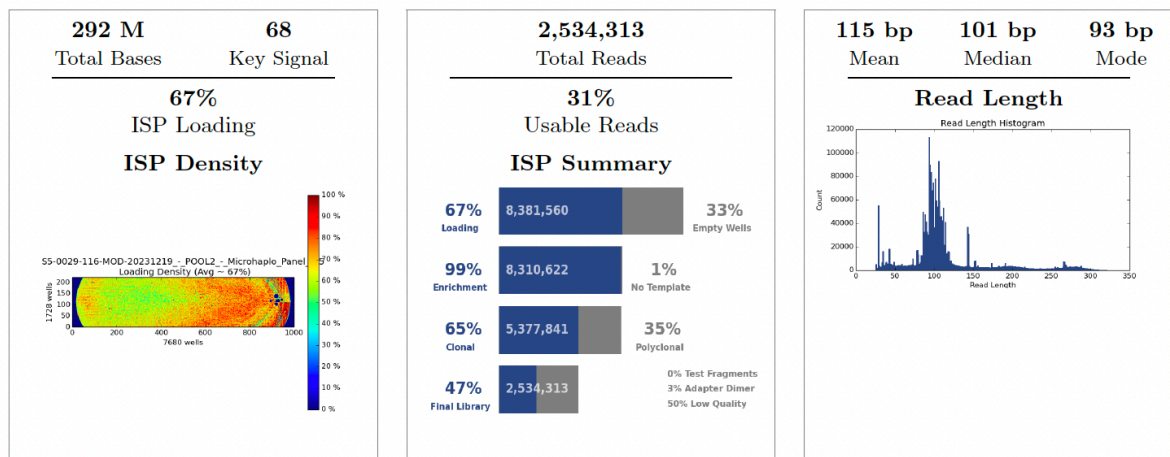
Il *Run Report* restituisce anche informazioni che riguardano le metriche post-allineamento (Fig.22), lo strumento è stato impostato per l'allineamento con *Homo Sapiens*.

Il grafico di allineamento riporta in blu le basi allineate, in questo caso 97% e in viola quelle non allineate, 3%.

Viene rappresentata, grafico a destra in *Fig.22*, l'accuratezza media calcolata per ciascuna posizione in ogni lettura che risultano essere del 98,2%. Infine, vengono visualizzati anche i parametri di qualità.

Le metriche di corsa del secondo chip sono rappresentate nella *Fig.23* e e nella *Fig.24*.

### Run Summary



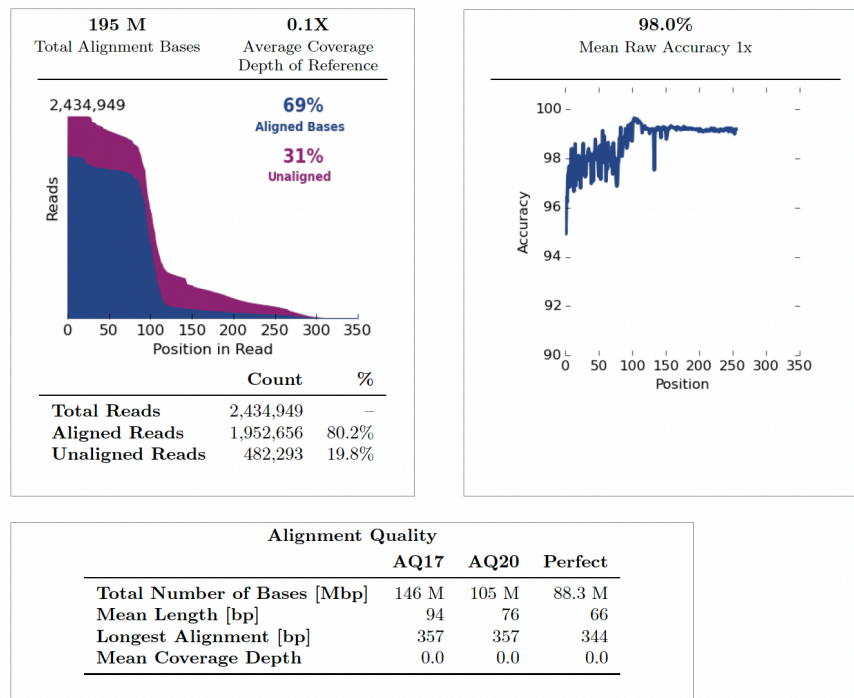
*Fig.23 Run summary della corsa di sequenziamento*

Nella *Fig.23* notiamo che il carico effettivo delle *Ion Sphere Particles* (ISPs) sul chip presenta un valore del 67%. Il numero di letture totali è di 2.534,313, di queste il 31%, secondo lo strumento, sono letture utilizzabili. Anche per questo pool è stato osservato un livello relativamente elevato di policlonalità, intorno al 35%. Nell'ultimo riquadro vediamo rappresentate le metriche post-



allineamento troviamo, come prima, la Media, Mediana e Moda della lunghezza delle letture riportate.

**Alignment Summary (aligned to *Homo sapiens*)**



*Fig.24 Metriche post-allineamento incluse nel report di esecuzione*

Nel primo riquadro è possibile osservare le basi totali incluse nell'allineamento. Il grafico di allineamento riporta in blu le basi allineate, in questo caso 69% e in viola quelle non allineate, 31%. Nel secondo e terzo riquadro vengono riportati rispettivamente: l'accuratezza media calcolata e i parametri di qualità dell'allineamento.

Nelle *Tab.20* e *Tab.21* vengono elencati, rispettivamente per i campioni presenti nel pool 1(campioni di DNA per studio popolazionistico) e nel pool 2

(campioni forensi reali), i *barcode* utilizzati, il numero di basi sequenziate, il numero di basi sequenziate *on target*, numero di *reads* e lunghezza media degli ampliconi di ogni campione.

BARCODE	CAMPIONE	BASI	≥Q20	READS	MEAN READ LENGTH
IonXpress_001	ANAS015	10,303,626	9,703,831	99,002	104 bp
IonXpress_002	ANAS018	7,257,305	6,846,377	68,707	105 bp
IonXpress_003	ANAS039	8,041,939	7,585,886	77,070	104 bp
IonXpress_004	ANAS040	7,479,883	7,053,254	72,290	103 bp
IonXpress_005	ANAS046	9,923,128	9,352,502	92,395	107 bp
IonXpress_006	ANAS047	9,151,673	8,566,800	88,767	103 bp
IonXpress_007	ANAS054	8,050,210	7,541,806	77,578	103 bp
IonXpress_008	ANAS062	8,175,578	7,630,672	80,154	101 bp
IonXpress_009	ANAS070	8,423,930	7,911,266	80,696	104 bp
IonXpress_010	ANAS071	9,382,084	8,778,229	90,874	103 bp
IonXpress_041	ANCS031	559,155	524,228	5,321	105 bp
IonXpress_042	ANCS032	361,317	343,139	3,592	100 bp
IonXpress_043	ANCS033	6,225,041	5,836,736	61,491	101 bp
IonXpress_044	ANCS034	11,338,607	10,686,698	111,622	101 bp
IonXpress_045	ANAS060	6,519,863	6,125,150	63,053	103 bp
IonXpress_046	ANAS061	7,093,103	6,664,546	67,094	105 bp
IonXpress_047	ANAS064	5,485,860	5,174,338	52,614	104 bp
IonXpress_048	ANAS092	5,728,877	5,396,574	56,431	101 bp
IonXpress_049	ANAS093	7,428,609	7,000,391	72,776	102 bp
IonXpress_050	ANAS103	7,266,583	6,832,859	70,723	102 bp
IonXpress_051	ANCS017	5,346,351	5,022,999	51,234	104 bp
IonXpress_052	ANCS018	6,716,512	6,293,716	64,646	103 bp
IonXpress_053	ANCS019	6,458,057	6,069,976	63,620	103 bp
IonXpress_054	ANCS020	7,349,954	6,941,562	71,969	102 bp
IonXpress_055	ANCS002	8,320,741	7,800,689	82,880	100 bp
IonXpress_056	ANCS005	6,713,359	6,351,609	66,006	101 bp
IonXpress_057	ANCS023	6,947,005	6,508,462	68,268	101 bp
IonXpress_058	ANCS001	6,056,991	5,694,878	59,566	101 bp
IonXpress_059	ANCS011	6,258,337	5,848,648	61,283	102 bp
IonXpress_060	ANCS021	6,827,343	6,400,501	68,054	100 bp
IonXpress_061	ANCS027	5,434,504	5,134,042	52,829	102 bp
IonXpress_062	ANCS029	6,941,597	6,487,909	67,295	103 bp
IonXpress_063	ANAS001	7,656,079	7,179,838	72,888	105 bp
IonXpress_064	ANAS014	7,722,719	7,180,771	71,627	107 bp

Tab.20 Numero di basi e di reads per ogni campione (Pool1)

BARCODE	CAMPIONE	BASI	≥Q20	READS	MEAN READ LENGTH
IonXpress_065	CAMPIONE 1	24,890,936	23,381,018	192,282	129 bp
IonXpress_066	CAMPIONE 2	14,322,861	13,554,727	141,596	101 bp
IonXpress_067	CAMPIONE 3	17,198,175	16,221,483	132,541	129 bp
IonXpress_068	CAMPIONE 4	14,100,152	13,376,992	121,464	116 bp
IonXpress_069	CAMPIONE 5	10,639,322	10,060,126	90,111	118 bp
IonXpress_070	CAMPIONE 6	15,692,599	14,827,207	113,734	137 bp
IonXpress_071	CAMPIONE 7	5,987,105	5,664,932	61,334	97 bp
IonXpress_072	CAMPIONE 8	11,403,930	10,651,077	73,512	155 bp
IonXpress_073	CAMPIONE 9	15,838,147	14,683,705	107,154	147 bp
IonXpress_074	CAMPIONE 10	12,980,326	12,269,788	126,109	102 bp
IonXpress_075	CAMPIONE 11	14,553,581	13,819,123	104,399	139 bp
IonXpress_076	CAMPIONE 12	21,840,742	20,684,910	175,100	124 bp
IonXpress_077	CAMPIONE 13	3,733,285	3,531,214	36,756	101 bp
IonXpress_078	CAMPIONE 14	3,880,515	3,669,782	32,232	120 bp
IonXpress_079	CAMPIONE 15	3,307,211	3,113,828	31,729	104 bp
IonXpress_080	CAMPIONE 16	12,136,938	11,505,813	117,925	102 bp
IonXpress_081	CAMPIONE 17	22,939,903	21,623,328	226,468	101 bp
IonXpress_082	CAMPIONE 18	19,031,047	17,975,874	185,565	102 bp
IonXpress_083	CAMPIONE 19	17,387,162	16,482,214	172,723	100 bp
IonXpress_084	CAMPIONE 20	17,980,882	16,987,548	178,884	100 bp

Tab.21 Numero di basi e di reads per ogni campione (Pool 2)

#### 4.4 Analisi del Coverage

Il plug-in del *Coverage Analysis V5.12.00* genera un Coverage Analysis Report in cui sono riportati i valori di convergenza ottenuti dal sequenziamento di ogni campione.

In particolare, nella *Tab.22* e *Tab.23* troviamo: i barcode utilizzati per ogni campione, *mapped reads*, *on target*, *mean depth* e *uniformity*<sup>(46)</sup>. Per *mapped reads*, si intende il numero totale di letture mappate all'interno del genoma di riferimento (hg19). Il valore di *on target* invece indica il numero di reads che ricadono nel target scelto. Invece i valori di *mean depth* e *uniformity*, indicano

rispettivamente: le statistiche di riepilogo per le letture assegnate agli ampliconi e che il target è stato sequenziato in maniera uniforme.

Barcode ID	Sample Name	Mapped Reads	On Target	Mean Depth	Uniformity
IonXpress_001	ANAS015	98930	95,55%	1212	96,33%
IonXpress_002	ANAS018	68681	94,88%	833,3	96,33%
IonXpress_003	ANAS039	77023	95,12%	947,4	92,92%
IonXpress_004	ANAS040	72241	96,27%	895,7	93,01%
IonXpress_005	ANAS046	92344	95,86%	1141	95,40%
IonXpress_006	ANAS047	88714	95,89%	1097	94,10%
IonXpress_007	ANAS054	77489	95,66%	948,8	97,36%
IonXpress_008	ANAS062	80070	96,16%	988,9	96,44%
IonXpress_009	ANAS070	80642	95,64%	991,9	94,05%
IonXpress_010	ANAS071	90801	95,76%	1115	96,31%
IonXpress_041	ANCS031	5318	94,08%	64,52	92,93%
IonXpress_042	ANCS032	3589	97,10%	44,86	94,08%
IonXpress_043	ANCS033	61450	97,24%	767,9	92,92%
IonXpress_044	ANCS034	111566	97,35%	1400	93,09%
IonXpress_045	ANAS060	63003	95,25%	778,5	91,96%
IonXpress_046	ANAS061	67057	95,73%	824,5	94,09%
IonXpress_047	ANAS064	52575	96,00%	647,3	95,37%
IonXpress_048	ANAS092	56367	95,78%	694,8	95,27%
IonXpress_049	ANAS093	72713	95,78%	894,8	94,98%
IonXpress_050	ANAS103	70681	96,06%	874,8	94,75%
IonXpress_051	ANCS017	51194	94,13%	623,6	91,95%
IonXpress_052	ANCS018	64600	94,99%	790,0	94,06%
IonXpress_053	ANCS019	63577	96,44%	791,5	92,87%
IonXpress_054	ANCS020	71899	95,49%	883,3	92,87%
IonXpress_055	ANCS002	82814	95,97%	1018	95,14%
IonXpress_056	ANCS005	65920	95,89%	809,3	95,02%
IonXpress_057	ANCS023	68217	96,57%	846,2	95,12%
IonXpress_058	ANCS001	59489	96,14%	734,1	94,13%
IonXpress_059	ANCS011	61226	96,77%	760,7	95,12%
IonXpress_060	ANCS021	67945	95,58%	832,0	95,10%
IonXpress_061	ANCS027	52795	96,33%	656,5	94,09%
IonXpress_062	ANCS029	67228	96,13%	831,6	95,00%

IonXpress_063	ANAS001	72850	95,90%	902,4	95,21%
IonXpress_064	ANAS014	71600	96,00%	884,4	93,99%

Tab.22 Risultati di coverage del sequenziamento dei campioni per studio popolazionistico.

Barcode ID	Sample Name	Mapped Reads	On Target	Mean Depth	Uniformity
IonXpress_065	CAMP_01	150238	47,65%	906,1	97,55%
IonXpress_066	CAMP_02	140066	97,21%	1745	94,95%
IonXpress_067	CAMP_03	58642	13,76%	88,32	97,89%
IonXpress_068	CAMP_04	23630	15,84%	42,67	96,44%
IonXpress_069	CAMP_05	14578	17,42%	27,24	96,73%
IonXpress_070	CAMP_06	22888	75,79%	186,1	93,35%
IonXpress_071	CAMP_07	58481	97,57%	709,2	95,18%
IonXpress_072	CAMP_08	72072	26,59%	242,4	97,43%
IonXpress_073	CAMP_09	102075	42,31%	543,0	97,44%
IonXpress_074	CAMP_10	125609	98,28%	1587	95,13%
IonXpress_075	CAMP_11	34656	11,59%	47,38	96,66%
IonXpress_076	CAMP_12	173217	95,80%	2128	94,93%
IonXpress_077	CAMP_13	36473	98,86%	458,9	95,99%
IonXpress_078	CAMP_14	28291	48,81%	175,9	97,23%
IonXpress_079	CAMP_15	31653	98,06%	396,9	94,96%
IonXpress_080	CAMP_16	114643	96,15%	1398	96,02%
IonXpress_081	CAMP_17	224267	96,00%	2726	97,00%
IonXpress_082	CAMP_18	183560	96,55%	2242	95,96%
IonXpress_083	CAMP_19	171884	98,55%	2160	96,17%
IonXpress_084	CAMP_20	177368	97,97%	2205	96,90%

Tab.23 Risultati di coverage del sequenziamento dei campioni reali forensi.

Vengono inoltre calcolati i valori del Relative Depth of Coverage (rDoC) normalizzati per singolo amplicone, calcolati come :

## Coverage MH per campione/Coverage totale del campione

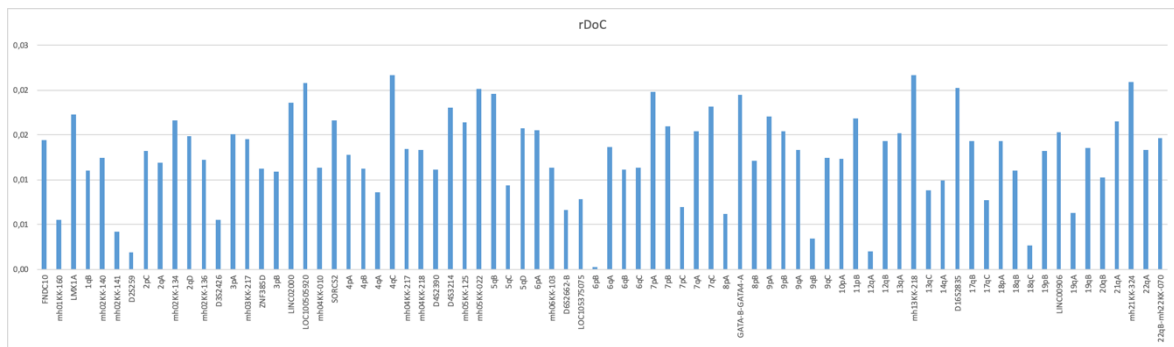


Grafico1. Risultati relativi al Relative Depth of Coverage (rDoC)

Questo è un importante criterio di qualità che ci permette di definire le prestazioni di ciascun amplicone. Infatti, alcuni possono essere considerati under performing, come il microaplotipo 6qB.

### 4.5 Risultati della genotipizzazione

L'analisi primaria delle varianti alleliche è stata eseguita con il software Torrent Suite utilizzando il plug-in *HID\_Microhaplotype\_Research\_PluginV1.5*. Al termine dell'analisi del plug-in, è disponibile il report riepilogativo di ciascun campione insieme ai file BAM (binary Allineamento Index) e BAI (binary Allineamento Index) per le letture mappate ed elaborate dal plug-in dei microaplotipi. Le varianti che non hanno superato il filtraggio del *HID\_Microhaplotype\_Research\_PluginV1.5* sono

state ispezionate manualmente tramite il software IGV (*Integrative Genomics Viewer*), il quale permette di allineare le reads ottenute dal sequenziamento con la sequenza di riferimento (GRCh37/hg19). IGV necessita di un file BED contenente i riferimenti delle regioni cromosomiche analizzate (*target file*) e di un file contenente la lista degli SNPs tipizzati con le relative localizzazioni cromosomiche. Nelle tabelle seguenti vengono riportati la genotipizzazione dei campioni relativi allo studio popolazionistico e la legenda.

	reads <20x
	campioni che presentavano più di due chiamate alleliche

*Tab.24 Legenda relativa alla tabella della genotipizzazione*

CAMPIONI	ANCS011	ANCS005	ANAS092	ANAS062	ANCS029	ANAS001	ANAS071	ANAS046	ANAS070	ANCS027	ANCS001
LMX1A											
FNDC10	GCCG/GCGG	GCCG/GCCG	GCCG	GCCG	GCCG	ACGC/GCCG	GCCG/GCCG	ACGC	GCCG/GCGG	GCCG/GCGG	GCCG
mh01KK-160	TTAA	CGAA/TGAA	CGAA	TGAA	CGAA	CGAA	CGAA/TTAA	CGAA/TTAA	CGAA/TGAA	CGAA/TTAA	CGAA/TGAA
1qB	CGGC/CGAC	TCGT/CGAC	TCGC/CGGC	CGGC/CGAC	CGGC	CGGC/TCGC	CGGC/CGAC	TCGT/CGGC	TCGT/TCGC	CGGC/CGAC	CGGC/CGAC
mh02KK-140	GAGG/CGGA	GAGG	GGAG/CGGA	CGGG/GAGG	CGGG/GGGG	CGGG/GAGG	GGAG/GAGG	CGGG/GGAG	GAGG/CGGG	CGGA/CGGG	CGGG/GAGG
mh02KK-141	GGTG	GGGG	GCTA/GGTG	GGGG	GGGG/GCTA	GGTG/GGGG	GCTG/GCTA	GGGG/GCTA	GGGG/GGTG	GGTG/GCTA	GGGG
D25259	ATGG/ACAG	GCCG/ACGC		ACGC/ATGG	ACGC/GCCG	ACGC/ACAG	ACAG/ATGG	GCCG/ACGC	ACAG/ACGC	ACAG	ACGC/ACAG
2pC	GCTT/GCGG	ACGT/GCCG	GTGT	GCGT/ACGT	GCTT	GCTT/GCGT	GCGT	GCGT/ACGT	GTGT/GCGT	GCTT/ACGT	GCGG
2qA	CTGC/TTAC	TTAC/CTGC	CCAC	CCAC/CTGC	TTAC/CCAC	TTAC/CTGC	CTGC	CTGC	CCAC/TTAC	CCAC/TTAC	TTAC/CCAG
mh02KK-134	ATC/ACT	ACC	ACT/ACC	ATT/ACC	TTT/ACC	TTT/ATC	ATC/ACC	ACT	ACT/ATC	ATC/TTT	TCC/ATC
2qD	GCG/GGG	GGG/AGG	GCG/GGG	AGG/GCG	AGG/GGG	GCG/GCA	GCG	GCA/GCG	AGG/GCA	GCG	AGG/GCG
mh02KK-136	GTC/GTA	TCC/GTA	GTC/TTT	TCA/TTT	GTA	TCA/TTT	GTA	GTC	GTA	TCA/TTT	TCC/TTA
D3S2426	GTC/ATC	GTC/ATC	ATC/ATA	ATC	ATC/ATA	GTC/ATA	GTC/ATC	ATC	ATC	GTC/ATC	ATC/ATA
3pA	CTAT/CCGT	CTGC	CTAC/CCGT	ACGT/CTAT	CTAC/CTAT	CTGC/CCGT	ACGT/CTAC	CTAT/CTAC	ACGT	CCGT/CTAC	CTAC/CTAT
mh03KK-217	AGCT	CGCT/AGCT	CGCT/AGCT	AGCT/CGCT	CGCT/AGCT	AGCT/CGCC	AGCT	AGCC/AGCT	AGCT/AGCC	AGCT	CGCT/AGCT
ZNF385D	CGTG/CATG	CGTG	AATG/CGTG	AATG	AATG	AATG/CGTG	AATG	CGTG/CATG	CATG	CGTG	CATG/AATG
3pB	AACCA/CGCCA	AACCC	AACCG	AACCA/CGCCA	AACCC/AACCA	AACCG/CGCCA	AACCG/CGCCA	AACCG	AACCA/AACCA	AACCC	AACCG/CGCCA
LIN02000	CGTC/CCTC	CCCC/GGTC	GGTC/CGTC	GGTC/CGTC	CCCC/CCTC	CCTT/CCCC	CGTC	CGTC/GGTC	CCCC/GGTC	CGTC	CGTC
LOC100505920	GCGT	GCGT/GCGA	GCGT/GCAT	GCGT	GCGT	GCGT	CTAT/GCGT	CTGA/GCGT	GCGT	CTGA/GCGT	GCGT/GCGA
mh04KK-010	AA	AG/GA	AG	AA/AG	AA	GA/AA	GA	GG/GA	GA	AG/GA	AG/GA
SORCS2	CCT	TAG/CCG	CCT	CCT/CCG	CCT	CCT	CCT	CCT/CCG	CCT	CCG/CCT	CCT/CCG
4pA	GCAC/GTAA	GCAG/GGCA	GTAG/GTAA	GTA/GGCA	CTCA/GGCA	GGCG/GGCA	GTA/GGCA	GTA/GTAG	CTCA/GGCG	GGCA/GTCA	CTCA/GGCA
4pB	GCAC/ATGC	GCAC	GCAC	GCAG/ACGC	GCAC/ATGC	GCAC	GCAG/ACGC	GCAC/ATGC	GCAG/ACGC	GCAC/ATGC	ACGC
4qA	CTT/CTC	CTT	CTT/GCG	GCG/CCG	CCG/CTT	CCT/CCG	GCG/CTT	CTT/CTC	CCT/CTT	GCG/CTT	CTT/CTC
4qC	TTAG	TTAG	TTAG/TAGG	TAGG/GGGG	TAGG/TTAG	GGGG/TTAG	TTAG/TAGG	TAGG/TTAG	TTAG/TAGG	TTAG	TAGG
mh04KK-217	ACT	ATC/ACC	ACT/ATC	ACC	ATC	ATT/GTC	GTC/ACC	ATC/ACC	ATC	ACT	GTC/ATC
mh04KK-218	GGTC	GGTC	GGCC/AACT	AACG/GGCC	GGTC/AACC	GGTC/AACC	AACG/GGCC	AACC/AACT	GGCC	GGCC/GGTC	AACG/GGTC
D4S2390	AGAGC	AGGCT	GGAGC/AGAGT	GGAGC/AGAGC	AGAGC	AAAGT/AGAGC	AGAGC/AAAGT	AGGGC/AAAGT	AGAGC	AGAGC/AAAGT	GGAGC
D4S3214	AGGG/AAAA	AAAG	AAAG/AAAA	GGAG/AAAA	GGAG/AGGG	GGAG/AGGG	AGGG/AAAG	GGAG/AAAA	AGGG	GGAG/AGAG	GGAG
mh05KK-125	AATG/GGTG	AATG/GATG	GATG/AATG	GATG/GATG	GGT/AATG	GATG	AATG/GGTG	GGTG/GATG	GGTG/GATG	AATG/GGTG	GATG
mh05KK-022	CA	CC	CA/TC	TC	CA/CC	TC	CC/CA	CC/TC	CA	CA	CC
5qB	CCAT	TTAC/TTGC	TTAC	TTAC/TTGC	TTAC/CCAC	TTAC/CCAC	TTAC/CAAC	CCAT	TTGC/CCAT	TTGC/CCAC	CCAT/CCAC
5qC	CCGT	CTGT/TCAC	CCGC/CCAC	CTGT/CCGT	CCGC	CTGT/TCAC	CTGT/CCGT	CCGT/CCGC	TTGT/CCGT	CCGT/CCGC	CCGC/CCGT
5qD	GAA/AAAG	GAA	GAA/GGG	GAA/GAG	GAA	AAG/GAG	GAA	GAA/GGG	AAG/GAG	GGG	GAA/GAA
6pA	CCT/CCA	GCT/CCA	CCA/GCT	CCT/GCT	CCT/CTT	CCA/GCT	GCT	CCT/CTT	CCT/CCA	CCT/GCT	CCT/CTT
mh06KK-103	CCA/CCTG	CCTG/CTCA	CTCA/CCTG	CCTG	CTCA/CCTA	CCTG/CTCA	CCTA/CCTG	CCTA/CCTG	CCTG	CTCA/CCA	CCTG
D6S2662-B	AGGT/TGGT	AGGT/TGGT	TGGT/AGGA	TGGT/AGGA	AGAA/TGGT	TGGT/AGGA	AGAA/TGGT	TGGT	AGAA/AGGT	TGGT	TGGT
LOC105375075	GTTT/TTCC	GTTT/TTCC	TTCT/GTTC	TTCT/GTTC	TCTT/TTCC	TCTT/TTCC	TCTT/TTCC	GTTT	TTT/TTCC	TCTT/GTTC	TCTT/GTTC
6pB	GCAA/GTAG	<20x	<20x	GTA/GTAG	GTAG/GTAA	GTA/GCAA	GCAA	<20x	GCAA	<20x	GCAA
6qA	GCT	CCT/GCC	GCC	CCT/GCT	GCT	GCT	GCC/GCT	GCT	CCT/GCC	GCT/GCT	GCT
6qB	ACAG/GTGG	GTGG	ACAG/GTGG	ACAG/GTGG	ACAG/GTGG	GCAC/ACAG	ACAG/GCGG	GCGG/GTGG	GTGG	GTGG	ACAG/GTGG
6qC	ATC/CTC	ATT/ATC	ATT/ATC	ATT/CTC	ATC/CTC	CTC	CTC	CTC	ATT/ACC	ATC	ATT
7pA	AGTC/TATC	AGTC/TATC	AAAGA/AGTC	AAAGA/AGTC	AAAGA/ATA	AAAGA/AGTC	AAAGA/AGTC	AAAGA/TATC	AAAGA/TATC	AAAGA/TATC	AAAGA/TATC
7pB	AGGC/AAGC	AAGC	AGGC	TAGT/TAAC	AGGC	TAGT/TAAC	AGGC/AAGC	AAGC	TAGT/AGGC	AGGC/AAAC	AGGC/AAAC
7pC	CTCGT/CCTAC	CTCGT/CCTAC	CTCGT	CTCGT/TCGCG	CTCGT	CTCGT	TCGCG/CCTAC	CTCGT/CCTAC	CCTAC	TCCGC	CTCGT
7qA	CGC	CAT/CAC	CGC	CAC	CGC	CAC/CAT	CAT	CAT	CGC/CAC	CAT	CGC
7qC	TACGC/AGTGC	AGTGC	AGCAC/AGTGC	TACGC/AGTGC	TACGC	TACGC/AACGC	TACGC/AGTGC	AGCAC/AGTGC	AGCAC/AGTGC	AGCAC/AGTGC	AGCAC/AGTGC
8pA	CGGT/CAGTG	CGGT/CAGTG	CGGT/CAGTG	CGGTA	CGGT/CAGTG	CGATG/CGGTA	CGGT/CAGTA	CGGT/CAGTG	CGGT/CAGTG	CGGT	CGGT
GATA-B-GATA4-A	CCTAG	CCTAG/TTTAA	TTTAA/CTCAA	CTCAA/CCCAA	CCTAG	CCTAG/CTCAA	CCTAG/CCCAAG	CCTAG/CTCAA	CTCAA/CCTAG	CTCAA/CCTAG	CTCAA
8pB	CATCA	CTCTG	CATCA	CACCA/CATCA	CATCA/CTCTG	CATCA	CATCA	CATCA/CTCTG	CATCA	CTCTG/CATCA	TTCTA/CATCA
9pA	AGTCAC/GGCCAC	AGTCAC/GATCAA	AGTCAC	GATCAA/AGTCAC	AGTCAC/GGCCAC	AATGC/GGTCGC	GATCAA/GGCCAC	GATTCG/GATCAA	AGTCAC/GATCAA	GATCAA/AGTCAC	GGTCGC/GATCAA
9pB	TCGC	TCGT	GTAC/TCGC	TCGT	TCGT	TCGT/TCGC	GTAC/TCGT	TCGT/TCGC	TCGT	GTAC	GTAC/TCGC
9qA	GGC/TGC	GGC	GGC/GGG	GGG/TCG	TGC	GGC/TTC	GGG/TCC	GGG/TCC	GGC/TCC	GGC	TGC/GGC
9qB	CATT	TGCT/TGTT	TGCT/CATT	TGCT/CGTT	CGTT	CGTT	TGCT/CGTT	TGCT	CGTT/CATT	TGCT/CATT	TGCT
9qC	GTGGA/ATCGC	ACCGC	ACCGC/GCTGA	ATCGC	ATCGC	ATCGC/ATCGA	GCTGA/ATCGC	GCTGA/ATCGC	ATCGC/GCCGC	ACCGC	ATCGC/GCCGC
10pA	CTAG/CTGG	CTGG	CTAG	TAGG/CTGG	CTGG/CTAG	CTGG/TAGG	TAGG/CTGG	CTAG/CTGG	CTGG/TAGG	CTAG	CTGG
11pB	TCGGG	TCGGG	TTGG/CCGTG	TCGGG/CCATG	TTGG/TCGGG	CCATG/TTGGG	TCGGG	TCGGG/CCATG	CCATG	TTGGG/CCATG	TTGGG/CCATG
12qA	ATGA/ATAA	GGGG	GTGG/ATGA	GTGG/ATAA	GTGG/ATGA	GTGG/ATAA	GTGG/ATGA	GGGG/ATGA	GGGG/ATGA	GTGG/ATAA	GTGG/ATAA
12qB	GAAC/AAAT	GCAC/GACC	GCAC/GAAT	GCAC/GAAC	GAAC/AAAT	GAAC/AAAC	AAAT	GAAC	GCAC/AAAT	GAAC/ACAC	GCAC/GAAT
13qA	GCAA/ATGG	ACAG/ATAA	GCAA/GTAA	ACAG/ATGG	ACAA/ATGG	ATGG/ACAG	GCAA/ATGG	ACAA/ATGG	ACAG/GTAA	GCAA/GCAC	GCAA/ATAA
mh13KK-218	CTC	CTC/CCC	CCC/TTT	TTT	TTT/TTT	TTT/TTT	CTT	CTC/TTT	TTT	TTT/CTT	CTT/CCC
13qC	TGGC	GCTG/TGGC	TGTC/GTTC	CGTC	GCTG/TGGT	TGTC/TGGC	TGGC	CCTC/TGGC	CCTC/CGTC	CCTC/TGGC	CCTC/CGTC
14qA	GCG/TCC	GTC	TCC/GTG	GCG	GCG/GTG	GCG/TCC	GCG	GCG/TCC	TCC/TGC	TCC/TGC	TGC/GCG
D16S2835	TCGTA	TCGTA	TCATG/GTGGC	TCGTA	TCGTC/TCATG	TCGTA/TCATG	TCGTA/TCGTC	TCGTA	TCGTA/TCATG	TCGTA	TCGTA
17qB	AGC/ACC	AGC/GGC	AGC/AGT	AGC/GGC	AGT/GGC	AGC/AGT	GGC	AGC	GGC/AGT	AGC	GGC/AGT
17qC	GCGT	GCAT/CCGC	GCAT/CCGC	GCGT/CCGC	CCGC/GCAT	GCAT	GCGT	GCAT	CCGC	GCAT/GCGT	GCAT
18pA	CAATG/TCGTG	CAATG/TCGTG	TCATG/CAATG	TCGTG/CAATG	TCGTG	CAATG/CCAGG	TCGTG	TCATG/CAATG	CAATG/TCATG	TCGTG	TCGTG
18qB	GGC/TAC	GAC/GGC	TAC/GGC	TAC/GGC	TAC	GAC/GGC	GGC	TAC/GGC	GGC/GAT	GGC	TAC/GAT
18qC	GAAG/TGAG	TAGG	GAAG/TGAT	TGAT/GAAG	TGAG/GAAG	GAAG	TAGG	TGAG/TAGG	GAAG	TAGG/TGAT	TAGG/TGAT
19pB	AGTCT/AGTCG	GGTTG	GGTTG	GGTTG/AACTG	GGTTG/AACTG	AACCTG	AGTCG/AACTG	AACTG/AGTCT	AACTG/AGTCG	AACTG	AGTCG/AACTG
LIN00906	ATCGG	ATCAG/ATCAA	ATCA /ATCGG	ATCA /ATCGG	ATTGG/ATCGG	ATCGG/ATCA	ATCGG	ATCGG/ATCA	ATCGG	ATCA /ATCGG	ATCGG/ATCA
19qA	CCACC	CTACC/CCACG	CCACC/TCGCC	TCGCC	CTACC/CCACC	CCACC/CCACC	CCACC/TCGCC	CCATC/CCACC	CCACC	TCGCC/CTACC	CCACC
19qB	AGCG/AGCA	AGGG/AGCG	AGGG/AGCG	AGCA/AGCG	AGGA/AGGG	AGCG/GAGA	AGGG/GGGA	AGCA/AGGG	AGCG/AGCA	AGGG	AGCG/AGGG
20qB	CCCAT/CCAAC	TTGCG/CCAAC	CCAAC/TTGCG	TTGCG/CCCGT	TTGCG/CCCGT	CCCGT/CCAAC	TTGCG/CCAAC	CCCAT/CCAAC	CCAAC	CCCAT/CCCGT	CCCAT/TTGCG
21qA	GTGG/GAGA	GAAG	GTGG/CAGG	GAGA	GAAG/GTGG	GAGA/CAGG	GTGG/GAGA	GTGG/GAGA	CAGG/GTGG	GTGG	GTGG/CAGG
mh21KK-324	TAA/CTG	CTG/TAA	CAG/CTG	CTA/CAG	CTA/CAG	CTG/CAG	CAG/TAA	CTG/TAA	CTG/TAA	TAA/CTG	CTG/TAA
22qA	ATGG/TCGG	TCGG/ACTA	TCGG	ACTA/ATGG	TCGG/ACTG	ACTA/ATGG	ATGG/ACGG	TCGG	ACTA/ATGG	ACTA/ATGG	ACTG/ATGG
22qB-mh22KK-070	AATATG/AGTGC	AGTGC/AGTGA	AATATG	GGCGTA	AGCATG/GCGTAG	AATATG	AGCATG/GCGTAG	GGCGTA/AGCATG	AGCATG/AATATG	AGCATG/AGTGC	GGCATG/AATATG

Tab.25 A Risultati genotipizzazione campioni di DNA per studio popolazioneistico

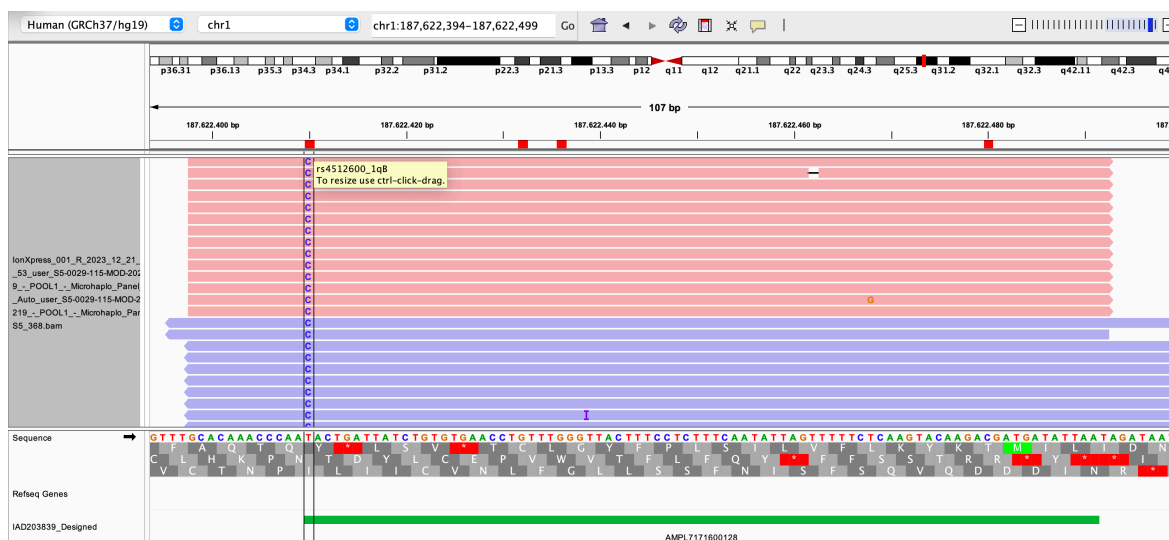




CAMPIONI	ANCS021	ANAS061	ANCS017	ANAS054	ANAS018	ANCS019	ANCS023	ANAS015	ANAS014	ANAS060	ANCS034
LMX1A											
FNDC10	GCGG/GCCG	GCCG	ATGC/ACGC	ATGC/GCGG	GCCG/ACGC	GCGG/ACGC	GCCG/GCGG	ACGC/GCGG	GCGG/ACGC	ACGC/ATGC	GCCG/ACGC
mh01KK-160	CGAA/TGAA	TTAA	CGAA/TTAA	CGAA	CGAA	CGAA	CGAA/TGAA	CGAA/TGAA	CGAA	TTAA	CGAA/TGAA
1qB	TCGT	TCGC/CGAC	TGTT/CGGC	CGGC/CGAC	TCGC/CGAC	CGGC/CGAC	CGGC/CGAC	CGGC/CGAC	CGGC/CAGC	TCGT/CGGC	CGGC/CGAC
mh02KK-140	GCGG	GCGG/GAGG	GCGG/GGAG	GCGG/GGAG	GGAG	GCGG/GAGG	GGAG/GGAG	GCGG/GGAG	GCGG	GGAG/GAGG	GCGG/GGAG
mh02KK-141	GCTA	GCGG	GCGG/GCTA	GCGG/GGTTG	GCGG/GCTA	GCTA/GGTTG	GCTA/GCTG	GCGG	GCGG/GCTA	GCGG	GCTA/GGTTG
D25259	ACGC/ACGA	ACCG/ACGC	<20x	ACAG	ATGG/ACAG	<20x	ACAG/ATGG	ACCG/GCGG	GCGG/ATGG	<20x	ACCG/GCGG
2pC	GCGG	ACGT/GCGG	GCGT/GCGG	GCGG	GCGT/ACGT	GTGT/GCGG	ACGT/GCGG	GCGG	GCTT/GTGT	GCGT/GCGG	GCTT/ACGT
2qA	TTAC/CTGC	CTGC	CCAC/CTGC	CCAG	CTGC/TTAC	CCAC/CCAG	CCAC/CCAG	CCAC	CTGC/CCAG	TTAC/CCAG	TTAC
mh02KK-134	ACT	ACC/ATT	TCT/ATC	ACC/ATT	ATC/ACC	ACC/ATT	ACC/ATT	ACC	ACT/TCT	ACC	ACT/ATC
2qD	GCG/GGG	GCG/GCA	AGG/GGG	GCA/GGG	GCG/GGG	GCG/AGG	GCG	GGG/GCG	GCA/GCG	GCA/GGG	GCG
mh02KK-136	TCC/GTC	TCC/GTC	GTC/GTA	GTC	TCC/GTC	TCA/TTT	TCA/GTC	GTC	TCC/GTA	TCC/TTC	TCC/GTA
D352426	ATC/ATA	GTC/ATC	ATC	GTC/ATC	ATA	ATC	GTC/ATC	ATC	GTC/ATA	GTC/ATA	GTC
3pA	CTAC/CCGT	CTAC	CTGC/CTAT	CTAC/CTAT	CTAC/CTAT	CTAC/CTAT	CTGC/CTAT	CTAC/CTGC	CTAT/CTAC	CTAT/CTAC	CTAT/CTAC
mh03KK-217	AGCT/CGCC	AGCT/CGCC	AGCT/AGCC	AGCT/AGCC	CGCT/AGCT	AGCC/CGCC	CGCC	AGCT/AGCC	AGCT	AGCT	CGCC/AGCC
ZNF385D	CGTG/CACG	AATG	CGTG	CATG/AATG	AATG/CATG	AATG	AATG	AATG	CGTG/CATG	AATG	CACG/CATG
3pB	CGCGA/AACCG	AACCG/CGCGA	AACCG/AACCA	CGCCA	AACCG/AACCA	CGCGA/CGCCA	AACCG	AACCG/AACCG	AACCG	AACCG	AACCG/CGCGA
LINC02000	GGTG/CCCC	CGTC	GGTG/CGTC	CGTC/CCCC	CGGT/CGTC	CGTC/CCTT	CGTC/GGTC	CGTC/CCTT	GGTC	GGTC/CCTT	CGTC
LOC100505920	CTGA/GCGA	GCGT/GCGA	GCGT/GCGA	GGT/GCGT	GCGT/GCGA	CTGA/GCGT	CTGA/GCGT	GCGT/GCGA	GCGT/GCGA	GCGT	GCGT/GCGA
mh04KK-010	AA/GA	GA/AA	AA/GA	AA	GA/AA	GG/GA	AG/GA	AA/GA	GA/AG	GG/AA	AA/GA
SORCS2	CCT/TAG	CCT/CCG	CCT/CCG	CCT/CCG	CCG	CCT	CCT/CCG	CCG	CCT	CCT	CCT
4pA	GGCA	GTA/GTAA	GTAA/GGCA	GGCG/GGCA	CCG/GGCA	GGCA/CTCA	GTA/GTAA	GGCA/CTCA	GTA/GGAA	GGCA	GTA/GGCA
4pB	GCAC/ACGC	GCAC/ATGC	ACGC/ATGC	ACGC	GCAG	GCAC/ACGC	GCAC	GCAG	ACGC	GCAC/ACGC	GCAC/ACGC
4qA	CCT	CTT	CCT	CCG	CTT/CCT	CTT/CCT	CCT/CTT	CCG	CCT	CCT	CCG/CTT
4qC	TGAG/TAGG	TAGG/GGGG	TGAG/TAGG	TGAG/GGGG	TGAG	TAGG	TAGG/TGAG	GGGG/TAGG	TAGG/TGAG	GGGG/TAGG	TAGG/TGAG
mh04KK-217	ACC/ACT	ATC	ACT	ACT	ATT/GTC	ATT/GTC	ATC/ACC	ACC/ACT	GTC/ATC	ATT/ACT	ACT/ATT
mh04KK-218	AACC	GGTG/GGCC	AACT/GGCC	AACT/AACC	GGTG/AACT	GGTG/AACT	GGTG/GGCC	GGTG	GGTG/GGCC	GGTG/GGCC	AACT/GGCC
D4S2390	GGAGC/GGGCT	AGAGC/GGAGC	GGAGC/AGAGC	AGAGT/AGGCT	AGAGT/AGGCT	AGAGT/AGGCT	AGAGC	AGAGC/GGAGC	AGGCT	AAAGT/GGAGC	AGAGT/AGGCT
D4S3214	AAAG	GGAG	AGGG	AAAG/AAAA	AAAG/GGAG	AAAG/AAAA	GGAG/AGGG	GGAG/AAAA	AGGG/AAAG	AGGG/GGAG	AGGG/GGAG
mh05KK-125	GGTG/GATG	GGTG/GATG	AATG/GATC	GATC/AATG	AATG	GGTG	GATG/GATG	AATG/GGTTG	GGTG/GATG	GATC/AAATG	AATG/GGTTG
mh05KK-022	CC	CA	TC	TC	CA/CC	CC	CA/TC	CA/TC	TC	CA	TC/CA
5qB	CCAT/CCAC	CCAT/TTAC	CCAT/CCAC	TTGC/CCAT	TTAC/CCAT	CAAC/TTAC	TTGC/TTAC	TTGC	TTAC/CAAC	CCAT	TTAC
5qC	CCGT/TCAC	TCGC/CCGT	CCGT/TCAC	CCGC/CCGT	CCGT	CCGC/TCAC	TCGC/CCGC	CTGC/CCGT	CCGT/TCAC	TCAC/CCGT	CTGT/TCAC
5qD	GGG	GGG/GAG	GAA/GAG	GGG	GGG	GAA/AGG	GAA	GAA/GGG	GAG/AGG	GAA/GGG	GAA
6pA	GCT/CCT	CCT/GCT	CCT/GCT	CCT	GCT/CCA	CCT/CCA	CCT/CCA	GCT/CCT	CCA/CCT	CCT	CCT/GCT
mh06KK-103	CTCA/CTGA	CCCA/CTTA	CTTG	CCCA/CTTG	CCCA/CTTG	CCTA	CCCA/CTTA	CCCA/CTTG	CCCA/CTTA	CCCA	CCCA/CTTG
D6S2662-B	AGAA/AGGA	TGGT	AGGA/AGGT	TGGT	AGAA/TGGT	TGGT	AGAA/TGGT	TGGT	TGGT/AGGT	TGGT	AGAA/TGGT
LOC105375075	TTCT	TTCC	TTCT/TTCC	TTCC/GTTT	TTT/GTTT	TTCT/TTCC	TCTC/GTTT	TTT/TTCC	GTTT	TTCT/TTCC	TCTC/GTTT
6pB	GCAA/GTAG	<20x	<20x	GCAA	GTAG/GCAA	<20x	<20x	GTAG/GCAA	<20x	<20x	GTAG
6qA	GCC/GCT	GCC/GCT	GCC/GCT	GCC/GCT	GCT/GCT	GCC/GCT	CCT/GCT	GCC/GCT	GCC/GCT	CCT/GCC	GCT/GCT
6qB	ACAG	ACAG	ACAG/GTGG	ACAG/GTGG	GGGG/GTGG	GCAG/GCAC	GTGG	GTGG	GTGG	ACAG/GCAC	GCAG/GTGG
6qC	CTC	ATT/ATC	CTC/ATC	ATT/ATC	ACC/CTC	CTC/ATC	ATC	ATC	ATT/ATC	CTC	CTC/ATC
7pA	AAGA/TATC	AGTC/TATC	AAGA/AGTC	AATA/AGTC	AAGA/AGTC	AGTC/TATC	AATA/AGTC	AAGA/TATC	AGTC/TATC	AGTC/TATC	TATC
7pB	AAAC/AGGC	TAGC/TAGT	AGGC	AGGC/AAAC	AGGC/TAGC	AAAC/AGGC	TAGT/AGGC	AGGC/AAAC	AAGT/TAGT	AAGT/AGGC	AAGC/TAGT
7pC	CTGT	CTGT/CCTAC	CTGT/CCTAC	CTGT/CCTAC	TCGC/CCTAC	CTGT/CCTAC	CTGT	CTGT	CTGT	CTGT	CCTAC/TCGC
7qA	CAT	CAT/CAC	CGC/CAT	CAT	CAT/CGC	CAC/GAT	CAT/CAC	CAT/CAC	CAT/GAT	GAT/CGC	GAT/CAT
7qC	TACGC/AGCAC	TACGC/AGTGC	AGTGC/AACGC	TACGC/AGTGC	CAACG	AACGC/AGCAC	TACGC/AGCAC	TACGC/AGCAC	TACGC/AGTGC	AGTGC	TACGC/AGTGC
8pA	CGGTA	CGGTA	CAGTG/CGGTA	CAGTA	CAGTA/CGGTA	CAGTG/CGGTA	CGGTA/CAGTG	CAGTA/CGGTA	CGGTA/CGGTA	CGGTA/CGGTA	CGGTA/CAGTA
GATA-B-GATA4-A	CTCAA/CTAG	CTCAA	CTAG/CTCAA	CCCA/CTAG	CTAG	CTAG/TTTAA	TTTAA	CTAG/CTCAA	CTAG/TTTAA	CTAG/CTCAA	TTTAA
8pB	CATCA/CTCTG	CATCA	CTCTG/CATCA	CATCA	CATCA/CATCA	TTCTA/CTCTG	CATCA	CATCA	CATCA/CATCA	CTCTG	CATCA
9pA	GATCAA	GATTG/CAGTAC	GATCAA/AATGCG	GATCAA/GTGTCC	AGTCAA/AATGCG	GATCAA/AATGCG	AATGCG/AGTCAA	GATGCG/GTGTCC	GATGCG/GATTGCG	GATCAA/GTGTCC	AGTCAA
9pB	TCGC/TCGT	TCGT	TCGT/GTAC	GTAC/TCGC	TCGC/TCGT	TCGC/TCGT	GTAC/TCGT	TCGC/GTAC	GTAC/TCGT	TCGT/TCGC	TCGT
9qA	GGC/TCC	GGG/GGG	GGG/TCC	GGG/GGC	GGG	GGG	TCC/GGC	GGG/GGC	GGG/GGC	GGG/GGC	GGC/TCC
9qB	CGTT	CATT	TGCT/CGTT	TGCC/CATT	CGTT/CATT	TGCT/CATT	TGCT/CATT	CATT	TGCT/CATT	CGTT	TGCT/CATT
9qC	GTGCA/ATCGC	ATGCG/ACCGC	GCTCA/ATCGC	ACCGC/ATCGC	ATGCG/ACCGC	ATGCG/ACCGC	ATGCG/ATCGA	ATGCG/GCCGC	ATCGC	ATCGC	ATGCG/ACCGC
10pA	CTGG	TAGG	CTGG/CTAG	CTGG/CTAG	CTGG	CTAG/CTGA	CTGG	CTAG/CTGA	CTGG	CTAG/TAGG	CTAG/CTGG
11pB	TTGCG/TCGCG	CCGTG/TTGCG	TTGCG	TTGCG/TCGCG	TCGCG	TCGCG/CCATG	TCGCG/CCATG	CCATG/TTGCG	TTGCG	TCGCG/CCATG	TCGCA/TCGCG
12qA	GGGG/ATAA	GGGG/ATAA	<20x	GTGG	GTGG	GTGG/ATGA	GTGG/ATGA	GGGG/ATAA	GGGG/ATAA	<20x	GGGG/ATGA
12qB	AAAT	GCAC/GAAC	GAAC/GAAT	AAAT	GAAC/AAAT	GCAC/GAAC	GCAC/GAAT	GAAC	GAAC/AAAT	GCAC/GAAC	GCAC/GAAC
13qA	GCAA	ACAA/ATAA	ACAG/ATGG	ATGG/GCAA	GCAA/ATAA	GCAA/ATAA	ATGG/ACAG	ATGG/ACAG	ATGG/ATAA	ACAG/ATAA	ACAG/ATGG
mh13KK-218	TTT/CTT	TCC/TTT	CCC/TTT	TTT/TTT	TTT	TTT/CTT	TTT/TTT	CTT/CCC	TTT	TTT	CTT/TTT
13qC	CCTC/TGGC	TGGC/CTGC	GCTC/TGGC	CCTC/TGGC	TGGC	CCTC/TGGC	CCTC/TGGC	TGTG/TGGC	TGGC	TGGC/CTGC	TGGC/CTGC
14qA	GCG/TCC	GTG/GCG	GTG/TCC	TCC	GCG/TCC	TCG/TCC	TCC/GCG	TCC/GCG	GCG/TCC	TCG/GCG	TCG/GCG
D16S2835	TCATG	TCGTA	TCGTA	TCGTG	TCATG	TCGTA/GTGGC	TCGTA/GCGCG	TCGTA/TCATG	TCGTA	TCGTA	TCATG/GTGGC
17qB	AGT/GGC	GGC	AGT	ACC/GGC	AGC/GGT	AGC/GGC	GGC	GGC	AGT/GGC	GGC	AGC/GGT
17qC	CCGC	CCGC/GCAT	GCAT/CCGC	GCAT/CCGC	GCAT/GGAG	GCAT/GGAG	GCAT/GGAG	GCAT/GGAG	CCGC/GCAT	CCGC	CCGC/GCAT
18pA	TCGTTG	TCATA/TCGTTG	CAATG	TCGTTG/CAATG	CAATG/TCGTTG	TCATA/CAATG	TCATA	TCGTTG	CAATG/TCGTTG	TCGTTG	CAATG/TCGTTG
18qB	GGC/GAC	GGC/GAT	GGC/TAC	GGC/TAC	TAC/GGC	TAC	GCC	TAC/GGC	TAC/GGC	GGC	GAC/GGC
18qC	TGAT/TAGG	TGAG/TAGG	TAGG	GAAG/TAGG	TGAT/TAGG	TAGG/GAAG	GAAG/TAGG	TAGG	TGAT	TGAT/TAGG	TGAG/TAGG
19pB	AGTGC/GGTTG	AACTG	AACTG/GGTTG	GGTTG	AGTCT/GGTTG	AGTGC/GGTTG	AACCTG	AGTGC/AACTG	GGTTG/AACTG	AGTGC/AACTG	GGTTG
LINC00906	ATCAG	ATCGG/ATCA	ATCGG	ATCA_/ATCGG	ATCGG/CTCGA	ATCAG/ATCAA	ATCGG	ATCA_/ATCGG	ATCGG	ATCGG/CTCGA	ATCGG
19qA	TGCC/CCACG	CCACC	CCACC/CCACG	CCACC/CCACG	TCGCC	CCACC/CTACC	CCACC	CTACC/CCATC	CCACC/TCGCC	CCACC	TCGCC/CCACC
19qB	AGCA/GAGA	AGGA/AGCA	AGGA	AGCA/AGCG	AGCA	GAGA/AGGG	AGCG/GAGA	AGCA/GAGA	AGGA/AGGA	AGCG/AGGG	AGCG/AGGG
20qB	CCCAT/TTCCG	TTCCG/CCCAT	CCCAT	CCCAT/CCAAC	CCCGT/CCAAC	CCAAC/TTCCG	CCCAT/CCAAC	CCAAC/TTCCG	CCCGT/TTCCG	CCCAT/CCAAC	TTCCG
21qA	CAGG/GAGA	GTTGG/CAGG	GAGA	GAAG/GAGA	GAAG	GTTGG/GAGA	GTGG/GAGA	GTTGG/GAGA	GTTGG/GAAG	CAGG	GAAG
mh21KK-324	CTG/CAG	CAG/CTG	CTG/TAA	CAG	CTG/CAG	TAA/CTA	TAA/CTA	TAA/CAG	CAG	CTG/CAG	CTG
22qA	ATGG/ACTA	ACTA	ATGG/ACGG	ACTG/TCGG	ACTG/ACTA	ATGG/ACTG	ACTG/TCGG	ACTG/ATGG	ATGG/ACTG	ATGG/ACTG	ATGG
22qB-mh22KK-070	AATATG/AGTGC	GGCATG	GGCATG/AGTGT	GGCGTA/AATATG	AATATG	AATATG/AGCATG	AGTGT/AGCATG	AATATG/GGCGTA	AATATG/AGCATG	AATATG	GGCGTA/GGCATG

Tab.25 C Risultati genotipizzazione campioni di DNA per studio popolazioneistico

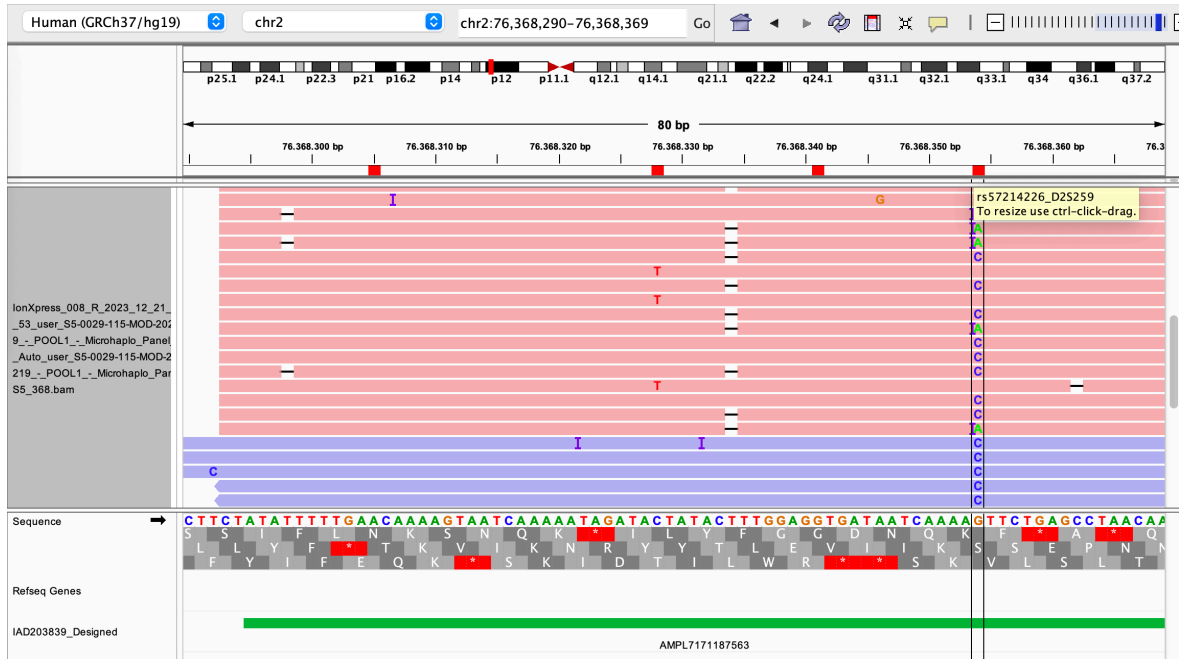
Da una prima analisi è emerso che i microaplotipi, 1qB, 4pB, 7pA, 12qB e 17qb non sono stati sequenziati in nessuno dei campioni testati. Un'ispezione più approfondita di tali microaplotipi su IGV ha permesso di evidenziare che il primo SNPs del microaplotipo corrisponde alla prima posizione dell'amplicone. Infatti, come è possibile vedere nella Fig.25, è stato necessario controllare tali campioni, per quei determinati microaplotipi su IGV.



*Fig.25 Rappresentazione di un microaplotipo rilevato tramite IGV. Preso come esempio il microaplotipo 1qB per il campione ANAS015, con Barcode 01,*

I campioni ANAS062, ANAS015, ANCS002, ANCS033 e ANAS093 dal risultato del plug-in, presentavano più di due chiamate alleliche per il microaplotipo D2S259. Dalla visualizzazione su IGV abbiamo notato che la terza chiamata allelica era un artefatto. Lo SNP, rs57214226 del microaplotipo D2S259 sul cromosoma 2, presentava un certo grado di incorporazioni

nucleotidiche non alleliche sui filamenti forward. Nel caso di questo SNP, le letture del nucleotide “A” non alleliche, potrebbero essere dovute alla presenza di tratti omopolimerici (AAAA [G/C] TT). (Fig.26)



*Fig.26 Rappresentazione di un microplotto rilevato tramite IGV. Preso come esempio il microplotto DS259 per il campione ANAS062, con Barcode 08,*

Anche i campioni ANAS071, ANAS046, ANCS027, ANCS021, ANCS019 e ANCS023, presentavano tre chiamate alleliche per il microplotto LOC100505920 e anche in questo caso dalla visualizzazione su IGV è risultato essere un artefatto, anche in questo caso dovuta alla presenza di tratti omopolimerici (ACACAC).

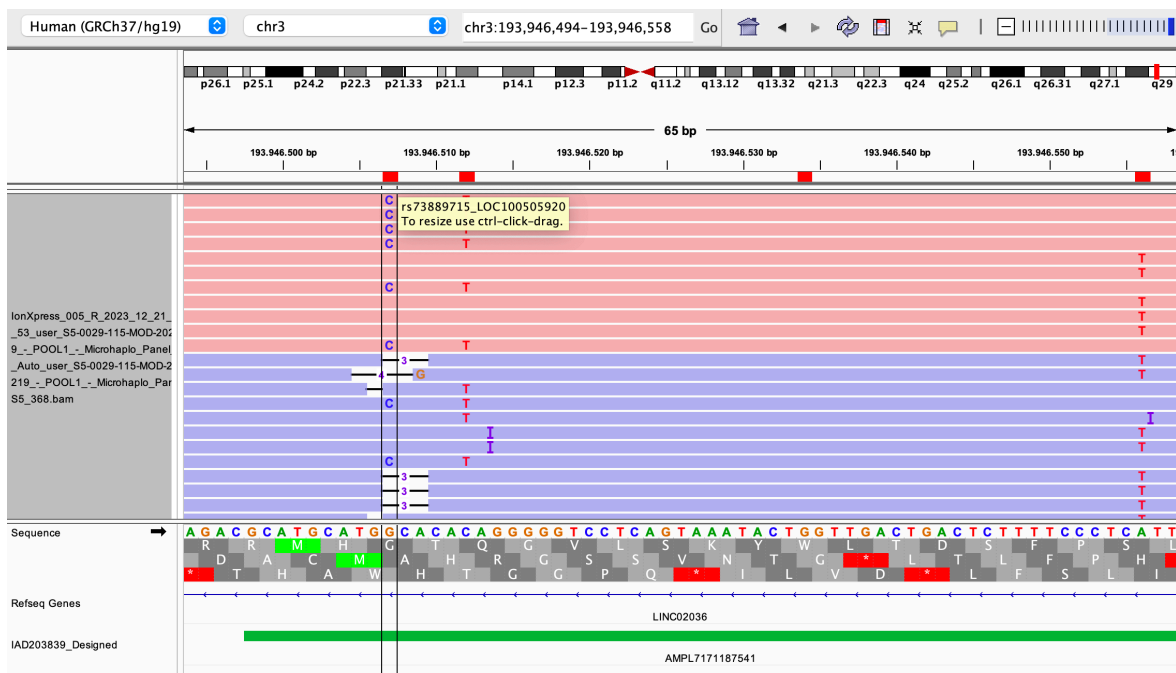


Fig.27 Rappresentazione di un microplotipo rilevato tramite IGV.  
 Preso come esempio il microplotipo LOC100505920 per il campione ANAS019, con Barcode  
 53

Nella tabella seguente vengono riportati gli artefatti eliminati, relativi ai campioni.

CAMPIONI	D2S259	LOC100505920
ANAS062	ACGA	
ANAS015	ACGA	
ANCS002	ACGA	
ANCS033	ACGA	
ANAS093	ACGA	
ANAS071		GTAT
ANAS046		GTGA
ANCS027		GTGA
ANCS021		GTGA
ANCS019		GTGA
ANCS023		GTGA

Tab. 26 Artefatti eliminati



Anche per i campioni del pool 2, ovvero i campioni forensi reali, abbiamo notato che i microaplotipi 1qB, 4pB, 7pA, 12qB e 17qb non venivano letti, per cui anche in questo caso i singoli campioni sono stati ispezionati su IGV e controllati manualmente.

CAMPIONI	CAMP_01	CAMP_02	CAMP_03	CAMP_04	CAMP_05	CAMP_06	CAMP_07	CAMP_08	CAMP_09	CAMP_10
LMX1A										
FND10	ATGC/GCGG	ATGC/GCGG	ATGC/GCGG	GCGG/ATGC	ATGC/GCGG	GCGG	GCGG/GCGG	ACGC/GCCG	ACGC/GCCG	GCGG/ACGC
mh01KK-160	CGAA/TTAA	CGAA/TTAA	TTAA/CGAA	TTAA/CGAA	TTAA/CGAA	CGAA/CGAA	CGAA/CGAA	CGAA/TTAA	CGAA/TTAA	CGAA/TTAA
1qB	TGGC/TCGT	TGGC/TCGT	TGGC/TCGT/CGGC	TGGC/TCGT/CGGC	<20x	TCGT/CGAC/CGGC	TCGT/CGGC/TCGC	CGGC/CGAC	TCGC/CGGC	TGGC/CGGC
mh02KK-140	CGGG/GAGG	CGGG/GAGG	CGGG/GAGG	GAGG/CGGG	<20x	CGGG/GGAG	CGGG/GGAG	CGGG/GAGG	GGAG/CGGA	GAGG
mh02KK-141	ACTG/GGAG	GGGG/ACTG	ACTG/GGGG	<20x	<20x	GGTG	GGTG/GCTA	GGGG/GGTG	GGTG/GCTA	GGGG/ACTA
D2S259	ATGG/ACAG	ATGG/ACAG	ACAG/ATGG	ATGG/ACAG	ACAG/ATGG	ACGC/ACGA	ACGC/ACGA	ATGG/ACAG	ACGG	ACGG/ACAG
2pC	GCTT	GCTT	GCTT	GCTT/GCGT/ACGT	<20x	GTGT/GCGG	GCGG/GTGT	GCGT/ACGT	GCGT/GCGG	GCGT/GCTT
2qA	TTAC/CCAC	TTAC/CCAC	TTAC/CCAC	<20x	CCAC/TTAC	CCAC/CTGC	CCAC/CTGC	TTAC/CCAG	CCAG	CCAC/CCAG
mh02KK-134	TCT/TCC	TCT/TCC	TCC/TCT	<20x	TCT/TCC	ACC/ACT	ACT/ACC	ACC/ATC	ATT/ACC	ATC/ATT
2qD	GGG	GGG	GGG	<20x	GGG	GGG/AGG/GCA	AGG/GGG	GGG/AGG	GGG/GCG	AGG
mh02KK-136	TCC/GTC	TCC/GTC	GTC/TCC	TCC/GTC	<20x	GTA/GTC	GTA/GTC	GTC	GTA/TCA	GTA
D3S2426	ATA/ACA	ACA/ATA	ATA/ACA	ACA/ATA	ACA/ATA	ATC	ATC	ATC/ATA	ATC/ATA	GTC/ATA
3pA	CTAT/CTGC	CTGC/CTAT	CTAT/CTGC	CTAT/CTGC	<20x	ACGT/CTGC	CTGC/ACGT	CTAC	CTAC/CCGT	CTGC/CTAC
mh03KK-217	AGTT/GCGC	AGTT/GCGC	GCGC/AGTT	AGTT/GCGC	AGTT/GCGC	CGCC/AGCC	CGCC/AGCC	CGCT	CGCT/ACCC	GCGT/CTCT
ZNF385D	AATG/CATA	AATG/CATA	CATA/AATG	AATG/CATA	<20x	CATG/AATG/CGTG	AATG/CATG	AATG/CACG	ATGG/ACGC	CATG/CGTG
3pB	AAACCG	AAACCG	AAACCG	AAACCG	<20x	AAACCG	AAACCG	CGCGA	AAACCG	AAACCG
LINC02000	CGTC	CGTC	CGTC	CGTC/GGTC	CGTC/GGTC	CGTC/CTCT	CGTC/CTCT	CGTC	CCTC/GGTC	CGTC/CTCT
LOC100505920	GCGT	GCGT	GCGT	GCGT	GCGT	GCGT/GCGA	GCGT/GCGA	GCGA/GCAT	GCGT	CTGA/GTGA
mh04KK-010	GA	GA	GA	GA	GA	AG/AA/GA	GA/AG	GA/AG	GA/AG	AA/GG
SORCS2	CCG/CCT	CCT/CCG	CCG/CCT	CCT/CCG	CCT/CCG	CTT	CCT	CCT/CCG	CCG	CCG/CCT
4pA	GGCG/GGCA	GGCG/GGCA	GGCG/GGCA	GGCG/GGCA	GGCA/GGGC	GTA	GTA/GGGC	GGCA	GTA	GTA/GGGCA
4pB	GCAC	GCAC	GCAC	<20x	GCAC	GCAC/ATGC	GCAC/ATGC/ACGC	GCAC/ACGC	GCAC/ACGC	GCAC/ATGC
4qA	CCG/CTT	CTT/CCG	CCG/CTT	CTT/CCG	CTT/CCG	GCG/CTT	GCG/CTT	GCG/CTT	CCG/GCG	CTT
4qC	GGGG/TGGC	GGGG/TGGC	GGGG/TGGC	GGGG/TGGC/TAGG	GGGG/TGGC	GGGG/TAGG	GGGG/TAGG	TAGG	TAGG	GGGG/TAGG
mh04KK-217	ATC	ATC	ATC	ATC	ATC	GTC/ATC	GTC/ATC	ACC	ACC/ACT	ATC
mh04KK-218	GGCC/GACC	GGCC/GACC	GACC/GGCC	GGCC/GACC	<20x	GGTC	GGTC/GGCC	GGCC/GGTC	GGCC/AACT	GGTC/GGCC
D4S2390	GGAGC/AAAGC	GGAGC/AAAGC	AAAGC/GGAGC	GGAGC/AAAGC	<20x	AGAGT/GGAGC	AGAGT/GGAGC	AGAGC	AGAGT/AGAGC	AGGCT
D4S3214	GGAG	GGAG	GGAG	GGAG	GGAG	GGAG/AGGG	GGAG/AGGG	GGAG/AGGG	AGGG	AAAG
mh05KK-125	GGTG/AATG	AATG/GGTG	AATG/GGTG	GGTG/AATG	AATG	AATG/GGTG/GATG	AATG/GGTG	GATG/AATG	GATG	AATG
mh05KK-022	CA/CC	CA/CC	CC/CA	CC/CA	CC/CA	CC/TC	CC/TC	TC/CC	CA/CC	TC/CC
5qB	CCAT/TTGC	CCAT/TTGC	TTGC/CCAT	CCAT/TTGC	TTGC/CCAT	CCAT/TTAC	CCAT/TTAC	TTAC	TTAC/CCAC	TTGC/CCAC
5qC	CCAC/CTGT	CTGT/CCAC	CTGT/CCAC	CTGT/CCAC	<20x	TCAC/CCGC	TCAC/CCGC	TCAC/CCGC	CCGT/CTGT	CCGT/TCAC
5qD	AAG/GGG	GGG/AAG	AAG/GGG	AAG/GGG	GGG/AAG	GGG/GAA	GGG/GAA	GGG	GGG	GGG
6pA	CCT/CTT	CTT/CCT	CCT/CTT	CCT/CTT	CCT/CTT	GCT	GCT/CTT	CCT	CCT/GCT	CTT/GCT
mh06KK-103	CCCA/CTGT	CCCA/CTGT	CCCA/CTGT	CCCA/CTGT	<20x	CCCA/CTA/CTGT	CCTA/CCCA	CCTA/CTGT	CCTA	CCCA
D6S2662-B	AGAA/AGGA	AGAA/AGGA	AGAA/AGGA	AGGA/AGAA	AGAA/AGGA	AGGA/TGGT	AGGA/TGGT	TGGT/AGAA	TGGT	TGGT
LOC105375075	TCTC/GTTC	GTTT/CTCT	TCTC/GTTC	GTTT/CTCT	<20x	TTTT/TTCC	TTTT/TTCC/TTCT	TTCC/TTCT	TTCT	TCTC/GTTC
6pB	<20x	GTAG	<20x	<20x	<20x	<20x	<20x	<20x	<20x	<20x
6qA	GCC/GTT	GTT/GCC	GCC/GTT	GCC/GTT	GCC/GTT	GTT/GCC	GTT/GCC	GCT/CCT	GCC	GCC/GCT
6qB	GCAC/GTGG	GACG/GTGG	GCAC/GTGG	GCAC	GCAC/GTGG	GCAC/GTGG	GCAC/GTGG	GATC/AATG	GTCG/GCGG	GCAC/GTGG
6qC	CTC	CTC	CTC	CTC	CTC	ATC/CTC	ATC/CTC	ATC	CTC	ATT
7pA	TATC	AAGA	TATC	TATC	TATC	AAGA/AATA	AAGA/AATA/TATC	AATA	AGTC/AAGA	AGTC
7pB	TAAC/TAGC	TAGC/TAAC	TAAC/TAGC	TAAC/TAGC	<20x	AAGT/AAGC	AAGT/AAGC/TAGC	AAAC	TAGT/TAAC	TAAC/AAGT
7pC	CTGT	CTGT	CTGT	CTGT	<20x	<20x	CTGT/TCGC	CCCG/TCGC	CTGT	CTGT/CTAC
7qA	CAT	CAT	CAT	CAT	CAT	CAC/CAT	CAC/CAT	CAT/CAC	CAC/GCG	CAC/CAT
7qC	TACGC/AGTGC	TACGC/AGTGC	TACGC/AGTGC	TACGC/AGTGC	AGTGC/TACGC	TACGC/AGTGC	TACGC/AGTGC	AGTGC	TACGC/AGTGC	AAACG
8pA	CAGTG/AGGGG	CAGTG/AGGGG	CAGTG/AGGGG	<20x	<20x	CAGTG/CAGTA	CAGTG/CAGTA	CAGTG/CGGTA	CAGTG/CGGTA	CGGTA/CAGTA
GATA-B-GATA4-A	CTAG	CTAG	CTAG	CTAG	CTAG	CTAG	CTAG	CTAA	CTAA/CTAG	CTAA
8pB	CTCTG	CTCTG	CTCTG	CTCTG	CTCTG	TTCTA	TTCTA	TTCTA/CATCA	TTCTA/CTCTA	CTCTG/CATCA
9pA	GGTCG/AGTCAC	AGTCAC/GGTCG	GGTCG/AGTCAC	AGTCAC/GGTCG	AGTCAC/GGTCG	GGTCG/AGTCAC	GGTCG/AGTCAC	GGTGCG	AATGTC/AGTCAC	GATTCG/AGTCAC
9pB	TCGT/TCG	TCGT/TCG	TCGT/TCGT	TCGT/TCGT	<20x	TCGT/GTAC	TCGT/GTAC	TGCT	GTAC	TCGT/GTAC
9qA	GGC	GGC	GGC	GGC	GGC	TCC/GGC	GGC/TCC	GGG/GGC	GGG/GGC	GGG/GGC
9qB	CATT	CATT	CATT	CATT	<20x	CGTT/TGCT	TGCT/CGTT	CATT	TGCT/CGTT	CATT/TGCT
9qC	ACCGC/ATCGA	ACCGC/ATCGA	ATCGA/ACCGC	ACCGC/ATCGA	<20x	ATCGC	ATCGC	ATCGC	ATCGC/GCTGA	ACCGC/ATCGC
10pA	CTGA/CTGG	CTGA/CTGG	CTGG/CTGA	CTGA/CTGG	CTGA/CTGG	CTAG	CTAG	CTGG/CTAG	CTGG/CTAG	CTGG/TAGG
11pB	CCATG/CCGTG	CCGTG/CCATG	CCATG/CCGTG	CCATG/CCGTG	<20x	TGCG/TTGCG	TGCG	TTGCG/CCATG	TGCG/TTGCG	TGCG/CCATG
12qA	GTGG/ATAA	ATAA/GTGG	ATAA/GTGG	ATAA/GTGG	<20x	GGGG	GGGG	GTGG/ATAA	GGGG/GTGG	GGGG
12qB	GCAC/GACC	GCAC/GACC	GCAC/GACC	<20x	GCAC/GACC	GAAC/AAAT	GAAC/AAAT	GCAC/AAAT	GCAC/AAAT	GCAC/GAAC
13qA	GCAA/ATAA	GCAA/ATAA	GCAA/ATAA	GCAA/ATAA	GCAA/ATAA	ACAG/GCAA	ACAG/GCAA	GCAA/ACAA	ATGG	ACAG/ATGG
mh13KK-218	TTT/CTT	CTT/TTT	TTT/CTT	TTT/CTT	TTT/CTT	TTT/TTT	TTT/TTT	TTT/CTT	TTT/CTT	TTT/CTT
13qC	CGTC	CGTC	CGTC	<20x	<20x	CGTC	CGTC	CGTC	CGTC/TGGC	CGTC/TGGC
14qA	GCG/TCC	GCG/TCC	TCC/GCG	GCG/TCC/TCG	TCC/GCG	GCG	GCG	GCG/TCC	TCC/GCG	TCC/GCG
D1S2835	TCATG/TCGTG	TCATG/TCGTG	TCGTG/TCATG	TCATG/TCGTG	TCATG/TCGTG	TCGTA	TCGTA	GTGG/TCATG	TCATG/TCGTG	TCATG/TCGTA
17qB	AGC/AGT	AGC/AGT	AGC/AGT	<20x	AGC/AGT	AGT/GGC	AGC/GGC/AGT	GGC	ACC/GGC	ACC/GGC
17qC	CCGC	CCGC	CCGC	<20x	<20x	GGAC/CCGC	CCGC/GGAC/GCGT	GGAC/GCAT	GCAT/CCGC	GCAT/GCGT
18pA	TCATG/TCGTG	TCGTG/TCATG	TCGTG/TCATG	TCGTG/TCATG	TCATG/TCGTG	TCGTG	TCGTG	TCGTG/CAATG	TCGTG/TCATG	CCAGG/CAATG
18qB	TAC	TAC	TAC	TAC	TAC	GCG/GAT/TAC	GCG/GAT	GAC/TAC	GAC/TAC	GCG/GAT
18qC	GAAG/TAGG	GAAG/TAGG	<20x	<20x	<20x	TAGG	TAGG/GAAG	TGAT/TAGG	TAGG	GAAG
19pB	AGCTG/AGTTG	AGTTG/AGCTG	AGTTG/AGCTG	AGTTG/AGCTG	<20x	GGTTG/AACTG	AACTG/GGTTG	AGTCT	AGTCT/GGTTG	GGTTG
LINC00906	ATCGG/ATCA	ATCGG/ATCA	ATCA/ATCGG	ATCA/ATCGG	ATCGG	CTCGA	ATCA/CTCGA	ATCGG	ATCGG/ATCA	ATCA/ATCGG
19qA	CCACG	CCACG	CCACG	<20x	<20x	TCGCG/CCACG	TCGCG/CCACG	TCGCG	CCATC/CCACG	CCCG/CCACG
19qB	AGGG/AGCG	AGGG/AGCG	AGGG/AGCG	AGCG/AGGG/AAGA	<20x	GAGA/AGCG	AGCG/GAGA	AGGG/AGCA	AGCG/AGGG	AGCG
20qB	CCAT/TTCCG	TTCCG/CCAT	TTCCG/CCAT	TTCCG/CCAT	<20x	CCAT/TTCCG	CCAT/TTCCG	CCAT/CCAAC	TTCCG/CCAT	CCAT/CCAAC
21qA	GAGG/GTGG	GAGG/GTGG	GAGG/GTGG	GAGG/GTGG/GAGG/CAGG	GAGG/GTGG	GAGG/GAGA	GAGG/GAGA	GAGA	GAGG/GAGA	GAGG/GTGG
mh21KK-324	TTA/CTG	TTA/CTG	CTG/TTA	TTA/CTG/TAA	TTA/CTG/TAA	CAG/CTG	CTG/CAG	CTG/TAA	CTG/CAG	CTA/CAG
22qA	TCGG	TCGG	TCGG	TCGG	TCGG	TCGG/ACTA	TCGG	ACTG/ATGG	ACTA/TCGG	TCGG/TCGG
22qB-mh22KK-070	AATATG/AGTGCG	AGTGCG/AATATG	AATATG/AGTGCG	AATATG/AGTGCG	AATATG/AGTGCG	AGCATG/AGTGCG	AGCATG/AGTGCG	AGCATG/AGTGCG	AATATG/GGCATG	AGCATG/AATATG

Tab.27 A Risultati genotipizzazione campioni reali forensi



## Capitolo 5

### DISCUSSIONI

Per molti anni, i marcatori del DNA per la pratica forense sono stati loci a ripetizione tandem breve (STR), ma nel caso di DNA degradato risultano essere poco efficienti, questo potrebbe essere dovuto alla lunghezza dell'amplicone e alla presenza di picchi stutter che rende difficile la risoluzione delle miscele. Ideale sarebbe quindi utilizzare un tipo di marcatore caratterizzato da piccole dimensioni dell'amplicone, come nel caso degli SNPs.

Gli SNP sono molto abbondanti nel genoma umano e sono caratterizzati da un tasso di mutazione molto inferiore rispetto agli STR. Recentemente, sono stati proposti nuovi marcatori costituiti da una combinazione di più SNP, denominati microaplotipi, la combinazione di più SNP può aumentare il livello di eterozigosità e la breve distanza tra gli SNP offre il vantaggio di mostrare un basso tasso di ricombinazione all'interno della regione MH. Inoltre, gli MH non solo possono fornire informazioni sull'identità e sull'ascendenza biogeografica, ma possono anche essere utilizzati per la deconvoluzione della miscela, poiché mancano della produzione di frammenti stutter, comunemente presenti negli STR.

Tali regioni sono meglio tipizzate mediante sequenziamento massivo parallelo (MPS). Le piattaforme in commercio consentono di sequenziare *reads* in



parallelo, invece di elaborare singoli ampliconi come avveniva per l'elettroforesi capillare. Negli ultimi anni sono stati effettuati diversi studi sui microplotipi tramite la tecnologia MPS, come quello di De La Puente et al.<sup>(31)</sup>, il quale ha proposto un pannello di 118 microplotipi costituito da 107 microplotipi autosomici e 11 microplotipi del cromosoma X, analizzati in multiplex per l'analisi del DNA forense con sequenziamento massivo parallelo (MPS). Interessante è stato anche il pannello da 74 microplotipi proposto da Oldoni et. al, analizzati su Ion Torrent S5™ (Thermo Fisher Scientific) e con applicazione per la deconvoluzione delle miscele.

Questo studio si è concentrato sulla valutazione della performance di un pannello di microplotipi particolarmente informativi per scopi forensi come per la deconvoluzione delle miscele, identificazione personale e strette relazioni biologiche. Il pannello in esame è composto da 78 microplotipi, uno dei parametri fondamentale per la scelta dei microplotipi, è stato valutare l'ampiezza dell'amplicone, infatti minore sarà l'ampiezza e maggiore sarà la possibilità di poter usare questo set di microplotipi su campioni di DNA degradato. Un locus microhap, LMXA, non è stato esaminato, poiché lo SNP rs149217113 presente sul cromosoma 1 presenta un INDEL. Per cui il totale di microplotipi analizzati è stato quindi di 77.

Il pannello di microaplotipi è stato testato inizialmente su campioni che presentavano condizioni ottimali per l'analisi, si trattava di campioni con buoni livelli di DNA non degradato. Questa prova è stata effettuata per valutare l'efficienza dei microaplotipi già studiati e per confermare che il disegno dei *primer* fosse corretto. Considerando che lo scopo del pannello è quello di poter essere usato, data la dimensione degli ampliconi, su campioni che contengono DNA degradato, sono state effettuate prove anche su campioni reali forensi derivati da scene del crimine.

Tutte le letture sequenziate sono state allineate con (GRCh37/hg19). I livelli di *coverage* e di *reads on target* ottenuti per i campioni del pool 1, ovvero per i campioni dello studio popolazionistico, sono risultati accettabili. Al contrario invece, i livelli di *coverage* e di *reads on target* per i campioni del pool 2, ovvero i campioni forensi reali, risultano essere più bassi, probabilmente dovuti alla bassa quantità di DNA input e alla diversa matrice da cui sono stati estratti, permettendo comunque una buona tipizzazione. E' stato inoltre calcolato il *relative Depth of Coverage (rDoC)* dei singoli ampliconi. Tali risultati ci hanno permesso di valutare la performance dei microaplotipi e ci hanno permesso di individuare che il marcatore, il 6pB mostra un *coverage* normalizzato più basso rispetto agli altri microaplotipi che presenta un rischio di drop-out di loci o alleli. (*Grafico 1*). Allo stesso modo, nello studio condotta da De La Puente et

al. <sup>(31)</sup> ha evidenziato che il marcatore 6pB aveva una copertura di circa 10 volte inferiore rispetto agli altri microplotipi esaminati, per cui considerato underperforming. Una soluzione, per migliorare le performance del microplotipo, potrebbe essere il disegno di un nuovo primer, così come proposto anche da De La Puente et al. <sup>(31)</sup>

Per quanto riguarda la genotipizzazione invece, a seguito dell'analisi di identificazione delle varianti da parte del plug-in *HID\_Microhaplotype\_Research\_PluginV1.5*, i file BAM e BAI sono stati ispezionati con il software IGV. La problematica riscontrata, per entrambe le due tipologie di campioni, è stata la mancata chiamata allelica per 5 microplotipi: 1qB, 4pB, 7pA, 12qB e 17qB. Per tali microplotipi, abbiamo effettuato un'analisi su IGV, andando a selezionare manualmente le varianti alleliche per quei microplotipi in tutti i campioni. La mancata chiamata allelica è dovuta alla posizione del primo SNP che combacia con la prima posizione dell'amplicone. Questa problematica è stata risolta modificando le dimensioni del target iniziale; ad esempio per il microplotipo 1qB la dimensione del target inizialmente era 187622409-187622491 modificata poi a 187622398-187622509. I campioni sono stati sottoposti nuovamente ad analisi, ottenendo la corretta lettura da parte del Ion Torrent.

Per i campioni di DNA dello studio popolazionistico, il confronto dei risultati ottenuti dall'analisi primaria tramite plug-in e la successiva ispezione manuale tramite IGV software, ha rilevato in totale 11 chiamate discordanti. Tuttavia, la maggior parte di queste sono dovute ad artefatti. Infatti dal plugin dei microplotipi, venivano considerate come mixture, poiché potrebbero essere dovute alla presenza di tratti omopolimerici (AAAA[G/C] TT; ACACACAC), su ciascun lato dello SNP, queste ripetizioni vengono considerate erroneamente dall' Ion Torrent il quale producendo false inserzioni, crea una chiamata allelica errata.

E' infatti noto in letteratura che ripetizioni di tratti omopolimerici nei pressi dello SNP, possono essere causa di false inserzioni e delezioni nel sito SNP, come anche affrontato in uno studio condotto da De la Puente et. al <sup>(48)</sup> .

Per quanto riguarda la genotipizzazione dei campioni reali forensi, è stato notato che alcuni campioni presentavano più chiamate alleliche per alcuni microplotipi. Essendo campioni provenienti da scene del crimine, si è ipotizzato che potessero essere una miscela di più contributori. Anche in questo caso si è effettuato un controllo su IGV, confermando che si trattasse effettivamente di una mistura di più contributori. La genotipizzazione di questi campioni ha sottolineato la forte prestazione dei microplotipi nel rilevare campioni misti e identificare i contributori. In diversi campioni, come il

campione 12, campione 13, campione 14 è stato possibile rilevare gli alleli dei diversi contributori, quindi si può dire che quanto osservato con i microaplotipi concorda con gli STR, ad esempio nel numero dei contributori.

I campioni 19 e 20 mostrano corrispondenza con i dati attesi, come campioni di DNA a fonte singola, utilizzati come campioni di riferimento.

## Capitolo 6

### CONCLUSIONI

L'introduzione della tecnologia *Massive Parallel Sequencing* consente di analizzare nuovi marker genetici, come i microaplotipi. Per definizione un microaplotipo corrisponde ad una piccola regione del DNA (<300 bp) costituito da 2 o più SNPs geneticamente vicini. I microaplotipi risultano molto informativi per scopi forensi quali identificazione personale, identificazione di parentele, deconvoluzione di miscele e determinazione dell'ascendenza biogeografica. Inoltre questi nuovi marcatori sono potenzialmente adatti per l'analisi di campioni di DNA degradati.

Il presente studio si è focalizzato sulla validazione di un nuovo pannello da 78 microaplotipi, caratterizzati da valori di Ae superiori a 3 che risultassero efficienti nella deconvoluzione delle miscele e nell'identificazione individuale. Tale pannello è stato testato su campioni di DNA per uno studio popolazionistico, ma anche su campioni reali forensi derivanti da scene del crimine.

L'analisi dei microaplotipi nel pannello è stata ridotta a 77, poiché LMXA, non è stato esaminato, poiché lo SNP rs149217113 presente sul cromosoma 1 presenta un INDEL. Inoltre è emerso che un secondo microaplotipo il 6qB

risulta essere poco performante e questo lo si evince con maggiore chiarezza nella genotipizzazione dei campioni reali forensi.

Nel complesso, possiamo concludere che i risultati ottenuti con il pannello testato nel presente studio confermano l'utilità dei microhaps in ambito forense, com'è possibile osservare per il campione 12, il campione 13 e il campione 14, nei quali è possibile distinguere i diversi contributori di una miscela.

Anche se sono necessarie ulteriori analisi, come valutare la quantità del DNA di partenza, fondamentale per la corretta tipizzazione, perché lo scopo principale è quello di poter utilizzare questi nuovi marker generici per campioni di DNA fortemente degradati. Infatti, spesso in genetica forense ci si trova a dover analizzare campioni in cui la quantità di DNA è veramente minima.

## Capitolo 6

### BIBLIOGRAFIA

1. Landsteiner K. Ueber *Agglutinationsersche inungen normalen menschlichen* Blutes. 1901;14:1132. Wien. Klin. Wschr
2. Butler, J. M. (2009). *Fundamentals of forensic DNA typing*. Academic press;
3. Tagliabracci A., Alessandrini F., Mazzarini L., Onofri V., Onori N., Turchi C., *Introduzione alla genetica forense. Indagini di identificazione personale e di paternità*, Ancona, Springer Biomed. 2010;
4. Juan J. Sanchez; Chris Phillips ; Claus Børsting<sup>1</sup>; 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;
5. Susan Walsh , Alexander Lindenbergh , Sofia B. Zuniga , Titia Sijen , Peter de Knijff, Manfred Kayser , Kaye N. Ballantyne, *Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence*, Forensic Science International: Genetics 5 (2011);



6. <https://www.visage-h2020.eu/>;
7. C. Xavier, M. de la Puente, A. Mosquera-Miguel, A. Freire-Aradas, V. Kalamara, A. Ralf, A. Revoir, T.E. Gross, P.M. Schneider, C. Ames, C. Hohoff, C. Phillips, M. Kayser, W. Parso, on behalf of the VISAGE Consortium, *Development and inter-laboratory evaluation of the VISAGE Enhanced Tool for Appearance and Ancestry inference from DNA*, *Forensic Science International: Genetics* 61 (2022);
8. Ge, J.,B.,Plaz,J. V., & Chakraborty, R. (2010). *Haplotype block: a new type of forensic DNA markers*. *International journal of legal medicine*, 124, 353-361;
9. Richard Cordaux, Deepa Srikanta, Jungnam Lee , Mark Stonekin, Mark A. Batzer, *In search of polymorphic Alu insertions with restricted geographic distributions* *Genomics*, 2007; 90:154-158;
10. Schneider PM. et al. *The use of forensic DNA Phenotyping in Predicting Appearance and Biogeographic Ancestry*. 2019;
11. Oldoni F., Kidd K.K., Podini D., *Microhaplotypes in forensic genetics*, *Forensic Science International: Genetics*. 38 (2019) 54-69;
12. Kidd K.K., *Proposed nomenclature for microhaplotypes*, *Human Genomics*. 10 (2016)16;
13. <https://www.genenames.org/> HUGO Gene Nomenclature Committee;

14. Jing Zhu, Bing Long, Shengqiu Qu, Lu Yin, Wang He, Jiong Mao, Hui Wang, Lin Zhang, Bo Jin, Dan Chenc,, Weibo Liang, *A new proposed nomenclature for microhaplotypes*, *Forensic Science International: Genetics Supplement Series* 7 (2019);
15. Rajeevan H., Soundararajan U., Kidd J.R., Pakstis A.J., Kidd K.K., *ALFRED: an allele frequency resource for research and teaching*, *Nucleic Acids Research*. 2012 Jan; 40(Database issue): D1010–D1015;
16. <http://alfred.med.yale.edu>;
17. Turchi C., Melchionda F., Pesaresi M., Tagliabracci A., *Evaluation of a microhaplotypes panel for forensic genetics using massive parallel sequencing technology*, *Forensic Science International: Genetics*. 41 (2019) 120-127;
18. Kenneth K. Kidd, Judith R. Kidd, William C. Speed, Rixun Fang, Manohar R. Furtado, F.C.L. Hyland, Andrew J. Pakstis, *Expanding data and resources for forensic use of SNPs in individual identification*, *Forensic Science International: Genetics* 6 (2012);
19. Kenneth K. Kidd and Andrew J. Pakstis, *State of the Art for Microhaplotypes*, *Genes* 2022, 13, 1322;

20. 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;
21. Kidd, K.K.; Pakstis, A.J.; 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;
22. Zhaochen Bai a,c,1, Nan Zhang a,c,1, Jiawei Liu a,c, Heng Ding a, Yongkang Zhang a, Tian Wang a, Jun Gao b, Xueling Ou, *Identification of missing persons through kinship analysis by microhaplotype sequencing of single-source DNA and two-person DNA mixtures*, Forensic Science International: Genetics 58 (2022);
23. Jiaming Xue 1, Mengyu Tan 1, Ranran Zhang , Dezhi Chen , Guihong Liu , Yazhi Zheng , Qiushuo Wu , Yuanyuan Xiao , Miao Liao , Shengqiu Qu, Weibo Liang, *Evaluation of microhaplotype panels for complex kinship analysis using massively parallel sequencing*, Forensic Science International: Genetics 65 (2023) 102887;
24. Fabio Oldoni, Chiara Della Rocca, Daniele Podini, *Investigation of 74 microhaplotypes for kinship testing in US populations*, Forensic Science International: Genetics Supplement Series 8 (2022) 40–41;

25. Jiawen Yang, Ji Chen, Qiang Ji, Youjia Yu, Kai Li, Xiaochao Kong, Sumei Xie, Wenxuan Zhan, Zhengsheng Mao, Yanfang Yu, Ding Li, Peng Chen, Feng Chen, *A highly polymorphic panel of 40-plex microhaplotypes for the Chinese Han population and its application in estimating the number of contributors in DNA mixtures*, Forensic Science International: Genetics 56 (2022) 102600;
26. 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*. Forensic Sci. Int. Genet. 2022, 58, 102687;
27. <https://www.thermofisher.com/it/en/home/brands/ion-torrent.html>;
28. <https://corefacilities.iss.it/dw/doku.php?id=aree:ngs:start>;
29. Fabio Oldoni, Drew Bader, Chiara Fantinato, Sharon C. Wootton, Robert Lagac'e, Kenneth K. Kidd, Daniele Podini, *A sequence-based 74plex microhaplotype assay for analysis of forensic DNA mixtures*, Forensic Science International: Genetics 49 (2020) 102367;
30. Kenneth K Kidd and William C Speed *Criteria for selecting microhaplotypes: mixture detection and deconvolution*, Investigative Genetics (2015) 6:1;

31. M. de la Puente<sup>a,b</sup>, 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;
32. <https://www.ensembl.org/index.html>;
33. <https://www.ampliseq.com>;
34. QIAamp® DNA Micro Handbook, Third Edition, December 2014, *QIAGEN®*;
35. EZ1&2 DNA Investigator Kit Handbook, Qiagen 08/2022;
36. Precision ID SNP Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System APPLICATION GUIDE, Rev. C. 0;
37. <https://www.centrodimedicinabiologica.it/polimorfismi/>;
38. <https://www.ncbi.nlm.nih.gov/snp/?term=rs10800082>;
39. E-Gel™ Agarose Gels with SYBR Safe™, *Invitrogen, ThermoFisher Scientific* Rev. A.0;
40. PowerQuant® System, Revised 8/22 TMD047, *Promega*;
41. Pamela Tozzo, Caterina Politi, Arianna Delicati, Andrea Gabbin, Luciana Caenazzo, *External visible characteristics prediction through SNPs analysis in the forensic setting: a review*, DOI: 10.52586/4991;

42. Taishan Hu, Nilesh Chitnis, Dimitri Monos Anh Dinh,, *Next-generation sequencing technologies: An overview*, Human Immunology, Volume 82, Issue 11, November 2021, Pages 801-81;
43. *Introduction to next-generation sequencing Ion Torrent technology and instrumentation*. (2023) Thermo Fisher Scientific Inc;
44. Sara Goodwin, John D. McPherson and W. Richard McCombie, *Coming of age: ten years of next-generation sequencing technologies* NATURE REVIEWS, **GENETICS** VOLUME 17 JUNE 201;
45. Thermo Fisher Scientific, Real-Time PCR handbook;
46. <https://ionreporter.thermofisher.com/ionreporter/help/GUID7355C2DB-5AC9-4EF6-B166-1F55ABE0F1BA.html>;
47. Daniel S. Standage and Rebecca N. Mitchell, *MicroHapDB: A Portable and Extensible Database of All Published Microhaplotype Marker and Frequency Data*. *Frontiers in Genetics*, August 2020, Volume 11 Article 781;
48. M. de la Puente, C. Phillipsa, C. Santos, M. Fondevila, Á. Carracedo, M.V.Lareu, *Evaluation of the Qiagen 140-SNP forensic identification multiplex for massively parallel sequencing*. *Forensic Science International: Genetics* 28 (2017) 35–43;

*49. TECHNICAL NOTE The Ion AmpliSeq™ MH-74 Plex Microhaplotype*

*Research Panel, Revision A, 20 October 2021.*