

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

Corso di Laurea Magistrale in Biologia Molecolare e Applicata

ANALISI DELLA METILAZIONE DEL DNA PER LA PREVISIONE DELL'ETÀ BIOLOGICA

DNA METHYLATION FOR AGE PREDICTION

Tesi di Laurea Magistrale di: Angela Palmieri

Relatore: Chiar.mo Prof. Mauro Pesaresi

Sessione Autunnale

Anno Accademico: 2020/2021

Indice

Capitolo primo	1
INTRODUZIONE	1
Capitolo secondo	5
METILAZIONE DEL DNA	5
2.1 Epigenetica	5
2.2 Metilazione del DNA	7
2.3 Ruolo delle proteine DNMTs nella metilazione del DNA	10
2Ruolo della metilazione del DNA nel processo di invecchiament	o14
2.5 Analisi del livello di metilazione del DNA correlato all'età: applicazione in genetica forense	17
Capitolo terzo	27
SCOPO DELLA TESI	27
Capitolo quarto	29
MATERIALI E METODI	29
4.1 Campioni	29
4.2 Estrazione del DNA	29
4.3 Quantificazione del DNA	
4.4 Trattamento del DNA con bisolfito	
4.5 Quantificazione del DNA convertito con bisolfito	46
4.6 Multiplex PCR dei geni target	48
4.7 Purificazione enzimatica con ExoSAP-IT TM	49
4.8 Single-base extension (SBE)	50
4.9 Sequenziamento tramite elettroforesi capillare	53
4.10 Analisi della metilazione	56

4.11 Analisi statistica dei risultati e costruzione di un modello di previsio	one
	57
Capitolo cinque	59
RISULTATI	59
5.1 Analisi della correlazione tra livello di metilazione ed età	59
5.2 Costruzione di un modello di regressione lineare per la stima dell'et	à
biologica a partire da dati relativi alla metilazione	66
5.3 Predizione dell'età biologica utilizzando il modello di regressione	
lineare costruito	71
Capitolo sei	76
CONCLUSIONI E PROSPETTIVE FUTURE	76
Ringraziamenti	82
Bibliografia	83

Capitolo primo

INTRODUZIONE

L'invecchiamento è un processo graduale tipico di tutti gli organismi viventi, in base al quale è possibile determinare uno dei tratti distintivi fondamentali per l'identificazione personale, l'età (1). Per invecchiamento si intende un declino funzionale dipendente dal tempo, riscontrabile a livello molecolare tramite una serie di modificazioni cellulari che si accumulano durante la vita di un individuo.

I processi molecolari implicati nell'invecchiamento sono molteplici e, nel corso degli anni, sono stati fatti innumerevoli sforzi per identificare e classificare tutte le caratteristiche molecolari collegate a questo processo. hanno identificato Lopez et al.. nel 2013 (2),nove marcatori dell'invecchiamento: instabilità genomica, logoramento dei telomeri, perdita di turnover proteico, deregolazione del sensing dei nutrienti, disfunzione mitocondriale, senescenza cellulare, diminuzione delle cellule staminali, alterata comunicazione tra le cellule e alterazioni dal punto di vista epigenetico. La possibilità di risalire all'età di in individuo da un campione di DNA può rappresentare un enorme vantaggio nel campo delle investigazioni forensi poiché questo tipo di determinazione fornirebbe un ulteriore restringimento del

1

campo di ricerca fornendo un dettaglio utilissimo all'identificazione personale, oltre al già utilizzato profiling tramite amplificazione e analisi di STR. La genetica forense, sviluppatasi su tecniche basate sulle proteine circa vent'anni fa, è diventata importantissima grazie a tecniche di profiling che sfruttano l'amplificazione simultanea di alcune ripetizioni in tandem (STR) che differiscono negli individui di una popolazione. Questo tipo di analisi attualmente rappresenta il "gold standard" per l'identificazione genetica. Nonostante ciò, la nuova idea di fingerprinting mira all'acquisizione di più informazioni possibili a partire da un'unica esigua traccia biologica e quindi dal DNA di partenza. (3)

Uno degli approcci più seguiti è quello che mira alla definizione di tratti fenotipici, come colore degli occhi, dei capelli, della pelle, etnia ed età, a partire da un campione di DNA. L'età, in particolare, è un elemento fenotipico che è stato fortemente ricercato tramite una serie di studi di genetica forense che includevano diversi approcci. Tra tutte le metodiche testate, l'analisi della metilazione di markers (ricchi di isole CpG) accuratamente selezionati, sembra essere l'approccio più promettente per risalire, in modo anche molto accurato (con una precisione di 3/4 anni), all'età di un individuo dal suo DNA. (3). Inoltre, oltre al campo della genetica forense, lo studio dell'invecchiamento dal punto di vista molecolare potrebbe aprire la strada ad ulteriori studi in ambito clinico.

La relazione tra i segnali ambientali e l'epigenetica non è stata ancora ben definita nei mammiferi ma la presenza di regioni del DNA ipo- ed ipermetilate è stata associata, in numerosi studi, all'età e, quindi, al processo di invecchiamento. Studi effettuati su gemelli omozigoti hanno definito relazioni tra l'invecchiamento ed effetti epigenetici a lungo termine sul fenotipo (4). I gemelli omozigoti rappresentano un'ottima base per lo studio dei fenomeni epigenetici, poiché condividono le stesse basi genetiche. Nei primi anni, i gemelli omozigoti mostrano lo stesso pattern di metilazione che cambia, però, nel corso della loro vita, mostrando livelli e regioni di metilazione diversi. Alcuni studi condotti da Pirazzini et al. (5) hanno permesso di individuare due regioni genomiche in cui, dopo i 60 anni di età, i gemelli mostravano forti differenze in termini di metilazione. Le differenze nel grado di metilazione, in realtà, interessano solo particolari regioni del genoma localizzate, spesso, all'interno di geni che codificano per proteine con funzioni diverse, non correlate tra loro. Negli ultimi anni, soprattutto nel campo della genetica forense, lo scopo principale è stato proprio quello di individuare queste regioni specifiche del DNA umano che possano rappresentare dei validi marker da investigare in termini epigenetici al fine di risalire, con una certa accuratezza, ad uno stretto range di età dell'individuo a cui il DNA appartiene.

Le differenze dei pattern di metilazione su regioni genomiche di individui di età diverse è correlata al processo di invecchiamento che rende meno efficiente il meccanismo coinvolto. E' stato, infatti, dimostrato che, con l'avanzare del tempo, il meccanismo di metilazione del DNA tende a perdere la sua abilità di mantenere un certo pattern durante le divisioni cellulari (6).

Questi primi studi che correlavano il grado di metilazione del DNA all'età sollevarono, quindi, nuove domande su come le alterazioni epigenetiche tessuto-specifiche e regione-specifiche potessero essere influenzate dal fenomeno dell'invecchiamento. E' stata proposta l'ipotesi che i diversi patterns di metilazione riscontrati fossero il frutto di un processo stocastico di deriva epigenetica casuale (7).

Capitolo secondo

METILAZIONE DEL DNA

2.1 Epigenetica

Il termine "epigenetica" descrive una serie di modificazioni ereditabili non attribuibili a modificazioni nella sequenza primaria del DNA (mutazioni). Le modificazioni epigenetiche giocano un ruolo fondamentale in numerose funzioni come la regolazione dell'espressione genica e la regolazione dello sviluppo cellulare.

La definizione precisa di epigenetica ha confuso gli scienziati per tanti anni. Al di sopra del codice genetico, il codice epigenetico comprende un ulteriore strato di informazioni. Mentre il primo fornisce una cornice di lettura per l'RNA e per la struttura proteica, il codice epigenetico controlla l'impacchettamento del DNA e la regolazione genica.

Il termine "epigenetica" è stato introdotto, per la prima volta, nei primi anni '40, da Conrad Waddington che definì l'epigenetica come "la branca della biologia che studia le interazioni casuali tra i geni e i loro prodotti che portano al fenotipo" (8). La definizione di epigenetica, da quel momento, con lo sviluppo della ricerca nel campo della genetica, è stata continuamente modificata e continua ad acquisire sempre nuove sfaccettature. Riggs et al. (9), nel 1996, fornirono una nuova definizione di epigenetica, intesa come "lo studio dei cambiamenti ereditabili mitoticamente o meioticamente della funzione di un gene che non possono essere spiegati da cambiamenti nella sequenza nucleotidica del DNA". Una definizione oggi largamente accettata descrive l'epigenetica come "lo studio dei processi che producono un fenotipo ereditabile che non dipende in maniera stretta dalla sequenza del DNA" (10). I processi epigenetici includono modificazioni istoniche, la metilazione del DNA, il rimodellamento della cromatina e meccanismi in cui sono coinvolti RNA non codificanti; tutti questi meccanismi giocano un ruolo specifico nella regolazione dell'espressione genica senza coinvolgere cambiamenti nella sequenza del DNA.

Le basi molecolari dell'epigenetica sono molteplici ma hanno, come linea generale, un ruolo nell'alterazione dell'attivazione di geni specifici.

Si pensa che il programma epigenetico inizi già a livello dello sviluppo fetale nell'utero. Così come il DNA viene trasmesso da una generazione a quella successiva, così i patterns epigenetici rimangono stabili e vengono quindi preservati durante la divisione cellulare. Cambiamenti nei pattern epigenetici vengono, invece, osservati durante la vita di un individuo. Essi sono dovuti ad una serie di fattori diversi: età, esposizione ad un certo tipo di ambiente, dieta e stile di vita in generale. Gli effetti finali dei processi epigenetici sono numerosissimi e vanno dall'imprinting genomico alla riprogrammazione genica, dal silenziamento genico alla carcinogenesi, fino all'inattivazione di interi cromosomi, come nel caso dell'inattivazione del cromosoma X, nella femmina. Nei mammiferi, un'importante funzione cellulare regolata da processi epigenetici è il differenziamento cellulare durante l'embriogenesi, a partire da cellule staminali completamente indifferenziate.

2.2 Metilazione del DNA

Tra le diverse modificazioni epigenetiche, la metilazione del DNA è quella che è stata meglio caratterizzata. Essa riveste un ruolo importantissimo nello sviluppo, tale che la perdita di metilazione conduce le cellule ad apoptosi o arresto della crescita.

Nel genoma dei mammiferi, la metilazione del DNA è un meccanismo epigenetico che consiste nel trasferimento di un gruppo metile a livello del carbonio 5 (C5) di una citosina a formare 5-metilcitosina (Fig. 2.2.1). Le citosine, generalmente, coinvolte nella metilazione si trovano in regioni specifiche del genoma conosciute come isole CpG. In una sequenza di DNA, quando una guanina (G) è separata da una citosina (C) tramite un solo residuo fosfato (p), si parla di sito CpG. A livello del genoma, esistono clusters caratterizzati da un'alta frequenza di siti CpG, noti come isole CpG. Le isole CpG vengono identificate come regioni con una lunghezza maggiore di 200 paia di basi, contenuto in G + C maggiore del 50% e rapporto tra siti CpG osservati e attesi di 0.6 (11). Queste regioni sono, spesso, associate ai promotori di alcuni geni tanto che si stima che il 70% dei promotori contengano isole CpG al loro interno. I promotori possono, infatti, essere classificati in base alla loro densità di isole CpG. Questi siti giocano un ruolo importantissimo nella regione del promotore, principalmente in termini di silenziamento genico. Inoltre, alcuni studi, come quello condotto da Yang et al. nel 1996 (12), dimostrano che l'insorgenza di mutazioni C \rightarrow T a livello dei siti CpG aumenta il rischio di sviluppare malattie ereditarie.

Le isole CpG mostrano una struttura ben precisa: esse sono costituite da estremità fiancheggianti ("shores"), le quali si trovano ad una distanza di 2kb dalle isole vere e proprie, e da ulteriori sequenze, anch'esse fiancheggianti le isole CpG, dette "shelves", a 2-4 kb dalle isole CpG. Le regioni "shores" contengono i pattern di metilazione tessuto-specifici più conservati (13).

La metilazione del DNA regola l'espressione genica tramite il reclutamento di proteine coinvolte nel silenziamento di geni o nell'inibizione del legame di fattori di trascrizione al DNA. Durante lo sviluppo, il pattern di metilazione del DNA cambia continuamente in seguito ad un processo dinamico fatto di continue metilazioni e demetilazioni. Da questo processo risulta che cellule differenziate sviluppano un pattern di metilazione del DNA stabile e, soprattutto, unico, in modo da regolare, in maniera tessuto-specifica, la trascrizione dei geni (5). Quindi, nonostante virtualmente tutte le cellule di un organismo contengano la stessa informazione genetica, non tutti i geni sono espressi in maniera simultanea da tutti i tipi cellulari. In un certo senso, i meccanismi epigenetici vanno a delineare un certo profilo di espressione in tipi cellulari e in organismi diversi.



Figura 2.2.1 Metilazione della citosina, da https://www.pianetachimica.it/mol_mese/mol_mese_2011/07_ DNA_metiltransferasi/DNA_metiltransferasi.htm

Nel 1948, Rollin Hotchkiss, per primo, scoprì l'esistenza di citosina modificata in una preparazione di timo di vitello, tramite cromatografia su carta. Hotchkiss ipotizzò che si potesse trattare di 5-metilcitosina (5mC) perché si separava in modo simile al modo in cui la timina (nota anche come metiluracile) si separava dall'uracile. Egli, inoltre, suggerì che questa citosina modificata potesse esistere naturalmente nel DNA.

Solo negli anni '80 numerosi studi dimostrarono che la metilazione del DNA era coinvolta nella regolazione genica e nel differenziamento cellulare (Holliday and Pugh, 1975; Compere and Palmiter, 1981).

La metilazione del DNA è un processo catalizzato da una famiglia di DNA metiltrasferasi (DNMTs) che trasferiscono un gruppo metile dall'S-adenil metionina (SAM) al carbonio 5 di un residuo di citosina, a formare 5mC (Figura 1.1). Dopo aver donato un gruppo metile alla citosina, il prodotto finale della reazione consiste in un residuo di S-adenosilomocisteina (SAH), che si comporta da inibitore delle DNMTs. Le molecole coinvolte nella donazione di gruppi metilici, essenziali nelle reazioni di metilazione, provengono, generalmente, dal metabolismo dei folati, Questi ultimi, oltre ad essere coinvolti nella donazione di gruppi metilici, agiscono come cofattori di enzimi chiave della sintesi di DNA ed RNA. Quindi, il supporto del folato e di altri donatori di gruppi metilici, come la metionina e la colina, ha una grandissima rilevanza biologica.

2.3 Ruolo delle proteine DNMTs nella metilazione del DNA

Mantainance methylation	DNMT1
RNA methylation	DNMT2
De novo methylation	DNMT3a
De novo methylation	DNMT3b

Tabella 2.3-1. Enzimi della famiglia delle DNMTs coinvolti nella metilazione, con rispettiva funzione

Come già accennato, nelle cellule, l'addizione di un gruppo metile è controllata a vari livelli da un insieme di proteine, la famiglia delle DNMTs (Figura 2.3-1).Nonostante questi enzimi condividano strutture molto simili tra loro, caratterizzate da un grande dominio N-terminale con funzione regolatrice e da un dominio catalitico C-terminale, ognuno di essi presenta è dotato di un meccanismo di azione e di una funzione unici (Tabella 2.3-1)

Dnmt1 è, probabilmente, l'enzima di questa famiglia più studiato, soprattutto a livello del sistema nervoso dei mammiferi, nei cui tessuti, questo enzima è altamente espresso. Una delle differenze con le altre Dnmts sta nel fatto che Dnmt1 agisce metilando, in maniera preferenziale, DNA emimetilato, ovvero DNA metilato a livello di un solo filamento e non di quello complementare. Durante la replicazione del DNA, infatti, Dnmt1 si posiziona a livello della forcella replicativa, dove viene emimetilato il nuovo filamento appena sintetizzato. Dnmt1 si lega al filamento e lo metila seguendo esattamente il pattern di metilazione originale presente prima della replicazione del DNA.

Oltre a questo tipo di azione, Dnmt1 possiede, inoltre, la capacità di riparare la metilazione del DNA, mantenendo un certo pattern di metilazione in una linea cellulare. Per questo motivo, Dnmt1 è conosciuto anche come Dnmt di mantenimento, giocando, quindi, un ruolo importantissimo nei processi di differenziamento e divisione cellulare.

Dnmt3a e Dnmt3b sono estremamente simili tra loro sia in struttura che in funzione. A differenza di Dnmt1, entrambi, se espressi ad opportuni livelli, sono capaci di metilare entrambi i filamenti (nativo e appena sintetizzato) senza mostrare alcuna preferenza per il DNA emimetilato. Per questa ragione, Dnmt3a e Dnmt3b sono dette *de novo* Dnmt (Figura 1.1a). Ciò che permette di distinguere Dnmt3a da Dnmt3b è, essenzialmente, il pattern di espressione dei loro geni: mentre il gene *Dnmt3a* è espresso in maniera relativamente ubiquitaria, *Dnmt3b* è poco espresso nella maggior parte dei tessuti differenziati, ad eccezione della tiroide, del testicolo e del midollo osseo. Mentre il knockout di *Dnmt1* e *Dnmt3b* risulta essere letale per gli embrioni di topo, il knockout di Dnmt3a porta ad una sopravvivenza del topo per circa 4 settimane dopo la nascita (15). Questi risultati suggeriscono che Dnmt3b sia

richiesto durante le prime fasi di sviluppo embrionale, mentre Dnmt3a sia richiesto, principalmente, per il normale differenziamento cellulare. Comunque, le tre Dnmts sono ampiamente coinvolte nello sviluppo e, nel momento in cui le cellule raggiungono il loro completo differenziamento, la loro espressione si riduce molto. Questo potrebbe suggerire che il pattern di metilazione nelle cellule in fase post-mitotica sia stabile anche se questo non è del tutto vero: individui di età biologica diversa sembrano avere pattern di metilazione diversi a livello di determinati loci, identificati come markers nella determinazione dell'età biologica di in individuo in base al loro livello di metilazione.

L'ultimo membro della famiglia delle Dnmts è Dnmt3L, una proteina che, a differenza degli altri membri, manca del dominio catalitico. Dnmt3L è espressa principalmente nelle prime fasi di sviluppo mentre, in età adulta, è limitata alle cellule germinali e al timo. Nonostante questa proteina non abbia una funzione catalitica in sé, si associa a Dnmt3a e Dnmt3b stimolando la loro attività di metiltrasferasi.



Fig. 1.3-1. Pathway di metilazione del DNA da: Lisa D Moore, Thuc Le and Guoping Fan, "DNA Methylation and Its Basic Function", Neuropsychopharmacology Reviews (2013) 38, 23–38

2.4 Ruolo della metilazione del DNA nel processo di invecchiamento

L'invecchiamento è un processo graduale ed inevitabile della vita umana ed è influenzato da numerosissimi fattori legati all'ereditarietà, all'ambiente, allo stile di vita e alle malattie. Il processo di invecchiamento modifica l'espressione genica, i telomeri e le strutture cellulari di un organismo. In particolare, una serie di modificazioni biochimiche che includono, quindi, la metilazione del DNA, sono fortemente correlate all'invecchiamento umano. L'individuazione di biomarkers associati all'invecchiamento è un tema fortemente investigato in numerosi studi. Per biomarker si intende una caratteristica naturale (non indotta artificialmente), che può essere valutata e analizzata per ottenere informazioni su un cambiamento biologico di un individuo ad essa correlato.

Studi precedenti hanno individuato determinate isole CpG, il cui grado di metilazione sembra essere correlato all'età dell'individuo e quindi al suo grado di invecchiamento. Queste regioni sono state identificate come "age-related CpG" (AR-CpG) e sono state suddivise in due categorie: ipermetilate ed ipometilate. I siti AR-CpG possono essere utili nella determinazione dell'età di un individuo a scopo forense o nella valutazione del rischio di incorrere in patologie associate all'invecchiamento, come il cancro. Numerosi studi rivelano, quindi, che numerose isole CpG potrebbero mostrare gradi di metilazione differenti correlati all'età. Questi siti sono, quindi, dei promettenti biomarkers per la predizione dell'età biologica a scopo forense. Numerosi sforzi sono ancora in atto da parte dei ricercatori di tutto il mondo, per poter individuare sempre più siti CpG da analizzare come biomarkers.

Sono stati sviluppati numerosi modelli di predizione dell'età biologica. Molti di questi modelli sono basati sull'analisi della lunghezza dei telomeri, i quali si trovano alla fine dei cromosomi eucariotici e la cui lunghezza aumenta ad opera

15

di telomerasi tramite l'aggiunta di basi alla fine del DNA, processo che aumenta ulteriormente il grado di metilazione. In altre parole, l'aumento della lunghezza dei telomeri aumenta il grado di metilazione del DNA, aumentando il tempo di sopravvivenza di un individuo. Al contrario, ad una diminuzione della lunghezza dei telomeri corrisponde un minore grado di metilazione del DNA, riducendo il tempo di sopravvivenza di una persona. In termini numerici, la 5mC rappresenta l'1.5% dell'intero DNA genomico. Questa percentuale decresce notevolmente durante l'ontogenesi, risultando massima negli embrioni e diminuendo gradualmente all'aumentare dell'età di un individuo (6) Quindi, trovare sempre più siti CpG il cui livello di metilazione è correlato all'età, aiuta, oltre a determinare l'età di un individuo a scopo forense (ad esempio per investigare l'età di un individuo in base a sangue o fluidi biologici trovati sulla scena del crimine), a capire la biologia del processo di invecchiamento e del rischio di sviluppare patologie ad esso associate come il cancro. L'ipermetilazione associata all'età è stata descritta per un certo numero di isole CpG presenti, spesso, in geni associati a tumori, allo sviluppo di metastasi e alla detossificazione. Tra tutti i markers individuati, il locus ELOV2 è stato quello maggiormente valutato per la predizione dell'età biologica, in quanto associato al grado di invecchiamento in numerosi organi e tessuti (7). Un altro modello di predizione dell'età biologica basato su ELOV2 e

analizzando la metilazione di 71 loci, ha mostrato una grande accuratezza evidenziando anche il ruolo della metilazione del DNA in patologie correlate all'invecchiamento (8). In uno studio di Mawlood et al. (9) è stata, invece, evidenziata l'importanza della metilazione del DNA mitocondriale come marker biologico per la determinazione dell'età biologica in campo forense. Altri studi analoghi si sono susseguiti negli ultimi anni. Quelli più recenti prevedono lo studio di patterns di metilazione su arrays che permettono di analizzare un elevato numero di isole CpG. Questo modello permette di risalire, con una certa accuratezza, all'età di un individuo a partire da un campione di sangue, e rappresenta, quindi, un ottimo candidato nel campo della genetica forense. ELOV2 sembra essere il marker migliore per la valutazione dell'età in base al livello della sua metilazione anche in campioni di sangue non recenti, la più comune fonte di DNA in ambito forense. Le previsioni tramite questo marker mostrano un coefficiente di correlazione del 96% con un errore standard (SE) di 3.9 anni (8).

2.5 Analisi del livello di metilazione del DNA correlato all'età: applicazione in genetica forense

Il DNA è la fonte biologica più frequentemente utilizzata per l'identificazione personale a scopo forense. Spesso, un'esigua quantità di fluido biologico è sufficiente per risolvere un caso quindi l'abilità giungere all'identificazione in una maniera non-distruttiva della fonte recuperata sulla scena del crimine rappresenta un imperativo al fine di poter preservare il campione per un ulteriore uso o conferma. I test di identificazione esistenti sono applicabili su specifici fluidi biologici e sono diversi a seconda del tipo di campione. Attualmente, per analizzare DNA estratto da un fluido biologico, si utilizzano test basati sulla PCR. I profili così ottenuti vengono confrontati con quelli depositati in databases. Esistono, a questo proposito, innumerevoli sistemi disponibili in commercio che sono stati approvati per applicazioni forensi.

I fluidi biologici maggiormente riscontrabili sulla scena di un crimine sono rappresentati da sangue, saliva, liquido seminale e vaginale, oltre a sudore e urina che possono rappresentare ulteriori fonti di evidenza. Ciascuno di questi fluidi contiene DNA quindi è importante che i test siano caratterizzati da un'estrema precisione e da misure precauzionali molto stringenti. Sono stati sviluppati numerosi test sia presuntivi che confermativi per l'identificazione personale da fluidi biologici. I test presuntivi sono utilizzati principalmente come test di screening e tendono ad avere limitazioni in termini di precisione, mentre i test confermativi sono utilizzati per un'identificazione assoluta a partire da un tessuto e possono essere utili per ricostruire gli eventi di un crimine. La stima dell'età biologica di un donatore di un determinato tipo di campione, gioca un ruolo importantissimo nel campo delle indagini forensi. Si tratta, infatti, di un elemento che fornisce importanti indizi nella ricerca di un soggetto sconosciuto per diverse ragioni: l'età biologica è un elemento essenziale per l'identificazione da resti scheletrici, ad esempio. Inoltre, la determinazione dell'età può essere un elemento critico in drammatici casi di immigrazione in cui l'identità e l'età di un numero, spesso, molto alto di cadaveri, non è chiara in seguito all'esame autoptico. Gli attuali schemi di determinazione dell'età includono analisi morfologiche di determinate caratteristiche dello scheletro (20). Queste modalità sono, senz'altro, ancora molto utili per le indagini su resti umani ma esse sono applicabili solo in casi in cui sono riscontrabili resti costituiti da tessuti solidi come ossa e denti. Le stesse limitazioni si applicano ai metodi chimici di stima dell'età, in cui è preferibilmente richiesta la presenza di campioni di denti (21).

I metodi molecolari per la determinazione dell'età biologica stanno, quindi, trovando campo poiché non hanno limitazioni di questo genere ma, al contrario, sono applicabili a qualsiasi tipo di tessuto dal quale è possibile estrarre il DNA. Il primo marker molecolare ad essere investigato, in questa direzione, è stato il DNA mitocondriale, poiché era noto che la sua sequenza nucleotidica cambiasse durante l'invecchiamento fisiologico di un individuo, mostrando un carico maggiore di eteroplasmie puntiformi e un numero maggiore di delezioni su larga scala (22). Le tecniche basate su questo target mostrano, però, una bassa accuratezza poiché il grado di associazione di queste modificazioni con l'età risulta essere piuttosto debole.

Per quanto riguarda il DNA nucleare, invece, il primo elemento ad essere individuato come possibile marker molecolare per la predizione dell'età, sono state le sequenze telomeriche conosciute per essere più corte ad ogni divisione cellulare. Nonostante questo, l'applicazione di questo concetto a tecniche che permettano di prevedere l'età biologica, non sembra così semplice. Infatti, i primi studi in questo senso hanno mostrato un margine di errore tra i 22 e i 10 anni circa (23), valori troppo elevati per un modello accurato di predizione applicabile a scopo forense. Erano, quindi, fortemente richieste tecniche più accurate che permettessero una stima precisa dell'età per applicazioni forensi. Un metodo promettente deriva dal campo dell'epigenetica. Come già ampiamente discusso, è stato osservato un cambiamento del livello di metilazione del DNA globale associato all'aumento dell'età, in numerosi studi. Dal momento che la metilazione del DNA è la componente più accessibile e meglio caratterizzata dei fenomeni epigenetici che identificano una certa regione genomica, ed è inoltre una modificazione altamente stabile, lo stato di metilazione di DNA di un certo locus è un ottimo candidato come biomarker

sia in ambito clinico che in ambito forense. L'identificazione personale da fluidi biologici basata sulla metilazione del DNA ha un grande potenziale, soprattutto in genetica forense, dove il riscontro di fluidi biologici sulla scena del crimine rappresenta un aspetto cruciale nelle indagini.

Da una prospettiva prettamente forense, l'analisi del profilo di metilazione di una certa regione genomica può fornire informazioni utili direttamente collegate alle circostanze della violenza o del decesso. Le tecnologie che permettono di ricavare profili di metilazione del DNA possono fornire dettagli su come e quando i fluidi biologici sono stati rilasciati sulla scena del crimine, oltre a permettere l'identificazione tramite importanti informazioni come età, genere e altre caratteristiche fenotipiche come la presenza di particolari patologie.

L'analisi dello stato di metilazione del DNA è un approccio vantaggioso poiché mostra un'alta sensibilità e specificità. Inoltre, si tratta di un approccio standardizzabile con i metodi di estrazione e purificazione del DNA attualmente utilizzati e accreditati in ambito forense che permettono un minimo utilizzo del materiale biologico di partenza che, quindi, può essere utilizzato più volte per ulteriori test. In altre parole, i test basati sulla metilazione del DNA risultano efficienti e convenienti in ambito forense perché permettono di non alterare completamente il campione, elemento cruciale di un'indagine. E' stata svolta un'estensiva ricerca sugli effetti dell'età sulla metilazione, da cui i genetisti forensi hanno attinto per poter stimare l'età di individui coinvolti nelle indagini o individuare la possibile età biologica al momento della morte da cadaveri rinvenuti dopo molti anni, tramite l'analisi di patterns di metilazione.

Molti studi hanno permesso di identificare una serie di markers CpG associati all'età e modelli di predizione dell'età specifici per determinati tipi di tessuto o di fluido biologico. L'analisi degli stessi markers in tessuti diversi ha rivelato errori di valutazione dell'età biologica di più di 10 anni, in alcuni casi. Questa osservazione suggerisce che i migliori markers per un'accurata predizione dell'età siano differenti nei vari tipi di cellule e tessuti.

L'approccio intrapreso nel campo forense è quello di predizione dell'età biologica utilizzando solo pochi markers, approccio che potrebbe essere vantaggioso, data l'esigua quantità di DNA generalmente trovata su una scena del crimine.

Attualmente, l'approccio più comunemente perseguito per l'analisi dei siti CpG è il sequenziamento del DNA precedentemente convertito con bisolfito (Figura 2.5-1). Questo approccio consiste nel trattamento del DNA a singolo filamento con bisolfito di sodio che conduce alla deaminazione delle citosine non metilate ad uracile, mentre le citosine metilate rimangono inalterate.



Figura 2.5-1. Passaggi di un tipico sequenziamento con bisolfito, confrontati con i passaggi di un sequenziamento classico da DNA Sequencing by Capillary Electrophoresis Chemistry Guide, Thermo Fisher Scientific

Nel 2011, tramite una semplice estrazione di DNA e arrays di tipo Illumina HumanMethylation27, un gruppo di ricercatori ha sviluppato un modello di regressione per ottenere una stima dell'età a partire da campioni biologici. In particolare, sono stati esaminati campioni di saliva appartenenti a partecipanti di età compresa tra i 21 e i 55 anni. Questo studio ha rivelato che il livello di metilazione di tre siti situati a livello dei promotori dei geni TOM1L1 e EDARADD diminuisce con l'età mentre il grado di metilazione di alcune regioni del gene NPTX2 aumenta all'aumentare dell'età biologica, in maniera progressiva. Questo modello di regressione aveva un'accuratezza di circa 5 anni (13). Studi più recenti hanno raggiunto un livello di accuratezza ancora maggiore, prendendo in considerazione molti più loci contemporaneamente.

Weidner et al. (20) hanno dimostrato la possibilità di predire l'età biologica con grande accuratezza tramite l'analisi della metilazione a livello di soli 3 siti CpG dei geni ITGA2B, ASPA e PDE4C. Il modello di Weidner et al. si è basato sul pirosequenziamento per misurare la metilazione e ha mostrato un'accuratezza di predizione molto alta, con una deviazione media assoluta (MAD) di 4.5 anni dall'età cronologica. Successivamente, Eipel et al. hanno applicato il modello su sangue, proposto da Weidner et al. a campioni ricavati tramite tamponi buccali. E' stata osservata una sovrastima dell'età biologica di 14.6 anni, tale da presupporre che la metilazione a livello dei 3 siti CpG dei geni ITGA2B, ASPA e PDE4C sia altamente tessuto-specifica. Occorre, quindi, rilevare dei siti CpG che siano analizzabili in maniera specifica nel caso in cui i campioni siano rappresentati da tamponi buccali. Un nuovo modello che include i 3 markers specifici per il sangue e l'aggiunta di due markers specifici per le cellule epiteliali buccali, mostra una diminuzione della MAD a 5.1 anni. Dati

simili sono stati ottenuti su campioni di saliva contenente sia cellule epiteliali buccali che leucociti, in percentuale diversa.

Hong et al. (21) hanno identificato 6 markers associati all'età (siti CpG dei geni SST, CNGA3, KLF14, TSSK6, TBR1 e SLC12A5), suggerendo, però, di considerare, insieme a questi markers, un ulteriore sito CpG del gene PTPN7, la cui metilazione risultava essere altamente specifica per un determinato tipo cellulare. Questo sistema rendeva la predizione più accurata su campioni di saliva: mentre il modello composto da sei markers mostrava un'accuratezza di predizione identificata con un valore di MAD di 4.1 anni, il modello con un ulteriore sito, specifico per il tipo di campione analizzato, mostrava una MAD leggermente migliore, di 3.2 anni. Per quanto riguarda questo studio, vi è una variazione di MAD minima rispetto a quella osservata nello studio di Eipel et al. e questo può essere attribuito al fatto che campioni di saliva, rispetto a tamponi buccali, non hanno percentuali di tipi cellulari tanto diverse quanto quelle di campioni di sangue rispetto a tamponi buccali. In altre parole, un tampone buccale ha una composizione cellulare simile ad un campione di saliva mentre un campione di sangue è composto da cellule diverse, in proporzioni estremamente diverse.

I modelli di predizione elaborati da Hong et al. sembrano, quindi, avere un potenziale maggiore rispetto a quelli precedenti perché utilizzabili su tessuti

25

diversi ed eterogenei: almeno 3 dei 6 markers studiati (siti CpG di SST, KLF14 e SLC12A5) funzionano bene sia su campioni di sangue che su campioni di saliva (25).

Dagli studi di Hong sull'epigenoma, sono stati identificati altri siti CpG, a livello dei geni ELOVL2, C1orf132 e TRIM59. Anch'essi hanno mostrato un elevato grado di correlazione tra metilazione ed età sia nel sangue che nella saliva.

Tra tutti i modelli avanzati in campo forense, quello che sembra essere più promettente è stato proposto da Zbiec-Piekarska et al. (17). Si tratta di un modello in cui vengono presi in considerazione, per l'analisi della metilazione, cinque diversi siti CpG di cinque geni diversi: ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C e TRIM59. Già i primi studi, del 2014, su questo modello hanno mostrato un'elevata accuratezza di predizione, con valori di MAD tra 3.9 anni e 4.2 anni, a seconda dell'etnia dei donatori.

Capitolo terzo

SCOPO DELLA TESI

La predizione dell'età di un soggetto ignoto è uno scopo molto perseguito in genetica forense. L'idea di poter definire, con una certa accuratezza, l'età di un individuo tramite un semplice tampone buccale potrebbe rappresentare una vera e propria rivoluzione nel campo della medicina legale in cui, per la determinazione dell'età di cadaveri ignoti si fa riferimento, principalmente, ad analisi morfologiche su tessuti solidi.

Già numerosi studi precedenti hanno identificato modelli di predizione dell'età biologica ottenuti grazie all'analisi di markers la cui metilazione risultava essere associata all'età. Da questi studi è emerso che il livello di metilazione di markers diversi è altamente condizionato dal tipo di tessuto in cui si trova, quindi, lo scopo di questi studi è quello di identificare per ciascun tipo di campione il marker o i markers che meglio possano fornire un output di età predetta il più accurato possibile.

Per fare ciò, è necessario raccogliere dati sul livello di metilazione di marker diversi da soggetti ad età nota. In base ai dati ottenuti è possibile costruire un modello di regressione lineare per ogni target considerato e testarlo confrontando l'output di età predetta con l'età cronologica dell'individuo. Studi

27

precedenti hanno raggiunto risultati incoraggianti con la costruzione di modelli di previsione con un'accuratezza di circa 4 anni per quanto riguarda l'analisi di campioni di sangue e saliva. L'analisi di tamponi buccali, invece, risulta ancora poco accurata rispetto alle altre, con una deviazione dai modelli di predizione costruiti che si aggira intorno ai 5 anni.

In questo lavoro di tesi sperimentale, sono stati presi in considerazione tre siti CpG di tre geni: ELOVL2, FHL2 e TRIM59, già identificati da studi precedenti come markers di metilazione correlata all'età, in campioni di sangue, saliva e tamponi buccali.

L'idea è quella di analizzare la metilazione a livello di questi siti in maniera simultanea al fine di realizzare un modello di regressione lineare che possa ricondurre ad un valore di età prevista partendo da dati relativi al grado di metilazione dei siti CpG presi in considerazione. Inoltre, assumendo che la metilazione sia un fenomeno estremamente dipendente dal tipo cellulare, lo scopo dello studio è quello di identificare, tra i tre siti, quello che mostra un'accuratezza maggiore nella previsione dell'età biologica rispetto agli altri, partendo da cellule della mucosa orale, ovvero quelle prelevate tramite un tampone buccale.

Capitolo quarto

MATERIALI E METODI

4.1 Campioni

Il DNA da analizzare è stato estratto da 55 tamponi buccali di uomini e donne sani tra i 18 e i 95 anni. Tramite un tampone buccale è possibile prelevare cellule epiteliali della mucosa orale, raschiando entrambi i lati della cavità orale stessa del soggetto donatore, ed evitando il contatto con denti e saliva.

Una volta prelevati, i tamponi sono stati congelati ad una temperatura di -20°C, fino al successivo step di estrazione.



Figura 4.1-1. Prelievo di cellule epiteliali della mucosa orale per l'estrazione del DNA da https://www.analisidna.it/istruzioni-tampone-buccale-per-test-di-paternita/

4.2 Estrazione del DNA

L'estrazione del DNA dai singoli tamponi, opportunamente identificati tramite un codice univoco, è stata eseguita mediante l'utilizzo del kit di purificazione NucleoMag® DNA Forensic (Macherey-Nagel). I passaggi di estrazione manuali sono stati sostituiti dall'utilizzo di un estrattore automatico, il KingFisher Ml automatic extractor (ThermoLab Systems). Il kit utilizzato contiene i seguenti reagenti:

	NucleoMag [®] DNA Forensic	
REF	1 x 96 preps 744660.1	4 x 96 preps 744660.4
NucleoMag [®] F-Beads	1.4 mL	2 x 2.6 mL
Lysis Buffer FOL	50 mL	250 mL
Binding Buffer FOB	100 mL	300 mL
Wash Buffer FOW1	80 mL	200 mL
Wash Buffer FOW2	25 mL	100 mL
Elution Buffer FOE	13 mL	60 mL
Liquid Proteinase K	2 x 1250 μL	9 mL
Reducing Agent TCEP	14 mg	2 x 14 mg
User manual	1	1

Figura 4.2-1. Reagenti contenuti in ciascun kit NucleoMag® DNA Forensic, con le rispettive quantità, da https://www.mn-net.com/media/pdf/02/1e/fb/Instruction-NucleoMag-DNA-Forensic.pdf

La procedura utilizzata si basa sull'adsorbimento reversibile degli acidi nucleici, come il DNA, a biglie paramagnetiche, in opportune condizioni, fornite dall'utilizzo di un buffer di binding.

Innanzitutto, le cellule raccolte tramite il tampone buccale, devono essere lisate per permettere la fuoriuscita del DNA. La lisi si ottiene tramite incubazione dei campioni, ovvero della parte in cotone del tampone, con la Proteinasi K, a 56°C. Per regolare le condizioni di legame del DNA alle biglie, oltre alle biglie stesse (NucleoMag® F-Beads), viene aggiunto al lisato il buffer di binding (FOB). L'applicazione di un campo magnetico permette, quindi, la separazione del DNA dal resto del lisato. Dopo la separazione, le biglie paramagnetiche vengono lavate per tre volte al fine di rimuovere eventuali contaminanti e sali. A questo scopo, vengono utilizzati due buffers di lavaggio: FOW1 e FOW2. L'etanolo residuo dai precedenti step di lavaggio viene rimosso tramite asciugatura.

Infine, il DNA purificato viene eluito tramite un Elution Buffer (FOE) a basso contenuto di sali. L'eluito può, quindi, essere utilizzato per applicazioni successive.

Il kit può essere utilizzato sia manualmente che tramite l'utilizzo di separatori magnetici automatici. In questo caso, è stato utilizzato l'estrattore automatico KingFisher Ml automatic extractor (ThermoLab Systems).



Figura 4.2-2. Estrattore automatico KingFisher Ml automatic extractor (ThermoLab Systems) da https://assets.thermofisher.com/TFS-Assets/LSG/manuals/D01464.pdf

Si tratta di un estrattore progettato per trasferire e processare, in maniera automatizzata, particelle magnetiche in una serie di provette da 1 ml. Il principio si basa sull'utilizzo di aste magnetiche coperte da punte monouso e



Figura 4.2-3. Pettine con punte monouso da https://assets.thermofisher.com/TFS-Assets/LSG/manuals/D01464.pdf

di strips di provette appositamente progettate per l'inserimento dei magneti. I campioni e i reagenti, incluse le biglie paramagnetiche, vengono fornite direttamente allo strumento, mentre i passaggi del protocollo vengono definiti

dall'utente direttamente sul display dello strumento, prima di avviarlo. Lo strumento è progettato per un massimo di 15 strips di provette compatibili con il pettine magnetico. E' importante che ciascuna strip venga inserita in modo che sia stabile: gli unici movimenti possibili devono essere quelli verticali delle due piattaforme, quella a cui sono ancorate le provette e quella a cui è ancorato il pettine.

Prima di avviare il processo automatizzato di estrazione, vengono forniti allo strumento sia i campioni che i reagenti (Figura 4.2-1) nelle provette di ciascuna strip. Dopodiché, le strips possono essere caricate e il processo avviato, tramite un apposito display.
Il principio su cui si basa l'estrattore è la tecnologia MPP (inverse magnetic particle processing). Anziché muovere i liquidi mediante aspirazione o infusione, sono le particelle magnetiche ad essere spostate da una provetta all'altra contenente ciascun reagente specifico. Le particelle magnetiche sono trasferite con l'aiuto di un pettine magnetico coperto da un cappuccio monouso in plastica (Figura 4.2-3).



Figura 4.2-4. MPP (inverse magnetic particle processing) da https://assets.thermofisher.com/TFS-Assets/LSG/manuals/D01464.pdf

L'estrazione del DNA tramite l'utilizzo di un estrattore automatico dotato di un pettine magnetico, in generale, può essere descritta dividendo l'intero processo in cinque fasi distinte:

- Raccolta delle particelle magnetiche

- Rilascio delle particelle magnetiche
- Lavaggio delle particelle magnetiche
- Incubazione
- Concentrazione

Durante la raccolta delle particelle magnetiche, il pettine magnetico si trova interamente all'interno del puntale in plastica. Pettine magnetico e puntali, insieme, si muovono lentamente, in maniera verticale, nelle provette, in modo che le particelle magnetiche vengano raccolte nelle pareti dei puntali. Le barre magnetiche, ricoperte dal cappuccio in plastica, che hanno raccolto le biglie, possono, quindi, essere tirate fuori dalle provette per essere trasferite a quelle successive.

Dopo la raccolta delle particelle magnetiche, il pettine viene, quindi trasferito, nelle provette successive, contenenti un reagente. Le particelle magnetiche, a questo punto, vengono rilasciate tramite un veloce movimento verticale dei puntali, per molte volte, finché le particelle non si sono mescolate bene con il reagente successivo.

Il lavaggio delle particelle magnetiche è un passaggio frequente e molto importante. Questa fase consiste in una combinazione di processi di raccolta e rilascio all'interno di una provetta contenente la soluzione di lavaggio, fornita dal kit. Per massimizzare l'efficienza del lavaggio, sia il pettine magnetico che la sua copertura in plastica sono progettati in modo da rendere minima l'affinità nei confronti di sostanze liquide.

L'incubazione dell'estratto avviene tramite un movimento del pettine verso il basso e, successivamente, verso l'alto così da mantenere le particelle magnetiche in sospensione durante tutta la durata delle reazioni.

L'ultima fase è un processo di concentrazione che avviene grazie all'utilizzo di provette successive a volume minore rispetto a quella precedente (Figura 4.2-5).



Figura 4.2-5. Step di concentrazione da https://assets.thermofisher.com/TFS- Assets/LSG/manuals/D01464.pdf

Con questa modalità di estrazione, da un tampone buccale, è possibile purificare 1-3 μ g di DNA genomico. A seconda del volume di eluizione scelto, si ottengono concentrazioni di 10-30 ng/ μ L. La concentrazione di ciascun estratto è stata, comunque, determinata successivamente tramite qPCR-RealTime, per poter stabilire con maggior accuratezza, le quantità di reagenti da utilizzare nelle fasi successive.

4.3 Quantificazione del DNA

La determinazione della concentrazione di DNA umano presente in ciascun estratto, prima della sua conversione con bisolfito, è stata effettuata tramite una PCR quantitativa in Real Time. Il sistema utilizzato è stato il Plexor® HY System della Promega, che permette di quantificare, in maniera specifica, sia il DNA autosomico che quello del cromosoma Y. Si tratta di un metodo molto sensibile capace di rilevare anche meno di 6.4 pg di DNA.



Figura 4.3-1. Plexor® HY System (Promega): kit di reagent per qPCR-RealTime da https://ita.promega.com/products/forensic-dna-analysis-ce/human-specific-dnaquantitation/plexor-hy-system/?catNum=DC1001

Si tratta di un sistema che sfrutta la PCR-RealTime per determinare, in maniera simultanea, sia il DNA umano totale che il DNA umano maschile, in un'unica reazione. Il kit è molto utile a scopo forense, perché permette di definire una prima informazione importante per l'identificazione personale. In questo studio, era semplicemente necessario quantificare il DNA umano, per avere una determinazione della concentrazione il più precisa possibile, escludendo eventuali contaminanti. Il kit contiene un controllo interno per la PCR, l'IPC, che permette di testare eventuali risultati falsi negativi che possono insorgere in presenza di inibitori della PCR.

Questo sistema funziona misurando la riduzione di un segnale fluorescente durante l'amplificazione. L'amplificazione di ciascun target avviene grazie all'utilizzo di due primers, uno dei quali contiene sia un tag fluorescente che una base modificata. Man mano che l'amplificazione procede, la fluorescenza si riduce a causa dell'incorporazione sito-specifica di un quencher fluorescente opposto alla base modificata complementare. Il quencher si trova nelle immediate vicinanze di un colorante fluorescente situato all'estremità del primer, con conseguente riduzione del segnale fluorescente (Figura 4.3-2).



Figura 4.3-2. Fenomeno del quenching che determina una riduzione della fluorescenza al procedere della reazione di PCR da https://ita.promega.com/products/forensic-dna-analysis-ce/human-specific-dna-quantitation/plexor-hy-system/?catNum=DC1001#specifications

La fluoresceina del sistema Plexor, è utilizzata per evidenziare la presenza di DNA autosomico. I primers di questo sistema amplificano un target multicopia di 99 bp presente sul cromosoma 17. I dati ottenuti da questa reazione vengono utilizzati per determinare la quantità totale di DNA umano all'interno di un campione. Il colorante CALFluor® Orange 560, invece, è utilizzato per rilevare il DNA associato al cromosoma Y. In questo caso, i primers amplificano un target multicopia di 133 bp sul cromosoma Y. I dati così ottenuti vengono utilizzati per quantificare, invece, il DNA totale appartenente a un soggetto di sesso maschile. Un altro colorante, il CALFluor® Red 610, viene, invece, utilizzato per rilevare il controllo interno della PCR, ovvero l'IPC, che viene aggiunto ad ogni reazione. L'IPC non è altro che una sequenza di DNA aggiunta alla mix di reazione. Il prodotto di questa amplificazione dovrebbe essere un amplicone di 150 paia di basi e la sua presenza serve a monitorare eventuali inibizioni durante la reazione. I dati dei tre canali di amplificazione (fluoresceina, CALFluor® Red 610 e CALFluor® Orange 560) possono essere normalizzati in base ad un ulteriore segnale fornito dal colorante IC5, usato come reference. Lo strumento utilizzato per la reazione (CFX96 Touch Real-Time PCR) è stato precedentemente calibrato.

Per la reazione, sono stati utilizzati 2 µl sia di standard che di campione da quantificare. Ogni campione è stato aggiunto in duplicato ed è stata considerata una media dei due valori ottenuti. Come standard, è stato utilizzato il Plexor® HY Male Genomic DNA, fornito dal kit, in concentrazione di 50ng/µl. Si tratta di una miscela di DNA umano maschile. Insieme ai campioni a concentrazione di DNA ignota, diluizioni seriali del DNA standard sono state amplificate allo stesso modo, e i risultati di queste amplificazioni sono stati utilizzati per generare una curva standard nei canali del DNA autosomico (canale della fluoresceina) e dell'Y (canale del CAL Fluor® Orange 560). In base alla curva standard, è stato possibile determinare la concentrazione dei campioni ignoti. Anche per lo standard, è stata effettuata un'amplificazione in duplicato, per

ogni diluizione. Le diluizioni del DNA standard sono state preparate aggiungendo un buffer TE, nelle quantità indicate in figura (Figura 4.3-3):

Concentration	Volume of DNA	Volume of TE-4 Buffer
50ng/µl	Use undiluted DNA	0µ1
10ng/µl	10µl of undiluted DNA	40µ1
2ng/µl	10µl of 10ng/µl dilution	40µ1
0.4ng/µl	10µl of 2ng/µl dilution	40µ1
0.08ng/µl	10µl of 0.4ng/µl dilution	40µ1
0.016ng/µl	10µl of 0.08ng/µl dilution	40µ1
0.0032ng/µl	10µl of 0.016ng/µl dilution	40µ1

Figura 4.3-3. Diluizioni seriali del DNA utilizzato come standard

A questo punto, si procede con la preparazione della reazione di amplificazione. E' stata preparata una mix contenente acqua, Plexor® HY 2X Master Mix e Plexor® HY 20X Primer/IPC Mix nelle proporzioni indicate nella Figura 5.3-4 e tenendo in considerazione il numero di reazioni:

Considerando di aggiungere 2 μ l di estratto per ciascuna reazione, il volume finale è stato considerato di 20 μ l.

Component	Volume (Per Reaction)
Plexor® HY 2X Master Mix	10µl
Water, Amplification Grade	7µ1
Plexor® HY 20X Primer/IPC Mix	1µl
final volume	18µl

Figura 5.3-4. Preparazione della mix di reazione per l'amplificazione assumendo 2 μ l di estratto per reazione

E' stata, quindi, avviata la reazione di amplificazione impostando il termociclatore con il programma mostrato in figura (Figura 5.3-5):



Figura 5.3-5. Programma di amplificazione impostato sul termociclatore CFX96 Touch Real-Time PCR

La chiave della tecnologia del Plexor è il fenomeno del quenching di un reporter fluorescente dovuto all'incorporazione sito-specifica di una base modificata, il dabcyl-iso-dGTP. Come risultato, il segnale fluorescente derivante decresce man mano che il prodotto della PCR si accumula. Il software dello strumento per la RealTime registra i dati del quenching e fornisce una risposta in termini di Ct (ciclo di threshold) e temperatura di melting.

La presenza dello standard a concentrazioni note permette di definire, quindi, una curva standard, in base alla quale, il software fornisce una risposta in termini di concentrazione del campione ignoto da quantificare.

4.4 Trattamento del DNA con bisolfito

Il sequenziamento del DNA convertito con bisolfito rappresenta il gold standard per l'analisi della metilazione del DNA, dal momento che, non è possibile, tramite il sequenziamento di DNA non trattato, poter distinguere una citosina metilata rispetto ad una non metilata, invece.

La chimica della deaminazione della citosina, ad opera del sodio bisolfito, include tre passaggi (Figura 4.4-1).



Figura 4.4-1. Passaggi schematizzati della conversione chimica del DNA con bisolfito da https://www.jove.com/t/3170/dna-methylation-bisulphite-modification-and-analysis

(1) *Solfonazione*, ovvero l'aggiunta del bisolfito al doppio legame 5-6 della citosina.

(2) *Deaminazione* idrolitica del risultante derivato citosina-bisolfito, a dare il derivato uracile-bisolfito.

(3) *Desolfonazione* alcalina, ovvero la rimozione del gruppo solfonato tramite un trattamento alcalino, per ottenere l'uracile.

Il bisolfito porta, preferenzialmente, alla deaminazione della citosina ad uracile nel DNA a singolo filamento mentre il 5-MeC è refrattario alla deaminazione mediata dal bisolfito. Durante l'amplificazione, l'uracile viene amplificato come la timina mentre i residui di 5-MeC rimangono come citosine, consentendo, quindi, di distinguere i siti CpG metilati rispetto a quelli non metilati proprio per la presenza di un residuo di citosina (C), rispetto ad un residuo di timina (T) riscontrabili in seguito a sequenziamento.

La conversione del DNA con bisolfito consiste in un protocollo ben consolidato che viene ampiamente sfruttato in moltissime analisi della metilazione del DNA. Il kit utilizzato a questo scopo, in questo studio, è stato l'EZ DNA Methylation-Lightning Kit, contenente i seguenti reagenti:

EZ DNA Methylation- Lightning™ Kit	D5030T (10 Rxns.)	D5030 (50 Rxns.)	D5031 (200 Rxns.)	Storage Temperature
Lightning Conversion Reagent*	1 Tube	5 Tubes	20 Tubes	Room Temp.
M-Binding Buffer	7 ml	30 ml	125 ml	Room Temp.
M-Wash Buffer**	6 ml	6 ml	24 ml	Room Temp.
L-Desulphonation Buffer	2 ml	10 ml	40 ml	Room Temp.
M-Elution Buffer	500 µl	1 ml	4 ml	Room Temp.
Zymo-Spin™ IC Columns	10	50	200	Room Temp.
Collection Tubes	10	50	200	Room Temp.
Instruction Manual	1	1	1	-

Si tratta di un sistema molto rapido perché utilizza un unico reagente di conversione, il Lightning Conversion Reagent, che non necessita di nessun tipo di preparazione ma va aggiunto semplicemente al campione di DNA.

I processi di denaturazione del DNA di conversione con bisolfito sono combinati, grazie all'apporto di un aumento di temperatura che facilita una rapida denaturazione. Infine, la desolfonazione e il lavaggio del DNA convertito, vengono effettuati utilizzando un'unica colonnina da eluizione in centrifuga.

Nel dettaglio, il protocollo prevede, innanzitutto, la preparazione del buffer di lavaggio tramite l'aggiunta di 24 ml di etanolo puro ai 6 ml di buffer concentrato M-Wash Buffer. A 20 μ l di estratto (corrispondente a 20-500 ng di DNA, basandosi sui risultati ottenuti tramite Real-Time PCR precedentemente descritta) è stato, quindi, aggiunto direttamente il Lightning Conversion Reagent (130 μ l). La provetta da PCR è stata quindi posizionata sul termociclatore, impostato sui seguenti step:

- 1. 98°C per 8 minuti
- 2. 54°C per 60 minuti

3. Conservazione (opzionale) a 4°C per più di 20 ore

A questo punto, è stata preparata una colonnina Zymo-Spin[™] IC Column per il lavaggio, aggiungendoci 600 µl di M-Binding Buffer. Il campione è stato

44

aggiunto alla colonnina, opportunamente posizionata in una collection tube, che è stata centrifugata a velocità massima. Un secondo lavaggio, con le stesse modalità, è stato effettuato aggiungendo 100 μ l di M-Wash Buffer alla colonnina. Dopo aver effettuato un'ulteriore centrifuga, sempre a velocità massima, sono stati aggiunti 200 μ l di buffer di desolfonazione, e la colonnina è stata lasciata a temperatura ambiente per 15-20 minuti per lasciar agire il buffer. Dopodiché, è stata effettuata una breve centrifuga a velocità massima.

A questo punto, prima dell'eluizione finale, è stato effettuato un ultimo lavaggio, utilizzando ancora una volta il M-Wash Buffer (200 μ l) e centrifugando per circa 30 secondi alla velocità massima.

La colonnina, pronta per l'eluizione, è stata quindi posta in una provetta da 1.5 ml, alla quale sono stati aggiunti 10 μ l di buffer di eluizione direttamente a livello della matrice della colonnina stessa.

Dopo una breve centrifuga di circa 30 secondi, a velocità massima, il DNA è pronto per le analisi successive, in quanto eluito.

Template:	A: B:	5'-GACCGTTCCAGGTCCAGCAGTGCGCT-3' 3'-CTGGCAAGGTCCAGGTCGTCACGCGA-5'
Bisulfite Converted:	A:	5'-GATCGTTTTAGGTTTAGTAGTGCGTT-3'
	B:	3' -TTGGCAAGGTTTAGGTTGTTATGCGA-5'

Figura 4.4-3. L'immagine illustra cosa accade ad un campione di DNA durante la conversione con bisolfito. Le citosine metilate sono sottolineate. In seguito alla conversione, i filamenti non sono più complementari. Da https://files.zymoresearch.com/protocols/_d5030t_d5030_d5031_ez_dna_methylation-lightning_kit.pdf

4.5 Quantificazione del DNA convertito con bisolfito

Prima di procedere con l'amplificazione del DNA convertito con bisolfito, è stata effettuata una seconda quantizzazione per poter partire da un'opportuna, e nota quantità di DNA in proporzione ai reagenti della PCR.

A questo scopo, è stato utilizzato il fluorimetro Qubit® 2.0 Fluorimeter (ThermoFisher Scientific) per la lettura di campioni preparati con il kit Qubit® 2.0 Fluorimeter (ThermoFisher Scientific).

Il kit include un reagente specifico per il DNA a doppio filamento, capace di emanare fluorescenza, una volta legatosi, un buffer di diluizione e due DNA standard a concentrazione nota pre-diluiti.

Il reagente è stato diluito con il buffer provvisto dal kit e, ad esso, sono stati aggiunti 2 μ L di ciascun campione. I due standard sono stati preparati allo stesso modo.

Prima di leggere misurare la concentrazione dei campioni sul fluorimetro, occorre calibrarlo inserendo i due standard a concentrazione nota. Questo permette di generare una retta in base alla quale lo strumento fornisce una risposta in termini di concentrazione di campioni sconosciuti.

Il fluorimetro fornisce valori in ng/ml. Il valore fornito corrisponde alla concentrazione di DNA dopo che il campione è stato diluito per la preparazione

alla lettura. Per calcolare la concentrazione del campione, è stata utilizzata la seguente equazione:

Concentration of your sample = QF value
$$\times \frac{200}{x}$$

dove il valore QF è quello fornito dallo strumento mentre x è il numero di microlitri di campione aggiunto alla provetta in cui è stata effettuata la lettura. L'unità di misura è la stessa del valore fornito dal fluorimetro.



Figura 4.5-1. Il fluorimetro fornisce un valore di concentrazione basato sulla relazione tra i due standard utilizzati per la calibrazione. Il grafico mostra la retta (determinata dai due standard), in base alla quale è possibile attribuire un valore di concentrazione ad un campione ignoto, misurando la fluorescenza emessa. Da https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FQubit_dsDNA_HS_Assay_UG.pdf&title=VXNlciBHdWlkZTogUX ViaXQgZHNETkEgSFMgQXNzYXkgS2l0cw==

4.6 Multiplex PCR dei geni target

La metilazione a livello dei siti CpG dei geni ELOVL2, FHL2 e TRIM59 è stata valutata nel DNA precedentemente convertito con bisolfito, tramite un'amplificazione con PCR multiplex.

Si è scelto di utilizzare le stesse sequenze dei primers utilizzati da Jung et al. (2019) in uno studio analogo (27). Le sequenze, le concentrazioni utilizzate e la lunghezza di ciascun amplicone risultate, sono mostrate nella Tabella 4.6-1.

Gene target	Localizzazione cromosomica (GRCh38)	Sequenza(5'→3')	Conc. (µM)	Lunghezza amplicone (bp)
ELOVL2	chr6: 11044628	Frw GGG GYG TAG GGT AAG TGA G Rv CAA CRA ATA AAT ATT CCT AAA ACT CC	1	187
FHL2	chr2: 105399282	Frw GGG TTT TGG GAG TAT AGT AGT Rv AAA ATA ACC CCC TCC TCC	1.5 1.5	191
TRIM59	chr3: 160450189	Frw TAT GGT ATY GGT GGT TTG GGG GAG A Rv ATA AAA AAC ACT ACR CTC CAC AAC ATA AC	1	148

Per la reazione di amplificazione è stato utilizzato 1 μ l (1-9 ng, in base alla quantizzazione con Qubit) di DNA convertito con bisolfito, fino ad un volume di reazione finale di 20 μ l, tramite l'aggiunta di 2X QIAGEN Multiplex PCR Master Mix (Qiagen), 5X Primer mix (preparati alle concentrazioni mostrate nella tabella precedente) e acqua ultrapura.

4.7 Purificazione enzimatica con ExoSAP-ITTM

Prima del sequenziamento, è necessario rimuovere i primer in eccesso e i nucleotidi non incorporati nella precedente reazione di PCR.

A questo scopo, è stato utilizzato un reagente di purificazione, ExoSAP-IT[™] Express (Applied Biosystems), contenente due enzimi (un'esonucleasi I modificata e una fosfatasi alcalina SAP), in un opportuno buffer.



Figura 4.7-1. Overview del processo di ExoSAP-IT[™] Express (Applied Biosystems), da https://www.thermofisher.com/document-connect/documentconnect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2F75001b.pdf&title=QnJpZWYgUHJvdG9jb2w6IEV4b1NBUC1JV CAmbHQ7ZW0mZ3Q7RXhwcmVzcyZsdDsvZW0mZ3Q7IFBDUiBQcm9kdWN0IENsZWFudXA=

Nel dettaglio, 5 μ L del prodotto della precedente PCR, sono stati incubati con 2 μ L di reagente ExoSAPITTM Express. Dopo un breve vortex, le provette sono state posizionate nel termociclatore dove:

- Una prima incubazione a 37°C per 4 minuti permette al reagente di degradare l'eccesso di primers e nucleotidi
- 2. L'incubazione a 80°C per 1 minuto è necessaria per inattivare il reagente

Una volta purificato, il campione è pronto per il sequenziamento.

4.8 Single-base extension (SBE)

Lo stato di metilazione a livello dei siti CpG presenti all'interno dei prodotti della precedente PCR, è stato determinato tramite una multiplex SBE SNaPshot[™] Assay (Applied Biosystems).

Si tratta di un metodo per determinare l'identità di una base nucleotidica in una posizione specifica lungo un acido nucleico. Nella single-base extension (SBE) un primer oligonucleotidico si appaia alla sua regione complementare lungo il DNA a formare un duplex, in modo che l'estremità 3' del primer sia direttamente adiacente alla base nucleotidica da identificare. Con l'ausilio di una DNA-polimerasi, il primer oligonucleotidico viene esteso enzimaticamente di una singola base in presenza di tutti e quattro i terminatori nucleotidici; il terminatore nucleotidico complementare alla base dello stampo da identificare, viene incorporato ed esso stesso identificato. La presenza di tutti i terminatori evita l'errata incorporazione di nucleotidi non complementari.

Esistono diversi metodi per marcare il terminatore nucleotidico incorporato, al fine di poterlo identificare: la chimica del reagente SNaPshot è basata su un'estensione a singola base con ddNTPs, a partire da primers oligonucleotidici non marcati. I ddNTPs sono marcati in maniera fluorescente. La Figura 4.8-1 mostra un tipico approccio di SBE per la determinazione di un polimorfismo, ma lo stesso principio vale per l'identificazione di basi metilate di siti CpG.



Figura 4.8-1. Single-base extension, da https://www.thermofisher.com/documentconnect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2Fcms_042107.pdf&title=TWFudWFsOiBVc2luZyB0aGUgU05hU HNob3QmcmVnOyBNdWx0aXBsZXggU3lzdGVtIHdpdGggdGhlIFBPUC03JnRyYWRlOyBQb2x5b

WVyIG9uIEFwcGxpZWQgQmlvc3lzdGVtcyAzNzMwLzM3MzB4bCBETkEgQW5hbHl6ZXJzIGFuZ CAzMTMwLzMxMzAgKEVuZ2xpc2ggKQ==

Per questa analisi, è stata effettuata una multiplex, quindi sono stati utilizzati 3 primers singoli (uno per ciascuna regione CpG dei tre geni analizzati). Sia per le sequenze oligonucleotiche, che per le concentrazioni dei primers, ci si è basati sullo studio di Jung et al. del 2019 (27).

Gene target	Sequenza (5'→3')	Conc. (µM)	Lunghezza (nt)	Orientamento
ELOVL2	(T)9GGGAGGAGATTTGTAGGTTTAGT	5.0	32	Forward
FHL2	(T)21GTTTTGGGAGTATAGTAGTTAT	0.5	43	Forward
TRIM59	(T)51CCTCAAAAACCRTCRACCACCRAC	0.1	75	Reverse

Per la reazione di SBE è stata preparata un mix dei primers indicati in tabella, alla quale è stato aggiunto 1 μl di ciascun prodotto purificato della PCR, il reagente 2X SNaPshotTM Multiplex Ready Reaction Mix (Applied Biosystems) ed acqua, per arrivare ad un volume finale di 10 μl.

La reazione di estensione è avvenuta, su termociclatore, alle seguenti condizioni:

- Annealing: 25 cicli a 96°C per 10 secondi
- Denaturazione: 50°C per 5 secondi
- Estensione: 60°C per 30 secondi

I prodotti della reazione di SBE sono stati, quindi, purificati, prima del sequenziamento, utilizzando fosfatasi alcalina intestinale bovina (CIAP), fornita dalla ThermoFisher Scientific, ad una concentrazione di 1U/μl.

4.9 Sequenziamento tramite elettroforesi capillare

Per ricostruire l'ordine delle basi azotate che si susseguono all'interno del tratto di DNA ottenuto tramite SBE, è stata utilizzata una tecnica di sequenziamento, tramite SeqStudio Genetic Analyzer (Applied Biosystems), che si basa sul metodo Sanger.

Il metodo Sanger permette di sintetizzare nuovi filamenti utilizzando ddNTPs marcati. Il sequenziamento è suddiviso in provette, ognuna delle quali contiene il DNA stampo, i primer, la DNA polimerasi, i quattro dNTPs e un solo ddNTPs radioattivo. Terminata la reazione, si avranno numerosi frammenti dotati un'estremità 5' in comune ma di lunghezza diversa poiché ciascun frammento è stato bloccato in una posizione precisa diversa, durante l'allungamento (Figura 4.9-1).



Figura 4.9-1. *Metodo di sequenziamento Sanger, da https://www.microbiologiaitalia.it/didattica/frederick-sanger/*

Per questo studio, è stato utilizzato un sequenziamento automatico che si basa sul metodo Sanger appena descritto ma prevede l'utilizzo di fluorocromi non radioattivi per la marcatura dei diversi ddNTPs. Si tratta di un metodo basato sull'utilizzo dell'elettroforesi capillare per separare i frammenti di lunghezza diversa e rilevarne la fluorescenza.

Il sequenziatore automatico utilizzato è stato il SeqStudio Genetic Analyzer (Applied Biosystems). Si tratta di uno strumento dotato di quattro capillari e, quindi, capace di effettuare un sequenziamento tramite un'elettroforesi capillare il cui mezzo di separazione è un polimero direttamente incluso nel capillare in silice.

La soluzione contenente l'amplificato, precedentemente purificata, viene posta a contatto con l'anodo del capillare, a livello del quale, lo strumento applica un potenziale di elettricità. Le molecole di DNA contenute all'interno del campione, sotto l'effetto del campo applicato, migrano a velocità diverse lungo il capillare, a seconda delle loro dimensioni, fino al raggiungimento della "finestra del capillare". In questa zona trasparente, i frammenti di DNA marcati vengono colpiti da una radiazione luminosa emessa da un laser ad argon, emettendo una fluorescenza, rilevata da un sensore. Il sensore è capace di distinguere la fluorescenza emessa in base a quattro differenti colori, ognuno associato ad una singola base nucleotidica. La preparazione dei campioni al sequenziamento è avvenuta direttamente nell'apposita piastra da 96 pozzetti dello strumento. 1 µl di prodotto purificato della precedente reazione di SNaPShot è stato aggiunto a 13.75 µl di formammide Hi-Di (Sigma-Aldrich) e 0.25 µl di GeneScan[™] 120 LIZ[™] Size Standard (Applied Biosystems), lo standard per la determinazione delle dimensioni dei frammenti in corsa elettroforetica.

Prima di andare incontro ad elettroforesi, i campioni sono stati denaturati ponendoli a 95°C per 3 minuti e, successivamente, ponendoli rapidamente in piastra ghiacciata (a circa -20°C) per circa 3 minuti.

Per la corsa è stato utilizzato il modulo E5 ed il set di matrici DS-02 (dR110, dR6G, dTAMRATM, dROXTM, LIZ® Dyes). Si tratta di oligonucleotidi marcati con coloranti, tipicamente utilizzati per generare la matrice richiesta dal sequenziatore per campioni derivati da SNaPshotTM Multiplex System.

Al termine del sequenziamento, un elettroferogramma mostra i risultati ottenuti. Ciascun picco distinguibile di un elettroferogramma mostra una base rilevata durante il passaggio nel capillare in base alla remissione dei fotoni di DNA marcato in seguito a sollecitazione con laser.

L'altezza di ogni picco è proporzionale al numero di molecole colpite dal laser. Grazie a software come GeneMapper® Software v 5.1 (Applied Biosystems) è, infine, possibile individuare la successione delle basi che formano la sequenza.

4.10 Analisi della metilazione

Il livello di metilazione di ciascun sito CpG individuato come target è stato determinato tramite l'analisi, mediante il software GeneMapper® Software v 5.1 (Applied Biosystems), dell'altezza dei picchi corrispondenti alle basi implicate nella metilazione.



Figura 4.10-1. Esempio di un tipico elettroferogramma ottenuto mediante il software GeneMapper®. Ciascuna base è indicata con un colore diverso mentre l'altezza di ogni picco possiede un valore quantitativo

In particolare, il livello di metilazione a livello di ogni sito CpG (valutato con un valore compreso tra 0 e 1) è stato calcolato dividendo l'intensità del nucleotide C (DNA metilato, non convertito dal bisolfito) per la somma delle intensità del nucleotide C e T (DNA non metilato, quindi convertito), dove per intensità si intende l'altezza di ciascun picco. Poiché la metilazione è un fenomeno che interessa entrambi i filamenti di DNA, si è considerata anche la situazione complementare, misurando anche l'intensità del nucleotide G (complementare a C) e dividendola per la somma delle intensità dei nucleotidi G ed A (complementare a T).

Le formule utilizzate sono rappresentate nella seguente figura:



4.11 Analisi statistica dei risultati e costruzione di un modello di previsione

I risultati ottenuti dall'analisi della metilazione dei 3 markers, sono stati elencati in un'unica tabella che permettesse di avere una prima panoramica dell'andamento del livello di metilazione in base all'età, per quanto riguarda tutti i siti CpG analizzati. Da questa, sono stati estrapolati i dati per la costruzione di grafici (uno per ogni gene) che mostrano la relazione tra le due variabili. Sia le tabelle che i grafici sono stati costruiti e analizzati tramite Excel versione 2020.

I dati sono stati, successivamente, divisi in due set: uno per la costruzione di un modello di regressione lineare (training set), uno per testare il modello (test set). La scelta dei componenti di ciascun set è avvenuta casualmente.

Per la costruzione di un modello di previsione, è stata effettuata, tramite il programma Excel e con l'ausilio di software statistici quali R Project e IBM ® Statistics, un'analisi di regressione lineare impostando i dati precedentemente ottenuti come variabili dipendenti ed indipendenti.

Tutte le operazioni, dalla raccolta dei campioni al loro sequenziamento, sono state svolte in collaborazione con l'Istituto di Medicina Legale dell'Università di Torino.

Capitolo cinque

RISULTATI

5.1 Analisi della correlazione tra livello di metilazione ed età

Età	Campio	%	%	%	Età	Campio	%	%	%
	ne	meth	meth	meth		ne	meth	meth	meth
		ELOVL	FHL2	TRIM5			ELOVL	FHL2	TRIM5
22	(DF	2	0 170	9	47	£C10	2	0 227	9
22	TD5	0,280	0,173	0,277	47	fC10	0,292	0,227	0,315
22	TF5	0,258	0,027	0,297	4/	fF10	0,317	0,140	0,387
23	fD6	0,191	0,037	0,324	51	tD11	0,459	0,161	0,372
24	fC5	0,322	0,063	0,293	52	fF11	0,476	0,088	0,532
25	fB3	0,159	0,134	0,244	53	fC12	0,499	0,162	0,440
26	fE5	0,210	0,226	0,229	53	fE11	0,466	0,172	0,360
27	fA3	0,315	0 <i>,</i> 080	0,330	54	fA6	0,512	0,192	0,442
28	fC6	0,204	0,196	0,228	54	fB6	0,470	0,260	0,489
29	fE6	0,338	0,075	0,302	54	fE12	0,247	0,256	0,333
29	fF6	0,257	0,088	0,276	54	fF12	0,368	0,075	0,521
30	fD7	0,309	0,077	0,431	56	fD12	0,475	0,221	0,384
31	fC7	0,191	0,177	0,281	61	fD13	0,633	0,171	0,502
31	fE7	0,319	0,176	0,301	62	fB7	0,463	0,232	0,482
32	fA4	0,354	0,155	0,292	63	fC13	0,605	0,140	0,561
34	fF7	0,454	0,085	0,364	63	fF14	0,374	0,172	0,557
35	fB4	0,390	0,175	0,369	65	fC14	0,474	0,245	0,480
35	fC8	0,258	0,092	0,301	67	fE13	0,357	0,368	0,334
38	fE8	0,324	0,138	0,457	68	fA7	0,495	0,185	0,464
38	fF8	0,406	0,259	0,225	68	fD14	0,543	0,188	0,491
41	fA5	0,435	0,094	0,399	71	fC15	0,713	0,104	0,621
41	fE9	0,410	0,097	0,358	74	fE15	0,712	0,042	0,731
41	fF9	0,345	0,073	0,373	77	fB8	0,691	0,164	0,549
42	fD9	0,308	0,142	0,381	77	fD15	0,518	0,410	0,447
43	fC9	0,389	0,066	0,448	80	fC16	0,482	0,052	0,565
43	fE10	0,296	0,130	0,381	83	fA8	0,686	0,144	0,713
44	fB5	0,315	0,224	0,348	84	fD16	0,538	0,057	0,491
45	fD10	0,458	0,125	0,415	88	fF16	0,632	0,137	0,549
					95	fE16	0,588	0,200	0,682

Tabella 5.1-1. Dati ottenuti dall'analisi della metilazione dei tre siti CpG utilizzati come markers

Tutti i campioni analizzati sono stati opportunamente codificati mediante un codice univoco e ordinati in maniera crescente in base all'età cronologica. Come illustrato in tabella 5.1-1, ciascun campione è stato descritto in termini di livello di metilazione, mediante i valori ottenuti nell'analisi precedente, per ogni marker analizzato.

Sono stati, quindi, costruiti dei grafici mettendo in relazione l'età cronologica di ciascun individuo con il livello di metilazione ad essa associato, per poter avere una prima panoramica dell'andamento della seconda variabile (% di metilazione) in base alla prima (età). E' stato costruito un grafico di dispersione per ogni marker considerato.



Figura 5.1-1. Il grafico mette in relazione tutti i dati relativi al livello di metilazione del marker ELOVL2 con l'età di ciascun individuo a cui il campione considerato appartiene.

Da una prima analisi, risulta evidente l'andamento crescente della dispersione: all'aumentare dell'età cronologica di ciascun campione, aumenta, in maniera progressiva, anche la percentuale di metilazione a livello del sito CpG di ELOVL2. La retta tratteggiata mostra, appunto, questa tendenza, in maniera lineare, tracciando un primo modello approssimato di quello che potrebbe essere il modello di regressione lineare associato a questo marker.

Il quadrato del coefficiente di correlazione (R quadro) tra le due variabili si attesta intorno a 0.7, un valore che attesta un buon indice di correlazione per la costruzione di un modello di regressione lineare.

Questa prima analisi conferma i risultati ottenuti da Jung et al. (27) in uno studio analogo (Figura 5.1-2).



Figura 5.1-2. Relazione tra età cronologia e percentuale di metilazione del marker ELOVL2. In verde, i dati relativi a tamponi buccali. Da "DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples" di Jung et al.

Lo studio di Jung et al. mostrava un valore di r = 0.684, per quanto riguarda i campioni rappresentati da tamponi buccali. In questo studio, si osserva, invece,

un valore di r = 0.814, che mostra un miglior indice di correlazione, probabilmente dovuto al minor numero di campioni analizzati.

Lo stesso tipo di analisi preliminare della correlazione tra le variabili (età e livello di metilazione) è stato effettuato per gli altri due markers presi in considerazione: FHL2 e TRIM59. I seguenti grafici mostrano i risultati ottenuti:



Figura 5.1-3. Grafico a dispersione relativo al marker FHL2

Per quanto riguarda il marker FHL2, è possibile osservare un aumento di metilazione corrispondente ad un aumento di età, minimo. La linea di tendenza mostra solo un leggero aumento di pendenza. I valori mostrano, quindi, una dispersione piuttosto marcata e un basso indice di correlazione, riscontrabile dal valore di R quadro = 0.0687 e di r = 0.26.

Anche in questo caso, l'analisi rispecchia i risultati ottenuti da Jung et al. (Figura 5.1-4).



Figura 5.1-4. Relazione tra età cronologica e percentuale di metilazione del marker FHL2. In verde, i dati relativi a tamponi buccali. Da "DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples" di Jung et al.

Nello studio di Jung et al. i valori relativi al marker FHL2 erano quelli che mostravano un livello di dispersione maggiore, confermato da un valore di r molto basso, per quanto riguarda campioni di tamponi buccali. Per quanto riguarda campioni di sangue, invece, l'indice di correlazione aumentava moltissimo suggerendo che il livello di metilazione potesse essere correlato o meno all'età a seconda del tipo di tessuto. Risultati analoghi sono stati ottenuti in questo studio, fornendo un'ulteriore conferma all'ipotesi secondo la quale il grado di metilazione di determinati siti CpG sia altamente tessuto-specifico. L'ultimo marker analizzato è stato TRIM59 che, da una prima analisi (Figura 5.1-5) sembra essere quello più promettente per la costruzione di un modello valido per la predizione dell'età biologica a partire da campioni ottenuti tramite la raccolta di tamponi buccali.



Figura 5.1-5. Grafico a dispersione relativo al marker TRIM59

La linea di tendenza mostra un andamento simile a quello osservato nel grafico relativo al marker ELOVL2 (Figura 5.1-1). Anche l'indice di correlazione, è molto simile a quello riscontrato nell'analisi del primo marker: in questo caso osserviamo un valore di R quadro di 0.6833, leggermente superiore a quello osservato per ELOVL2, così come il valore r = 0.827.

Dal grafico è possibile osservare che gli outliers rispetto alla linea di tendenza sembrano aumentare all'aumentare dell'età. Questo suggerisce che una stima dell'età di individuo di 20-60 anni possa essere più accurata rispetto a quella di individui di età maggiore di 60 anni.

Anche i risultati mostrati nel grafico a dispersione precedente (Figura 5.1-5) sono stati confrontati con gli stessi risultati, relativi a tamponi buccali, raggiunti da Jung et al.:



TRIM59

Figura 5.1-6. Relazione tra età cronologica e percentuale di metilazione del marker TRIM59. In verde, i dati relativi a tamponi buccali. Da "DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples" di Jung et al.

Anche dallo studio di Jung et al., risulta evidente un'elevata correlazione tra età cronologica e livello di metilazione, a livello del marker TRIM59 con valori di r, per quanto riguarda i tamponi buccali, paragonabili a quelli ottenuti da campioni di sangue.

Gene	Localizzazione Cromosomica	r	<i>R</i> ²
ELOVL2	chr6:11044628	0.814	0.6628
FHL2	chr2:105399282	0.262	0.0687
TRIM59	chr3:160450189	0.827	0.6833

Tabella 5.1-2. Indici di correlazione relativi ai dati ottenuti per i tre markers analizzati

Mentre a livello dei markers ELOVL2 e TRIM59 è osservabile una forte correlazione tra livello di metilazione ed età (r > 0.7), a livello del sito CpG di FHL2, numerosi tamponi buccali mostrano un'ipometilazione riscontrabile in un debole livello di correlazione tra le due variabili (r < 0.3).

Probabilmente la metilazione a livello di questo sito è altamente tessutospecifica in quanto, a livello delle cellule prelevate tramite tampone buccale, non si osserva un elevato grado di metilazione. Ulteriori studi su questo marker andrebbero effettuati utilizzando campioni di origine diversa.

5.2 Costruzione di un modello di regressione lineare per la stima dell'età biologica a partire da dati relativi alla metilazione

Già da una prima analisi, tramite la semplice costruzione di una linea di tendenza su ciascun grafico a dispersione, risultava evidente la forte correlazione tra i livelli di metilazione e l'età, per quanto riguarda i due markers ELOVL2 e TRIM59. Comunque, nonostante l'analisi precedente abbia mostrato una debole correlazione tra le due variabili, a livello del gene FHL2,

i dati relativi ad essi sono stati considerati ugualmente significativi per la

costruzione di un modello di predizione.

I 55 campioni sono stati suddivisi, casualmente, in due gruppi: un "training set" per la costruzione del modello di regressione lineare, e un "test set" per verificare l'accuratezza del modello costruito.

TRAINI	NG SET			
Età	Campione	%Met ELOVL2	%Met FHL2	%Met TRIM59
22	fD5	0,28	0,173	0,277
23	fD6	0,191	0,037	0,324
25	fB3	0,159	0,134	0,244
27	fA3	0,315	0,08	0,33
29	fE6	0,338	0,075	0,302
30	fD7	0,309	0,077	0,431
31	fE7	0,319	0,176	0,301
34	fF7	0,454	0,085	0,364
35	fC8	0,258	0,092	0,301
38	fF8	0,406	0,259	0,225
41	fE9	0,41	0,097	0,358
42	fD9	0,308	0,142	0,381
43	fE10	0,296	0,13	0,381
45	fD10	0,458	0,125	0,415
47	fF10	0,317	0,14	0,387
52	fF11	0,476	0,088	0,532
53	fE11	0,466	0,172	0,36
54	fB6	0,47	0,26	0,489
54	fF12	0,368	0,075	0,521
61	fD13	0,633	0,171	0,502
63	fC13	0,605	0,14	0,561
65	fC14	0,474	0,245	0,48
68	fA7	0,495	0,185	0,464
71	fC15	0,713	0,104	0,621
77	fB8	0,691	0,164	0,549
80	fC16	0,482	0,052	0,565
84	fD16	0,538	0,057	0,491
95	fE16	0,588	0,2	0,682

 Tabella 5.2-1. Set di dati utilizzati per la costruzione del modello di regressione lineare

Un'analisi di regressione multipla di questi dati permette di costruire un modello per poter prevedere la variabile "Y" dell'equazione della retta ottenuta, ovvero l'età stimata. Inserendo i valori del "training set" come input e ponendo i valori relativi all'età come dipendenti dai valori relativi al livello di metilazione dei tre markers, è stato ottenuto un modello, con le seguenti caratteristiche statistiche.

Statistica della r	egressione
R multiplo	0,899
R al quadrato	0,808201
R al quadrato	
corretto	0,784226
Errore standard	9,312934
Osservazioni	28

Tabella 5.2-2. Valori indicatori della bontà del modello statistico costruito

Il valore di R multiplo, prossimo all'1, indica una forte relazione lineare tra le due variabili. Anche il valore di R², coefficiente di determinazione, è di poco minore di 1. Questo valore, utilizzato come indicatore della bontà dell'adattamento, mostra il grado di dispersione rispetto alla retta di regressione. In questo caso, il valore di R al quadrato è 0,808201. In altre parole, l'81% delle variabili indipendenti (X, percentuali di metilazione) sono spiegate da variabili indipendenti (Y, età).

Il valore relativo all'errore standard risulta piuttosto alto (9,312934). Poiché questo valore mostra la distanza media attorno alla retta di regressione, si può ipotizzare (considerata anche l'analisi su grafico a dispersione precedente) che
questa misura assoluta così alta sia dovuta principalmente ai dati relativi al marker FHL2, in cui è evidente un elevato grado di dispersione, con una bassa linearità dei valori.

Il numero di osservazioni, pari a 28, indica semplicemente il numero di campioni in base ai quali il modello di previsione è stato costruito.

Per ciascun marker è stato, quindi, determinato un modello matematico che potesse fornire un output in termini di età stimata, fornendo un input relativo alla percentuale di metilazione di quel determinato marker.



Figura 5.2-1. Modello di regressione lineare relativo al marker ELOVL2 con modello matematico corrispondente per la predizione dell'età biologica a partire da dati sulla percentuale di metilazione



Figura 5.2-2. Modello di regressione lineare relativo al marker FHL2 con modello matematico corrispondente per la predizione dell'età biologica a partire da dati sulla percentuale di metilazione



Figura 5.2-3. Modello di regressione lineare relativo al marker FHL2 con modello matematico corrispondente per la predizione dell'età biologica a partire da dati sulla percentuale di metilazione

5.3 Predizione dell'età biologica utilizzando il modello di regressione lineare costruito

In base ai modelli matematici determinati analizzando statisticamente i dati dei campioni costituenti il "training set", è stata, quindi, stimata l'età biologica di individui ad età nota, utilizzando unicamente i dati relativi al livello di metilazione dei tre markers ELOVL2, FHL2 e TRIM59. I campioni utilizzati per testare i modelli di regressione sono i seguenti (Tabella 5.3-1).

TEST SET						
		% Met		% Met	% Met	
Età	Campione	ELOVL2		FHL2	TRIM59	
22	fF5		0,258	0,027		0,297
24	fC5		0,322	0,063		0,293
26	fE5		0,21	0,226		0,229
28	fC6		0,204	0,196		0,228
29	fF6		0,257	0,088		0,276
31	fC7		0,191	0,177		0,281
32	fA4		0,354	0,155		0,292
35	fB4		0,39	0,175		0,369
38	fE8		0,324	0,138		0,457
41	fA5		0,435	0,094		0,399
41	fF9		0,345	0,073		0,373
43	fC9		0,389	0,066		0,448
44	fB5		0,315	0,224		0,348
47	fC10		0,292	0,227		0,315
51	fD11		0,459	0,161		0,372
53	fC12		0,499	0,162		0,44
54	fA6		0,512	0,192		0,442
54	fE12		0,247	0,256		0,333
56	fD12		0,475	0,221		0,384
62	fB7		0,463	0,232		0,482
63	fF14		0,374	0,172		0,557
67	fE13		0,357	0,368		0,334
68	fD14		0,543	0,188		0,491
74	fE15		0,712	0,042		0,731
77	fD15		0,518	0,41		0,447
83	fA8		0,686	0,144		0,713
88	fF16		0,632	0,137		0,549
Tabella 5.3-1. Set di dati utilizzati per testare i modelli di regressione linear						

71

Per ogni marker, i valori di metilazione ottenuti dal sequenziamento, sono stati sostituiti in ciascuna equazione che descrive il modello di predizione corrispondente al gene analizzato, ottenendo così un valore di età stimata che è stato, successivamente, confrontato con l'età cronologica nota di ciascun individuo.

	% Met	Età	Età		
Campione	ELOVL2	cronologica	stimata	Differenza	
fF5	0,258	22	30,88	8,88	
fC5	0,322	24	38,19	14,19	
fE5	0,21	26	25,4	0,6	
fC6	0,204	28	24,71	3,29	
fF6	0,257	29	30,76	1,76	
fC7	0,191	31	23,23	7,77	
fA4	0,354	32	41,84	9,84	
fB4	0,39	35	45,95	10,95	
fE8	0,324	38	38,41	0,41	
fA5	0,435	41	51,09	10,09	
fF9	0,345	41	40,81	0,19	
fC9	0,389	43	45,83	2,83	
fB5	0,315	44	37,39	6,61	
fC10	0,292	47	34,76	12,24	
fD11	0,459	51	53,83	2,83	
fC12	0,499	53	58,39	5,39	
fA6	0,512	54	59,88	5,88	
fE12	0,247	54	29,62	24,38	
fD12	0,475	56	55 <i>,</i> 65	0,35	
fB7	0,463	62	54,28	7,72	
fF14	0,374	63	44,12	18,88	
fE13	0,357	67	42,18	24,82	
fD14	0,543	68	63,42	4,58	
fE15	0,712	74	82,71	8,71	
fD15	0,518	77	60,56	16,44	
fA8	0,686	83	79,74	3,26	
fF16	0,632	88	73,58	14,42	

Tabella 5.3-2. Età stimata in base ai dati sul livello di metilazione del marker ELOVL2

In base ai valori dell'ultima colonna (Tabella 5.3-2), ottenuti confrontando l'età

stimata con quella reale di ciascun campione, è stata determinata l'accuratezza

del modello, in termini di MAD (Deviazione Mediana Assoluta), per avere una minore influenza dei valori estremi ("outliers"). In base alla mediana (6,61 anni) è stato, quindi, calcolata la MAD, pari a 4,34 anni.

Il risultato ottenuto è coerente con quello ottenuto da Jung et al. (27), il cui modello per la predizione dell'età biologica da tamponi buccali mostrava una MAD di 3,823.

L'età biologica degli stessi campioni è stata determinata anche utilizzando gli altri due modelli di regressione, basandosi sui valori ottenuti dall'analisi dei markers FHL2 e TRIM59.

Rispetto a modello basato sulla metilazione di ELOVL2, quello relativo al marker FHL2 mostra risultati meno incoraggianti. Già dalla prima analisi di regressione, risultava, infatti, evidente l'elevata dispersione dei valori che non mostrava un andamento lineare e a metilazione crescente in base all'età. Il valore di R quadro = 0.0687 (Figura 5.1-3), infatti, indicava una debole correlazione tra le due variabili: livelli di metilazione di FHL2 ed età.

Ciò che ci si aspettava era, quindi, un modello di predizione dell'età biologica non particolarmente accurato, probabilmente a causa dell'elevata tessutospecificità della metilazione del sito CpG di FHL2.

Testando il modello, tali aspettative sono state confermate (Tabella 5.3-3).

	%			
	Met	Età	Età	
Campione	FHL2	cronologica	stimata	Differenza
fF5	0,027	22	43,32	21,32
fC5	0,063	24	45,45	21,45
fE5	0,226	26	55,08	29,08
fC6	0,196	28	53,31	25,31
fF6	0,088	29	46,92	17,92
fC7	0,177	31	52,18	21,18
fA4	0,155	32	50,88	18,88
fB4	0,175	35	52,07	17,07
fE8	0,138	38	49,88	11,88
fA5	0,094	41	47,28	6,28
fF9	0,073	41	46,04	5,04
fC9	0,066	43	45,62	2,62
fB5	0,224	44	54,96	10,96
fC10	0,227	47	55,14	8,14
fD11	0,161	51	51,24	0,24
fC12	0,162	53	51,3	1,7
fA6	0,192	54	53,1	0,9
fE12	0,256	54	56,85	2,85
fD12	0,221	56	54,78	1,22
fB7	0,232	62	55,43	6,57
fF14	0,172	63	51,89	11,11
fE13	0,368	67	63,47	3,53
fD14	0,188	68	52 <i>,</i> 83	15,17
fE15	0,042	74	44,21	29,79
fD15	0,41	77	65,96	11,04
fA8	0,144	83	50,23	32,77
fF16	0,137	88	49,82	38,18

 Tabella 5.3-3. Età stimata in base ai dati sul livello di metilazione del marker FHL2

Anche di questo modello è stata valutata l'accuratezza, calcolando la Deviazione Mediana Assoluta (MAD) = 8,26 anni. Chiaramente, un valore così alto, denota una bassa accuratezza di predizione dell'età biologica utilizzando il modello basato sull'analisi del sito CpG di FHL2.

Risultati decisamente migliori, sono stati ottenuti testando l'ultimo modello di regressione lineare ottenuto, quello basato sui dati di metilazione relativi al marker TRIM59. Già in studi precedenti, era stato reso noto che questo marker fosse il più promettente per questo tipo di predizione su tessuti diversi tra loro, in quanto, probabilmente, il suo livello di metilazione è meno influenzato dal tipo cellulare e, quindi, dal tipo di campione.

	% Met	Età	Età	
Campione	TRIM59	cronologica	stimata	Differenza
fF5	0,297	22	31,06	9,06
fC5	0,293	24	30,47	6,47
fE5	0,229	26	21,03	4,97
fC6	0,228	28	20,88	7,12
fF6	0,276	29	27,96	1,04
fC7	0,281	31	28,7	2,93
fA4	0,292	32	30,32	1,68
fB4	0,369	35	41,67	6,67
fE8	0,457	38	54,65	16,65
fA5	0,399	41	46,1	5,1
fF9	0,373	41	42,26	1,26
fC9	0,448	43	53,33	10,33
fB5	0,348	44	38,58	5,42
fC10	0,315	47	33,71	13,29
fD11	0,372	51	42,12	8,88
fC12	0,44	53	52,15	0,85
fA6	0,442	54	52,44	1,56
fE12	0,333	54	36,37	17,63
fD12	0,384	56	43,89	12,11
fB7	0,482	62	58,34	3,66
fF14	0,557	63	69,4	6,4
fE13	0,334	67	36,51	30,49
fD14	0,491	68	59,67	8,33
fE15	0,731	74	95 <i>,</i> 06	21,06
fD15	0,447	77	53,18	23,82
fA8	0,713	83	92,41	9,41
fF16	0,549	88	68,22	19,78

 Tabella 5.3-4. Età stimata in base ai dati sul livello di metilazione del marker TRIM59

Da un confronto tra i valori di età stimata e quelli reali dei campioni, è possibile, anche in questo caso, determinare l'accuratezza del modello, in base ad un valore di MAD pari a 4,19 anni, il più basso tra i tre ottenuti.

Capitolo sei

CONCLUSIONI E PROSPETTIVE FUTURE

La parte della genetica forense nota come "DNA-fingerprinting", si è sviluppata a partire da tecniche basate sulle proteine, un quarto di secolo fa, passando, successivamente, a tecniche basate sullo studio dei polimorfismi in base alla lunghezza di frammenti di restrizione (RFLPs). L'amplificazione di piccole ripetizioni in tandem (STR), utilizzando la reazione a catena della polimerasi (PCR), presto, ha preso il posto dell'analisi dei RFLP, diventando il "gold standard" nel campo dell'identificazione genetica. Nel frattempo, sono state sviluppate e rese commercialmente disponibili numerose multiplex STR che permettono di amplificare simultaneamente più di 30 STR da un minimo di 15 cellule, o addirittura meno.

L'enorme contenuto di informazioni proveniente dalla grandissima varietà di genotipi STR osservati consente, quindi, l'individualizzazione genetica (ad eccezione dei gemelli omozigoti).

Alcuni loci STR principali, accuratamente selezionati, oggi costituiscono le basi di database a DNA, grazie ai quali, spesso, è possibile collegare scene del crimine irrisolte e criminali tramite profili STR corrispondenti. Tuttavia, il successo del moderno DNA-fingerprinting dipende strettamente dalla disponibilità del materiale di riferimento da sospettati.

Al fine di fornire nuovi indizi investigativi nei casi in cui tali campioni di riferimento siano assenti, gli scienziati forensi hanno iniziato ad esplorare il campo del phenotyping, ovvero la possibilità di prevedere tratti fenotipici di un soggetto partendo da un campione di DNA. Questo nuovo paradigma attualmente è basato principalmente sull'analisi di marcatori epigenetici che permettano di prevedere caratteristiche utili al lavoro investigativo.

Finora, le caratteristiche visibili esternamente meglio studiate sono il colore degli occhi, dei capelli e della pelle, così come l'etnia ed, infine, l'età.

La possibilità di avere informazioni relative all'età biologica di un soggetto donatore è un elemento essenziale in campo forense. Questo aspetto è stato, ed è tutt'ora, quindi, ampiamente studiato. Tra i diversi approcci metodologici ad oggi testati, l'analisi sensibile alla metilazione di markers del DNA accuratamente selezionati (siti CpG) ha portato a risultati molto promettenti fornendo livelli di precisione nella previsione dell'età di \pm 3-4 anni, valori paragonabili o addirittura superiori a quelli relativi a testimonianze oculari.

La stima dell'età biologica del donatore ignoto di un campione di DNA gioca un ruolo importantissimo nel campo delle investigazioni forensi, in quanto, oltre a fornire spunti investigativi nella ricerca di un autore sconosciuto di un crimine, è essenziale per l'identificazione umana da resti scheletrici. Questo tipo di analisi, inoltre, può essere critico nei casi di immigrazione in cui sia l'identità che l'età degli individui non è chiara.

Lo studio della metilazione per stimare l'età cronologica basato sull'analisi del DNA convertito con bisolfito, attualmente, sembra essere la tecnica più promettente, pur richiedendo quantità di DNA nettamente più elevate rispetto a quelle normalmente disponibili in campioni tipici della medicina legale. Tuttavia, la ricerca di siti CpG in base ai quali costruire dei modelli statistici di predizione dell'età biologica è continuata, portando alla scoperta di alcuni di essi il cui grado di metilazione risulta essere strettamente correlato all'età.

Due dei tre markers analizzati in questo studio (ELOVL2 e TRIM59) hanno mostrato un elevato grado di correlazione tra i livelli di metilazione riscontrati e l'età di ciascun soggetto, permettendo di costruire dei modelli accurati, anche partendo da campioni rappresentati da tamponi buccali, i quali da studi precedenti, sembravano essere quelli la cui analisi risultava meno precisa. ELOVL2 e TRIM59 si confermano, quindi, come ottimi candidati a costituire eventuali core-markers per la predizione dell'età da campioni di origine diversa.

Al contrario, la metilazione del marker FHL2, rispetto a studi precedenti basati su campioni di sangue e di saliva, mostra un grado di correlazione con l'età

78

molto più basso: il modello di predizione costruito in base ai dati sulla metilazione di questo sito ha un'accuratezza quasi doppiamente peggiore rispetto ai due precedentemente descritti. Considerando studi precedenti che dimostrano l'elevata tessuto-specificità di siti diversi alla metilazione, è lecito supporre che FHL2, con l'invecchiamento, subisca un certo effetto epigenetico diverso a seconda del tipo cellulare. Questo spiegherebbe il diverso grado di correlazione tra metilazione ed età in campioni costituiti da percentuali cellulari diverse come sangue e cellule della mucosa orale prelevate mediante un tampone.

In un campo come la genetica forense in cui la standardizzazione delle tecniche e dei parametri utilizzati ne rappresenta la chiave, la selezione di core-markers per l'analisi della metilazione è un aspetto fondamentale. La selezione dei marcatori principali richiede ulteriori ricerche per risolvere questioni ancora aperte. A tale riguardo, non è ancora chiaro (ed è proprio l'aspetto su cui questo studio si è focalizzato) il diverso comportamento dei marcatori di metilazione in tessuti di rilevanza forense differenti come sangue, cellule epiteliali, sperma, capelli, ossa e denti; come già accennato, è noto che tessuti e cellule diversi dimostrino differenti pattern di metilazione.

Inoltre, sono necessari studi dettagliati per stabilire se i modelli di metilazione rimangano stabili per periodi di tempo più o meno lunghi. Questo aspetto è

79

molto importante nell'ambito della medicina legale in cui, spesso, i campioni riscontrati mostrano diversi gradi di decomposizione dovuti a tempi e condizioni variabili prima della loro raccolta. Numerosi studi di antropologia forense si occupano dell'identificazione, tramite analisi genetica, di resti umani risalenti al Medioevo. Tessuti solidi risalenti a periodi così antichi sono noti per mostrare determinati pattern di danno, utili a confermare l'autenticità del risultato genetico ma che possono porre difficoltà per l'analisi della metilazione, in quanto i processi di deaminazione agiscono naturalmente in questo genere di campioni antichi.

Anche il sesso e range definiti di età sono variabili che possono influenzare la precisione della previsione. Mentre per quanto riguarda il sesso ci sono numerose contraddizioni, la maggior parte dei ricercatori è d'accordo sulla possibilità di avere stime migliori per le fasce di età più giovani (neonati e adolescenti), che sono proprio le più critiche nelle decisioni legali.

L'etnia è un altro aspetto importante. Finora, la maggior parte della ricerca si è concentrata su popolazioni dell'Eurasia occidentale, con pochissime eccezioni. La stima dell'età basata sulla metilazione è un aspetto da indagare su aree geografiche più ampie e rappresentative per definire meglio le caratteristiche e i limiti dei set di dati ottenuti. Popolazioni diverse hanno, infatti, differenti condizioni sanitarie, stili di vita ed esposizione alle condizioni climatiche e a fattori ambientali, tutte condizioni che possono influenzare le stime e che devono essere tenute in considerazione nell'analisi. Studi recenti mostrano che gli stessi markers, analizzati in popolazioni differenti, mostrano delle leggere differenze. Questo conferma che i laboratori di genetica forense hanno bisogno di valutare e validare individualmente dei metodi con specifici basati sulla popolazione rilevante in una certa area in cui operano.

Nonostante queste criticità, si può concludere che l'attenzione della genetica forense nella ricerca e nella validazione di markers che mostrano una forte, robusta e riproducibile associazione con l'età biologica sta ottenendo numerosi risultati incoraggianti che sfociano inevitabilmente in un altro fondamentale campo della medicina, la ricerca sui processi di invecchiamento.

Ringraziamenti

Ringrazio il mio relatore, il prof. Mauro Pesaresi per avermi seguito durante l'ultima fase del mio percorso universitario nonché prima fase del mio percorso professionale.

Desidero fare un ringraziamento particolare al professor Carlo Robino e alle dottoresse Anna Vittoria Buzzetti ed Elena Chierto, del Dipartimento di Scienze della Sanità Pubblica e Pediatriche dell'Università di Torino, per il loro contributo, fondamentale alla stesura di questa tesi.

Ringrazio le dottoresse Eleonora Ciarimboli, Lucia Fiordelmondo e Filomena Melchionda, letteralmente fondamentali durante tutto l'anno trascorso in laboratorio.

Ringrazio la professoressa Chiara Turchi, il dottor Valerio Onofri, e tutto il reparto di Medicina Legale dell'Ospedale Torrette di Ancona per avermi accolta e guidata con competenza e professionalità rare.

A tutti voi,

Grazie

Angela

Bibliografia

DNA methylation-based age prediction from various tissues. Sang-Eun Jung,
 Kyoung-Jin Shin & Hwan Young Lee. Department of Forensic Medicine,
 Yonsei University College of Medicine, Seoul 03722, Korea : s.n., 2017.

2. The hallmarks of aging. López-Otín C, Blasco MA, Partridge L, Serrano M and. 153, 2013, Vol. Cell.

3. Age Estimation with DNA: From Forensic DNA Fingerprinting to Forensic (Epi)Genomics: A Mini-Reviw. Parson, Walther. Gerontology, Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria; Forensic Science Program, The Pennsylvania State University, University Park, PA, USA : s.n., 2018. DOI: 10.1159/000486239.

4. Epigenetic differences arise during the lifetime of monozygotic twins. M.F.
 Fraga, E. Ballestar, M.F. Paz, S. Ropero, F. Setien, M.L. Ballestar, D. Heine-Suner, J.C. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T.D. Spector,
 Y. Wu, C. Plass. USA : s.n., 2005, Vols. Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 10604–10609.

5. Space/population and time/age in DNA methylation variability in humans; a study on IGF2/H19 locus in different Italian populations and in mono- and dizygotic twins of different age. C. Pirazzini, C. Giulana, M.G. Bacalini, A. Boattini, M. Capri, E. Fontanesi, E. Marasco, V. Mantovani, M. Pierini, E. Pini,
D. Luiselli, C. Franchesci, P. Garagnani. 2012, Vols. Ageing 4 (2012) 509-520.
6. Accuracy of DNA methylation patterns preservation by the DNMT1 methyltransferase. R. Goyal, R. Reinhardt, A. Jeltsch. 2006, Vols. Nucleic Acids Res. 34 (2006) 1182-1188.

7. Epigentic regulation of gene expression; how the genome integrates intrinsic and environmental signals. R. Jaenisch, A. Bird. 2003, Vols. Nat. Genet. 33 (2003) 245-254.

8. The epigenotype. Waddington, C.H. 1942, Vol. Endeavour 1.

Epigenetic Mechanism of Gene Regulation. A.D. Riggs, R. Martienssen, U.
 Russo. 1996, Vols. Cold Spring Harbor Laboratory Press pp 1-4.

 Applying whole-genome studies of epigenetic regulation to study human disease. D. Lieb, S. Beck, M.L. Bulyk, P. Farnham, N. Hattori, S. Henikoff, Z.S. Liu, K. Okumura, K. Shiota, T. Ushijima, J.M. Greally. 2006, Vol. Cytogenet. Genome Res. 114.

11. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Saxonov, S., Berg, P., Brutlag, D.L. USA : s.n., 2006, Vols. Proc. Natl. Acad. Sci. USA 103, 1412-1417.

12. The rate of CpG mutation in Alu repetitive elements within the p53 tumor suppressor gene in the primate germline. Yang, A.S., Gonzalgo, M.L., Zingg,

J.-M., Millar, R.P., Buckley, J.D., Jones, P.A. 1996, Vols. J. Mol. Biol. 258, 240–250.

DNA methylation and healthy human aging. Jones, M.J., Goodman, S.J.,
 Kobor, M.S. 2015, Vols. Aging Cell 14, 924-932.

14. DNA Methylation and Its Basic Function. Lisa D Moore, Thuc Le and Guoping Fan. s.l. : Neuropsychopharmacology Reviews, 2013.

15. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. M. Okano, D. W. Bell, D. A. Haber,
E. Li. Cardiovascular Research Center, Massachusetts General Hospital,
Department of Medicine, Harvard Medical School, Charlestown 02129, USA : s.n., 1999, Vols. Cell 1999 Oct 29;99(3):247-57.

16. Role of DNA methylation in human age prediction. Neelam Goela, Priya Karira, Vivek Kumar Gargb. Department of Information Technology, UIET,
Panjab University, Chandigarh 160014, UT, India: s.n., 2017, Vols.
Mechanisms of Ageing and Development 166 (2017) 33–41.

17. Examination of DNA methylation status of the ELOVL2 marker may be useful for human age prediction in forensic science. Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Makowska, Z., Spas, A., Parys-Proszek, A. s.l.: Forensic Sci. Int. Genet. 14, 161–167. http://dx.doi.org/10.1016/j.fsigen., 2014. Genome-wide methylation profiles reveal quantitative views of human aging rates. Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes. 2014, Vols. Mol. Cell 49, 359–367.

19. Quantification of global mithocondrial DNA methylation levels inverse correlation with age at two CpG sites. Mawlood, K., Dennany, L., Watson, N., Dempster, J., Benjamin, S. Albany, NY : s.n., 2016, Vols. Aging 8, 1-6.

20. Molecular pathology and age estimation. Meissner C, Ritz-Timme S. 2010,Vols. Forensic Sci Int Genet 2010; 18: 33–48.

21. Age estimation in forensic sciences: application of combined aspartic acid racemization and radiocarbon analysis. Alkass K, Buchholz BA, Ohtani S, Yamamoto, T, Druid H, Spalding KL. 2010, Vols. Mol Cell Proteomics 2010;
9: 1022–1030.

22. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. Calloway CD, Reynolds RL, Herrin GL Jr, Anderson WW. 2000, Vols. Am J Hum Genet 2000; 66: 1384–1397.

23. Estimating human age in forensic samples by analysis of telomere repeats. Karlsson AO, Svensson A, Marklund A, Holmlund G. s.l. : Forensic Sci Int Genet Suppl Ser 2008; 1: 569–571, 2008.

24. Epigenetic predictor of age. S. Bockland, W. Lin, M.E. Sehl, F.J. Sanchez,J.S. Sinsheimer, S. Horvath, E. Vilain. 2011, Vols. PLoS ONE 6 (2011) 1–6.

25. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. C. Weidner, Q. Lin, C.M. Koch, L. Eisele, F. Beier, P. Ziegler, D.O. Bauerschlag, K.H. Jöckel, R. Erbel, T.W. Mühleisen, M. Zenke, T.H. Brümmendorf, W. Wagner. 2014, Vol. Genome Biol. 15 (2014) R24.

26. DNA methylation-based age prediction from saliva: high age predictability by combination of 7CpG markers. S.R. Hong, S.E. Jung, E.H. Lee, K.J. Shin, W.I. Yang, H.Y. Lee. 2017, Vols. Forensic Sci Int Genet. 29 (2017) 118–125. 27. DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples. Sang-Eun Jung, Seung Min Lim, Sae Rom Hong, Eun Hee Lee, Kyoung-Jin Shin. 38 (2019) 1-8, 2019, Vol. Forensic Science International: Genetics.

28. DNA-methylation based age prediction from saliva: high age predictability
by combination of 7 CpG markers. S.R. Hong, S.E. Jung, E.H. Lee, K.J. Shin,
W.I. Yang, H.Y. Lee. 2017, Vols. Forensic Sci Int Genet. 29 (2017) 118–125.
29. DNA Methylation: Bisulphite Modification and Analysis. Kate Patterson,
Laura Molloy, Wenjia Qu, Susan Clark. 2011.

30. Thermo Scientific. Thermo Scientific KingFisher® mL User Manual. 2007. User Manual Rev. 1.3; August 2007, Cat. no. 1508260. 31. ThermoFisher Scientific. DNA Sequencing by Capillary Electrophoresis Chemistry Guide. Third Edition.

32. MACHEREY-NAGEL – 04/2017, Rev. 01. Genomic DNA from forensic samples (User Manual). April 2017 / Rev. 01.

33. ThermoFisher Scientific. SeqStudio[™] Genetic Analyzer Instrument and Software User Guide. Publication Number MAN0016138.

34. ZymoResearch. EZ DNA Methylation-Lightning[™] Kit.

35. Life Technologies. Qubit® dsDNA HS Assay Kits. MAN0002326.

36. Applied Biosystems. User Bulletin Applied Biosystems 3730/3730xl DNA Analyzers and 3130/3130xl Genetic Analyzers. July 2005.

37. Genetic fingerprinting. AJ, Jeffreys. 2005; 11: 1035–1039, Vol. Nat Med.

38. Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. M, Kayser. Vols. Forensic Sci Int Genet 2015; 18: 33–48.

39. Molecular pathology and age estimation. Meissner C, Ritz-Timme S. Vols.Forensic Sci Int 2010; 203: 34–43.

40. Huang Y, Yan J, Hou J, Fu X, Li L, Hou Y: Developing a DNA methylation assay for human age prediction in blood and bloodstain. Forensic Sci Int Genet 2015; 17: 129–136. 41. Vidaki A, Ballard D, Aliferi A, Miller TH, Barron LP, Syndercombe Court D: DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. Forensic Sci Int Genet 2017; 28: 225–236.

42. Xu C, Qu H, Wang G, Xie B, Shi Y, Yang Y, et al: A novel strategy for forensic age prediction by DNA methylation and support vector regression model. Sci Rep 2015; 5: 17788.

43. Yi SH, Jia YS, Mei K, Yang RZ, Huang DX: Age-related DNA methylation changes for forensic age-prediction. Int J Legal Med 2015;

129: 237–244.

44. Yi SH, Xu LC, Mei K, Yang RZ, Huang DX: Isolation and identification of age-related DNA methylation markers for forensic age prediction. Forensic Sci Int Genet 2014; 11: 117–125.

45. Soares Bispo Santos Silva D, Antunes J, Balamurugan K, Duncan G, Sampaio Alho C, McCord B: Evaluation of DNA methylation markers and their potential to predict human aging. Electrophoresis 2015; 36: 1775–1780.

46. Lee HY, Jung SE, Oh YN, Choi A, Yang WI, Shin KJ: Epigenetic age signatures in the forensically relevant body fluid of semen: a preliminary study. Forensic Sci Int Genet 2015; 19: 28–34.

47. Giuliani C, Cilli E, Bacalini MG, Pirazzini C, Sazzini M, Gruppioni G, et al: Inferring chronological age from DNA methylation patterns of human teeth.Am J Phys Anthropol2016; 159: 585–595.

48. Naue J, Hoefsloot HCJ, Mook ORF, Rijlaarsdam- Hoekstra L, van der Zwalm MCH, Henneman P, et al: Chronological age prediction based on DNA methylation: massive parallel sequencing and random forest regression.

Forensic Sci Int Genet 2017; 31: 19–28.

49. Eduardoff M, Xavier C, Strobl C, Casas-Vargas A, Parson W: Optimized mtDNA control region primer extension capture analysis for forensically relevant samples and highly compromised mtDNA of different age and origin. Genes (Basel) 2017, DOI: 10.3390/ genes8100237.

50. Cho S, Jung SE, Hong SR, Lee EH, Lee JH, Lee SD, et al: Independent validation of DNAbased approaches for age prediction in blood. Forensic Sci Int Genet 2017; 29: 250–256.