



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova
Dipartimento di *Medicina Molecolare*

SCUOLA DI DOTTORATO DI RICERCA IN BIOMEDICINA
XXV CICLO

**“SVILUPPO DI UN METODO ACCURATO PER L’IDENTIFICAZIONE E
GENOTIPIZZAZIONE DEL PAPILOMAVIRUS UMANO BASATO SUL
SEQUENZIAMENTO 454 DI NUOVA GENERAZIONE”.**

Direttore della Scuola : Ch.mo Prof. Giorgio Palù

Supervisore :Ch.ma Prof.ssa Luisa Barzon

Dottorando : Valentina Militello

(**firma** del dottorando)

SOMMARIO

RIASSUNTO	3
INTRODUZIONE	5
1. Caratteristiche dei papillomavirus, ciclo replicativo, patogenesi.	5
2. Classificazione e genotipizzazione di HPV	7
3. Diagnosi di infezione da HPV	12
3.1. Polymerase chain reaction (PCR) con primer consensus	13
3.2. Test commerciali per la diagnosi di hpv	15
3.3. Sequenziamento diretto	17
3.4. Nuovi metodi di sequenziamento: pirosequenziamento e <i>ultradeep sequencing</i>	18
4. Utilità clinica ed epidemiologica della genotipizzazione di HPV	19
5. Tecnologie di Next-generation sequencing nella diagnostica virologica	23
OBBIETTIVI	27
METODI	28
1. Genotipizzazione di HPV mediante 454ultradeep sequencin	28
2. Metodi del primo protocollo di sequenziamento 454	39
3. Metodi del secondo protocollo di sequenziamento 454	40
RISULTATI	44
1. Risultati del primo protocollo di sequenziamento 454	44
2. Risultati del secondo protocollo di sequenziamento 454	50
DISCUSSIONE	62
BIBLIOGRAFIA	66
LISTA PUBBLICAZIONI E PRESENTAZIONI AI CONGRESSI PRODOTTE DURANTE I TRE ANNI DI DOTTORATO	72

RIASSUNTO

Un'accurata tipizzazione di HPV è essenziale per la valutazione e il monitoraggio della vaccinazione, come test di secondo livello nello screening del carcinoma cervicale e per le indagini epidemiologiche. In questo studio, è stato messo a punto e testato su campioni clinici un nuovo metodo di tipizzazione di HPV basato sul sequenziamento 454 next-generation (NGS) dell'amplicone L1 di HPV.

A tale scopo è stato sviluppato un primo protocollo che utilizzava dei primer ottenuti modificando i primer degenerati MY09/11; tale protocollo è stato testato su 164 campioni citologici cervicali fornendo una prima dimostrazione che il NGS può essere utilizzato per la tipizzazione di HPV. Il protocollo presentava però alcuni limiti legati soprattutto al disegno dei primer. È stato quindi studiato un secondo protocollo. Tale protocollo utilizzava un set di primer PGMY modificati in modo da ottenere sensibilità migliorata per alcuni tipi di HPV che non erano riconosciuti dai primer PGMY standard.

Usando una median coverage di 12,000-fold, il metodo NGS è arrivato a identificare correttamente tutti i tipi di HPV ad alto rischio sia in infezioni singole che multiple, con una sensibilità di 50 genomi equivalenti, come dimostrato testando il pannello di campioni del WHO LabNet EQA. L'analisi di miscele di linee cellulari positive per HPV16 e per HPV18 ha dimostrato che il metodo NGS può quantificare in modo riproducibile la proporzione con cui ogni tipo virale è presente nelle infezioni multiple con un ampio range. Nell'analisi di campioni clinici il metodo NGS ha identificato correttamente un ampio numero di tipi di HPV nelle infezioni multiple ed è stato anche utile nell'analisi di campioni cervicali con risultati discordanti al test hybrid capture 2 e line probe assays. Il metodo ha permesso inoltre di riconoscere la presenza di varianti intratipo (sostituzioni nucleotidiche) nello stesso campione clinico. In conclusione, è stato messo a punto un nuovo metodo di tipizzazione dell'HPV basato sul pirosequenziamento 454. Questo metodo è sensibile, specifico, quantitativo, e preciso sia nelle

infezioni singole che nelle infezioni multiple. Permette inoltre di identificare un ampio range di tipi di HPV e potenzialmente consente di scoprire nuovi tipi di HPV e varianti.

INTRODUZIONE

1. Caratteristiche dei papillomavirus, ciclo replicativo, patogenesi.

Il papillomavirus, originariamente incluso insieme ai polyomavirus nella famiglia dei *Papovaviridae* per le comuni caratteristiche capsidiche e genomiche, è attualmente riconosciuto ufficialmente dall'International Council on Taxonomy of Viruses (ICTV) come appartenente alla famiglia dei *Papillomaviridae* (de Villiers *et al.*, 2004).

I papillomavirus sono molto diffusi in natura e possono infettare un ampio numero di vertebrati, avendo però una specie-specificità molto alta ed uno specifico tropismo per le cellule epiteliali squamose.

HPV è un virus di piccole dimensioni (52-55 nm di diametro), il virione è privo di envelope; il capside icosaedrico è composto da 72 capsomeri formati da due proteine strutturali L1 ed L2; il genoma è a DNA circolare a doppio filamento (circa 8.000 bp), il sito di replicazione del DNA virale e di assemblaggio capsidico è nucleare. Il genoma dei papillomavirus presenta almeno una decina di *open reading frames* (ORF) localizzati in un solo filamento di DNA. Un solo filamento quindi funge da stampo per la trascrizione delle proteine precoci (E, Early) e tardive (L, Late). Le regioni precoci del genoma virale sono espresse nelle cellule con infezione non produttiva e nelle cellule trasformate, mentre le regioni tardive, incluse le proteine capsidiche L1 ed L2, sono espresse solo nelle cellule in cui avviene un'infezione produttiva. Questo si riflette su una caratteristica essenziale del virus, cioè sulla sua capacità di dare un'infezione produttiva esclusivamente nelle cellule squamose in differenziamento terminale e quindi sul fatto che il ciclo vitale del virus sia strettamente correlato con il programma di differenziamento dell'epitelio squamoso.

I papillomavirus sono caratterizzati da un particolare tropismo per le cellule degli epiteli squamosi e l'infezione ha solitamente inizio nelle cellule basali e parabasali dell'epitelio, dotate di spiccata attività proliferativa. Per dare infezione, il virus deve poter accedere al compartimento "germinativo" dell'epitelio ed è necessaria la presenza di soluzioni di continuo della cute e delle mucose.

Nella fase iniziale dell'infezione, quando cioè il virus colonizza le cellule basali e parabasali dell'epitelio, il genoma virale va incontro ad una replicazione in forma episomiale, essendo presente come frammento extracromosomico di DNA circolare. Durante l'infezione produttiva, l'espressione dei geni virali procede sequenzialmente seguendo le fasi di differenziamento dell'epitelio, a partire dalle cellule basali e parabasali, dove i geni precoci del virus sono più attivi, fino agli strati superiori dell'epitelio (intermedio e superficiale), in cui si assiste ad una

significativa produzione di proteine capsidiche e alla formazione del virione completo, ovvero la particella virale infettante (Howley&Douglas, 2007).

Il genoma di HPV (fig. 1) contiene una regione principale regolatoria, costituita da un lungo segmento di controllo (*long control region*, LCR), e una regione codificante le proteine virali strutturali e non strutturali. La maggior parte degli HPV hanno sei geni non strutturali ampiamente studiati e descritti (*E1*, *E2*, *E4*, *E5*, *E6* ed *E7*) e due ORFs (*E3* e *E8*) la cui funzione non è ancora ben nota. La maggior parte dei geni virali sono multifunzionali.

La proteina E1 ha un'attività di DNA replicasi. E2 regola l'attivazione e la repressione della trascrizione genica e partecipa alla regolazione della replicazione del DNA virale. Tale attività è mediata dal legame della proteina E2 alla regione di controllo del genoma virale (LCR). E4 codifica per una fosfoproteina citoplasmatica, la cui funzione non è stata ancora identificata. HPV ha tre oncogeni principali: *E5*, *E6* ed *E7*. La proteina codificata da *E5* esplica la propria funzione oncogena interagendo e attivando recettori specifici per fattori di crescita. La proteina E6, in particolare quella dei *tipi* ad alto rischio, induce la degradazione ubiquitina-dipendente del soppressore tumorale p53, mentre la proteina E7 lega ed inattiva la forma non fosforilata di un altro soppressore tumorale, la proteina del retinoblastoma pRb.

Le proteine strutturali di HPV sono essenzialmente due: L1 ed L2. Le proteine L1 costituiscono il capsido virale, il quale è composto da 72 pentameri di L1. Gli epitopi immunodominanti neutralizzanti sono localizzati su L1. La proteina L2 è presente in minore quantità rispetto ad L1 e partecipa con quest'ultima alla formazione del capsido.

I geni virali possono essere anche classificati in base all'ordine con cui vengono espressi durante il ciclo replicativo, e suddivisi quindi in geni precoci (*early*) e tardivi (*late*). I geni *E1*, *E2*, *E4*, *E5*, *E6* ed *E7*, che codificano proteine coinvolte nella regolazione della replicazione virale e nella trasformazione cellulare, sono espressi precocemente negli strati basali epiteliali, mentre *L1* ed *L2*, essendo geni codificanti le proteine strutturali, sono espressi quasi esclusivamente negli strati superficiali degli epitelii in attiva replicazione.

Nello stato episomiale del virus, l'espressione degli oncogeni *E6* ed *E7* è inibita dalla proteina E2. L'integrazione del genoma virale nel DNA della cellula provoca un'interruzione della regione codificante la proteina E2 e di conseguenza viene a mancare la sua funzione inibitoria sui promotori degli oncogeni E6 ed E7. Il sito di integrazione del DNA virale nei cromosomi della cellula ospite è casuale, ma il tipo di integrazione è clonale: lo stesso in tutte le cellule che compongono una determinata neoplasia.

Nell'epitelio non infettato dal virus, le cellule basali escono dal ciclo cellulare subito dopo essere migrate nello strato parabasale, dove subiscono un processo che porta alla differenziazione. In questo strato si verificano una serie di cambiamenti come il cross-linking dei filamenti intermedi

della cheratina, la formazione dello strato corneo e la secrezione di lipidi, che contribuiscono alla formazione di una barriera fisica protettiva contro l'ambiente.

Durante l'infezione da HPV, invece, l'espressione di E6 ed E7 nella cellula blocca i sistemi d'inibizione del ciclo cellulare e, di conseguenza, il normale differenziamento cellulare è inibito (Howley&Douglas, 2007). Le due proteine oncogene, E6 ed E7, collaborano per ottenere tali effetti, e sono espresse dai *tipi* virali ad alto rischio come un mRNA bicistronico (Stacey *et al.*, 2000) sotto il controllo del promotore virale precoce p97. Nei tipi virali non oncogeni invece E6 ed E7 possono essere espressi su differenti messaggeri (Howley &Douglas, 2007).

L'attività oncogena degli HPV definiti ad alto rischio è mediata dalle proteine virali E6 ed E7, che legano ed inattivano gli onco-soppressori p53 ed Rb, rispettivamente, oltre ad interagire con molte altre proteine cellulari implicate nella proliferazione e segnalazione cellulare e nell'adesione intercellulare.

La normale funzione di Rb è di bloccare la cellula in uno stadio del ciclo cellulare prevenendone errate o dannose divisioni. Così, quando Rb è difettosa, alcune cellule mutate possono continuare a dividersi indisturbate dando origine al tumore. E7 può associarsi anche ad altri fattori coinvolti nella proliferazione cellulare come l'enzima istone deacetilasi (Longworth & Laimins, 2004).

La proteina virale E6 complementa la funzione di E7, impedendo l'induzione dell'apoptosi mediante inattivazione di p53. P53 è un fattore di trascrizione che regola il ciclo cellulare con funzione di soppressore tumorale per la sua capacità di preservare la stabilità del genoma attraverso la prevenzione delle mutazioni.

2. Classificazione e genotipizzazione di HPV

I papillomavirus descritti e caratterizzati sono più di 180 e si stima che negli anni futuri molti altri ne verranno identificati. Essi sono inoltre molto diversi tra loro per ospite e per tipo e sede di lesione cui sono correlati. Per tale motivo la classificazione di questo gruppo di virus, che comprende importanti patogeni umani, ha richiesto tre decenni di studio e di dibattito.

I papillomavirus isolati vengono tradizionalmente identificati come "tipi". La classificazione si basa, oltre che sulla specie da cui è stato isolato il tipo virale, soprattutto sull'estensione e il grado di omologia tra i genomi virali. Un tipo virale viene classificato comparando la sequenza nucleotidica di regioni specifiche del genoma particolarmente conservate tra i vari membri della famiglia, quali E6, E7 ed L1 ORF.

I nuovi metodi di sequenziamento e i progressi di biologia molecolare (Snijders *et al.*, 1990; Manos *et al.*, 1989) hanno permesso di migliorare la classificazione dei papillomavirus. Tali analisi hanno rivelato inoltre come il genoma di questi virus sia estremamente stabile e come gli

eventi di ricombinazione o di mutazione siano talmente rari da avere una frequenza molto simile a quella degli organismi ospiti dell'infezione (de Villiers, 2004).

In base ai criteri adottati dal *Papillomavirus Nomenclature Committe*, per definire un nuovo tipo virale, l'omologia di sequenza dei suoi geni virali deve essere inferiore al 90% rispetto a quelli dei tipi noti; gradi di omologia superiori individuano invece una variante di un tipo noto o di un suo sottotipo (International Papillomavirus Workshop, Quebec 1995). Al nuovo HPV viene assegnato un numero, che viene stabilito solo dopo che il genoma è stato isolato e completamente caratterizzato; quindi il genoma viene depositato al *Reference Center for Papillomaviruses* (Heidelberg).

In base all'allineamento della sequenza altamente conservata del ORF di L1 quindi, i virus HPV risultano classificati in 12 generi, indicati con le prime dodici lettere dell'alfabeto greco. Tutti i membri appartenenti allo stesso genere hanno un'omologia di sequenza almeno del 60%. HPV dello stesso genere con una omologia superiore al 60-70% vengono raggruppati in specie, quelli con omologia tra il 70 e il 90% sono raggruppati in tipi, quelli con più del 90% di omologia in sottipi, e quelli con omologia superiore al 98% sono considerati varianti. Cinque dei dodici generi, alfa, beta, gama, mu e nu sono papillomavirus umani mentre gli altri sette producono infezione negli animali (fig. 1). Ai diversi sottogruppi virali corrispondono differenti caratteristiche clinico-patologiche. Gli HPV appartenenti al genere alfa sono di maggiore importanza medica in quanto causano tumori genitali e delle mucose. Gli HPV del genere beta, gamma, mu e nu hanno un tropismo cutaneo e mucoso, provocando verruche volgari, lesioni orali e cutanee ed epidermodisplasia verruciforme.

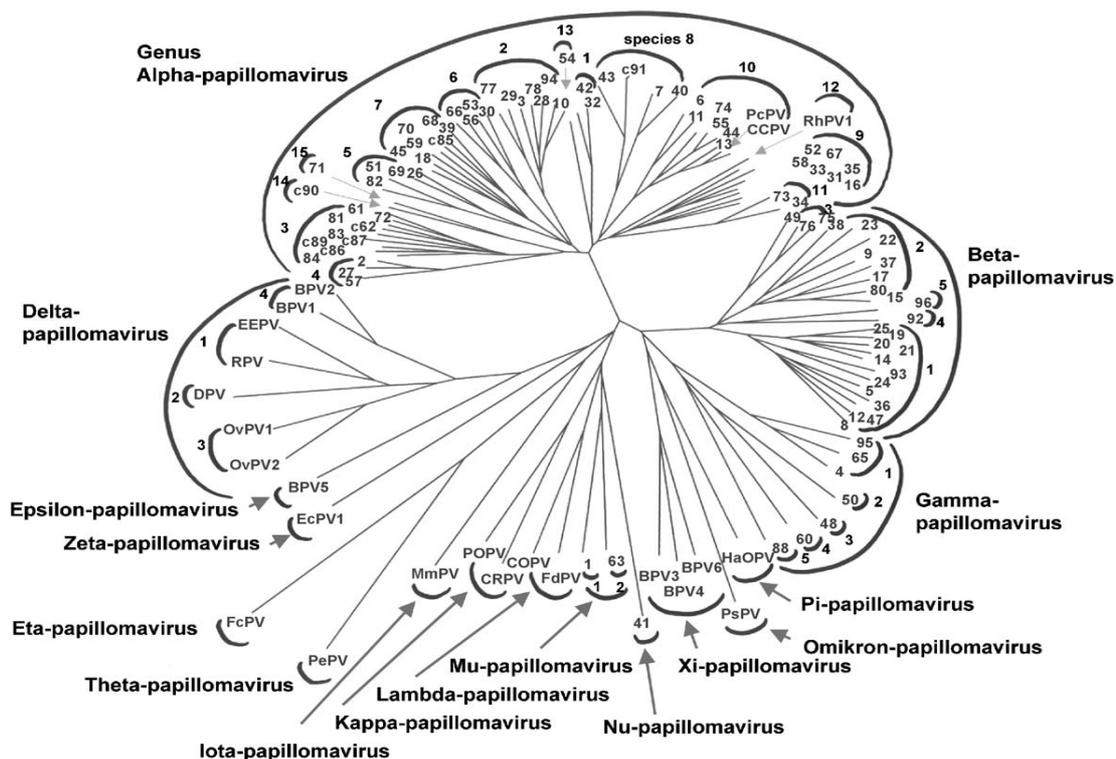


Fig.1. Classificazione e albero filogenetico di HPV costruito confrontando la sequenza dell'ORF L1 (da Hoory *et al.*, 2008).

Lo scopo della classificazione genotipica è quello di stabilire correlazioni tra gli innumerevoli tipi virali, creare un linguaggio comune che permetta di distinguere genere e specie (de Villiers, 2004), ma soprattutto investigare le relazioni tra le classificazioni tassonomiche e le proprietà biologiche e patologiche del virus.

Studi *in vitro* hanno dimostrato come alcuni HPV dotati di tropismo per le mucose e correlati tra loro filogeneticamente abbiano un potenziale oncogenico comune: le proteine E6 ed E7 di tale gruppo di virus interferiscono infatti con il ciclo cellulare interagendo rispettivamente con le proteine cellulari p53 ed Rb. Da studi epidemiologici si è visto però che, all'interno di tale gruppo di virus, alcuni sono dotati di un potenziale oncogeno maggiore e sono più frequentemente associati a tumore, altri sono invece dotati di capacità oncogene minori; questo dato è probabilmente da imputare ad una diversa competenza delle proteine E6 ed E7 nell'interferire con le proteine cellulari.

Quindi dal punto di vista delle capacità oncogene possiamo distinguere HPV ad alto rischio, tra cui sono inclusi i tipi 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, HPV a possibile alto rischio, tra cui sono inclusi i tipi 26, 53 e 66, 68, 73, 82 e gli HPV a basso rischio, tra cui sono inclusi i tipi 6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81, CP6108. I tipi virali ad alto rischio, anche dal punto di vista filogenetico, risultano vicini tra loro, mentre risultano filogeneticamente distanti dai tipi a basso rischio (de Villiers *et al.*, 2004). Per alcuni HPV il rischio oncogeno non è tuttora ben

chiarito, ma gli studi filogenetici possono dare un indirizzo in tal senso. Ad esempio HPV-67 è stato inizialmente isolato da neoplasie vaginali intraepiteliali (VIN) e si è successivamente appurato come esso sia filogeneticamente associato ad HPV-16, uno dei più diffusi HPV ad alto rischio (Brink *et al.*, 2007).

L'*International Agency for Research on Cancer* (IARC) l'ente del World Health Organization (WHO) che si occupa di coordinare e condurre le ricerche sul cancro, ha recentemente classificato i principali tipi di HPV in base alla capacità oncogena, come riassunto in tab. I (Bouvard *et al.*, 2009; Schiffman *et al.*, 2009). In tale classificazione i tipi virali sono divisi in agenti cancerogeni certi (gruppo 1), probabili agenti cancerogeni per l'uomo (gruppo 2A), possibili agenti cancerogeni per l'uomo (gruppo 2B) e non classificabili come agenti cancerogeni per l'uomo (gruppo 3).

Tab. I: Classificazione IARC di HPV in gruppi a seconda del potere oncogeno (Bouvard *et al.*, 2009).

<i>Alpha-papillomavirus</i>		
Gruppo IARC	Tipo di HPV	
1	16	Il tipo di HPV più potente come agente cancerogeno, riconosciuto come causa di carcinoma in diverse sedi
1	18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59	Sufficienti evidenze di associazione causale con il carcinoma cervicale
2°	68	Limitate evidenze nell'uomo e forti evidenze meccanicistiche di associazione causale con il carcinoma cervicale
2B	26, 53, 66, 67, 70, 73, 82	Limitate evidenze nell'uomo di associazione causale con il carcinoma cervicale
2B	30, 34, 69, 85, 97	Classificati con possibili agenti cancerogeni per analogia filogenetica con tipi HPV con sufficienti o limitate evidenze di cancerogenicità nell'uomo
3	6, 11	Inadeguate evidenze epidemiologiche e assenza di potenziale carcinogenico in studi meccanicistici

Beta-papillomavirus		
2B	5, 8	Limitate evidenze di associazione causale con il carcinoma cutaneo in EV
3	Altri tipi dei generi beta e gamma	Inadeguate evidenze epidemiologiche e assenza dipotenziale carcinogenico in studi meccanicistici

In una recente revisione della letteratura (Schiffman *et al.*, 2009) è discussa tale classificazione e la sua ricaduta sulle decisioni riguardanti le scelte nei test di screening e sullo sviluppo dei vaccini. La nuova classificazione IARC assegna un ruolo di vero cancerogeno solo ad un'infezione persistente da HPV-16; definisce anche l'importanza dell'HPV-18, soprattutto in relazione con l'adenocarcinoma; include sei tipi alpha-7 (HPV-45) e alpha-9 (HPV-31, HPV-33, HPV-35, HPV-52, HPV-58) tra gli otto tipi più importanti globalmente, nonostante alcune variabilità regionali nella percentuale di carcinomi dovuti a ciascun tipo; stabilisce un piccolo aumento del contributo eziologico di un altro gruppo di tipi cancerogeni alpha-5 (HPV-51), alpha-6 (HPV-56), e alpha-7 (HPV-39 e HPV-59) che causano una più o meno piccola percentuale di casi di carcinoma cervicale nel mondo. Evidenzia il fatto che rimane irrisolta a causa dei pochi dati disponibili la questione riguardante alcuni tipi virali probabili o possibili cancerogeni come l'alpha-7 HPV-68 e l'alpha-11 HPV-73). L'aumento dei dati su HPV-66 permette di rivalutare tale tipo virale, nonostante le evidenze vengano giudicate ancora limitate; i dati disponibili mostrano che il ritrovamento nel cancro di HPV-66 è molto raro mentre è comune nelle donne con citologia negativa. HPV-53, specie alpha-6, mostra la stessa prevalenza nella popolazione comune e in casi estremamente rari si associa al cancro. Il gruppo di lavoro dello IARC sottolinea, in particolare, che l'inclusione di HPV-53 tra i cancerogeni nei pannelli di screening possa ridurre la specificità e il valore predittivo positivo del test senza accrescerne sensibilità e valore predittivo negativo.

Ci sono molti tipi per i quali ci sono isolate o nessuna evidenza di cancerogenicità. Per alcuni tipi ci sono casi aneddotici, ma molto interessanti, che meritano approfondimento. Ad esempio, ci sono alcuni report su singole infezioni da tipo alpha-9 HPV-70, e pochi casi di HPV-67 nel cancro (Gudleviciene *et al.*, 2006; Wentzensen *et al.*, 2009; Gargiulo *et al.*, 2007; Anderson *et al.*, 2005). Quest'ultimo è il solo tipo noto della specie alpha-9 non catalogato come cancerogeno. La classificazione IARC non enfatizza il possibile ruolo dell'immunosoppressione nella cancerogenicità di HPV (Schiffman *et al.*, 2009). Per causare il cancro, un'infezione da HPV deve persistere ed è possibile che alcuni tipi di HPV siano solo deboli cancerogeni poichè

hanno scarsa capacità di persistenza. Ad esempio, la cancerogenicità del tipo alpha-5 HPV-26 è stata supportata da un recente caso di cancro periungueale multiplo in un soggetto immunocompromesso, avendo un alta carica virale e un attiva trascrizione (De Vuyst *et al.*, 2008). HPV-26 è un tipo raro e probabilmente l'immunosoppressione è un contributo importante alla cancerogenesi.

Tale riclassificazione IARC è fondamentale e va tenuta in considerazione per tutte le decisioni riguardanti i test di screening e lo sviluppo dei vaccini, come punto di partenza per gli studi molecolari ed epidemiologici futuri.

3. Diagnosi di infezione da HPV

HPV non può essere coltivato, pertanto i metodi per rivelare la presenza del virus si basano esclusivamente sulla ricerca degli acidi nucleici.

Un buon metodo, oltre che specifico e sensibile, dovrebbe essere in grado di distinguere il genotipo virale e di individuare il più ampio numero di tipi virali. Non tutti i test attualmente disponibili rispondono a tali caratteristiche: alcuni permettono di individuare solo i genotipi virali più frequenti e ne escludono alcuni più rari, considerati meno rilevanti per programmi di screening; altri permettono di distinguere esclusivamente tra HPV ad alto e basso rischio. La metodica deve inoltre permettere di individuare le coinfezioni da tipi multipli di HPV, poiché il 20-30% delle infezioni da HPV sono dovute a più tipi virali ed è importante determinare il contributo di ciascun virus nell'infezione.

Diversi test *in house* e protocolli sono stati sviluppati nell'ultimo ventennio e applicati a studi randomizzati e trial clinici; alcuni laboratori li utilizzano tuttora anche per la pratica clinica e possono essere indispensabili in alcune situazioni particolari. Questi test si basano essenzialmente sull'amplificazione tramite PCR usando primer consensus e sequenziamento. La maggior parte dei laboratori diagnostici utilizza però per la routine uno dei molteplici test commerciali ormai disponibili per il test HPV. Esistono almeno 125 test disponibili in commercio e 84 varianti di alcuni di questi che permettono di identificare un pannello di tipi virali più ampio rispetto al test originale. Una recente review (Poljak *et al.*, 2012) ha diviso i test HPV presenti in commercio in sei diversi gruppi e relativi sottogruppi: *hrHPV DNA test*, *hrHPV DNA test* che consentono una genotipizzazione parziale dei principali tipi di HPV ad alto rischio in contemporanea o in differita, *HPV DNA test* che consentono la genotipizzazione di tutti i tipi di HPV, *HPV DNA type- o group-specific genotyping test* di genotipizzazione di HPV tipo- o

gruppo -specifici, test che valutano l'*mRNA E6/E7 di hrHPV*, HPV test basati su *in situ hybridization*.

3.1. Polymerase chain reaction (PCR) con primer consensus

La PCR è la metodica più utilizzata per individuare DNA di HPV, ed esistono due possibili approcci: utilizzare una PCR tipo-specifica o una PCR ad ampio spettro.

I primer tipo-specifici permettono di rilevare la presenza di un singolo tipo virale per reazione. È in tal caso molto importante il disegno dei primer che, se mappano in una regione troppo conservata, potrebbero interagire anche con altri tipi virali. Tale approccio è quindi utile qualora si indaghi o si sospetti uno specifico tipo virale (es. HPV-16 in un tumore della cervice uterina), ma qualora sia necessaria un'indagine genotipica completa tale approccio risulterebbe indaginoso, dovendo allestire reazioni separate per ogni tipo virale (Molijn *et al.*, 2005).

L'utilizzo di primer ad ampio spettro risulta quindi essere l'approccio più impiegato. In tal caso i primer hanno come target una regione del genoma virale altamente conservata nei differenti tipi di HPV, in modo da riconoscere un ampio spettro di genotipi. La regione genomica più adatta in tal senso è la regione *L1*, anche se sono stati descritti in letteratura primer che mappano nel gene *E1* (Tieben *et al.*, 1993).

I sistemi di primer possono essere disegnati in tre modi per ottenere una PCR ad ampio spettro (Molijn *et al.*, 2005). Il primo utilizza una coppia di primer che mappa in una regione conservata, ma che è pienamente complementare ad un solo o a pochi genotipi. Perché possano essere riconosciuti un numero maggiore di genotipi, la reazione deve essere condotta ad una bassa temperatura di annealing. Un esempio di tale approccio è il sistema GP5+/GP6+ che amplifica una regione di L1 di 150 pb (temperatura di annealing di 40°C) (Jacobs *et al.*, 1997).

Il secondo approccio utilizza primer che contengono uno o più basi degenerate per compensare le differenze di sequenza tra i diversi tipi virali. Tali primer non richiedono basse temperature di annealing (Tieben *et al.*, 1993). Esempi di primer degenerati sono il sistema MY09/MY11 e il sistema CPI/II, che comprendono una miscela complessa di primer contenenti basi degenerate (differenti oligonucleotidi) per compensare le differenze di sequenza tra i diversi tipi virali. Il primo sistema amplifica una regione di L1 di 450 pb (temperatura di annealing di 54°C), il secondo amplifica invece una regione di E1 ORF di 188 pb. Lo svantaggio di tali sistemi è che le degenerazioni non sono altamente riproducibili e pertanto presentano variazioni tra un lotto di primer e l'altro, e soprattutto differenze nell'efficacia di amplificazione di ciascun genotipo. Il sistema PGMY è stato sviluppato proprio per aumentare la riproducibilità e sensibilità del sistema MY09/MY11, utilizzando un pool di oligonucleotidi che mappino nella stessa regione di

L1 ORF. I tipi virali sono stati raggruppati in base all'omologia di sequenza di ciascuna delle regioni di legame dei due primer; sulla base di tali raggruppamenti sono stati disegnati un set di 5 oligonucleotidi che compone il pool di primer PGMY11 e un set di 13 oligonucleotidi che compone il pool PGMY09 (Gravitt *et al.*, 2000).

Il terzo metodo combina un numero di primer distinti che mappano nella stessa regione genomica. Essi non contengono siti degenerati, ma inosina la quale ha la capacità di appaiarsi con tutti i nucleotidi presenti nei punti in cui ci sono differenze di sequenza. Questa miscela di primer ha il vantaggio che gli oligonucleotidi sono sintetizzati in modo altamente riproducibile e che la PCR può essere condotta a temperature ottimali. Esempi di tali set di primer pubblicati in letteratura sono il sistema SPF10, che amplifica un piccolo frammento di 65 pb dell'L1 ORF di 43 tipi virali (Kleter, 1998) e il sistema Amplicor, che amplifica una sequenza di 170 pb sempre in L1 (genotipi ad "alto rischio"), utilizzando 5 primer forward e 7 reverse (Monsonogo *et al.*, 2005).

Anche la scelta della dimensione del prodotto di amplificazione è importante, perché l'efficienza della reazione aumenta quanto più l'amplificato è piccolo.

Ciascun set di primer permette inoltre di amplificare i diversi tipi virali con diversa efficienza. Ad esempio i primer CPI/II rilevano i tipi virali -1, -2, -3, -4, -5, -6b, -7, -8, -9, -10a, -11, -12, -14a, -16, -17, -18, -19, -20, -21, -22, -24, -25, -31, -33, -36, -37, -38, -39 e -46, mentre il -15, -23, -49 e -50 sono scarsamente amplificati ed il -41 non è amplificato (Tieben *et al.*, 1993). L'associazione di due PCR con diversi primer come CPI/II e MY09/11 o di GP5+/6+ e MY09/11, permette di aumentare la sensibilità del test. La sensibilità può anche essere aumentata con una metodica di PCR nested. Ad esempio è possibile effettuare una PCR con primer MY09/11 seguita da una PCR con primer GP5+/6+, i quali fiancheggiano una sequenza interna al primo amplificato.

In figura 2 sono riportati i sistemi di primer più utilizzati, la lunghezza dell'amplificato e la sua posizione nel genoma virale (Molijn, 2005).

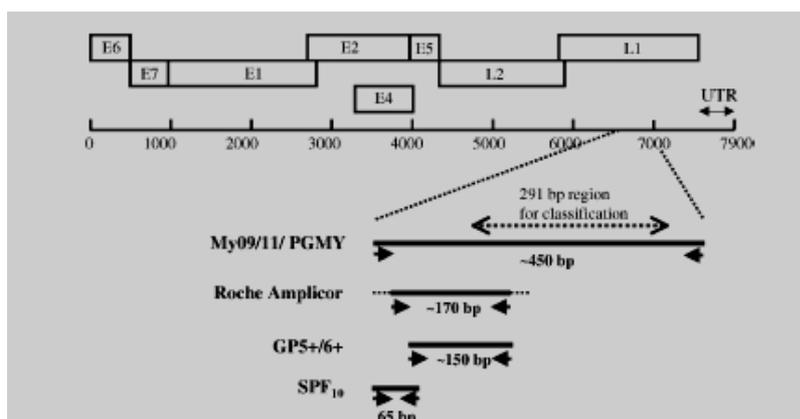


Fig. 2: Genoma di HPV in forma lineare; è indicata la posizione delle regioni early (E), late (L) e della

regione UTR e la posizione delle coppie di primer più utilizzate: MY09/11, GP5+/6+, SPF₁₀ e del sistema Roche Amplicor (localizzazione precisa ignota) con le rispettive dimensioni degli amplificati. Il frammento di 291bp usato per la classificazione formale di HPV è localizzato in L1 (Molijn *et al.*, 2005).

3.2. Test commerciali per la diagnosi di hpv

hrHPV DNA test

In questa categoria sono compresi test qualitativi o semiquantitativi, che permettono di individuare tipi di HPV considerati oncogeni in aggregato, usando diverse tecnologie, senza distinguere il singolo tipo o i diversi tipi di HPV presenti nel campione. A questo gruppo appartengono due test entrambe FDA approved e di uso comune per il triage di donne con citologia dubbia, nel follow up post trattamento delle neoplasie intraepiteliali di alto grado (CIN) e nello screening primario: Hybrid Capture 2 (HC2) HPV DNA test (HC2, Digene Corp., USA), marcato Qiagen (MD, USA) e Cervista HPV HR Test (Hologic, Madison, WI).

Il sistema HC2 si basa su un metodo di amplificazione del segnale che utilizza due differenti miscele di sonde ad RNA: la prima contiene le sonde per i tipi ad alto (16, 18, 31,33,35, 39, 45, 51, 52, 56, 58, 59, 68), la seconda quelli a basso rischio (6, 11, 42, 43, 44). Il DNA denaturato viene ibridato con le sonde a RNA presenti nella miscela. Gli ibridi DNA-RNA eventualmente formati vengono catturati in fase solida in appositi pozzetti e fatti reagire con anticorpi monoclonali marcati con perossidasi; quindi il segnale viene rilevato per chemiluminescenza, ottenendo una misurazione semiquantitativa. La sensibilità analitica del metodo è di 1 pg/ml di DNA che corrisponde a 10⁵ copie genomiche. Tale metodica presenta alcune limitazioni legate alla sensibilità inferiore rispetto alle metodiche in PCR, presenta rischi di cross reattività tra i due cocktail di sonde e soprattutto distingue genotipi ad alto e basso rischio, ma non consente una genotipizzazione (Brink *et al.*, 2007). D'altro canto, questo test è stato ampiamente validato in campo clinico in programmi di screening, dimostrando la sua efficacia nella prevenzione del cervico-carcinoma e delle lesioni cervicali pre-invasive. HC2 è il test attualmente più utilizzato, e le linee guida raccomandano che qualsiasi nuovo test, per essere utilizzabile in un trial clinico per il carcinoma cervicale, dovrebbe possedere caratteristiche cliniche almeno equivalenti ad HC2 (Mijer *et al.*, 2009).

Il test Cervista HPV HR si basa su ibridazione ed amplificazione del segnale con chimica Invader e identifica 14 tipi di HPV ad alto rischio attraverso una metodica totalmente automatizzata. È indicato da FDA per il triage di pazienti con ASCUS o in associazione alla citologia cervicale per lo screening di donne a partire dai trentenni.

hrHPV DNA test che consentono una genotipizzazione parziale dei principali tipi di HPV ad alto rischio in contemporanea o in differita.

In questo gruppo sono inclusi alcuni nuovi test che permettono di rilevare 13 o 14 tipi virali ad alto rischio e in contemporanea o in una reazione separata, di valutare la presenza di HPV16 e HPV18. Questi test sono stati sviluppati sulla base dei risultati di studi clinici, che dimostrano come questi due tipi virali abbiano un potenziale oncogeno molto superiore agli altri genotipi.

In questa categoria è incluso il test cobas 4800 HPV (Roche Molecular Systems Inc., Alameda, CA, USA). Il test, approvato FDA, consente fare uno screening per 12 genotipi ad alto rischio e di valutare singolarmente la presenza di HPV16 e 18. Il test Cervista HPV16/18 (Hologic), approvato FDA, viene utilizzato nei casi di positività al test Cervista HPV HR e viene pertanto inserito nel gruppo di test eseguiti “in differita”.

A questi due test approvati FDA, si aggiunge il test clinicamente validato Abbott RealTime High Risk HPV (Abbott Molecular, Des Plaines, IL). Il test consente di rilevare in contemporanea tramite real-time PCR 14 HPV ad alto rischio e HPV16 e HPV18.

Test che consentono la genotipizzazione di tutti i tipi di HPV

La maggior parte dei test inclusi in questa categoria utilizza la metodica dell’ibridazione inversa per la genotipizzazione di un ampio spettro di tipi di HPV. Una sequenza di genoma virale viene prima amplificata tramite PCR, l’amplicone ottenuto viene denaturato e rilevato tramite ibridazione con specifiche sonde immobilizzate su *strip*, filtri o micropiastre e reazione colorimetrica o chemiluminescenza.

I test di questo gruppo per i quali esistono più dati in letteratura sono il Linear Array HPV Genotyping Test (Roche Molecular Systems Inc., Alameda, CA, USA), l’HPV SPF10 LiPA25 (Labo Bio-Medical Products, Ev Rijnswijk, Netherlands) e diverse versioni dell’INNO-LiPA HPV Genotyping test (Innogenetics NV, Gent, Belgium).

Il principio dell’ibridazione inversa è stato anche impiegato per lo sviluppo di test basati sui microarray. In questo caso sonde tipo-specifiche possono essere immobilizzate su vetrini da microarray o micro chip, permettendo di arrivare a distinguere simultaneamente più di 100 tipi di HPV (Pagliusi et al., 2007). I test di questo tipo su cui ci sono più dati in letteratura sono il PapilloCheck HPV-screening Test/Papillocheck High-risk Test (greiner Bio-One, Frickenhausen, Germany) e il Clart HPV2-Papillomavirus Clinical Arrays (Gnomica, Coslada, Spain).

Le sonde di oligonucleotidi possono essere fissate su superfici anche molto diverse, come nel caso del sistema multiplex papillomavirus genotyping (MPG), recentemente descritto, in cui il prodotto di PCR ottenuto con i primer GP5+/GP6+ è rilevato mediante sonde fluorescenti legate a perle di polistirene (luminex suspension array technology). Le perle sono colorate con diverse combinazioni di due distinti fluorofori ed è possibile creare fino a 100 diversi tipi di perle che

assorbano a spettri diversi e su ciascun tipo è possibile legare un diverso tipo di sonda. Tale metodo permetterebbe quindi di distinguere più di 100 tipi virali contemporaneamente (Schmitt *et al.*, 2006).

Test di genotipizzazione dell'HPV tipo- o gruppo-specifici

Come già detto prima, primer tipo-specifici permettono di rilevare la presenza di un singolo tipo virale per reazione. Tale approccio è quindi utile qualora si indaghi o si sospetti uno specifico tipo virale (es. HPV-16 in un tumore della cervice uterina).

Primer tipo specifico vengono utilizzati soprattutto tramite real-time PCR, la quale ha il vantaggio di permettere una quantificazione della carica virale (Brink *et al.*, 2007). L'utilità di quantificare la carica del tipo virale identificato è stata suggerita da alcuni autori per i tumori della cervice uterina HPV-16-positivi. Un'alta carica di HPV-16 sembrerebbe correlata con CIN di alto grado e carcinoma invasivo o con un maggior rischio di sviluppare CIN2-3 (Josefsson *et al.*, 2000). Tale correlazione non sembrerebbe dimostrabile per gli altri genotipi virali (Lorincz, 2002). La real-time PCR può essere impiegata anche per distinguere i diversi genotipi o allestendo reazioni separate per ogni tipo virale (Lindh *et al.*, 2007) o tramite una miscela di sonde; spesso le sonde presentano caratteristiche di ibridazione differenti e la standardizzazione della reazione risulta in tal caso piuttosto complessa (Hart *et al.*, 2001). Sono comunque stati pubblicati in letteratura alcuni lavori che utilizzano multiplex real-time PCR con sonde Beacon, in sistemi che permettono contemporaneamente di genotipizzare e quantificare i virus (Szuhai *et al.*, 2001).

HrHPV E6/E7 mRNA test

Molti studi hanno dimostrato l'utilità clinica e la specificità di un'analisi che consenta di rilevare l'mRNA di HPV invece che il DNA. Tali test valutano l'mRNA delle oncoproteine E6 ed E7 virali tramite PCR inversa e metodica NASBA. I test in commercio di questo tipo sono attualmente APTIMA HPV Assay (Gene-Probe Inc., San Diego, CA) approvato FDA, e ilPreTect HPV-Proofer (NorChip, Klokkarstua, Norway) e il NucliSens EasyQ HPV V1 test (Biomerieux, Marcy l'Etoile, France).

HPV test basati su *in situ* hybridization

L'ibridazione *in situ* è l'unico metodo molecolare che consente di valutare HPV topograficamente nelle lesioni patologiche tramite valutazione microscopica. Test commerciali basati su questo principio sono stati sviluppati ma sono considerati laboriosi e con una sensibilità e specificità clinica troppo bassa.

3.3. Sequenziamento diretto

Le metodiche di sequenziamento rapido vengono utilizzate per l'identificazione di HPV dopo amplificazione con primer *consensus* in molti protocolli *in house*. Queste metodiche consentono

di identificare quasi tutti i genotipi noti, ma presentano anche limiti importanti. Il sequenziamento infatti non permette di identificare il genotipo qualora la carica virale sia bassa e non è utilizzabile per i casi di coinfezioni da genotipi multipli. La tecnica del sequenziamento tradizionale consiste nella generazione di frammenti a singola elica attraverso un'interruzione dell'estensione della catena. Ciò si realizza con una miscela contenente un primer complementare a parte della sequenza nucleotidica; l'enzima DNA polimerasi; i desossiribonucleotidi trifosfato (dATP, dTTP, dCTP, dGTP); un piccolo quantitativo di analoghi 2'-3'-didesossi (ddNTP), i quali possiedono un solo atomo di idrogeno al posto del gruppo 3'-OH e sono marcati con diversi fluorofori. Il primer si appaia al filamento stampo e viene esteso dalla DNA polimerasi. Questa aggiunge all'estremità libera del primer i dNTP complementari al filamento stampo, ma, di tanto in tanto, casualmente invece di aggiungere un dNTP, aggiunge un ddNTP. Con l'incorporazione del ddNTP la sintesi del filamento complementare si ferma, poiché i ddNTP non permettono la formazione del legame con il successivo nucleotide trifosfato. I dNTP hanno un 3'OH nello zucchero desossiribosio che consente l'aggiunta di nucleotidi, i ddNTP invece hanno un 3'-H che non consente tale aggiunta. Si ottengono in questo modo frammenti di DNA di varia lunghezza, che differiscono per il tipo di ddNTP con il quale terminano. Tali frammenti vengono separati mediante elettroforesi capillare e la loro migrazione è rilevata grazie all'emissione di fluorescenza provocata da un laser. Le informazioni sono poi integrate e trasformate in picchi colorati (elettroferogramma).

Una volta identificata la sequenza di HPV, il genotipo può essere dedotto in due modi. Il metodo più veloce consiste nel consultare un database di sequenze con cui ricercare l'omologia tra la sequenza trovata e quelle note. Il più noto esempio di database presente in Internet è BLAST (<http://www.ncbi.nlm.nih.gov>). L'altro metodo consiste invece in un'analisi filogenetica in cui la sequenza trovata viene allineata con un certo numero di sequenze note, rappresentative dei diversi genotipi. Viene così costruito un albero filogenetico da cui è possibile dedurre il genotipo in base alle relazioni evolutive tra nuova sequenza e sequenze note.

3.4. Nuovi metodi di sequenziamento: pirosequenziamento e *ultradeep sequencing*

Il pirosequenziamento è un metodo bioluminometrico, alternativo al sequenziamento per elettroforesi (Ronaghi *et al.*, 1998). Esso si basa su una cascata di reazioni enzimatiche accoppiate, che utilizzano DNA polimerasi, ATP sulfotransferasi e luciferasi per monitorare la sintesi del DNA dosando il pirofosfato liberato in seguito all'attacco di un dNTP al filamento polimerizzato. Rispetto al sequenziamento classico tale tecnica presenta maggior accuratezza, flessibilità ed evita l'uso di primer e nucleotidi marcati e l'elettroforesi su gel.

Già nel 2001 la metodica è stata applicata da alcuni autori alla tipizzazione di HPV (Gharizadeh *et al.*, 2001) per sequenziare amplificati di 20-40 pb (GP5+/6+). La comparazione con il sequenziamento classico ha dato risultati concordanti al 100%. Il pirosequenziamento consente una genotipizzazione veloce ed efficiente; permette inoltre di identificare nuovi tipi di HPV, che con le metodiche di ibridazione non vengono riconosciuti o danno reazioni crociate, e di identificare varianti di sequenza dovute a mutazioni. Secondo gli stessi autori, i vantaggi del pirosequenziamento sono maggiormente apprezzabili quando si utilizza per il sequenziamento un pool di primer multipli tipo-specifico (Gharizadeh *et al.*, 2005). Con tale approccio è possibile: a) rilevare le infezioni multiple e discriminare i genotipi dominanti e sub dominanti al loro intero; b) individuare genotipi di interesse in amplificati contenenti prodotti di amplificazione non specifici; c) rilevare specie/tipi/DNA-bersaglio di interesse con primer di sequenza specifici in amplificati ottenuti con primer degenerati o con PCR multiplex; d) rilevare amplificati con bassa resa la cui PCR normalmente andrebbe ripetuta.

Un sistema di pirosequenziamento è il 454 *deep sequencing system* (Life Science, Roche), che attua un sequenziamento massivo in parallelo high-throughput del DNA, usando un nuovo approccio di sequenziamento tramite sintesi e che sequenzia fino a 10Mbp di genoma. Tale sistema e le sue applicazioni vengono descritte più approfonditamente nei successivi paragrafi in questo lavoro di tesi.

4. Utilità clinica ed epidemiologica della genotipizzazione di HPV

Lo studio del genoma dei papillomavirus è fondamentale per la classificazione del virus e per l'analisi delle correlazioni filogenetiche tra i diversi tipi virali. L'evoluzione filogenetica del virus ha sicuramente ripercussioni sull'evoluzione delle capacità biologiche dei diversi tipi virali. Essi sembrano essersi evoluti per occupare nicchie biologiche differenti, diventando specie specifici, acquisendo la capacità di infettare diversi distretti anatomici, pur essendo tutti epiteliotropi. Per capire meglio come questo sia possibile è fondamentale lo studio dei meccanismi molecolari e di come i diversi tipi virali interagiscano con la cellula. In tal modo è possibile capire come le differenze nel genoma si esplichino sui meccanismi virali e come alcuni virus filogeneticamente distanti possano dare lesioni analoghe e tipi vicini lesioni differenti (Doorbar, 2005).

La genotipizzazione ha però anche un'enorme importanza clinica ed epidemiologica.

Un'infezione persistente da parte di HPV ad alto rischio è riconosciuta come causa necessaria di carcinoma cervicale. Con l'implementazione di un programma di vaccinazione profilattica contro HPV, si è reso necessario un metodo accurato di tipizzazione per valutare l'efficacia del vaccino e per monitorare la distribuzione dei tipi di HPV nella popolazione generale e nelle

coorti vaccinate (Ferguson et al., 2009). Inoltre, la tipizzazione di HPV è stata introdotta nel programma di screening del carcinoma cervicale come second line test per il management di donne con infezione da HPV ad alto rischio (Barzon et al., 2008; Saslow et al., 2012).

In tutto il mondo, la prevalenza di infezione da HPV nella donna varia dal 2% al 44%. Tale variabilità è attribuibile a differenze nell'età della popolazione studiata, area geografica e sensibilità del test utilizzato per ricercare l'infezione da HPV. Si stima però che più del 50% delle donne sessualmente attive siano entrate in contatto con uno o più tipi virali, mentre poco si conosce sull'infezione nell'uomo (Baseman *et al.*, 2005). Dato che il potenziale oncogeno varia tra differenti tipi di HPV ad alto rischio, la genotipizzazione di HPV è da considerare per il management di donne con infezioni da HPV ad alto rischio. La definizione del potenziale oncogeno di un singolo tipo virale si basa principalmente su evidenze epidemiologiche di associazione tra l'identificazione di HPV DNA e carcinoma invasivo (Bouvard et al., 2009; Guan et al., 2012) e dipende dall'accuratezza dei metodi di identificazione e genotipizzazione di HPV. A questo proposito, il WHO Global Proficiency Studies ha dimostrato la necessità di migliorare i metodi per una genotipizzazione di HPV affidabile, dato che molti risultati falsi positivi e falsi-negativi sono stati riportati dai laboratori partecipanti (Eklund et al., 2010; de Sanjose et al., 2010). Recentemente un ampio studio sul carcinoma cervicale invasivo ha sottolineato l'importanza di un'accurata identificazione e tipizzazione di HPV ad ampio spettro. In tale studio, dei 10.575 campioni di cervico carcinoma invasivo complessivamente raccolti, è stato possibile identificare la presenza di HPV DNA in 84.9% dei campioni e il tipo di HPV è stato identificato in 96.3% dei casi positivi (de Sanjose et al., 2010). Ulteriori analisi dei carcinoma invasive HPV-positivi con unidentified HPV types, utilizzando un nuovo primer set per la genotipizzazione, ha dimostrato la presenza di tipi di HPV possibili carcinogeni gruppo 2B IARC e rare varianti di tipi di HPV carcinogeni che vengono identificati infrequentemente o mai identificati nei carcinomi invasivi e che non sono rilevati dai test di genotipizzazione basati sulla ibridazione comunemente usati (Geraets et al., 2012).

È molto importante quindi utilizzare un test sensibile, per evitare di sottostimare il problema e poter conoscere la reale prevalenza ed incidenza dei diversi genotipi nelle diverse aree geografiche. È noto infatti che HPV-16 è il più comune tipo virale nel cancro della cervice in tutto il mondo, seguito da HPV-18, -45, -31, -33; mentre la prevalenza degli altri tipi ad alto rischio varia nelle diverse aree geografiche, ad esempio HPV-58 e -52 sono i tipi più comuni dopo il -16 e il -18 in Asia, mentre sono più rari in altre aree geografiche (Bosch *et al.*, 2008; Giuliano *et al.*, 2008; De Vuyst *et al.*, 2009; Dunne *et al.*, 2006; Meijer *et al.*, 2006). Un altro dato importante è quello riguardante le coinfezioni che risultano più frequenti in alcune popolazioni a

rischio (soggetti con infezione da HIV). Queste non devono essere sottostimate poiché sembrano esporre ad un maggior rischio di sviluppare tumore in tempi brevi.

Tutti questi dati diventano ancora più importanti in epoca vaccinale per poter decidere chi vaccinare, verso quali tipi virali deve essere protettiva la vaccinazione e quali tipi virali deve ricercare un test di screening. Tali decisioni potrebbero essere differenti in diverse aree geografiche. Informazioni sull'infezione sia negli uomini che nelle donne sarebbero utili per valutare il potenziale impatto clinico del vaccino profilattico contro HPV.

La genotipizzazione di HPV infine diventata essenziale nello screening e nella gestione del carcinoma del collo dell'utero e dei suoi precursori (CIN 2 e 3). I test per la ricerca degli HPV oncogeni si sono dimostrati più sensibili e con un valore predittivo negativo (VPN) più alto nel rilevare un CIN2 rispetto alla citologia cervicale, con una minima riduzione della specificità e del valore predittivo positivo (VPP). Inoltre consentono di rilevare un end point precoce nello sviluppo delle anomalie cervicali e di allungare i tempi di follow up nelle donne con test negativo (Meijer *et al.*, 2006). Come riportato (Huhet *al.*, 2010), HPV 16/18 sono responsabili del 77% dei carcinomi cervicali e del 54% delle lesioni di alto grado negli Stati Uniti in epoca post vaccinale. È plausibile attendersi una riduzione dei casi di Atypical Squamous cells of undetermined significance (ASCUS), low-grade e high-grade squamous intraepithelial lesions (LSIL and HSIL) così come del numero di soggetti inviati alla colposcopia. Meta-analisi stimano una riduzione nei LSIL del 40% per le donne vaccinate per HPV tipo 16/18 e del 50% per quelle vaccinate per HPV types 6/11/16/18 (Clifford *et al.*, 2005). Queste proporzioni sono probabilmente destinate ad aumentare con l'adozione nella pratica clinica dei vaccini multivalenti. Con l'incremento della popolazione vaccinata, la prevalenza di anomalie cervicali continuerà a diminuire influenzando in maniera avversa il valore predittivo positivo (VPP) di tutti i test di screening per il cervicocarcinoma. Il test HPV è probabilmente quello che soffrirà meno di questo effetto rispetto al pap test. L'incrementata proporzione di popolazione vaccinata determina anche un declino della specificità del Pap test. Una strategia più sensibile per lo screening del cervico carcinoma, suggerita dalla Società di Ginecologia oncologica (Society of Gynecologic Oncologists; SGO), consiste nell'utilizzare un test ad alta sensibilità HPV DNA test come test di screening di primo livello e il Pap test come triage di secondo livello in donne risultate positive al test HPV DNA. Questa strategia di triage per lo screening del cervico carcinoma può giocare anche un importante ruolo nella sorveglianza post vaccinale. Nell'epoca del vaccino, l'HPV test può essere considerato utile per stabilire la prevalenza di specifici tipi di HPV e per determinare genotipi nascosti (ad es., un incremento nella prevalenza di tipi di HPV non inclusi nel vaccino) nei registri e negli studi di popolazione. Nelle donne vaccinate, test tipo

specifico possono anche permettere più agevolmente di valutare le infezioni persistenti e stimare il loro contributo nello sviluppo della neoplasia intraepiteliale cervicale (CIN) (Huh et al., 2010)

La tipizzazione virale consente di individuare le donne con rischio oncogeno più elevato, quali quelle con infezioni da HPV-16 e -18. Alcune varianti intratipo asiatiche-americane ed alcune non europee di tali tipi virali sembrano dare più frequentemente infezioni persistenti e quindi esporre ad un maggiore rischio di sviluppare il cancro (Berumen *et al.*, 2001; Sichiero *et al.*, 2007). Questo sembra dimostrare che la genotipizzazione è utile non solo per studi epidemiologici ma anche per fini prognostici.

In conclusione la genotipizzazione nello studio del cancro della cervice uterina è utile per fornire ulteriori dati sulla correlazione causale cancro-infezione virale, sugli HPV oncogeni e sulle modificazioni cellulari che essi inducono. Sulla popolazione vaccinata permette di monitorare la durata e l'efficacia della protezione ed è utile nello stratificare le donne HPV positive in base al rischio di sviluppare lesioni precancerose o cancerose e di guidare quindi le decisioni cliniche (Meijer *et al.*, 2006).

Come evidenziato nel paragrafo 3 i test di genotipizzazione in uso oggi differiscono nelle loro capacità analitiche per sensibilità e specificità tipo specifico e molti studi hanno comparato differenti metodi di tipizzazione di HPV usando diversi campioni clinici. Usando un metodo standardizzato, il 2010 WHO Global Proficiency studies (Eklund et al., 2012) ha dimostrato la necessità di migliorare i metodi per una tipizzazione di HPV più affidabile e ha rivelato alcuni limiti della maggior parte dei test di tipizzazione di HPV usati dai laboratori partecipanti allo studio. Per esempio vari test di tipizzazione sottostimano la prevalenza di HPV-39, -56 e -59. La corretta tipizzazione di campioni infettati da multiple HPV types è riportata in questo studio dal 44% al 78% dei data sets, in comparazione con una media di 86% quando solo un tipo di HPV era presente nel campione. La sottostima della prevalenza di infezioni multiple introduce un bias sistematico di individuazione negli studi epidemiologici e questo comporta un problema nella sorveglianza delle vaccinazioni. Eklund et al. (2012) riportano una sistematica falsa positività nel test INNO-LiPA per HPV-52 identificato in campioni positivi per HPV-58. Linear Array non esclude HPV-52 quando il campione è positivo per HPV-33, -35, -58 e riporta falsi positivi per HPV-66. INNO-LiPA riporta HPV-52 in campioni con plasmidi contenenti HPV-58 o HPV-68b.

In epoca vaccinale, è necessaria un'alta sensibilità analitica, poiché un fallimento nell'identificare le infezioni prevalenti all'ingresso in un trial può comportare un falso risultato di fallimento del vaccino nei protocolli vaccinali.

5. Tecnologie di Next-generation sequencing nella diagnostica virologica

Le nuove tecniche di sequenziamento del DNA, dette anche “di nuova generazione” o “next-generation” sequencing (NGS), sono caratterizzate da alta velocità ed elevato “throughput”, potendo produrre un enorme volume di sequenze rapidamente e ad un costo relativamente basso. Il più importante vantaggio di queste piattaforme è che riescono a determinare la sequenza di un singolo frammento di DNA di una libreria, segregato in un chip, evitando di doverlo clonare in un vettore prima di poter acquisire la sequenza.

La prima tecnologia di NGS high-throughput sequencing, sviluppata da 454 Life Sciences nel 2005, è stata la piattaforma 454 FLX pyrosequencing (<http://www.454.com/>). Successivamente sono state sviluppate e immesse in commercio piattaforme sempre nuove e migliorate con differenti biochimiche, protocolli di sequenziamento, throughput e differente lunghezza della sequenza ottenuta.

Come conseguenza di tali differenze ogni tecnologia può essere più idonea per diverse applicazioni. Ad esempio la piattaforma 454 è particolarmente adatta per il *de novo assembly* e per il sequenziamento di ampliconi lunghi (*long length amplicons sequencing*).

Le tecnologie NGS vengono comunemente usate per sequenziamento *whole genome*, per studi di diversità genomica, metagenomica, epigenetica, scoperta di RNA non codificanti e *protein-binding sites*, e *gene-expression profiling* tramite il sequenziamento dell'RNA. Come riportato in Barzon et al. (2011), le applicazioni delle metodiche di NGS in virologia, oltre all'high-throughput whole genome sequencing, sono:

1) scoperta di nuovi microorganismi e virus utilizzando un approccio di metagenomica.

Le tecniche NGS *high throughput* possono essere utilizzate attraverso strategie di metagenomica per identificare virus associati a malattie non note o per scoprire nuovi virus umani. Questi metodi offrono il vantaggio di avere alta sensibilità e potenzialmente sono in grado di identificare tutto lo spettro dei virus inclusi virus sconosciuti o inaspettati. Ad esempio differenti strategie sono state utilizzate per lo studio di casistiche di pazienti durante epidemie stagionali di influenza e durante *outbreaks* di norovirus, per l'identificazione di arenavirus correlati con lymphocytic choriomeningitis viruses e per identificare gli arenavirus associati alla febbre emorragica del Sud Africa (Lujo virus). Tali metodiche hanno permesso inoltre l'identificazione e il sequenziamento dell'intero genoma di nuovi papillomavirus umani e polyomavirus, inclusi i polyomavirus umani 6 e 7 (HPyV6 and HpyV7) ed hanno consentito la scoperta dei nuovi

polyomavirus associati con Merkel cell carcinoma (MCC) e di altri *Polyomaviridae* e *Papillomaviridae* associati a tumori.

2) valutazione di comunità virali ambientali e di nicchie del corpo umano confrontate tra soggetti sani e soggetti in condizioni patologiche.

Il microbioma umano è l'intera popolazione di microbi (batteri, funghi, virus) che colonizza il corpo umano. Il termine metagenomica si riferisce a studi non dipendenti da coltura, dell'intera raccolta di genomi di una comunità microbica mista e si applica allo studio di tutti i genomi microbici che risiedono in nicchie ambientali, piante e ospiti animali, incluso il corpo umano.

Non esistono sequenze virali ubiquitariamente conservate che consentano un'amplificazione ad ampio spettro dei genomi virali, come è possibile fare per batteri e funghi, ed è indispensabile quindi utilizzare metodi per l'arricchimento dei campioni contenenti particelle virali. Le analisi di metagenomica condotte sui virus, che sono state applicate in particolare in campioni ambientali come acqua dolce, acque reflue, e acque dell'oceano, hanno individuato molte sequenze virali uniche e specie virali non note.

3) analisi della variabilità del genoma virale nello stesso ospite (*quasispecie*).

L'alta percentuale di mutazione che caratterizza la replicazione dei virus a RNA crea un'ampia varietà di mutanti che sono presenti nelle popolazioni virali e che vengono definite *quasispecie*. Questa caratteristica permette alla popolazione virale di adattarsi rapidamente ai cambiamenti replicativi dell'ambiente, selezionando quelle varianti preesistenti con il migliore fitness. Per questa ragione molte importanti proprietà del virus non possono essere spiegate da una sola sequenza *consensus*, ma rendono necessaria la conoscenza delle varianti presenti nella popolazione virale. Queste varianti di sequenza possono essere criticamente rilevanti per l'evoluzione del virus e la sua diffusione, la virulenza, l'evasione dalla risposta immunitaria, la resistenza ai farmaci antivirali e per lo sviluppo e produzione di un vaccino.

Tra i virus a RNA, le *quasispecie* di HIV sono state investigate approfonditamente per la loro rilevanza al fine di disegnare un vaccino e per valutare la risposta ai farmaci antivirali. Le analisi dell'intero genoma virale e delle *quasispecie* sono state applicate ad altri virus ad RNA, ad esempio nello studio della polmonite virale dovuta alla pandemia di influenza A del 2009 (A/H1N1/2009) o del rhinovirus umano (HRV).

Le popolazioni dei virus a DNA sono considerate meno complesse e variabili paragonate con quelle dei virus a RNA. Tuttavia, studi di deep sequencing di virus con genoma a DNA hanno rilevato che nei soggetti infettati può essere presente una complessa mistura di genotipi virali e

una selezione positiva può aver contribuito alla divergenza dei diversi ceppi di uno stesso virus. Questo è il caso del citomegalovirus umano (HCMV), che stabilisce un'infezione latente nell'uomo per tutta la vita e può riattivarsi e dare patologie severe nei soggetti immunocompromessi. Una alta variabilità intra ospite del genoma di HCMV è stata dimostrata nei soggetti trapiantati di polmone e nei neonati con infezione congenita da HCMV.

Il deep sequencing ha dimostrato anche variabilità nel genoma di herpes simplex virus 1 (HSV-1) e può essere utile per valutare geni di virulenza di questo virus.

4) ricerca di mutazioni minoritarie che inducono resistenza ai farmaci antivirali in pazienti con infezione virale (infezioni da HIV e da epatite virale).

Recentemente è stato incrementato l'uso nella pratica clinica di tecniche di Deep sequencing tramite NGS per individuare le varianti resistenti di HIV presenti in basso numero di copie ed anche per individuare le varianti minoritarie di HCV.

Il convenzionale sequenziamento diretto di prodotti di RT-PCR (ovvero “*population sequencing*”) è il gold standard come test di resistenza di HIV. Il maggior limite del sequenziamento diretto di PCR, è la sua incapacità di individuare le varianti resistenti presenti in percentuale inferiore al 20-25% della popolazione virale eterogenea presente nel plasma di soggetti infetti. Molti studi hanno dimostrato che le varianti resistenti minoritarie, che non vengono identificate dal sequenziamento di popolazione, sono clinicamente rilevanti in quanto responsabili del fallimento virologico di un nuovo regime di trattamento antiretrovirale.

Il sequenziamento clonale di prodotti di RT-PCR tramite 454 pyrosequencing offre il vantaggio di un'alta sensibilità per le varianti minoritarie e sequenze relativamente lunghe, facilitando la caratterizzazione del legame tra mutazioni resistenti. Evita inoltre il rischio di perdere mutazioni dovute a variazioni di sequenza circostanti il sito in esame. Diversi studi condotti impiegando il 454 pyrosequencing per l'analisi di mutazioni nel gene della proteasi di HIV, transcriptasi inversa e recentemente integrasi, hanno dimostrato l'accuratezza di questa tecnica nell'identificare tutte le mutazioni che inducono resistenza identificate dal sequenziamento di popolazione ed inoltre la capacità di rilevare anche le mutazioni a bassa frequenza non identificabili con il sequenziamento di popolazione. Inoltre vari studi dimostrano che le mutazioni di resistenza individuate tramite 454 hanno un significativo impatto nel fallimento virologico, mentre altre metodiche non hanno trovato una forte associazione tra mutazioni a bassa frequenza e risposta clinica.

Deep sequencing basato sulla tecnologia 454 è stato anche utilizzato per identificare le resistenze agli inibitori nucleosidici e nucleotidici della trascrittasi inversa in infezioni da HBV. La

metodica NGS è più sensibile per l'identificazione di rare mutazioni che provocano resistenza ai farmaci rispetto ai metodi convenzionali basati su sequenziamento e ibridazione inversa.

Infine, con la disponibilità di nuovi farmaci aventi come bersaglio la proteasi o la polimerasi di HCV, le esperienze acquisite per HIV nell'analisi di mutazioni che inducono resistenza e nelle analisi delle *quasispecie* potranno essere traslate allo studio dell'infezione da HCV.

La diagnostica virologica e gli studi di metagenomica costituiscono le applicazioni più apprezzabili della tecnologia NGS e con la riduzione dei costi e tempi di applicazione, questi metodi diventeranno essenziali per la diagnostica e la routine clinica.

OBBIETTIVI

Lo scopo di questo studio è sviluppare e validare un nuovo metodo per l'identificazione e genotipizzazione di HPV in campioni clinici sfruttando la tecnologia 454 next-generation sequencing (NGS).

Questo nuovo protocollo di genotipizzazione, paragonato ai test in commercio, dovrebbe essere estremamente sensibile, tipo specifico, quantitativo, e capace di discriminare tutti i tipi virali presenti nelle coinfezioni, eventuali varianti virali, *quasispecie* e nuovi tipi non noti.

Queste caratteristiche renderebbero tale metodo estremamente utile per diverse applicazioni, da un'accurata analisi delle infezioni nei campioni di carcinoma e nelle lesioni preneoplastiche, a studi epidemiologici nella popolazione generale e in particolari sottogruppi di popolazione, e ogni applicazione che richieda accuratezza nel metodo di sequenziamento e sensibilità per le infezioni multiple.

Per il raggiungimento di questo obiettivo è stato sviluppato un primo protocollo che utilizza fusion primer costruiti modificando i primer degenerati MY09/MY11 per creare una libreria di ampliconi. Questo protocollo è stato utilizzato per l'analisi di 164 campioni citologici cervicali positivi per HPV; tale libreria è stata sequenziata in una reazione *multiplex* su piattaforma 454 Life-Sciences. I risultati sono stati confrontati con quelli del sequenziamento Sanger e del test InnoLiPA (Barzon et al., J Clin Virol 2011).

Questa parte di studio costituisce una prima dimostrazione che l'NGS può essere usato per la tipizzazione di HPV ed è utile per identificare sottotipi virali e varianti. Il protocollo presentava però alcuni limiti legati soprattutto ad alcune caratteristiche dei primer utilizzati per costruire la libreria di ampliconi. È stato pertanto sviluppato e testato un secondo protocollo che utilizza il set di primer consenso PGMY09/11 che sono stati disegnati per eliminare la degenerazione e aumentare la sensibilità, specificità e riproducibilità dei primer MY09/11 (Gravitt et al., 2000). Per la validazione sono state seguite le linee guida e i protocolli del WHO (Human papilloma virus laboratori manual – First edition, 2009). Le performance analitiche del nuovo protocollo sono state testate su controlli e campioni clinici.

METODI

1. Genotipizzazione di HPV mediante 454 ultradeep sequencing

Generalità sul metodo di pirosequenziamento con tecnologia 454 Life Sciences.

Il sistema di pirosequenziamento 454 Life Sciences (Roche), consente di effettuare *deep sequencing* ovvero sequenziamento massivo in parallelo high-throughput di ampliconi di DNA.

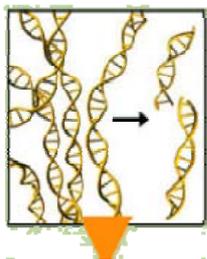
Flusso di lavoro del sistema: One Fragment = One Bead = One Read

Il completo workflow di sequenziamento del 454 Genome Sequencing System comprende quattro fasi principali, che vanno dalla purificazione del DNA all'analisi dei risultati.

Tali fasi includono:

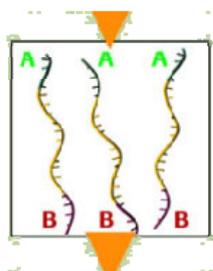
- Generazione di una libreria di DNA template a singolo filamento
- Un'amplificazione clonale della libreria tramite *emulsion-based PCR*
- Generazione dei dati per sequenziamento tramite sintesi
- Analisi dei dati utilizzando differenti strumenti bioinformatici

Campioni utilizzati e frammentazione



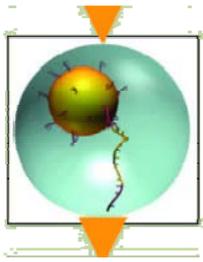
Il sequenziatore genomico 454 Genome Sequencing System supporta il sequenziamento di campioni partendo da una ampia varietà di materiali incluso DNA genomico, prodotti di PCR, BACs e cDNA. Campioni come DNA genomico e BACs devono essere frammentati in piccoli segmenti da 300 a 800 paia di basi (Pb). Per campioni più piccoli, come small non-coding RNA o ampliconi di PCR la frammentazione non è richiesta. Invece i prodotti di PCR devono essere amplificati usando Genome Sequencer fusion primers che servono poi per immobilizzare il DNA sulle biglie di cattura come mostrato sotto al paragrafo "*One Fragment = One Bead*".

Preparazione di una libreria



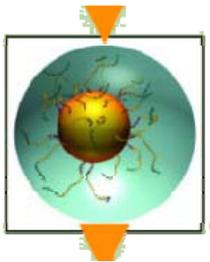
Usando una serie di tecniche standard di biologia molecolare, dei corti adattatori (A e B), specifici per l'estremità 3' e 5', vengono aggiunti ad ogni frammento. Gli adattatori vengono usati per gli step di purificazione, amplificazione e sequenziamento. I frammenti a singolo filamento con gli adattatori A e B compongono la libreria campione usata per i successivi step.

Un frammento = una biglia



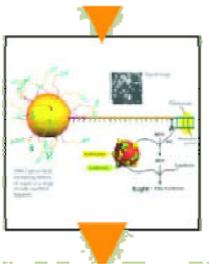
La libreria di DNA a singolo filamento è immobilizzata su biglie di cattura del DNA appositamente disegnate. Ogni biglia carica un unico frammento di DNA a singolo filamento della libreria. Le biglie leganti la libreria sono emulsionate con i reagenti di amplificazione in una miscela olio in acqua che forma un microreattore contenente solo una biglia con un unico frammento della libreria.

Amplificazione con emPCR (emulsion PCR)

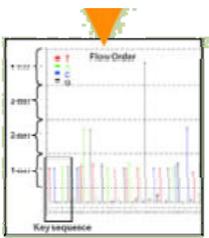


Ogni unico frammento della libreria è amplificato dentro un unico microreattore, escludendo competizioni o contaminazioni di sequenze. L'amplificazione dell'intera collezione di frammenti avviene in parallelo; ogni frammento, è ottenuto in un numero di copie di vari milioni per biglia. Successivamente, la emPCR si interrompe e i frammenti amplificati rimangono legati alla specifica biglia.

One Bead = One Read



I frammenti amplificati clonalmente sono arricchiti e caricati in un PicoTiterPlate (PTP) appositato per il sequenziamento. Il diametro del PTP contiene solo una biglia per pozzetto. Dopo l'aggiunta degli enzimi di sequenziamento, il subsistema fluidico del Genome Sequencer FLX Instrument produce un flusso individuale di nucleotidi in un ordine fisso attraverso le centinaia di migliaia di pozzetti contenenti ognuno una biglia. L'aggiunta di uno (o più) nucleotidi complementari al filamento template produce un segnale chemiluminescente registrato dalla CCD camera dello strumento Genome Sequencer FLX.



Analisi dei dati

La combinazione dell'intensità del segnale e delle informazioni posizionali generate attraverso il sistema PTP permette al software di

determinare la sequenza di più di 1.000.000 singole reads simultaneamente in 10 ore di corsa dello strumento. Per l'analisi dei dati di sequenziamento sono disponibili diversi tools bioinformatici che supportano le diverse applicazioni: *de novo assembly*; *resequencing* di genomi; e l'individuazione di varianti di ampliconi per confronto con le sequenze di riferimento note.

Chimica del pirosequenziamento 454 Life Sciences

Il sistema di Sequenziamento 454 Genome Sequencer FLX si basa sul sequenziamento-tramite-sintesi (Fig. 3):

- I nucleotide fluiscono in un ordine sequenziale fisso attraverso il sistema PicoTiterPlate durante la corsa di sequenziamento.
- Durante il flusso di nucleotidi, centinaia di migliaia di biglie leganti ciascuna milioni di copie di un unico singolo filamento di DNA sono sequenziate in parallelo.
- Se un nucleotide complementare al filamento template fluisce nel pozzetto, la polimerasi allunga l'esistente filamento di DNA aggiungendo il o i nucleotidi.
- L'aggiunta di uno o più nucleotidi risulta in una reazione che genera un segnale luminoso che viene registrato da una CCD camera nello strumento.
- La forza del segnale è proporzionale al numero di nucleotidi incorporati in un singolo flusso di nucleotidi.

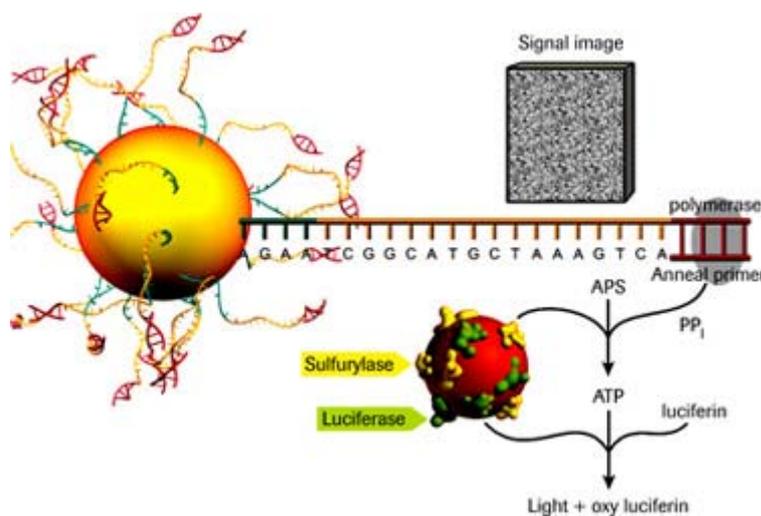


Fig. 3: reazione di sequenziamento del Genome Sequencer System. Milioni di copie di un singolo frammento clonale è contenuto in ciascuna biglia di cattura del DNA.

Processazione dell'immagine

Il software GS FLX System registra la posizione delle biglie leganti il DNA su un asse XY. Ogni biglia corrisponde ad una coordinata XY in una serie di immagini. L'intensità di segnale per flusso di nucleotide è registrata per ogni biglia tutto il tempo ed è tracciata a generare un grafico di flusso. Ogni 10 ore di corsa di sequenza nel GS FLX Titanium series può produrre più di un milione di diagrammi di flusso (*flowgrams*). (Fig. 4)

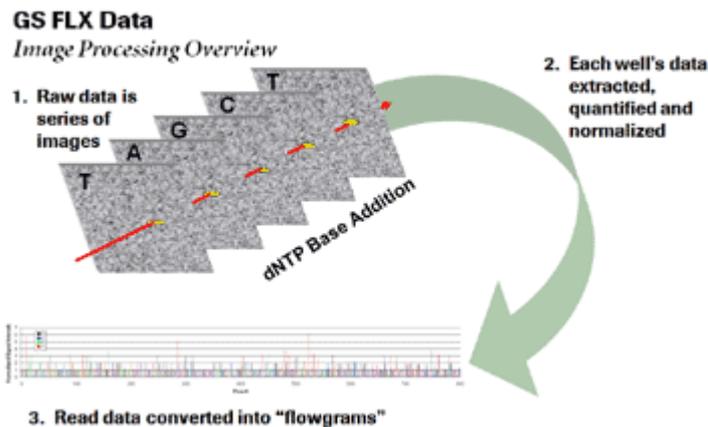


Fig. 4: sistema di processazione dell'immagine GS FLX.

Preparazione della libreria di ampliconi

Disegno dei Fusion Primers. I fusion primer devono essere disegnati e sintetizzati secondo regole precise, in modo da essere adeguati per l'esperimento con il sequenziatore 454. I primers devono avere alle estremità una sequenza A ed una sequenza B in grado di legare le biglie di sequenziamento ed i primer di amplificazione e sequenziamento (vedi Fig. 5). Le estremità 5' dei fusion primers sono sempre identiche secondo le caratteristiche richieste dal sistema di sequenza:

- Lunghezza: 19 nt
- Sequenza:
Primer A: 5' GCCTCCCTCGCCAT**TCAG** 3'
Primer B: 5' GCCTTGCCAGCCCG**TCAG** 3'.

Essi hanno la funzione di:

1. legare le biglie di cattura del GS emPCR Kit,
2. legare i primer di sequenza del GS emPCR Kit: il primer A lega il primer di sequenza di GS emPCR kitII, il primer B lega il primer di sequenza di GS emPCR kit III.

3. l'estremità con la chiave di sequenza "TCAG" è usata dal software come base di richiamo per riconoscere le reads della libreria.

La seconda parte di ogni fusion primer è invece specifica per ogni amplicone e nel nostro studio è costituita dai primers MY09 ed MY11 nel primo protocollo sviluppato e da PGMY11/09 modificati nel secondo protocollo.

Nel disegno dei primer vanno considerati i normali vincoli di specificità e le condizioni di annealing, tenendo conto che le sequenze A e B durante la reazione rimangono non appaiate e quindi non dovrebbero influire sulla reazione. Inoltre la distanza tra i due primer sulla sequenza templato non deve superare 500 bp.

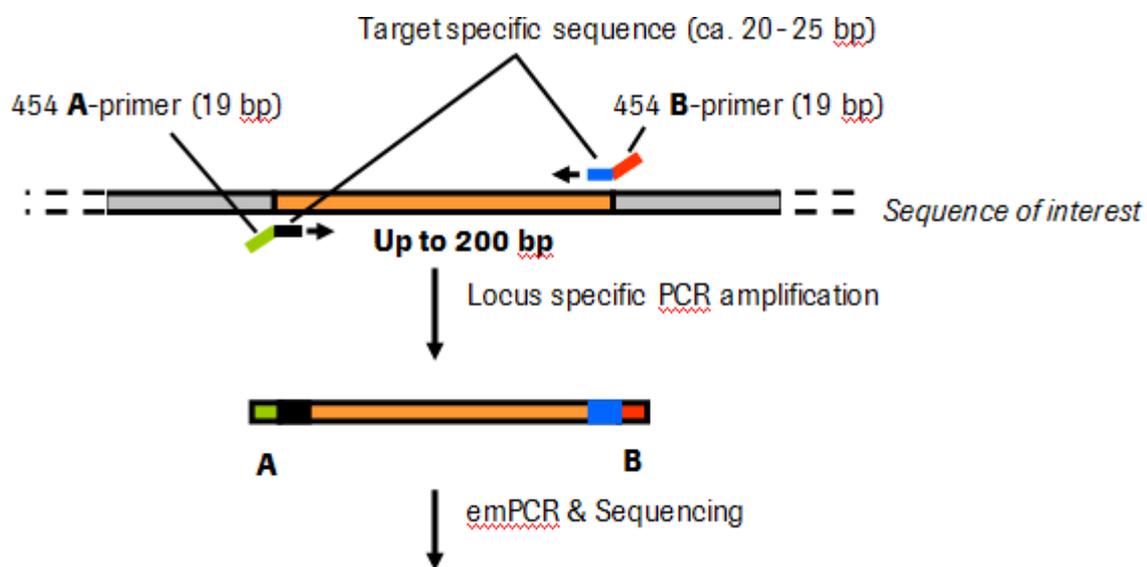


Fig. 5: la preparazione del DNA campione per il sequenziamento di Ampliconi con il GS FLX System consiste di una semplice reazione di amplificazione (PCR) con specifici Fusion Primers. I Fusion Primer sono costituiti da una sequenza target specifica di 20-25 bp (3' end) e una sequenza fissa di 19 bp Multiplexing sequence (Primer A o Primer B in 5').

Multiplex Identifiers (MIDs). A ciascuna coppia di primers è stata poi legato uno specifico tag (Ligation Multiplex Identifiers, MID, adaptors) ovvero una sequenza che consenta il riconoscimento del singolo amplicato ottenendo così 12 coppie di primers con tag diverso. Ogni MID contiene una sequenza di 10 nucleotidi e consente di mescolare 12 ampliconi diversi e di sequenziarli contemporaneamente in un singolo supporto. Il software AVA che analizza le sequenze, riconosce i MID come il vero inizio della lettura della sequenza, quindi ciascun MID deve essere posizionato tra il primer

specifico (in questo caso MY09 o MY11) e la sequenza chiave, dopo gli adattatori A e B (Figg. 6-7).

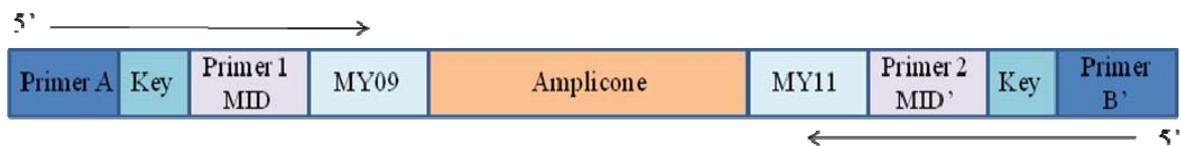


Fig. 6: schema dei Fusion primer; fusion primer disegnati per il primo protocollo dello studio.

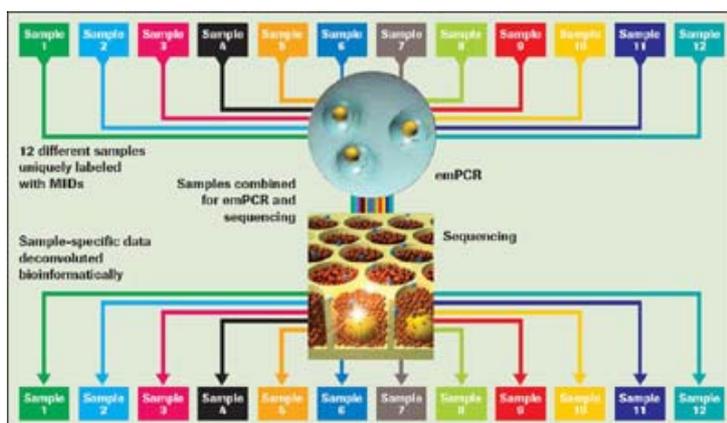


Fig.7: Ligation MID workflow. Usando il Ligation MIDs, 12 campioni differenti possono essere mescolati per l'emPCR e sequenziamento. Al termine del sequenziamento, il software GS FLX automaticamente identifica ogni lettura tramite la sequenza MID campione specifica.

Separazione fisica dei campioni

Il supporto PicoTiterPlate (PTP) su cui verranno caricati gli ampliconi può essere diviso fisicamente in regioni più piccole tramite un Gasket (1, 2, 4, 8, e 16 regioni). È così possibile caricare sequenze di campioni multipli contemporaneamente in una singola PTP mantenendo l'identità unica di ogni campione durante il sequenziamento e l'analisi dei dati.

Gasket Format (regions)	Reads/Region (x10 ³)
2	450-650
4	160-250
8	80-120
16	25-40



Fig. 8: opzioni di formato del Gasket e reads per regione per GS FLX Titanium series kits.

Nel primo protocollo si è scelto di dividere il PTP con un gasket in 16 regioni in modo tale da poter analizzare grazie a 12 coppie di primer con tag diversi, fino a 192 campioni simultaneamente. È stato quindi necessario assegnare a ciascuno dei 192 campioni una specifica coppia di primers e dividere i campioni in gruppi di 12. Nel secondo protocollo invece il PTP è stato diviso in 8 regioni, ottenendo un maggiore efficienza di sequenziamento per singolo campione, riducendo però il numero di campioni sequenziabili in parallelo in una corsa.

Estrazione di banda da gel. Gli amplificati sono stati caricati su gel di agarosio al 2% e la banda corrispondente a 450 pb è stata excisa con un bisturi sterile e purificata da gel con il kit Quiagen QIAquick® Gel Extraction Kit, come da protocollo, al fine di eliminare eventuali altre bande (frammenti di DNA, aspecifici, dimeri di primers) presenti e ottenere un amplificato più pulito. 5 µl di DNA così purificato sono stati ricaricati in gel per verificare di avere una quantità di DNA sufficiente per i successivi passaggi di purificazione e sequenziamento.

Purificazione degli amplificati mediante kit AMPure™. Gli amplificati vengono ulteriormente purificati tramite il kit AMPure_System, PCR Purification System™ (Beckman Coulter, Inc), secondo il protocollo per ampliconi >100 bp (Fig. 9). Tale passaggio ha lo scopo di eliminare eventuali residui di dNTPs non incorporati, sali, primers, DNA polimerasi, e altri contaminanti, ottenendo ampliconi altamente purificati. Il kit utilizza biglie paramagnetiche per legare i frammenti di DNA permettendo di recuperare più del 90% degli ampliconi di dimensioni superiori a 200 bp e più dell'85% degli ampliconi di dimensioni comprese tra 100 e 200 bp.

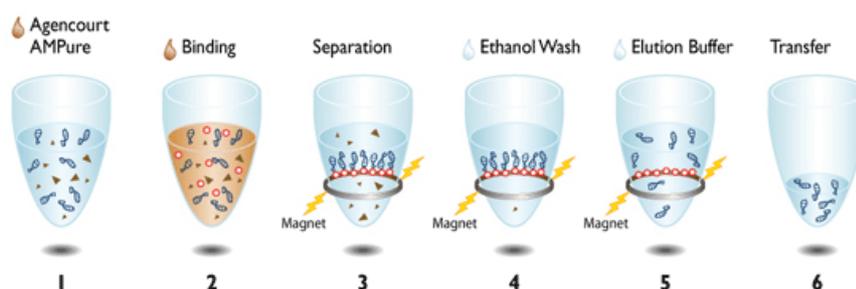


Fig. 9: 1. Reazione di PCR; 2. Legame degli ampliconi di PCR alle biglie magnetiche; 3. Separazione degli ampliconi legati alle biglie magnetiche dai contaminanti; 4. Lavaggio degli ampliconi con etanolo; 5. Eluizione degli ampliconi dalle particelle magnetiche; 6. Trasferimento dell'eluato in una eppendorf pulita per separarlo dalle biglie.

Analisi qualitativa degli ampliconi. La qualità e quantità degli amplificati è stata quindi valutata tramite Agilent 2100 Bioanalyzer. 1µl di ciascun campione della libreria di ampliconi viene fatto correre in un BioAnalyzer DNA 1000 LabChip. È così possibile:

- verificare che il prodotto di amplificazione sia delle dimensioni attese (in tal caso circa 506 bp), sapendo che la misura indicata dal BioAnalyzer solitamente si discosta da quella calcolata al massimo di poche bp;

- esaminare attentamente le tracce di prodotti estranei come dimeri di primer, di contaminazioni che potrebbero ridurre il numero di *reads* utili per il sequenziamento della libreria di ampliconi.

Qualora siano presenti dimeri di primer è necessario ripetere la purificazione con AMPure.

Quantificazione della libreria di ampliconi. La libreria deve quindi essere quantificata tramite fluorimetro usando il Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen), secondo le istruzioni del kit. Il fluorimetro è in grado di dare una misura molto precisa e riproducibile ed è importante una determinazione molto accurata della concentrazione individuale di ogni amplicone della libreria prima di mescolarli per l'emulsion PCR. In tal modo si assicura che ogni target sia ugualmente rappresentato nella reazione di sequenza.

Calcolo delle molecole e diluizioni. Dalla concentrazione della libreria espressa in ng/µl, bisogna calcolare l'equivalenza in molecole/µl usando l'equazione:

$$\text{Molecole}/\mu\text{l} = \frac{(\text{concentrazione del campione; ng}/\mu\text{l}) \times (6,022 \times 10^{23})}{(656,6 \times 10^9) \times (\text{lunghezza dell'amplicone; bp})}$$

Dove $6,022 \times 10^{23}$ è il numero di Avogadro (molecole/moli) e 656,6 è il peso molecolare medio delle coppie di nucleotidi in g/moli.

Quindi un'aliquota di stock di libreria concentrata deve essere diluita 2×10^5 molecole/µl in buffer TE. Lo stock di lavoro va poi suddiviso in aliquote da 50 µl conservate insieme allo stock concentrato a -15/-25°C.

Gli ampliconi così preparati sono pronti per l'*emulsion PCR* (emPCR) e per la corsa di sequenziamento.

2.4.5. GS FLX emPCR

Nel presente lavoro sono stati utilizzati il GS emPCR Kit II per il sequenziamento dal Primer A, e il GS emPCR Kit III per il sequenziamento dal Primer B, in modo da ottenere sia la sequenza forward che reverse. Il protocollo usato è lo Small emPCR Procedure.

Il processo di emPCR per small preparations prevede 7 step principali, che devono essere condotti in due aree dedicate e separate del laboratorio: una stanza controllata e una stanza di amplificazione (*Controlled Room* e *Amplicon Room*). Gli step che vanno condotti nella *Controlled Room* consistono in:

- 1 Preparazione della *live amplification mix*;
- 2 Cattura della libreria di DNA;
- 3 Preparazione dell'emulsione;
- 4 Dispensazione dell'emulsione

Gli step che avvengono nella *Amplicon Room* consistono in:

- 5 Reazione di amplificazione
- 6 Recupero delle biglie
- 7 Arricchimento delle *DNA library bead*
- 8 Annealing dei primer di sequenza.

1 Preparazione della *live amplification mix*: la quantità di *live amplification mix* richiesta per l'esperimento è determinata dal numero di reazioni di emulsione/amplificazione (in questo caso 16) e preparata come descritto nei protocolli.

2 Cattura della libreria di DNA: i frammenti che compongono la libreria di DNA vengono immobilizzati sulle biglie di cattura, tenendo conto della loro segregazione nell'emulsione. L'obiettivo è di ottenere una molecola effettiva (un clone funzionale) di DNA per biglia e una biglia per microreattore acqueo (micella), isolato dalle altre biglie grazie all'olio circostante. È la segregazione di ogni biglia nel proprio microreattore (con i reagenti di amplificazione) che mantiene la clonalità durante la fase di amplificazione.

Il processo di cattura comporta l'ibridazione dell'adattatore (che è stato legato ad ogni molecola di templatato durante la preparazione della libreria) agli oligonucleotidi complementari che sono covalentemente attaccati alle biglie di cattura del DNA:

_ le biglie di cattura del DNA del GS emPCR Kit II legano il Primer "B" localizzato ad una estremità di ogni frammento della libreria di ampliconi (quindi qui il sequenziamento partirà dal Primer "A").

_ le biglie di cattura del DNA del GS emPCR Kit III legano il Primer "A" localizzato ad un'estremità di ogni frammento della libreria di ampliconi (quindi qui il sequenziamento partirà dal Primer "B").

_ la libreria di ampliconi è solo aggiunta alle biglie in questa fase, la cattura fisica avviene solo durante la reazione di amplificazione.

3 Preparazione dell'emulsione: la libreria di DNA catturata è risospesa nella mix di amplificazione e olio, per formare una miscela olio in acqua. Questo passaggio richiede un vigoroso “shaking” meccanico a condizioni strettamente controllate. Il prodotto di emulsione è costituito da una fase acquosa “microreattore” da 50 a 100 µm di diametro, ognuna contenente tutti i componenti della miscela di amplificazione e non più di una biglia.

4 Amplificazione: l'emulsione di biglie viene sottoposta a PCR per amplificare in modo clonale ogni molecola di DNA template. Il DNA template si ibrida ai primers oligonucleotidici legati alle biglie durante la cattura dei primers (step 2). Questi primers di cattura, doppi come i primers di PCR, ancorano i filamenti complementari neo-sintetizzati alle biglie. Quando la reazione di PCR procede, i filamenti complementari legati alle biglie legano i primer di cattura liberi in eccesso e dirigono la sintesi di molti filamenti complementari non legati alle biglie di cattura; questi secondi primer di amplificazione solubili però sono biotinilati e questo tiene conto dell'arricchimento delle biglie che legano il DNA e che avviene successivamente nella fase 6. Dopo l'amplificazione, sono immobilizzate su ogni biglia un numero di copie di template tra 10 e 30×10^6 copie per biglia.

5 Recupero delle biglie: dopo la fase di amplificazione, l'emulsione viene rotta chimicamente, e le biglie che legano la libreria di DNA amplificata (a doppio-filamento a questo punto) sono recuperate e lavate per filtrazione.

6 Arricchimento della libreria di biglie-DNA: la procedura fin qui descritta genera una certa proporzione di biglie che legano DNA non amplificato, o perché non hanno catturato una molecola di template dall'inizio, o perché il DNA template non si è amplificato. Per ridurre la percentuale di biglie senza template, questa sesta fase consiste in una procedura che arricchisce la popolazione di biglie che hanno legato DNA amplificato. Questo step di arricchimento comporta il legame dei primer “A” di amplificazione biotinilati (i primer di PCR solubili, vedi step 4) a biglie magnetiche rivestite di streptavidina, e la separazione delle biglie che legano DNA amplificato da quelle vuote con un “collettore” di particelle magnetiche. Le biglie con la libreria di DNA sono separate dalle biglie magnetiche separando i prodotti di amplificazione a doppio filamento, lasciando una popolazione di biglie legate ad un singolo filamento di DNA template: la libreria è immobilizzata e amplificata.

7 Annealing dei primer di sequenza: lo step finale nel processo di emPCR è l'appaiamento dei primer di sequenza ai template di DNA immobilizzati e amplificati. A questo punto i campioni nel PTP sono pronti per il sequenziamento. Tutti i passaggi descritti sono stati eseguiti secondo i protocolli forniti su www.roche-applied-science.com e forniti con i kit in uso.

Sequenziamento

Dopo l'amplificazione, le biglie leganti il DNA devono essere caricate nei pozzetti di un PTP in modo che ogni pozzetto contenga una sola biglia-DNA. La PTP carica viene quindi inserita nello strumento Genome Sequencer FLX, e i reagenti di sequenziamento fluiscono sequenzialmente sul piatto. Lo strumento Genome Sequencer FLX automaticamente esegue e monitora la reazione di sequenziamento in ogni pozzetto del PTP simultaneamente.

L'esperimento di sequenziamento con il Genome Sequencer FLX Instrument prevede 3 parti principali: Pre-lavaggio del sub-sistema fluidico, preparazione della PTP e corsa di sequenziamento. Lavaggio, impostazione dello strumento e preparazione della PTP avvengono in parallelo.

Il primo step nella corsa di sequenza è stato quindi il prelavaggio del sistema fluidico e il caricamento dei reagenti di sequenziamento nello strumento. La cartuccia PTP e la cameretta del piatto frontale sono state pulite, in modo che lo strumento fosse pronto per la corsa. Sono stati preparati i buffer richiesti combinando le biglie di reagenti con gli altri componenti forniti nel GS Sequencing Kit secondo protocollo; è stata quindi preparata la PTP, le biglie di DNA, le biglie di *packing*, e le biglie di enzima. Dopo che il sistema di deposizione delle biglie (*bead deposition device*) è stato assemblato, con la PTP ed il *bead loading gasket* appropriato (fig. 10), le biglie sono state depositate in tre strati, uno per ogni tipo di biglie incluse nel sistema:

- Il primo strato è composto dalle biglie contenenti il DNA da sequenziare che vengono fatte depositare per gravità;
- Il secondo strato è costituito dalle *packing beads*, che servono a stabilizzare e mantenere tutti i componenti del sistema immobilizzati nei pozzetti del PTP, durante la corsa di sequenza. Questo viene fatto depositare tramite centrifugazione;
- Il terzo strato è formato dalle biglie di enzima che legano i componenti enzimatici per il sistema di chemiluminescenza e viene depositato anche questo per centrifugazione.

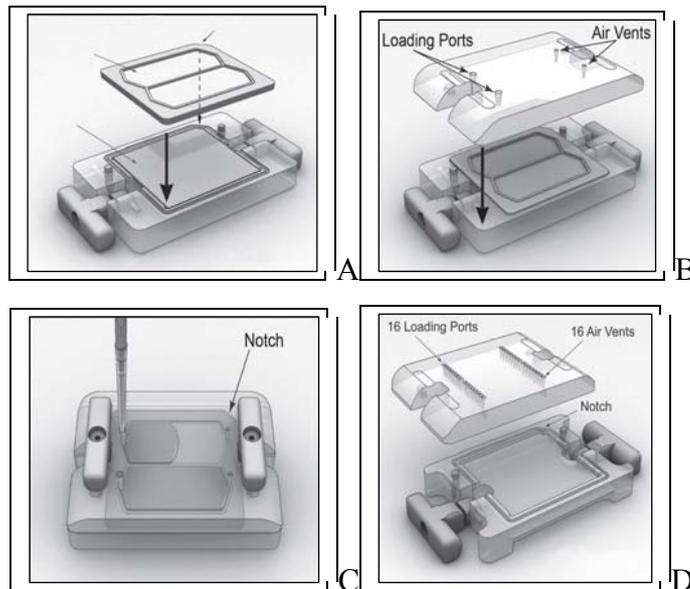


Fig. 10: assemblaggio del sistema di deposizione delle biglie (*Bead Deposition Device*), *PTP* e *Bead Loading Gasket* (A;B); le biglie vengono caricate attraverso le *loading ports* (C) che sono in numero diverso a seconda del numero di porzioni in cui è diviso il *PTP* dal *gasket*; nel presente lavoro è stato utilizzato un *BDB* da 16 *loading ports* (D).

Per avviare la corsa di sequenza, è stata selezionata la *Run script* idonea e specificato lo schema di processazione dei dati appropriato per l'esperimento. Infine, la *PTP* completamente pronta, contenente il DNA da sequenziare, è stata inserita nella *PTP Cartridge* nello strumento, e la corsa è stata avviata. Il *Genome Sequencer FLX Instrument* registra, cattura i dati grezzi, e processa i dati autonomamente.

Analisi dei dati

I dati grezzi di una corsa di sequenza consistono in una serie di immagini digitali (*PIF files*) per ciascuna sequenza dei frammenti della libreria di DNA ("reads") e devono essere sottoposti ad una analisi sofisticata a seconda del tipo di libreria e dell'obiettivo dell'esperimento.

La tipizzazione di HPV è stata fatta usando una combinazione di custom programmed Perl scripts con altri software esistenti (*Clustalw2*, *Jalview 2.4* etc). Per prima cosa, le sequenze specifiche di ogni campione sono state identificate in base al loro *MID-tags* associato e la sequenza dei primer è stata eliminata. Quindi è stato creato un database custom, contenente le sequenze appartenenti a tutti i tipi di HPV presenti in *GenBank*. Sequenze per ogni campione sono state allineate con questo e con *BLAST* e divise in sottogruppi tipo specifico. Per poter essere assegnata ad uno specifico tipo di HPV, una sequenza deve allinearsi almeno al 90% della

lunghezza, per evitare errori di tipizzazione dovute all'allineamento di porzioni corte di sequenze, e con una identità di sequenza minima del 90%, in accordo con le linee guida della tipizzazione di HPV. Con questi criteri, è stata clusterizzata la maggior parte delle sequenze e sono state allineate con le sequenze di riferimento con una identità >95%. Sebbene l'accuratezza dell'allineamento possa essere limitata dal tasso di errore del pirosequenziamento, specialmente nelle regioni omopolimeriche, i tipi virali presentano un delta nelle sequenze similarity alta sufficiente per ottenere una corretta identificazione. (Barzon et al., 2011)

2. Metodi del primo protocollo di sequenziamento 454

Campioni clinici

Sono stati selezionati 154 campioni citologici cervicali positivi per HPV DNA e 10 negativi, raccolti da pazienti rivoltisi all'Unità operativa di Microbiologia Clinica per effettuare il test dell'HPV e genotipizzazione, come test di secondo livello per lo screening del carcinoma cervicale o per investigare lesioni correlate ad HPV. In questi campioni la ricerca dell'HPV e genotipizzazione è stata condotta usando una PCR consensus con primer MY09/11, seguita da sequenziamento tradizionale, come precedentemente descritto (Barzon et al., 2010). Il sequenziamento aveva identificato singole infezioni da HPV in 126 campioni e suggerito infezioni multiple in 28, inclusi 23 campioni non tipizzabili. Tutti i 164 campioni sono stati analizzati tramite sequenziamento NGS come descritto nel precedente paragrafo. Inoltre un sottogruppo di 148 campioni è stato analizzato usando il test basato su ibridazione inversa INNO-LiPA HPV Genotyping Extra assay (Innogenetics, Gent, Belgium), secondo protocollo.

Protocollo per la tipizzazione di HPV tramite 454 NGS

Il DNA totale è stato purificato dai tamponi cervicali usando il kit QIAmp DNA blood in uno strumento BioRobot 9604 Workstation (QIAGEN GmbH, Hilden, Germany). Per generare ampliconi per la libreria 454 NGS, 100ng circa di DNA sono stati amplificati tramite PCR usando fusion primer e la polimerasi PWO SuperYield DNA (Roche Applied Sciences, Monza, Italy). I fusion primer sono stati disegnati in modo da contenere la stessa sequenza dei primer degenerati MY09/11 usati per il sequenziamento tradizionale (MY09 5'-CGTCCMARRGGAWACTGATC-3' e MY11 5'-CMCAGGGWCATAAYAATGG-3'), le sequenze adattatori A e B e la sequenza chiave richieste per il sequenziamento 454 NGS ed uno dei 12 differenti identificatori multipli (MID) in accordo con il manufacturer's GS FLX Standard sequencing method. Le condizioni di PCR sono 1 ciclo a 95°C per 10 min, seguito da 40 cicli a 95°C per 30 s, 54°C per 45 s, e 72°C per 1 min, e uno step finale di 5 min a 72°C. I prodotti di

PCR sono stati purificati tramite AMPure beads (Agencourt Bioscience Corporation, Beverly, MA), quantificati, diluiti 2×10^5 molecole/ μL , e poolate secondo la compatibilità dei barcode (MID). Ogni pool è stato amplificato tramite emulsion PCR (emPCR) usando il GS-FLX emPCR Kit II e III e sequenziato sia in direzione forward che reverse a $5000\times$ coverage in uno strumento Genome Sequencer FLX usando la chimica GS FLX Standard (Roche 454 Life Sciences, Branford, CT).

La lunghezza media delle sequenze era di 248 nucleotidi, in accordo con la dimensione media di sequenza attesa secondo il protocollo GS FLX standard.

Controlli di DNA positivi per HPV16 e HPV18 sono stati preparati con DNA purificato da cellule CaSki e HeLa (che contengono approssimativamente 500 e 50 copie genomiche di HPV16 e HPV18 per cellula, rispettivamente) e diluite in DNA carrier ottenuto da cellule HPV-negative HEK293, in accordo con il WHO Human Papillomavirus Laboratory Manual. DNA di cellule CaSki e HeLa sono stati anche mescolati in differenti proporzioni per generare campioni che fungessero da controlli di infezioni multiple.

3. Metodi del secondo protocollo di sequenziamento 454

Campioni

Controlli positivi. In accordo con quanto riportato nel WHO Human Papillomavirus Laboratory Manual (2009), i controlli positivi sono stati preparati con DNA purificato da cellule CaSki e HeLa, che contengono approssimativamente 500 e 50 copie di genomi di HPV16 e HPV18 per cellula, rispettivamente. Il DNA delle cellule CaSki e HeLa è stato diluito in DNA carrier ottenuto da cellule HEK293 HPV negative per ottenere 200 HPV16 o HPV18 GE per μL in un background di ≈ 1670 copie di DNA cellulare; si ottiene così un input di ≈ 1000 HPV GE e 8400 cell equivalenti per $5 \mu\text{L}$. I controlli positivi sono stati quindi diluiti per ottenere 100 e 10 HPV GE per $5 \mu\text{L}$. DNA ottenuto dalle cellule CaSki e HeLa è stato quindi mescolato in differenti proporzioni 10:1, 5:1, 1:1, 1:10, 1:100 e 1:500 di 200 GE HeLa-DNA:CaSki-DNA per generare campioni controllo di infezioni multiple.

Campioni del WHO proficiency panel. Sono stati utilizzati come controllo campioni di HPV DNA dello studio di genotipizzazione del WHO del 2010 e 2011 (Eklund et al., 2010). I campioni del WHO proficiency study sono composti di DNA plasmidico purificato in cui il DNA genomico di differenti tipi di HPV sono stati clonati e diluiti in un background di DNA umano placentare (Eklund et al., 2012). Questi campioni sono stati usati per preparare il pannello di 43 campioni di riferimento che contengono diverse percentuali di plasmidi (5 e 50 GE per 5

µL per HPV16 e HPV18 e 50 e 500 GE per 5 µL per gli altri tipi ad alto rischio HPV 31, 33,35, 39, 45, 51, 52, 56, 58, 59, 66, 68a, 68b, e a basso rischio HPV 6 e 11) e 3 campioni di riferimento che contengono linee cellulari HPV-positive e HPV-negative come descritto nel report del WHO proficiency study (Eklund et al., 2012). Alcuni campioni contengono differenti tipi di HPV per mimare infezioni multiple.

Pools di tamponi anogenitali HPV-positivi. È stato analizzato un gruppo di 30 tamponi anogenitali raccolti da pazienti afferenti all'Unità operativa della Clinica Microbiologica di Padova per eseguire il test dell'HPV e genotipizzazione e risultati positivi al test INNO-LiPA HPV Genotyping Extra (Innogenetics, Gent, Belgium), che si basa su SPF10 primers (SPF10-LiPA). Sono stati selezionati campioni con infezioni multiple e rappresentativi di differenti tipi di HPV. Gli acidi nucleici purificati da questi 30 campioni sono stati mescolati in uguali quantità per generare 10 campioni con un gran numero di differenti tipi di HPV al fine di testare la capacità del metodo di pirosequenziamento 454 di identificare le infezioni multiple.

Tamponi cervicali con risultati discordanti ottenuti tramite hybrid capture e SPF10 – LiPA.

Una selezione di 8 campioni citologici cervicali raccolti da pazienti afferenti all'Unità operativa della Clinica Microbiologica di Padova e con risultati discordanti ottenuti tramite i test hybrid capture 2 (HC2; Quiagen Corporation, Gaithersburg, MD) e INNO-LiPA (Innogenetics) sono stati selezionati per essere tipizzati con il metodo di pirosequenziamento 454. I campioni discordanti sono stati testati tramite real-time PCR specifiche per HPV 16, 18, 31, 33, 45, 52, 58 and 68 (Militello et al., 2009; Schmitz et al., 2009; Depuydt et al., 2006).

Protocollo per la tipizzazione di HPV tramite 454 pyrosequencing

Il DNA totale è stato purificato da campioni cervicali usando MagNA Pure 96 Viral NA Small Volume Kit® in uno strumento MagNA Pure 96™ (Roche Diagnostics, Monza, Italy). Quantità e integrità del DNA purificato è stata testata amplificando il gene della β-globina tramite real-time PCR quantitativa (Militello et al, 2008). Per generare gli ampliconi per la libreria per il 454 NGS, sono stati amplificati approssimativamente 100ng di DNA usando per la PCR i fusion primers e la PWO SuperYield DNA Polymerase (Roche Applied Sciences, Monza Italy). I multiplex fusion primers sono stati disegnati in modo da contenere le sequenze dei primer PGMY riportate nel WHO Human Papillomavirus Laboratory Manual (2009), le sequenze adattatori A e B e la sequenza chiave richiesta per il 454 NGS, e uno di 8 differenti barcodes multiplex identifier (MID) della lunghezza di 10 bp, in accordo con il protocollo del GS FLX

Standard sequencing method (Fig. 12). Dato che il set di primer PGMY non identifica HPV68a, sono stati disegnati dei primer aggiuntivi specifici per questo tipo di HPV ad alto rischio (i.e., PGMY09 5'-CGTCCTAATGGGAATTGGTC-3' e PGMY11 5'-GCACAGGGACACAACAATGG-3').

Questi oligo modificati sono stati purificati tramite cromatografia liquida ad alta risoluzione (HPLC).

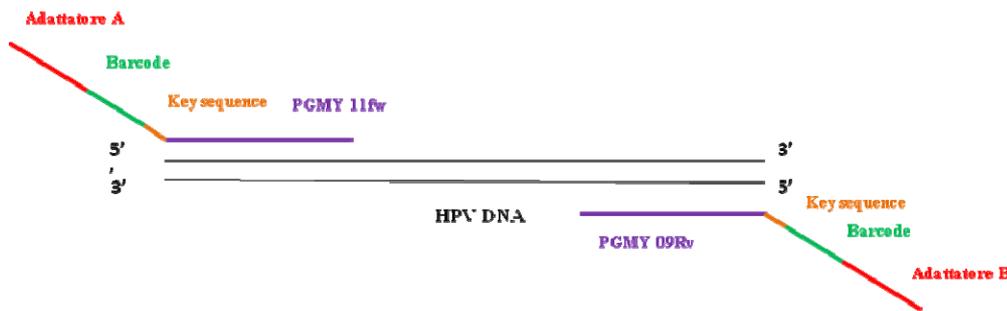


Fig. 12: disegno dei Fusion Primer.

Per mettere a punto le condizioni di PCR, i controlli positivi sono stati amplificati usando i primer PGMY11/09 e i fusion primer a differenti concentrazioni di $MgCl_2$ nella mix di reazione riportata dal Human papillomavirus Laboratory Manual (2009) e usando differenti diluizioni di controlli di HPV per testare la sensibilità della PCR. Per amplificare i campioni per la libreria i fusion primer sono stati mescolati con i primer PGMY11/09 in una proporzione di 1:10 per favorire la reazione.

Il programma di amplificazione è: 95°C, 9 min; 45 cicli (95°C, 30s; 55°C, 1 min 30s, 72°C, 2 min); 72°C, 5 min.

Ogni campione è stato amplificato e processato per la preparazione e quantificazione secondo il protocollo 454 NGS come riportato in Barzon et al. (2011) e come descritto precedentemente in questo lavoro di tesi.

Gli ampliconi ottenuti da ogni campione sono stati purificati tramite AMPure beads (Agencourt Bioscience Corporation, Beverly, MA), quantificati e diluiti 4×10^6 molecole/ μL , poolati (8 campioni per pool) in base alla compatibilità dei MID barcode. Ogni pool è stato amplificato in una singola emulsion PCR (emPCR) usando il GS-FLX emPCR Kit II e III e sequenziato in una singola lines, sia in direzione forward che reverse a 5000x coverage, in uno strumento Genome Sequencer FLX usando la chimica GS FLX Titanium (Roche 454 Life Sciences, Brandford, CT).

RISULTATI

1. Risultati del primo protocollo di sequenziamento 454

Performance analitica del metodo di tipizzazione 454 NGS

Per determinare la sensibilità del metodo NGS in infezioni singole e multiple erano stati utilizzati campioni di DNA purificato da cellule CaSki, positive per HPV16, e da cellule HeLa, positive per HPV18, e campioni contenenti solo DNA carrier. I campioni erano stati analizzati in triplo usando il protocollo descritto. In ogni campione i tipi di HPV erano stati identificati correttamente con una sensibilità superiore a 100 genomi equivalenti di HPV16 e di HPV18 per μl di DNA, sia in infezioni singole che multiple. Inoltre nelle infezioni multiple i tipi di HPV erano stati identificati correttamente quando rappresentavano più del 1% dei genomi equivalenti e la percentuale di sequenze ottenute era proporzionale all'input di DNA di HPV16 e HPV18.

Confronto della tipizzazione di HPV tramite 454 NGS con il sequenziamento tradizionale.

Tutti i 154 campioni nei quali era stato identificato DNA di HPV tramite PCR MY09/11 erano risultati positivi anche usando i fusion primer NGS, mentre i campioni che erano risultati negativi per HPV, lo erano anche con i fusion primer. Inoltre il metodo NGS riusciva a genotipizzare tutti i genotipi dei campioni positivi alla PCR. Nonostante la normalizzazione della quantità di amplicone per la preparazione della libreria NGS, il numero di sequenze ottenuto variava molto tra i diversi campioni (mediana delle sequenze per campione 624; range 99-3409), ma senza significative differenze tra campioni con diversi tipi di HPV, tra campioni con HPV ad alto o basso rischio o tra infezioni singole e multiple. Come atteso il metodo NGS classificava più campioni con infezioni da tipi multipli rispetto al sequenziamento tradizionale (68 con singola infezione e 86 con infezione multipla). In tutti i 68 casi di infezione singola i risultati di tipizzazione NGS erano in perfetto accordo con quelli del sequenziamento tradizionale. HPV16 e HPV53 erano i tipi più comunemente individuati, seguiti da HPV33, 58, 62, 66 (tab. II). Degli 86 campioni con infezioni multiple, i risultati dell'NGS erano compatibili con il sequenziamento tradizionale in 57 casi, cioè in 57 casi gli HPV identificati tramite sequenziamento tradizionale erano inclusi tra i tipi identificati da NGS. In 70% di questi casi il tipo di HPV identificato tramite sequenziamento tradizionale corrispondeva al tipo di HPV presente per $\geq 80\%$ delle sequenze ottenute tramite NGS. Risultati discordanti erano stati osservati nei rimanenti 29 campioni, inclusi 23 non tipizzabili tramite sequenziamento tradizionale. Dei campioni con infezioni multiple 61,2% avevano 2 differenti tipi di HPV, 28,2% avevano 3 differenti tipi virali, 7,1% avevano 4 tipi virali e 3,5% ne avevano 5. HPV53, 66 e 58 seguiti da HPV16 erano i più comunemente identificati (tab. II).

Confronto tra tipizzazione di HPV ottenuta tramite 454 NGS e tramite LiPA

In un sottogruppo di 148 pazienti (67 con infezione singola e 81 con infezioni multiple identificate dall'NGS) la tipizzazione di HPV era stata condotta anche usando INNO-LiPA HPV Genotyping Extra kit (Innogenetics). Complessivamente i livelli di concordanza per i singoli genotipi di HPV rilevati da entrambe i metodi era compreso tra 85% e 99%, ma i livelli di concordanza nei positivi era compreso tra 0 e 70% (tabella III). In particolare veniva dimostrata una buona concordanza tra i due metodi per HPV11, 31, 53, 54 e HPV82; il metodo NGS confrontato con il test LiPA, rilevava in percentuale significativamente maggiore HPV58 rispetto ad HPV52; HPV35, HPV44, HPV51, HPV56, e HPV74, rilevati da LiPA in 5.4%–10.8% dei campioni, non erano mai stati identificati dal metodo NGS. Complessivamente, il metodo LiPA classificava un alto numero di campioni con infezioni da HPV ad alto rischio rispetto al metodo NGS. In qualche caso, il metodo LiPA produceva risultati falso-positivi, incluse diagnosi di infezione da HPV ad alto rischio, in presenza di HPV che non erano tipizzabili dal test, come HPV61, HPV62, HPV81, e HPV84. Inoltre, il metodo LiPA identificava raramente i tipi di HPV che rappresentano meno del 10% delle sequenze di HPV ottenute tramite NGS. Per valutare la sensibilità del metodo LiPA nelle infezioni multiple, il test era stato condotto in campioni contenenti differenti proporzioni di DNA purificato da cellule CaSki e HeLa. Il limite di identificazione del metodo LiPA era di 10 genomi equivalenti di DNA di HPV16 e HPV18 per μ L. I tipi di HPV erano correttamente identificati quando rappresentavano il 5% dei genomi equivalenti di HPV nelle infezioni multiple, mentre i tipi che rappresentavano una più bassa proporzione non erano consistentemente individuati.

Individuazione di varianti intratipo nei campioni clinici

Per trovare la presenza di varianti/mutanti intratipo di HPV nello stesso campione, era stato condotto un allineamento multiplo di sequenza tra tutte le sequenze raggruppate per lo stesso tipo di HPV usando ClustalW211, e includendo anche la sequenza di riferimento del genotipo corrispondente. Gli allineamenti multipli venivano scansionati posizione per posizione, e varianti che rappresentavano almeno il 10% delle sequenze per un certo tipo di HPV venivano arbitrariamente considerate una potenziale reale variante.

Per verificare se queste varianti fossero un errore del metodo di pirosequenziamento o reali varianti/mutazioni nel genoma virale, quattro campioni rappresentativi contenenti inserzioni o delezioni e sostituzioni di un singolo nucleotide erano state successivamente analizzate tramite sub clonaggio di amplicone in un plasmide vettore. Successivamente erano stati sequenziati almeno 20 cloni per clonaggio con primer tipo specifici per il gene L1 di HPV. Mentre varianti

con inserzione o delezione di nucleotidi non erano confermate dal sequenziamento, 3 delle 4 sostituzioni nucleotidiche erano state confermate, ad esempio le varianti intratipo identificate in HPV31 (tab. IV) già descritte in letteratura [12] (GenBank accession number HM596540 e DQ218249).

TAB. II. Distribuzione dei tipi di HPV in infezioni singole e in infezioni multiple

HPV type	infezioni singole		infezioni multiple		Totale	
	No.	(%)	No.	(%)	No.	(%)
HPV53	7	10.3	27	12.6	34	12.0
HPV66	6	8.8	24	11.2	30	10.6
HPV58	6	8.8	21	9.8	27	9.5
HPV16	7	10.3	17	7.9	24	8.5
HPV6	5	7.4	15	7	20	7.1
HPV31	5	7.4	15	7	20	7.1
HPV62	6	8.8	14	6.5	20	7.1
HPV54	2	2.9	15	7	17	6.0
HPV61	4	5.9	10	4.7	14	4.9
HPV33	6	8.8	7	3.3	13	4.6
HPV81	3	4.4	9	4.2	12	4.2
HPV59	1	1.5	7	3.3	8	2.8
HPV83	3	4.4	5	2.3	8	2.8
HPV84	2	2.9	6	2.8	8	2.8
HPV52	3	4.4	3	1.4	6	2.1
HPV73	1	1.5	4	1.9	5	1.8
HPV89	0	0	4	1.9	4	1.4
HPV70	1	1.5	2	0.9	3	1.1
HPV11	0	0	2	0.9	2	0.7
HPV18	0	0	2	0.9	2	0.7
HPV68	0	0	2	0.9	2	0.7
HPV39	0	0	1	0.5	1	0.4
HPV42	0	0	1	0.5	1	0.4
HPV72	0	0	1	0.5	1	0.4
HPV82	0	0	1	0.5	1	0.4
Total infections	68	100	215	100	283	100

TAB. II. Confronto dei risultati ottenuti dal metodo NGS e dal test LiPA per l'individuazione dei singoli tipi di HPV.

HPV genotype ^a	No. of specimens				Prevalence (%)		% Agreement		κ^b	β^c
	NGS-/LiPA-	NGS-/LiPA+	NGS+/LiPA-	NGS+/LiPA+	NGS	LiPA	Overall	Positive		
High-risk HPVs										
HPV16	120	8	5	15	13.5	15.5	91.2	53.6	0.647	0.581
HPV18	142	4	0	2	1.4	4.1	97.3	33.3	0.490	0.125
HPV31	125	4	3	16	12.8	13.5	95.3	69.6	0.793	1
HPV33	129	6	4	9	8.8	10.1	93.2	47.4	0.606	0.754
HPV35	145	3	0	0	0.0	2.0	98.0	0.0	na	na
HPV39	142	5	1	0	0.7	3.4	95.9	0.0	0	0.219
HPV45	147	1	0	0	0.0	0.7	99.3	0.0	na	na
HPV51	136	12	0	0	0.0	8.1	91.9	0.0	na	na
HPV52	127	15	1	5	4.1	13.5	89.2	23.8	0.344	0.001
HPV56	143	5	0	0	0.0	3.4	96.6	0.0	na	na
HPV58	123	0	13	12	16.9	8.1	91.2	48.0	0.605	<0.0001
HPV59	138	2	5	3	5.4	3.4	95.3	30.0	0.438	0.453
HPV68	143	3	0	2	1.4	3.4	98.0	40.0	0.563	0.25
Any high-risk HPV	47	22	10	69	53.4	61.5	78.4	68.3	0.561	0.05
Low-risk HPVs										
HPV6	126	3	9	10	12.8	8.8	91.9	45.5	0.581	0.146
HPV11	144	2	0	2	1.4	2.7	98.6	50.0	0.661	0.500
HPV26	147	1	0	0	0.0	0.7	99.3	0.0	na	na
HPV40	147	1	0	0	0.0	0.7	99.3	0.0	na	na
HPV43	147	1	0	0	0.0	0.7	99.3	0.0	na	na
HPV44	133	15	0	0	0.0	10.1	89.9	0.0	na	na
HPV53	103	12	5	28	22.3	27.0	88.5	62.2	0.692	0.143
HPV54	132	0	5	11	10.8	7.4	96.6	68.8	0.797	0.063
HPV66	103	16	7	22	19.6	25.7	84.5	48.9	0.559	0.093
HPV69/71	145	3	0	0	0.0	2.0	98.0	0.0	na	na
HPV70	145	0	2	1	2.0	0.7	98.6	33.3	0.495	0.5
HPV73	142	1	3	2	3.4	2.0	97.3	33.3	0.487	0.625
HPV74	138	10	0	0	0.0	6.8	93.2	0.0	na	na
HPV82	146	1	0	1	0.7	1.4	99.3	50.0	0.664	1.000

TAB. III. Risultati della ricerca di varianti di sequenza L1 di HPV ottenuti tramite NGS e confermati tramite PCR clonaggio e sequenziamento tradizionale in quattro campioni citologici cervicali.

Sample #/HPV type	Intratype variant ^a	Aminoacid change ^a	Average value of quality scores ^b	No. (%) of reads with variant at NGS	No. (%) of clones with variant at cycle sequencing
Sample 1/HPV16	1308insC	Frameshift	19.29	68 (40.2%)	0/20 (0%)
Sample 2/HPV31	T1095G	No change	36.52	151 (63.4%)	10/20 (50%)
	G1245A	No change	36.22	178 (74.8%)	16/20 (80%)
	C1266A	No change	36.04	46 (47.4%)	14/20 (70%)
Sample 3/HPV52	1281delT	Frameshift	22.92	74 (36.8%)	0/20 (0%)
Sample 4/HPV81	C1441G	Leu481Val	35.65	29 (55.8%)	0/20 (0%)

^a HPV L1 sequence variants and amino acid changes relative to the L1 gene of prototype genomes.

^b The average is calculated for the quality values associated to all nucleotides in the corresponding position of multiple alignment.

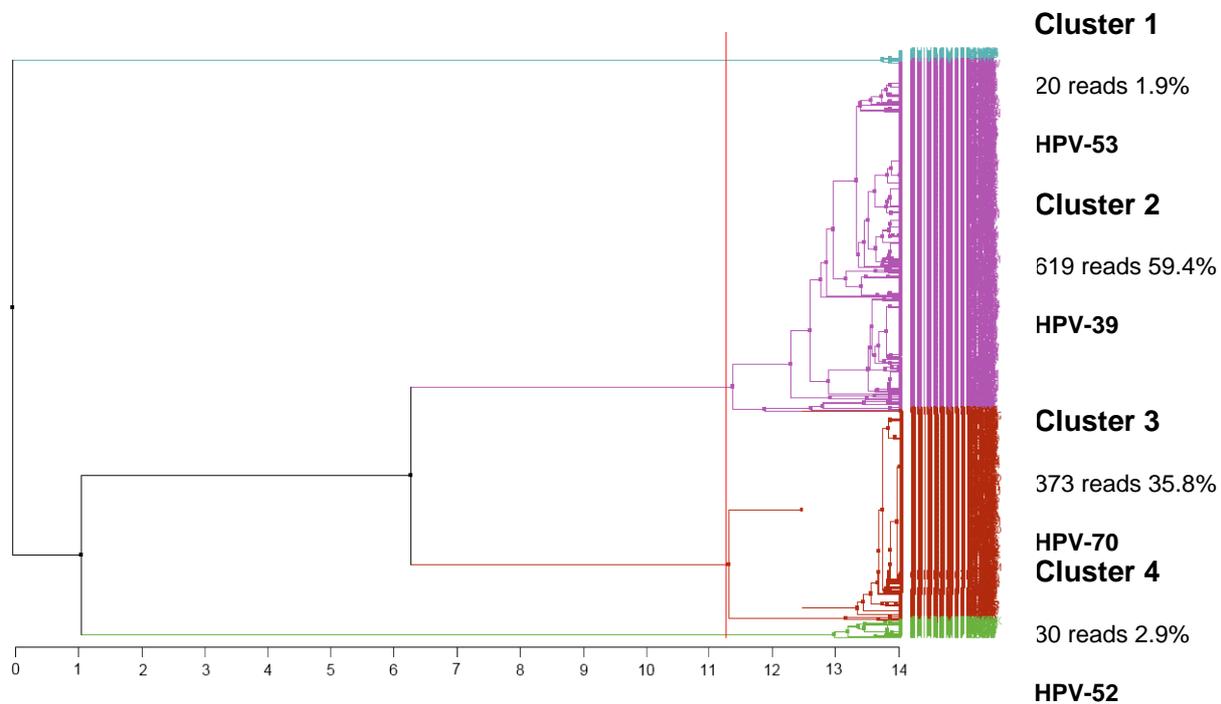
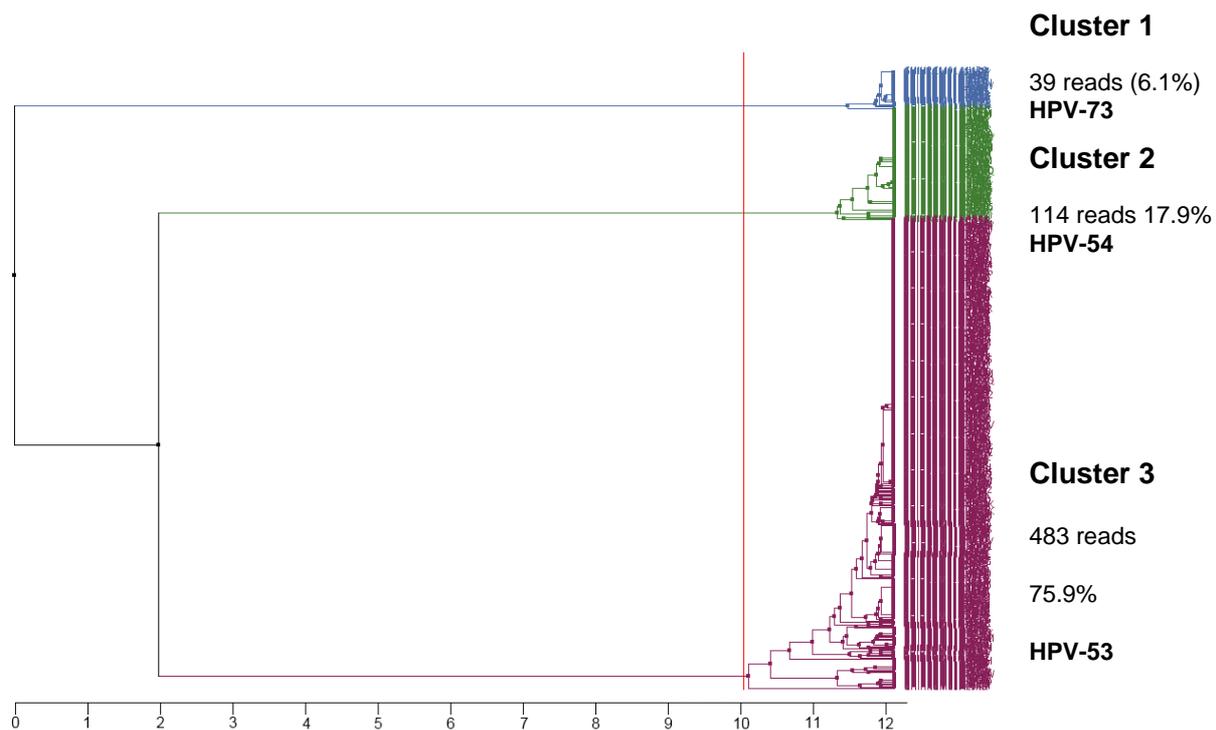


Figure 13. esempi di analisi di cluster di campioni con infezioni multiple; rappresentazione schematica della distanza (%) tra le sequenze ottenuta tramite Jalview (Waterhouse AM, Procter JB, Martin DM,

Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009 May 1;25(9):1189-91. Epub 2009 Jan 16.

2. Risultati del secondo protocollo di sequenziamento 454

Messa a punto del protocollo.

La messa a punto del protocollo aveva richiesto di testare diverse condizioni di PCR:

1. prima PCR condotta usando i primers PGMY per 10 cicli; 3µl della prima PCR sono usati in una nested PCR con fusion primers per 45 cicli.
2. prima PCR condotta usando i primers PGMY per 16 cicli; 3µl della prima PCR sono usati in una nested PCR con fusion primers per 45 cicli.
3. una mix di PGMY-primers:fusion-primers 1:10 è usata in una sola PCR per 45 cicli.

L'efficienza dei tre protocolli era stata confrontata con l'efficienza della PCR condotta usando i primer PGMY normali (amplicone A in fig.14).

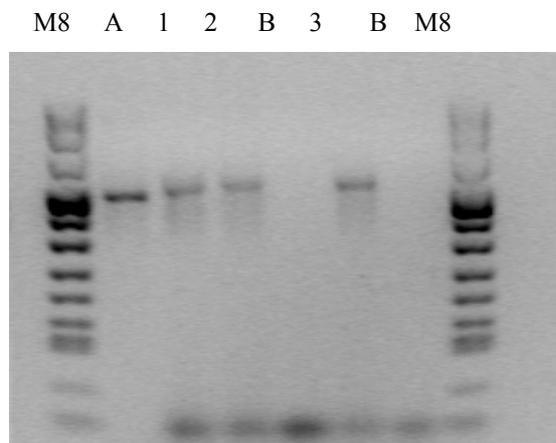


Fig. 14. *Differenti protocolli di PCR testati in controlli di DNA-positivi.*

Differenti concentrazioni di MgCl₂ e differenti concentrazioni di primer erano state provate per stabilire le condizioni ottimali di PCR (Fig.14). Il miglior risultato era stato ottenuto usando 3 µl MgCl₂ e 0.4 µl di primer.

MgCl ₂ (µl)	3	4	2	4	2	B
Primers (µl)	0.4	0.6	0.6	0.4	0.4	

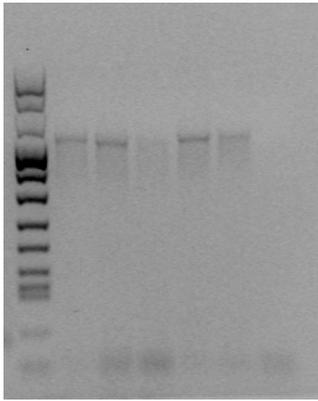


Fig.15. differenti concentrazioni di $MgCl_2$ e di primer.

La sensibilità della PCR condotta con i fusion primer (protocollo 3, $4\mu l$ $MgCl_2$ e $0.4\mu l$ primers) era stata paragonata con la sensibilità della PCR condotta con i primer PGMY in un campione di cellule CasKi positivo per HPV16 e di cellule HeLa positive per HPV18, diluiti da 1000 GE a 10 GE (fig.16).

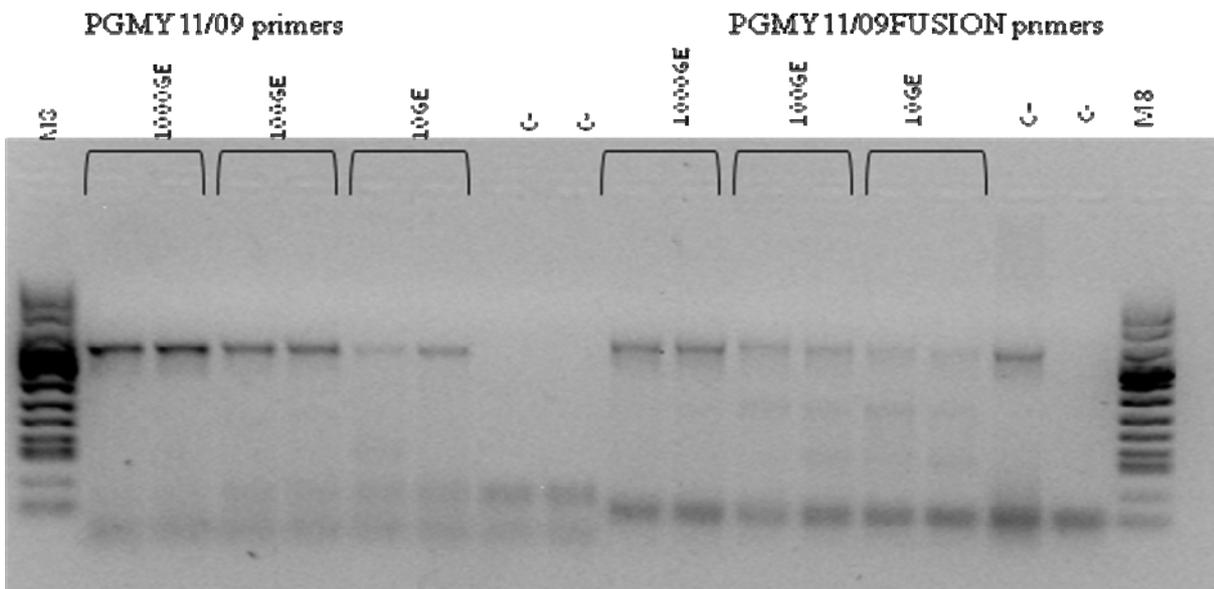


Fig. 16. Sensibilità della PCR con fusion primer in differenti diluizioni di diversi GE genomi equivalenti di DNA di HPV16.

Il protocollo di PCR era stato testato prima in 4 campioni clinici positivi al test INNO-LiPA rispettivamente per HPV33, per HPV56 e HPV54, per HPV66, e per HPV51 (fig. 17) ed infine in ogni campione selezionato.

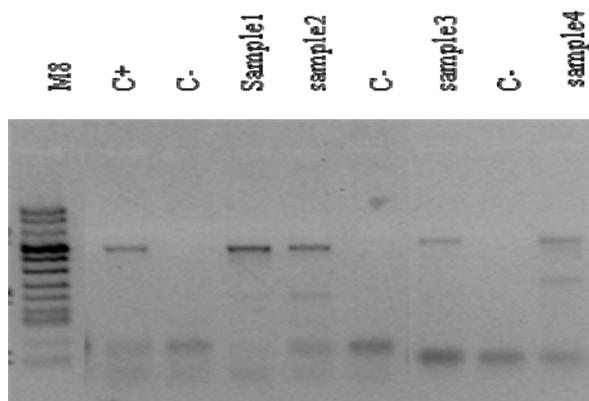


Fig.17. preparazione della libreria di ampliconi da 4 campioni.

Risultati della corsa NGS.

La lunghezza media delle sequenze ottenute dalla corsa NGS era 358.5 ± 131.4 nucleotidi, in accordo con la le sequenze attese dal protocollo GS FLX Titanium. Il numero medio di sequenze di high-quality per campione era 12806, range 4686-34990.

Sensibilità e specificità nel WHO Panel 2010 e 2011

In una prima corsa di pirosequenziamento, per analizzare i campioni del WHO 2010 e 2011 proficiency panel era stato utilizzato un set di primer PGMY multiplex disegnato basandosi sulle sequenze raccomandate nel WHO Laboratory manual (2011). In ogni campione, i tipi di HPV erano identificati e tipizzati correttamente sia a 10 che a 100 GE/ μ L di concentrazione, con l'eccezione di HPV68a, che non era stato identificato. Nei campioni con infezioni multiple di HPV, il numero di sequenze ottenute per HPV31, HPV33, HPV51 e HPV56 era approssimativamente dieci volte più basso degli altri tipi di HPV, suggerendo una differente sensibilità del metodo per questi tipi di HPV.

Poichè i primer PGMY si erano dimostrati incapaci di identificare HPV68a a causa dei diversi mismatches anche nel WHO proficiency study (Eklund et al., 2012), si era deciso di disegnare un nuovo paio di primer aventi come target HPV68a da aggiungere alla miscela di primer PGMY. La sensibilità e specificità del set di primer PGMY così modificato per HPV68a era stata studiata con una nuova corsa NGS in 8 campioni del WHO 2010 proficiency panel contenenti HPV68a o HPV68b in infezione singola o multipla e in 8 campioni negativi per HPV68a e HPV68b. I risultati della corsa 454 NGS dimostravano che la la miscela di primer PGMY modificata permetteva la corretta identificazione di HPV68a sia ad una concentrazione di 10 che di 100 GE/ μ L, sia in infezioni singole che multiple (Table V-VI).

Tab. V. Risultati della genotipizzazione di HPV dei campioni del WHO 2010 proficiency panel usando la mix di primerPGMY modificata e il pirosequenziamento 454.

HPV type	HPV GE/ μ l	Primer MID	454 pyrosequencing results (no. reads)
HPV11	10	MID4	HPV11 (6498)
HPV11	100	MID2	HPV11 (5177)
HPV18	1	MID1	HPV18 (6208)
HPV18	10	MID7	HPV18 (9756)
HPV59	10	MID1	HPV59 (6754)
HPV59	100	MID8	HPV59 (12535)
HPV68a	10	MID2	HPV68a (8655)
HPV68a	100	MID3	HPV68a (6780)
HPV68b	10	MID6	HPV68b (10710)
HPV68b	100	MID5	HPV68b (8675)
HPV6, 16, 15, 51	10	MID8	HPV6b (950), HPV16 (825), HPV18 (454), HPV51 (52)
HPV6, 16, 15, 51	100	MID6	HPV6b (1237), HPV16 (972), HPV18 (740), HPV51 (67)
HPV11, 16, 31, 33, 58	10	MID7	HPV11 (3447), HPV16 (3326), HPV31 (21), HPV33 (86), HPV58 (1655)
HPV11, 16, 31, 33, 58	100	MID4	HPV11 (2876), HPV16 (3165), HPV31 (144), HPV33 (236), HPV58 (2054)
HPV39, 45, 52, 56, 68a	10	MID5	HPV39 (553), HPV45 (7580), HPV52 (3879), HPV56 (58), HPV68a (2676)
HPV39, 45, 52, 56, 68a	100	MID1	HPV39 (332), HPV45 (6181), HPV52 (5552), HPV56 (27), HPV68a (864)
HPV35, 59, 66, 68b	10	MID2	HPV35 (200), HPV59 (3773), HPV66 (1249), HPV68b (1512)
HPV35, 59, 66, 68b	100	MID3	HPV35 (372), HPV59 (5418), HPV66 (1780), HPV68b (2112)

Tab VI. Risultati della genotipizzazione di HPV dei campioni del WHO 2011 proficiency panel usando la mix di primerPGMY modificata e il pirosequenziamento 454.

WHO samples (HPV type)	HPV GE/ μ l	Primer MID	INNO-LiPA results	454 pyrosequencing results (no. reads)
WHO25 (16,45, 51, 33)	100	MID4	HPV16, 33, 45, 51	HPV16 (3592), 33 (38), 45 (2328), 51 (320)
WHO26 (11, 18, 31, 52)	100	MID8	HPV18, 31, 52, 11	HPV18 (1424), 31 (3), 52 (2041), 11 (5032)
WHO28 (6, 56, 58, 68a)	100	MID6	HPV56, 58, 6	HPV56 (1), 58 (966), 6b (4593)
WHO37 (35, 39, 59, 66, 68b)	100	MID7	HPV 35, 68, 66	HPV 35 (129), 68 (1822), 66 (4026), 59 (6800), 39 (591)
WHO4 (11)	10	MID4	HPV11	HPV11 (6498)
WHO14 (59)	100	MID8	HPV59	HPV59 (12535)
WHO15 (11)	100	MID6	HPV11	HPV11 (5177)
WHO38 (18)	10	MID8	HPV18	HPV18 (6208)

Riproducibilità e analisi quantitativa dei campioni positivi per HPV16 e HPV18

Per analizzare la riproducibilità del metodo di pirosequenziamento 454, acidi nucleici purificati da linee cellulari CaSki e HeLa erano stati mescolati in differenti proporzioni per ottenere differenti percentuali di copie genomiche di HPV16 e HPV18, comprese tra 0.1% e 80% di copie di HPV16 del numero totale di copie genomiche di HPV16 e HPV18. I campioni erano stati sia amplificati che sequenziati in doppio. I risultati di questo test sono mostrati in figura 18. La percentuale ed il numero assoluto di sequenze dei due tipi di HPV ottenuti tramite il pirosequenziamento 454 era altamente riproducibile e la percentuale di sequenze ottenute era correlata con la percentuale di ciascun tipo di HPV presente nella mix di campione.

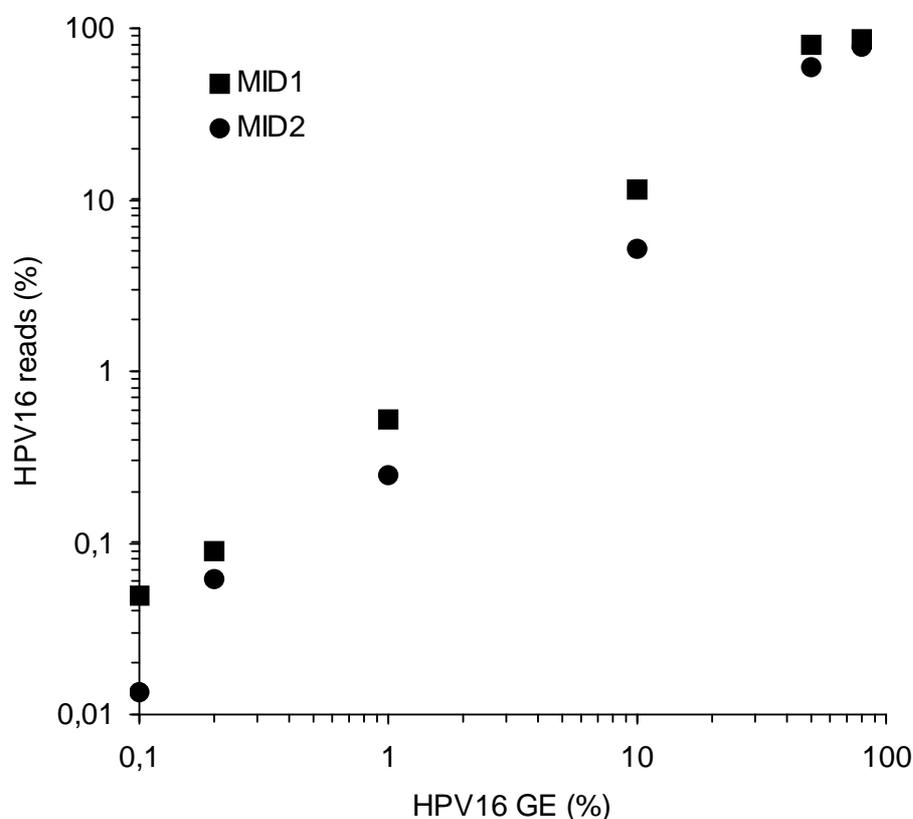


Fig. 18: L'analisi della miscela di linee cellulari positive per HPV16 e HPV18 dimostra che il metodo NGS può quantificare in modo riproducibile la proporzione di ogni tipo di HPV nelle infezioni multiple.

Analisi delle infezioni multiple di HPV

Erano stati generati dieci pool di campioni con infezioni multiple di HPV identificate con SPF10-LiPA allo scopo di stimare l'abilità del metodo di pirosequenziamento 454 di individuare

i tipi di HPV nelle infezioni multiple. I risultati di questa analisi sono riportati in tab.VII. In molti casi, il metodo di pirosequenziamento 454 identificava tutti i tipi di HPV attesi identificati tramite SPF10-LiPA. Inoltre, in alcuni campioni, il metodo di pirosequenziamento 454 identificava i tipi di HPV che non rientrano nel target del metodo SPF-10 LiPA (ad esempio HPV55, HPV83), mentre in altri casi, il pirosequenziamento 454 non identificava alcuni tipi di HPV rilevati dal SPF10-LiPA. In questi casi, il pirosequenziamento 454 poteva aver perso tipi di HPV a bassa concentrazione, ma non si potevano escludere risultati falsi positivi dell'SPF10-LiPA. A questo proposito, per escludere contaminazioni, due campioni bianchi erano stati analizzati in ognuna delle 8 lane della corsa di pirosequenziamento 454 e non si erano osservate cross-contaminazioni.

Table VII. Risultati della genotipizzazione di HPV dei pool di tamponi genitali HPV positivi usando una miscela di primer PGMY modificata e il pirosequenziamento 454.

HPV types (SPF-10 LiPA)	primer MID	454 PGMY results (no. reads)
HPV6, 16, 31, 51, 52, 56	MID3	HPV6 (6557), HPV16 (203), HPV51 (18), HPV56 (22), HPV58 (21), HPV61 (92), HPV73 (38)
HPV6, 31, 56, (52)*	MID3	HPV6 (11451), HPV31 (52), HPV42 (38), HPV56 (33), HPV83 (32)
HPV16, 18, 52, 66, (51)*	MID3	HPV16 (11894), HPV18 (1316), HPV66 (41), HPV55 (25), HPV89 (37), HPV11 (25)
HPV16, 44, 53, 66	MID3	HPV16 (4357), HPV53 (251), HPV66 (14107), HPV44 (64)
HPV11, 16, 52, 58	MID3	HPV11 (44), HPV16 (8003), HPV58 (29)
HPV11, 52, 59, 66	MID3	HPV11 (23), HPV52 (30), HPV59 (9276), HPV66 (47)
HPV6, 11, 16, 44, 52	MID4	HPV6 (12805), HPV11 (19160), HPV16 (218), HPV35 (10), HPV39 (11), HPV44 (298), HPV52 (345), HPV54 (11), HPV58 (29), HPV61 (176), HPV62 (21), HPV83 (33)
HPV16, 44, 51, 52	MID5	HPV16 (9076), HPV44 (69), HPV51 (54), HPV52 (255), HPV61 (105)
HPV16, 52, 58, 70	MID7	HPV16 (1453), HPV42 (21), HPV52 (2506), HPV53 (27), HPV58 (1074), HPV61 (44), HPV66 (48), HPV70 (16063), HPV89 (32)
HPV6,11, 16, 31, 44, 51, 53, 66	MID8	HPV6 (7950), HPV11 (870), HPV16 (1402), HPV44 (124), HPV45 (335), HPV51 (212), HPV53 (117), HPV66 (5188)

*La presenza dei tipi di HPV tra parentesi è definite come possibile dal test SPF10-LiPA.

Analisi di campioni con risultati discordanti ai test HC2 e SPF10-LiPA

8 campioni con risultati discordanti ai test HC2 e SPF10-LiPA erano stati analizzati tramite pirosequenziamento 454. Real-time PCR tipo-specifiche erano state condotte per investigare ulteriormente risultati discrepanti. I risultati sono riportati in tab. VIII. Il pirosequenziamento 454 sembrava essere un metodo accurato e sensibile per l'identificazione e tipizzazione di HPV; permetteva di identificare la presenza di infezioni da HPV ad alto rischio non individuabili tramite HC2 o la presenza di alcuni tipi di HPV ad alto rischio, come HPV59, per i quali il sistema SPF10-LiPA ha bassa sensibilità. Il sistema di pirosequenziamento 454 rivelava anche risultati falsamente positivi ottenuti sia con HC2 (ad esempio positività ad HC2 in presenza di HPV a basso rischio o a possibile alto rischio come HPV53 e HPV66) che con SPF10-LiPA (ad esempio HPV51). Non era però esclusa completamente la presenza di infezioni a bassa carica da parte di qualche tipo di HPV non identificabile dal pirosequenziamento 454. A tale proposito testando i campioni del WHO Proficiency Panel si era visto che la sensibilità del metodo di pirosequenziamento 454 per HPV51 era relativamente bassa come riportato sopra. In fine, il metodo di pirosequenziamento 454 consentiva di individuare tipi di HPV che non possono essere identificati tramite SPF10-LiPA (ad esempio HPV61, HPV62, HPV67).

Tab. VIII. Risultati di genotipizzazione di HPV di tamponi cervical con risultati discordanti ottenuti tramite HC2 e SPF10-LiPA usando una mix di primer PGMY modificata e il pirosequenziamento 454.

HC2	SPF10-LiPA	Real-time PCR*	454 NGS typing (genotype/No. reads)
Negative	HPV31, 51, 66	HPV16 pos, HPV31 pos	HPV6 (6410), HPV16 (137), HPV31 (2118), HPV51 (1044), HPV66 (67)
Negative	HPV16, 31, 33, X	HPV16 pos	HPV11 (5253), HPV16 (38), HPV31 (37), HPV58 (28), HPV59 (770), HPV66 (17)
Negative	HPV16, 51	HPV16 pos	HPV16 (638), HPV51 (947), HPV61 (3448), HPV81 (5710)
Negative	HPV11, 51, 66	Negative	HPV11 (10117), HPV66 (538)
Positive	Negative	Negative	HPV59 (392), HPV62 (9015)
Positive	HPV66	Negative	HPV66 (12445)
Positive	HPV53	Negative	HPV53 (4423), HPV61 (4709)
Positive	HPVX	Negative	HPV62 (9142), HPV67 (2481)

*HPV types detected by type-specific real-time PCR: HPV16, HPV18, HPV31, HPV33, HPV45, HPV52, HPV58, HPV68b.

DISCUSSIONE

In questo studio, è stato sviluppato un nuovo metodo per la genotipizzazione di HPV basato sul pirosequenziamento 454 NGS dell'amplicone L1.

La scelta di utilizzare la piattaforma 454 FLX per lo sviluppo di nuovo un metodo di sequenziamento di HPV era stata valutata particolarmente idonea, per la lunghezza delle sequenze ottenibili tramite questa tecnologia, le quali possono arrivare a fino a 700 nucleotidi (lunghezza media delle sequenze 500 nucleotidi); tale dimensione non è raggiungibile qualora si utilizzino altre piattaforme NGS. Questa caratteristica consente di tipizzare HPV tramite una larga porzione del gene L1, incrementando l'affidabilità del risultato. Le altre piattaforme NGS possono essere impiegate con protocolli simili di tipizzazione di HPV, ma è necessario utilizzare primer che producono corti ampliconi, come gli SPF10 che mappano in una regione del gene L1 di 65pb.

Un primo protocollo è stato messo a punto e testato su campioni citologici cervicali, rappresentativi dei campioni testati nella pratica clinica. Tale protocollo disegnato utilizzando i primer degenerati MY09/11, è il primo esempio descritto in letteratura di sequenziamento di HPV tramite NGS (Barzon et al., 2011) e ne dimostra i vantaggi e l'utilità a fini clinici, epidemiologici e di ricerca.

Per incrementare il throughput e ridurre i costi della metodica, si è scelto di sviluppare un protocollo di sequenziamento multiplex usando primer dotati di barcode, che consentisse di genotipizzare fino a 192 campioni in una singola corsa. Questo protocollo ha dimostrato, grazie ai test condotti sui controlli positivi per DNA di HPV16 e di HPV18, di poter rilevare tipi di HPV presenti in infezioni multiple con una frequenza dell'1% e con una sensibilità di 100 genomi equivalenti/ μ l.

La metodica così sviluppata ha però dimostrato di avere alcuni limiti importanti, legati in particolare alla scelta dei primer degenerati utilizzati per il disegno dei fusion primer. I primer MY09/11 hanno infatti una scarsa sensibilità per alcuni tipi virali, come HPV51 e HPV52 (Gravitt et al., 2000), dovuta alla presenza di *mismatches* tra la sequenza stessa del primer e le sequenze di questi tipi virali (Qu et al., 1997).

Il disegno dei primer rimane quindi uno dei punti critici per la tipizzazione di HPV anche quando si utilizza un sistema di sequenziamento estremamente sensibile. Per tale ragione si è deciso di modificare il protocollo e di ridisegnare i fusion primer, utilizzando i primer PGMY09/PGMY11. Questa miscela di primer, che mappano nella stessa regione del gene L1,

sono stati disegnati per eliminare la degenerazione dei primer MY09/11 ed aumentarne la sensibilità, specificità e riproducibilità (Gravitt et al., 2000). Per la messa a punto di questo protocollo sono state seguite strettamente le linee guida e i protocolli del WHO (Human papillomavirus laboratory manual-First edition, 2009) e la metodica è stata validata su campioni clinici, ma soprattutto sui controlli del WHO proficiency panel del 2010 e del 2011.

Il metodo può identificare correttamente tutti i tipi di HPV ad alto rischio, incluso HPV68a, che non è amplificabile con i primer PGMY. L'amplificazione di HPV68a è stata ottenuta grazie all'aggiunta di nuovi primer specifici al set di primer multipli PGMY. L'aggiunta di questi nuovi primer non riduce la sensibilità del metodo per gli altri tipi di HPV ad alto rischio.

Testando diverse mix di linee cellulari positive per HPV16 e HPV18, il metodo di pirosequenziamento 454 ha dimostrato di identificare in modo riproducibile le proporzioni di ogni tipo di HPV nelle infezioni multiple, con un range dinamico relativamente alto. Tuttavia, alcuni tipi di HPV, come HPV31, HPV33, HPV51 e HPV56, sembrano essere meno efficientemente amplificati, con una conseguente loro sottostima. Tuttavia il metodo di pirosequenziamento 454 può identificare correttamente ogni tipo di HPV del WHO proficiency panel ad una bassa concentrazione di GE/5 µl. Un ampio numero di differenti tipi di HPV può essere trovato in campioni con infezioni multiple.

Il sequenziamento tramite NGS ha dimostrato una più alta specificità nella genotipizzazione rispetto ai metodi basati sull'ibridazione, essendo meno pronò ad errori di miss-identificazione. Nella prima parte di questo studio i risultati ottenuti con il sequenziamento 454 sono stati confrontati con quelli del sequenziamento tradizionale, dimostrando una buona concordanza e quindi una buona accuratezza. Al contrario, confrontati con i risultati ottenuti dal test LiPA, hanno rilevato errori di identificazione in alcuni campioni nei quali il sequenziamento 454 aveva identificato, come singola infezione, un tipo di HPV non incluso tra quelli identificati dal test LiPA. Altri studi mostrano come il test LiPA abbia dei problemi di cross reattività in presenza di tipi virali non inclusi tra le sonde del test. Inoltre il metodo NGS ha dimostrato di avere una più alta sensibilità dei metodi basati sull'ibridazione inversa in presenza di infezioni multiple, poiché il test LiPA ha una scarsa sensibilità per i tipi di HPV che rappresentano meno del 5% del numero di copie virali nelle infezioni multiple.

Questo studio dimostra che il metodo NGS è capace di identificare la presenza di differenti varianti dello stesso tipo di HPV in un campione. Nonostante siano state descritte raramente in letteratura (Gagnon et al., 2005; Ho et al., 1991; Xi et al., 1995), le infezioni con varianti multiple dello stesso tipo di HPV hanno una certa rilevanza epidemiologica e diagnostica. Alcune varianti

intratipo sono state associate a specificità geografica, alle migrazioni (Ho et al., 1993; Ong et al., 1993), ad un'incrementata oncogenicità (Bernard et al., 2006; Lichtig et al., 2006) e possono essere responsabili di un fallimento della vaccinazione profilattica. (Pillai et al., 2009).

L'analisi delle varianti di sequenza di L1, che sono state condotte nella prima parte dello studio, dimostrano che il sequenziamento 454 può generare dei falsi positivi, specialmente nel caso di inserzioni/delezioni. Queste ultime sono infatti caratterizzate da un basso *quality score*. Utilizzando la chimica 454 FLX pyrosequencing, sostituzioni puntiformi sono più rare rispetto alle inserzioni/delezioni e più raramente dovute ad artefatti di PCR, non identificabili tramite il *quality score*. La tecnologia NGS è in continua evoluzione con continui miglioramenti ed è probabile che questo problema di accuratezza venga risolto in tempi rapidi.

Nel corso di questo studio non è stato possibile identificare nuovi tipi di HPV. L'applicazione di tale metodica, utilizzando in particolare la seconda versione del protocollo dimostratasi più sensibile, ad un'ampia casistica potrebbe permettere di individuare nuovi tipi virali. Lo dimostra anche il lavoro condotto da Ekstrom et al., (2011) che, usando i primer FAP specifici per gli HPV a tropismo cutaneo ed un approccio analogo su piattaforma 454, ha individuato nuovi tipi virali in lesioni cutanee.

Il metodo è stato utile per l'analisi di campioni con risultati discordanti ottenuti con i test HC2 e SPF10-LiPA. Per questi due metodi, basati sull'ibridazione, sono stati riportati rispettivamente problemi di cross ibridazione e di bassa sensibilità per alcuni tipi di HPV (Eklund et al., 2010-2012; Castle et al., 2008-2002; Barzon et al., 2012). Infatti, i metodi basati sul sequenziamento sono più specifici di quelli basati sull'ibridazione e sono considerati il *gold standard* per la genotipizzazione. Il nostro metodo che utilizza il sequenziamento di un frammento di 450-bp di lunghezza del gene L1, permette di ottenere un'ottima specificità di genotipizzazione e potenzialmente di scoprire nuovi tipi di HPV per i quali una sequenza lunga può fornire sufficienti informazioni per un'analisi filogenetica (Bernard et al., 1994; de Villiers et al 2004). Questo, come già detto, rappresenta un vantaggio rispetto a metodi basati sull'analisi di sequenze molto corte. Un esempio è il test commerciale HPV sign® Genotyping (Qiagen), che si basa sull'amplificazione ad ampio spettro di una regione variabile di L1 di HPV e sul pyrosequenziamento di corte sequenze dell'amplicone, usando il sistema Pyromark Q24 (Qiagen). Il test HPV sign® Genotyping ha dimostrato di avere una bassa sensibilità per alcuni tipi di HPV e per la loro discriminazione HPV in infezioni multiple. Il test ha però il vantaggio che, essendo basato sul sequenziamento di corti ampliconi, può essere utile per analizzare biopsie incluse in paraffina e fissate in formalina (Barbieri et al., 2012).

Meiring et al. (2012) hanno recentemente dimostrato, usando rolling circle amplification (RCA) e Illumina sequencing technology, che ci sono dei tipi di HPV con sostanziale prevalenza in donne HIV positive i quali non vengono comunemente identificati negli studi molecolari epidemiologici condotti usando i kit commerciali. Il significato di questi tipi in relazione alla malattia cervicale rimane da investigare, ma dimostra la presenza di un *bias* negli studi epidemiologici condotti con metodi di tipizzazione a bassa sensibilità.

Nella nostra esperienza, confrontato con i metodi di tipizzazione convenzionali, il metodo NGS ha un tempo di *turnaround* per campione più lungo e costi più alti. Tali costi sono principalmente correlati al costo della strumentazione, mentre il costo dei reagenti è simile per i tre metodi ed è stimato circa di 40 euro per campione. I tre metodi richiedono 2 h per la PCR; dopo l'amplificazione il test LiPA è abbastanza veloce e necessita di altre 3 ore per ibridazione e analisi di 48 campioni; il metodo di sequenziamento tradizionale richiede 4 ore per la purificazione tramite big-dye e per l'elettroforesi capillare di 16 campioni; il metodo 454 richiede infine 22 ore in 3 giorni per una corsa di sequenziamento, ma consente di analizzare potenzialmente fino a 192 campioni in una singola corsa. Le *strip* del test LiPA vengono scannerizzate e valutate tramite Line Reader e Analysis Software (Innogenetics), che permettono un'analisi più rapida dei risultati di tipizzazione rispetto al sequenziamento tradizionale e al metodo 454 NGS. Non appena sarà disponibile una maggiore automazione e standardizzazione dei protocolli, una riduzione dei costi con un incremento del *throughput* e con lo sviluppo di strumenti meno costosi, come il GS Junior System (Roche 454 Life Sciences), è ipotizzabile che le piattaforme NGS diventeranno di uso comune nella routine della diagnostica molecolare.

Le possibili applicazioni di questo metodo di genotipizzazione includono: (i) test di secondo livello in pazienti con lesioni correlate ad HPV, per la stratificazione del rischio, (ii) accurata analisi di tipi di HPV in carcinomi correlati con HPV, (iii) monitoraggio epidemiologico della distribuzione mucosale dei tipi di HPV, (iv) identificazione di tipi di HPV non comuni o nuovi; (v) qualsiasi altra applicazione che richieda un'accurata tipizzazione di HPV e la stima del relativo carico nelle infezioni multiple.

In conclusione, è stato sviluppato un metodo molto accurato per la tipizzazione di HPV basato sul pirosequenziamento di ampliconi del gene L1 virale. Questo metodo ha un'alta sensibilità e specificità nell'identificazione di HPV ad alto rischio sia in infezioni singole che multiple. Inoltre, permette di stimare la relativa percentuale di ogni tipo di HPV presente nelle infezioni multiple, identifica un ampio range di tipi di HPV e può potenzialmente scoprire nuovi tipi virali.

BIBLIOGRAFIA

Barbieri D, Nocera M, Gallinella G, Gentilomi GA, Plazzi M, Costa S, Santini D, Venturoli S. Comparison of HPV sign Genotyping Test with INNO-LiPA HPV Genotyping Extra assay on histologic and cytologic cervical specimens. *Diagn Microbiol Infect Dis* 2012; 74:43-48.

Barzon L, Giorgi C, Buonaguro FM, Palù G, Italian Society for Virology. Guidelines of the Italian Society for Virology on HPV testing and vaccination for cervical cancer prevention. *Infect Agent Cancer* 2008; 3:14.

Barzon L, Lavezzo E, Militello V, Toppo S, Palù G. Applications of Next-Generation Sequencing Technologies to Diagnostic Virology. *Int. J. Mol. Sci.* 2011, 12, 7861-7884.

Barzon L, Militello V, Lavezzo E, Franchin E, Peta E, Squarzon L, Trevisan M, Pagni S, Dal Bello F, Toppo S, Palù G. Human papillomavirus genotyping by 454 next generation sequencing technology. *J Clin Virol.* 2011;52(2):93-7.

Barzon L, Militello V, Pagni S, Franchin E, Dal Bello F, Mengoli C, Palù G. Distribution of human papillomavirus types in the anogenital tract of females and males. *J Med Virol.* 2010 Aug;82(8):1424-30.

Barzon L, Militello V, Pagni S, Palù G. "Comparison of INNO-LiPA genotyping extra and hybrid capture 2 assays for detection of carcinogenic human papillomavirus genotypes." *J Clin Virol.* 2012 Nov;55(3):256-61. Epub 2012 Aug 9.

Baseman JG, Koutsky LA. "The epidemiology of human papillomavirus infections." *J Clin Virol,* 2005;32 Suppl 1:S16-24. Review.

Bernard HU, Chan SY, Manos MM, Ong CK, Villa LL, Delius H, Peyton CL, Bauer HM, Wheeler CM. Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction amplification restriction fragment length polymorphisms, nucleotide sequences and phylogenetic algorithms. *J Infect Dis* 1994; 170:1077-1085.

Berumen J, Ordoñez RM, Lazcano E, Salmeron J, Galvan SC, Estrada RA, Yunes E, Garcia-Carranca A, Gonzalez-Lira G, Madrigal-de la Campa A. Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study. *J Natl Cancer Inst.* 2001; 93(17):1325-30.

Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Benbrahim-Tallaa L, Guha N, Freeman C, Galichet L, Coglianò V; WHO International Agency for Research on Cancer Monograph Working Group. "A review of human carcinogens - Part B: biological agents." *Lancet Oncol* 2009;10:321-2.

Brink AATP, Snijders PJF, Meijer CJLM. "HPV detection methods." *Disease Markers,* 2007: 273-281.

Castle PE, Schiffman M, Burk RD, Wacholder S, Hildesheim A, Herrero R, et al. Restricted cross-reactivity of hybrid capture 2 with nononcogenic human papillomavirus types. *Cancer Epidemiol Biomarkers Prev* 2002; 11:1394-1399.

Castle PE, Solomon D, Wheeler CM, Gravitt PE, Wacholder S, Schiffman M. Human papillomavirus genotype specificity of hybrid capture 2. *J Clin Microbiol* 2008; 46:2595-2604.

Clifford GM, R.K. Rana, S. Franceschi, J.S. Smith, G. Gough, J.M. Pimenta Human papillomavirus genotype distribution in low-grade cervical lesions: comparison by geographic region and with cervical cancer. *Cancer epidemiol Biomarkers Prev*, 14 (5) (May 2005), pp. 1157–1164

de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, Tous S, Felix A, Bravo LE, Shin HR, Vallejos CS, de Ruiz PA, Lima MA, Guimera N, Clavero O, Alejo M, Llombart-Bosch A, Cheng-Yang C, Tatti SA, Kasamatsu E, Iljazovic E, Odida M, Prado R, Seoud M, Grce M, Usubutun A, Jain A, Suarez GA, Lombardi LE, Banjo A, Menéndez C, Domingo EJ, Velasco J, Nessa A, Chichareon SC, Qiao YL, Lerma E, Garland SM, Sasagawa T, Ferrera A, Hammouda D, Mariani L, Pelayo A, Steiner I, Oliva E, Meijer CJ, Al-Jassar WF, Cruz E, Wright TC, Puras A, Llave CL, Tzardi M, Agorastos T, Garcia-Barriola V, Clavel C, Ordi J, Andújar M, Castellsagué X, Sánchez GI, Nowakowski AM, Bornstein J, Muñoz N, Bosch FX; Retrospective International Survey and HPV Time Trends Study Group. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol* 2010; 11:1048-1056.

de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. *Virology* 2004; 324:17-27.

De Vuyst H, Gichangi P, Estambale B, Njuguna E, Franceschi S, Temmerman M. “Human papillomavirus types in women with invasive cervical carcinoma by HIV status in Kenya.” *Int J Cancer* 2008, 122(1):244-6.

Depuydt CE, Benoy IH, Bailleul EJ, Vandepitte J, Vereecken AJ, Bogers JJ. “Improved endocervical sampling and HPV viral load detection by Cervex-Brush Combi”. *Cytopathology*. 2006 Dec;17(6):374-81.

Doorbar J. “The papillomavirus life cycle.” *J. Clin. Virol*, 2005; 32S: S7–S15.

Eklund C, Forslund O, Wallin KL, Zhou T, Dillner J; for the WHO Human Papillomavirus Laboratory Network. Global proficiency study of human papillomavirus genotyping in vaccinology. *J Clin Microbiol* 2012; 50:2289-2298.

Eklund C, Zhou T, Dillner J; WHO Human Papillomavirus Laboratory Network. Global proficiency study of human papillomavirus genotyping. *J Clin Microbiol* 2010; 48:4147-4155.

Ekstrom, J; Bzhalava D; Svenback D; Forslund O; Dillner J. High throughput sequencing diversity of Human Papillomaviruses in cutaneous lesions. *Int J Cancer*. 2011 Dec 1;129(11):2643-50

Ferguson M, Wilkinson DE, Zhou T. WHO meeting on the standardization of HPV assays and the role of the WHO HPV Laboratory Network in supporting vaccine introduction held on 24-25 January 2008, Geneva, Switzerland. *Vaccine* 2009; 27: 337-347.

Gagnon S, Hankins C, Tremblay C, Pourreaux K, Forest P, Rouah F, et al. Polymorphism of human papillomavirus type 31 isolates infecting the genital tract of HIV-seropositive and HIV-seronegative women at risk for HIV infection. *J Med Virol* 2005;75:213–21.

Gargiulo F, De Francesco MA, Schreiber C, Ciravolo G, Salinaro F, Valloncini B, Manca N. “Prevalence and distribution of single and multiple HPV infections in cytologically abnormal cervical samples from Italian women.” *Virus Res* 2007, 125(2):176-82.

Geraets D, Alemany L, Guimera N, de Sanjose S, de Koning M, Molijn A, Jenkins D, Bosch X, Quint W; on behalf of the RIS HPV TT study group. Detection of rare and possibly carcinogenic human papillomavirus genotypes as single infections in invasive cervical cancer. *J Pathol* 2012 Jun 19. doi: 10.1002/path.4065. [Epub ahead of print]

Gudleviciene Z, Didziapetriene J, Ramael M, Uleckiene S, Valuckas KP. “Human papillomavirus and p53 polymorphism in Lithuanian cervical cancer patients.” *Gynecol Oncol* 2006, 102(3):530-3.

Gravitt P, Peyton CL, Alessi TQ, Wheeler C, Coutlée F, Hildesheim A, Schiffman M, Scott D R, Apple RJ. Improved amplification of genital human papillomaviruses. 2000.*J. Clin. Microbiol.* 38:357–361.)

Guan P, Howell-Jones R, Li N, Bruni L, de Sanjosé S, Franceschi S, Clifford GM. Human papillomavirus types in 115,789 HPV-positive women: A meta-analysis from cervical infection to cancer. *Int J Cancer* 2012; 131:2349-2359.

Jacobs MV, Snijders PJ, van den Brule AJ, Helmerhorst TJ, Meijer CJ, Walboomers JM. “A general primer GP5+/GP6(+)-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings”. *J Clin Microbiol*, 1997;35:791-5.

Josefsson AM, Magnusson PK, Ylitalo N, Sørensen P, Qwarforth-Tubbin P, Andersen PK, Melbye M, Adami HO, Gyllensten UB. “Viral load of human papilloma virus 16 as a determinant for development of cervical carcinoma in situ: a nested case-control study”. *Lancet* 2000;355:2189 – 93.

Hart KW, Williams OM, Thelwell N, Fiander AN, Brown T, Borysiewicz LK. “Novel method for detection, typing, and quantification of human papillomavirus in clinical samples.” *J Clin Microbiol*, 2001; 39:3204–12.

Howley PM, Douglas RL. "Papillomaviruses". In: *Fields Virology 5th edition*, Ed. Knipe D & Howley PM. *Lippincott-Williams & Wilkins, Philadelphia*, 2007; p. 2299-2354.

Ho L, Chan SY, Burk RD, Das BC, Fujinaga K, Icenogle JP, et al. "The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations". *J Virol* 1993;**67**:6413–23.

Ho L, Chan SY, Chow V, Chong T, Tay SK, Villa LL, et al. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. *J Clin Microbiol* 1991;**29**:1765–72.

Xi LF, Demers GW, Koutsky LA, Kiviat NB, Kuypers J, Watts DH, et al. Analysis of human papillomavirus type 16 variants indicates establishment of persistent infection. *J Infect Dis* 1995;**172**:747–55.

Hoory T, Monie A, Gravitt P, Wu TC. "Molecular epidemiology of human papillomavirus". *J Formos Med Assoc.* 2008; 107:198-217.

Huh W, Einstein MH, Herzog TJ, Franco EL. What is the role of HPV typing in the United States now and in the next five years in a vaccinated population? *Gynecol Oncol.* 2010 Jun;117(3):481-5.

Kleter B, van Doorn LJ, ter Schegget J. "A novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses." *Am J Pathol*, 1998, 153:1731-9.

Lindh M, Gorander S, Andersson E, Horal P, Mattsby-Balzer I, Ryd W. "Real-time PCR targeting 14 human papilloma virus types." *J Clin Virol*, 2007;40(4):321-4.

Longworth MS, Laimins LA. "The binding of histone deacetylases and the integrity of zinc finger-like motifs of the E7 protein are essential for the life cycle of human papillomavirus type 31". *J Virol*, 2004; 78:3533-41.

Lorincz AT, Castle PE, Sherman ME, Scott DR, Glass AG, Wacholder S, Rush BB, Gravitt PE, Schussler JE, Schiffman M. "Viral load of human papillomavirus and risk of CIN3 or cervical cancer." *Lancet*, 2002; 360(9328):228-9.

Manos MM, Ting Y, Wright DK, Lewis AJ, Broker TR, Wolinsky SM. "The use of polymerase chain reaction amplification for detection of genital human papillomaviruses." *Cancer Cells* 1989; 7:209–14.

Meiring TL, Salimo AT, Coetzee B, Maree HJ, Moodley J, Hitzeroth II, Freeborough MJ, Rybicki EP, Williamson AL. Next-generation sequencing of cervical DNA detects human papillomavirus types not detected by commercial kits. *Virol J.* 2012 Aug 16;9:164.

- Meijer CJ, Heideman DA, Berkhof H, Snijders PJ. "Prevention of cervical cancer: where immunology meets diagnostics". *Immunol Lett*. 2009 21;122(2):126-7.
- Meijer CJ, Snijders PJ, Castle PE. "Clinical utility of HPV genotyping." *Gynecol Oncol*, 2006; 103:12-7.
- Militello V, Trevisan M, Squarzon L, Biasolo MA, Rugge M, Militello C, Palù G, Barzon L. Investigation on the presence of polyomavirus, herpesvirus, and papillomavirus sequences in colorectal neoplasms and their association with cancer. *Int J Cancer*. 2009;124(10):2501-3.
- Molijn A, Kleter B, Quint W, van Doorn LJ. "Molecular diagnosis of human papillomavirus (HPV) infections." *J Clin Virol*, 2005; 32 suppl 1:S43-S51.
- Monsonogo J, Bohbot JM, Pollini G. "Performance of the Roche AMPLICOR human papillomavirus (HPV) test in prediction of cervical neoplasia (CIN) in women with abnormal PAP smear." *Gynecol Oncol*, 2005; 99(1):160-168.
- Pagliusi SR, Garland SM. "International standard reagents for HPV detection." *Disease Markers*, 2007; 23(4):283-96.
- Poljak M, Marin IJ, Seme K, Vince A. Hybrid Capture II HPV test detects at least 15 human papillomavirus genotypes not included in its current high-risk probe cocktail. *J Clin Virol* 2002; 25 Suppl 3:S89-S97.
- Poljak M, Cuzick J, Kocjan BJ, Iftner T, Dillner J, Arbyn M. "Nucleic acid test for the detection of alpha Human Papillomaviruses" *Vaccine* 2012;30S F100-6.
- Ong CK, Chan SY, Campo MS, Fujinaga K, Mavromara-Nazos P, Labropoulou V, et al. Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. *J Virol* 1993;**67**:6424–31.
- Bernard H-U, Calleja-Macias EI, Dunn ST. Genome variation of human papillomavirus types: phylogenetic and medical implications. *Int J Cancer* 2006;**118**:1071–6.
- Lichtig H, Algrisi M, Botzer LE, Abadi T, Verbitzky Y, Jackman A, et al. HPV16 E6 natural variants exhibit different activities in functional assays relevant to the carcinogenic potential of E6. *Virology* 2006;**350**:216–27.
- Pillai MR, Hariharan R, Babu JM, Lakshmi S, Chiplunkar SV, Patkar M, et al. Molecular variants of HPV-16 associated with cervical cancer in Indian population. *Int J Cancer* 2009;**125**:91–103.
- Qu W, Jiang G, Cruz Y, Chang CJ, Ho GY, Klein RS, et al. PCR detection of human papillomavirus: comparison between MY09/MY11 and GP5+/GP6+ primer systems. *J Clin Microbiol* 1997;**35**:1304–10.

Ronaghi M, Uhlén M, Nyrén P. “A sequencing method based on real-time pyrophosphate.” *Science*, 1998; 281(5375):363-365.

Saslow D, Solomon D, Lawson HW, Killackey M, Kulasingam SL, Cain J, Garcia FA, Moriarty AT, Waxman AG, Wilbur DC, Wentzensen N, Downs LS Jr, Spitzer M, Moscicki AB, Franco EL, Stoler MH, Schiffman M, Castle PE, Myers ER; ACS-ASCCP-ASCP Cervical Cancer Guideline Committee. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *CA Cancer J Clin* 2012; 62:147-172.

Schiffman M, Clifford G, Buonaguro FM. “Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline.” *Infect Agent Cancer*. 2009;4:8.

Schmitz M, Scheungraber C, Herrmann J, Teller K, Gajda M, Runnebaum IB, Dürst M. Quantitative multiplex PCR assay for the detection of the seven clinically most relevant high-risk HPV types. *J Clin Virol* 2009; 44:302-307.

Snijders PJF, Van Den Brule AJC, Schrijnemakers HFJ, Snow G, Meijer CJLM, Walboomers JMM. “The use of general primers in the polymerase chain reaction permits the detection of a broad spectrum of human papillomavirus genotypes.” *J General Virol*, 1990; 71:173-181.

Stacey SN, Jordan D, Williamson AJ, Brown M, Coote JH, Arrand JR. “Leaky scanning is the predominant mechanism for translation of human papillomavirus type 16 E7 oncoprotein from E6/E7 bicistronic mRNA”. *J Virol*. 2000; 74:7284-97.

Wentzensen N, Schiffman M, Dunn ST, Zuna RE, Walker J, Allen RA, Zhang R, Sherman ME, Wacholder S, Jeronimo J, Gold MA, Wang SS. “Grading the severity of cervical neoplasia based on combined histopathology, cytopathology, and HPV genotype distribution among 1,700 women referred to colposcopy in Oklahoma.” *Int J Cancer*, 2009, 124(4):964-9.

WHO. Human Papillomavirus Laboratory Manual. First Edition, 2009. Available at www.who.int/vaccines-documents/

LISTA PUBBLICAZIONI E PRESENTAZIONI AI CONGRESSI PRODOTTE DURANTE I TRE ANNI DI DOTTORATO

Pubblicazioni in extenso:

1. Barzon L, **Militello V**, Pagni S, Franchin E, Dal Bello F, Mengoli C, Palù G. “Distribution of human papillomavirus types in the anogenital tract of females and males” *J Med Virol* 82(8):1424-30 (2010). **IF. 2.82**
2. Barzon L, Squarzon L, **Militello V**, Trevisan M, Palù G. “WU and KI viruses still remain orphan in adults.” Author reply. *J Infect Dis* (2010). **IF 6.41**
3. Barzon L, Lavezzo E, **Militello V**, Toppo S, Palù G. “Applications of next-generation sequencing technologies to diagnostic virology.” *Int J Mol Sci.*;12(11):7861-84 (2011). **IF. 2.598**
4. Barzon L, **Militello V**, Lavezzo E, Franchin E, Peta E, Squarzon L, Trevisan M, Pagni S, Dal Bello F, Toppo S, Palù G. “Human papillomavirus genotyping by 454 next generation sequencing technology.” *J Clin Virol.*; 52(2):93-7 (2011). **IF. 3.969**
5. Laroni A , Giacomazzi CG, Grimaldi L, Gallo P, Sormani MP, Bertolotto A, McDermott JL, Gandoglia I, Martini I, Vitello G, Rinaldi F, Barzon L., **Militello V**, Pizzorno M, Bandini F, Capello E, Palù G, Uccelli A, Mancardi GL, and Varnier OE “Urinary JCV-DNA testing during natalizumab treatment may increase accuracy of PML risk stratification” *j. of Neuroimmune Pharmacol.*; 7(3):665-72 (2012). **IF. 4.57**
6. *Gervasi MT, Romero R, Bracalente G, Chaiworapongsa T, Erez O, Dong Z, Hassan SS, Yeo L, Yoon HB, Mor G, Barzon L, Franchin E, **Militello V**, Palù G.* “Viral invasion of the amniotic cavity (VIAC) in the midtrimester of pregnancy.” *J Matern Fetal Neonatal Med.*;25(10):2002-13 (2012). **IF. 1.36**
7. *Barzon L, **Militello V**, Pagni S, Palù G.* “Comparison of INNO-LiPA Genotyping Extra and Hybrid Capture 2 assays for detection of carcinogenic human papillomavirus genotypes.” *J Clin Virology.*; 55(3):256-61 (2012). **IF. 3.969**
8. *Casarotto A, **Militello V**, Piatto G, Gruppo M, Militello C.* “Hernioplasty in elderly high-risk adults: efficacy of fibrin glue.” *J Americ Geriatrics Society.* 06/2012; 60(6):1193-4. · **IF. 3.74**

Abstract pubblicati su rivista internazionale:

1. L. Barzon, **V. Militello**, E. Peta, E. Lavezzo, S. Toppo, M. Trevisan, L. Squarzon, E. Franchin, G. Palù “Deep sequencing for accurate and high-throughput HPV genotyping in clinical samples” 20th ECCMID congress 2010, CMI Clinical Microbiology and Infection, Volume 16, Issue Supplement s2, P. S26.
2. L. Barzon, **V. Militello**, S. Pagni, G. Palu` “Comparison of the INNO-LiPA genotyping extra and the Hybrid Capture 2 assays for detection of carcinogenic human papillomavirus genotypes” 22nd ECCMID London 2012, CMI Clinical Microbiology and Infection, April 2012 Volume 18, Issue Supplement s3 Pages 210. P937
3. P1373 M.C. Scarpa, E. Lavezzo, E. Franchin, **V. Militello**, G. Costanzi, M. Pacenti, M. Saetta, P. Maestrelli, G. Palu`, L. Barzon “Dysbiosis of the lung bacterial community in patients with stable

chronic obstructive pulmonary disease” 22ndECCMID London 2012, CMI Clinical Microbiology and Infection, April 2012 Volume 18, Issue Supplement s3 Pages 357.

Poster e presentazioni:

1. L. Barzon, **V. Militello**, E. Peta, E. Lavezzo, S. Toppo, M. Trevisan, L. Squarzon, E. Franchin, G. Palù “Deep sequencing for accurate and high-throughput HPV genotyping in clinical samples” (Abstr. 2224) **oral presentation** at the ECCMID meeting -Vienna 10 - 13 April 2010. **Oral presentation.**
2. E. Lavezzo, **V. Militello**, E. Peta, M. Trevisan, L. Squarzon, E. Franchin, S. Toppo, L. Barzon, G. Palù. “Deep sequencing for accurate and high-throughput HPV genotyping in clinical samples” (Abstr. 763) has been accepted as a poster presentation at the 4th European Congress of Virology to be held in Cernobbio (Como) from 7 - 11 April 2010. **Poster presentation.**
3. E. Peta, **V. Militello**, A. Sinigaglia, M. Trevisan, G. Masi, L. Barzon, G. Palù. “High-risk human papillomavirus E6 and E7 genes modulate expression of cellular microRNAs involved in tumorigenesis” (Abstr. 330) has been accepted as a poster presentation at the 4th European Congress of Virology to be held in Cernobbio (Como) from 7 - 11 April 2010. **Poster presentation.**
4. **V.Militello**, Peta E., Sinigaglia A., Trevisan M., Barzon L., Palù G. “Analysis of microRNA expression in cytological cervical samples and correlation with HPV infection” (Abstr. 700) has been accepted as a poster presentation at the 4th European Congress of Virology to be held in Cernobbio (Como) from 7 - 11 April 2010. **Poster presentation.**
5. **V.Militello**, Lavezzo E., Franchin E., Peta E., Squarzon L., Trevisana M., Pagnia S., Dal Bello F., Toppo S., Barzon L., Palù G. “Human Papillomavirus Genotyping by 454 Next Generation Sequencing Technology” Orvieto SIV 2011. **Poster presentation.**
6. Applications of 454 next-generation sequencing to clinical microbiology and virology. DOTTORANDO: Dr. Valentina MILITELLO - TUTOR: Prof.ssa Luisa BARZON Scuola di Dottorato in Biomedicina CONVEGNO DI PRESENTAZIONE DELL’ATTIVITÀ DI RICERCA DEI DOTTORANDI DEL XXV CICLO DELLA MACROAREA 2 SCIENZE DELLA VITA Area Medica (II anno di corso)- 30 Novembre 2011, Padova. **Poster presentation.**
7. P937 L. Barzon, **V. Militello**, S. Pagni, G. Palu` “Comparison of the INNO-LiPA genotyping extra and the Hybrid Capture 2 assays for detection of carcinogenic human papillomavirus genotypes” 22ndECCMID London 2012, CMI April 2012 Volume 18, Issue Supplement s3 Pages 210. **Poster presentation.**
8. P1373 M.C. Scarpa, E. Lavezzo, E. Franchin, **V. Militello**, G. Costanzi, M. Pacenti, M. Saetta, P. Maestrelli, G. Palu`, L. Barzon “Dysbiosis of the lung bacterial community in patients with stable chronic obstructive pulmonary disease” 22ndECCMID London 2012 April 2012 Volume 18, Issue Supplement s3 Pages 357. **Poster presentation.**
9. P026 **V. Militello**, Lavezzo E., Costanzi G., Franchin E., Toppo S., Palù G., Barzon L. “Accurate genotyping of human papillomavirus by 454 next generation Sequencing” 11th National Congress of the Italian Society for Virology. Orvieto, September 17-19, 2012. **Poster presentation.**
10. P044 Peta E., **V.Militello**, Sinigaglia A., Cappellesso R., Fassan M., Fassina A., Palù G., Barzon L. “Expression of Human Papillomavirus type 16-related microRNAs in penile squamous carcinoma”

11th National Congress of the Italian Society for Virology. Orvieto, September 17-19, 2012. **Poster presentation.**