

# **Vaccinegate:**

## Report analisi metagenomiche su Gardasil 9



## Presentazione in breve dei risultati

Con queste analisi siamo arrivati alla conclusione dello screening di primo livello del vaccino Gardasil 9.

### Risultati

Vi è materiale genetico avventizio presente in quantità residuali. Si possono riassumere i seguenti punti essenziali:

#### Materiale genetico avventizio presente come DNA:

- **Batteri:** la percentuale è rilevante: 54% del DNA totale, questa contaminazione può provenire principalmente dalla coltura di lievito, ma anche da contaminanti presenti nel laboratorio; sono stati effettuati più bianchi per ridurre al minimo l'errore dovuto alla contaminazione ambientale, ma avremo un dato più accurato quando faremo i replicati con altri laboratori. Il DNA batterico potrebbe interagire con l'alluminio adiuvante e causare allergie, infiammazione e autoimmunità. Dato da confermare.
- **DNA umano e di topo:** la loro origine non è nota! Può essere che il DNA umano provenga dai residui presenti nelle materie prime del virus HPV prelevato da tessuti umani. Il DNA di topo potrebbe invece essere una contaminazione crociata da altre linee cellulari utilizzate per la produzione di vaccini (è un'ipotesi). Questi DNA potrebbero interagire con l'alluminio adiuvante e causare reazioni infiammatorie e autoimmuni.
- **Virus avventizi:** Frammento L1 del virus HPV di DNA a doppia catena - proviene dal processo di fabbricazione degli antigeni; è un contaminante perché pone dei problemi di sicurezza in quanto non viene degradato e rimane a lungo nei macrofagi legato all'alluminio adiuvante; il suo effetto biologico non è del tutto noto ma può probabilmente integrarsi nel DNA dell'ospite, stimolare l'infiammazione attraverso la produzione di citochine proinfiammatorie e reazioni autoimmunitarie (vedi ricerche del prof. Lee).
- **Fagi:** provengono dal processo di produzione, sono contaminanti avventizi di pericolosità non nota. Anticorpi prodotti contro i fagi possono interagire con i batteri della flora batterica intestinale? Possono integrarsi nel DNA batterico della flora batterica?
- **Molluscum contagiosum virus:** appartiene alla famiglia Poxviridae, sottofamiglia Chordopoxvirinae, genere Molluscipoxvirus. Il termine pox contenuto nel nome di questi virus, deriva dalle vescicole (in inglese: poxes) prodotte dal virus del vaiolo (smallpox).
- **Retrovirus:** potenzialmente integrati nel DNA; possono causare trasformazione neoplastica e mutazioni del genoma dell'ospite; provengono dalle contaminazioni di DNA umano e di topo, quali possibili cross contaminazioni con altre linee cellulari.
- **Virus della leucemia del topo.**
- **Retrovirus endogeno umano K.**

#### Materiale genetico avventizio presente come RNA:

- **Batteri:** i trascritti stanno ad indicare che il DNA è funzionante.
- **Costrutti sintetici (sequenze artificiali):** possono provenire dal processo di produzione degli antigeni mediante ricombinazione genetica con il plasmide; sono potenzialmente in grado di ricombinarsi con DNA umano; il legame con l'alluminio adiuvante può prolungarne e potenziarne l'effetto biologico (infiammazione e autoimmunità).
- **Lievito e suoi virus (virus L-BC e narnavirus):** l'RNA di lievito può dare origine a proteine allergeniche (che si possono legare all'alluminio adiuvante), mentre i virus non sono noti per gli effetti sulle cellule umane e sul microbiota.
- **Virus dell'anemia equina infettiva e virus della leucemia del topo:** (quest'ultimo è presente sia come DNA che RNA e quindi è un virus funzionante); questi virus provengono dalle contaminazioni delle materie prime e non devono essere presenti.



# Report analisi metagenomiche su Gardasil 9

## Introduzione

Come è noto, i vaccini sono farmaci biologici utilizzati per la prevenzione di alcune malattie infettive e sono costituiti da più componenti: gli antigeni (virus, batteri inattivati o attenuati, tossine inattivate, proteine o molecole complesse derivati dai virus e batteri, in grado di stimolare la risposta immunitaria), gli adiuvanti (sostanze che aumentano la capacità degli antigeni vaccinali di indurre la risposta immunitaria anticorpale), gli eccipienti (sostanze necessarie per formulare il vaccino, o per preservarlo dalle contaminazioni batteriche) e le contaminazioni (sostanze presenti in tracce provenienti dalle materie prime, es. linee cellulari per la crescita dei batteri e virus, o dal processo di lavorazione, es. formaldeide, antibiotici). Durante la fase di registrazione di un farmaco biologico, il vaccino viene sottoposto ai controlli previsti dalle linee guida dell'EMA e concordati con l'ente regolatore in base al tipo specifico di vaccino. Tali controlli vengono poi effettuati su un numero rappresentativo di campioni su ciascun lotto prima della commercializzazione.

*La responsabilità della conformità del prodotto venduto è quindi del produttore e degli enti regolatori preposti al controllo.*

Poiché la sicurezza di un vaccino dipende dalla sua conformità ai criteri di qualità, soprattutto riguardanti il controllo dell'assenza di contaminazioni tossiche o potenzialmente tossiche (cioè per le quali non sono noti gli effetti sull'uomo) è di grande importanza che tale conformità venga rispettata in modo molto rigoroso.

Vari studi in letteratura hanno posto il problema della presenza di vari tipi di contaminazioni, sia di tipo chimico che microbiologico, aprendo quindi il quesito se effettivamente i vaccini sono conformi alle direttive imposte dagli enti regolatori, se a loro volta gli enti regolatori applicano il controllo per il rispetto di tali direttive e se gli enti regolatori hanno definito con linee guida efficaci i criteri per il controllo e il contenimento di tali contaminazioni. Per rispondere a tali domande il Corvelva ha commissionato l'analisi delle contaminazioni biologiche, che non dovrebbero mai essere presenti nei vaccini, ad un centro altamente qualificato di servizi specializzato nel sequenziamento genomico di DNA e di RNA.

Lo studio commissionato dal Corvelva si è articolato su due tipi di analisi:

1. **Test di presenza di acidi nucleici (DNA/RNA)** di origine umana e animale e da microrganismi (virus, batteri) utilizzando la metodica Next Generation Sequencing, che ha permesso di quantificare in maniera altamente specifica e accurata la sequenza del materiale genetico contenuto nei vaccini esaminati
2. **Verifica di corrispondenza delle sequenze genomiche** dei batteri e virus vivi attenuati o inattivati presenti nei vaccini (presenza di varianti genetiche)

## Descrizione del metodo utilizzato per l'analisi

Il Next Generation Sequencing, noto anche come **deep sequencing**, genera una singola sequenza da ciascun frammento di DNA, o cDNA, presente in un campione. L'analisi bioinformatica a valle consente poi la differenziazione tra l'origine dei frammenti di sequenza, ad esempio umana, specie batteriche o un particolare virus. Questo significa che campioni biologici misti possono essere agevolmente risolti con questa tecnologia, ormai entrata nella routine della ricerca genomica e della diagnostica. Inoltre da dati NGS è possibile ricostruire l'intera sequenza di genomi virali a DNA e RNA e di genomi batterici presenti nel campione e confrontarlo con i genomi di riferimento presenti nei database pubblici.

I campioni esaminati sono riportati di seguito insieme ai risultati ottenuti, raggruppandoli per classi di vaccini simili:

\* ssRNA: single strand RNA, RNA a filamento singolo; dsDNA: double strand DNA, DNA a doppio filamento.

I termini sottolineati sono costituiti o contengono materiale genetico (DNA e/o RNA)



## Lotti Analizzati

### Lotto #1 - 9R009338

<b>Nome prodotto:</b>	<b>Gardasil 9</b>
<b>Tipo di prodotto:</b>	Vaccino Papilloma virus umano, HPV (ceppi 6, 11, 16, 18, 31, 33, 45, 52 e 58)
<b>Produttore:</b>	MSD Vaccins - fabbricato Merck Sharp and Dohme
<b>Composizione:</b>	Sostanze attive: 9 proteine L1 non infettive, altamente purificate di Papilloma virus umano (ceppi 6, 11, 16, 18, 31, 33, 45, 52 e 58) in quantità variabili tra 20 e 60 microgrammi. Le proteine L1 sono nella forma di pseudo particelle virali prodotte su cellule di <i>Saccaromyces cerevisiae</i> CANADE 3C-5 mediante tecnologia del DNA ricombinante. Adsorbite su adiuvante alluminio (Al: 0.5 milligrammi); cloruro di sodio, L-istidina, polisorbato 80, borato di sodio.

## Analisi richieste

Test di presenza di acidi nucleici (DNA/RNA) di origine umana e animale e da microrganismi (virus, batteri), utilizzando un approccio di tipo metagenomico/metatrascrittomico su piattaforma Illumina di tipo Next Generation Sequencing.

### Dal confronto di questi tre vaccini è possibile evidenziare le seguenti criticità:

L'estrazione del DNA genomico è stata effettuata con il kit Maxwell® 16 Blood DNA Purification Kit rivenduto dall'azienda Promega e con l'estrattore automatico Maxwell® 16 IVD (Promega), seguendo il protocollo del produttore. L'estrazione dell'RNA è stata fatta utilizzando il kit PureLink™ Viral RNA/DNA Mini Kit (Invitrogen) seguendo il protocollo del produttore.

Le quantità di partenza utilizzate per le estrazioni sono le seguenti:

- Estrazione DNA: 290 µl di soluzione iniettabile
- Estrazione RNA: 290 µl di soluzione iniettabile

La quantificazione e il controllo di qualità del DNA estratto è stata effettuata rispettivamente con il fluorimetro Qubit 2.0 (Invitrogen, Carlsbad, CA) e con lo spettrofotometro NanoDrop 1000 (Thermo Fisher Scientific, Waltham, USA).

Di seguito il risultato delle quantificazioni del DNA (ND= NanoDrop 1000; QB= Qubit 2.0; HS= dsDNA HS Assay Kit)

Sample ID	ND A260/280	ND A260/230	QB_HS_ng/µL	volume_µl	Tot_amount_ng
DNA lot 9R009338	1,06	0,41	NA	55	NA

La misurazione della concentrazione con fluorimetro QuBit ha evidenziato che il DNA contenuto in Gardasil 9 lotto 9R009338, **non è quantificabile con metodi fluorimetrici standard**.

La quantificazione e il controllo di qualità dell'RNA estratto è stata eseguita su Agilent 2100 Bioanalyzer utilizzando l'Agilent RNA 6000 Pico Kit (Agilent Technologies, Santa Clara, CA). Di seguito i valori di concentrazione e il RIN (RNA Integrity Number) misurati al Bionalyzer:

Sample ID	Bioanalyzer_pico_totale_ng/µL	RIN	volume_ul	Tot_amount_ng
RNA lot 9R009338	0,236	1,4	37	8,732

La quantità di **RNA** contenuta nella fiala di Gardasil 9 lotto 9R009338 è **risultata essere di 8.732 ng** così calcolata: 0,236 ng/µl (concentrazione determinata al Bionalyzer) x 37 (volume di risospensione finale dell'RNA dopo estrazione, espresso in microlitri) x 2 (volume totale della soluzione iniettabile/volume utilizzato per l'estrazione). Il RIN pari a 1,4 indica un RNA degradato. Il tracciato appare piatto considerando la bassissima quantità di RNA contenuta nel campione.

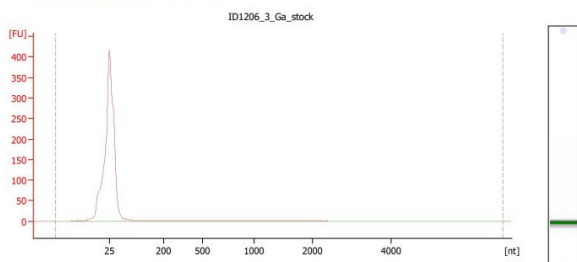


2100 expert\_Eukaryote Total RNA Pico\_DE72901783\_2018-11-22\_12-38-01.xad

Page 3 of 4

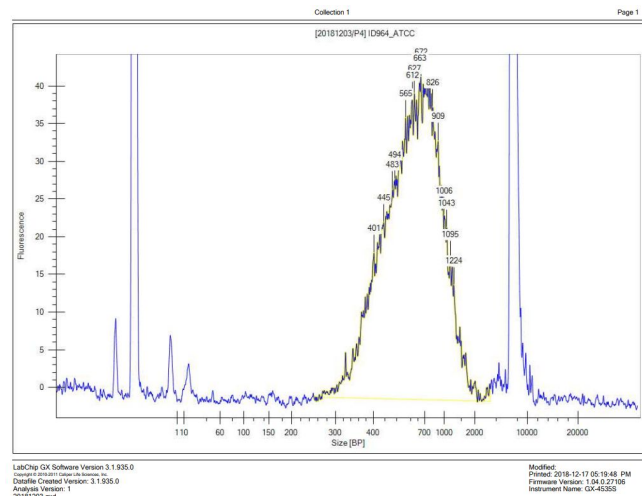
Assay Class: Eukaryote Total RNA Pico  
Data Path: I:\\_Eukaryote Total RNA Pico\_DE72901783\_2018-11-22\_12-38-01.xad  
Created: 11/22/2018 12:38:01 PM  
Modified: 11/22/2018 12:56:44 PM

Electropherogram Summary Continued ...



Overall Results for sample 7 : ID1206\_3\_Ga\_stock

RNA Area:	75.1	RNA Integrity Number (RIN):	1.4 (B.02.08)
RNA Concentration:	236 pg/ul	Result Flagging Color:	
rRNA Ratio (28s / 18s):	0.0	Result Flagging Label:	



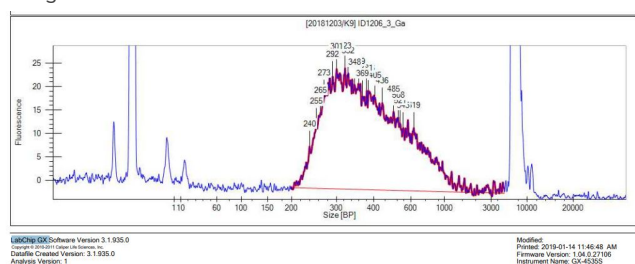
### Preparazione della libreria DNA-seq con tecnologia Illumina

Per la preparazione delle librerie è stato usato il kit Ovation® Ultralow System V4 1–96 (Nugen, San Carlos, CA) seguendo le indicazioni del produttore, a partire da una quantità minore di 1ng di DNA genomico. La libreria finale è stata quantificata con il fluorimetro Qubit 2.0 (Invitrogen, Carlsbad, CA) e testata per qualità con il sistema Agilent 2100 Bioanalyzer, DNA High Sensitivity Analysis kit (Agilent technologies, Santa Clara, CA). Di seguito il tracciato della libreria ottenuta:

### Preparazione della libreria RNA-seq con tecnologia Illumina

Per la preparazione delle librerie RNA-seq è stato usato il kit Ovation® RNA-Seq System V2 (Nugen, San Carlos, CA) per preparare il cDNA e il kit Ovation® Ultralow System V4 1–96 per preparare la libreria a partire da 10ng di cDNA. La libreria finale è stata quantificata con il fluorimetro Qubit 2.0 (Invitrogen, Carlsbad, CA) e testata per qualità con il sistema Caliper GX (PerkinElmer, Waltham, MA).

Di seguito il tracciato della libreria ottenuta:



### Preparazione di una libreria DNA-seq con tecnologia Illumina da un mix di DNA genomici di una comunità metagenomica a composizione nota.

Al fine di validare il workflow di lavoro dalla preparazione della libreria fino all'analisi dei dati, è stato utilizzato uno standard ATCC (mix di DNA genomici a composizione nota, 20 Strain Staggered Mix Genomic Material, ATCC® MSA-1003™) per costruire una libreria con il kit Ovation® Ultralow System V4 1–96 a partire da 10ng di DNA. Di seguito il tracciato su Caliper GX (PerkinElmer, Waltham, MA) della libreria ottenuta:

### Sequenziamento

Le librerie sono state sequenziate su strumento Illumina HiSeq2500 in modalità 'paired-end 125bp', secondo le indicazioni standard Illumina. E' stata utilizzata la versione 1.8.2 della pipeline Illumina CASAVA per processare le sequenze grezze.

Sono state prodotte 9.194.982 sequenze paired-end Illumina pari a 4.597.491 frammenti/reads identificative per la libreria DNA-seq; 12.495.858 sequenze paired-end Illumina pari a 6.247.929 frammenti/reads identificative per la libreria RNA-seq; 9.938.490 sequenze paired-end Illumina pari a 4.969.245 di frammenti/reads identificative per la libreria DNA-seq costruita a partire dal DNA dello standard Mix Genomic Material, ATCC® MSA-1003.

### Analisi bioinformatica

#### Pulizia delle sequenze (trimming)

Le sequenze degli adattatori (cioè sequenze oligonucleotidiche 'artificiali' che vengono introdotte durante la preparazione della libreria Illumina) e delle basi di DNA lette a bassa qualità sono state rimosse usando i software ERNE<sup>1</sup> e Cutadapt<sup>2</sup>.

#### Identificazione degli organismi di origine delle sequenze di DNA e cDNA/RNA con il software Kraken.

L'analisi metagenomica è stata eseguita utilizzando il software Kraken<sup>3</sup> sul database 'Human-Virus-Bacteria\_25mer' (<https://ccb.jhu.edu/software/kraken/>). Kraken è un classificatore che assegna etichette tassonomiche a brevi letture del DNA. Lo fa esaminando i k-mers all'interno di una lettura e interrogando un database con quei k-mers.

La presenza di DNA e RNA è espressa in termini di numero di reads/frammenti e di percentuale di reads/frammenti sul totale delle reads/frammenti prodotti, attribuite dai database pubblici ai diversi organismi.

Le sequenze classificate con software Kraken sono state inoltre confermate manualmente con il software BLAST (Basic Local Alignment Search Tool, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>).



### Gardasil 9 lot. 9R009338 – analisi DNA

DNA Seq total fragments/reads	4,597,491		
<b>Classification</b>		<b>n° reads</b>	<b>% reads</b>
Bacteria		2405197	54%
Homininae (Homo sapiens)		411670	9%
Murinae (Mus musculus)		400972	9%
Viruses*		1253	0,03%
Unassigned		1378399	30%
<b>*Viruses classification</b>		<b>n° reads</b>	<b>% reads</b>
dsDNA viruses		1215	0,03%
	Human papillomavirus virus-L1 fragment	752	0,02%
	Myoviridae (phages)	350	0,008%
	Molluscum contagiosum virus	113	0,002%
Retroviridae		38	0,0008%
	Murine leukemia virus	37	0,0008%
	Human endogenous retrovirus K	1	0,00001%



**Gardasil 9 lot. 9R009338 – analisi RNA**

<b>RNA Seq total fragments/reads</b>	<b>6,247,929</b>		
<b>Classification</b>		<b>n° reads</b>	<b>% reads</b>
Bacteria		736256	15%
Artificial sequences (synthetic construct)		540799	10%
Saccharomyces		2216355	35%
Viruses*		61254	1%
Unassigned		2693265	43%
<b>*Viruses classification</b>		<b>n° reads</b>	<b>% reads</b>
<b>dsDNA viruses</b>		<b>53184</b>	<b>1%</b>
	Human papillomavirus virus-L1 fragment	53184	0,9%
<b>dsRNA viruses</b>		<b>7595</b>	<b>0,1%</b>
	Saccharomyces cerevisiae virus L-BC (La)	7582	0,1%
<b>ssRNA viruses</b>		<b>469</b>	<b>0,007%</b>
	Saccharomyces 20S RNA narnavirus	469	0,007%
<b>Ortervirales</b>		<b>6</b>	<b>0,0001%</b>
	Equine infectious anemia virus	3	0,00005%
	Murine leukemia virus	3	0,00005%



**Risultati dell'analisi DNA-seq eseguite con il software Kraken su uno standard genomico a composizione nota (20 Strain Staggered Mix Genomic Material, ATCC® MSA-1003™)**

DNA Seq total fragments/reads	4,969,245		
Classification	n° reads	% reads	% declared by ATCC
Acinetobacter baumannii	10735	0,2%	0,18%
Actinomyces odontolyticus	2	0,00004%	0,18%
Bacillus cereus	176327	3,5%	1,8%
Bacteroides vulgatus	1088	0,02%	0,02%
Bifidobacterium adolescentis	489	0,01%	0,02%
Clostridium beijerinckii	123609	2,5%	1,8%
Cutibacterium acnes	6528	0,13%	0,18%
Deinococcus radiodurans	745	0,02%	0,02%
Enterococcus faecalis	704	0,01%	0,02%
Escherichia coli	929837	19%	18%
Helicobacter pylori	4738	0,1%	0,18%
Lactobacillus gasseri	4491	0,1%	0,18%
Neisseria meningitidis	9820	0,19%	0,18%
Porphyromonas gingivalis	578294	12%	18%
Pseudomonas aeruginosa	152307	3%	1,8%
Rhodobacter sphaeroides	1135927	23%	18%
Staphylococcus aureus	72598	1,5%	1,8%
Staphylococcus epidermidis	634940	13%	18%
Streptococcus agalactiae	31622	0,6%	1,8%
Streptococcus mutans	526420	11%	18%
Unassigned	14248	0,3%	0%





## Riferimenti bibliografici

1. Del Fabbro, C et al. 2013 An extensive evaluation of read trimming effects on Illumina NGS data analysis. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM. PLoS One. 2013 Dec 23;8(12):e85024. doi: 10.1371/journal.pone.0085024. eCollection 2013
2. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, [S.l.], 17 (1): 10-12 (2011). ISSN 2226-6089. Date accessed: 02 Apr. 2015. doi:<http://dx.doi.org/10.14806/ej.17.1.200> paper
3. Wood and Salzberg. Kraken: ultrafastmetagenomicsequence classification using exact alignments Genome Biology 2014, 15:R46

