

Vaccinegate:

Approfondimenti sull'analisi della presenza di
sequenze del virus della rosolia e di virus
avventizi nei dati RNA-seq ottenuti per Priorix
Tetra, lotto A71CB256A



Approfondimenti sull'analisi della presenza di sequenze del virus della rosolia e di virus avventizi nei dati RNA-seq ottenuti per Priorix Tetra, lotto A71CB256A

Ricerca di Rubella virus

La presenza del virus della rosolia è stata dimostrata sequenziando una libreria RNA-seq ad altissima profondità (circa 260 milioni di sequenze Illumina prodotte). Sono state rilevate 114 sequenze su 260 milioni, pari allo 0.00004% del totale delle sequenze. Le sequenze del genoma della rosolia sono state quindi confermate manualmente con il software BLAST (Basic Local Alignment Search Tool, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>). La libreria sequenziata a più bassa profondità (circa 12 milioni di sequenze accoppiate Illumina, pari a 6 milioni di frammenti di libreria sequenziati) non aveva rilevato la presenza di NESSUNA reads attribuibile alla rosolia, in questo lotto. Anche le sequenze degli altri virus presenti nel vaccino (varicella, morbillo e parotite) sono state validate con la stessa modalità manuale, confermando di essere state correttamente assegnate.

Conferma della presenza di virus estranei

Analogamente a quanto fatto per rosolia, varicella, morbillo e parotite, le sequenze provenienti dal sequenziamento standard effettuato a profondità di circa 12 milioni di sequenze accoppiate Illumina (pari a 6 milioni di frammenti di libreria sequenziati), che il software Kraken aveva classificato come appartenenti ad altri virus, sono state validate manualmente con software BLAST.

I virus avventizi di tipo ssRNA evidenziati dall'analisi Kraken NON sono stati confermati dall'analisi manuale BLAST. Le sequenze si sono dimostrate essere a bassa complessità e l'algoritmo utilizzato da Kraken (diverso da quello utilizzato dal software BLAST) le ha classificate erroneamente. Questo problema si riscontra SOLO su numeri molto bassi di sequenze quando non vengono costruiti dei filtri di alcun genere nella pipeline di analisi (nel nostro caso i filtri non erano stati volutamente inseriti nella prima versione delle analisi fatte).

La presenza di virus avventizi che Kraken aveva classificato come Retroviridae, è stata invece IN PARTE confermata dall'analisi BLAST. Si fa notare che la ricerca dei Retroviridae è stata effettuata su dati di RNA-seq, quindi il loro riscontro, potrebbe indicare che queste sequenze di retrovirus sono trascritte e non necessariamente inserite nel genoma umano presente nel campione come provirus/retrovirus endogeni integrati.

La quantità di sequenze attribuibili a Retroviridae, confermata con software BLAST, è risultata essere pari a 40 su 12 milioni totali di sequenze, ovvero lo 0.002% delle sequenze totali. In particolare si conferma, tramite validazione manuale con BLAST, la presenza di sequenze appartenenti ai seguenti retrovirus:

Human endogenous retrovirus K	32 sequenze
Equine infectious anemia virus	2 sequenze
Avian leukosis virus	2 sequenze
HERV-H/env62	4 sequenze

Per quanto riguarda la conferma le tracce di altri genomi trovate nei dati RNA-seq quali 'unclassified bacterial and environmental viruses' che include sequenze fagiche, 'dsRNA virus' e 'unclassified RNA viruses ShiM-2016' che include genomi virali poco conosciuti, è necessaria un'ulteriore analisi manuale di approfondimento per dimostrare la corretta assegnazione e l'effettiva presenza. La presenza del virus dell'epatite B NON è stata confermata dall'analisi BLAST.

Di seguito una tabella che riassume di quanto descritto nel testo:

RNA analysis Batch #2 A71CB256A

		Profondità 12 M		Profondità 260M		
		Software Kraken				
Viruses classification		n° reads	% reads	n° reads	% reads	Confermato*
dsDNA viruses		836208	7%			
	Human alphaherpesvirus 3	835736	7%			sì
ssRNA viruses		1450	0,01200%			
	Mumps rubulavirus	1016	0,008%			sì
	Measles morbillivirus	434	0,004%			sì
	Rubella virus	0	nd	114	0,00004%	sì
	Influenza A virus	18	0,0001%			no
	Cupixi mammarenavirus	60	0,0005%			no
	Pneumoviridae	12	0,0001%			no
	Jamestown Canyon virus	24	0,0002%			no
	Hepacivirus C	60	0,0005%			no
	Kobuvirus	40	0,0003%			no
	Enterovirus	4	0,00003%			no
	Porcine reproductive and respiratory syndrome virus	6	0,00005%			no
	Coronavirinae	10	0,00008%			no
	Potyvirus	6	0,00005%			no
Retroviridae		40	0,002%			
	Human immunodeficiency virus 1	72	0,0006%			no
	Human endogenous retrovirus K	32	0,0003%			sì
	Simian immunodeficiency virus	2	0,00002%			no
	Equine infectious anemia virus	2	0,00002%			sì
	Lymphoproliferative disease virus	50	0,0004%			no
	Avian leukosis virus	2	0,00002%			sì
	Rous sarcoma virus	2	0,00002%			no
	HERV-H/env62	4	0,00003%			sì
	Red clover bacilliform virus	4	0,00003%			no
unclassified bacterial and environmental viruses (phages)		176	0,001%			?



dsRNA viruses		28	0,0002%			?
Hepatitis B virus		4	0,00003%			no
unclassified RNA viruses ShiM-2016		18	0,0002%			?
Mollivirus sibericum		2	0,0002%			no

* Confermato manualmente con software BLAST

