

# Flusso di lavoro di arricchimento per rilevare il coronavirus utilizzando i sistemi NGS Illumina

Rilevamento e caratterizzazione  
altamente sensibili di virus respiratori  
comuni, inclusi i ceppi di coronavirus



## Sorveglianza virale con NGS

Le infezioni virali sono un'importante questione di salute globale, con nuove varianti di malattie infettive che emergono continuamente. L'epidemia del nuovo coronavirus nel 2019 (SARS-CoV-2) iniziata a Wuhan, in Cina, e rapidamente diffusasi in diversi Paesi è un esempio importante dei potenziali rischi delle pandemie virali globali. I coronavirus (CoV) sono un'ampia famiglia di virus che possono infettare gli esseri umani, causando malattie respiratorie che vanno dal comune raffreddore a patologie più gravi, come la sindrome respiratoria mediorientale (MERS-CoV, Middle East Respiratory Syndrome Coronavirus) e la sindrome respiratoria acuta grave (SARS-CoV, Severe Acute Respiratory Syndrome Coronaviruses). L'Organizzazione Mondiale della Sanità (OMS) ha classificato la malattia associata a SARS-CoV-2 (COVID-19) come un'emergenza sanitaria pubblica di interesse internazionale, evidenziando la necessità di un rilevamento virale rapido e accurato.<sup>1</sup>

Il sequenziamento di nuova generazione (NGS, Next-Generation Sequencing) fornisce un modo efficace e innovativo per sottoporre a screening i campioni e rilevare i virus senza una precedente conoscenza dell'agente infettivo.<sup>2</sup> L'arricchimento mirato è un metodo di sequenziamento che cattura le regioni genomiche di interesse mediante ibridazione degli acidi nucleici a sonde biotinate specifiche per il target. L'arricchimento mirato mediante ibridazione-cattura consente un rilevamento altamente sensibile, senza richiedere l'elevata profondità di lettura necessaria per il sequenziamento con metagenomica shotgun. Inoltre, il flusso di lavoro NGS per l'arricchimento mirato consente di ottenere dati di sequenza quasi completi dei target e apre applicazioni come l'analisi delle varianti per l'evoluzione virale o la sorveglianza virale.<sup>3</sup> Rispetto ad altri metodi di sequenziamento mirati, come il sequenziamento degli ampliconi, l'arricchimento mediante ibridazione-cattura consente di ottenere pannelli sonda notevolmente più grandi con una profilazione più completa delle regioni

target. Inoltre, le sonde oligonucleotidiche utilizzate per i protocolli di ibridazione-cattura sono efficaci anche nelle regioni altamente mutagene e consentono di mirare a virus dalla rapida evoluzione, come i virus a RNA.

## Flusso di lavoro di arricchimento virale

Questa nota sull'applicazione evidenzia un flusso di lavoro semplificato per il rilevamento e l'analisi del coronavirus mediante il kit di preparazione della libreria Illumina DNA Prep with Enrichment, i pannelli di targeting virale, il sequenziamento Illumina comprovato e l'analisi semplificata dei dati (Figura 1). Questo flusso di lavoro ha lo scopo di arricchire i target virali di DNA e RNA dall'estrazione totale dell'acido nucleico. Il protocollo inizia con l'estrazione di campioni di acido nucleico e la successiva trascrizione inversa dell'RNA estratto in un DNA complementare (cDNA, complementary DNA) a doppio filamento. La preparazione della libreria viene eseguita utilizzando Illumina DNA Prep with Enrichment con IDT for Illumina DNA/RNA UD Indexes. In questo protocollo, DNA e cDNA vengono sottoposti a tagmentazione, pulizia e amplificazione pre-arricchimento. Dopo l'amplificazione, è possibile raggruppare in pool fino a 12 campioni per una reazione di arricchimento utilizzando un pannello di oligonucleotidi che mirano alle sequenze virali. L'ibridazione della sonda è seguita da cattura della sonda, amplificazione dell'arricchimento, quantificazione e sequenziamento.

## Metodi

### Preparazione dei campioni

Per dimostrare le prestazioni del pannello di arricchimento virale mirato, è stato analizzato un ceppo di campione di coronavirus disattivato, ceppo CoV OC43 di Microbiologics

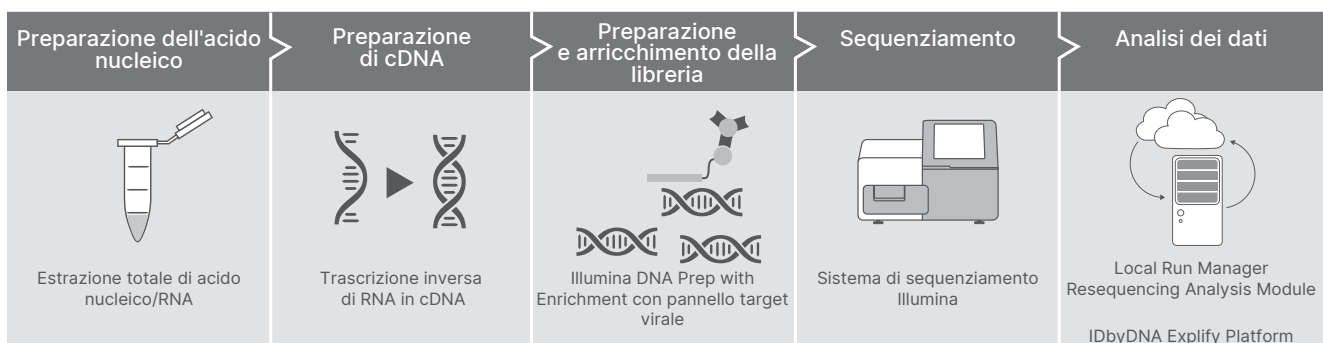


Figura 1: flusso di lavoro di arricchimento per il rilevamento del coronavirus. Il flusso di lavoro NGS ottimizzato per il rilevamento del coronavirus integra la preparazione dei campioni, la preparazione delle librerie, l'arricchimento mirato, il sequenziamento e l'analisi dei dati.

(set e pannelli QC: Helix Elite, n. di catalogo 8217) (Tabella 1). Il campione di coltura virale per il coronavirus è stato estratto utilizzando il QIAGEN QIAmp Viral Mini Kit (QIAGEN, n. di catalogo 52904) in un ambiente di laboratorio BSL2. 150 ng di RNA estratto sono stati retrotrascritti in cDNA utilizzando due diversi flussi di lavoro: uno derivato dai reagenti Illumina TruSeq RNA e l'altro utilizzando il Thermo Fisher Scientific Maxima H Minus Double-Stranded cDNA Synthesis Kit (Thermo Fisher Scientific, n. di catalogo K2561). Per simulare i risultati del sequenziamento ottenuti da campioni di pazienti, il campione virale è stato anche addizionato all'RNA di fondo del riferimento umano universale (UHR, Universal Human Reference) al 95% (Agilent Technologies, n. di catalogo 74000) e retrotrascritto in cDNA utilizzando i metodi descritti sopra (Tabella 1).

Inoltre, è stato estratto un pool virale respiratorio contenente quattro virus a RNA e un virus a DNA (Tabella 2) utilizzando il QIAGEN AllPrep PowerViral DNA/RNA Kit (QIAGEN, n. di catalogo 28000-50) in un ambiente di laboratorio BSL2. Il volume massimo di acido nucleico totale è stato retrotrascritto in cDNA (senza trattamento con DNasi) utilizzando un flusso di lavoro derivato dai reagenti Illumina TruSeq RNA. Per simulare i risultati del sequenziamento ottenuti da campioni di pazienti con più patogeni presenti, il campione del pool virale è stato anche addizionato all'RNA di fondo del riferimento umano universale (UHR) al 95% (Agilent Technologies, n. di catalogo 74000) e retrotrascritto in cDNA (Tabella 1).

Tabella 1: composizione di campioni virali per l'analisi

Campione	Composizione <sup>a</sup>	Genoma di riferimento
CoVOC43	150 ng di RNA di CoV OC43	AY391777.1 Coronavirus umano OC43
CoVOC43_95UHR	7,5 ng di RNA di CoV OC43 e 142,5 ng di RNA di UHR	AY391777.1 Coronavirus umano OC43
ViralPool	Volume massimo di acido nucleico totale del pool virale (8,5 µl)	Vedere Tabella 2
ViralPool_95UHR	Acido nucleico totale del pool virale al 5% e RNA di UHR al 95% per volume	Vedere Tabella 2

a. L'input minimo raccomandato di RNA/acido nucleico totale per la trascrizione inversa è 10 ng. Per ottenere i migliori risultati, la trascrizione inversa deve essere eseguita su campioni di acido nucleico appena estratti.

Tabella 2: composizione del pool virale

Virus	Tipo di acido nucleico	Genoma di riferimento
Virus dell'influenza A (H1N1)	RNA	Influenza A/ Michigan/45/2015
Virus dell'influenza B	RNA	Influenza B/ Colorado/06/2017
Virus parainfluenzale umano 3	RNA	NC_001796.2
RSV B9320	RNA	AY353550.1
Adenovirus 7	DNA	AY594255.1

## Preparazione delle librerie

Le librerie pronte per il sequenziamento sono state preparate con Illumina DNA Prep with Enrichment (Illumina, n. di catalogo 20025524) e IDT for Illumina DNA UD Indexes (Illumina, n. di catalogo 20027213)\* utilizzando cDNA dal campione CoV (CoVOC43), dal campione del pool virale (ViralPool) e da entrambi i campioni addizionati (CoVOC43\_95UHR e ViralPool\_95UHR). L'input totale di DNA raccomandato per la tagmentazione è 10-1.000 ng.

Dopo l'amplificazione, i campioni sono stati suddivisi in pool separati per l'arricchimento in base al tipo di campione (CoVOC43, CoVOC43\_95UHR, ViralPool e ViralPool\_95HR). Le reazioni di arricchimento sono state eseguite con il Respiratory Virus Oligos Panel (Illumina, n. di catalogo 20042472) che dispone di circa 7.800 sonde progettate per rilevare virus respiratori, ceppi influenzali recenti e SARS-CoV-2, nonché sonde umane per fungere da controlli positivi (Tabella 3). Dopo l'arricchimento (Figura 2), le librerie preparate sono state quantificate, raggruppate in pool e caricate sul MiSeq™ System per il sequenziamento.

\* Il n. di catalogo 20027213 è attualmente obsoleto, sostituito da Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 indexes, 96 samples), n. di catalogo 20091654.

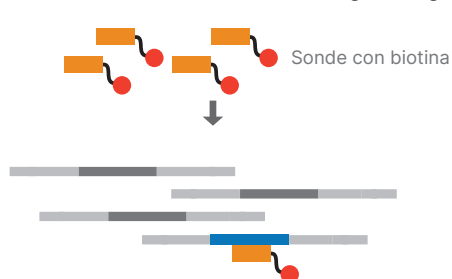
Tabella 3: virus target del Respiratory Virus Oligos Panel

Coronavirus umano 229E	Virus parainfluenzale umano 1
Coronavirus umano NL63	Virus parainfluenzale umano 2
Coronavirus umano OC43	Virus parainfluenzale umano 3
Coronavirus umano HKU1	Virus parainfluenzale umano 4a
SARS-CoV-2	Metapneumovirus umano (CAN97-83)
Adenovirus umano B1	Virus respiratorio sinciziale (tipo A)
Adenovirus umano C2	Virus respiratorio sinciziale umano 9320 (tipo B)
Adenovirus umano E4	Virus dell'influenza A (A/Porto Rico/8/1934(H1N1))
Bocavirus umano 1 (isolato st2 di Primate bocaparvovirus 1)	Virus dell'influenza A (A/Corea/426/1968(H2N2))
Isolato PK-5510 di bocavirus umano 2c PK	Virus dell'influenza A (A/New York/392/2004(H3N2))
Bocavirus umano 3	Virus dell'influenza A (A/oca/Guangdong/1/1996(H5N1))
Ceppo NI di bocavirus umano 4 HBoV4-NI-385	Virus dell'influenza A (A/Zhejiang/DTID-ZJU01/2013(H7N9))
Poliomavirus KI Stoccolma 60	Virus dell'influenza A (A/Hong Kong/1073/99(H9N2))
Poliomavirus WU	Virus dell'influenza A (A/Texas/50/2012(H3N2))
Parechovirus umano di tipo 1	Virus dell'influenza A PicoBank/HPeV1/a (A/Michigan/45/2015(H1N1))
Parechovirus umano 6	Virus dell'influenza B (B/Lee/1940)
Rinovirus umano A89	Virus dell'influenza B (B/Wisconsin/01/2010)
Rinovirus umano C (ceppo 024)	Virus dell'influenza B (B/Brisbane/60/2008)
Rinovirus umano B14	Virus dell'influenza B (B/Colorado/06/2017)
Ceppo C104 umano di enterovirus: AK11	Virus dell'influenza B (B/Washington/02/2019)
Isolato C109 umano di enterovirus NICA08-4327	Geni di controllo umano

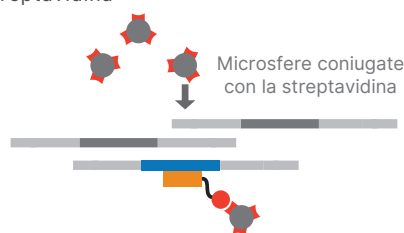
A. Raggruppamento in pool delle librerie Illumina DNA Prep with Enrichment



B. Ibridazione delle sonde biotilate sulle regioni target



C. Cattura mediante microsfere coniugate con la streptavidina



D. Eluizione dalle microsfere

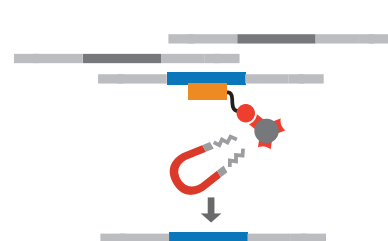


Figura 2: chimica di arricchimento. Le librerie preparate vengono sottoposte ad arricchimento con un pannello di target virali, seguito da amplificazione, quantificazione e sequenziamento.

### Sequenziamento

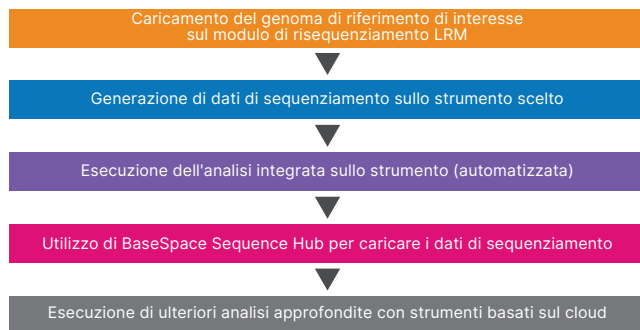
Le librerie preparate possono essere sequenziate su qualsiasi strumento Illumina. I sistemi da banco iSeq™ 100 System, MiniSeq™ System e MiSeq System sono adatti grazie ai requisiti di lettura bassi per questi campioni. In questo studio, le librerie preparate dai campioni virali sono state denaturate e diluite a una concentrazione di caricamento finale di 10 pM, secondo la Guida alla diluizione e alla denaturazione delle librerie con MiSeq System (documento n. 15039740 v10) e sequenziate sul MiSeq System a una lunghezza di lettura di 2 × 151 bp utilizzando i reagenti MiSeq v3 (Illumina, n. di catalogo MS-102-3001). Anche le aliquote di tutte le librerie sono state sequenziate

prima dell'ibridazione della sonda per determinare la variazione della reazione di arricchimento.

Il titolo del virus, la qualità dell'RNA e il numero di letture per campione influiscono sul numero di letture specifiche del virus ottenute e sulla copertura del genoma virale. Come linea guida generale, la raccomandazione di lettura per questo flusso di lavoro è di 500.000 letture per campione, ma questi numeri possono variare e si tratta solo di un punto di partenza raccomandato.

### Analisi dei dati

Per l'analisi automatica integrata sullo strumento, le corse di sequenziamento sono state impostate in Local Run Manager utilizzando il Resequencing Analysis Module. Questo modulo consente l'input di tutte le informazioni sulla corsa e la selezione del genoma di riferimento per il successivo allineamento della sequenza. Gli utenti possono caricare i genomi di riferimento direttamente sullo strumento, consentendo una facile personalizzazione. L'analisi viene avviata automaticamente al termine del sequenziamento, in modo che gli utenti possano interpretare i risultati il più rapidamente possibile. Il modulo di risequenziamento fornisce i dati di allineamento, copertura e varianti piccole, nonché i file FASTQ, BAM e VCF da utilizzare in altre pipeline di analisi dei dati, se lo si desidera (Figura 3).



**Figura 3:** flusso di lavoro dell'analisi dell'arricchimento. Le corse di sequenziamento sono impostate in Local Run Manager e l'analisi dei dati procede automaticamente al termine della corsa. Se lo si desidera, è disponibile un'analisi aggiuntiva con applicazioni di terze parti.

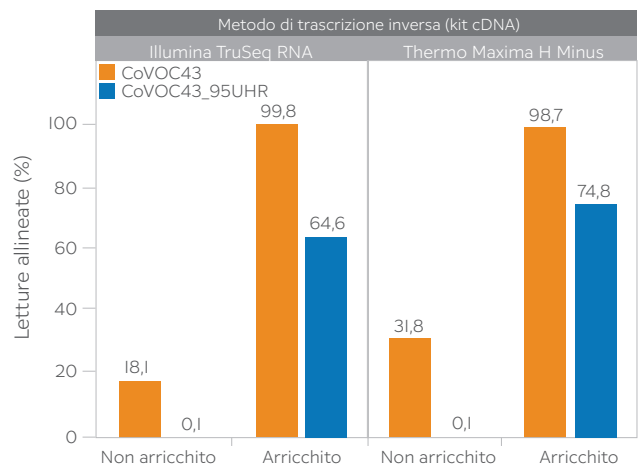
IDbyDNA Explify Platform è una soluzione di facile utilizzo per l'analisi approfondita dei dati che offre un solido controllo qualità (QC, Quality Control) dei dati, un'interpretazione standardizzata dei risultati, database accuratamente selezionati e generazione di report personalizzati. L'analisi dei dati si basa sui k-mer e sulle fasi di allineamento, incluso il rilevamento dei virus a livello proteico, che aumenta la capacità di identificare virus nuovi e altamente divergenti. È possibile accedere alla piattaforma in BaseSpace™ Sequence Hub.

## Risultati

Dopo la preparazione delle librerie e il sequenziamento, è stato utilizzato il Local Run Manager Resequencing Analysis Module (v2.5.56.11) per allineare ciascun campione al genoma di riferimento OC43 per il coronavirus, come descritto nella [Guida al flusso di lavoro del Local Run Manager Resequencing Analysis Module](#).

### Trascrizione inversa e rilevamento di CoV con Local Run Manager Resequencing Module

In questo studio, sono stati utilizzati due diversi kit di trascrizione inversa per generare cDNA per la preparazione delle librerie: un kit di Thermo Fisher Scientific disponibile in commercio e un flusso di lavoro derivato dai reagenti Illumina TruSeq RNA (vedere Metodi per i dettagli). Per i campioni CoVOC43, i risultati hanno mostrato che l'allineamento percentuale al genoma di riferimento era maggiore del 99% per il metodo di trascrizione inversa Illumina e maggiore del 97% per il metodo Thermo Fisher Scientific (Figura 4). L'allineamento percentuale dei campioni addizionati era maggiore del 64% e maggiore del 19% per il Respiratory Virus Oligos Panel, indipendentemente dal metodo di trascrizione inversa (Figura 4).



**Figura 4:** prestazioni equivalenti con diversi kit di cDNA. La percentuale di letture allineate per i campioni virali retrotrascritti con i due diversi kit di cDNA (vedere la sezione Metodi) viene tracciata prima (non arricchita) e dopo (arricchita) l'arricchimento con il Respiratory Virus Oligos Panel.

Come misura del successo della reazione di cattura e ibridazione del target, le letture allineate percentuali dei campioni prima e dopo l'arricchimento sono state confrontate per calcolare la variazione di arricchimento. I risultati hanno dimostrato aumenti significativi nel valore di arricchimento per i campioni CoV analizzati indipendentemente dal metodo di trascrizione inversa. Questo è stato particolarmente evidente per i campioni CoVOC43\_95UHR, che hanno mostrato un arricchimento maggiore di 640 volte per il Respiratory Virus Oligos Panel (Tabella 4).

La copertura media fornisce una panoramica della profondità di copertura per ciascuna base nel genoma di riferimento, identificando eventuali regioni che potrebbero non essere state sequenziate. L'arricchimento con sonde virali target ha aumentato notevolmente la copertura media della sequenza virale per tutti i campioni, indipendentemente dal metodo di trascrizione inversa (Tabella 5).

Tabella 4: metriche di arricchimento del Respiratory Virus Oligo Panel

Campione	Illumina TruSeq RNA		Thermo Maxima H Minus	
	CoVOC43	CoVOC43_95UHR	CoVOC43	CoVOC43_95UHR
Valore di arricchimento	5,5x	646x	3,1x	748x

Tabella 5: copertura media con il Respiratory Virus Oligos Panel<sup>a</sup>

Kit cDNA	CoVOC43		CoVOC43_95UHR	
	Non arricchito	Arricchito	Non arricchito	Arricchito
Illumina TruSeq RNA	427,1	3.283,5	2,6	2.437,3
Thermo Maxima H Minus	987,0	3.861,4	4,2	3.035,0

a. Le metriche di copertura medie sono state normalizzate a 1 milione di letture per campione.

### Analisi con la piattaforma IDbyDNA Explify

La piattaforma Explify ha identificato il coronavirus addizionato (Figura 5) e tutti e cinque i virus nel pool virale (Figura 6). Inoltre, ha fornito sequenze consenso di genomi virali, grafici di copertura e la capacità dimostrata di rilevare le coinfezioni con altri virus, batteri, funghi o parassiti (Figura 5 e Figura 6).

La piattaforma Explify ha inoltre dimostrato che i dati di arricchimento CoVOC43\_95UHR ottenuti dal Respiratory Virus Oligos Panel hanno mostrato una copertura equivalente ai dati di sequenziamento con metagenomica shotgun<sup>4</sup> (Figura 7), dimostrando che l'arricchimento con pool di sonde target virali non introduce distorsioni di copertura rispetto al metodo shotgun.

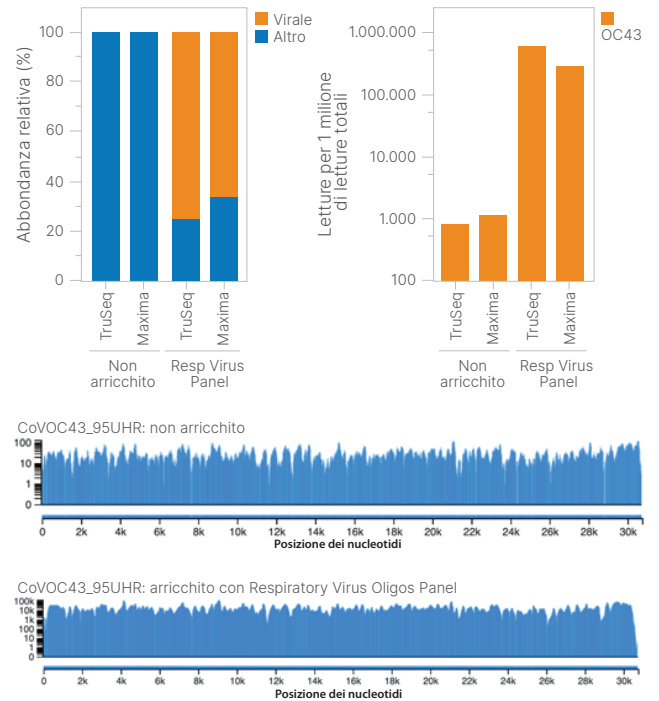
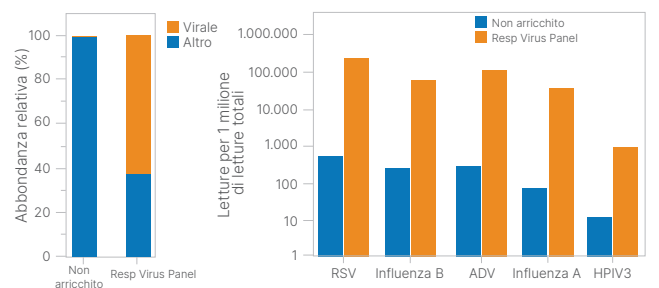


Figura 5: identificazione di OC43. L'OC43 è stato identificato, indipendentemente dal kit cDNA, con un'elevata copertura genomica. I dati sono stati sottocampionati a 1 milione di letture per campione.



MD	Alert	Pathogen	Comment	Organism Name	Evidence	Type	% Coverage	ANI	Median Depth
				Human respiratory syncytial virus B <sup>9</sup>	13,658	RNA	100.0%	98.8%	22,444
				Influenza A virus (H1N1) <sup>9</sup>	16,950	RNA	99.3%	97.6%	2,337
				Influenza B virus <sup>9</sup>	3,687	RNA	99.2%	98.1%	8,457
				Human adenovirus 7 <sup>9</sup>	198	DNA	99.1%	99.7%	3,833
				Human parainfluenza virus 3 <sup>9</sup>	3	RNA	85.6%	95.0%	15

Figura 6: caratterizzazione dei pool virali: tutti i ceppi virali di RNA e DNA sono stati identificati nel pool virale. Il pool virale non è stato verificato con tecniche ortogonali, pertanto non vi è alcuna garanzia di rappresentazione equivalente di ciascun virus. I dati sono stati sottocampionati a 1 milione di letture per campione.

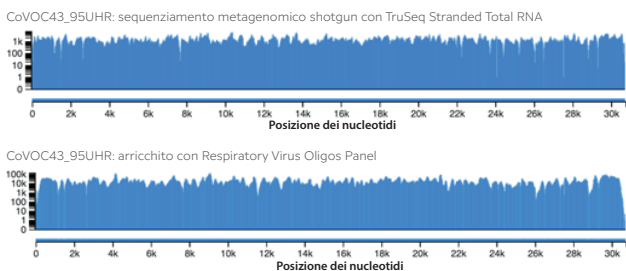
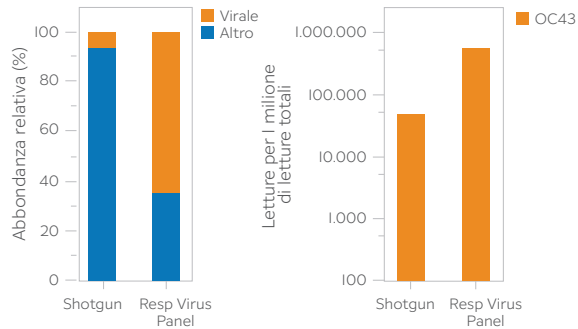


Figura 7: prestazioni equivalenti con il metodo shotgun. L'arricchimento ha mostrato prestazioni e copertura equivalenti al sequenziamento con metagenomica shotgun.

## Riepilogo

L'identificazione e la caratterizzazione dei virus emergenti è fondamentale per migliorare la salute pubblica. In queste situazioni, NGS è un metodo potente per il rilevamento ad ampio raggio per identificare i virus noti e quelli emergenti. L'utilizzo di Illumina DNA Prep with Enrichment con pannelli che mirano ai patogeni virali consente ai ricercatori di

ottenere dati genomici che possono confermare la presenza di CoV e far progredire le analisi come la genotipizzazione e l'analisi delle varianti. Il design imparziale consente l'identificazione diffusa di virus patogeni su tutti i tipi di campioni di interesse e l'uso di indici doppi unici riduce il rischio di qualsiasi crossover di indicizzazione da campioni in multiplex. Questo flusso di lavoro facile da seguire consente il rilevamento e la caratterizzazione delle epidemie di patogeni, come il nuovo SARS-CoV-2.

## Maggiori informazioni

Metodi di sequenziamento virale

## Bibliografia

1. World Health Organization. COVID 19 Public Health Emergency of International Concern (PHEIC) Global research and innovation forum: towards a research roadmap. Pubblicato il 12 febbraio 2020. Consultato il 9 settembre 2023.
2. Bulcha B. Review on viral metagenomics and its future perspective in zoonotic and arboviral disease surveillance. *J Biol Agric Healthcare*. 2017;7(21):35–41.
3. Gaudin M, Desnues C. Hybrid Capture-Based Next Generation Sequencing and Its Application to Human Infectious Diseases. *Front Microbiol*. 2018;9:2924. Pubblicato il 27 novembre 2018. doi:10.3389/fmicb.2018.02924.
4. Illumina. Comprehensive workflow for detecting coronavirus using Illumina benchtop systems. Pubblicato nel 2020. Consultato il 29 agosto 2023.

illumina®

Numero verde 1.800.809.4566 (U.S.A.) | Tel. +1.858.202.4566  
techsupport@illumina.com | www.illumina.com

© 2023 Illumina, Inc. Tutti i diritti riservati. Tutti i marchi di fabbrica sono di proprietà di Illumina, Inc. o dei rispettivi proprietari. Per informazioni specifiche sui marchi di fabbrica, visitare la pagina web [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).  
M-GL-02146 ITA v1.0