Utility of long read Nanopore sequencing as a tool for rapid insight into the host response in patients infected with SARS-CoV-2 or influenza A virus at point-of-care

Natasha Y. Rickett
University of Liverpool
Catherine Hartley
University of Liverpool
Ghada T. Shawli
University of Liverpool
Andrew George Shapanis
University of Southampton
Aaron Ions Gardner
TopMD Precision Medicine Ltd
Nicholas Harding
TopMD Precision Medicine Ltd
Jelmer Legebeke
University of Southampton
Jenny Lord
University of Southampton
Stephen Poole
National Institute for Health Research (NIHR), University of Southampton, University Hospital Southampton National Health Service (NHS) Foundation Trust
Nathan J. Brendish
National Institute for Health Research (NIHR), University of Southampton, University Hospital Southampton National Health Service (NHS) Foundation Trust
Anthony P. Williams
University Hospital Southampton
Fabio Strazzeri
TopMD Precision Medicine Ltd
James P. R. Schofield
TopMD Precision Medicine Ltd
Paul J. Skipp
University of Southampton
Tristan W. Clark
National Institute for Health Research (NIHR), University of Southampton, University Hospital Southampton National Health Service (NHS) Foundation Trust
Diana Baralle
University of Southampton
Julian A. Hiscox
University of Liverpool
Rebekah Penrice-Randal
R.Penrice-Randal@liverpool.ac.uk

University of Liverpool

Research Article

Keywords: COVID-19, critical care, RNA-sequencing, point-of-admission, transcriptomics

Posted Date: August 11th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7300653/v1

Additional Declarations: Competing interest reported. TC has received speaker fees, honoraria, travel reimbursement, and equipment and consumables free of charge for the purposes of research from BioFire diagnostics LLC and BioMerieux. TC has received discounted equipment and consumables for the purposes of research from QIAGEN. TC has received consultancy fees from Biofire diagnostics LLC, BioMerieux, Synairgen research Ltd, Randox laboratories Ltd and Cidara therapeutics. TC has been a member of advisory boards for Roche and Janssen and has received reimbursement for these. TC is member of two independent data monitoring committees for trials sponsored by Roche. TC has previously acted as the UK chief investigator for trials sponsored by Janssen. TC is currently a member of the NHSE COVID-19 Testing Technologies Oversight Group and the NHSE COVID-19 Technologies Validation Group. JS is a founding director, CEO, employee and shareholder in TopMD Precision Medicine Ltd. FS is a founding director, CTO, employee and shareholder in TopMD Precision Medicine Ltd. PS is a founding director, employee and shareholder in TopMD Precision Medicine Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abstract

Long read sequencing with Nanopore provides a platform for rapid, accessible, and flexible sequencing of nucleic acids, which is advantageous when faced with an outbreak or biological threat, particularly in situations where there is a lack of and/or demands on specific infrastructure. Investigation of the blood transcriptome can provide a snapshot of the systemic responses within a patient with a view to categorise and/or predict disease progression and outcome. For patients infected with SARS-CoV-2, analysis of the blood transcriptome using Illumina short read sequencing showed that there was a delayed adaptive immune response at the transcript level in those that went on to have a fatal outcome, compared to survivors. The profile of transcript expression data could also be used to predict admission to the intensive care unit (ICU). The use of Nanopore sequencing as a tool for rapid insight into the host response was explored using samples taken at point-of-care from patients infected with SARS-CoV-2 or influenza. This was compared to the same samples that had been previously sequenced using Illumina. Although an overall lower abundance of transcripts was identified using Nanopore sequencing in this study, when combined with bioinformatic analysis of expression pathways and functional analysis results, the findings were comparable. This work demonstrated that Nanopore sequencing is a feasible method for gaining rapid insight into the host response, which may contribute to effective patient care and help inform policy decisions in the face of a novel communicable disease, especially in a low-resource setting.

Introduction

The emergence of SARS-CoV-2 in late 2019 and the trajectory of the resultant pandemic reinforced the need for robust pandemic preparedness. In the face of a novel disease, rapid characterisation of the host response and the disease-causing agent can aid in the development of medical countermeasures (MCMs), particularly in therapeutics. Long read Nanopore sequencing has been an essential tool for genomic epidemiology worldwide, quickly providing real time data that can inform public health response(1–5).

As was shown in Ebola virus disease (EVD), the host response can play an instrumental role in disease progression and outcome(6). Organ damage observed in lung tissue in post-mortem samples from patients who died of severe coronavirus disease 2019 (COVID-19) highlighted a virus-independent immunopathologic response(7). Respiratory viral infections are a leading cause of disease globally. The impact on the health of both adults and children can be significant and pose a threat to health services, especially during the peak of seasonal infection rates(8–10). How the host responds to infection can dictate disease progression and severity and understanding this response can provide insight into the clinical management of patients and the selection and design of therapeutic interventions. For example, in hospitalised patients infected with SARS-CoV-2, dexamethasone was used to reduce mortality(11, 12). This compound is a glucocorticoid that is thought to decrease inflammation and the immune pathology associated with COVID-19 (13, 14).

Previously, sequencing of the blood from patients infected with influenza A virus (IAV) or SARS-CoV-2 revealed overlapping transcriptomic signatures. The two infections could be distinguished by elevated innate immune responses in patients infected with IAV and an elevated adaptive immune response in patients infected with SARS-CoV-2 (15). By using topological methods, pathways enriched with genes expressed by the transcription factor Myc were also identified as differentially activated (15). Activated Myc supresses Interferon Regulatory Factor (IRF) 7 expression, which was reported in the COVID-19 cohort (15). In addition to characterising the molecular differences between patients with a confirmed COVID-19 or IAV diagnosis at point-of-admission, severe cases (i.e. ICU admissions) (16) were explored to identify potential differentiating markers tied to prognostic indicators. An adaptive immune response was associated with survival (15), and pathway activation in those admitted to the ICU were consistent with acute lung injury and fibrosis (15, 17).

To further this work, we employed Nanopore cDNA-PCR sequencing of RNA in a peripheral blood sample, to determine the feasibility of characterising host responses in patients infected with SARS-CoV-2 or IAV at point of admission. As sequencing technologies and bioinformatic analyses continue to develop, there may be future scope for use of transcriptomics to guide clinical decision making, especially as precision and personalised medicine continue to gain traction. To assess the utility of Nanopore transcriptomic sequencing in these cohorts (16). Nanopore devices and technology provide opportunities for decentralised, remote, and rapid sequencing of nucleic acids(18–23). We found that a global level transcriptomic analysis of blood was equivalent to that delivered by Illumina. The work suggested that Nanopore sequencing could be used in an outbreak/resource limited setting to evaluate how the patient responds to an infectious disease. This will help in the more rapid development of MCM.

Methods

Samples

Samples were obtained from patients infected with SARS-CoV-2 or IAV enrolled into clinical studies of hospitalised adults, as previously described (15, 16). All patients were recruited prior to receiving any treatment and within the first 24 hours of admission to hospital.

Healthy participants were recruited from an ongoing study at the University of Liverpool: “Immune responses to infection of white blood cells from healthy people”. The research protocol was approved by the Ethics Committee of the Institute of Infection and Global Health, University of Liverpool (RETH000685). Informed consent was obtained from each subject. 4 ml venous blood was collected aseptically from each individual in standard EDTA tubes.

RNA was extracted from PAXgene tubes and sequenced on the Illumina NovaSeq, as described previously (16, 24).

Nanopore sequencing

cDNA sequencing libraries were prepared using the PCR-cDNA Barcoding sequencing kit (SQK-PCB109), starting with 50 ng of total RNA. Barcoded samples were pooled together ensuring that samples were equimolar and did not exceed 100 fmol per flow cell. For adapter ligation, 1 µl of Rapid Adapters (RAP) were added to pooled sequencing libraries and incubated at room temperature for 5 minutes at room temperature. The MinKNOW GUI was used to initiate sequencing and was ran for up to 72 hours. Fast5 files were base called using the high accuracy calling model. For each sequencing run, 5–10 samples were multiplexed per flow cell.

Bioinformatics

Multiplexed sequencing reads were base called and demultiplexed by Guppy basecaller. Minimap2 was used to index and map reads to the reference transcriptome Homo_sapiens.GRC.38.cdna.all.fa to generate alignment files using the –ax map-ont -N 100 -p 1.0 parameters (25). Alignment files were sorted and indexed with samtools before counting bam files using Salmon in alignment model using –noErrorModel -l U parameters (25, 26). For Illumina data, paired reads were inputted into Salmon.

Gene expression inference

The quant.sf files produced by salmon were imported into RStudio (v.4.4.1) with the tximport library and the Homo_sapiens.GRCh38.107.gtf was imported with the rtracklayer library to annotate the count matrix. Nanopore and Illumina count data were imported into separate matrices prior to filtering and normalisation.

The edgeR package was used to filter lowly expressed transcripts and normalise sequencing libraries using the TMM method. Samples with small libraries were removed and differentially expressed transcripts were identified, with these being defined as at least a 2-fold difference from the healthy group (n = 5) and a false discovery rate (FDR) less than 0.05 (27).

Biological themes

Different approaches to pathway analysis were explored, including CAMERA(28), gene set enrichment (GSE) analysis (29) and Gene Ontology (GO). Differential transcript expression data was used for transcript ontology enrichment analysis of biological process terms in each group using the ClusterProfiler programme in R (30). The compareCluster function was used to compare increasing and decreasing DE transcripts in each group with a q-value cut-off of 0.05 was used with a Benjamini-Hochberg-FDR correction. The simplify function was used to remove redundant GO terms and the top 20 biological processes are presented for each condition.

Results

Using blood samples from patients infected with SARS-CoV-2 or IAV and those from healthy participants, long read and short read sequencing was compared by data acquired on Nanopore and Illumina platforms, respectively. Note the data acquired on the Illumina platform had been previously published by our team(15, 16). A subset of 40 samples were identified as meeting the minimum library size threshold (1,000,000 reads) and were taken forward for bioinformatic analysis (healthy volunteers; n=5, patients infected with SARS-Cov-2; n=20, patients infected with IAV; n=15). The characteristics and comorbidities of the patients are

reported in Table 1, with available characteristics of healthy participants in Supplementary Table 1 and further detail published previously (24). As described, patients infected with SARS-CoV-2 were generally older than patients infected with IAV and experienced a longer stay in hospital, though neither variable reached the threshold of statistical significance. As expected due to the similar risk factors for respiratory illnesses, no significant difference in burden of comorbidities was observed between patients in each cohort (Table 1). Detailed characteristics, comorbidities and outcomes are reported in Supplementary Tables 1 and 2. To further explore the differences between cohorts in molecular phenotype, blood samples were taken from these patients with 24 hours of hospital admission. RNA was extracted from the blood samples and sequenced by Illumina (reported previously) (15, 17). In this study, the same samples were sequenced on a Nanopore platform to explore the feasibility of employing Nanopore as a point-of-care transcriptomics tool. As expected, samples sequenced with Illumina generated a dataset containing a higher abundance of transcripts in comparison to Nanopore after filtering by expression level (i.e. retaining genes which had sufficiently large counts to be used in statistical analyses) (Figure 1A, Table 2).

Table 1: Basic characteristics and comorbidities of patients infected with SARS-CoV-2 (COVID-19 cohort) or influenza virus (IAV cohort) that were included in this study.

Variable	N	COVID-19, N = 20[1]	Influenza (IAV), N = 15[1]	p-value[2]
Age (years)	35			0.010
Median (IQR)		72 (67, 77)	45 (38, 68)	
Sex	35			0.4
F		8 (40%)	4 (27%)	
M		12 (60%)	11 (73%)	
Hypertension	35			0.080
no		9 (45%)	12 (80%)	
yes		10 (50%)	3 (20%)	
unknown		1 (5.0%)	0 (0%)	
Cardiovascular disease	35			0.2
no		12 (60%)	13 (87%)	
yes		7 (35%)	2 (13%)	
unknown		1 (5.0%)	0 (0%)	
Respiratory disease	35			0.6
no		13 (65%)	8 (53%)	
yes		6 (30%)	7 (47%)	
unknown		1 (5.0%)	0 (0%)	
Renal disease	35			>0.9
no		17 (85%)	14 (93%)	
yes		2 (10%)	1 (6.7%)	
unknown		1 (5.0%)	0 (0%)	
Diabetes mellitus	35			0.2
no		12 (60%)	13 (87%)	
yes		7 (35%)	2 (13%)	
unknown		1 (5.0%)	0 (0%)	
Immunosuppression	35			>0.9
no		17 (85%)	14 (93%)	
yes		2 (10%)	1 (6.7%)	
unknown		1 (5.0%)	0 (0%)	
C-reactive protein (i.e. inflammation / injury)	34			0.6
Median (IQR)		108 (48, 140)	118 (55, 182)	
O_2 saturation	34			0.6
Median (IQR)		95.00 (91.50, 96.00)	95.00 (93.25, 96.75)	
National early warning score 2 on admission	34			0.5

Median (IQR)		5.00 (4.00, 6.50)	4.50 (2.25, 6.75)	
Symptom duration (days)	35			0.019
Median (IQR)		7.5 (5.0, 11.8)	4.0 (3.5, 5.5)	
[1] **n (%)**				
[2] **Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test**				

Table 2: EdgeR was used to filter transcripts expressed at lower levels and to normalise sequencing libraries. Differentially expressed (DE) transcripts were identified between patients infected with SARS-CoV-2 (COVID-19 cohort) or influenza (IAV cohort) and healthy participants. The abundance of shared DE transcripts are reported.

Dataset	Total number of transcripts kept after filterbyExpr	DE transcripts (increase)			DE transcripts (decrease)		
		COVID-19	IAV	Shared DE transcripts	COVID-19	IAV	Shared DE transcripts
Nanopore data set	5751	316	361	3	112	137	0
Illumina data set	15761	2257	2210	12	270	322	8

Similarities were observed at the transcript level between patients infected with SARS-CoV-2 and those infected with IAV

Sequencing data derived from blood samples from patients admitted with SARS-CoV-2 or IAV was acquired from Nanopore and Illumina platforms. These files were mapped to the human transcriptome before calculating gene expression. Of interest, a higher mapping rate to the human transcriptome was observed in the Nanopore dataset (~90%), whereas the dataset generated by Illumina had a much lower mapping rate (30%).

The Nanopore dataset was smaller than the Illumina dataset, therefore differences would be expected (Figure 1A, Table 2). To determine how the groups clustered over 2-dimensional space to explore initial trends and outliers, a principal component analysis (PCA) was used. The data indicated that both Nanopore and Illumina datasets clustered separately from healthy participants (Figure 1B). Overlap between transcriptional profiles were observed between patients infected with SARS-CoV-2 or IAV, as reported previously (15). The library size acquired during these sequencing experiments differed between platforms, however, the normalised expression moderately correlated at the individual level (Supplementary Figure 1). Upon calculating log2FC values, the datasets were more strongly correlated (Figure 1C). In both datasets, identified transcripts associated with immunoglobulins are upregulated (as previously described) (15), although these did not reach appropriate levels of significance.

Library preparation and library size may drive differences between datasets

The transcriptional profiles from patients infected with SARS-CoV-2 and IAV overlap, therefore indicating that the differences between these groups may not be distinct (Figure 2A). The observed differences between Nanopore and Illumina data are largely driven by the differences in dataset sizes, library preparation and short read and long read transcript quantification. Just over half of the genes identified in the Nanopore dataset were also identified in the Illumina dataset (Figure 2B), leaving just under half in disagreement or not found.

Comparing biological themes derived from Nanopore and Illumina datasets

Nanopore sequencing of RNA extracted from patient blood provides a shallower insight of the transcriptional activity at the gene expression level when compared to Illumina, demonstrating that different transcripts are identified depending on which sequencing platform is used. Therefore, in order to reduce bias, multiple methods were employed to assess similarities and differences in biological pathways - including CAMERA and GSE analysis.

Correlation Adjusted MEan RAnk gene set testing (CAMERA) shows some consistency between Illumina and Nanopore datasets

Gene set analysis was conducted with CAMERA to compare the datasets(28). Significant results from the Illumina and Nanopore dataset from patients infected with SARS-CoV-2 versus IAV were joined and then ordered by gene set size from Nanopore (Table 3) and then Illumina (Table 4). Results from patients infected with SARS-CoV-2 or IAV versus healthy participants are in Supplementary Tables 4 and 5, respectively. The top ten gene sets from Nanopore were also identified in the Illumina dataset and agreed in direction of expression - i.e. up or downregulation in the infected cohort - with the exception of "GO:0000398: mRNA splicing, via spliceosome" (Tables 3 and 4). This disagreement may be due to the different library preparation and the higher abundance of ribosomal transcripts present in the Nanopore dataset. When ordering by the Illumina dataset (Table 4), the top pathway, innate immune response, remains consistent across datasets, however, other pathways were not identified as significant in the Nanopore data. Though generally more abundant transcripts were identified in Illumina compared to Nanopore datasets (Figure 3A), the size of the gene sets significantly correlated between platforms (Figure 3B).

Table 2: CAMERA geneset enrichment results ordered by Nanopore gene scores (COVID-19 cohort versus IAV cohort). Gene sets with a direction of 1 were upregulated in patients infected with SARS-CoV-2, those with a direction of -1 were downregulated.

gene_set	nr_genes_illumina	dir_illumina	pval_illumina	nr_genes_nanopore	dir_nanopore	pval_nanopore
innate immune response	397	1	1.03e-05	231	1	1.50e-03
inflammatory response	299	1	3.39e-02	137	1	6.55e-03
defense response to virus	192	1	9.65e-12	134	1	8.06e-06
mRNA splicing, via spliceosome	173	1	2.90e-02	126	-1	2.78e-03
positive regulation of canonical NF-kappaB signal transduction	200	1	1.87e-03	117	1	1.09e-02
cytoplasmic translation	88	1	4.69e-02	86	-1	4.53e-21
G protein-coupled receptor signaling pathway	347	-1	3.92e-02	79	1	7.46e-04
protein polyubiquitination	126	1	2.89e-03	68	1	2.36e-03
response to virus	91	1	8.81e-05	62	1	3.87e-02
positive regulation of tumor necrosis factor production	87	1	4.17e-05	61	1	2.45e-02

Table 3: CAMERA geneset enrichment results ordered by Illumina gene scores (COVID-19 cohort versus IAV cohort). Gene sets with a direction of 1 were upregulated in patients infected with SARS-CoV-2, those with a direction of -1 were downregulated.

gene_set	nr_genes_illumina	dir_illumina	pval_illumina	nr_genes_nanopore	dir_nanopore	pval_nanopore
innate immune response	397	1	1.03e-05	231	1	1.50e-03
G protein-coupled receptor signaling pathway	347	-1	3.92e-02	79	1	7.46e-04
protein transport	325	1	3.94e-03	NA	NA	NA
cell adhesion	324	-1	2.51e-02	NA	NA	NA
inflammatory response	299	1	3.39e-02	137	1	6.55e-03
adaptive immune response	242	-1	4.59e-02	NA	NA	NA
positive regulation of canonical NF-kappaB signal transduction	200	1	1.87e-03	117	1	1.09e-02
ubiquitin-dependent protein catabolic process	197	1	3.40e-02	NA	NA	NA
defense response to virus	192	1	9.65e-12	134	1	8.06e-06
mRNA processing	177	1	3.09e-02	NA	NA	NA

Exploring the intersect between Nanopore and Illumina gene set analysis

GSEA analysis was conducted in clusterProfiler after genes were ranked by the log2FC value. Analysis of the Illumina dataset identified 118 significant gene sets based on differentially expressed genes between patients infected with SARS-CoV-2 or IAV, whereas analysis of the Nanopore dataset identified 78, with 17 being shared between both (Figure 4A). Gene sets that are shared between the technologies are biologically relevant when contrasting COVID-19 and IAV cohorts. Terms are mostly in agreement, for example "defense response to virus", "regulation of response to biotic stimulus", "regulation of defense response" and "regulation of innate immune response", among other pathways associated with cytokine production and response (Figure 4B). However, there was disagreement in directionality of gene sets; "nervous system process", "disruption of anatomical structure in another organism", “killing of cells of another organism" and "disruption of cell in another organism" (Figure 4B).

Discussion

Severe and fatal COVID-19 is contributed to by the host response, whereby the presence of virus and inflammation does not always co-exist in tissues (7). The involvement of the host response in patient outcome is further supported by the impact of dexamethasone on the outcome of patients within intensive care units which is able to reduce inflammation through immunosuppression (11, 12). This work utilised samples from the COVID-19 and influenza point-of-care testing trials to determine the blood transcriptome in patients who had just been admitted to hospital (31–33). All samples were sequenced using Illumina to acquire a substantial read depth for differential transcript expression; however, a subset of samples was selected at random to sequence on the MinION to determine how the rapid, long read technology compared to the gold standard Illumina sequencing.

Both IAV and SARS-CoV-2 infections in humans can result in severe disease and the need for hospitalisation. The blood transcriptomes from patients infected with SARS-CoV-2 or IAV at point-of-care were compared to examine differences in host response. There were no fatal cases within the IAV cohort, however, there were 15 fatal cases amongst patients infected with SARS-CoV-2. Through differential gene expression analysis, a plethora of genes were identified that were unique to the COVID-19 cohort versus IAV. Interestingly, through both Illumina and Nanopore sequencing, immunoglobulin transcripts were identified at an

increased abundance when comparing the blood transcriptome of patients infected with SARS-CoV-2 to IAV. The significant over-representation of immunoglobulin transcripts associated with the heavy chain and light chain V transcripts has been reported previously(34).

A previous study that utilised single cell sequencing of peripheral blood mononuclear cells (PBMCs) from ten patients infected with SARS-CoV-2 identified an over-representation of the immunoglobulin heavy chain variable region (IGHV)3 family, specifically, IGHV3-30, IGHV3-7, IGHV3-15, IGHV3-21 and IGHV3-23 (36). IGHV3-30 has been shown to facilitate the encoding of primary antibodies to neutralise human cytomegalovirus (37, 38). Additional studies into SARS-CoV-2 reveal that IGHV3-30 was overrepresented in convalescent patients and that IGHV3-30 facilitated kappa over lambda chains, it has also been hypothesised that IGHV3-30 is part of the initial response within the immunological repertoire under emergency situations (39, 40). This has been further validated in other investigations where it has been concluded that S-reactive IgG positive B-cell responses are readily developed after infection where the same B-cell clones are detectable over time with a preference for the IGHV3-30 transcript segment (41). Influenza viruses have also been recorded to utilise this transcript for the transcription of broadly neutralising antibodies, and with previous studies the importance of this transcript seems to be important for the B-cell response to both respiratory viruses, although abundance is higher in COVID-19 patients (42).

In addition to investigating the blood transcriptome of patients infected with SARS-CoV-2 or IAV at point-of-care, sequencing methodologies were compared and contrasted due to the extended turnaround time associated with Illumina sequencing. It should be noted that only a subset of samples, reaching a threshold of certain library size, was sequenced on a Nanopore platform, with matched samples were considered throughout analysis. This selection of larger datasets could potentially limit the generalisability of these results to all patient samples. Although Nanopore sequencing returned fewer sequencing reads in comparison with Illumina, the reads acquired were approximately 750 base pairs long and were associated with a high mapping rate to the human transcriptome. Due to a much lower read depth, fewer transcripts were identified and thus fewer differentially expressed transcripts were reported within the Nanopore dataset. Filtered and normalised transcript counts were compared between matched Illumina and Nanopore datasets. There was a moderate correlation when comparing the transcript counts, however, when contrasts were formed against samples collected from healthy participants to determine the log2FC of transcripts, a much stronger correlation was observed (Fig. 1). Following independent analysis of both datasets, transcripts identified as increasing or decreasing in abundance were compared using Venn diagrams. Illumina datasets were richer in data, however, both sequencing technologies acquired unique transcripts that were changing in abundance between conditions. Therefore, allowing tentative suggestion that a Nanopore platform can be used to identify similar patterns, at the pathway level, to sequencing data by Illumina (Fig. 3B). Further validation on larger datasets would be required for concrete conclusions to be drawn. In addition, the recruitment of infected and healthy participants at different time points should be noted as a potential study limitation.

Nanopore sequencing generated longer reads and its data had a higher mapping rate to the human genome than Illumina sequencing reads, which could be due to the advantage of long reads in genome assembly, with particular advantage reported in areas of high repetition or homology (20, 35, 36). This also raises a number of questions about how we consider and handle Nanopore data. Given that these platforms generate datasets with different properties, it brings into question if handling them in the same manner is appropriate. If the generation of long reads leads to a higher mapping rate, then perhaps more lenient cut offs downstream could be implemented, allowing data integrity to be uncompromised whilst maximising the impact of the data generated. Critically, the advance in Nanopore technologies in the past decade, in accuracy, throughput and read length, the next decade could see a refinement in these properties.

With its rapid turnaround, generation of long reads and field deployable nature, the benefits of Nanopore sequencing both in low resource settings and clinical practice are clear. With transcriptomics now and in the future, any decision-making is likely to be a balance between depth of coverage, cost, and time.

Declarations

Conflicts of Interests

TC has received speaker fees, honoraria, travel reimbursement, and equipment and consumables free of charge for the purposes of research from BioFire diagnostics LLC and BioMerieux. TC has received discounted equipment and consumables for the purposes

of research from QIAGEN. TC has received consultancy fees from Biofire diagnostics LLC, BioMerieux, Synairgen research Ltd, Randox laboratories Ltd and Cidara therapeutics. TC has been a member of advisory boards for Roche and Janssen and has received reimbursement for these. TC is member of two independent data monitoring committees for trials sponsored by Roche. TC has previously acted as the UK chief investigator for trials sponsored by Janssen. TC is currently a member of the NHSE COVID-19 Testing Technologies Oversight Group and the NHSE COVID-19 Technologies Validation Group. JS is a founding director, CEO, employee and shareholder in TopMD Precision Medicine Ltd. FS is a founding director, CTO, employee and shareholder in TopMD Precision Medicine Ltd. PS is a founding director, employee and shareholder in TopMD Precision Medicine Ltd.

Competing Interests

TC has received speaker fees, honoraria, travel reimbursement, and equipment and consumables free of charge for the purposes of research from BioFire diagnostics LLC and BioMerieux. TC has received discounted equipment and consumables for the purposes of research from QIAGEN. TC has received consultancy fees from Biofire diagnostics LLC, BioMerieux, Synairgen research Ltd, Randox laboratories Ltd and Cidara therapeutics. TC has been a member of advisory boards for Roche and Janssen and has received reimbursement for these. TC is member of two independent data monitoring committees for trials sponsored by Roche. TC has previously acted as the UK chief investigator for trials sponsored by Janssen. TC is currently a member of the NHSE COVID-19 Testing Technologies Oversight Group and the NHSE COVID-19 Technologies Validation Group. JS is a founding director, CEO, employee and shareholder in TopMD Precision Medicine Ltd. FS is a founding director, CTO, employee and shareholder in TopMD Precision Medicine Ltd. PS is a founding director, employee and shareholder in TopMD Precision Medicine Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: the CoV-19POC trial was funded by University Hospital Southampton Foundation Trust (UHSFT) and the FluPOC trial by the National Institute for Health Research (NIHR) Post-Doctoral Fellowship Programme. In addition, the CoV-19POC and FluPOC trials were supported by the NIHR Southampton Clinical Research Facility and NIHR Southampton Biomedical Research Centre (BRC). JLe was supported by a PhD studentship from the NIHR Southampton BRC (no. NIHR-INF-0932). RP-R was supported by a PhD studentship from the Medical Research Council Discovery Medicine North Doctoral Training Partnership (no. MR/N013840/1). NB was supported by the NIHR Clinical Lecturer scheme. JHi,RPR, CH and XD were supported by the US Food and Drug Administration (no. 75F40120C00085), and this work was partly supported by U.S. Food and Drug Administration Medical Countermeasures Initiative (no 75F40120C00085) awarded to JHi. TC was supported by a NIHR Post-Doctoral Fellowship (no. 2016-09-061). DB and her laboratory are supported by a NIHR Research Professorship (no. RP-2016-07-011). This work was also supported by a UK MRC grant MR/Y004205/1 (G2P2-UK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contribution

Conceptualization (JPRS, PJS, TWC, DB, JAH and RP-R). Methodology (JLe, JLo, RP-R). Software (JLE, JLo, RP-R). Validation (NYR and RP-R). Formal analysis (NYR, JLE, JLo, RP-R). Investigation (all authors). Resources (PJS, TWC, DB). Data curation (RP-R). Writing – Original Draft (NYR, JAH and RP-R). Writing – review and editing (NYR, JAH and RP-R). Visualisation (NYR and RP-R). Supervision (JPRS, PJS, TWC, DB, JAH). Project Administration (JPRS, PJS, TWC, DB, RP-R, JAH). Funding acquisition (JPRS, PJS, TWC, DB, JAH).

Data Availability

The transcriptomic datasets derived from Illumina sequencing used in this study are available by request. Code used to perform the analysis and generate figures is available: [https://github.com/Hiscox-lab/Nanopore_illumina_transcriptomics_comp]

(https:/github.com/Hiscox-lab/Nanopore_illumina_transcriptomics_comp) .

References

1. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228-32.
2. Kafetzopoulou LE, Efthymiadis K, Lewandowski K, Crook A, Carter D, Osborne J, et al. Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. Eurosurveillance. 2018;23(50):1800228.
3. Kafetzopoulou LE, Pullan ST, Lemey P, Suchard MA, Ehichioya DU, Pahlmann M, et al. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. Science. 2019;363(6422):74-7.
4. Freed NE, Vlková M, Faisal MB, Silander OK. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. Biol Methods Protoc. 2020;5(1):bpaa014.
5. Aljabr W, Alruwaili M, Penrice-Randal R, Alrezaihi A, Harrison AJ, Ryan Y, et al. Amplicon and Metagenomic Analysis of Middle East Respiratory Syndrome (MERS) Coronavirus and the Microbiome in Patients with Severe MERS. mSphere. 2021;6(4):e0021921.
6. Liu X, Speranza E, Muñoz-Fontela C, Haldenby S, Rickett NY, Garcia-Dorival I, et al. Transcriptomic signatures differentiate survival from fatal outcomes in humans infected with Ebola virus. Genome Biology. 2017;18(1):4.
7. Dorward DA, Russell CD, Um IH, Elshani M, Armstrong SD, Penrice-Randal R, et al. Tissue-specific Immunopathology in Fatal COVID-19. American Journal of Respiratory and Critical Care Medicine. 2020.
8. Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the Global Burden of Disease Study 2017. Lancet Respir Med. 2019;7(1):69-89.
9. Morbey RA, Harcourt S, Pebody R, Zambon M, Hutchison J, Rutter J, et al. The burden of seasonal respiratory infections on a national telehealth service in England. Epidemiol Infect. 2017;145(9):1922-32.
10. Dawood FS, Iuliano AD, Reed C, Meltzer MI, Shay DK, Cheng P-Y, et al. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. The Lancet Infectious Diseases. 2012;12(9):687-95.
11. Sterne JAC, Murthy S, Diaz JV, Slutsky AS, Villar J, Angus DC, et al. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. Jama. 2020;324(13):1330-41.
12. Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. N Engl J Med. 2020.
13. Chimote AA, Alshwimi AO, Chirra M, Gawali VS, Powers-Fletcher MV, Hudock KM, Conforti L. Immune and ionic mechanisms mediating the effect of dexamethasone in severe COVID-19. Front Immunol. 2023;14:1143350.
14. Sinha S, Rosin NL, Arora R, Labit E, Jaffer A, Cao L, et al. Dexamethasone modulates immature neutrophils and interferon programming in severe COVID-19. Nat Med. 2022;28(1):201-11.
15. Legebeke J, Lord J, Penrice-Randal R, Vallejo AF, Poole S, Brendish NJ, et al. Evaluating the immune response in treatment-naive hospitalised patients with influenza and COVID-19. Frontiers in Immunology. 2022(13).
16. Penrice-Randal R, Dong X, Shapanis A, Gardener AI, Harding N, Legebeke J, et al. Blood gene expression predicts intensive care unit admission in hospitalised patients with COVID-19. Frontiers in immunology. 2022.
17. Penrice-Randal R, Dong X, Shapanis AG, Gardner A, Harding N, Legebeke J, et al. Blood gene expression predicts intensive care unit admission in hospitalised patients with COVID-19. Frontiers in Immunology. 2022;13.
18. Sun X, Song L, Yang W, Zhang L, Liu M, Li X, et al. Nanopore Sequencing and Its Clinical Applications. In: Huang T, editor. Precision Medicine. New York, NY: Springer US; 2020. p. 13-32.
19. Oehler JB, Wright H, Stark Z, Mallett AJ, Schmitz U. The application of long-read sequencing in clinical settings. Human Genomics. 2023;17(1):73.
20. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348-65.

21. Ying YL, Hu ZL, Zhang S, Qing Y, Fragasso A, Maglia G, et al. Nanopore-based technologies beyond DNA sequencing. Nat Nanotechnol. 2022;17(11):1136-46.
22. Sheka D, Alabi N, Gordon PMK. Oxford nanopore sequencing in clinical microbiology and infection diagnostics. Brief Bioinform. 2021;22(5).
23. Schmidt J, Blessing F, Fimpler L, Wenzel F. Nanopore Sequencing in a Clinical Routine Laboratory: Challenges and Opportunities. Clin Lab. 2020;66(6).
24. Legebeke J, Lord J, Penrice-Randal R, Vallejo AF, Poole S, Brendish NJ, et al. Distinct immune responses in patients infected with influenza or SARS-CoV-2, and in COVID-19 survivors, characterised by transcriptomic and cellular abundance differences in blood. medRxiv. 2021:2021.05.12.21257086.
25. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094-100.
26. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods. 2017;14(4):417-9.
27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.
28. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. Nucleic Acids Res. 2012;40(17):e133.
29. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.
30. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284-7.
31. Beard K, Brendish N, Malachira A, Mills S, Chan C, Poole S, Clark T. Pragmatic multicentre randomised controlled trial evaluating the impact of a routine molecular point-of-care 'test-and-treat' strategy for influenza in adults hospitalised with acute respiratory illness (FluPOC): trial protocol. BMJ Open. 2019;9(12):e031674.
32. Brendish NJ, Poole S, Naidu VV, Mansbridge CT, Norton NJ, Wheeler H, et al. Clinical impact of molecular point-of-care testing for suspected COVID-19 in hospital (COV-19POC): a prospective, interventional, non-randomised, controlled study. The Lancet Respiratory Medicine. 2020;8(12):1192-200.
33. Clark TW, Beard KR, Brendish NJ, Malachira AK, Mills S, Chan C, et al. Clinical impact of a routine, molecular, point-of-care, test-and-treat strategy for influenza in adults admitted to hospital (FluPOC): a multicentre, open-label, randomised controlled trial. Lancet Respir Med. 2021;9(4):419-29.
34. Robbiani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z, Cho A, et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. Nature. 2020;584(7821):437-42.
35. Warburton PE, Sebra RP. Long-Read DNA Sequencing: Recent Advances and Remaining Challenges. Annu Rev Genomics Hum Genet. 2023;24:109-32.
36. Jain M, Abu-Shumays R, Olsen HE, Akeson M. Advances in nanopore direct RNA sequencing. Nat Methods. 2022;19(10):1160-4.

Figures

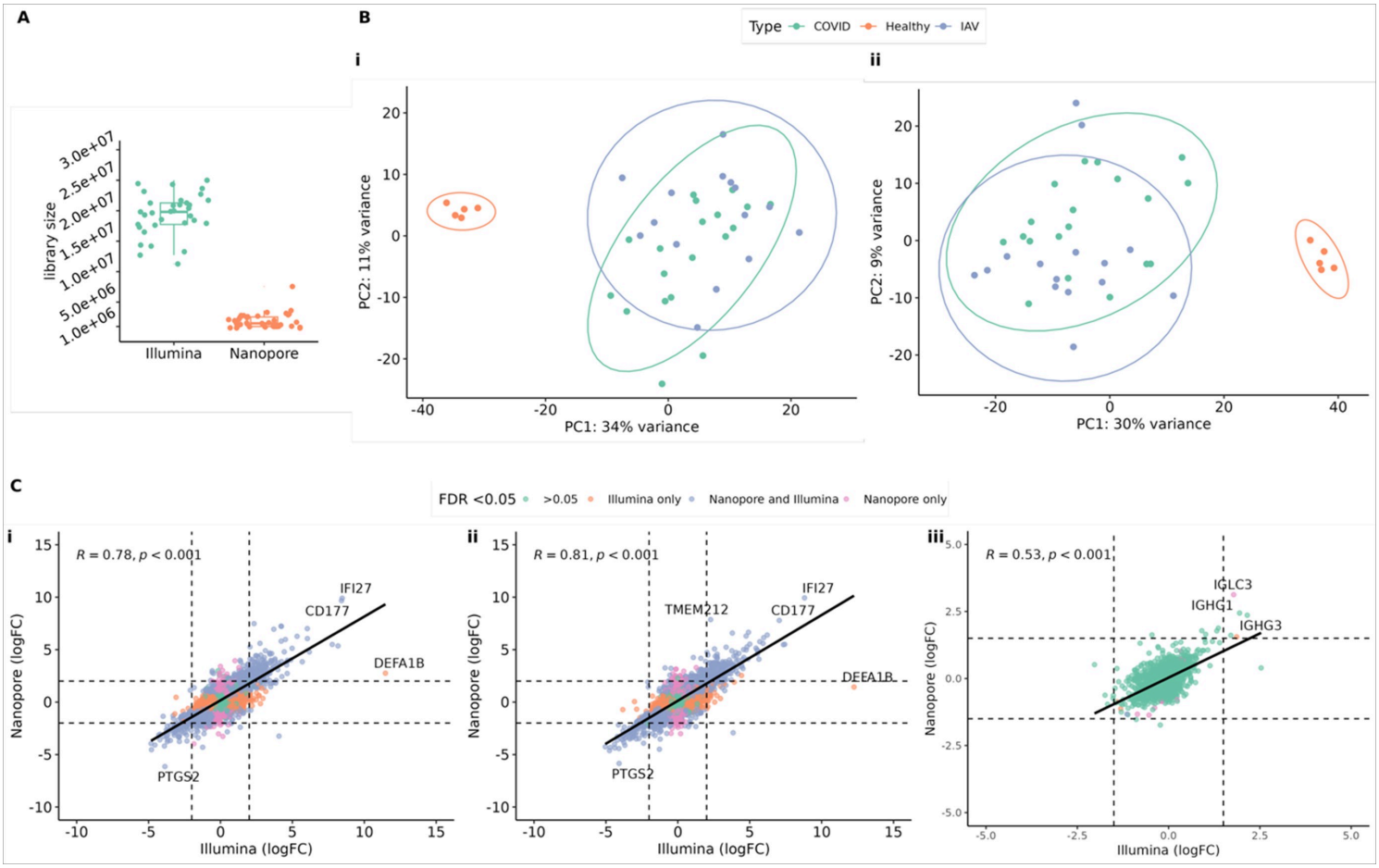

Figure 1

RNA extracted from the blood of patients infected with SARS-CoV-2 (COVID cohort) or influenza virus (IAV cohort). Samples sequenced either by Illumina (described previously) (15, 17)or Nanopore platform (this study) demonstrated reasonably comparable characteristics. A) Illumina datasets had library sizes of approximately 20 million, and Nanopore 1 million. B) Principal component analysis of Nanopore (i) and Illumina (ii) datasets showed that IAV and COVID-19 patient cohorts clustered together, away from healthy participants. C) Strong correlation was observed between log2 fold change (FC) values of transcript abundance in Illumina and Nanopore datasets contrasting (i) COVID-19 cohort/healthy samples, (ii) IAV cohort/healthy samples and moderate correlation contrasting (iii) COVID-19 cohort/IAV cohort samples. Colours indicate whether the gene was <0.05 FDR in one or both datasets.

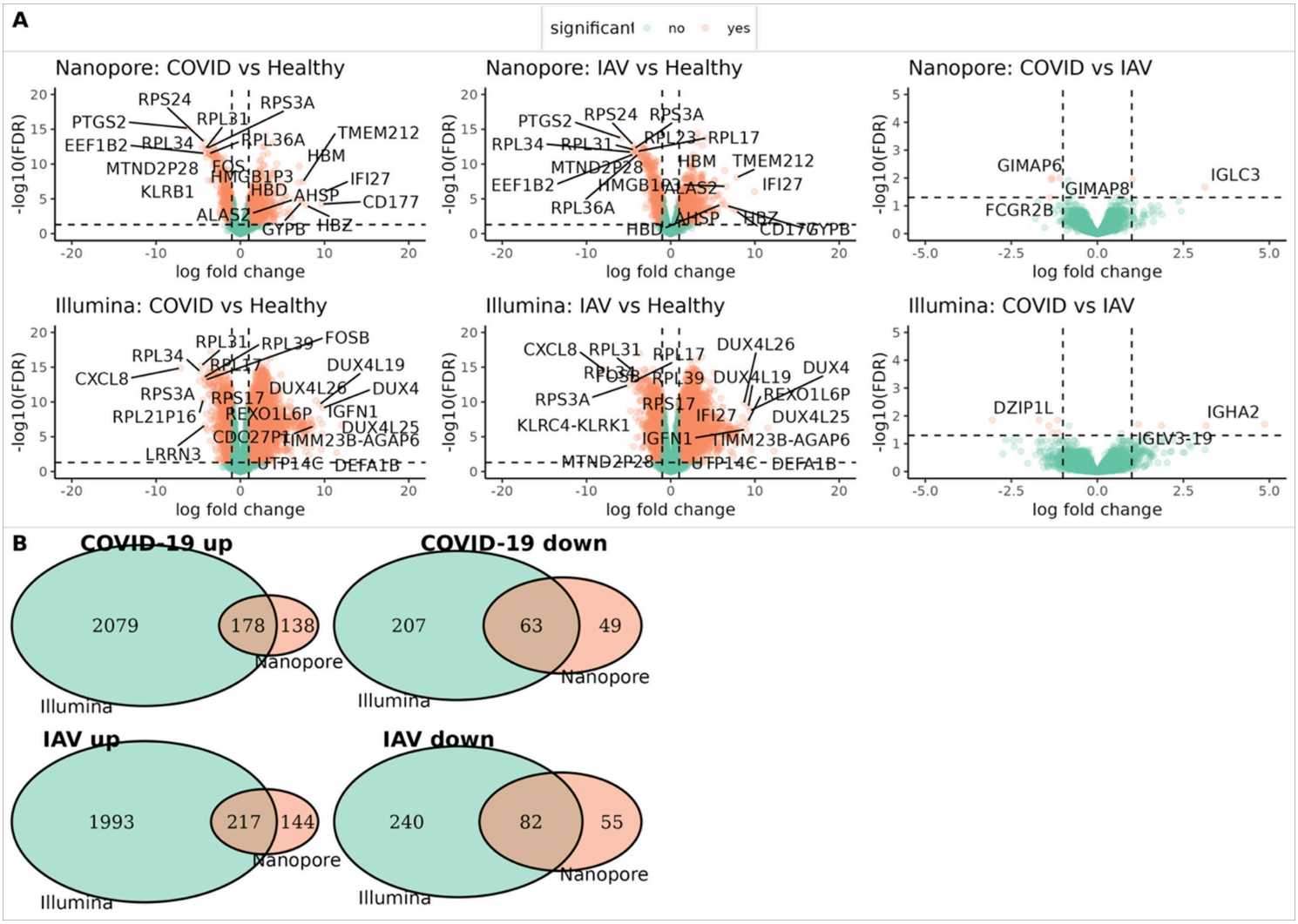

Figure 2

Differentially expressed genes in patients infected with SARS-CoV-2 (COVID-19 cohort), influenza virus (IAV cohort) when compared to healthy participants. A) Volcano plots show differentially expressed genes in the COVID-19 and IAV cohorts, identified by Nanopore and Illumina sequencing. The nodes represent genes, which are labelled. The orange represents the significant differentially expressed genes and green represents non-significant expressed genes. The horizontal black dashed line represents a threshold of FDR <0.05 whereas the vertical line represents thresholds for log2FC ≥ 2 or ≤ −2. B) Differentially expressed genes identified by different sequencing platforms are shown in each dataset. The intersection between the two circles represents the genes differentially expressed in both datasets.

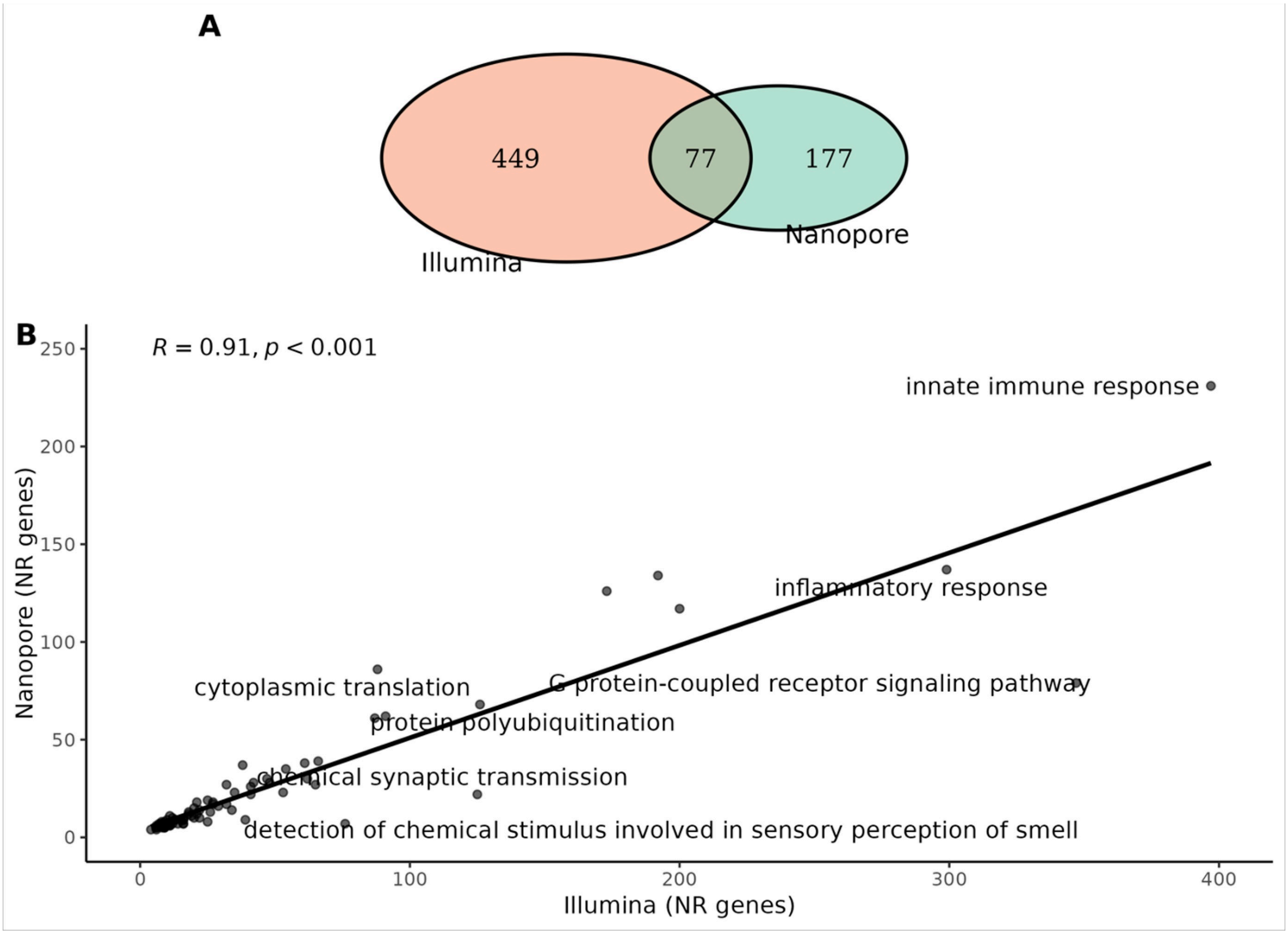

Figure 3

CAMERA was used to identify enriched gene sets when comparing patients infected with SARS-CoV-2 or IAV. 526 gene sets were identified in the Illumina data set and 254 in the Nanopore dataset with an intersect of 77 shared between them (A). The number (NR) of genes identified within the gene sets correlated between the Illumina and Nanopore datasets (B).

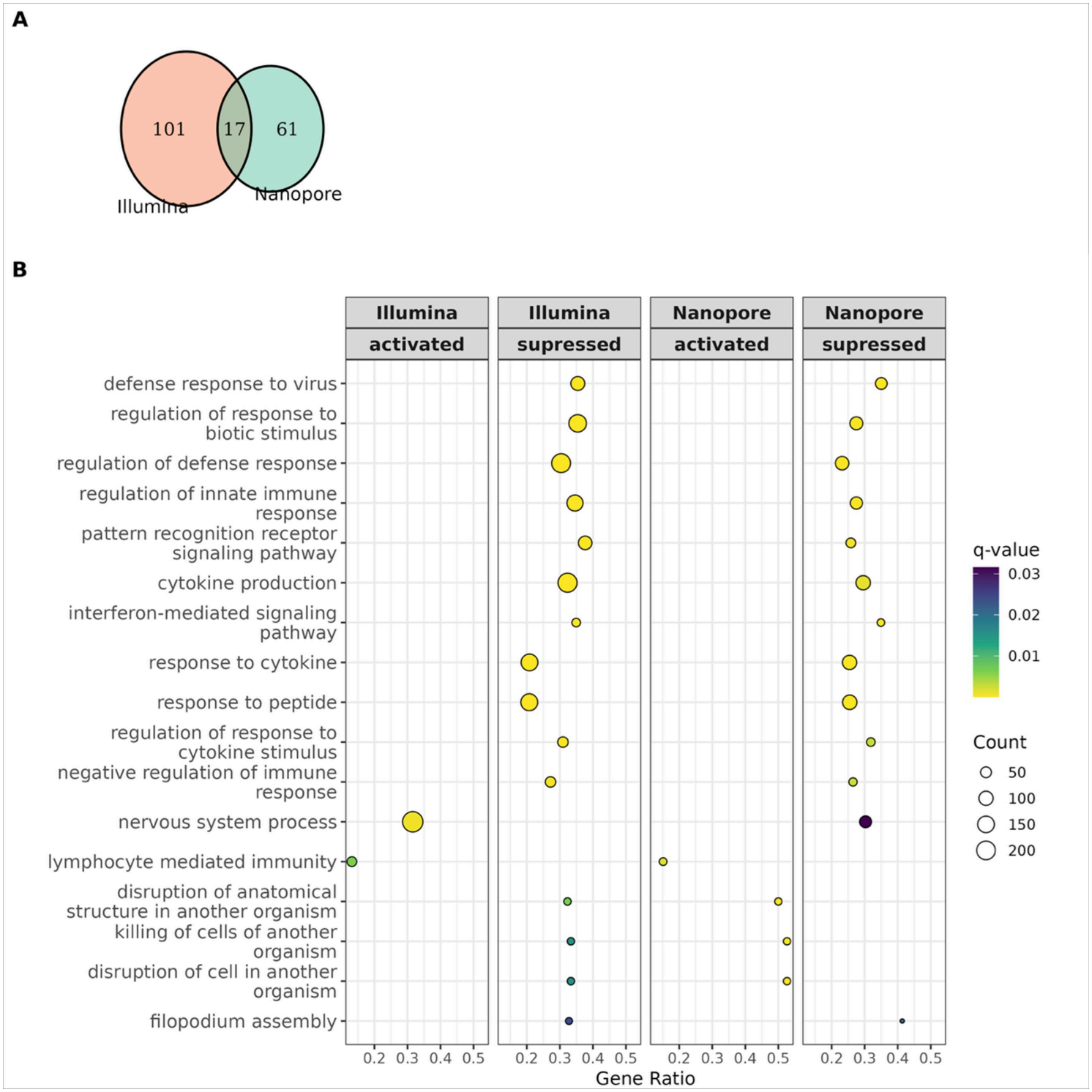

Figure 4

Gene set enrichment analysis was conducted in clusterProfiler on ranked log2FC data when comparing COVID-19 and IAV cohorts. Gene sets sizes were set between 15 and 800, and a p value of 0.05 was used with a BH correction, the simplify function was used to remove redundancy of enriched gene sets. Illumina and Nanopore datasets shared 17 gene sets that were identified as significant (A). The 17 shared significantly activated (activation in patients infected with SARS-CoV-2) and suppressed (activated in patients infected with IAV) are shown (B).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.docx
- SupplementaryTables.docx