

Real-time reverse transcription–polymerase chain (RT-PCR) for detection of SARS-CoV-2

All COVID tests were done by the TrueNat real-time reverse transcription–polymerase chain reaction (RT-PCR) test. Oropharyngeal or nasopharyngeal swabs were collected using a standard nylon-flocked swab and inserted into the viral transport medium (VTM) provided by the same company (Molbio Diagnostics Pvt. Ltd., Goa, India). Samples were transported immediately to the laboratory maintaining proper temperature and processed as per the manufacturer's guidelines (TrueNat Beta CoV chip-based real-time PCR test for betacoronavirus; Molbio Diagnostics Pvt. Ltd.). The target sequence for this assay is E gene of the Sarbeco virus and human RNase P (internal positive control). Confirmatory genes were RdRP gene and ORF1A gene.

Analysis of the relative efficacy of Mw in prevention of COVID-19

Table S1: Statistical Method for calculation of efficacy of Mw

	Infected	Non-infected	Total
Mw + ChAdOx1 group	a	b	n1
ChAdOx1 group	c	d	n2

Attack rate in ChAdOx1 group (ARU) = $c/n2$

Attack rate in Mw + ChAdOx1 group (ARV) = $a/n1$

Incidence Risk Ratio (IRR)/ Relative Risk = ARV/ARU

Absolute Risk Reduction (ARR) = $ARU - ARV$

Number needed to treat (NNT) = $1/ARR$

Vaccine Efficacy (%) = $((ARU - ARV)/ARU) * 100$

95% CI Vaccine efficacy (%) = $1 - IRR\ CI$

Table S2. SARS-CoV-2 infections in Mw + ChAdOx1 and ChAdOx1 group

	Infected	Non-infected	Total
Mw + ChAdOx1 group	2	48	50
ChAdOx1 group	36	165	201
Total	38	213	251

Table S3. Incidence rate of SARS-CoV-2 infections.

	Incidence Rate
Mw + ChAdOx1 group	6.7/10,000 person days
ChAdOx1 group	29.85/10,000 person days

Table S4. Incidence Rate Ratio of SARS-CoV-2 infections and Mw efficacy.

Parameter	Estimate	p-value	95% CI
------------------	-----------------	----------------	---------------

Incidence Rate Ratio (IRR)	0.22	0.02	0.05 to 0.9
Attack rate in ChAdOx1 group (ARU)	0.18		
Attack rate in Mw + ChAdOx1 group (ARV)	0.04		
Mw efficacy (%)	77.78%		
Absolute Risk Reduction (ARR)	0.14		0.06 to 0.21
Number Needed to Treat (NNT)	7.2		4.7 to 15.8

Table S5. Multivariate analysis of outcomes

Variable	Hazard Ratio (95% CI)	<i>p</i> -value
Mw	0.46 (0.11-1.93)	0.008
ChAdOx1 both dose		

Methods for RNA Sequencing and Differential Gene Expression Analysis (DGE)

RNA isolation using Trizol method and poly(A) RNA selection

Peripheral blood mononuclear cells (PBMC) were isolated from whole blood samples by density gradient centrifugations using HiSep™ LSM 1077 media (Himedia, Mumbai, India). 5 x 10⁶ PBMCs were used for RNA isolation using the Trizol method and followed by DNase treatment for RNA purification. Purified RNA was quantified using a Qubit 4.0 fluorometer. Take 5µg of total RNA used for the poly(A) RNA selection using NEB NEXT oligo d(T)25 beads (NEB, MA, USA). Poly(A) RNA was quantified and integrity assessed using Qubit 4.0 fluorometer (Invitrogen, Waltham, Massachusetts, United States) according to manufacturer instructions.

cDNA library preparation and whole transcriptome sequencing using MinION 2.0-Oxford Nanopore Technologies (ONT)

For Direct cDNA Native Barcoding Sequencing (SQK-DCS109 with EXP-NBD104, Oxford Nanopore Technologies, Oxford, UK), 100ng of poly(A) RNA was used for the library preparation. Using H minus Reverse Transcriptase (Invitrogen, Waltham, Massachusetts, United States), complementary DNA (cDNA) was synthesised, followed by RNA degradation and second-strand synthesis of cDNA. Double-stranded cDNA was used for the end preparation, followed by native barcoding and adaptor ligation (all steps were followed, according to the manufacturer's instructions). Ligated cDNA was loaded on the flow cell (R9) in MinION Libraries were sequenced specifying 72 hours on the ONT MinION using R9.4.1 flow cells and MinKNOW (v21.06.10, Microsoft Windows OS based) to generate FAST5 files. FAST5 files were base-called with CPU based Guppy basecaller (v.5.0.11) (ONT) to create summary text files and FASTQ files of the reads for further downstream analysis.

Sample pre-processing and Quality assessment

Demultiplexing of the pooled samples and adapter removal was carried out using inbuilt algorithm of Minknow. Linux Long Time Support (v.20.04) Operating System was used for all the analysis. A comprehensive report of the sequencing was generated by NanoComp (<https://github.com/NanoComp/h5utils>) tool and sequencing quality assessment was done by FastQc tool (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Sample reads were subjected to a minimum phred quality score of 9 and reads with lower quality were filtered out using NanoFilt (<https://github.com/wdecoster/nanofilt>) tool. Some initial reads are usually prone to low base-calling quality. Hence, 50bp of each initial reads from every sample were

filtered out for quality maintenance using NanoFilt tool only. All the samples were subjected to quality assessment before and after quality filtering.

Table S6. Sequencing and mapping metrics

Average base calling phred quality score (q) was found to be more than 10 for each sample (range 8-30). Read base quality ≥ 9 was considered for DGE analysis. N50 of read length for sequenced pooled samples was more than 900 bp (range 900bp-1.4kb). Average read-length was more than 800bp for each sample (range 200bp-4.6kb). Numbers of reads were more than 500,000 for each sample. Total number of passed (>7 q score) base sequenced for each sample was found to be more than 500 million with one sample going as much as 1 billion implying a coverage of 16X to 35X. More than 60% mapping coverage was found for each sample.

Sample	Mean read length (bp)	Mean read quality (q-score)	Read number (N)	Total base number (N)	Read length N50 (bp)
Cs1	873.1	10.9	825,250	720,547,331	1,209
Cs2	808.8	11.1	904,886	731,834,535	962
Cs3	955.5	11	526,857	503,435,179	1,251
Cs4	1045.3	11	1,066,901	1,115,184,646	1,393
Mw_Cs1	941.5	10.8	766,585	721,729,757	1,196
Mw_Cs2	857.8	10.7	1,053,022	903,322,821	1,134
Mw_Cs3	881.6	10.7	580,930	512,171,601	1,109
Mw_Cs4	817.7	11.1	846,667	692,33,369	956

Differential Gene Expression analysis

DGE was done using “pipeline-transcriptome-de” (<https://github.com/nanoporetech/pipeline-transcriptome-de>) pipeline. This pipeline from nanopore tech uses snakemake, minimap2, salmon, edgeR, DEXSeq and stageR to automate DGE workflows on long read data. Pipeline was set to make only reads aligned to minimum 3 samples to be considered for analysis. A

separate conda (<https://docs.conda.io/en/latest/#>) environment was created on Linux OS to host this pipeline. Quantification and DGE was done by R language based (<https://www.r-project.org/>) tool edgeR (<https://bioconductor.org/packages/release/bioc/html/edgeR.html>) employing gene-wise negative binomial regression model and normalisation factor (Transcript Mean of M-value) for each sequence library. Differentially expressed genes (DEGs) with log2Fold Change ($\log_2\text{Fc}$) ≥ 0.5 , ≤ -0.5 and associated p-value < 0.05 were selected as significant for further analysis. Annotation of DEGs was fetched from ENSEMBL database (<https://asia.ensembl.org/index.html>) using R based biomaRt package (<https://bioconductor.org/packages/release/bioc/html/biomaRt.html>). All the file compilation was ultimately done using Microsoft Excel and Linux Libre Office calc.

Hierarchical clustering analyses was performed between all four samples of each group to generate Heatmap from normalised log2 counts per million (log2CPM) expression values using R/Shiny based rnaseqDRaMA (https://hssgenomics.shinyapps.io/RNAseq_DRaMA/) package. Comparative gene expression boxplot on the basis of log2CPM expression values between Mw + ChAdOx1 and ChAdOx1 group were generated from START (<https://kcv.shinyapps.io/START/>) tool. Volcano plot involving all DEGs were created using R based ggplot2 (<https://ggplot2.tidyverse.org/>) package.

Table S7: Statistics of ANK-ADCC related pathway DEGs

Gene	log2CPM_ChAdOx1 group	log2CPM_Mw+ChAdOx1 group	p-value	FDR
KLRC1	7.443592115	6.494524546	0.02851	0.34456
KLRC2	3.430179664	5.458179183	0.00189	0.04876
EAT2	4.534002259	3.441975526	0.03610	0.59835
KIT	5.556271995	3.962163365	0.00397	0.04856
KLRC4	3.472858505	5.392449484	0.01103	0.74389
ARID5B	3.053976522	5.776082056	3.06E-04	0.00721

B3GAT1	3.302880956	5.530822266	5.22E-04	0.00104
BCL11B	6.680114141	7.826602734	0.00233	0.04731
ZBTB16	6.59092053	3.361821493	6.57E-05	0.00093
FCGR2A	6.057301045	7.89179857	0.00077	0.00967
FCGR1A	4.020489378	7.088221547	0.00041	0.07843
FCER1G	8.368291497	6.762609915	0.00150	0.05934
CD247	6.553818095	7.858964687	0.00986	0.04296
FCGR3A	6.25150747	7.524711647	0.00682	0.08426
FCGR2C	7.356255432	8.669027505	0.00885	0.09854

Table S8: Statistics of innate immune inflammatory related pathway DEGs

Gene	log2CPM_ChAdOx1 group	log2CPM_Mw+ChAdOx1 group	p-value	FDR
JAK2	7.379807441	6.514382609	0.02807	0.56874
MAPK3	7.008513364	5.187758026	0.00081	0.00957
MAPK8	7.438225964	5.595844271	0.00966	0.04437
NFKB1	8.113694347	6.06757918	0.00869	0.03842
IFNG	6.47017729	3.071481171	0.00086	0.00544
IL1B	8.146030118	5.835769987	0.00417	0.06395
IL6	6.729883946	4.185704642	0.00565	0.03928
TLR2	7.302998258	5.892284383	0.00042	0.00245
IL18	6.284423115	2.749841978	9.26E-05	0.00083
NLRP3	7.400315421	5.692328804	0.00338	0.04964
STAT3	7.158620649	6.137838244	0.00758	0.05376
MYD88	7.001107125	6.012094282	0.00862	0.04165
TLR7	7.32518637	4.621653945	0.00999	0.04453

TNF	6.94930147	5.229051455	0.00312	0.02645
-----	------------	-------------	---------	---------

Gene Ontology (GO) Pathway Analysis

Pathway enrichment analysis was done for significant DEGs using R based clusterProfiler (<https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>, v.4.2.1) tool employing Gene Ontology database. Pathways related to ADCC and innate immune inflammatory pathways were considered for focused analysis. An adjusted *p*-value threshold of ≤ 0.05 was considered for this study.

Figure S1:

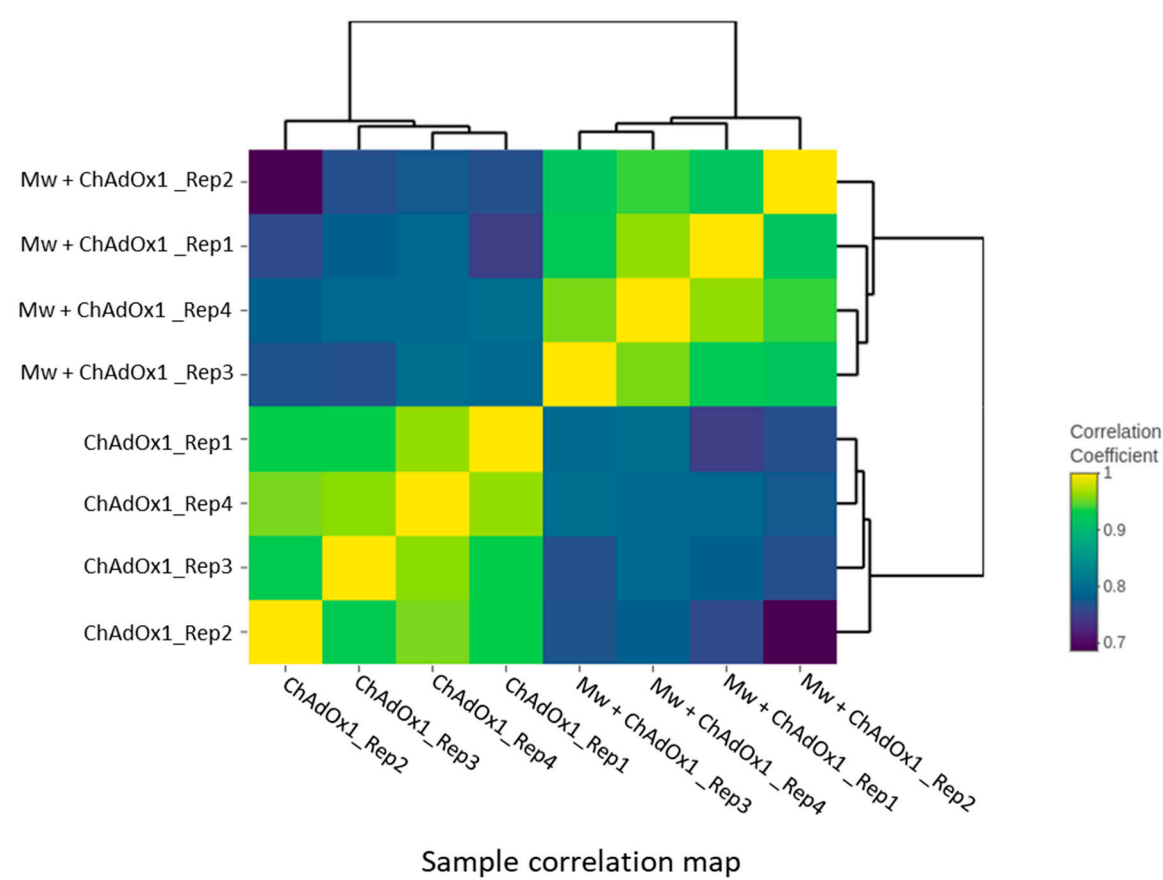


Figure S1: Correlation matrix of 8 RNA-Seq libraries. Pairwise Pearson correlation coefficients (PCC) were calculated for comparison among transcriptomes of Mw + ChAdOx1 and ChAdOx1 group samples. Samples were hierarchically clustered with the Euclidean distance method. The colour scale indicates the degree of correlation. The sample correlation matrix was generated using R/Shiny based rnaseqDRaMA package.

Figure S2:

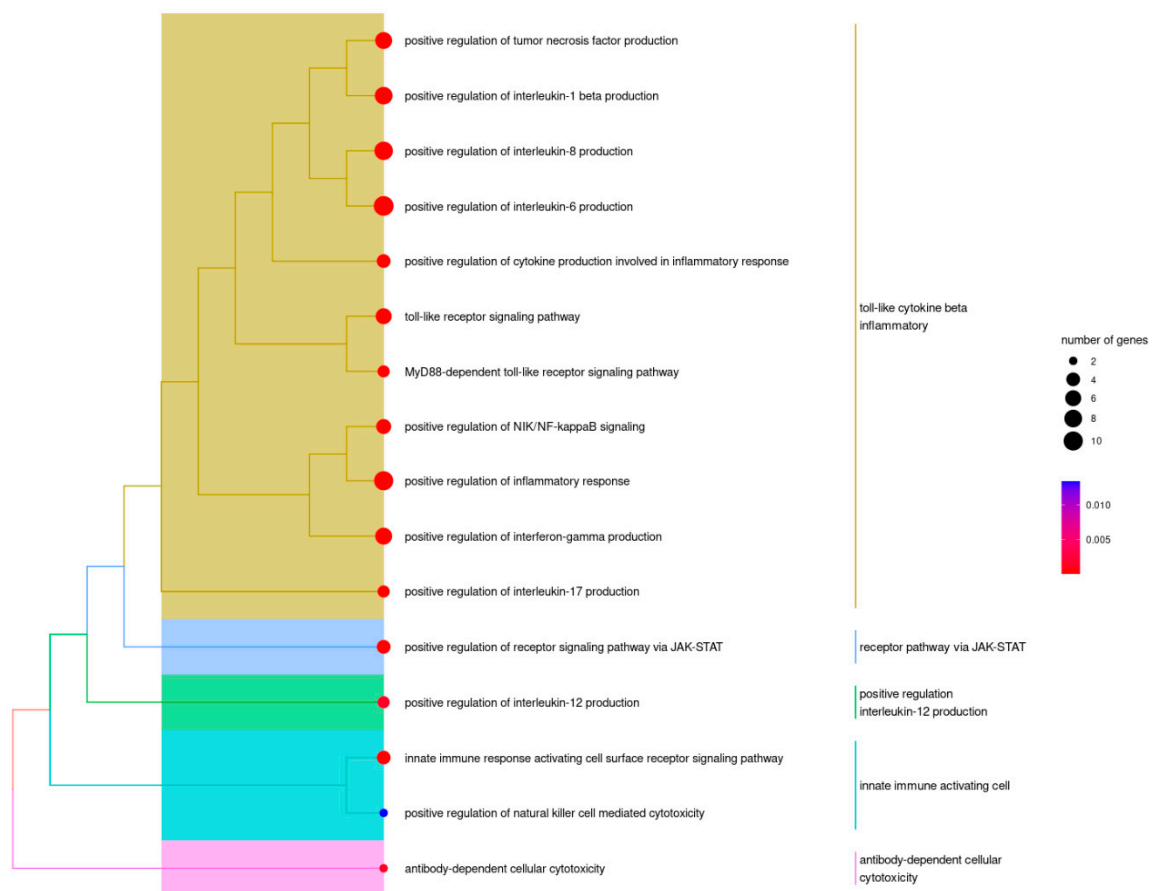


Figure S2: Tree plot depicting clustering of enriched GO terms based on the ‘average’ hierarchical clustering method. Subtrees are labelled using high-frequency GO words. The colour gradient scale indicates the adjusted p-value of enriched GO terms. (Fisher’s exact test using treeplot function in R based clusterProfiler package, multiple test correction by Benjamini-Hochberg method, adj. p-value <0.05).

Table S9: Enriched pathways related to ANK-ADCC pathways

GO ID	Description	p-value	p.adjusted value
-------	-------------	---------	------------------

GO:0002220	innate immune response activating cell surface receptor signaling pathway	1.47E-06	1.96E-05
GO:0001788	antibody-dependent cellular cytotoxicity	0.000223195	0.001268656
GO:0045954	positive regulation of natural killer cell mediated cytotoxicity	0.004013321	0.013457959

Table S10: Enriched pathways related to innate immune inflammatory pathways

GO ID	Description	p-value	p.adjusted value
GO:0032755	positive regulation of interleukin-6 production	1.50E-15	1.04E-12
GO:0032757	positive regulation of interleukin-8 production	7.70E-14	1.45E-11
GO:0050729	positive regulation of inflammatory response	2.83E-12	2.62E-10
GO:0032731	positive regulation of interleukin-1 beta production	1.67E-11	1.15E-09
GO:0032729	positive regulation of interferon-gamma production	5.43E-09	1.56E-07
GO:1903557	positive regulation of tumor necrosis factor superfamily cytokine production	1.04E-08	2.62E-07
GO:0002224	positive regulation of toll-like receptor signaling pathway	6.41E-07	9.09E-06
GO:1901224	positive regulation of NIK/NF-kappaB signaling	8.28E-07	1.16E-05
GO:1900017	positive regulation of cytokine production involved in inflammatory response	1.62E-06	2.09E-05
GO:0046427	positive regulation of receptor signaling pathway via JAK-STAT	4.50E-06	4.75E-05
GO:0032740	positive regulation of interleukin-17 production	1.58E-05	0.000131
GO:0002755	MyD88-dependent toll-like receptor signaling pathway	1.79E-05	0.000147
GO:0032735	positive regulation of interleukin-12 production	0.000279	0.001519

