



## OPEN ACCESS

## EDITED BY

Namal P. M. Liyanage,  
Wexner Medical Center, The Ohio State  
University, United States

## REVIEWED BY

Fabio Romero,  
Johns Hopkins Medicine, Johns Hopkins  
University, United States  
Patrick L. Collins,  
The Ohio State University, United States

## \*CORRESPONDENCE

Melinda A. Judge  
✉ melindaajudge@gmail.com

†These authors contributed  
equally to this article and share  
first authorship

‡These authors share senior authorship

## SPECIALTY SECTION

This article was submitted to  
Microbes and Innate Immunity,  
a section of the journal  
Frontiers in Cellular and  
Infection Microbiology

RECEIVED 20 October 2022

ACCEPTED 14 March 2023

PUBLISHED 03 April 2023

## CITATION

Parker E, Judge MA, Pastor L,  
Fuente-Soro L, Jairoce C, Carter KW,  
Anderson D, Mandomando I, Clifford HD,  
Naniche D and Le Souëf PN (2023) Gene  
dysregulation in acute HIV-1 infection –  
early transcriptomic analysis reveals the  
crucial biological functions affected.  
*Front. Cell. Infect. Microbiol.* 13:1074847.  
doi: 10.3389/fcimb.2023.1074847

## COPYRIGHT

© 2023 Parker, Judge, Pastor, Fuente-Soro,  
Jairoce, Carter, Anderson, Mandomando,  
Clifford, Naniche and Le Souëf. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Gene dysregulation in acute HIV-1 infection – early transcriptomic analysis reveals the crucial biological functions affected

Erica Parker<sup>1†</sup>, Melinda A. Judge<sup>1\*†</sup>, Lucia Pastor<sup>2,3,4</sup>,  
Laura Fuente-Soro<sup>2</sup>, Chenjerai Jairoce<sup>4</sup>, Kim W. Carter<sup>5</sup>,  
Denise Anderson<sup>5</sup>, Inácio Mandomando<sup>4</sup>, Holly D. Clifford<sup>5</sup>,  
Denise Naniche<sup>2,4‡</sup> and Peter Neils Le Souëf<sup>1,5‡</sup>

<sup>1</sup>School of Medicine, University of Western Australia, Crawley, WA, Australia, <sup>2</sup>ISGlobal, Barcelona Institute for Global Health, Hospital Clinic–Universitat de Barcelona, Barcelona, Spain, <sup>3</sup>AIDS Research Institute–IrsiCaixa, Institut Germans Trias i Pujol (IGTP), Hospital Germans Trias i Pujol, Universitat Autònoma de Barcelona, Badalona, Spain, <sup>4</sup>Centro de Investigação em Saúde da Manhiça (CISM), Maputo, Mozambique, <sup>5</sup>Telethon Kids Institute, Perth, WA, Australia

**Introduction:** Transcriptomic analyses from early human immunodeficiency virus (HIV) infection have the potential to reveal how HIV causes widespread and lasting damage to biological functions, especially in the immune system. Previous studies have been limited by difficulties in obtaining early specimens.

**Methods:** A hospital symptom-based screening approach was applied in a rural Mozambican setting to enrol patients with suspected acute HIV infection (Fiebig stage I–IV). Blood samples were collected from all those recruited, so that acute cases and contemporaneously recruited, uninfected controls were included. PBMC were isolated and sequenced using RNA-seq. Sample cellular composition was estimated from gene expression data. Differential gene expression analysis was completed, and correlations were determined between viral load and differential gene expression. Biological implications were examined using Cytoscape, gene set enrichment analysis, and enrichment mapping.

**Results:** Twenty-nine HIV infected subjects one month from presentation and 46 uninfected controls were included in this study. Subjects with acute HIV infection demonstrated profound gene dysregulation, with 6131 (almost 13% of the genome mapped in this study) significantly differentially expressed. Viral load was correlated with 1.6% of dysregulated genes, in particular, highly upregulated genes involved in key cell cycle functions, were correlated with viremia. The most profoundly upregulated biological functions related to cell cycle regulation, in particular, CDCA7 may drive aberrant cell division, promoted by overexpressed E2F family proteins. Also upregulated were DNA repair and replication, microtubule and spindle organization, and immune activation and response. The interferome of acute HIV was characterized by broad activation of interferon-stimulated genes with antiviral functions, most notably IFI27 and OTOF. BCL2 downregulation alongside upregulation of several apoptotic trigger genes and downstream effectors may contribute to cycle arrest and apoptosis. Transmembrane protein 155 (TMEM155)

was consistently highly overexpressed during acute infection, with roles hitherto unknown.

**Discussion:** Our study contributes to a better understanding of the mechanisms of early HIV-induced immune damage. These findings have the potential to lead to new earlier interventions that improve outcomes.

#### KEYWORDS

HIV, gene expression profiling, sub-Saharan Africa (SSA), Mozambique, acute HIV infection, host microbial interactions, innate immunity

## Introduction

Acute HIV infection encompasses the period after HIV transmission and prior to seroconversion, typically 3-12 weeks (Fiebig et al., 2003), following an initial 8-10 day undetectable eclipse period (Konrad et al., 2017). The rate and severity of acute infection may be predictive of disease progression and long-term sequelae, reflecting the extent of immune damage and hyperinflammatory state during this time (Lavreys et al., 2006). Furthermore, the early failure of immune responses to curb the viral threat and the latent viral reservoirs established during the acute stage set the challenging environment in which researchers seek a therapeutic vaccine and cure (Johnston and Fauci, 2008; Davenport et al., 2019).

Enhanced understanding of the pathogenesis of acute HIV infection may offer insights toward personalized therapies, vaccine development or curative treatment (Chun et al., 2015). Additionally, diagnostic testing that enables earlier detection may reduce onward transmission during this time of heightened risk due to peak viremia (Pilcher et al., 2004; Wawer et al., 2005; Powers et al., 2011). Transcriptomic analyses can pinpoint specific molecular mechanisms and pathways altered by HIV. We recently reviewed research of gene expression during HIV infection, and this revealed that studies to date have provided fragmentary insight into the host response to HIV, with conclusions limited by small sample sizes, and variation in study designs and data reporting (Judge et al., 2020). Case-control study designs comparing individuals with HIV to healthy controls were a common approach to describing major transcriptomic changes attributable to infection. Despite acknowledgement that gene expression changes are most profound during acute infection (Li et al., 2009; Judge et al., 2020), and the potential benefits of such research, relevant information is scarce, largely due to the well-described clinical and resource-intensive screening challenges of identifying individuals with acute infection (Cohen et al., 2010; Miller et al., 2010).

We identified only four published studies examining human host gene expression during acute HIV infection. Li et al. (2009) used microarray to describe stage-specific changes in gene expression in inguinal lymphoid tissue, with greatest dysregulation observed during acute infection (n=9 samples with

358 differentially expressed genes (DEG) identified). Zhang et al. used RNA-seq and compared peripheral blood mononuclear cells (PBMC) from 3 early HIV infection subjects to 3 uninfected controls and also identified many DEG (n=2049) (Zhang et al., 2018). Hyrcza et al. reported transcriptomic changes in CD4+ and CD8+ T cells from 5 subjects with early infection and 10 with longer-term infection, and found that some changes persisted during chronic infection (Hyrcza et al., 2007). Unsurprisingly, dysregulation centred on immune activation, innate immune defenses, adaptive immunity (Li et al., 2009; Zhang et al., 2018) and interferon-stimulated genes (Hyrcza et al., 2007). A fourth paper, by Kazer et al. provides a detailed description of gene modules in PBMC of 4 individuals with acute HIV, but the focus was more on temporal changes in module weighting over time, rather than describing differential gene expression against pre-infection samples (Kazer et al., 2020). Although these studies offered key insights, conclusions are hampered by variable definitions of early infection (with sampling up to several months post-transmission), small sample sizes, varied tissue types, use of superseded technologies, and participant bias in terms of sex, geographical location, and HIV subtypes.

This study aimed to use gold-standard RNA-seq transcriptomic profiling to describe the human host response to acute HIV infection, using PBMC from a large cohort of recently infected individuals as compared with HIV-negative controls. Importantly, the study aimed to recruit both men and women from Eastern and Southern Africa (ESA) where HIV-1 subtype C is predominant, reflective of the generalized epidemic. We hypothesized that examining a cohort early in HIV infection would reveal a consistent pattern of gene dysregulation among crucial biological functions, and highlight early “gateway” paths and processes leading to known HIV pathologies.

## Materials and methods

### Human ethics approval

This research was approved by ethical review boards at Barcelona Clinic Hospital (2011/6264), the Ministry of Health of Mozambique (461/CNBS/12) and the University of Western

Australia (2019/RA/4/1/6296). Written informed consent was obtained from all subjects.

## Sample collection

The present analysis is a sub-study of the Gastrointestinal biomarkers in Acute HIV-infected Mozambican Adults (GAMA) study. Procedures of recruitment for this cohort study have been described elsewhere (Pastor et al., 2017a; Pastor et al., 2017b). In brief, the cohort was enrolled in a semi-rural area of Southern Mozambique with an HIV prevalence of 39.9% (Gonzalez et al., 2012). Patients presenting with febrile illness or for voluntary HIV testing were invited to be screened for study eligibility using antibody-based fingerprick HIV rapid tests as per national diagnostic guidelines (Determine HIV 1/2 (Abbott Laboratories, Illinois) and Uni-Gold (Trinity Biotech Co., Ireland)). Those testing negative or indeterminate had HIV viral load measured *via* RT-PCR (Abbott RealTime HIV-1 assay, Illinois). Viral load positive, seronegative or indeterminate patients were diagnosed with acute HIV and invited for follow-up four weeks later. Time-matched HIV-uninfected controls were selected from patients who tested negative on rapid tests and viral load PCR, and also invited back four weeks later.

At this follow-up visit, clinical data and specimens for immunological and microbiological evaluation were collected. The Manhica Health Research Centre (CISM) is co-located with the Manhica District Hospital and contains a well-equipped clinical research laboratory (Saccoor et al., 2013). PBMC were isolated from peripheral blood on the day of collection, stored in RNAProtect (Qiagen, Germany) at -80°C, then shipped to Western Australia on dry ice.

## Fiebig staging

HIV-specific antibodies were analysed by Western blot assay using INNO-LIA HIV I/II Score (Innogenetics, Belgium). Staging of primary HIV infection was completed according to Fiebig and colleagues, 2003 (Fiebig et al., 2003). Date since infection was estimated by adding the mean time since infection for those in Fiebig stages I-IV (Fiebig et al., 2003) to the number of days between recruitment and sample collection.

## T-cell and plasma protein determinations

As previously described (Pastor et al., 2017b), CD4+ and CD8+ T cell counts were determined using CD3, CD8, CD4 and CD45 fluorochrome-labelled antibodies on fresh whole blood using TruCount tubes and FaCSCalibur flow cytometry (BD Biosciences, New Jersey). Cytokine levels in plasma were determined using enzyme-linked immunosorbent assay (Orgentec, Germany; Hycult Biotech, Pennsylvania; and Immunodiagnostik, Germany) and Luminex multianalyte profiling: Human Cytokine Magnetic 30-plex panel (Invitrogen, California), Bio-Plex Pro Human Th17

cytokine assay (Biorad, California), and Human Magnetic Luminex Screening Assay (R&D, Minnesota) (Pastor et al., 2017a).

## Subtyping

Previous work has demonstrated that circulating virus in Mozambique is almost exclusively HIV-1 Group M subtype C (Engelbrecht et al., 1998; Bellocchi et al., 2005; Parreira et al., 2006). One acute HIV case was the subject of a published case report (Velasco et al., 2015) and had sequencing completed to confirm subtype C.

## RNA extraction

RNA was extracted using a modified protocol for the RNeasy MinElute Clean-up kit (Qiagen, Germany). Briefly, PBMC pellets were resuspended in 1ml of TRIzol, shaken vigorously, 200µl of chloroform was added and shaken again. After centrifugation at 14000xg for 5 minutes at 4°C, the aqueous phase was added to an equal volume of 70% ethanol. After mixing by pipetting, samples were transferred to an RNeasy MinElute spin column and centrifuged at 800xg for 30 seconds. The columns were washed with 500µl of RPE buffer by centrifugation at 8000xg for 30 seconds, then the process was repeated using 80% ethanol. An additional 500µl of 80% ethanol was added and centrifuged at 8000xg for 2 minutes. Once the flow through was discarded, samples were again centrifuged at 8000xg for 5 minutes. To elute the RNA, 20µl of RNase-free water was pipetted directly onto the column membrane and stood for 1 minute prior to centrifugation at maximum speed for 2 minutes. Quality control was performed using nanodrop (Thermo Fisher Scientific, Massachusetts) and Bioanalyzer (Agilent, California); samples with a minimum concentration of 20ng/µl (total input 250ng) and an RNA integrity number  $\geq 8$  were sequenced.

## RNA-seq

Illumina 50 base-pair, single-end RNA-seq was completed, using HiSeq2000 at the Australian Genome Research Facility, with minimum 20 million reads/sample. Read quality was assessed using FastQC. Reads were aligned to human reference genome hg19 using HISAT. Alignment quality was assessed using SAMStat. Quantification and organization of read counts was performed using summarizeOverlaps (*GenomicAlignments* package) and voom (*limma* package), both from Bioconductor. Sequencing data was functionally validated against protein levels in plasma. Data was deposited in Gene Expression Omnibus (GSE199911).

## Statistical analysis

Sample cellular composition was estimated from gene expression data using CIBERSORT, an *in silico* gene expression

deconvolution tool. Estimates of CD4+ and CD8+ T cell counts from CIBERSORT were examined for consistency against flow cytometry counts (Spearman's rank order correlation). Cohort descriptive statistics were completed in R 3.4.2. Independent samples T-tests were used to compare parametric data, Mann-Whitney U-tests and Spearman's rank order correlations were used for non-parametric data, and Chi-square or Fisher's exact tests were used for categorical data. Comparison of acute HIV-infected versus uninfected control clinical data was performed using Spearman's rank order correlation, with statistical significance at  $p < 0.05$ .

Differential gene expression analysis was completed using TopHat and Cufflinks (Trapnell et al., 2012). Briefly, reads were aligned by TopHat (using the BowTie engine), and transcripts were assembled using Cufflinks. The latter involved analysis of immature transcriptomes and final transcriptome assembly with Cuffmerge; mapping of reads with Cuffdiff to quantify genes and transcripts and determine statistical significance of gene expression between samples; comparison with reference databases using Cuffcompare to determine presence of novel genes and transcripts; and management, visualization and integration of Cuffdiff data with CummeRbund. Data were exported to R for further analyses. Adjusted p values were calculated by Benjamini-Hochberg false discovery correction (5%). Genes with adjusted p values  $< 0.05$  were considered differentially expressed.

Differential gene expression was adjusted for participant age and sex, as well as RNA extraction year (a proxy for two individual extractors) to mitigate plausible batch effect, using an empirical Bayes framework. Viral load correlations were assessed using Spearman correlation with FDR adjustment (5%) using the psych package in R. Network Analyst generated a heatmap of raw expression counts. Cytoscape v3.8.1 and v3.9.1 were used to visualize known interactions between DEG, and between DEG and upstream regulators, using the STRING database (Szklarczyk et al., 2021) and CyTargetFinder with the ENCODE transcription factor database. Gene Set Enrichment Analysis (GSEA) was completed in GSEA 4.1.0, using a list of DEG ranked by the negative  $\text{Log}_{10}$  of adjusted p value divided by sign of the fold change (Ziemann, 2014; Dębski et al., 2016), with identifiers from Human\_ENSEMBL\_Gene\_ID\_MSigDB.v7.0.chip and gene sets from the Gene Ontology database (c5.go.bp.v7.2.symbols.gmt). Enrichment maps were created in Cytoscape v3.8.1 using GSEA results, with clusters generated using Community Cluster (GLay) Annotation Set in the AutoAnnotate plugin, with minor manual adjustments to improve language.

## Results

### Study population

Primary HIV cases ( $n=85$ ) and uninfected controls ( $n=58$ ) were recruited among adults aged 18–60 of both sexes, with ethnicities comprised from Indigenous African subgroups. From these subjects, 59 cases and 52 controls returned for a clinical follow-up visit one month later, during which PBMC samples were collected for gene expression analyses. RNA sequencing was completed

satisfactorily for 50 primary HIV samples and 46 HIV negative samples (Figure 1) with an average 23 million reads/sample, and 80% mapping accuracy to human reference genome hg19. To examine batch effect, three samples were sequenced on both RNA-seq runs, with highly similar results between batches (Supplementary Figure 1). Given our objective was to study the earliest stages of infection, further down-selection of cases was completed based on Fiebig staging at enrolment (Fiebig et al., 2003), with only the  $n=29$  most acute cases (Fiebig I-IV) included in further analyses.

At PBMC collection, participants enrolled as acute HIV cases had an estimated median time since infection of 8.0 weeks (range 5.9–10.6 weeks) and 83% of cases tested positive on HIV point-of-care rapid tests [which have median window periods of 19 and 32 days, respectively (Delaney et al., 2017)], while the remainder tested negative or indeterminate. While these rapid tests cannot confirm full seroconversion, they indicate initiation of an adaptive immune response among most cases. Median  $\text{Log}_{10}$  HIV viral load at PBMC collection was 5.03 (IQR 4.48–5.49) copies/mL; viral load at recruitment one month earlier had been higher, at median  $\text{Log}_{10}$  6.92 (IQR 6.15–7.52) copies/mL. CD4+ T cell count, and CD4+/CD8+ ratio were significantly lower in cases compared with controls (Table 1); there were no other significant differences between groups.

### Differential gene expression in acute HIV

Of 47,971 mapped genes, there were 6131 (12.8%) genes significantly differentially expressed during acute HIV-1 infection. Of these, 3307 genes (6.9%) were upregulated and 2824 (5.9%) were downregulated. Such high numbers of DEG have been previously described for other acute infections, such as influenza (Zhai et al., 2015), dengue (Popper et al., 2012), and tuberculosis (Wang et al., 2018). The top 50 DEG by significance are presented in Table 2 (full list in Supplementary Table 1). Notably, upregulated genes tended to have greater significance, with the first downregulated gene appearing at position 123 when ranked by adjusted p value. Upregulation was also more profound than downregulation, with 221 genes upregulated to  $\text{Log}_2$  fold change  $> 1.5$  (the highest at  $\text{Log}_2 4.84$ ) compared with 10 genes downregulated to  $\text{Log}_2$  fold

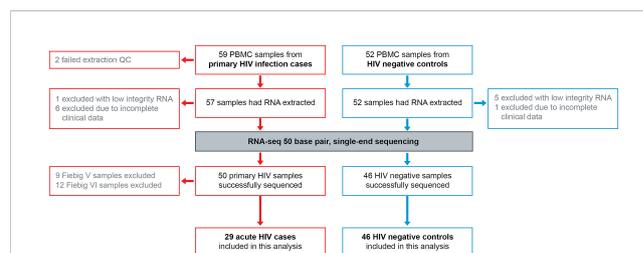


FIGURE 1

RNA-seq 50 base pair single-end sequencing was completed for PBMC from acute HIV cases and uninfected controls. PBMC, peripheral blood mononuclear cells; RNA, ribonucleic acid; QC, quality control.

TABLE 1 Characteristics of study population one month from presentation.

Characteristic	Acute HIV cases (n=29)	HIV negative controls (n=46)	P value
Sex, n (%) female	18 (62.1)	35 (76.1)	0.194
Age, median (IQR)	27.0 (19.0-35.0)	25.0 (20.7-35.5)	0.699
CD4 T cell count (flow cytometry), median (IQR)	599.0 (440.0-700.5)	995.5 (764.5-1156.0)	<0.001 ***
CD4:CD8 ratio (flow cytometry), median (IQR)	0.50 (0.27-0.69)	1.62 (1.24-2.36)	<0.001 ***
Fever in previous 24 hours (self-reported), n (%)	2 (6.9)	3 (6.5)	0.726
Diarrhoea in previous week (self-reported), n (%)	2 (6.9)	3 (6.5)	0.726
Malaria in previous month (self-reported), n (%)	5 (19.2)	9 (20.0)	0.308
Hepatitis B, n (%)	4 (14.3)	2 (4.3)	0.191
Syphilis, n (%)	2 (6.9)	3 (6.8)	0.990
<i>Entamoeba histolytica</i> , n (%)	1 (3.6)	3 (6.8)	0.826
<i>Giardia lamblia</i> , n (%)	2 (7.1)	5 (11.4)	0.083
<i>Cryptosporidium</i> spp., n (%)	0 (0.0)	0 (0.0)	NA
<i>Clostridium difficile</i> , n (%)	0 (0.0)	1 (2.3)	0.655
<i>Strongyloides stercoralis</i> , n (%)	1 (0.0)	1 (2.2)	0.424
Pregnancy, n (%)	1 (3.4)	0 (0.0)	0.259

\*\*\*p < 0.001.

change < -1.5 (the lowest at Log<sub>2</sub>-1.89). The top 50 DEG were highly consistent among cases and controls, respectively (Figure 2).

## Differential gene expression correlation with viral load

The correlation between gene expression and Log<sub>10</sub>HIV viral load was examined for the 29 acute HIV cases, with FDR adjustment for multiple comparisons ( $\alpha=0.05$ ). Among the top 50 DEG, 42 showed positive association with viral load (8 significantly correlated). None were negatively correlated. The FDR-adjusted correlation of viral load with the top ten upregulated and top five downregulated genes is depicted in Figure 3.

When all 6131 DEG were examined, 51 genes (0.85%) showed significant positive correlation with viral load, while 50 genes (0.82%) showed significant negative correlation (i.e. 1.67% of all DEG correlate with viral load, Supplementary Table 2). As mentioned, eight of the genes positively correlated with viral load were also in the Top 50 over-expressed DEG by fold change (Table 2). With the exception of SKA3, these eight genes have established protein-protein interactions (Figure 4), with protein activities implying that viral load is associated with numerous core functions across the cell cycle process. Specifically, these core functions include initiation of DNA replication [via MCM4-led increased DNA helicase unwinding (Bochman and Schwacha, 2009), and loading of DNA polymerase alpha onto chromatin through CDC45 and possibly GINS complex activity (Bochman and Schwacha, 2009)]; and positive regulation of progression

through G1, S and G2 phases through CCNA2 (Jiang et al., 2022). Notably, CCNA2 has been associated with aberrant cell cycle progression in several types of cancer, and has been posited as a potential therapeutic target that could be inhibited to repress cell proliferation (Ma, 2019; Yang et al., 2020). Heightened viral load may also be associated with sister chromatid cohesion during S phase via ESCO2 (Bender et al., 2020); increased microtubule activity [via KIF15-led motor activity and spindle assembly (Tanenbaum et al., 2009), and attachment to kinetochores during mitosis via SKA3 (Abad et al., 2016)], as well as functions in DNA repair and transcriptional regulation (CDCA7 and KIAA0101) (Gill et al., 2013; Zhang et al., 2021). CDCA7 is notable as a Myc-sensitive gene, through which apoptotic pathways and cellular proliferation can be regulated (Gill et al., 2013; Jimenez et al., 2018).

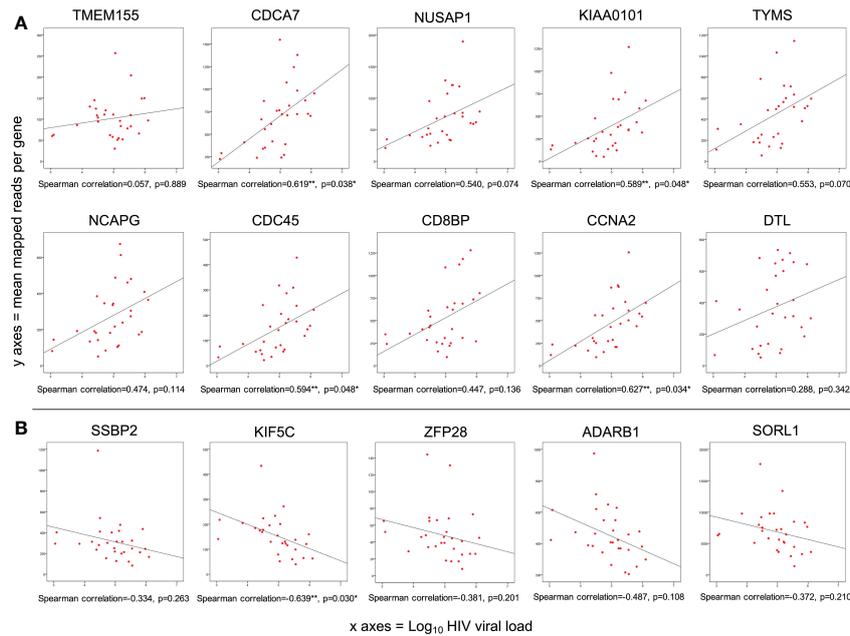
The genes that demonstrated the strongest negative correlation with viral load (Supplementary Table 2) also have key functional overlaps. For example, PTPN13, SFMBT1, ARM CX1, ZNF10, and ZBTB18 are all involved in repression of transcription (Huang et al., 2008; Nishitsuji et al., 2015; Liu et al., 2020; Heng et al., 2022; Xie et al., 2022). Downregulation of ZNF10 during high viremia is noteworthy as this protein has demonstrated functional capacity to limit HIV-1 gene expression *in vitro* (Nishitsuji et al., 2015). Several negatively correlated genes, including FGF9, ARM CX1, PMP22, and ZBTB18 are involved in nervous system functions, such as glial cell growth, axon regeneration, myelin development and brain tumour suppression (Li et al., 2013; Cartoni et al., 2016; Deng et al., 2018; Heng et al., 2022). High viral load may therefore be associated with inhibition of actors that regulate transcription and protect both the central and peripheral nervous systems.

TABLE 2 The top 50 genes differentially expressed during acute HIV infection, eight of which were positively correlated with viral load.

	Gene name	HGNC	Log <sub>2</sub> fold change	adjusted P value	viral load correlation coefficient	viral load correlation adj P value
1	transmembrane protein 155	TMEM155	3.267514	4.46E-35		
2	cell division cycle associated 7	CDCA7	2.47209	9.42E-34	0.619**	0.038*
3	nucleolar and spindle associated protein 1	NUSAP1	1.979959	1.08E-26		
4	KIAA0101	KIAA0101	2.784338	3.2E-26	0.589**	0.048*
5	thymidylate synthetase	TYMS	2.473498	3.2E-26		
6	non-SMC condensin I complex, subunit G	NCAPG	2.334234	3.2E-26		
7	cell division cycle 45	CDC45	2.836749	1.12E-25	0.594**	0.048*
8	CD8b molecule pseudogene	CD8BP	1.570053	1.12E-25		
9	cyclin A2	CCNA2	2.321098	3.88E-25	0.627**	0.034*
10	denticleless E3 ubiquitin protein ligase homolog (Drosophila)	DTL	2.886901	6.18E-25		
11	maternal embryonic leucine zipper kinase	MELK	2.489211	6.18E-25		
12	kinesin family member 11	KIF11	2.168095	6.18E-25		
13	CD8b molecule	CD8B	1.50107	3.77E-24		
14	BUB1 mitotic checkpoint serine/threonine kinase B	BUB1B	2.214816	6.3E-24		
15	myosin, light chain 6B, alkali, smooth muscle and non-muscle	MYL6B	1.562934	6.3E-24		
16	marker of proliferation Ki-67	MKI67	2.577557	6.65E-24		
17	anillin, actin binding protein	ANLN	2.351843	1.37E-23		
18	chemokine (C-C motif) ligand 5	CCL5	1.389217	1.54E-23		
19	ribonucleotide reductase M2	RRM2	2.648566	1.78E-23		
20	PDZ binding kinase	PBK	2.758444	1.99E-23		
21	ubiquitin-like with PHD and ring finger domains 1	UHRF1	1.923329	2.07E-23		
22	claspin	CLSPN	2.304573	2.77E-23		
23	cell division cycle 6	CDC6	2.411704	4.19E-23		
24	minichromosome maintenance complex component 10	MCM10	2.817904	5.57E-23		
25	cell division cycle associated 5	CDCA5	2.485974	1E-22		
26	topoisomerase (DNA) II alpha 170kDa	TOP2A	2.114054	1E-22		
27	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3H	APOBEC3H	1.665933	1E-22		
28	fatty acid binding protein 5 (psoriasis-associated)	FABP5	1.642369	1E-22		
29	kinesin family member 15	KIF15	2.38185	1.25E-22	0.663**	0.027*
30	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)	ASPM	2.263024	1.57E-22		
31	BRCA1 interacting protein C-terminal helicase 1	BRIP1	1.402205	1.97E-22		
32	spindle and kinetochore associated complex subunit 3	SKA3	2.784406	2.39E-22	0.612**	0.042*
33	GINS complex subunit 2 (Psf2 homolog)	GINS2	2.33612	4.25E-22		
34	kinesin family member 7	KIF7	2.730583	4.96E-22		
35	E2F transcription factor 8	E2F8	2.679807	4.96E-22		
36	exonuclease 1	EXO1	2.342272	4.96E-22		

(Continued)





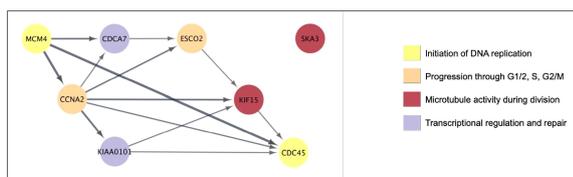
**FIGURE 3**  
Of genes differentially expressed during acute HIV infection, 1.6% were correlated with viral load. **(A)** Spearman correlation of  $\text{Log}_{10}$  HIV viral load with mean mapped reads of the top ten upregulated genes; **(B)** Spearman correlation of  $\text{Log}_{10}$  HIV viral load with mean mapped reads of the top five downregulated genes.  $p < 0.05$ ;  $**p < 0.01$ ;  $***p < 0.001$  (FDR adjusted).

GSEA was used to examine the biological relevance of changes that may be attributable to PBMC cell population differences between cases and controls. The 2530 genes ‘lost’ as a result of adjustment contributed to 671 upregulated gene sets and 301 downregulated gene sets. **Supplementary Figure 2** demonstrates the top 20 positively and negatively enriched processes, which are important pathways in the immune response, with many directly relevant to viral pathogenesis.

The 274 additional DEG resulting from adjustment for CD4+ T cell population contributed to 96 differentially enriched gene sets. The implication is that by not adjusting for CD4+ T cell count, these potentially important biological processes may have been missed. However, GSEA analysis of these genes indicated that all 23 upregulated gene sets, and all but 11 of the 73 downregulated gene sets, were captured by analyses without CD4+ adjustment. The

11 gene sets potentially missed pertain to downregulation of cellular organization processes (macromolecule localization, and protein complex subunit organization), intracellular signal transduction, and nucleobase compound metabolic processes.

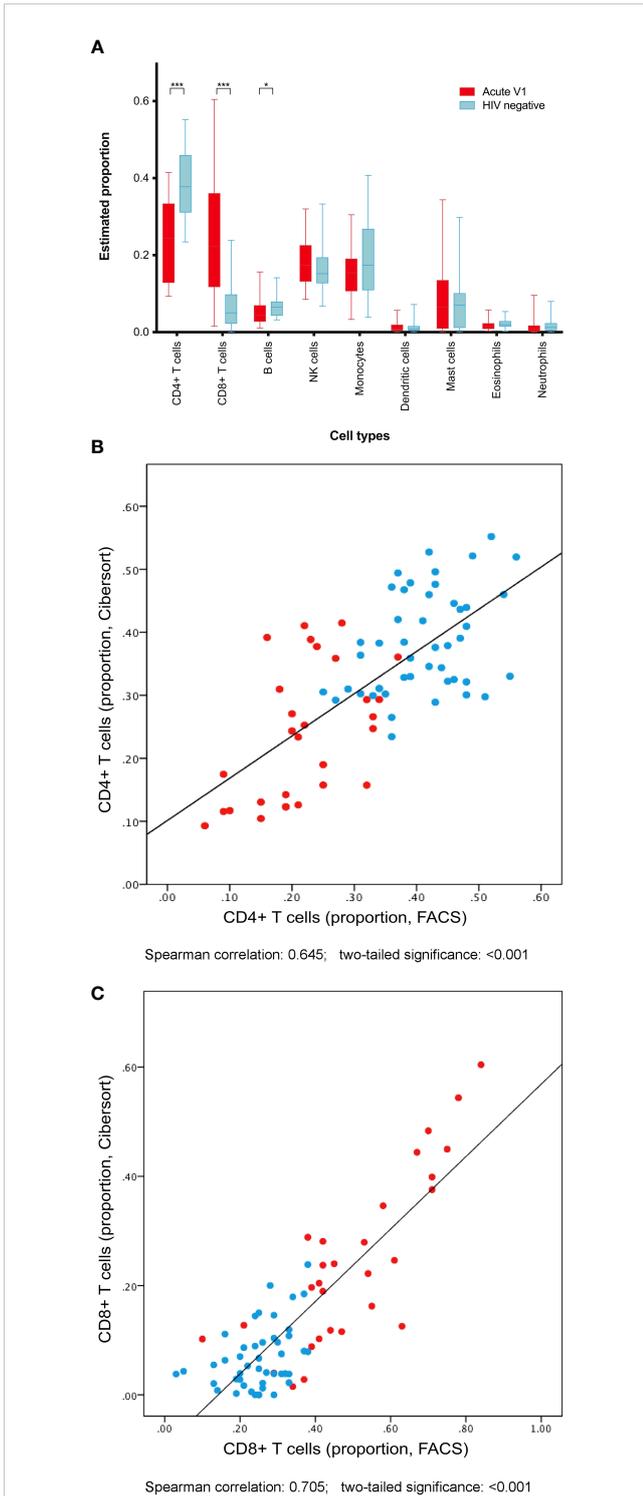
In summary, cell population frequencies were associated with both independent variables (viral load) and outcome measures (gene expression) in this analysis, making them an effect modifier rather than a confounder. Given that the primary objective of this analysis was to provide a global overview of host PBMC gene expression and pathogenic pathways during acute infection, inclusive of changes reflective of diversity in viral load and cell populations, all subsequent analyses of biological implication were not further adjusted for cell populations. This approach is in keeping with the previous mixed cell gene expression studies during early infection that have employed a case-control study design (Li et al., 2009; Zhang et al., 2018). We continued to adjust analyses for essential confounders (age, sex, and RNA extraction personnel). We acknowledge that studies with repeat measures, or comparing different timepoints along the course of HIV infection, would necessitate cell population adjustment.



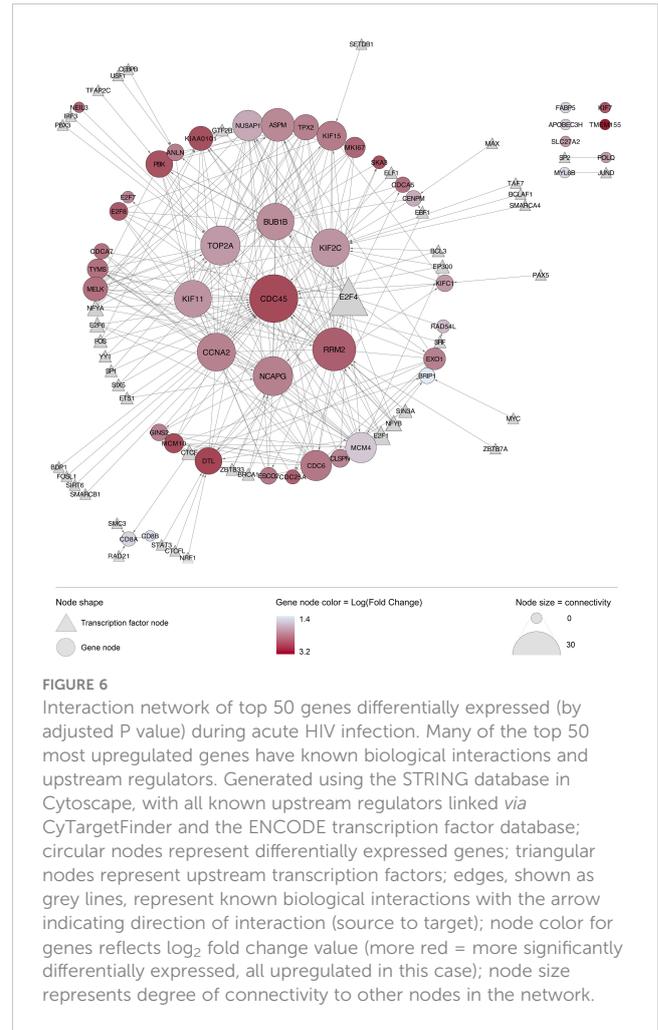
**FIGURE 4**  
The eight genes that are both highly overexpressed during acute HIV and correlated with viral load have core cell cycle functions. Generated using Cytoscape. Nodes were colored according to broad function as per the legend. Edges represent known protein-protein interactions in the STRING database, with the arrow indicating direction of interaction (source to target), and the weight indicating confidence of association (minimum 80%).

### Biological interactions and upstream regulators

The STRING database identified 35,759 established biological interactions between the 6,131 DEG. A closer look at the top 50 DEG revealed high connectivity and co-upregulation between genes (Figure 6). The most highly connected genes included those encoding cell division-related proteins (CDC45, BUB1B, KIF2C, NCAPG, KIF11), cell cycle regulation-related proteins (CCNA2), and DNA



**FIGURE 5**  
Cell populations among PBMC were predicted by interrogation of differential gene expression between acute HIV cases and uninfected controls. (A) Acute HIV was associated with predictable declines in CD4+ T cell and B cell populations, and expansion of CD8+ T cells. Both CD4+ T cell counts (B) and CD8+ T cell counts (C) estimated using gene expression data correlated well with cell counts via flow cytometry. \* $p<0.05$ . \*\*\* $p<0.001$ .

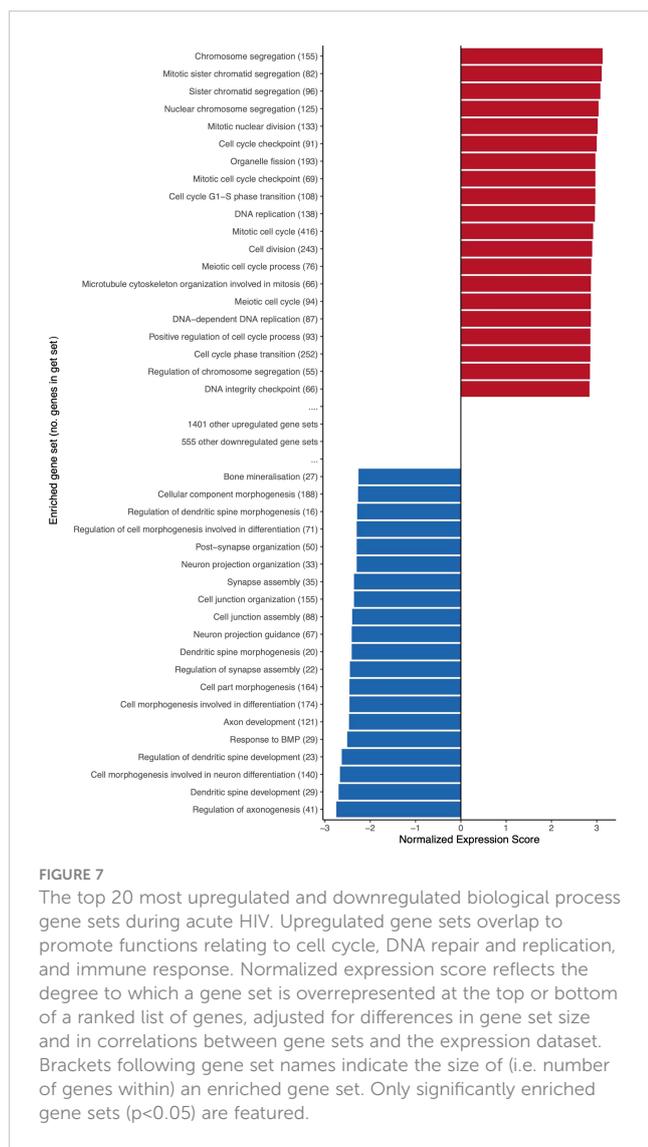


**FIGURE 6**  
Interaction network of top 50 genes differentially expressed (by adjusted P value) during acute HIV infection. Many of the top 50 most upregulated genes have known biological interactions and upstream regulators. Generated using the STRING database in Cytoscape, with all known upstream regulators linked via CyTargetFinder and the ENCODE transcription factor database; circular nodes represent differentially expressed genes; triangular nodes represent upstream transcription factors; edges, shown as grey lines, represent known biological interactions with the arrow indicating direction of interaction (source to target); node color for genes reflects log<sub>2</sub> fold change value (more red = more significantly differentially expressed, all upregulated in this case); node size represents degree of connectivity to other nodes in the network.

replication-related proteins (RRM2 and TOP2A). Transcription factor E2F4 was identified as an upstream regulator of many genes highly over-expressed in this cohort of individuals with acute HIV infection. E2F4 was not found to be overexpressed during acute HIV in our analysis, though functionally-similar E2F1-3, as well as E2F7-8 were overexpressed. Notably, the most significantly over-expressed gene, TMEM155, had no known interactions with other DEG, and no established upstream regulators.

## Dysregulated gene sets

GSEA was performed to determine over-represented Gene Ontology (GO) biological processes among DEG. 1421 gene sets were upregulated in acute HIV infection (405 reaching  $p<0.01$ ), compared with 575 biological processes downregulated (150 reaching  $p<0.01$ ). Normalized enrichment scores of the top 20 most up- and down-regulated gene sets are presented in Figure 7, and for all highly significant gene sets in Supplementary Table 4. The 20 most upregulated sets all related to cell cycle functions. The



top 20 downregulated categories relate to a variety of biological processes, with commonalities around growth and differentiation, and synaptic transmission.

## The 'interferome' of acute HIV infection

Within the GSEA results, ten pathways relating to interferon stimulation were prominent (beginning at position 187 by FDR-adjusted  $q$  value, normalized enrichment score 2.04). To explore this further, we collated from existing literature a list of interferon pathway genes, as well as interferon-stimulated genes (ISGs) with known antiviral activity (Liu et al., 2011; Schoggins and Rice, 2011; Schoggins et al., 2014; Kane et al., 2016; Yang and Li, 2020; Ullah et al., 2021; Ding et al., 2022). We mapped our differential gene expression results over this list to visualize the interferome pattern of acute HIV infection (Figure 8).

Of all interferon types, only IFN $\gamma$  was overexpressed within PBMC of individuals with acute HIV (noting that extracellular distribution is not captured in this type of study). Of the interferon

receptor heterodimers, only IFNLR1 (part of the IFN $\lambda$  receptor) was upregulated. Kinases TYK2, JAK1 and JAK2 were not differentially expressed, however signal transduction molecules STAT1, STAT2 and complex subunit IRF9 were all overexpressed compared with healthy controls. Many antiviral ISGs were upregulated in early stages of infection, with IFI27, OTOF, ISG15, MX1 and USP18 the most highly overexpressed. IFI27, in particular, was the single most overexpressed gene in individuals with HIV by fold change (+4.8). Only four of the ISGs were downregulated, namely PAK3, AKT3, CLEC4A and STING1. Most ISGs can be induced by multiple types of interferon, and the antiviral ISG analysis did not reveal patterns of upregulation specific to any type of interferon response. However, GSEA results indicated that Type 1 and Type 2 pathways were most recognisable among upregulated DEG overall.

## Enrichment mapping

GSEA results were used to inform an unbiased enrichment map of all up- and down-regulated GO biological process gene sets (Figure 9). The largest gene set clusters related to DNA replication and repair, regulation of cell cycle, microtubule and spindle organization, and immune activation and response. These functional clusters also have a high degree of gene overlap and similarity. Conversely, acute HIV downregulated functions related to regulation of growth and development, as well as numerous other functions in comparatively discrete pathways. These include bone mineralization, spinal cord and retina development, and regulation of systemic blood pressure, among others.

## Plasma protein validation

The relationship between gene expression and protein expression is complex, however any observed correlation validates the reliability of sequencing data, as gene expression changes are observed to be associated with functional protein changes. Of the 29 acute HIV cases and 46 HIV negative controls in this analysis,  $n=15$  and  $n=34$  respectively had corresponding protein data from the same blood draw. Quantification of  $n=49$  protein-based biomarkers in plasma was completed for this cohort as previously reported (Pastor et al., 2017a). Correlations between proteins in peripheral blood and their respective genes in circulating PBMC were significant in 20.4% of all relationships tested (Supplementary Table 5).

## Discussion

This study describes the global PBMC gene expression profile of humans acutely infected with HIV. By using the largest number of subjects for an early HIV transcriptomic study to date, we have been able to describe the massive gene expression dysregulation taking place in circulating PBMC, detailing the top DEG, frequency of correlation between DEG and viral load, up- and down-regulated gene sets, and enriched biological processes.

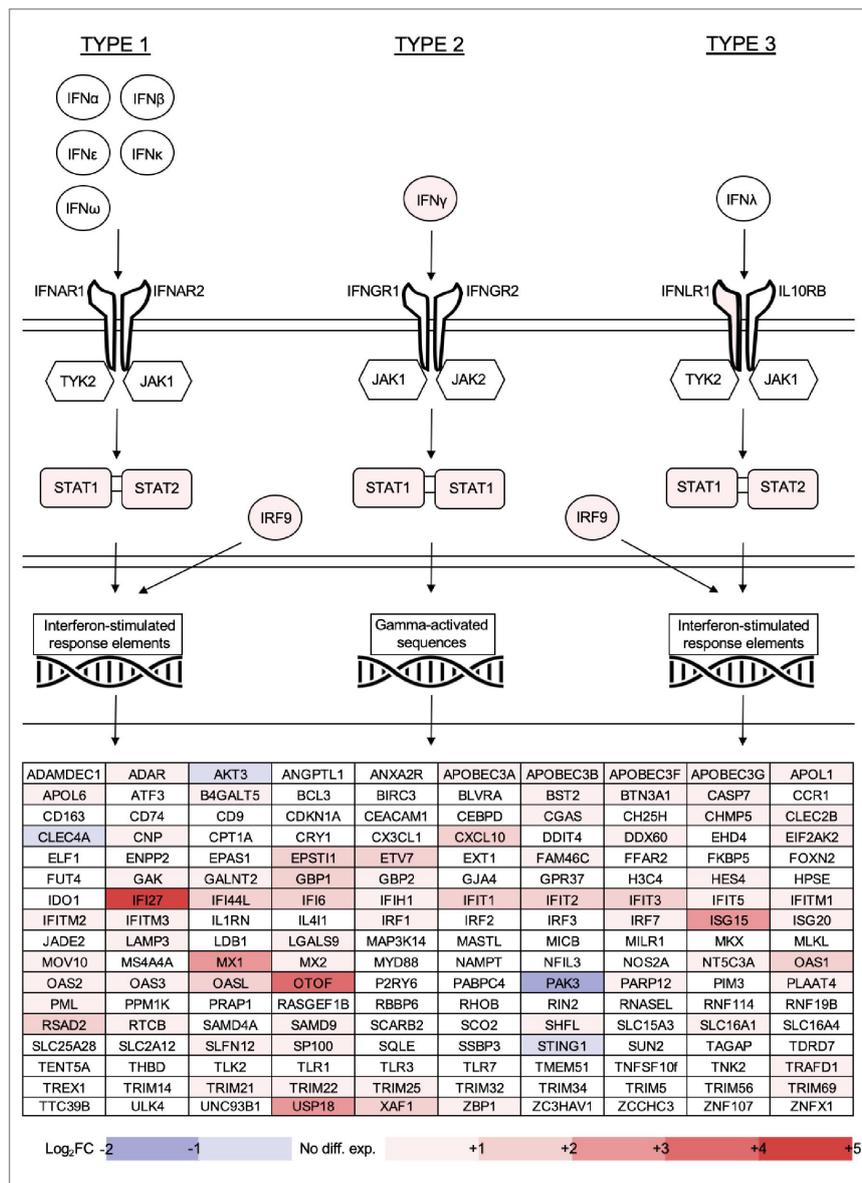
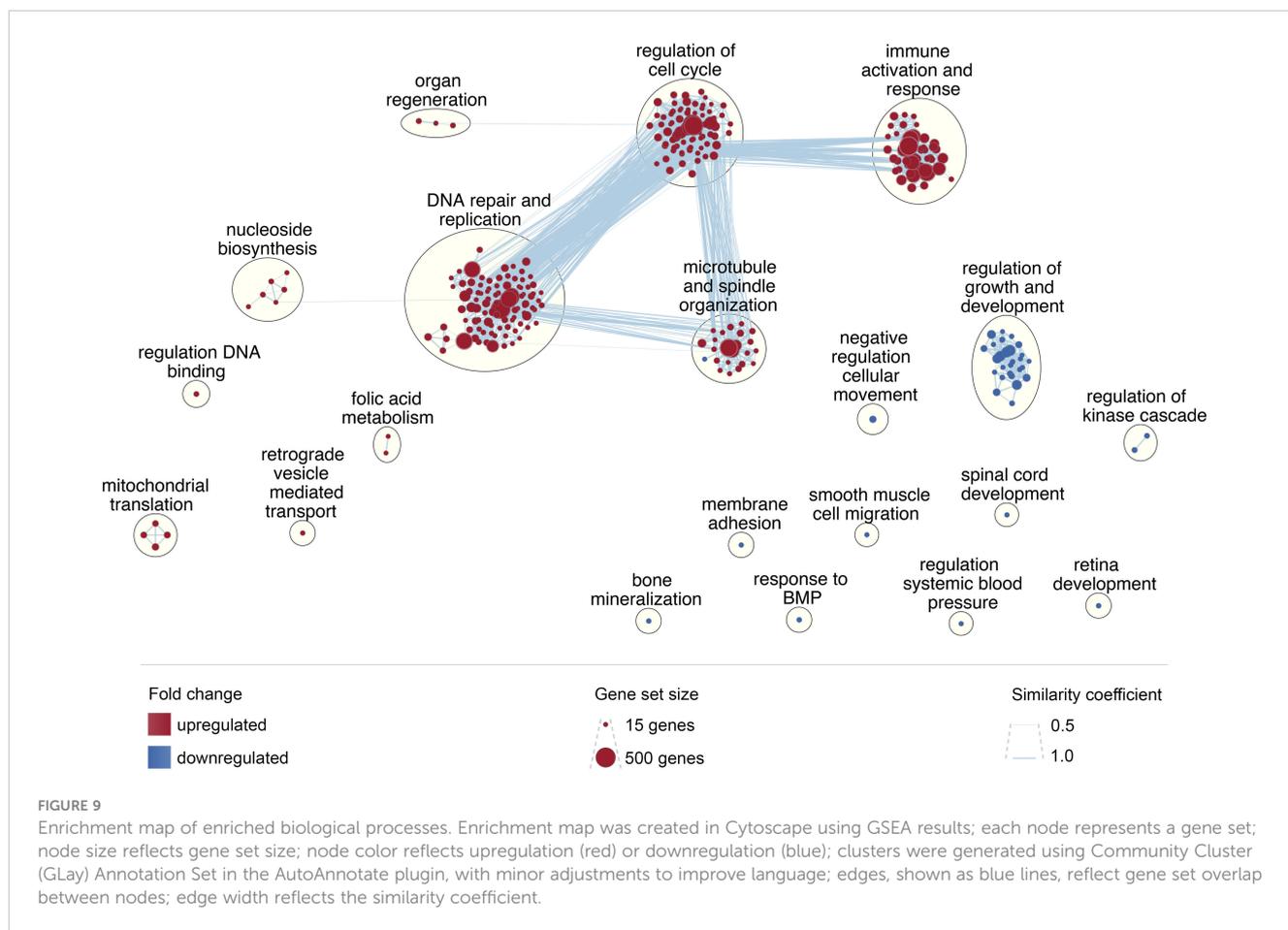


FIGURE 8 Gene expression changes for all types of interferon, key components of the Type 1, 2 and 3 interferon signalling pathways, as well as 170 other interferon-stimulated genes (ISGs) with established antiviral functions. ISGs listed in alphabetical order.

Previous studies of gene expression during early HIV expression have had six notable limitations. First, the non-specific symptoms of acute HIV make prompt diagnosis difficult and this is reflected in the time of sample collection, ranging from 1-5 months (Hyrca et al., 2007) or within 6 months (Zhang et al., 2018) post-infection, or within 4 months post-documented seroconversion (Li et al., 2009). Kazer et al. were able to collect earlier samples but this required the resource intensive approach of monitoring a prospective cohort of at-risk women (Kazer et al., 2020). Second, sample sizes were small (between 3 and 9 individuals) which hinders interpretation and reliability. Third, almost all subjects with HIV were male, although almost half of new infections worldwide occur in women (The Joint United Nations Programme on HIV/AIDS (UNAIDS), 2020). Fourth, the study

sites of North America (Hyrca et al., 2007; Li et al., 2009) and China (Zhang et al., 2018) are regions that represent minority HIV subtypes (Hemelaar et al., 2019; Gartner et al., 2020), with low HIV burden and mainly a concentrated HIV epidemic (in comparison with the generalized epidemic present across countries in ESA which bear two-thirds of the worldwide burden of HIV [The Joint United Nations Programme on HIV/AIDS (UNAIDS), 2020]). Fifth, each study used different sample types (T cells, lymph node tissue and PBMC) limiting comparability and meaningful conclusions. Finally, Li and Hyrcza utilized microarray technology, now largely superseded by RNA-seq (Zhao et al., 2014; Judge et al., 2020).

To address these methodological and external validity issues, this study used RNA-seq profiling to describe gene expression



changes in a large cohort of recently infected individuals, compared with contemporaneously recruited, HIV-negative controls who had presented with similar symptoms and were demographically indistinguishable. The study recruited men and women from ESA where HIV-1 subtype C is predominant. PBMC were identified as an ideal tissue for analysis, being readily collectable in resource-limited settings, and able to provide a comprehensive view of the host response, encompassing both target cells of HIV, and other cells directly and indirectly involved in the immune and inflammatory response to HIV. Although RNA-seq has the capacity to generate read lengths and depth of coverage greater than those applied in our experiment, sequencing using single-end 50 base pairs reads has been shown to produce high quality differential gene expression results, with negligible difference compared to up to 100 base paired reads (Chhangawala et al., 2015) and therefore was used for this study to maintain both quality and economy. Studies looking in detail at structural rearrangements to the genome, alternate splicing, or allele-specific expression may necessitate a greater depth of sequencing.

Changes in gene expression during acute HIV infection drive early immune system dysregulation, setting the stage for lifelong infection and progressive immunological decline. Our study identified that processes relating to cell cycle and cell division are the most profoundly upregulated functions during acute infection (Figure 9). By design, our study is likely to have captured urgent

attempts at cell regeneration triggered by CD4+ T cell loss (Mehandru et al., 2007), and rapid expansion of CD8+ T cells and immature transitional B cells, required for cytotoxic and early cellular immune responses. The quadruplicate evidence of E2F-sensitive CDCA7 ranking as the second most upregulated gene, the strong correlation of CDCA7 with viral load, the identification of an E2F-family protein as a common upstream regulator of many other overexpressed genes during acute infection, and the over-expression of E2F1-3 among early HIV cases, points toward a key role of E2F proteins in promoting CDCA7-led cell division during early HIV.

Lymphocyte survival during acute HIV infection is a very delicate balance of pro-apoptotic and pro-survival signals. Pathways related to cell cycle entry are significantly upregulated in PBMC, perhaps reflecting premature halting of cell cycle, or the posited 'tap and drain' hypothesis (Ho et al., 1995; Vidya Vijayan et al., 2017), as a mechanism for widespread cell death. The BCL2 family of proteins is known to protect cells from apoptosis (Schinzel et al., 2004; Kale et al., 2018). In this study, BCL2 and BCL2L2 genes were both downregulated among HIV cases, whereas pro-apoptotic genes BIM, BMAIP1 and BIK (involved in triggering the apoptosis cascade), as well as BAX and BAK (downstream apoptosis effector molecules) were all upregulated. In short, two of the key genes that protect lymphocytes from apoptosis are downregulated and five of those that promote apoptosis are upregulated. Gene set enrichment analysis systematically corroborated upregulation of pathways

related to cellular stress responses, suggesting that cell death in these early stages is predominantly pathological rather than physiological.

At this stage in the course of infection, the immune response is characterized by upregulated interferon-driven inflammatory and antiviral pathways, complement activation, humoral immunity (particularly upregulated antigen processing and presentation), and increased phagocytic activity (Supplementary Table 4). We described the interferome of acute HIV infection, which was characterized by a broad upregulation of antiviral ISGs, with IFI27, OTOF, ISG15, MX1 and USP18 the most overexpressed. Of all DEG, the most upregulated gene by fold change was interferon, alpha-inducible protein 27 (IFI27). High levels of expression of IFI27 have been reported previously, across different stages of infection (Liu et al., 2022; Palm et al., 2022), and the gene was even put forth as a 'potential therapeutic target for HIV infection' based on its downregulation after ART commencement (Huang et al., 2022). Among HIV cases in our analysis, IFI27 was not correlated with viral load. The interferome is of dual interest, as a potential cause of aberrant inflammation, and also in understanding the mechanisms by which HIV evades these interferon effectors and establishes long-term infection.

We did not expect to find significant differential gene expression relating to establishment of latent reservoirs. Although the reservoir is thought to exist from the earliest stages of infection, only small numbers (one per million) of memory and functional CD4+ T cells go on to be latently infected (Kuo and Lichterfeld, 2018), and at this stage of infection there is already a diminishing population of CD4+ T cells relative to total lymphocytes. Other cell types contributing to the reservoir would comprise an even smaller proportion within our samples (Kuo and Lichterfeld, 2018). Of four markers thought to be overexpressed in infected, transcriptionally-active T cells (Descours et al., 2017; Pardons et al., 2019), only MKI67 was upregulated; FCGR2A and TNFRSF9 were not differentially expressed and IL2RA was downregulated. However, LAG3, TIGIT and HAVCR2 [three of four immune checkpoint markers found to be enriched in cells with chromosomally-integrated HIV-1 proviruses (Chew et al., 2016; Fromentin et al., 2016; Pardons et al., 2019)] were over-expressed among HIV cases in this analysis. All three genes have a broad array of roles in immune responses, so their presence likely has alternate explanations. Gene expression analyses would be of benefit to the ongoing investigation of reservoirs, though single-cell approaches would be better equipped for this task than our global PMBC study.

Acute HIV infection, particularly subtype C, is associated with tremendously high viral load. Among all DEG, 0.8% positively correlated with viral load, including 8 of the top 50 most profoundly upregulated genes (16%) which have core cell cycle functions. This suggests that high viremia may drive key pathological processes seen during acute infection, particularly cell division disruption. Gene expression correlating with viral load has been previously reported, with strong associations between expression of interferon-stimulated genes and antiretroviral defense genes with viremia (Rotger et al., 2010). An almost equal amount (0.8%) of genes differentially expressed in PBMC during acute HIV infection were negatively correlated with viral load. The relatively balanced pattern of positive and negative correlation of viral load with gene

expression in PBMC contrasts previous work by Smith et al. (2009), whereby gene expression in mixed cell lymphatic tissue showed an overwhelmingly negative correlation with viral load (95% of associations) (Smith et al., 2010).

The genes consistently differentially expressed during acute HIV, but which are not correlated with viral load (even when conservative pre-FDR adjusted p value thresholds are used), are particularly interesting for exploration of biomarkers of recent infection. These candidates have the potential to remain stable during acute infection even as viral load fluctuates. Genes fitting this profile, warranting further investigation, include cell cycle transcription factor genes E2F7 and E2F8, cell cycle regulation gene DTL, DNA repair gene BRIP1, and antiretroviral gene APOBEC3H.

TMEM155 (also known as SMIM43) is a notable addition to this category. Uncorrelated with viral load, and not known to interact with any other over-expressed genes, TMEM155 was unexpectedly found atop the list of most highly over-expressed genes (in terms of significance, but also ranked 3rd in terms of fold change of overexpression). Figure 2 highlights that upregulation of this gene was consistent among all individuals with acute HIV. TMEM155 was identified as significantly differentially expressed (a 1.7-fold increase) in Li et al., 2009, but was not highlighted further as an important gene. The chromosomal region encoding TMEM155 was found to be one of many differentially methylated during HIV infection (Shiau, 2017). Outside of HIV, TMEM155 has been associated with hepatitis C liver cirrhosis (Ijaz et al., 2019) and brain tumours (Wang et al., 2010; Grupenmacher et al., 2013), and is possibly associated with essential tremor (Odgerel et al., 2019), basal cell nevus syndrome fibroblasts (Phatak et al., 2019), nystagmus (Toyono et al., 2015), cerebellar atrophy with motor neuropathies (Burns et al., 2018), and regulation of female fertility (Barragan et al., 2017).

The biological functions downregulated during acute HIV infection are remarkable for their resemblance to the longer-term consequences of HIV infection. Hypertension, retinopathy, osteopenia and muscular atrophy are well characterized sequelae of HIV, and this study demonstrates that the gene sets potentially predictive of these changes may well be disturbed within the first few weeks of infection (Figure 5B).

Our study has a number of inherent limitations. Completing analyses on mixed cell samples (as opposed to single cell) does introduce a level of effect modification which we have attempted to describe using cell population deconvolution analysis to estimate population frequencies, and describing the effects of CD4+ T cell population adjustment. As with all genome-wide association studies, we cannot rule out an element of reverse causation; for example, there is a possibility that some genes were over-expressed as a by-product of heightened cell turnover, rather than because they contribute mechanistically to the turnover. Almost certainly our DEG results reflect a combination of both. Others have attempted to untangle this issue *in vitro* for specific genes, arguing that individual genes are essential mediators of cell proliferative signalling, rather than consequences of said proliferation (Jimenez et al., 2018; Zhang et al., 2021).

Compared to previous studies, this study offers a large sample size of PMBC collected *ex vivo* from patients in the earliest stages of HIV infection, prior to treatment. Although sequencing of samples from the recruitment visit would have been ideal, it was unfeasible for our rural laboratory to process more than 3000 blood samples collected at this time point, when only 2.7% of participants would be diagnosed with acute HIV infection (Pastor et al., 2017a). Additionally, most study participants presented with a non-specific febrile illness, and recall at four weeks post-recruitment allowed time for resolution of other illnesses, thereby enabling comparisons with relatively healthy, time-matched controls. The burden of acute infections is high in ESA; however, Table 1 clearly shows that there were no significant differences in infections other than HIV between the groups.

In this cohort, gene expression was measured intracellularly in PBMC, whereas protein biomarkers were measured in plasma (Pastor et al., 2017a). Despite the limitations of correlating intracellular gene expression with peripherally circulating proteins, 20% of measured proteins correlated with gene expression, approximating previous reports (Chen et al., 2002; Du et al., 2014). Although RNA-seq does not require functional validation in the classical sense (Coenye, 2021), these findings support the biological plausibility of our results.

## Conclusions

This work demonstrates that acute HIV severely disrupts circulating immune cells, profoundly upregulating many genes, while downregulating others. The degree of dysregulation was frequently, but not universally, correlated with viral load. Analysis of the impacted genes, gene sets and pathways reveals in detail the mechanisms behind the shift toward a system of high cell turnover, cell death and cell expansion, and unwitting viral propagation. This work highlights a need for further investigation of the role of E2F transcription factors in driving exhaustive cell division during acute infection; the role of BCL2 family downregulation in permitting unchecked lymphocyte death, and the unknown role of highly over-expressed TMEM155 during acute infection. CCNA2, CDCA7, and IFI27 are being explored as therapeutic targets within oncology spheres, and heightened expression of these transcripts in acute infection suggest that learnings may be relevant for application in HIV treatments. Additionally, the high degree of consistency in gene expression among HIV cases hints at the possibility that transcriptomic biomarkers might be useful as indicators of recent incident HIV infection. These findings illuminate the molecular mechanisms underlying the pathogenesis of acute HIV infection *in vivo*, and have implications for designing therapeutic interventions aiming to prevent or halt this early immune damage.

## Data availability statement

The transcriptomic datasets presented in this study can be found in online repositories. The names of the repository/

repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/geo/>, GSE199911.

## Ethics statement

The studies involving human participants were reviewed and approved by ethical review boards at Barcelona Clinic Hospital (2011/6264), the Ministry of Health of Mozambique (461/CNBS/12) and the University of Western Australia (2019/RA/4/1/6296). The patients/participants provided their written informed consent to participate in this study.

## Author contributions

Conceptualization, EP, MJ, and PS; Methodology, EP and MJ; Software, EP, MJ, KC, and DA; Formal Analysis, EP and MJ; Investigation, all authors; Resources, EP, MJ, LP, LF-S, and CJ; Data Curation, EP and MJ; Writing – Original Draft, EP and MJ; Writing – Review & Editing, all authors; Visualization, EP and MJ; Supervision, DN and PS; Project Administration, EP and MJ; Funding Acquisition, HC, DN and PS. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by Bill and Melinda Gates Foundation [OPP1068252] to DN, The Spanish Ministry of Science (Mineco) [SAF-2011-27901] to DN, the Australian Government Research Training Program Scholarship to EP and to MJ, the University of Western Australia Safety Net Top-Up Scholarship to EP and to MJ, and The Spanish Ministry of Health through the Institute of Health Carlos III (ISCIII) [FI12/00096] to LP. ISGlobal is a member of the CERCA Programme, Generalitat de Catalunya. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Acknowledgments

The authors are grateful for the continued support of the clinical staff at the Manhiça District Hospital, as well as the research, laboratory and administrative staff of the Centro de Investigação de Saúde de Manhiça (CISM). We would particularly like to thank the study participants who contributed to this research. We also thank Julià Blanco and Jorge Carrillo for their work on plasma biomarkers which is referenced in this paper.

## Conflict of interest

The 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.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2023.1074847/full#supplementary-material>

### SUPPLEMENTARY TABLE 1

Full list of differentially expressed genes – acute HIV cases versus HIV negative controls. # = Rank by adj.P.Val; EnsemblID = Ensembl Stable ID name of the transcript; HGNC = HUGO Gene Nomenclature Committee name of the gene; logFC = Log<sub>2</sub> fold change; AveExpr = average expression across all samples, in log<sub>2</sub> counts per million reads; adj.P.Val = Benjamini-Hochberg false discovery rate adjusted p-value.

### SUPPLEMENTARY TABLE 2

Full list of differentially expressed genes that correlate with viral load in acute HIV cases. # = Rank by rho; EnsemblID = Ensembl Stable ID name of the transcript; HGNC = HUGO Gene Nomenclature Committee name of the gene; rho = Spearman correlation coefficient; adj.P.Val = false discovery rate adjusted p-value.

### SUPPLEMENTARY TABLE 3

Effect of adjustment for CD4+ T cell count (%) on Top 50 DEG. Pos. = Position ranked by adj.P.Val; HGNC = HUGO Gene Nomenclature Committee name of the gene; logFC = Log<sub>2</sub> fold change; AveExpr = average expression across all samples, in log<sub>2</sub> counts per million reads; adj.P.Val = Benjamini-Hochberg false discovery rate adjusted p-value; no-adj pos. = comparative position in the previous table, without adjustment for CD4+ T cell count (%). Yellow = top 50 DEG concordant in both analyses; orange = top 50 DEG discordant between analyses (though all within top 100).

### SUPPLEMENTARY TABLE 4

Full list of differentially expressed gene sets – acute HIV cases versus HIV negative controls as determined by gene set enrichment analysis (GSEA). (A) Overexpressed gene sets. (B) Underexpressed gene sets. # = Rank by NES; Name = Gene Ontology name of the geneset; Size = the number of genes in the geneset; NES = Normalized Enrichment Score; NOM p-val = nominal P value (statistical significance of the observed ES relative to null distribution); FDR q-val = False Discovery Rate Q value (the P value has been adjusted for FDR).

### SUPPLEMENTARY TABLE 5

Intracellular gene expression changes correlated moderately with measured plasma proteins (significant results only, of forty-nine quantifiable soluble biomarkers).

### SUPPLEMENTARY FIGURE 1

Mean mapped reads of three case samples as sequenced on two different RNAseq batches.

### SUPPLEMENTARY FIGURE 2

Top 20 most up and down-regulated gene sets, associated with adjustment for cell populations.

## References

- Abad, M. A., Zou, J., Medina-Pritchard, B., Nigg, E. A., Rappsilber, J., Santamaria, A., et al. (2016). Ska3 ensures timely mitotic progression by interacting directly with microtubules and Ska1 microtubule binding domain. *Sci. Rep.* 6 (1), 34042. doi: 10.1038/srep34042
- Barragán, M., Pons, J., Ferrer-Vaquero, A., Cornet-Bartolomé, D., Schweitzer, A., Hubbard, J., et al. (2017). The transcriptome of human oocytes is related to age and ovarian reserve. *Mol. Hum. Reprod.* 23 (8), 535–548. doi: 10.1093/molehr/gax033
- Bellocchi, M. C., Forbici, F., Palombi, L., Gori, C., Coelho, E., Svicher, V., et al. (2005). Subtype analysis and mutations to antiviral drugs in HIV-1-infected patients from Mozambique before initiation of antiretroviral therapy: Results from the DREAM programme. *J. Med. Virol.* 76 (4), 452–458. doi: 10.1002/jmv.20382
- Bender, D., Da Silva, E. M. L., Chen, J., Poss, A., Gaway, L., Rulon, Z., et al. (2020). Multivalent interaction of ESCO2 with the replication machinery is required for sister chromatid cohesion in vertebrates. *Proc. Natl. Acad. Sci.* 117 (2), 1081–1089. doi: 10.1073/pnas.1911936117
- Bochman, M. L., and Schwacha, A. (2009). The mcm complex: Unwinding the mechanism of a replicative helicase. *Microbiol. Mol. Biol. Rev.* 73 (4), 652–683. doi: 10.1128/MMBR.00019-09
- Burns, D. T., Donkervoort, S., Müller, J. S., Knierim, E., Bharucha-Goebel, D., Faqeih, E. A., et al. (2018). Variants in EXOSC9 disrupt the RNA exosome and result in cerebellar atrophy with spinal motor neuronopathy. *Am. J. Hum. Genet.* 102 (5), 858–873. doi: 10.1016/j.ajhg.2018.03.011
- Cartoni, R., Norsworthy, M. W., Bei, F., Wang, C., Li, S., Zhang, Y., et al. (2016). The mammalian-specific protein Armcx1 regulates mitochondrial transport during axon regeneration. *Neuron* 92 (6), 1294–1307. doi: 10.1016/j.neuron.2016.10.060
- Chen, G., Gharib, T. G., Huang, C.-C., Taylor, J. M. G., Misek, D. E., Kardias, S. L. R., et al. (2002). Discordant protein and mRNA expression in lung adenocarcinomas. *Mol. Cell Proteomics* 1 (4), 304–313. doi: 10.1074/mcp.M200008-MCP200
- Chew, G. M., Fujita, T., Webb, G. M., Burwitz, B. J., Wu, H. L., Reed, J. S., et al. (2016). TIGIT marks exhausted T cells, correlates with disease progression, and serves as a target for immune restoration in HIV and SIV infection. *PLoS Pathogens* 12 (1), e1005349. doi: 10.1371/journal.ppat.1005349
- Chhangawala, S., Rudy, G., Mason, C. E., and Rosenfeld, J. A. (2015). The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome Biol.* 16 (1), 131. doi: 10.1186/s13059-015-0697-y
- Chun, T.-W., Moir, S., and Fauci, A. S. (2015). HIV Reservoirs as obstacles and opportunities for an HIV cure. *Nat. Immunol.* 16 (6), 584–589. doi: 10.1038/ni.3152
- Coenye, T. (2021). Do results obtained with RNA-sequencing require independent verification? *Biofilm* 3, 100043. doi: 10.1016/j.biofilm.2021.100043
- Cohen, M. S., Gay, C. L., Busch, M. P., and Hecht, F. M. (2010). The detection of acute HIV infection. *J. Infect. Dis.* 202 Suppl 2, S270–S277. doi: 10.1086/655651
- Dębski, K. J., Pitkanen, A., Puhakka, N., Bot, A. M., Khurana, I., Harikrishnan, K. N., et al. (2016). Etiology matters – genomic DNA methylation patterns in three rat models of acquired epilepsy. *Sci. Rep.* 6 (1), 25668. doi: 10.1038/srep25668
- Davenport, M. P., Khoury, D. S., Cromer, D., Lewin, S. R., Kelleher, A. D., and Kent, S. J. (2019). Functional cure of HIV: The scale of the challenge. *Nat. Rev. Immunol.* 19 (1), 45–54. doi: 10.1038/s41577-018-0085-4
- Delaney, K. P., Hanson, D. L., Masciotra, S., Ethridge, S. F., Wesolowski, L., and Owen, S. M. (2017). Time until emergence of HIV test reactivity following infection with HIV-1: Implications for interpreting test results and retesting after exposure. *Clin. Infect. Dis.* 64 (1), 53–59. doi: 10.1093/cid/ciw666
- Deng, B., Lv, W., Duan, W., Liu, Y., Li, Z., Song, X., et al. (2018). FGF9 modulates schwann cell myelination in developing nerves and induces a pro-inflammatory environment during injury. *J. Cell Biochem.* 119 (10), 8643–8658. doi: 10.1002/jcb.27105
- Descours, B., Petitjean, G., Loez-Zaragoza, J. L., Bruel, T., Raffel, R., Psomas, C., et al. (2017). CD32a is a marker of a CD4 T-cell HIV reservoir harbouring replication-competent proviruses. *Nature* 543 (7646), 564–56+. doi: 10.1038/nature21710
- Ding, H., Zhang, X., Zhu, Z., Wang, S., Xiong, Y., Shang, H., et al. (2022). Membrane protein OTOF is a type I interferon-induced entry inhibitor of HIV-1 in macrophages. *mBio* 13 (4), e0173822. doi: 10.1128/mbio.01738-22
- Du, H. W., Chen, G. Y., Yang, C. H., Li, C. B., Xun, Y. P., Liu, J., et al. (2014). Multiple correlations of mRNA expression and protein abundance in human cytokine profile. *Mol. Biol. Rep.* 41 (10), 6985–6993. doi: 10.1007/s11033-014-3585-8
- Engelbrecht, S., Koulińska, I., Smith, T.-L., Barreto, J., and Van Rensburg, E. J. (1998). Sequence note: Variation in HIV type 1 V3 region env sequences from Mozambique. *AIDS Res. Hum. Retroviruses* 14 (9), 803–805. doi: 10.1089/aid.1998.14.803
- Fiebig, E. W., Wright, D. J., Rawal, B. D., Garrett, P. E., Schumacher, R. T., Peddada, L., et al. (2003). Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *Aids* 17 (13), 1871–1879. doi: 10.1097/00002030-200309050-00005
- Fromentin, R., Bakeman, W., Lawani, M. B., Khoury, G., Hartogensis, W., DaFonseca, S., et al. (2016). CD4+ T cells expressing PD-1, TIGIT and LAG-3 contribute to HIV persistence during ART. *PLoS Pathogens* 12 (7), e1005761. doi: 10.1371/journal.ppat.1005761

- Gartner, M. J., Roche, M., Churchill, M. J., Gorry, P. R., and Flynn, J. K. (2020). Understanding the mechanisms driving the spread of subtype c HIV-1. *EBioMedicine* 53, 102682. doi: 10.1016/j.ebiom.2020.102682
- Gill, R. M., Gabor, T. V., Couzens, A. L., and Scheid, M. P. (2013). The MYC-associated protein CDCA7 is phosphorylated by AKT to regulate MYC-dependent apoptosis and transformation. *Mol. Cell. Biol.* 33 (3), 498–513. doi: 10.1128/MCB.00276-12
- Gonzalez, R., Munguambe, K., Aponte, J., Bavo, C., Nhalungo, D., Macete, E., et al. (2012). High HIV prevalence in a southern semi-rural area of Mozambique: A community-based survey. *HIV Med.* 13 (10), 581–588. doi: 10.1111/j.1468-1293.2012.01018.x
- Gruppenmacher, A. T., Halpern, A. L., Bonaldo Mde, F., Huang, C. C., Hamm, C. A., de Andrade, A., et al. (2013). Study of the gene expression and microRNA expression profiles of malignant rhabdoid tumors originated in the brain (AT/RT) and in the kidney (RTK). *Child's nervous system ChNS* 29 (11), 1977–1983. doi: 10.1007/s00381-013-2268-4
- Hemelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Fleminger, I., Kirtley, S., et al. (2019). Global and regional molecular epidemiology of HIV-1, 1990–2015: A systematic review, global survey, and trend analysis. *Lancet Infect. Dis.* 19 (2), 143–155. doi: 10.1016/S1473-3099(18)30647-9
- Heng, J. L., Viti, L., Pugh, K., Marshall, O. J., and A. M. (2022). Understanding the impact of ZBTB18 missense variation on transcription factor function in neurodevelopment and disease. *J. Neurochem.* 161 (3), 219–235. doi: 10.1111/jnc.15572
- Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M., and Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373 (6510), 123–126. doi: 10.1038/373123a0
- Huang, H., Lv, J., Huang, Y., Mo, Z., Xu, H., Huang, Y., et al. (2022). IFI27 is a potential therapeutic target for HIV infection. *Ann. Med.* 54 (1), 314–325. doi: 10.1080/07853890.2021.1995624
- Huang, W., Zhu, C., Wang, H., Horvath, E., and EA, E. (2008). The interferon consensus sequence-binding protein (ICSBP/IRF8) represses PTPN13 gene transcription in differentiating myeloid cells. *J. Biol. Chem.* 283 (12), 7921–7935. doi: 10.1074/jbc.M706710200
- Hycza, M. D., Kovacs, C., Loutfy, M., Halpern, R., Heisler, L., Yang, S., et al. (2007). Distinct transcriptional profiles in ex vivo CD4+ and CD8+ T cells are established early in human immunodeficiency virus type 1 infection and are characterized by a chronic interferon response as well as extensive transcriptional changes in CD8+ T cells. *J. Virol.* 81 (7), 3477–3486. doi: 10.1128/JVI.01552-06
- Ijaz, B., Ahmad, W., Das, T., Shabbiri, K., Husnain, T., and Hassan, S. (2019). HCV infection causes cirrhosis in human by step-wise regulation of host genes involved in cellular functioning and defense during fibrosis: Identification of bio-markers. *Genes Dis.* 6 (3), 304–317. doi: 10.1016/j.gendis.2019.04.007
- Jiang, A., Zhou, Y., Gong, W., Pan, X., Gan, X., Wu, Z., et al. (2022). CCNA2 as an immunological biomarker encompassing tumor microenvironment and therapeutic response in multiple cancer types. *Oxid. Med. Cell. Longev.* 2022, 5910575. doi: 10.1155/2022/5910575
- Jimenez, R., Martin-Cortazar, C., Kourani, O., Chiodo, Y., Cordoba, R., Dominguez-Franjo, M. P., et al. (2018). CDCA7 is a critical mediator of lymphomagenesis that selectively regulates anchorage-independent growth. *Haematologica* 103 (10), 1669–1678. doi: 10.3324/haematol.2018.188961
- Johnston, M. I., and Fauci, A. S. (2008). An HIV vaccine — challenges and prospects. *New Engl. J. Med.* 359 (9), 888–890. doi: 10.1056/NEJMp0806162
- Judge, M., Parker, E., Nanche, D., and Le Souéf, P. (2020). Gene expression: The key to understanding HIV-1 infection? *Microbiol. Mol. Biol. Rev.* 84 (2), e00080–e00019. doi: 10.1128/MMBR.00080-19
- Kale, J., Osterlund, E. J., and Andrews, D. W. (2018). BCL-2 family proteins: Changing partners in the dance towards death. *Cell Death Differ.* 25 (1), 65–80. doi: 10.1038/cdd.2017.186
- Kane, M., Zang, T. M., Rihn, S. J., Zhang, F., Kueck, T., Alim, M., et al. (2016). Identification of interferon-stimulated genes with antiretroviral activity. *Cell Host Microbe* 20 (3), 392–405. doi: 10.1016/j.chom.2016.08.005
- Kazer, S. W., Aicher, T. P., Muema, D. M., Carroll, S. L., Ordovas-Montanes, J., Miao, V. N., et al. (2020). Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. *Nat. Med.* 26 (4), 511–518. doi: 10.1038/s41591-020-0799-2
- Konrad, B. P., Taylor, D., Conway, J. M., Ogilvie, G. S., and Coombs, D. (2017). On the duration of the period between exposure to HIV and detectable infection. *Epidemics* 20, 73–83. doi: 10.1016/j.epidem.2017.03.002
- Kuo, H. H., and Lichterfeld, M. (2018). Recent progress in understanding HIV reservoirs. *Curr. Opin. HIV AIDS* 13 (2), 137–142. doi: 10.1097/COH.0000000000000441
- Lavreys, L., Baeten, J. M., Chohan, V., McClelland, R. S., Hassan, W. M., Richardson, B. A., et al. (2006). Higher set point plasma viral load and more-severe acute HIV type 1 (HIV-1) illness predict mortality among high-risk HIV-1-infected African women. *Clin. Infect. Dis.* 42 (9), 1333–1339. doi: 10.1086/503258
- Li, J., Parker, B., Martyn, C., Natarajan, C., and G, J. (2013). The PMP22 gene and its related diseases. *Mol. Neurobiol.* 47 (2), 672–698. doi: 10.1007/s12035-012-8370-x
- Li, Q., Smith, A. J., Schacker, T. W., Carlis, J. V., Duan, L., Reilly, C. S., et al. (2009). Microarray analysis of lymphatic tissue reveals stage-specific, gene expression signatures in HIV-1 infection. *J. Immunol.* 183 (3), 1975–1982. doi: 10.4049/jimmunol.0803222
- Liu, M., Guo, S., Hibbert, J. M., Jain, V., Singh, N., Wilson, N. O., et al. (2011). CXCL10/IP-10 in infectious diseases pathogenesis and potential therapeutic implications. *Cytokine Growth factor Rev.* 22 (3), 121–130. doi: 10.1016/j.cytogtr.2011.06.001
- Liu, X., Lin, L., Lu, L., Li, X., Han, Y., Qiu, Z., et al. (2022). Comparative transcriptional analysis identified characteristic genes and patterns in HIV-infected immunological non-responders. *Front. Immunol.* 13. doi: 10.3389/fimmu.2022.807890
- Liu, X., Simon, J. M., Xie, H., Hu, L., Wang, J., Zurlo, G., et al. (2020). Genome-wide screening identifies SFMBT1 as an oncogenic driver in cancer with VHL loss. *Mol. Cell* 77 (6), 1294–306.e5. doi: 10.1016/j.molcel.2020.01.009
- Ma, Q. (2019). MiR-219-5p suppresses cell proliferation and cell cycle progression in esophageal squamous cell carcinoma by targeting CCNA2. *Cell. Mol. Biol. Lett.* 24 (1), 4. doi: 10.1186/s11658-018-0129-6
- Mehandru, S., Poles, M. A., Tenner-Racz, K., Manuelli, V., Jean-Pierre, P., Lopez, P., et al. (2007). Mechanisms of gastrointestinal CD4+ T-cell depletion during acute and early human immunodeficiency virus type 1 infection. *J. Virol.* 81 (2), 599–612. doi: 10.1128/JVI.01739-06
- Miller, W. C., Rosenberg, N. E., Rutstein, S. E., and Powers, K. A. (2010). Role of acute and early HIV infection in the sexual transmission of HIV. *Curr. Opin. HIV AIDS* 5 (4), 277–282. doi: 10.1097/COH.0b013e32833a0d3a
- Newman, A., Alizadeh, A., and Liu, C. (2015). *CIBERSORT* (Stanford, CA: Stanford University). Available at: <https://cibersort.stanford.edu/>.
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi: 10.1038/nmeth.3337
- Nishitsuji, H., Sawada, L., Sugiyama, R., and T, H. (2015). ZNF10 inhibits HIV-1 LTR activity through interaction with NK-κB and Sp1 binding motifs. *FEBS Lett.* 589 (15), 2019–2025. doi: 10.1016/j.febslet.2015.06.013
- Ogderel, Z., Sonti, S., Hernandez, N., Park, J., Ottman, R., Louis, E. D., et al. (2019). Whole genome sequencing and rare variant analysis in essential tremor families. *PLoS One* 14 (8), e0220512–e. doi: 10.1371/journal.pone.0220512
- Palm, A. A., Veerla, S., Lindman, J., Isberg, P.-E., Johansson, E., Biague, A., et al. (2022). Interferon alpha-inducible protein 27 expression is linked to disease severity in chronic infection of both HIV-1 and HIV-2. *Front. Virol.* 2. doi: 10.3389/fviro.2022.929053
- Pardons, M., Baxter, A. E., Massanella, M., Pagliuzza, A., Fromentin, R., Dufour, C., et al. (2019). Single-cell characterization and quantification of translation-competent viral reservoirs in treated and untreated HIV infection. *PLoS Pathogens* 15 (2), e1007619. doi: 10.1371/journal.ppat.1007619
- Parreira, R., Piedade, J., Domingues, A., Lobão, D., Santos, M., Venenno, T., et al. (2006). Genetic characterization of human immunodeficiency virus type 1 from beira, Mozambique. *Microbes Infect.* 8 (9–10), 2442–2451. doi: 10.1016/j.micinf.2006.05.005
- Pastor, L., Parker, E., Carrillo, J., Urrea, V., Fuente-Soro, L., Respeito, D., et al. (2017a). A cytokine pattern that differentiates preseroconversion from postseroconversion phases of primary HIV infection. *J. Acq. Imm. Def.* 74 (4), 459–466. doi: 10.1097/QAI.0000000000001272
- Pastor, L., Urrea, V., Carrillo, J., Parker, E., Fuente-Soro, L., Jairoce, C., et al. (2017b). Dynamics of CD4 and CD8 T-cell subsets and inflammatory biomarkers during early and chronic HIV infection in Mozambican adults. *Front. Immunol.* 8, 1925. doi: 10.3389/fimmu.2017.01925
- Phatak, A., Athar, M., Crowell, J. A., Leffel, D., Herbert, B.-S., Bale, A. E., et al. (2019). Global gene expression of histologically normal primary skin cells from BCNS subjects reveals “single-hit” effects that are influenced by rapamycin. *Oncotarget* 10 (14), 1360–1387. doi: 10.18632/oncotarget.26640
- Pilcher, C. D., Tien, H. C., Eron, J. J., Vernazza, P. L., Leu, S. Y., Stewart, P. W., et al. (2004). Brief but efficient: Acute HIV infection and the sexual transmission of HIV. *J. Infect. Dis.* 189 (10), 1785–1792. doi: 10.1086/386333
- Popper, S. J., Gordon, A., Liu, M., Balmaseda, A., Harris, E., and Relman, D. A. (2012). Temporal dynamics of the transcriptional response to dengue virus infection in Nicaraguan children. *PLoS Negl. Trop. Dis.* 6 (12), e1966. doi: 10.1371/journal.pntd.0001966
- Powers, K. A., Ghani, A. C., Miller, W. C., Hoffman, I. F., Pettifor, A. E., Kamanga, G., et al. (2011). The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: A modelling study. *Lancet* 378 (9787), 256–268. doi: 10.1016/S0140-6736(11)60842-8
- Rotger, M., Dang, K. K., Fellay, J., Heinzen, E. L., Feng, S., Descombes, P., et al. (2010). Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. *PLoS Pathogens* 6 (2), e1000781. doi: 10.1371/journal.ppat.1000781
- Saco, C., Nhalungo, A., Nhalungo, D., Aponte, J. J., Bassat, Q., Augusto, O., et al. (2013). Profile: Manhica health research centre (Manhica HDSS). *Int. J. Epidemiol.* 42 (5), 1309–1318. doi: 10.1093/ije/dyt148
- Schinzler, A., Kaufmann, T., and Borner, C. (2004). Bcl-2 family members: Intracellular targeting, membrane-insertion, and changes in subcellular localization. *Biochim. Biophys. Acta (BBA) - Mol. Cell Res.* 1644 (2), 95–105. doi: 10.1016/j.bbamcr.2003.09.006
- Schoggins, J. W., MacDuff, D. A., Imanaka, N., Gainey, M. D., Shrestha, B., Eitson, J. L., et al. (2014). Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature* 505 (7485), 691–695. doi: 10.1038/nature12862

- Schoggins, J. W., and Rice, C. M. (2011). Interferon-stimulated genes and their antiviral effector functions. *Curr. Opin. Virol.* 1 (6), 519–525. doi: 10.1016/j.coviro.2011.10.008
- Shiau, S. (2017). *Early treatment initiation and outcomes in perinatally HIV-infected infants and young children in Johannesburg, south Africa: Age, aging, and antiretrovirals* (New York City, NY: Columbia University).
- Smith, A. J., Li, Q., Wietgreffe, S. W., Schacker, T. W., Reilly, C. S., and Haase, A. T. (2010). Host genes associated with HIV-1 replication in lymphatic tissue. *J. Immunol.* 185 (9), 5417–5424. doi: 10.4049/jimmunol.1002197
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49 (D1), D605–DD12. doi: 10.1093/nar/gkab835
- Tanenbaum, M. E., Macúrek, L., Janssen, A., Geers, E. F., Alvarez-Fernández, M., and Medema, R. H. (2009). Kif15 cooperates with Eg5 to promote bipolar spindle assembly. *Curr. Biol.* 19 (20), 1703–1711. doi: 10.1016/j.cub.2009.08.027
- The Joint United Nations Programme on HIV/AIDS (UNAIDS) (2020). *UNAIDS data 2020* Vol. 2020 (Geneva: UNAIDS). Report No.: JC2997E.
- Toyono, T., Usui, T., Yokoo, S., Taketani, Y., Nakagawa, S., Kuroda, M., et al. (2015). Angiopoietin-like 7 is an anti-angiogenic protein required to prevent vascularization of the cornea. *PLoS One* 10 (1), e0116838. doi: 10.1371/journal.pone.0116838
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* 7 (3), 562–578. doi: 10.1038/nprot.2012.016
- Ullah, H., Sajid, M., Yan, K., Feng, J. P., He, M., Shereen, M. A., et al. (2021). Antiviral activity of interferon alpha-inducible protein 27 against hepatitis b virus gene expression and replication. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.656353
- Velasco, C., Parker, E., Pastor, L., Nhama, A., Macuacua, S., Mandomando, I., et al. (2015). Case report: Rapid HIV progression during acute HIV-1 subtype c infection in a mozambican patient with atypical seroconversion. *Am. J. Trop. Med. Hygiene.* 92 (4), 681–683. doi: 10.4269/ajtmh.14-0554
- Vidya Vijayan, K. K., Karthigeyan, K. P., Tripathi, S. P., and Hanna, L. E. (2017). Pathophysiology of CD4+ T-cell depletion in HIV-1 and HIV-2 infections. *Front. Immunol.* 8 (580). doi: 10.3389/fimmu.2017.00580
- Wang, Z., Arat, S., Magid-Slav, M., and Brown, J. R. (2018). Meta-analysis of human gene expression in response to mycobacterium tuberculosis infection reveals potential therapeutic targets. *BMC Syst. Biol.* 12 (1), 3. doi: 10.1186/s12918-017-0524-z
- Wang, H.-W., Wu, Y.-H., Hsieh, J.-Y., Liang, M.-L., Chao, M.-E., Liu, D.-J., et al. (2010). Pediatric primary central nervous system germ cell tumors of different prognosis groups show characteristic miRNome traits and chromosome copy number variations. *BMC Genomics* 11, 132. doi: 10.1186/1471-2164-11-132
- Wawer, M. J., Gray, R. H., Sewankambo, N. K., Serwadda, D., Li, X., Laeyendecker, O., et al. (2005). Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in rakai, Uganda. *J. Infect. Dis.* 191 (9), 1403–1409. doi: 10.1086/429411
- Xie, A., Wang, P., Chen, D., and Z, H. (2022). Aberrant ARMCX1 expression is an independent predictor of poor prognosis in gastric cancer. *J. Oncol.* 2022, 9348917. doi: 10.1155/2022/9348917
- Yang, R., Du, Y., Wang, L., Chen, Z., and Liu, X. (2020). Weighted gene co-expression network analysis identifies CCNA2 as a treatment target of prostate cancer through inhibiting cell cycle. *J. Cancer.* 11 (5), 1203–1211. doi: 10.7150/jca.38173
- Yang, E., and Li, M. M. H. (2020). All about the RNA: Interferon-stimulated genes that interfere with viral RNA processes. *Front. Immunol.* 11. doi: 10.3389/fimmu.2020.605024
- Zhai, Y., Franco, L. M., Atmar, R. L., Quarles, J. M., Arden, N., Bucacas, K. L., et al. (2015). Host transcriptional response to influenza and other acute respiratory viral infections – a prospective cohort study. *PLoS Pathogens* 11 (6), e1004869. doi: 10.1371/journal.ppat.1004869
- Zhang, Q., Yuan, J., Liu, Y., Liu, X., Lv, T., Zhou, K., et al. (2021). KIAA0101 knockdown inhibits cell proliferation and induces cell cycle arrest and cell apoptosis in chronic lymphocytic leukemia cells. *Ann. Trans. Med.* 9 (6), 487. doi: 10.21037/atm-21-626
- Zhang, Y., Zhang, H., An, M., Zhao, B., Ding, H., Zhang, Z., et al. (2018). Crosstalk in competing endogenous RNA networks reveals new circular RNAs involved in the pathogenesis of early HIV infection. *J. Trans. Med.* 16 (1), 332. doi: 10.1186/s12967-018-1706-1
- Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 9 (1), e78644. doi: 10.1371/journal.pone.0078644
- Ziemann, M. *Genome spot - data analysis step 8: Pathway analysis with GSEA*. Available at: <http://genomespot.blogspot.com/2014/09/data-analysis-step-8-pathway-analysis.html>.