Transcriptome analysis of peripheral blood of *Schistosoma mansoni* infected children from the Albert Nile region in Uganda reveals genes implicated in fibrosis pathology.

Joyce Namulondo[1], Oscar Asanya Nyangiri[1], Magambo Phillip Kimuda[1], Peter Nambala[2], Jacent Nassuuna[3], Moses Egesa[3,4], Barbara Nerima[2], Savino Biryomumaisho[1], Claire Mack Mugasa[1], Immaculate Nabukenya[1], Kato Drago[1], Alison Elliott[3,4], Harry Noyes[5], Robert Tweyongyere[1], Enock Matovu[1], Julius Mulindwa[2]* for the TrypanoGEN+ research group of the H3Africa consortium

[1]College of Veterinary Medicine, Animal Resources and Biosecurity, Makerere University, Kampala, Uganda

[2]College of Natural Sciences, Makerere University, Kampala, Uganda.

[3]Vaccine Research Theme, MRC/UVRI and LSHTM Uganda Research Unit, Entebbe, Uganda

[4]Department of Infection Biology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom

[5]Centre for Genomic Research, University of Liverpool, L69 7ZB, United Kingdom

*Corresponding author: Julius Mulindwa julius.mulindwa@mak.ac.ug

Abstract

Over 290 million people are infected by schistosomes worldwide. Schistosomiasis control efforts focus on mass drug treatment with praziquantel (PZQ), a drug that kills the adult worm of all *Schistosoma* species. Nonetheless, re-infections have continued to be detected in endemic areas with individuals living in the same area presenting with varying infection intensities. Our objective was to characterize the transcriptome profiles in peripheral blood of children between 10 - 15 years with varying intensities of *Schistosoma mansoni* infection living along the Albert Nile in Uganda.

RNA extracted from peripheral blood collected from 44 *S. mansoni* infected (34 high and 10 low by circulating anodic antigen [CAA] level) and 20 uninfected children was sequenced using Illumina NovaSeq S4 and the reads aligned to the GRCh38 human genome. Differential gene expression analysis was done using DESeq2 and enriched pathways in differentially expressed genes (DEGs) were identified using REACTOME. Principal component analysis revealed clustering of gene expression by gender when *S. mansoni* infected children were compared with uninfected children. In addition, we identified 14 DEGs between *S. mansoni* infected and uninfected individuals, 56 DEGs between children with high infection intensity and uninfected individuals, 33 DEGs between those with high infection intensity and low infection intensity and no DEGs between those with low infection and uninfected individuals. We also observed upregulation and downregulation of some DEGs that are associated with fibrosis and its regulation. These data suggest expression of fibrosis associated genes as well as genes that regulate fibrosis in *S. mansoni* infection. The relatively few significant DEGS observed in children with schistosomiasis suggests that chronic *S. mansoni* infection is a stealth infection that does not stimulate a strong immune response.

Author Summary

Schistosomiasis is a neglected tropical disease transmitted via an intermediate snail host through contact with contaminated fresh water. Even with routine Mass Drug Administration for treatment of the infection, re-infections are still common and variations in infection intensity and pathology are still observed in individuals in the same location. These may be due to differences in individuals' response to *S. mansoni* infection. In this study, we used RNAseq to identify differentially expressed genes associated with *S. mansoni* infection in children between 10-15 years. We conducted comparisons between phenotypes including infection intensities measured by circulating anodic antigen, wasting by body mass index and

stunting by height-for-age z score. Our data showed very low numbers of significant differentially expressed genes in all comparisons. Some of the few differentially expressed genes that were observed were associated with fibrosis which is the cause of pathology in humans and has been observed in late stages of *S. mansoni* infection in murine studies.

Introduction

Over 290 million people are infected by schistosomes worldwide. The infection is widespread in tropical and sub-tropical regions with over 78 countries reporting transmission (1). The WHO estimated that approximately 236.6 million people required treatment in 2019 (2). During their lifespan, schistosomes live in blood vessels and females lay eggs which migrate through and lodge in organs of the host provoking local and systemic responses(1). Despite mass administration of praziquantel (PZQ) to control the infection, communities still have high prevalence of schistosomiasis with variations in intensity of infection among individuals (3,4). We recently reported a high prevalence of *S. mansion* infection and stunting among school age children along the Albert Nile in Uganda, more in boys than in girls, coupled with variation in infection intensity (4). However, there is limited information on gene expression in humans infected with schistosomes that could underpin the observed varying infection intensity. Variations in infection intensity may be linked to the ability of the host to respond to the infection which may be driven by a number of underlying molecular mechanisms employed by the host. Animal studies have shown upregulation of immune genes early in the infection and of metabolic related genes later in the infection (5–7) as well as differences in expression profiles between susceptible and less susceptible hosts (8). Additionally, responses of human males and females to infection appear to differ with each having distinct sets of DEGs as observed in *Schistosoma haematobium* infection (9). To date, there is no study that has profiled gene expression in the blood of humans infected with *S. mansoni*. This study set out to

characterise the transcriptome profiles of genes expressed in children 10 - 15 years of age infected with *S. mansoni* along the Albert Nile. We hypothesized that the expression of genes in peripheral blood of *S. mansoni* infected children varies between children with high infection compared to those with low infection intensity and that both would differ from the uninfected.

Methods

Ethics statement

The study protocol was reviewed by the institutional review board of the Ministry of Health, Vector Control Division Research and Ethics Committee (Reference No. VCDREC106) and approved by the Uganda National Council for Science and Technology (Reference No. UNCST HS 118). The study was conducted with guidance from the district health officials, including the selection and training of the village health teams that were involved in the mobilisation and recruitment of the children into the study. The objectives, potential risks and benefits of the study were explained to the parents/ guardians who signed informed consent, and later explained to the school age children in English and Alur dialect who provided assent for participation in the study. Written formal consent from parents and written assent from the children were obtained. If a child was observed to have *S. mansoni* eggs in their stool, they were offered free treatment, which consisted of praziquantel at a dosage of 40mg/kg administered by trained Ministry of Health personnel, assisted by a district health worker.

Study design and study sites

This was a cross-sectional study carried out in communities along the Albert Nile. Samples were collected from school children aged between 10-15 years in Pakwach District located in the Northern part of Uganda near the Albert Nile. The selected areas of sampling were in sub-counties of Pakwach, Panyingoro, Panyimur and Alwi all of which are within 10km of the Albert Nile.

Screening and recruitment

Screening and recruitment were done as previously described (4). Participants were mobilized and educated about schistosomiasis by village health teams. Those between 10-15 years were registered and recruited based on ability to provide urine for screening by the point-of-care circulating cathodic antigen (POC-CCA). Briefly, 2 drops (100µl) of urine were placed on the POC-CCA test cassette and left at room temperature for 20 minutes prior to visualization. The results were scored by modifying the G scores as previously described (4). The modified scores included: 0 (G1), trace (G2, G3), 1+ (G4, G5), 2+ (G6, G7), 3+ (G8, G9) or 4+ (G10). Initial sampling for RNAseq was done based on POC-CCA scores. These included individuals with high POC-CCA (4+ and 3+), low (1+ and trace) and negative (0). These were further classified by infection intensity using CAA as explained in the section below. Participants who provided urine for screening were recruited to the study after being interviewed and informed consent signed by the parent/guardian and assent by the child. The height and weight were obtained from each participant to obtain estimates of body mass index (BMI) and stunting (Height for Age Z-scores HAZ) as previously described (4).

Sample collection

Following the interview, each selected participant was requested to provide peripheral blood which was collected in PAXgene® Blood RNA (PreAnalytiX, US) tubes for transcriptional analysis and in EDTA tubes (BD Biosciences, US) for plasma separation for circulating anodic antigen (CAA) analysis and for DNA extraction.

CAA assay for *S. mansoni* infection

To measure infection intensity, Circulating Anodic Antigen (CAA) levels were measured in plasma using the UCP-LF CAA test as described previously by Mulindwa and colleagues (4). CAA concentrations > 30 pg/mL were classified as negative (CAA < 30 pg/mL), low infection intensity (CAA 30 to <1000 pg/mL) and high infection intensity (CAA > 1000 pg/mL) (10).

RNA extraction and purification

RNA was extracted from blood collected in PAXgene® Blood RNA tubes using Trizol (Invitrogen, USA) protocol (11). The RNA was quantified using Qubit (Invitrogen, USA) and samples with concentration >1µg were shipped to the Centre for Genomics Research at the University of Liverpool for sequencing where the quality of the samples was checked using an Agilent Bioanalyser.

Library preparation and RNASeq

The QIAseq FastSelect rRNA HMR kits (Qiagen) were used to remove rRNA from total RNA and libraries prepared using the NEBNext Ultra II Directional RNA Library Prep Kit (NEB, New England Biolabs). The libraries were sequenced on an Illumina NovaSeq S4 (Illumina) in the 2x150 read configuration to a target depth of 30m read pairs per sample at the Centre for Genomic Research at the University of Liverpool. FASTQ reads were aligned to the GRCh38 release 84 human genome sequence obtained from Ensembl (12) using HiSat2 (13) and annotated using the human genome reference, *Homo_sapiens* GRCh38.104 from Ensembl.

Identification of DEGs

Differentially expressed genes between phenotypes were identified using DESeq2 (14). Analysis of read counts for each gene considered CAA as the independent variable with age and sex as covariates. Principal component analysis was done using PCA Explorer to identify samples that appeared as outliers (15). Genes with adjusted p-value <0.05, Log2 (FC)> 1.0 for up-regulated genes and Log2 (FC) < -0.8 for down regulated genes were selected as significant differentially expressed. We compared gene expression between different infection intensities by pairwise analysis of the different infection intensities described above; 1) all infected vs uninfected (IU), 2) high infection intensity vs low infection intensity (HL), 3) high infection intensity vs uninfected (HU) and 4) low infection intensity vs uninfected (LU) while including

sex and age as covariates. Gene ontology (GO) enrichment analysis of the DEGs for each pair was done using the GOnet (16) online resource and REACTOME (17).

In addition to high prevalence of schistosomiasis, our previous study found high levels of stunting and under nutrition in this study population (4). To understand the association of BMI and stunting with gene expression, we conducted linear regression analysis with BMI and stunting as dependent variables and sex and age as covariates for each analysis.

To identify the changes in relative abundance of different cell types with *S. mansoni* infection, the proportions of different cell types in each sample were estimated from the expression data using Bisque (Jew et al., 2020). Single cell reference sequence data from bone marrow and peripheral blood from Chinese donors was obtained from 7551 individual human blood cells representing 32 leukocyte cell types (18).

Association of DEGs with schistosomiasis or fibrosis

We searched Pubmed abstracts and titles using the name of each of the 63 differentially expressed genes and the terms "schistosomiasis" or "schistosoma" for association of the DEGs with schistosomiasis. We further searched using the term "fibrosis" with each of the unique DEGs for association with fibrosis.

Results

Following the screening of 914 children aged between 10 – 15 years in Pakwach District, 727 children were recruited for further studies (Mulindwa et., al., 2022) of which 152 children were recruited to participate in this gene expression study based on POC-CCA scores. Eighty (80) of the 152 samples passed the RNAseq quality control and were selected for sequencing to represent the extremes of infection intensity. Of the 80 sequenced, 11 lacked CAA results and were excluded (**Figure 1**). PCA analysis identified five samples which did not cluster closely with the remaining samples and these were also removed (**Figure S1**), as these samples may

be from participants with different ethnicity or suffer from poor-quality of RNA. Therefore, 64 samples were analysed for gene expression in peripheral blood of children aged between 10 – 15 years of which; 34 (17: high infection intensity, 5: low infection intensity, 12: uninfected) were male and 30 (17: high infection intensity, 5: low infection intensity, 8: uninfected) were female (**Table 1).**

Differential expression of genes by comparing *S. mansoni* infection intensity categories

PCA analysis of gene expression data showed clear separation by gender on principal components 1 and 2 (**Figure 2**). Genes were defined as significantly differentially expressed between conditions using the following criteria: adjusted p-value <0.05; fold change (FC) Log2 (FC) > 1.0 for up-regulated genes and Log2 (FC) < -0.8 for down regulated genes. The numbers of differentially expressed up and down regulated genes for each of four contrasts: 1) all infected vs uninfected (IU), 2) high infection vs uninfected (HU), 3) high infection vs low infection (HL), and 4) low infection vs uninfected (LU) are shown in **Table 2**.

Gene expression differences between infected and uninfected children

We found that 14 genes were significant differentially expressed between infected and uninfected children of which 9 (64%) were upregulated and 5 (36%) were down regulated among infected individuals compared to uninfected (IU) (**Tables 2 and 3** and **Figure 3A).** Enriched pathways (**Table 4**) were identified using the REACTOME online tool and GOnet **Table S1**. Of the upregulated genes, when infected were compared to uninfected, only the CAP-Gly domain containing linker protein 1 (*CLIP1*) gene was found in REACTOME. Among the significant downregulated genes, two genes (*SUZ12* and *TRBC2*) were found in REACTOME. Seventeen (17) enriched pathways were identified. These included pathways involved in senescence, T cell signalling, PRC2 methylation of histones and DNA by PRC2, defective pyroptosis, transcriptional regulation, and embryogenesis **(Table 4).**

Pathways enriched by significant DEGs between varying infection intensities.

We identified 56 significant DEGs (**Table 2**) listed in **Table 5** among the children with high *S. mansoni* infection compared to the uninfected of which 43 (77%) were upregulated and 13 (23%) were downregulated (**Figure 3B**). Of the significant DEGs, five (5) upregulated genes; *PGD*, *BLM*, *CLIP1*, *SP110* and *ACSM1* and six (6) downregulated genes; *ITGA4*, *SLC12A1*, *SUZ12*, *A1CF*, *AKR1A1* and *OGT* were found in enriched pathways in REACTOME (**Table S2**). The enriched pathways for the upregulated genes were associated with cellular metabolism, protein regulation and inflammation whereas the pathways of downregulated genes were associated with immune response and cell migration, fibrosis, and necrosis.

We identified 33 significant DEGs **(Table S3** and **Figure 3C)** among individuals with high infection compared with those with low infection (HL). Of these, two (2) upregulated genes; 8-oxoguanine DNA glycosylase (*OGG1*) and proteasome 20S subunit beta 7 (*PSMB7*) were found in enriched pathways. The enriched pathways (**Table S4**) were associated with pro-inflammatory response, fibrosis and chemostasis control. None of the downregulated genes, were found in REACTOME. Of note, there were no significant differentially expressed genes when individuals with low infection were compared with the uninfected (LU) (**Figure 3D)**.

Association of DEGs with schistosomiasis or fibrosis

We searched Pubmed abstracts and titles with each gene name and the term schistosomiasis or fibrosis. None of the 63 genes had informative associations with schistosomiasis. 27 gene names appeared in articles that also mentioned fibrosis (**Table S5**) and 13 of these had well documented associations with fibrosis. Seven genes were also associated with *TGFB1* (**Table S6)**.

Expression of genes associated with stunting and BMI in *S. mansoni* infected children

The mean height for age of the children in the study was in the bottom 3 percent for all children worldwide and 48% of children met the WHO definition of being stunted (height for age < - 2 SD of global mean). To identify the association of stunting and BMI with gene expression in *S. mansoni* infected children, we conducted a linear regression analysis using height for age Z-scores (HAZ) and body mass index (BMI) scores respectively. We identified significant differential expression of the neutral cholesterol ester hydrolase 1 (*NCEH1*) in stunting which was up regulated with log fold change of 1.28 and p-adj value <0.05. Additionally, we identified four genes with expression that was significantly associated with increase in BMI of which three (*MUC5B*, *DMD* and *REXO1L1P*) were upregulated while one (*SERPINA10*) was downregulated **(Figure 4 and Table S7).** Functional analysis of the expressed genes showed that stunting was linked to pathways of catabolic process and lipid metabolic processes. The genes upregulated in children with high BMI were linked to pathways involved with biosynthesis, signal transduction, cellular nitrogen compound metabolic processes and cellular protein modification among others (**Table S8**).

Cell type analysis

To identify the changes in cell type expression with *S. mansoni* infection, the proportions of different cell types in each sample were estimated using Bisque (Jew et al., 2020). We used a 2-sided T-test to identify cell types with significant differences in the proportions of each cell type in the blood of infected and uninfected children (**Figure S2**). Of the 32 cell types evaluated, only Multipotent Progenitors (MPPs) had a significant difference in abundance and were significantly downregulated with a p-value of 0.012 in *S. mansoni* infected individuals compared to the uninfected individuals (**Table S9**).

Discussion

Previous studies have shown that mammalian host response to *Schistosoma* infection varies, with some individuals being more susceptible to infection than others. We previously showed high prevalence of *S. mansoni* infection among children living along Lake Albert in Uganda and the same children also had high levels of wasting and stunting although this was not correlated with *Schistosoma* infection (4). In this study, we present for the first time, data on gene expression in peripheral blood of *S. mansoni* infected children. Gene expression was compared between all infected vs uninfected (IU), highly infected vs uninfected (HU), highly infected vs low infected (HL), and lightly infected vs uninfected (LU). Similar to findings by Dupnik and colleagues, we observed sufficient difference in gene expression between males and females for the two sexes to cluster separately in the principal components analysis (**Figure 2**) (9). Like other chronic infections (19,20), we observed only 63 DEGs in the *S. mansoni* infected. Additionally, our data showed gene expression differences when children with high *S. mansoni* infection intensity were compared with low infection intensities and uninfected whereas there was no significant differential gene expression between children with low infection compared to the uninfected. This suggests that gene expression is scarcely perturbed in individuals with low *S. mansoni* infection. We searched Pubmed abstracts and titles using each of the 63 unique differentially expressed gene names and the terms “schistosomiasis” or “schistosoma” and none of the differentially expressed genes had previously been associated with the response to *Schistosoma* infection.

Our comparison of *S. mansoni* infected children with uninfected highlighted enrichment of pathways through interaction with *CLIP1* among the upregulated genes. *CLIP1* mediates microtubule capture through the interaction with RHO GTPases activating the IQGAPs pathway (21). Microtubule capture may be linked to wound healing through stimulation of cell

271 migration and fibroblasts (22,23) and is involved in T Cell regulation (23–25). This finding is
272 similar to previous murine findings that demonstrated the upregulation of fibrosis linked genes
273 in the late stages of *S. japonicum* infection (5,6,26). Furthermore, enrichment of the
274 SUMOylation of DNA replication protein pathways indicate a potential role of *CLIP1* in the
275 host immune response. Pathogens have been shown to modulate this pathway to evade the
276 immune system (27). The upregulation of *CLIP1* was coupled with downregulation of immune
277 related genes of which we identified T cell receptor beta constant 2 (*TRBC2*) and SUZ12
278 polycomb repressive complex 2 subunits (*SUZ12*) to be linked to pathways in REACTOME.
279 The T cell receptor beta constant 2 (TRBC2) was found to be associated with T cell signalling
280 and regulation pathways.

281 **Genes associated with fibrosis.**

282 Whilst schistosome infections cause lethargy and other non-specific symptoms, death is mainly
283 caused by the fibrosis accumulating around eggs lodged in the tissues, particularly in the
284 hepatic portal vein. A search through Pubmed abstracts and titles using the term "fibrosis" with
285 each of the 63 genes in turn identified 27 gene names that appeared in articles that also
286 mentioned fibrosis of which 13 had well documented associations with fibrosis (**Table S5**).
287 Although our study was of whole blood and schistosomiasis associated fibrosis occurs in
288 extracellular matrix, six of the 13 of the well documented DEGS associated with fibrosis had
289 been previously found to be differentially expressed in comparisons of liver fibroses with
290 healthy tissues (28) suggesting that expression data from whole blood could be informative.
291 For 11 genes it was possible to predict a direction of effect on fibrosis that would be caused by
292 a change in gene expression. The expression changes in five genes (*OGG1, OGT, ITGA4, PRMT7,*
293 *SUZ12*), were predicted to increase fibrosis in participants with high parasitaemia and the
294 remaining 6 genes (*MALAT1, TPT1, A1CF, SSPN, SUV39H1, ZNF217)* were predicted to reduce
295 fibrosis. Therefore, it is not possible to predict the risk of fibrosis in these participants from

their expression profiles. A prospective study to determine whether these genes could be useful biomarkers of morbidity risk may be more informative.

TGFB1 has been described as the master regulator of fibrosis (29). Although it was not differentially expressed in our study, seven of the 13 genes associated with fibrosis were also associated with *TGFB1*. *SUZ12* suppresses p27 (30) and p27 promotes *TGFB1* mediated pulmonary fibrosis (31). Furthermore, THBS1 modulates the TGFB1/Smad2/3 signalling pathways (32), OGG1 promotes *TGFB1* induced cell transformation and activated Smad2/3 by interacting with Smad7 (33). *ITGA4* and *TGFB1* have been found to be co-expressed in four cancer studies (34–37). Additionally, *MALAT1* has been found to modulate *TGFB1* induced epithelial to mesenchymal transition in keratinocytes (38) and *TPT1* negatively regulates the *TGFB1* signalling pathway by preventing *TGFB1* receptor activation (39). These findings indicate an interplay of fibrosis enhancers and modulators in *S. mansoni* infected children living along the Albert Nile in Uganda.

Gene expression and stunting and BMI

We identified significant differential expression of one gene, the neutral cholesterol ester hydrolase 1 (*NCEH1*) in stunting as measured by HAZ scores. This gene is involved in the breakdown of cholesterol esters in macrophages which makes them available for export and recycling to the liver (40). It is possible that in stunted children there is a higher rate of cholesterol recycling to make best use of limited lipid resources. Additionally, four DEGS were associated with BMI of which three (*MUC5B*, *DMD* and *REXO1L1P*) were upregulated whereas one (*SERPINA10*) was downregulated. The increased expression of *MUC5B* with BMI is consistent with previous observations on the role of obesity in lung function (41) since increased expression of *MUC5B* has been reported to mediate chronic obstructive pulmonary disease development through regulation of inflammation and goblet cell differentiation (42).

Changes in cell type frequency

Our analysis of the relative abundance of different cell types showed significant downregulation of Multipotent Progenitor cells (MPPs) in the children infected with *S. mansoni*. MPPs are thought to regulate HSC proliferation in response to inflammation (43) and play a role in regulation of immune response (44). *S. mansoni* antigens are known to suppress Th1 and Th17 pathways whilst stimulating Th2, B regs and T regs (45). The reduction in relative abundance of MPPs may contribute to this process.

Conclusion

Our study shows evidence of differential expression of genes in *S. mansoni* infection in children living in endemic areas. The low number of differentially expressed genes may be due to *S. mansoni* having to avoid stimulating a strong immune response in order to survive for years in the host. As such the parasite is known to suppress inflammatory responses leading to a relatively weak effect of the parasite on the host transcriptome (45). Furthermore, many of the children in this study suffered from severe stunting and most were underweight. Malnutrition is associated with increased risk of death from infections and may impair the immune response (46). Therefore, malnutrition may have reduced the response to *S. mansoni* infection and the number of differentially expressed genes. Importantly we have also identified genes that may be involved in the development of fibrosis which is the principal pathology associated with schistosomiasis. Follow up studies will be required to determine if expression of these genes correlates with the development of hepatic fibrosis in these children. We also show that there is no significant difference in gene expression between individuals with low levels of infection and the uninfected; further studies are required to elucidate this finding.

Competing interests

The authors declare no conflict of interests.

346

347 Author contributions

348 **Joyce Namulondo**: Conceptualization, Methodology, Investigation, Administration, Formal
349 analysis, Writing – Original Draft preparation; **Oscar Nyangiri:** Formal analysis, Writing –
350 review & editing; **Phillip Kimuda**: review & editing; **Peter Nambala**: Formal analysis,
351 Writing – review & editing; **Jacent Nassuuna:** Writing – review & editing**, Alison Elliott:**
352 Writing – review & editing **Moses Egesa;** Writing – review & editing; **Barbara Nerima**:
353 review & editing; **Savino Biryomumaisho**: review & editing; **Claire Mack Mugasa:** review
354 & editing; **Immaculate Nabukenya**: review & editing; **Kato Drago Charles**: review &
355 editing; **Harry Noyes:** Investigation, Supervision, Formal analysis, Writing – review &
356 editing; **Robert Tweyongyenre:** review & editing; **Enock Matovu:** Investigation,
357 Supervision, Writing – review & editing; **Julius Mulindwa**: Conceptualization, Methodology,
358 Investigation, Supervision, Formal analysis, Writing – review & editing.

359

360 **Grant Information**

361 This work was supported by Human Heredity and Health in Africa (H3Africa) [H3A-18-004].
362 H3Africa is managed by the Science for Africa Foundation (SFA Foundation) in partnership
363 with Wellcome, NIH and AfSHG. The views expressed herein are those of the author(s) and
364 not necessarily those of the SFA Foundation and her partners.

365

366 Acknowledgement

367 We do acknowledge and thank all the children and the parents/guardians that participated in
368 this study. We do appreciate the efforts by the Village health team members and the local
369 council administrators of the villages of Panyigoro, Kivuje, Nyakagei, Kayonga, Dei, Pamitu

and Alwi. Membership of the TrypanoGEN+ Research group of the H3Africa Consortium: Annette MacLeod, Bruno Bucheton, Gustave Simo, Dieudonne N. Mumba, Mathurin Koffi, Ozlem T. Bishop, Pius V. Alibu, Janelisa Musaya, Christiane Hertz-Fowler

References

1. Colley DG, Bustinduy AL, Secor WE, King CH. Human schistosomiasis. Lancet. 2014 Jun 6 ;383(9936):2253. /pmc/articles/PMC4672382/
2. WHO. Schistosomiasis and soil-transmitted helminthiases: numbers of people treated in 2019 – Schistosomiase et géohelminthiases: nombre de personnes traitées en 2019. Weekly Epidemiological Record = Relevé épidémiologique hebdomadaire.2020;95(40):629–40. https://apps.who.int/iris/handle/10665/337573
3. Tukahebwa EM, Magnussen P, Madsen H, Kabatereine NB, Nuwaha F, Wilson S, et al. A Very High Infection Intensity of *Schistosoma mansoni* in a Ugandan Lake Victoria Fishing Community Is Required for Association with Highly Prevalent Organ Related Morbidity. PLoS Negl Trop Dis. 2013:e2268. https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002268
4. Mulindwa J, Namulondo J, Kitibwa A, Nassuuna J, Nyangiri OA, Kimuda MP, et al. High prevalence of *Schistosoma mansoni* infection and stunting among school age children in communities along the Albert-Nile, Northern Uganda: A cross sectional study. WEBSTER JP, editor. PLoS Negl Trop Dis. 2022 Jul 27;16(7):e0010570. https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0010570
5. Burke ML, McManus DP, Ramm GA, Duke M, Li Y, Jones MK, et al. Co-ordinated gene expression in the liver and spleen during *Schistosoma japonicum* infection regulates cell migration. PLoS Negl Trop Dis. 2010 May;4(5). /pmc/articles/PMC2872641/

6. Burke ML, McManus DP, Ramm GA, Duke M, Li Y, Jones MK, et al. Temporal expression of chemokines dictates the hepatic inflammatory infiltrate in a murine model of schistosomiasis. PLoS Negl Trop Dis. 2010 Feb; 4(2). https://pubmed.ncbi.nlm.nih.gov/20161726/

7. Chuah C, Jones MK, Burke ML, Owen HC, Anthony BJ, McManus DP, et al. Spatial and temporal transcriptomics of *Schistosoma japonicum* -induced hepatic granuloma formation reveals novel roles for neutrophils. J Leukoc Biol. 2013 Aug;94(2):353–65.

8. Yang J, Fu Z, Hong Y, Wu H, Jin Y, Zhu C, et al. The differential expression of immune genes between water buffalo and yellow cattle determines species-specific susceptibility to *Schistosoma japonicum* infection. Langsley G, editor. PLoS One 2015 Jun 30; 10(6):e0130344. https://dx.plos.org/10.1371/journal.pone.0130344

9. Dupnik KM, Reust MJ, Vick KM, Yao B, Miyaye D, Lyimo E, et al. Gene expression differences in host response to *Schistosoma haematobium* infection. Infect Immun. 2019 Jan 1;87(1):1–11.

10. Tamarozzi F, Ursini T, Hoekstra PT, Silva R, Costa C, Gobbi F, et al. Evaluation of microscopy, serology, circulating anodic antigen (CAA), and eosinophil counts for the follow-up of migrants with chronic schistosomiasis: a prospective cohort study. Parasit Vectors. 2021 Dec 1;14(1):1–11. https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-021-04655-z

11. Mulindwa J, Leiss K, Clayton C. High-Throughput Sequencing for Trypanosome Transcriptome Characterization. Methods Mol Biol. 2020; 2116:83–98. https://pubmed.ncbi.nlm.nih.gov/32221915/

12. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Ridwan Amode M, et al. Ensembl 2021. Nucleic Acids Res. 2021;49(D1): D884–91. https://pubmed.ncbi.nlm.nih.gov/33137190/

13. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019 Aug 1;37(8):907–15. https://pubmed.ncbi.nlm.nih.gov/31375807/

14. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5 ;15(12):1–21. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8

15. Marini F, Binder H. PcaExplorer: An R/Bioconductor package for interacting with RNA-seq principal components. BMC Bioinformatics. 2019 Jun 13;20(1):1–8. https://link.springer.com/articles/10.1186/s12859-019-2879-1

16. Pomaznoy M, Ha B, Peters B. GOnet: A tool for interactive Gene Ontology analysis. BMC Bioinformatics. 2018 Dec 7;19(1):1–8. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2533-3

17. Sidiropoulos K, Viteri G, Sevilla C, Jupe S, Webber M, Orlic-Milacic M, et al. Reactome enhanced pathway visualization. Bioinformatics. 2017 Nov 1 ;33(21):3461–7. https://pubmed.ncbi.nlm.nih.gov/29077811/

18. Xie X, Liu M, Zhang Y, Wang B, Zhu C, Wang C, et al. Single-cell transcriptomic landscape of human blood cells. Natl Sci Rev. 2021 Mar 1;8(3).

19. Ahn R, Schaenman J, Qian Z, Pickering H, Groysberg V, Rossetti M, et al. Acute and Chronic Changes in Gene Expression After CMV DNAemia in Kidney Transplant Recipients. Front Immunol. 2021 Nov 15;12:4669.

20. Bouquet J, Gardy JL, Brown S, Pfeil J, Miller RR, Morshed M, et al. RNA-Seq Analysis of Gene Expression, Viral Pathogen, and B-Cell/T-Cell Receptor Signatures in Complex Chronic Disease. Clinical Infectious Diseases. 2017 Feb 15;64(4):476–81. https://academic.oup.com/cid/article/64/4/476/2919064

444 21. Gundersen GG. Microtubule capture: IQGAP and CLIP-170 expand the repertoire.
445 Current Biology. 2002 Oct 1;12(19):R645–7.
446 http://www.cell.com/article/S0960982202011569/fulltext
447 22. Charafeddine RA, Nosanchuk JD, Sharp DJ. Targeting Microtubules for Wound Repair.
448 Adv Wound Care (New Rochelle). 2016 Oct 10 ;5(10):444. /pmc/articles/PMC5067841/
449 23. Zaoui K, Duhamel S, Parachoniak CA, Park M. CLIP-170 spatially modulates receptor
450 tyrosine kinase recycling to coordinate cell migration. Traffic. 2019 Mar 1;20(3):187.
451 /pmc/articles/PMC6519375/
452 24. Bamidele AO, Kremer KN, Hirsova P, Clift IC, Gores GJ, Billadeau DD, et al. IQGAP1
453 promotes CXCR4 chemokine receptor function and trafficking via EEA-1+ endosomes.
454 J Cell Biol. 2015 ;210(2):257–72. https://pubmed.ncbi.nlm.nih.gov/26195666/
455 25. Ilan-Ber T, Ilan Y. The role of microtubules in the immune system and as potential
456 targets for gut-based immunotherapy. Mol Immunol. 2019 Jul 1;111:73–82.
457 26. Perry CR, Burke ML, Stenzel DJ, McManus DP, Ramm GA, Gobert GN. Differential
458 expression of chemokine and matrix re-modelling genes is associated with contrasting
459 schistosome-induced hepatopathology in murine models. PLoS Negl Trop Dis. 2011 Jun
460 ;5(6). https://pubmed.ncbi.nlm.nih.gov/21666794/
461 27. Sajeev T. K, Joshi G, Arya P, Mahajan V, Chaturvedi A, Mishra RK. SUMO and
462 SUMOylation Pathway at the Forefront of Host Immune Response. Front Cell Dev Biol.
463 2021 Jul 14;9:1762.
464 28. Zhan Z, Chen Y, Duan Y, Li L, Mew K, Hu P, et al. Identification of key genes,
465 pathways and potential therapeutic agents for liver fibrosis using an integrated
466 bioinformatics analysis. PeerJ.2019 ;7(3). https://pubmed.ncbi.nlm.nih.gov/30923657/
467 29. Meng XM, Nikolic-Paterson DJ, Lan HY. TGF-β: the master regulator of fibrosis. Nat
468 Rev Nephrol. 2016 Jun 1;12(6):325–38. https://pubmed.ncbi.nlm.nih.gov/27108839/

469 30. Huang W, Feng Y, Liang J, Yu H, Wang C, Wang B, et al. Loss of microRNA-128
470 promotes cardiomyocyte proliferation and heart regeneration. Nat Commun. 2018 Dec
471 1 ;9(1). https://pubmed.ncbi.nlm.nih.gov/29453456/
472 31. Dai YH, Li XQ, Dong DP, Gu HB, Kong CY, Xu Z. P27 Promotes TGF- β-Mediated
473 Pulmonary Fibrosis via Interacting with MTORC2. Can Respir J. 2019 ;2019.
474 https://pubmed.ncbi.nlm.nih.gov/31641391/
475 32. Liu K, Wang J, Gao X, Ren W. C1q/TNF-Related Protein 9 Inhibits Coxsackievirus B3-
476 Induced Injury in Cardiomyocytes through NF- κ B and TGF- β 1/Smad2/3 by
477 Modulating THBS1. Mediators Inflamm. 2020;2020.
478 https://pubmed.ncbi.nlm.nih.gov/33414684/
479 33. Wang Y, Chen T, Pan Z, Lin Z, Yang L, Zou B, et al. 8-Oxoguanine DNA glycosylase
480 modulates the cell transformation process in pulmonary fibrosis by inhibiting Smad2/3
481 and interacting with Smad7. FASEB J . 2020 Oct 1;34(10):13461–73.
482 https://pubmed.ncbi.nlm.nih.gov/32808374/
483 34. Singh R, De Sarkar N, Sarkar S, Roy R, Chattopadhyay E, Ray A, et al. Analysis of the
484 whole transcriptome from gingivo-buccal squamous cell carcinoma reveals deregulated
485 immune landscape and suggests targets for immunotherapy. PLoS One . 2017 Sep 1
486 ;12(9). https://pubmed.ncbi.nlm.nih.gov/28886030/
487 35. Friedlander P, Wassmann K, Christenfeld AM, Fisher D, Kyi C, Kirkwood JM, et al.
488 Whole-blood RNA transcript-based models can predict clinical response in two large
489 independent clinical studies of patients with advanced melanoma treated with the
490 checkpoint inhibitor, tremelimumab. J Immunother Cancer. 2017 Aug 15;5(1).
491 https://pubmed.ncbi.nlm.nih.gov/28807052/

492 36. Huang C, Nie F, Qin Z, Li B, Zhao X. A snapshot of gene expression signatures
493 generated using microarray datasets associated with excessive scarring. Am J
494 Dermatopathol. 2013 Feb;35(1):64–73. https://pubmed.ncbi.nlm.nih.gov/22785331/
495 37. Garlet GP, Horwat R, Ray HL, Garlet TP, Silveira EM, Campanelli AP, et al. Expression
496 analysis of wound healing genes in human periapical granulomas of progressive and
497 stable nature. J Endod. 2012 Feb;38(2):185–90.
498 https://pubmed.ncbi.nlm.nih.gov/22244633/
499 38. Zhang L, Hu J, Meshkat BI, Liechty KW, Xu J. LncRNA *MALAT1* Modulates TGF-β1-
500 Induced EMT in Keratinocyte. Int J Mol Sci. 2021 Nov 1;22(21).
501 https://pubmed.ncbi.nlm.nih.gov/34769245/
502 39. Pinkaew D, Martinez-Hackert E, Jia W, King MD, Miao F, Enger NR, et al. Fortilin
503 interacts with TGF-β1 and prevents TGF-β receptor activation. Commun Biol. 2022 Dec
504 1 ;5(1). https://pubmed.ncbi.nlm.nih.gov/35197550/
505 40. Igarashi M, Osuga JI, Uozaki H, Sekiya M, Nagashima S, Takahashi M, et al. The
506 critical role of neutral cholesterol ester hydrolase 1 in cholesterol removal from human
507 macrophages. Circ Res. 2010 Nov 26 [cited 2023 Mar 24];107(11):1387–95.
508 https://www.ahajournals.org/doi/abs/10.1161/circresaha.110.226613
509 41. Dixon AE, Peters U. The effect of obesity on lung function. Expert Rev Respir Med.
510 2018 Sep 2 ;12(9):755. /pmc/articles/PMC6311385/
511 42. Huang X, Guan W, Xiang B, Wang W, Xie Y, Zheng J. *MUC5B* regulates goblet cell
512 differentiation and reduces inflammation in a murine COPD model. Respir Res. 2022
513 Dec 1;23(1):1–12. https://respiratory-
514 research.biomedcentral.com/articles/10.1186/s12931-021-01920-8
515 43. Pietras EM. Inflammation: a key regulator of hematopoietic stem cell fate in health and
516 disease. Blood. 2017 Oct 10;130(15):1693. /pmc/articles/PMC5639485/

517 44. Korniotis S, D'Aveni M, Hergalant S, Letscher H, Tejerina E, Gastineau P, et al.
518 Mobilized Multipotent Hematopoietic Progenitors Stabilize and Expand Regulatory T
519 Cells to Protect Against Autoimmune Encephalomyelitis. Front Immunol. 2020 Dec
520 23;11:3342.
521 45. Cleenewerk L, Garssen J, Hogenkamp A. Clinical Use of Schistosoma mansoni
522 Antigens as Novel Immunotherapies for Autoimmune Disorders. Front Immunol. 2020
523 Aug 13;11:1821.
524 46. Rytter MJH, Kolte L, Briend A, Friis H, Christensen VB. The Immune System in
525 Children with Malnutrition—A Systematic Review. PLoS One. 2014 Aug
526 25;9(8):e105017.
527 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0105017
528

529 ## Figures

530 **Figure 1:** Screening and recruitment of children for differential gene expression. Of 727
531 children recruited for the schistosomiasis study project, 152 were recruited for gene expression
532 studies. Seventy-one (71) were excluded for not meeting the sequencing criteria hence 81
533 samples from recruited children were sequenced. One (1) sample did not pass the QC; 64
534 samples were analysed for differentially expressed genes following removal of outliers and
535 samples without CAA results.

536 **Figure 2:** PCA showed clustering of differentially expressed genes by gender when *S.*
537 *mansoni* infected children were compared to uninfected.

538 **Figure 3**: Differential expression of genes between *S. mansoni* infected and uninfected
539 individuals. **3A** Fourteen (14) significant DEGs (9 upregulated and 5 down regulated) were
540 identified among the *S. mansoni* infected children compared with the uninfected. **3B Fifty**-six

(56) significant DEGs (43 upregulated and 13 downregulated) were identified among children with high *S. mansoni* infection intensity compared to the uninfected. **3C** Thirty-three (33) significant DEGs (30 upregulated and 3 downregulated) were identified among children with high *S. mansoni* infection intensity compared to those with low infection intensity. **3D** No significant DEGs were identified among children with low *S. mansoni* infection intensity compared to the uninfected.

Figure 4: A shows differential expression of genes with stunting. One gene (*NCEH1*) was significantly upregulated in stunting. **B** shows differential expression of genes with BMI. Two genes (*MUC5B* and *DMD* were upregulated whereas one gene (*SERPINA10)* was downregulated by increased BMI.

Supplementary data

Supplementary figures

Figure S1: Principal component analysis of all sequenced samples with CAA results

Figure S2: Difference in abundance of cell types in *S. mansoni* infected compared to uninfected individuals

Supplementary tables

Table S1. GO terms and pathways associated with differentially expressed genes and infection status

Table S2: REACTOME enriched pathways for significant DEGs identified between children with high *S. mansoni* infection intensity compared to the uninfected

Table S3: Significant DEGs between children with high *S. mansoni* infection intensity compared to those with low infection intensity.

563 **Table S4**: REACTOME enriched pathways for significant DEGs identified between children
564 with high *S. mansoni* infection intensity compared to those with low infection intensity.
565 **Table S5:** DEGs that may be associated with fibrosis.
566 **Table S6:** DEGs associated with TGFB1.
567 **Table S7:** Expressed genes in stunting and BMI.
568 **Table S8:** GO terms and pathways associated with differentially expressed genes by stunting
569 and BMI.
570 **Table S9:** Cell types that differ in relative abundance between *S. mansoni* infected and
571 uninfected children.

572 ## Tables

573 **Table 1:** Summary of the study sample data by sex and *S. mansoni* infection status using
574 CAA concentration in pg/ml

Gender **Infection Intensity**	**Male (n=34)**	**Female (n=30)**	**Total (n=64)**
High (>1000 pg/ml)	17	17	34
Low (25 to <1000 pg/ml)	5	5	10
Negative (<25 pg/ml)	12	8	20
Total	**34**	**30**	**64**

575
576
577 **Table 2:** Summary of number of significant upregulated and downregulated DEGs in the
578 different comparisons

Pair (total observations)	***Details***	***DEGs***	***Up regulated***	***Down regulated***

All infected vs uninfected (n=64)	Infected: 44 Uninfected: 20	14	9 (64%)	5 (36%)
High vs uninfected (n=54)	High: 34 Uninfected: 20	56	43 (77%)	13 (23%)
High vs low infected (n=44)	High: 34 Low: 10	33	30 (91%)	3 (9%)
Low vs uninfected (n=30)	Low: 10 Uninfected: 20	-	-	-

579

580 **Table 3:** List of the significant DEGs between *S. mansoni* infected and uninfected children

GeneName	log2FoldChange	pvalue	padj	GeneType	State
CCDC168	2.27522671	4.66E-05	0.0323618	protein_coding	Upregulated
AP003498.2	1.88348776	1.67E-05	0.02174276	lncRNA	
AC115485.1	1.74002755	1.05E-05	0.02004567	lncRNA	
VN1R110P	1.46438111	8.65E-05	0.03969168	processed_pseudogene	
SRGAP3-AS2	1.29781804	3.09E-05	0.02674836	lncRNA	
AC104035.1	1.26033894	5.55E-05	0.03463305	lncRNA	
CLIP1	1.22043839	1.20E-05	0.02004567	protein_coding	
LINC00945	1.06990044	8.81E-05	0.03969168	lncRNA	
AC005153.1	1.02430899	8.67E-05	0.03969168	lncRNA	
ZNF217	-0.8746193	0.00010464	0.03980514	protein_coding	Downregulated
RN7SKP203	-1.1980276	4.13E-06	0.02004567	misc_RNA	
AC234775.2	-1.4148175	0.00011825	0.04390895	processed_pseudogene	
SUZ12	-1.5855189	3.03E-05	0.02674836	protein_coding	

TRBC2	-1.6366035	8.82E-06	0.02004567	TRCgene	

581

582

583 **Table 4:** REACTOME enriched pathways for the significant DEGs between *S. mansoni*

584 infected and uninfected children

Pathway	Pathway name	Genes	pValue	Comparison
R-HSA-5626467	RHO GTPases activate IQGAPs	*CLIP1*	0.027	IU - up
R-HSA-4615885	SUMOylation of DNA replication proteins	*CLIP1*	0.042	IU - up
R-HSA-2559580	Oxidative Stress Induced Senescence	*SUZ12*	8.23E-04	IU - down
R-HSA-2559583	Cellular Senescence	*SUZ12*	0.0025	IU - down
R-HSA-202430	Translocation of ZAP-70 to Immunological synapse	*TRBC2*	0.0164	IU - down
R-HSA-212300	PRC2 methylates histones and DNA	*SUZ12*	0.0172	IU - down
R-HSA-202427	Phosphorylation of CD3 and TCR zeta chains	*TRBC2*	0.0176	IU - down
R-HSA-389948	PD-1 signaling	*TRBC2*	0.0176	IU - down
R-HSA-8953750	Transcriptional Regulation by E2F6	*SUZ12*	0.0179	IU - down
R-HSA-3214841	PKMTs methylate histone lysines	*SUZ12*	0.0199	IU - down
R-HSA-9710421	Defective pyroptosis	*SUZ12*	0.0203	IU - down
R-HSA-202433	Generation of second messenger molecules	*TRBC2*	0.0230	IU - down
R-HSA-4551638	SUMOylation of chromatin organization proteins	*SUZ12*	0.0242	IU - down
R-HSA-8943724	Regulation of PTEN gene transcription	*SUZ12*	0.02724	IU - down
R-HSA-9645723	Diseases of programmed cell death	*SUZ12*	0.03796771	IU - down
R-HSA-388841	Costimulation by the CD28 family	*TRBC2*	0.038348851	IU - down
R-HSA-5617472	Activation of anterior HOX genes in hindbrain development during early embryogenesis	*SUZ12*	0.044809026	IU - down
R-HSA-5619507	Activation of HOX genes during differentiation	SUZ12	0.044809026	IU - down
R-HSA-202424	Downstream TCR signaling	TRBC2	0.04783657	IU - down

585

586

587

588 **Table 5:** lists the significant DEGs between children with high *S. mansoni* infection intensity compared to uninfected. A total of 56 (43 upregulated and 13
589 downregulated) DEGs between children with high *S. mansoni* infection intensity and the uninfected were observed**.**

Comparison	GeneName	log2FoldChange	pvalue	padj	GeneType	GeneDescription
High vs unifected	CCDC168	2.59295386	4.09E-06	0.00444671	protein_coding	coiled-coil domain containing 168
	ZNF385B	1.70052058	0.00014699	0.01526606	protein_coding	zinc finger protein 385B
	GLOD5	1.60135226	2.41E-05	0.00690081	protein_coding	glyoxalase domain containing 5
	SSPN	1.49913759	0.000167	0.01597466	protein_coding	sarcospan
	AP000484.1	1.48363933	0.00018252	0.01657683	processed_pseudogene	zinc finger pseudogene
	BLM	1.45917067	3.70E-05	0.0077654	protein_coding	BLM RecQ like helicase
	AC127521.1	1.45692795	0.00026549	0.02095318	lncRNA	novel transcript, antisense to SPNS3
	AC008505.1	1.4544474	0.00011933	0.01384351	lncRNA	novel transcript
	SRGAP3-AS2	1.43254479	2.45E-06	0.00418023	lncRNA	SRGAP3 antisense RNA 2
	AC104035.1	1.42114006	1.48E-05	0.00556939	lncRNA	novel transcript
	CLIP1	1.39712612	2.68E-06	0.00418023	protein_coding	CAP-Gly domain containing linker protein 1
	AC092745.1	1.37263866	0.00035827	0.02504414	lncRNA	novel transcript
	AC005703.6	1.31559586	6.95E-06	0.005411	TEC	novel transcript
	AL049548.1	1.28684828	1.27E-05	0.00554087	lncRNA	novel transcript
	SP110	1.26219484	2.53E-05	0.00690081	protein_coding	SP110 nuclear body protein
	LINC00945	1.23765997	1.59E-05	0.00559496	lncRNA	long intergenic non-protein coding RNA 945
	CHDH	1.20212922	0.00039212	0.0262338	protein_coding	choline dehydrogenase
	PRMT7	1.20008928	4.40E-05	0.00812336	protein_coding	protein arginine methyltransferase 7
	SRGAP3-AS3	1.17424132	3.02E-05	0.00742572	lncRNA	SRGAP3 antisense RNA 3
	EPN2-AS1	1.16898483	0.00015706	0.01572652	lncRNA	EPN2 antisense RNA 1
	AL031717.1	1.16649869	0.00012754	0.01433814	lncRNA	novel transcript, antisense to MAPK8IP3

	CCDC141	1.14625368	1.89E-05	0.0059506	protein_coding	coiled-coil domain containing 141
	OLIG1	1.13126775	2.32E-05	0.00690081	protein_coding	oligodendrocyte transcription factor 1
	AC037198.2	1.11971342	0.00116746	0.04762653	lncRNA	novel transcript, antisense to THBS1
	CCDC86	1.11914281	8.39E-05	0.01203565	protein_coding	coiled-coil domain containing 86
	OLIG2	1.10802485	3.06E-05	0.00742572	protein_coding	oligodendrocyte transcription factor 2
	AL049647.1	1.1042126	1.64E-05	0.00559496	lncRNA	novel transcript
	CUEDC1	1.08992873	0.00025955	0.02095318	protein_coding	CUE domain containing 1
	ACSM1	1.08276336	1.74E-05	0.00575836	protein_coding	acyl-CoA synthetase medium chain family member 1
	AC091117.2	1.0777956	0.0012653	0.04916389	lncRNA	novel transcript, antisense to SORD
	AC005153.1	1.0732048	0.00014873	0.01530049	lncRNA	novel transcript, antisense to GRB10
	AC025278.1	1.06872568	0.00030718	0.02294416	lncRNA	novel transcript, antisense to EMR4P
	AC007327.2	1.05198324	0.00048388	0.02967028	lncRNA	novel transcript
	AC124283.4	1.04452886	0.00070679	0.03652861	processed_pseudogene	ADP-ribosylation factor-like 2 binding protein (ARL2BP) pseudogene
	PGD	1.03309012	0.00056168	0.03190168	protein_coding	phosphogluconate dehydrogenase
	AC090907.1	1.03257997	1.04E-05	0.00554087	lncRNA	novel transcript, antisense to LRRK1
	FAM170B-AS1	1.02621198	6.37E-06	0.00534105	lncRNA	FAM170B antisense RNA 1
	KIF1B	1.02534951	2.85E-07	0.00103554	protein_coding	kinesin family member 1B
	AC135048.4	1.02439271	0.00020387	0.0176447	TEC	novel transcript
	AL136090.1	1.01959472	0.00082991	0.03966962	lncRNA	novel transcript
	SETP11	1.01939623	2.53E-05	0.00690081	processed_pseudogene	SET pseudogene 11
	SUV39H1	1.00861108	7.93E-05	0.01168798	protein_coding	suppressor of variegation 3-9 homolog 1
	AL132642.1	1.00219447	1.42E-05	0.00554087	lncRNA	novel transcript, antisense to ASB2
	ITGA4	-0.8029889	4.69E-05	0.00829841	protein_coding	integrin subunit alpha 4

	LINC01712	-0.804547	0.00117351	0.04762653	lncRNA	long intergenic non-protein coding RNA 1712
	OGT	-0.8985712	3.58E-05	0.00769077	protein_coding	O-linked N-acetylglucosamine (GlcNAc) transferase
	A1CF	-0.8993234	0.00099447	0.04372856	protein_coding	APOBEC1 complementation factor
	AKR1A1	-0.945478	0.00017262	0.01609319	protein_coding	aldo-keto reductase family 1 member A1
	MS4A6A	-0.9573266	0.00075701	0.03769507	protein_coding	membrane spanning 4-domains A6A
	ZNF217	-0.9864463	5.94E-05	0.00981429	protein_coding	zinc finger protein 217
	MALAT1	-1.0053315	1.12E-05	0.00554087	lncRNA	metastasis associated lung adenocarcinoma transcript 1
	TPT1	-1.0178706	0.00013564	0.01479141	protein_coding	tumor protein, translationally controlled 1
	MED7	-1.1085708	0.00034823	0.02465889	protein_coding	mediator complex subunit 7
	RNY1	-1.413451	0.00057363	0.03207941	misc_RNA	RNA, Ro60-associated Y1
	SUZ12	-1.9684361	9.71E-07	0.00264642	protein_coding	SUZ12 polycomb repressive complex 2 subunit
	SLC12A1	-2.5873493	4.89E-05	0.00834908	protein_coding	solute carrier family 12 member 1

590

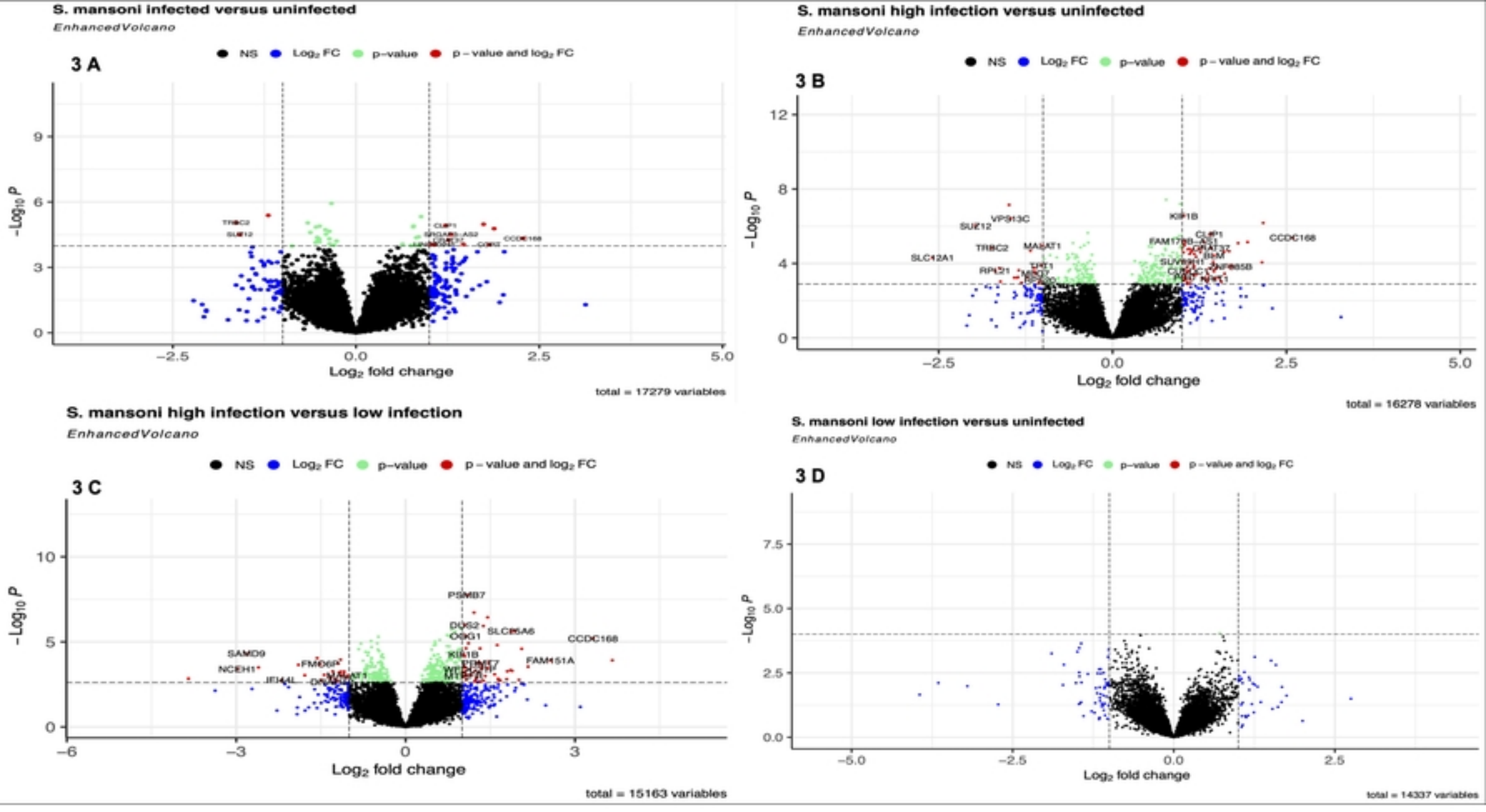
S. mansoni infected versus uninfected
EnhancedVolcano
NS
Log2 FC
p-value
p – value and log2 FC
3 A
-Log10 P
Log2 fold change
total = 17279 variables
S. mansoni high infection versus uninfected
EnhancedVolcano
3 B
total = 16278 variables
S. mansoni high infection versus low infection
EnhancedVolcano
3 C
total = 15163 variables
S. mansoni low infection versus uninfected
EnhancedVolcano
3 D
total = 14337 variables

Figure 3

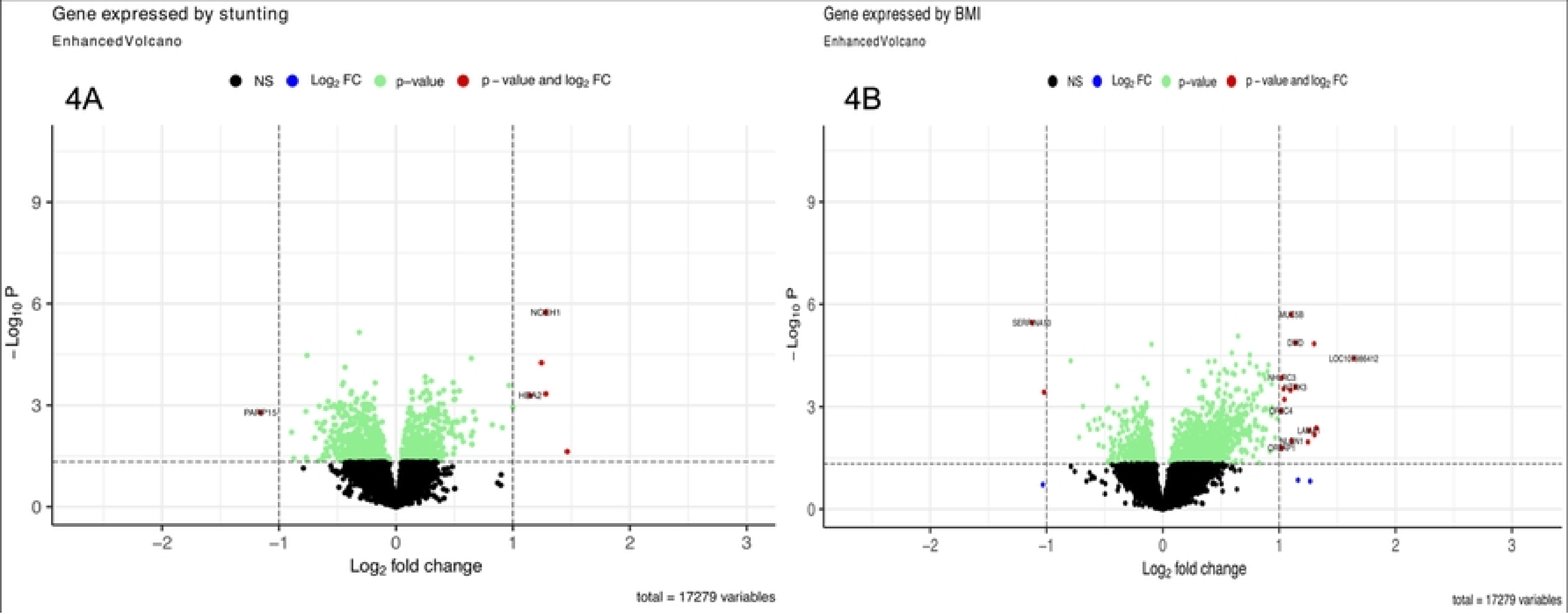
Gene expressed by stunting
EnhancedVolcano
NS
Log2 FC
p−value
p − value and log2 FC
4A
9
6
3
0
−Log10 P
NOTCH1
PAIP15
−2
−1
0
1
2
3
Log2 fold change
total = 17279 variables
Gene expressed by BMI
EnhancedVolcano
NS
Log2 FC
p−value
p − value and log2 FC
4B
9
6
3
0
−Log10 P
−2
−1
0
1
2
3
Log2 fold change
total = 17279 variables

Figure 4

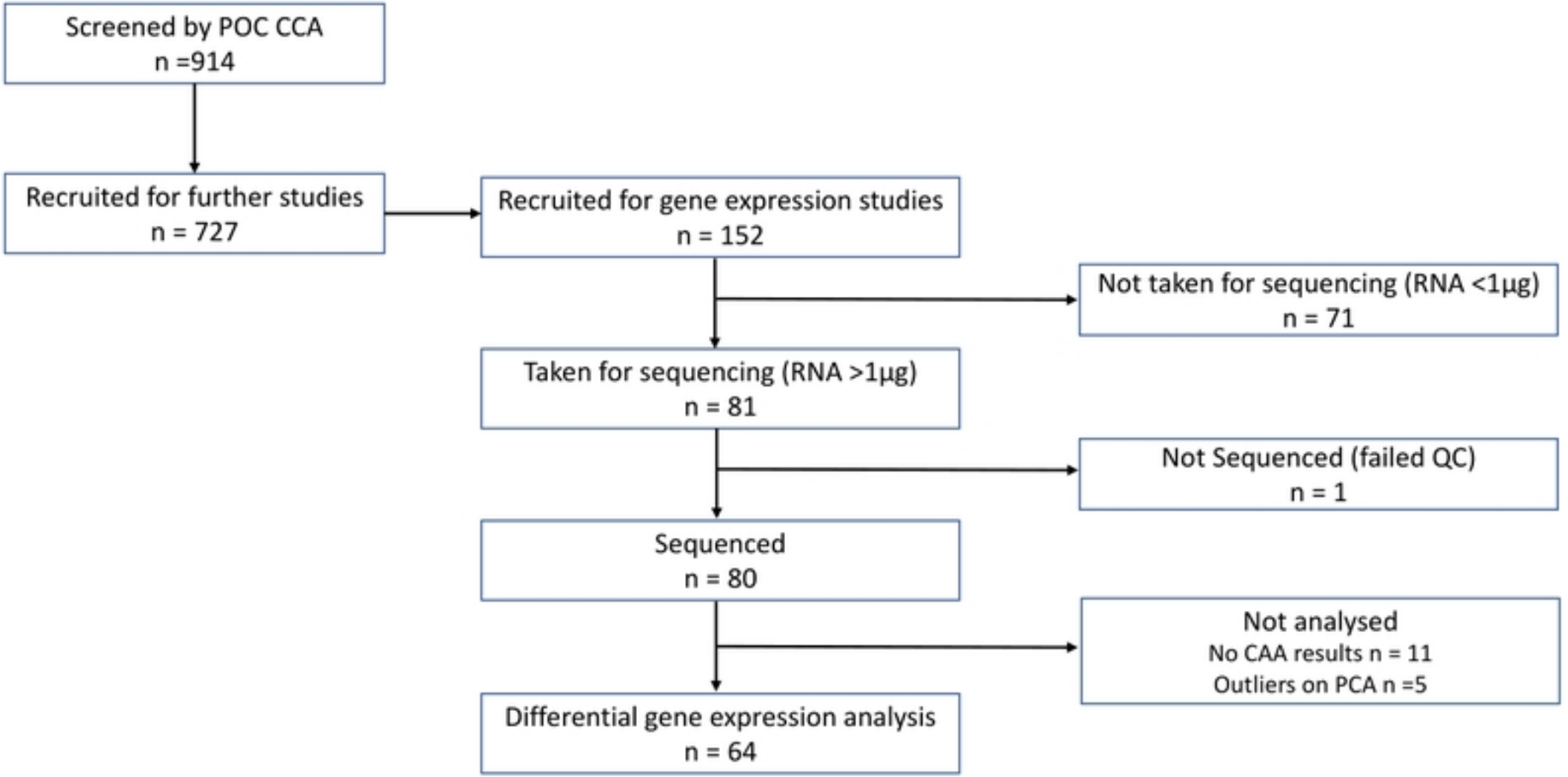
Screened by POC CCA
n =914
Recruited for further studies
n = 727
Recruited for gene expression studies
n = 152
Not taken for sequencing (RNA <1µg)
n = 71
Taken for sequencing (RNA >1µg)
n = 81
Not Sequenced (failed QC)
n = 1
Sequenced
n = 80
Not analysed
No CAA results n = 11
Outliers on PCA n =5
Differential gene expression analysis
n = 64

Figure 1

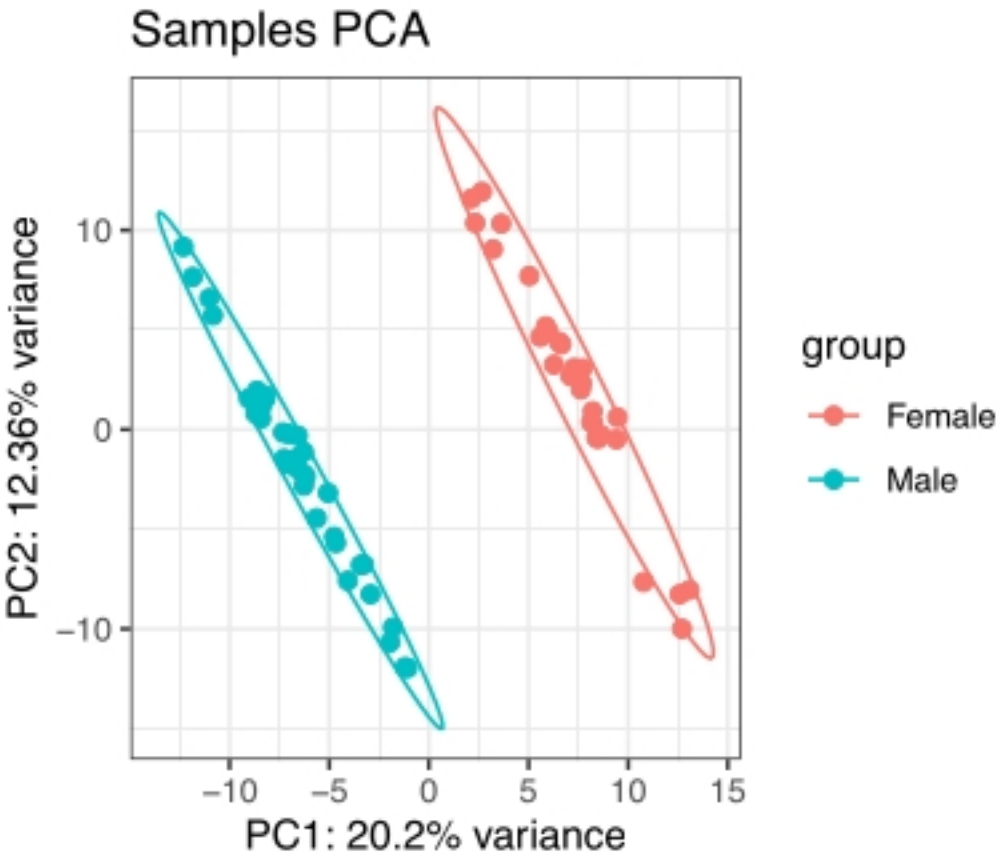
Samples PCA
PC2: 12.36% variance
PC1: 20.2% variance
10
0
−10
−10
−5
0
5
10
15
group
Female
Male

Figure 2