
# Transcriptional Comparison Between the Immune System of HIV-Exposed Uninfected And HIV-Unexposed Uninfected Infants in Kilifi, Kenya

**KIDIAVAI ZANETA MUSIMBI**

**I56/82231/2015**

**Center for Biotechnology and Bioinformatics**

**University of Nairobi**

**Thesis submitted for the partial fulfillment of the requirements of the degree**

**Master of Science in Bioinformatics**

**Submitted on 19th January, 2018**

Word count: 9,165

**Declaration**

I declare that this thesis is my own work, and has never been submitted as work of study or examined for the award of degree in any other institution of learning to the best of my knowledge.

**Kidiavai Zaneta Musimbi**

I56/82231/2015

University of Nairobi

Signature: _____ Date: _____

**Approved by;**

We confirm that this thesis has our approval to be presented to the board of postgraduate studies as per the University of Nairobi regulations.

**Dr. Eunice Nduati,**

Kenya Medical Research Institute/ Wellcome Trust Research Programme.

Signature: _____ Date: _____

**Dr. Etienne de Villiers,**

Kenya Medical Research Institute/Wellcome Trust Research Programme.

Signature: _____ Date: _____

**Dr. Isabella Oyier,**

Center for Biotechnology and Bioinformatics, University of Nairobi

Signature: _____ Date: _____

**Dedication**

I dedicate my thesis to my family without whom this thesis might not have been written.

To my loving parents, Mary Madiga and Dr. Ronald Lwegado for their love, support and words of encouragement. I would like to express my immense gratitude to my mother for always being there during those difficult and trying times throughout my life. For nurturing and helping me realize my potential.

To the Kidakwa family (Mr. and Mrs. Kidakwa, Phelim, Paula, Grainne, Darya and Nicholas), I will forever appreciate the big role they have played in my life especially Darya Kidakwa for the moral support and the push to finish the write up.

With special thanks, to my fiancé and best friend, Bernard Karani, for the moral, physical and emotional support and the many hours of proofreading. Though still an infant, to my dear son, Curtis who has filled me with profound love.

My love for you all cannot be quantified. God bless you all.

**Acknowledgments**

I wish to thank those that made this thesis possible.

The Centre of Biotechnology and Bioinformatics/University of Nairobi for accepting me into the Bioinformatics course. The lecturers for making the course informative and interesting. The KEMRI training department directed by Dr. Sam Kinyanjui for the opportunity (IDeAL studentship) to conduct my research in the institute and providing any assistance I required.

My supervisors, Dr. Eunice Nduati, Dr. Etienne de Villiers and Dr. Isabella Oyier, for their expertise, time and guidance to this final report. A special thanks to Dr. Eunice Nduati for accepting me to this project, her patience throughout the entire process and for challenging me - knowing when to push to achieve my potential.

Those that assisted in the entire process giving insightful suggestions. Grateful to James Otieno, Dr. Nelson Kibinge and Dr. Martin Rono for engaging in making sense of the programming languages on various platforms. Their excitement and willingness to provide feedback made the completion of this research an enjoyable experience

Finally my friends and colleagues; Accadeus, Dennis, Leonard, Lillian, John, Maria, Jaquie, Wandia, Rita, Shaban, Carol, Prophet and Ian as we mutually encouraged each other through the entire process.

**Abstract**

Vertical transmission of HIV-1 has reduced over the years due to the introduction of prevention mechanisms such as HAART, obstetrics management and reduced breastfeeding. This has led to a decrease in the HIV infected infant population giving rise to a new population termed as the HIV exposed uninfected infants (HEU); uninfected infants of HIV positive mothers. The HEU infant has been found to have an increased hospitalization and mortality rate relative to the HIV-unexposed uninfected (HUU) infants; uninfected infants of HIV negative mothers. Whether this is as a result of environmental exposure or intrinsic immunological alterations is not clear. This study focuses on the effect of HIV exposure on the immune system of the HEU infant that may have led to the immunological alterations.

In this study, a transcriptomic approach was used to determine whether the immunological gene expression profiles of HEU differ from that of HUU infants in the absence of an active infection. The gene expression profiles were determined using RNA samples from peripheral blood mononuclear cells (PBMCs) collected from 47 age-matched infant samples taken between 1 and 2 years of age, recruited between 2011 and 2013. Functional enrichment was done using the gene set enrichment analysis (GSEA v2.2.4) and the ingenuity pathway analysis software.

A total of 166 genes were significantly differentially expressed in HEU and HUU infants. Gene set enrichment analysis resulted in the G-protein coupled receptor signalling pathway being upregulated and the defense response to bacteria being downregulated while canonical pathway analysis indicated an altered Granzyme A signalling pathway.

It appears that exposure to HIV despite the absence of infection does contribute to distorting the HEU's immune system predisposing them to infections. This may be contributing to the prolonged hospitalization and mortality observed in HEU.

# Table of Contents

## List of Tables



## List of Figures

**List of Abbreviations**

| | |
|---|---|
| ARV: | Antiretroviral therapy |
| AZT: | antiretroviral therapy azidothymidine |
| BGI: | Beijing Genome Institute |
| CAMP: | Cathelicidin antimicrobial proteins |
| CCRC: | Comprehensive care and research clinic |
| CD4: | Cluster of Differentiation 4 |
| CD8: | Cluster of Differentiation 8 |
| DNA: | Deoxyribonucleic acid |
| FFAR2: | Free fatty acid receptor 2 |
| GBS: | Group B *Streptococcus* |
| GTF: | Gene transfer format |
| GPCR: | G-protein coupled receptors |
| GSEA: | Gene set enrichment analysis |
| HAART: | Highly active antiretroviral therapy |
| HCAR3: | Hydrocarboxyl axid receptor |
| HEU: | HIV exposed uninfected |
| HIV: | Human immunodeficiency virus |
| HMGB2: | High mobility group box 2 |
| HP: | Haptoglobin |
| HUU: | HIV unexposed uninfected |
| IgG: | Immunoglobulin gamma |
| IL: | Interleukin |
| IPA: | Ingenuity pathway analysis |
| LCN2: | Lipocalin2 |

| | |
|---|---|
| LRTI: | Lower respiratory tract infections |
| NGAL: | Neutrophil gelatinase-associated lipocalin |
| ncRNA: | non-coding RNA |
| P (MTCT): | Prevention of mother to child transmission |
| PBMC: | Peripheral blood mononuclear cells |
| PCR: | Polymerase chain reaction |
| RNA: | Ribonucleic acid |
| RseQC: | RNA-seq quality control |
| TNFSF14: | Tumor necrosis factor superfamily 14 |
| WHO: | World health organization |

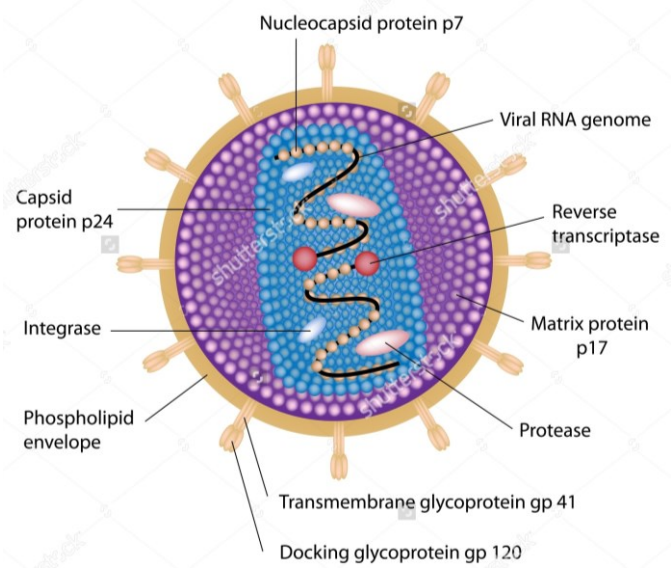
# 1.0 CHAPTER ONE: INTRODUCTION

## 1.1 Human Immunodeficiency Virus

The human immunodeficiency virus (HIV) is classified in Retrovirus family; *lentivirus* genus. HIV is known to be the cause of Acquired Immunodeficiency Syndrome (AIDS) having, allegedly, originated from the simian immunodeficiency virus (SIV) of the primates, reviewed by Paul M. Sharp and Beatrice H. Hahn (1). There are two strains of HIV: HIV-1 and HIV-2.

## 1.2 Virology of HIV-1

HIV-1, as shown in **figure 1.0**, is a double stranded retrovirus enclosed in an envelope composed of a bi-lamellar lipopolysaccharide layer with globular proteins responsible for the attachment of the virus to the antigen presenting cells (2). HIV-1 consists of 4 distinct groups; M, N, O and P. Group M is, however, the pandemic form of HIV-1 and is subdivided into 9 subtypes (A-D, F-H, J and K).

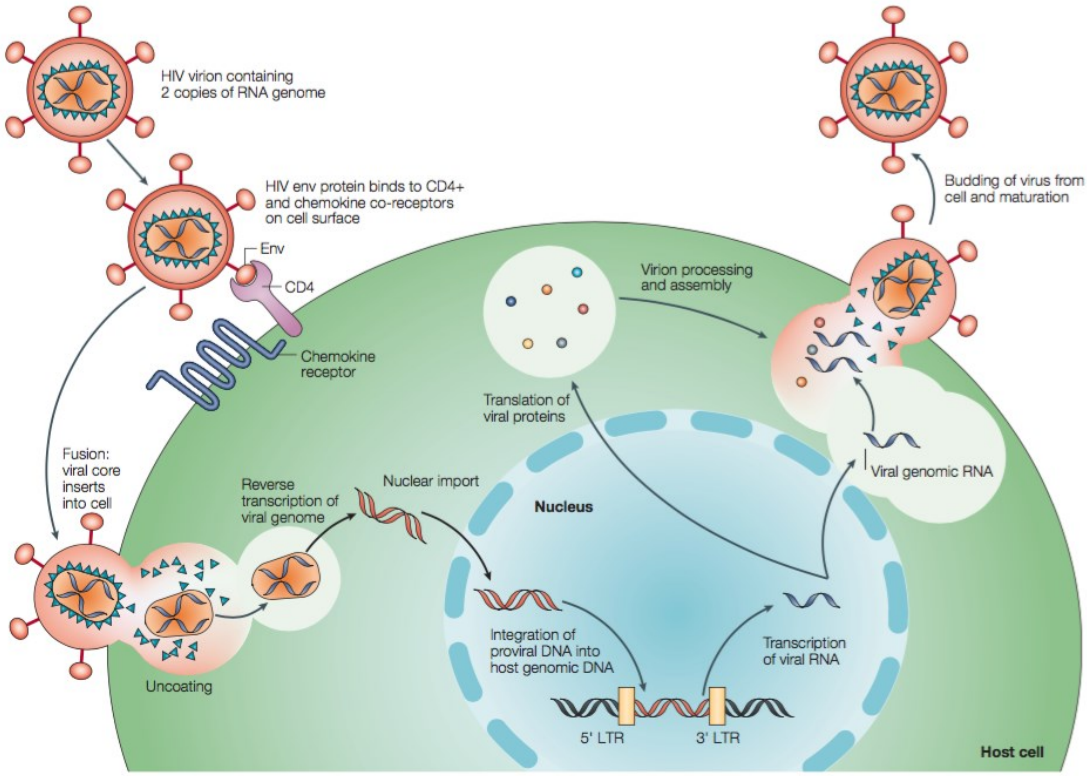



Source: Shutterstock vector images ID: 96426935

**Figure 1.0: Structure of HIV-1.** The HIV virion is enveloped with a phospholipid with globular proteins. It has two duplicate viral genomes and enzymes: a reverse transcriptase, protease and integrase.

### 1.2.1 Life cycle and phases of HIV

In summary, the HIV-1 life cycle, elaborated in **figure 1.1,** commences when the virion encounters its primary receptor, CD4, and chemokine co-receptors CXCR4 and CCR5. The CD4 receptor is mainly expressed on T lymphocytes and macrophages. Once the virion binds to the CD4 receptor, fusion between the virion and the host cell membranes is triggered via either of the co-receptors CXCR4 and CCR5. The contents of the virion are released into the host cell and viral RNA reverse transcription takes place forming a double stranded (ds) proviral DNA. Entry of the ds proviral DNA to the leads to its integration into the host genomic DNA and undergoes transcription to viral mRNA. The viral mRNA is then released into the cytoplasm whereby translation takes place to form the viral proteins. The viral proteins together with viral RNA are processed and assembled, and eventually released through budding to allow for virion maturation outside of the host cell(3)

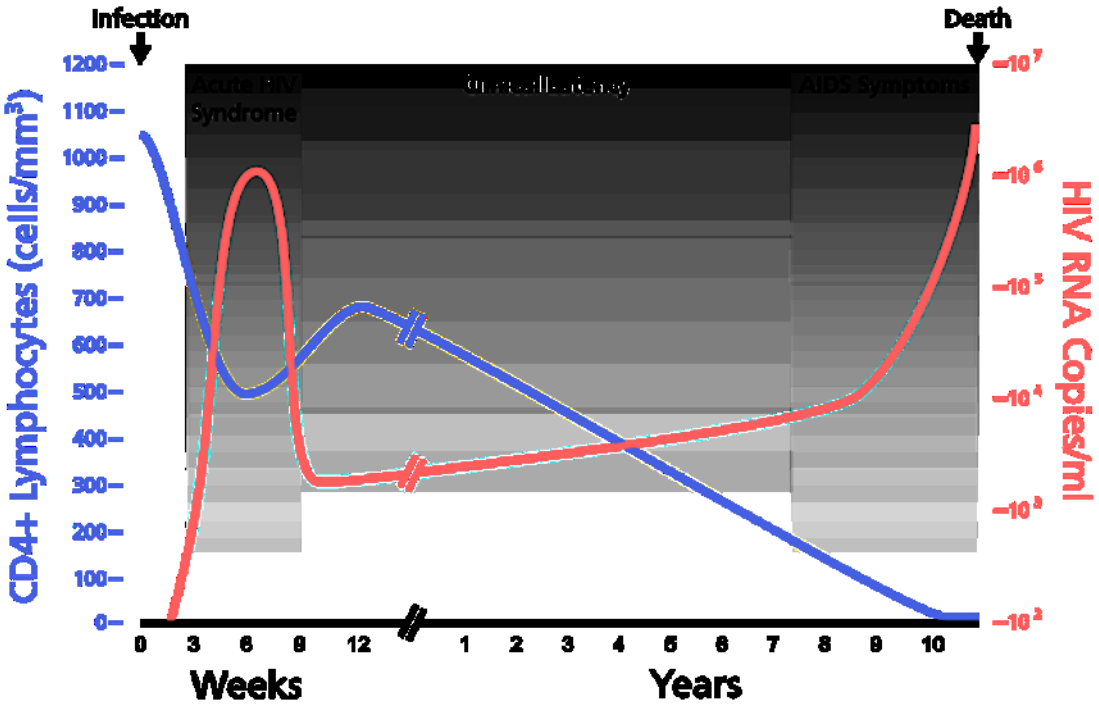



Source: Laskey & Siliciano, Nature Reviews Microbiology 12, 772-780

**Figure 1.1: HIV-1 life cycle.** The HIV viral replication takes place within the host cell through integration with the host cell DNA. The viral proteins and RNA bud out of the host cell using its membrane to form a membrane of its own.

HIV compromises the immune system by depleting the $CD4^+$ T lymphocytes, gradually weakening the body's defenses against secondary infections. The budding out of the

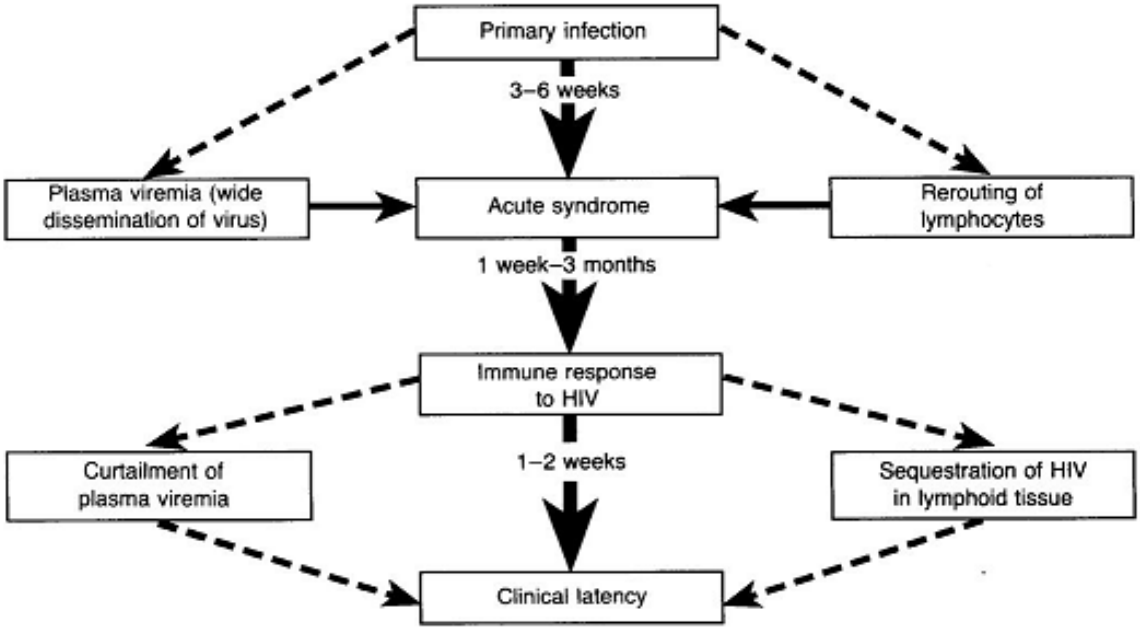
assembled and processed viral proteins and RNA uses the host's cell membrane directly lysing it partially causing the depletion effect of the $CD4^+$ T lymphocytes. The natural course of HIV-1 disease has three stages; the acute viremia, clinical latency and the clinical AIDS phase, showed below in **Figure 1.2**.




Source: Wikimedia

**Figure 1.2: Time course of HIV-1 correlating viral load and $CD4^+$ T lymphocyte.** After primary infection, there is gradual decrease in CD4 cells and increase in viremia. Once the viremia is detectable in blood plasma, there is exponential increase in HIV viremia leading to a high CD4 cell turnover. This immune response reduces the viremia ushering in the clinical latency stage of the HIV infection. The stage lasts for years as the viremia slowly increases and CD4 cells slowly decrease eventually leading to clinical AIDS. The x axis represents time in both weeks and years. The y axis in red represents viral load in copies/ml and in blue represents CD4 cell count in cells/$mm^3$.

After primary infection, the eclipse phase takes place between day 7 and day 21. This phase is known to be clinically silent; a period between the first infected cell and the first viral detection in the blood plasma. With the detection of the virus in the blood plasma, it increases exponentially ushering in the acute viremia stage. The acute viremia stage occurs after primary infection and is characterized by a high viral load with a drop in CD4 cell count. It is at this stage that replication of the HIV RNA takes place and is the main stage of transmission. Due to the high viral load, an immune response follows leading to a high turnover of $CD4^+$ T cells and a curtailment in viral load present in blood plasma resulting in the clinical latency stage (4–6). **Figure 1.3** illustrates the disease progression from primary infection to the clinical latency stage.




Source: The immunopathogenesis of Human Immunodeficiency Virus (5)

**Figure 1.3: Disease progression of HIV-1 from primary infection to clinical latency.** After primary infection, there is wide dissemination of the virus in blood plasma up to detectable levels and rerouting of lymphocytes resulting in acute HIV syndrome. Acute HIV syndrome results in a sharp increase in viremia that leads to an immune response. The immune response results in the curtailment of plasma viremia and the sequestration of HIV in lymphoid tissue. This marks the clinical latency stage.

The clinical latency stage involves the gradual decline of the $CD4^+$ T cells and a high virion production. It is at this stage that an inversion of $CD4^+/CD8^+$ T cell ratio occurs. The clinical latency stage can last for years but with the gradual decline in CD4 cell count and gradual increase in viremia, a $CD4^+$ T cell count of below 200 cells/$mm^3$ and at least an opportunistic infection marks the clinical AIDS stage. During clinical AIDS, the viremia increases overwhelming the immune system hence death as a result of opportunistic diseases and not the virus itself (4–6).

### 1.2.2 HIV epidemiology

According to the UNAIDS regional statistics factsheet (7), a total of 36.7 million [30.8M – 42.9M] people had HIV in 2016 (globally) with new infections tallying up to 1.8 million [1.6 million–2.1 million] and 1 million [830 000–1.2 million] people having succumbed to AIDS-related ailments. In Eastern and Southern Africa, 19.4 million [17.8 M – 21.1M] people live with HIV with more than half (59%) being women and girls.

#### 1.2.2.1 Impact of antiretroviral treatment on HIV epidemiology

HIV infection has no cure but it can be controlled by intake of antiretroviral drugs. There was  a global increase in antiretroviral (ARV) access as of June 2017 (20.9 million [18.4 million–21.7 million] people globally) compared to the year 2015 and 2010 (17.1 million [15.1 million–17.8 million] and 7.7 million [6.8 million–8.0 million] respectively) (7). Access to ARV medication has reduced the spread of HIV infection globally with a 16% (to 1.8M) reduction in people newly infected with HIV between 2010 and 2016. Despite the global decline in the spread of HIV, an increased HIV prevalence in women of childbearing age (15-49 years) has been observed. ARV access has also led to a 48% (1.9M in 2005 to 1.0M in 2016) decline in deaths caused from AIDS-related illnesses. Although 51% of those living with HIV globally are females, there has been a rapid decline in deaths among females when compared to males. However, it is concerning that ARV intervention has not had a significant impact on HIV incidence (8).

# 2.0 CHAPTER TWO: LITERATURE REVIEW

## 2.1 HIV transmission

HIV-1 is the global cause of HIV infections and is more aggressive compared to HIV-2 which is predominantly found in West African countries (9). A possible explanation of the lower transmission rate of HIV-2 could be because of the lower viral loads observed in a HIV-2 infected individual when compared to a HIV-1 infected individual (10,11). Plasma viremia and CD4 cell count are known to be the predictors of HIV progression. HIV transmission mainly occurs during the acute viremia phase when replication of the virion is at its peak leading to very high plasma viral loads. There are two ways of transmitting HIV, either vertically or horizontally. Horizontal transmission occurs mainly through genital fluids and infected blood. Vertical transmission on the other hand, involves transmission of HIV from mother to child.

### 2.1.1 Mother to child transmission

Adverse pregnancy outcomes such as low birth weights, spontaneous abortions, preterm-birth and, most importantly, mother to child transmission (MTCT) of the virus (12) have been reported in women with HIV. MTCT is the vertical transmission of HIV-1 infection that occurs when replication-competent virus is transmitted from the mother to the foetus or infant and establishes infection (13). The vertical transmission rates of the virus differ depending on; mode of delivery, viral phenotype, frequency and length of breastfeeding, maternal disease status and ART intake (14). Mechanisms through which MTCT could occur include: *in utero*, during labor, delivery (*intrapartum*) or breastfeeding (postnatally) (15).

## 2.2 Prevention of mother to child transmission

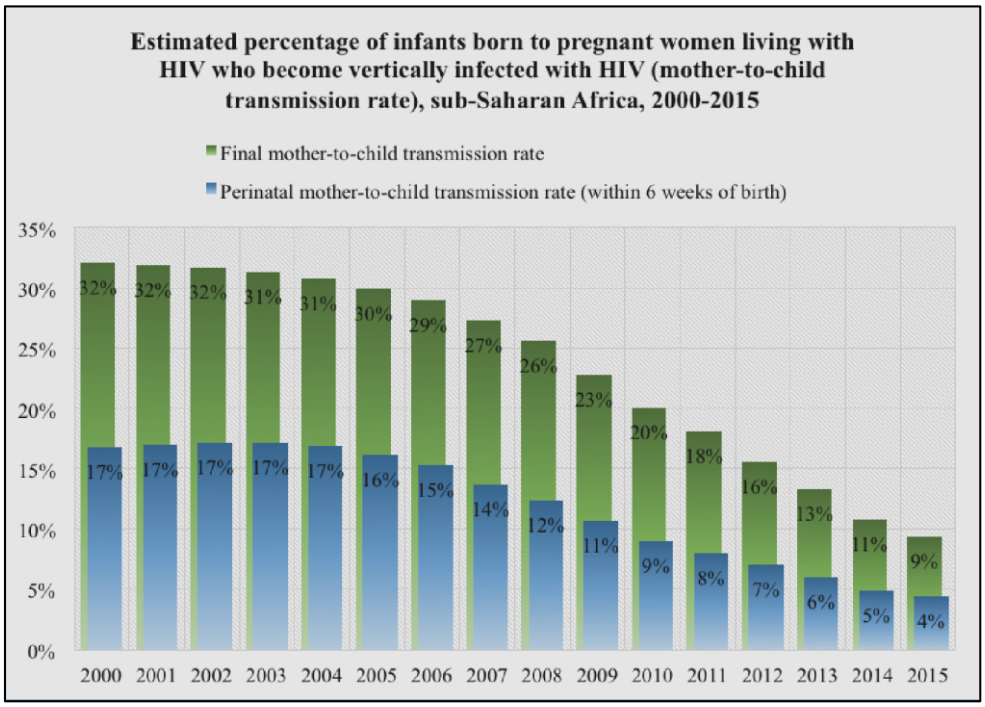
Prevention mechanisms put in place to reduce, if not eliminate, MTCT have been a success. In 2015, there was a reduction of about 35% in children who became infected with HIV, about 150,000 children (16). The risk of MTCT ranges from 20% to 45% without prophylactic intervention but if a considerable amount of intervention is put in place, the risk reduces to less than 2%. In 2016, around 76% [60–88%] of HIV positive pregnant women were able to access ARVs to prevent transplacental HIV transmission.

The Highly Active Antiretroviral Therapy (HAART) entails a combination of two or more classes of antiretroviral drugs so as to achieve maximal viral load reduction. The key risk

factor for MTCT of HIV is a high maternal viral load necessitating the use of HAART. Women who start HAART before pregnancy are able to achieve undetectable HIV RNA levels close or at delivery as compared to those who are put on therapy during pregnancy (17). In Kenya, regardless of gestation, WHO clinical stage and CD4 T cell count, HAART is initiated in all pregnant women upon diagnosis (18).

Other mechanisms used to prevent MTCT (PMTCT) include; improved obstetric management such as elective caesarian section that avoids exposure to contaminated maternal secretions during delivery and the shortened duration or avoidance of breastfeeding (19). Elective caesarian has been shown to reduce the risk of MTCT to about two-thirds independent of maternal viral load and ARV use, if the mother is on HAART the risk is reduced further. This is also true for the pregnant woman with undetectable viral loads as it has been shown that HIV is generally compartmentalized in the genital tract (17).

Breastfeeding as a way of reducing MTCT is dependent on the status of the settings, whether they are resource-rich or resource-limited (20). In resource rich settings, total avoidance of breast feeding has been used to further reduce the chances of MTCT, however in resource-limited settings lack of affordable, feasible, accessible, safe and sustainable alternatives limits total avoidance of breastfeeding. Exclusive breastfeeding, on the other hand, avoids increased risk of death from malnutrition, diarrhea and pneumonia. These mechanisms have reduced vertical transmission to < 2 % in developed countries with remarkable progress being observed in the developing countries (15) leading to an emergence of a new infant population, HIV-exposed uninfected. **Figure 2.0** shows the reduction in vertical transmission of HIV-1 from the year 2000 to 2015 indicating the success observed after the introduction of prevention mechanisms.




Source: UNICEF analysis of UNAIDS 2016 estimates.

**Figure 2.0: Estimated percentage of mother to child transmission rate of HIV-1 in sub-Saharan Africa between 2000 and 2015.** Mother to child transmission can be either through *in utero* or perinatal. The final mother to child transmission rate is indicated by the shade of green while the transmission rate within 6 weeks of birth is indicated by the shade of blue.

## 2.3 The HIV-exposed but uninfected (HEU) infant

Between 2009 and late 2014, there was a drop in children infected by HIV through MTCT by over 60% in many countries; these countries are close to virtual elimination of MTCT as a public health concern. With improved interventions leading to the large reduction of pediatric HIV, the proportion of uninfected infants of HIV-infected mothers (HIV exposed uninfected [HEU] infants) is bound to rapidly increase.

The recommended cascade to ascertain presence/lack of infection in infants born to HIV infected mothers involves; HIV testing by polymerase chain reaction (PCR) 6 weeks after birth, which if negative for HIV DNA, is followed by an antibody test at 9 months and then a confirmatory antibody test at 18 months (21). All HEU infants are placed on a short ART course between birth and six weeks when it is ascertained, through PCR, that they were not infected with HIV, either in utero or during, delivery. If uninfected then, ART is stopped (15) but if infected; at any point of the test cascade, the infant is placed on life time ART.

### 2.3.1 Increased morbidity and mortality in HEU infants

The HEU infants have been shown to have a high rate of prolonged hospitalization and mortality as relative to the HIV-unexposed uninfected (HUU) infants (22,23), a mortality rate of about 4-fold higher. With the rapidly increasing HEU infant population, the increased rate of hospitalization and mortality has become a public health concern. Several reviews have been done on the possible causes of the higher morbidity and mortality rates of the HEU when compared to the HUU infant (20,24). The ZVITAMBO study by Koyanagi et al (25) confirmed a Zambian study (26) that HEU infants whose mothers had severe disease and immune dysfunction had a higher morbidity and mortality rate. The HEU infants with mothers who had <800 CD4 cell counts also had a significantly higher all-cause sick clinic visit rate.

During the first 12 months of the HEU infants' life, the risk of serious bacterial infection is inversely proportional to maternal CD4 values during pregnancy (27). Lower respiratory tract infections (LRTI) and infections caused by encapsulated pathogens, including *Haemophilus influenzae* and *Streptococcus pneumonia* were the main cause of the serious bacterial infection (27,28). The infants experience elevated incidence of LRTI hospitalization as early as <6 months from birth, with prolonged hospitalization and possibly death relative to HUU infants (29,30).

Invasive pneumococcal disease and its associated case fatality rate is increased in the HEU infants, especially as young infants of <6 month of age (31,32), such is the case with group B streptococcal infection (33).Two studies done in Botswana reported that the higher mortality rate of the HEU infants were as a result of infectious diseases, especially the diarrheal diseases and the LRTIs (34) and treatment failures among children with pneumonia(32). The diarrheal disease rate was shown to be similar to those experienced with HUUs but were often severe leading to hospitalization (35).

Although breastfeeding has major benefits on growth and infant immunity, HIV-infected mothers may expose their infants to HIV virions. This fact has made breastfeeding by infected mothers controversial, with some being advised to avoid breastfeeding or breastfeed for a shorter duration. Reduced breastfeeding duration however has an impact on the mortality rate with HEU infants breastfed for shorter durations having a higher mortality rate through the 18 months of life (36,37) while continued breastfeeding had a

significant effect on the reduction of morbidity and mortality cases associated with diarrhoea(38).

It is also possible that the high rates of morbidity and mortality in the HEU infants are as a result of socio-economic factors such as poverty, unhygienic conditions, and lack of attention from an ailing infected mother, but may also be caused by defects in the developing infant's immune system. My study focussed on 'defects in the infant's developing immune system' as a possible cause of the high rates of morbidity and mortality.

**2.4 The immune system of the HEU infant**

Studies have been done to assess the immune system of the HEU infant in association to the increased morbidity and mortality in a bid to answer the question "Does exposure to HIV compromise the development of the immune system in HEU infants leading to the increased hospitalizations and death observed in these infants?" The HEU infants may have an impaired physiologic immune system development (39) presenting many differences in cellular mediated immune parameters at birth that persist for at least a year (40) hence dubbed as a population with a vulnerable immune system (41).

It is speculated that viral particles may traverse the placental membrane. This intrauterine exposure is suggested to mostly be due to the trans-placental diffusion of HIV soluble proteins and not live and replication-competent HIV. More than 20% as reviewed by Dauby and colleagues (42) of HEU infants have been shown to possess an immunological memory to HIV proteins (43,44). Analysis of antigen- stimulated interleukin-2 (IL-2) confirmed this when they observed high frequencies of HIV specific $CD4^+$ T helper cells in the HEU new-borns (39).

Due to the fact that the immune system develops during intrauterine life, it has been hypothesized that exposure to the HIV soluble proteins sensitizes the immune system leading to a negative impact on intra-thymic maturation and the selection of T lymphocytes (39,40) and an impairment of T-lymphocyte function and/or development would be expected (39). This impact on foetal immune maturation and responses to novel antigens in infants (45) could possibly lead to the modification of the immune functions of HEU infants (24). When compared to HUU infants, the HEU infants present an enhanced expression of CD40L on activated T lymphocytes (46), higher numbers of $CD3^+$ cells (47), a complex pattern of defects in $CD4^+$ and $CD8^+$ T-lymphocyte subpopulations, with a shift

from naïve to memory phenotypes, a peripheral increase in immature T lymphocytes (39,40,48) and altered dendritic cells (46). Lower naïve CD4 counts is presumed to be as a result of a reduced thymic output due to impaired progenitor cell function (49).

Contrarily, no difference in naïve and memory $CD4^+$ and $CD8^+$ T-cell distributions were observed between HEU and HUU infants (50). It is possible that the use of antiretroviral therapy in the mothers of these infants may have reduced exposure of HIV antigens in these infants in turn diminishing any naïve T-cell priming. However, other studies have differences in the immunologic profile of $CD4^+$ and $CD8^+$ T cells, with a shift from naïve to memory phenotypes and higher activation denoted by CD38 expression in HEU neonates despite the fact that the mothers had undergone antiretroviral prophylaxis therapy (40,51).

A study conducted at Cape Town demonstrated that the T cells of HEU infants have the inability to produce functional cytokine responses despite the fact that they proliferate more readily upon activation with key antigens; Bacillus Calmette-Guerin (BCG) and acellular pertussis, and staphylococcal enterotoxin B (SEB). The HEU infants, as compared to their counterparts, had significantly lower frequencies of IL17 and IL12 secreting poly-functional $CD8^+$ T cells (45).These abnormalities in cell mediated immunity and T-cell development result in lower CD4 counts in HEU infants (46).

Similarly the humoral arm of the adaptive immune system is affected in HEU infants with a significant reduction in the placental transfer of maternal-specific antibodies in women infected with HIV (52). HEU infants had an increased vaccine response to pertussis and pneumococcus/ pneumococcal conjugate as compared to their HUU counterparts (53,54). This could have been attributed to the fact that they had reduced trans-placental antibody transfer. This reduced transplacental antibody transfer leads to low levels of antibodies at birth compromising infant immunity, hence, the HEU infant tends to be more prone to initial infections such as encapsulated bacteria (24,27). Low proportions of *Haemophilus influenzae* type b-specific IgG (Hib-IgG) antibody were observed in the new born HEU infants (55,56) increasing their vulnerability to the disease. Any change in the immune response towards vaccines in the HEU infants could contribute to the increased morbidity and mortality. Impaired humoral response has also been reported in HEU infants as a result of both HIV and ARV exposure. These infants had a poor response (anti-HBs titre, <10 mIU/ml) to hepatitis B vaccine with lower geometric mean antitoxin titres to tetanus (46) increasing their susceptibility to these diseases.

No evidence of $CD19^+$ B cell subset alteration was observed in HEU infants in a study done in Malawi suggesting no alterations in the B cell compartment (48). In contrast, an increase in cord blood B lymphocytes driven by an increase in $CD19^+/CD5^+$ cells (47) and advanced absolute numbers and CD19 B cells percentages at 3 to 12 months of age (57) have been reported. HIV exposure was also shown to result in a reduction in the resting memory B cells, largely due to the changes in the unswitched memory B cells (56). However this did not alter the concentrations of antibodies specific to vaccine antigens such as diphtheria toxoid, tetanus toxoid, pneumococcal capsular polysaccharides in HEU infants as compared to the HUU infants after 18 months (56), in agreement with a previous report (46), implying that these infants generated lasting responses to antigens they had previously been vaccinated. HEU were also reported to present a higher B cell apoptosis level (58) when compared to the HUU infant. It therefore appears that exposure to HIV even in the absence of infection may compromise the developing infant's immune system.

## 2.5 Rationale

PMTCT programs have been a huge success globally (15) and in rural Kenya too(59). The consequence of this dramatic success, especially in areas with a high HIV prevalence, is the ever increasing emergence of infants who have been exposed to ART drugs (60) and to the viral antigens or virus, either *in utero* or through breastfeeding, and but are themselves not infected. Despite not being infected, exposure has been associated with high morbidity and mortality (61–63) making it a public health concern. Several studies have reported on the altered immune responses in HEU infants in comparison to the HUU infant, reviewed by Filteau (24) with probably the T-lymphocyte compartment being the most affected. The debate on the potential role of HIV exposure on the developing immune system and whether these defective immune responses contribute to the observed high morbidity and mortality in HEU infants is however inconclusive.

Determining if at all there is a difference in the HEU developing immune system and understanding whether indeed the immunological defects play a role in driving increased morbidity and mortality will be important in the design of potential control measures. Numerous studies have concentrated on differences in the phenotypic and the functional aspect of the immune system, with controversial outcomes being reported(39–43,45–51,53,55–58,64,65). With improved technology, peripheral blood transcriptome profiling may provide a more definitive approach that may identify immunological differences between HEUs and HUU infants. The aim of this study was to determine whether using a transcriptomic approach would reveal differences in the expression profiles of genes relevant for immune response between the HEU and the HUU infants.

### 2.5.1 Null Hypothesis

Expression profiles of genes relevant for immune responses do not differ between HEU and HUU infants.

## 2.6 Objectives

### 2.6.1 General Objective

The overall objective of this study is to determine whether expression profiles of genes relevant for immune responses differ between HEU and HUU infants.

### 2.6.2 Specific Objectives

1. To design an RNA-Seq analysis pipeline to be used in determining whether expression profiles of genes relevant for immune responses differ between HEU and HUU infants.
2. To identify differentially expressed genes in age-matched HEU and HUU infants in the first two years of life using the RNASeq analysis pipeline developed in specific aim 1 above.
3. To compare gene expression profiles between HEU and HUU infants for genes relevant to the developing immune system.

## 2.7 Research Question

1. What tools and software can be used to generate an RNASeq analysis pipeline that would effectively determine whether expression profiles of genes relevant for immune responses differ between HEU and HUU infants?
2. Are there any genes that are differentially expressed in age-matched HEU and HUU infants in the first two years of life?
3. Do expression profiles of genes relevant for immune responses differ between HEU and HUU infants?

# 3.0 CHAPTER TWO: METHODOLOGY

## 3.1 Study site and participants

This sub-study is nested within a parent study that was conducted at the comprehensive care and research clinic (CCRC), Kilifi County Hospital (KCH) situated on the Northern coast of Kenya (59).

Infants born to HIV-infected mothers between November 2011 and December 2012 were recruited and followed up longitudinally during the first two years of their life to assess the impact of HIV exposure on the infant's developing immune system (56). The infants' HIV-1 status were determined by polymerase chain reaction (PCR) at six weeks of age or at first contact and by rapid antibody test at 9 weeks, and a confirmatory antibody test at 18 months of age to ascertain HIV negativity. Community control HUU infants age-matched to the various sampling time points were recruited from three localities within the catchment area of Kilifi County. (50,56). Due to ethical reasons community controls were sampled at a single time-point and could not be followed up longitudinal. Forty-seven samples from the total number of samples collected during the parent study were made available for the current sub-study. Of these, 32 were from infants born to HIV-1 positive mothers and 15 were from infants born to HIV-1 negative mothers. Samples were collected from 19 of the HEU infants (boys, n=13) and 14 age-matched HUU infants (boys, n=8) in this cohort. These 19 HEU infants contributed 32 samples, 18 at an early time point (approximately 12 months) and 14 at a late time point (approximately 24 months). The 14 HUU infants contributed 15 samples.

Routine clinical data for the HIV infected mothers were available as they had been receiving HIV care at this clinic as per the current HIV care guidelines at that time. HIV infected pregnant women with CD4 count less than 350 cells/$mm^3$ were placed on HAART and those with higher CD4 levels were placed on prophylactic antiretroviral therapy azidothymidine (AZT) during pregnancy.

### 3.1.1 Ethical considerations

The protocol was approved by the Kenyan Medical Research Institute Ethical Research Committee (protocol no.SSC 2085). Written informed consent were also obtained from mothers or guardians of the infants recruited.

### 3.2 Sample collection and processing

Sample collection had previously been reported (50,56). In brief, 5ml whole blood was drawn and processed within four hours. 300µl of the whole blood was used for RNA extraction, a process that was carried out in two parts: initially lysing the red blood cells and storing PBMCs in buffer and downstream RNA extraction completed by Beijing Genomics Institute (BGI) China, where the RNA sequencing was done. The sequencing raw data were then made available for the current project.

In brief the processes followed during the RNA extraction and sequencing were as follows:

### 3.3 RNA extraction

Total RNA was isolated from whole blood using the Qiagen RNeasy kit. Following the Qiagen protocol, the blood cells were disrupted by lysis and homogenization done in the presence of a highly denaturing guanidine-thiocyanate-containing buffer, which immediately inactivates RNases to ensure purification of intact RNA. Ethanol was added to the lysate and the sample applied to an RNeasy Mini spin column where the total RNA bound to the membrane and contaminants efficiently washed out. Total RNA was then eluted in 30µl -100µl water. All steps were performed by centrifugation in a micro centrifuge. The quality of total RNA extracted was then ascertained using the Agilent Bioanalyser.

### 3.4 Library preparation and sequencing

DNA library preparation and sequencing was outsourced to BGI, China. Library construction (200bp short-insert library) was done according to manufacturer's recommendations after enrichment of mRNA. RNA sequencing was done using the Hi-seq Illumina platform whereby the service provider generated RNA-Seq data using nine lanes of single ended reads at a read length of 50 base-pairs (50bp).

In brief; mRNA enrichment was done by annealing total RNA to oligo-dt beads. The mRNA was then incubated with fragmentation reagent and the fragments primed with random hexamer primers. cDNA fragments were produced by reverse transcription of the primed mRNA fragments and the complementary strands of the cDNA fragments synthesized, forming double stranded cDNA (dscDNA). The RNA was discarded and the dscDNA purified from free nucleotides, enzymes, buffers and RNA. The ends of the dscDNA were then repaired. End-repaired dscDNA were adenylated and adaptors ligated

using a different index for each library. The library was then amplified using polymerase chain reaction (PCR) and libraries validated and quality control done. The libraries were then diluted to the same concentration (normalized) and pooled for sequencing.

After library preparation, cDNA sequencing was done. The ds cDNA was passed through a flow cell whereby individual molecules were hybridized based on their complementarity with adapter sequences. The hybridized sequences then formed a bridge with the next adapter after which they were amplified. This happened in many cycles creating a cluster. After cluster generation, one strand was removed from the ds cDNA and reagents introduced to allow for sequencing by synthesis. During this synthesis a single nucleotide was added at each round and the exact nucleotide determined by a fluorescent signal. Data filtering was done and then converted into fastQ format for bioinformatics analysis.

### 3.5.1 Bioinformatics analysis

FastQC software was used to check for the quality of the raw reads. These fastq files were then aligned to the splice-aware Tophat2 software (66) using the human reference and gene transfer format (GTF) file GRCh38vs86. The estimated abundance levels were then quantified at exon level using HTSeq (67). These quantified abundance estimates were then combined into a DESeqDataSet object and normalization done on them using the geometric mean method incorporated in the DESeq2 package (68). This was done to control the difference in the sequencing depth and the quality of the samples. Differential gene expression of the samples was analysed using a negative binomial distribution model as per the DESeq2 package.

### 3.5.2 Functional enrichment analysis

GSEA v2.2.4, Broad institute, open source software, was used to perform gene set enrichment. The input was a ranked list of differentially expressed genes obtained from DeSeq2 at a padj of <0.05 ordered from the most positive to the most negative. The enrichment score (ES) was calculated and its significance level estimated by the nominal p value calculated by permuting the genes 1000 times. To account for the size of the set, the ES for each gene set was normalised to obtain a normalised enrichment score (NES). The proportion of false positives was then controlled through calculating the false discovery rate (FDR) corresponding to each NES. The Molecular Signatures Database (MsigDB) catalogue c5.bp.v6.0 that includes 4436 gene sets for gene ontology was used. Cytoscape software v.3.5.0 was used to visualise the GSEA output. Enrichment map plugin was used

to draw networks using thresholds of p value < 0.05, FDR < 0.05 and Jaccard + overlap combined < 0.5.

Canonical pathways associated with the HEU infant's expression profiles were determined using the Ingenuity Pathway Analysis (IPA) software. The statistically significant differentially expressed genes (166 genes in total) were used as input in the IPA software and mapped to the Ingenuity knowledge base. The threshold used to determine the significant altered pathways was a p-value < 0.05.

#### 3.5.2.1 Interpretation of the functional enrichment analysis results

The networks drawn by the cytoscape software using the enrichment map plugin were then used to isolate the immune system process from the other biological processes. Within a network, colored nodes are used to identify whether the enrichment is upregulated or downregulated. The nodes represent gene ontologies formed from gene sets. A gene set contains a group of genes that perform the same function .The gene ontologies within the immune system were then isolated and the gene sets identified. Those that had some of the 166 differentially expressed genes were given priority.

# 4.0 CHAPTER THREE: RESULTS

## 4.1 Infant and Mother Characteristics

HIV exposed infants were recruited at the CCRC, KCH in a parent study that is described elsewhere (59). In the parent study, 97 HIV exposed uninfected infants were recruited and followed up longitudinally and some of these infant samples used to describe T and B cell phenotypes and functions (50,56). In addition, cross-sectional recruitment of 98 age-matched HIV unexposed infants was done as a comparative group. The current study is nested in this parent study and due to the high cost of whole transcriptome profiling analysis, only a small number of infants were included in this sub-study as opposed to the entire parent study participants. Due to ethical reasons it was not possible to follow the HUU infants longitudinally and only one HUU infant contributed two samples, one at an early and the other at a late time point. The remaining 13 HUU infants provided a single sample, 10 at an early time point and three at a late time point.

For the HIV exposed infants, HIV care data routinely collected during their clinic visits was available. Overall, the infants were well nourished with a MUAC median 14.1 [IQR 13.6 – 15] and were all on Cotrimoxazole prophylaxis until 18 months of life. Maternal clinical data for the HEU infants was also available. At the time of the study, HIV infected adults were placed on ART if their CD4 counts were below 350cells/$mm^3$, based on this we categorized maternal CD4 counts of the 19 mothers, six mothers had a CD4 count below 350cells/ $mm^3$ while eight mothers had a CD4 count above 350cells/ $mm^3$. Six mothers had viral loads that were less than 300copies/ml while 12 had viral loads greater than 500copies/ml. Of the 19 mothers, nine had been on HAART for more than two years before childbirth while six were not on HAART prior to starting PMTCT. However, all mothers were placed on HAART once they begun PMTCT care. The 19 mothers were well nourished and had BMI median 22.55 [IQR 19.65 – 23.58]. **Table 4.1** describes the characteristics of the HEU infants and their mothers.

**Table 4.1: Characteristics of HIV-exposed uninfected infants and their HIV-1 infected mothers**

| **Infant characteristics** | | **N = 19** |
|---|---|---|
| Gender | Male | 13 |
| | Female | 6 |
| Age at sample collection (Months) | Early time point (IQR) | ~12 (12.02 – 12.6) |
| | Late time point (IQR) | ~ 24 (17.82 – 21.05) |
| MUAC | Median (IQR) | 14.1 (13.6 – 15) |
| | | |
| **Mothers characteristics (at infant's birth)** | | **N = 19** |
| BMI | Median (IQR) | 22.55 (19.65 – 23.58) |
| CD4 count (no. of cells/$mm^3$) categories | <350 | 6 |
| | >350 | 8 |
| | Missing* | 5 |
| Duration on continuous HAART before infant's birth (Months) | Not on HAART | 6 |
| | On HAART | 9 |
| | Missing* | 4 |
| Viral load (no. of copies/ml) | <300 | 6 |
| | >500 | 12 |
| | Missing* | 1 |

BMI: Body mass index, HAART: Highly active antiretroviral therapy, IQR: Interquartile range
Mothers not on continuous HAART before infant's birth received prophylactic antiretroviral therapy azidothymidine (AZT) during, through and after delivery.
Mothers on continuous HAART did not receive any PMTCT prophylaxis
*Missing refers to mothers whose data was missing from the records

### 3.5 Expression data analysis

The Bioinformatics analysis to determine the differentially expressed genes was done using a pipeline that is a bash script in which R script and ssh script have been integrated. It makes use of the command line in any -Unix based system. This is represented in **Appendix I and II.**

### 4.2 Determining whether gene expression profiles in PBMC samples taken at the early time point (~12 months) differ from those taken at late time point (~24 months) in HEU infants

Immunological changes occur very rapidly during the first two years of life when the infant's immune system is developing. I therefore compared gene expression profiles in paired samples taken at 12 months with those taken at 24 months. Of importance there were no differences in the gene expression profiles for HEU samples taken at 12 months or at 24 months, with only 9 differentially expressed genes at p.adj < 0.05 and an absolute log 2-fold change threshold of 1(**Figure 4.0a**).

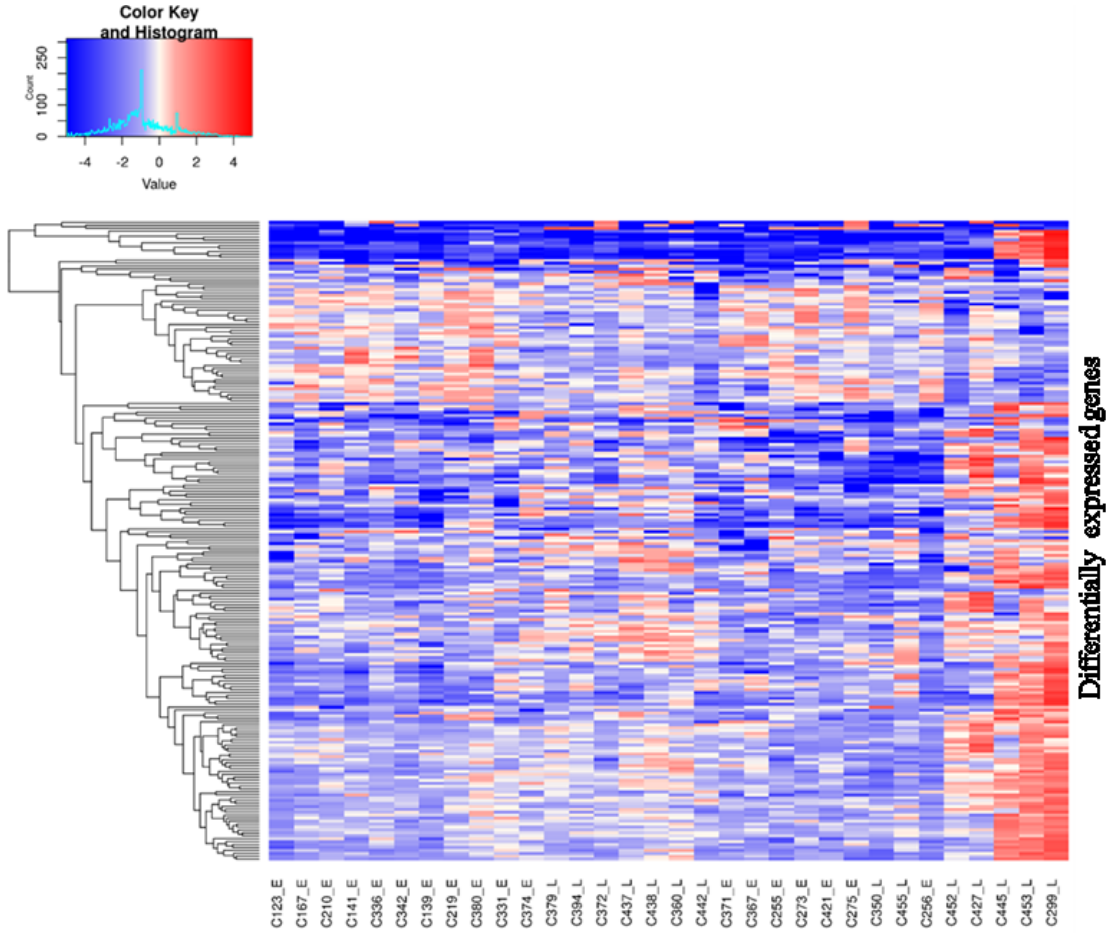



**Figure 4.0a: Differentially expressed genes in HEU infants at 12 and 24 months.** Heatmap of the differentially expressed genes in HEU (C) at 12 (E) and 24 (L) months at a threshold of Padj of <0.05 and Log2FoldChange >|1. The dendogram indicates similarity in gene expression in the HEU infants in the first two years of life. Red indicates upregulated genes while blue indicates downregulated genes in HIV-exposed uninfected infants. The y axis represents the differentially expressed genes and the x axis represents the sample ID.

However, there were differences in the expression profiles of HEU vs HUU at both time points, 12 months (**Figure 4.0b**) and 24 months (**Figure 4.0c**) with 211 and 308

differentially expressed genes at p.adj < 0.05 and an absolute log 2-fold change threshold of 1 respectively. Based on these observations, further analysis was done on combined expression profiles for 12 and 24 months in both HEU and HUU infants. Using samples collected at an early and late time point possibly provides a better coverage of the events that occur in the first 2 years of life.

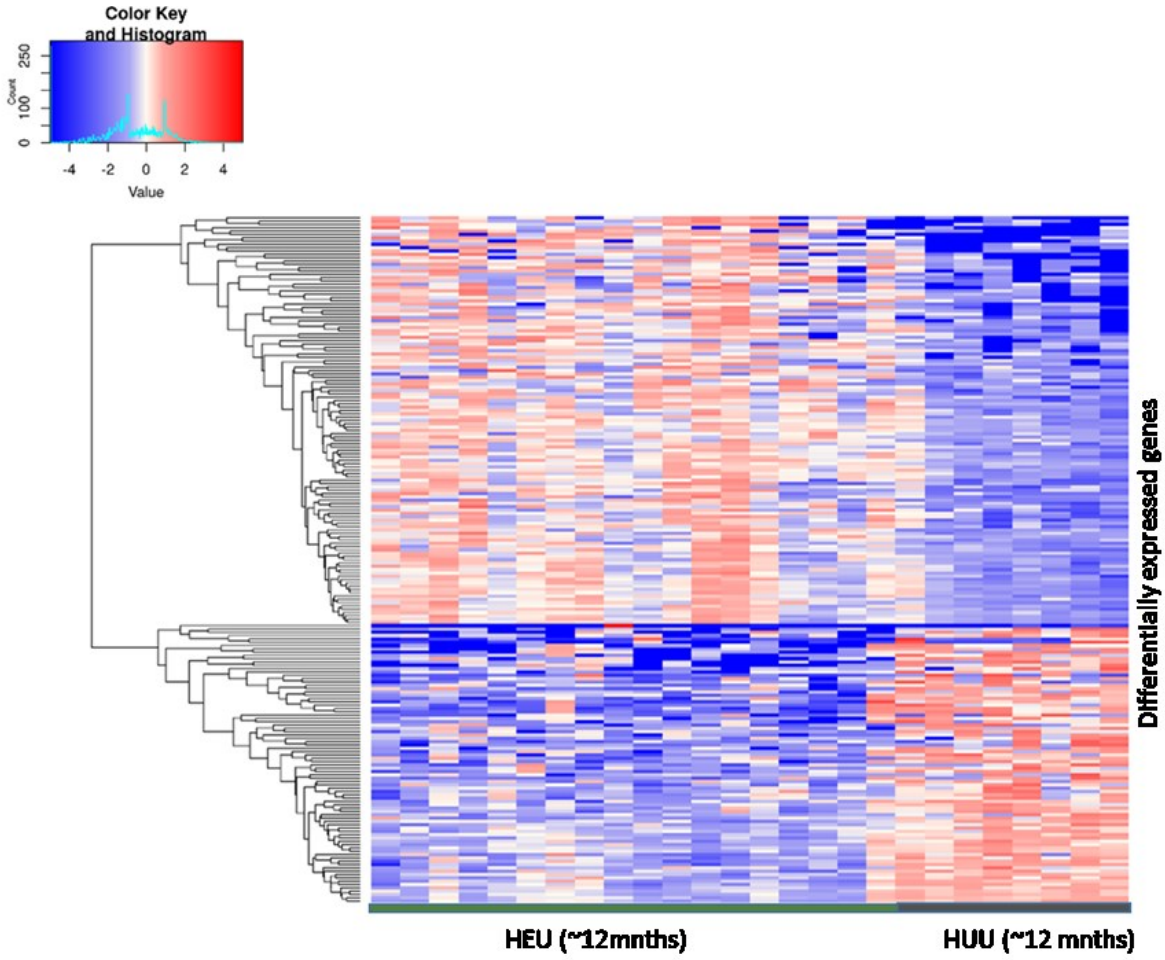



**Figure 4.0b: Differentially expressed genes in HEU infants relative to HUU infants at 12 months.** Heatmap of the differentially expressed genes in HEU infants relative to HUU infants at 12 (E) months. P.adj of <0.05 and Log2FoldChange >|1| resulted in 211 genes in total implying difference in gene expression in the first year of life of an HEU infant when compared to that of an HUU infant. Red indicates upregulated genes while blue indicates downregulated genes in HIV-exposed uninfected infants. The y axis represents the differentially expressed genes and the x axis represents the infant groups.

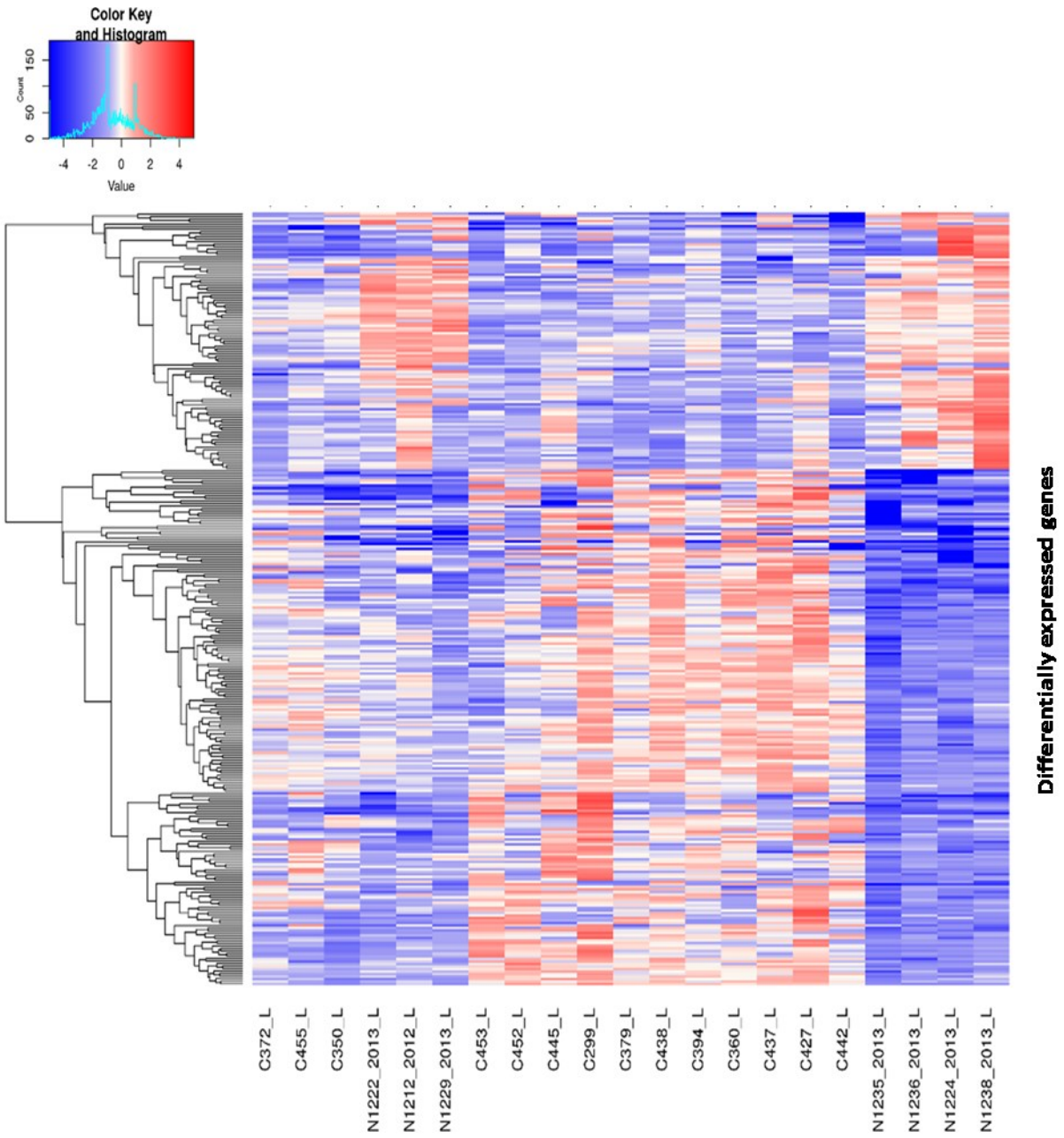



**Figure 4.0c: Differentially expressed genes in HEU infants relative to HUU infants at 24 months.** Heatmap of the differentially expressed genes in HEU infants relative to HUU infants at 24 (L) months. HEU infants indicated by prefix C while HUU infants are indicated by prefix N. Padj of <0.05 and Log2FoldChange >|1| resulted in 308 genes in total implying difference in gene expression in the second year of life of an HEU infant when compared to that of an HUU infant. Red indicates upregulated genes while blue indicates downregulated genes in HIV-exposed uninfected infants. The y axis represents the differentially expressed genes and the x axis represents the sample ID.

### 4.3 Gene expression profiles in HEU vs HUU in the first two years of life

A total of 166 genes (**APPENDIX III**) showed statistically significant differential expression in the HEU infants with 56 genes being down regulated and 110 genes being up regulated. To determine the statistically significant set of differentially expressed genes, a p.adj < 0.05 and absolute log 2-fold change threshold of > 1 was set as depicted by the volcano plot in **figure 4.1** and heatmap in **figure 4.2** below. The adjusted p value was obtained after multiple testing using Benjamini-Hochberg (69).

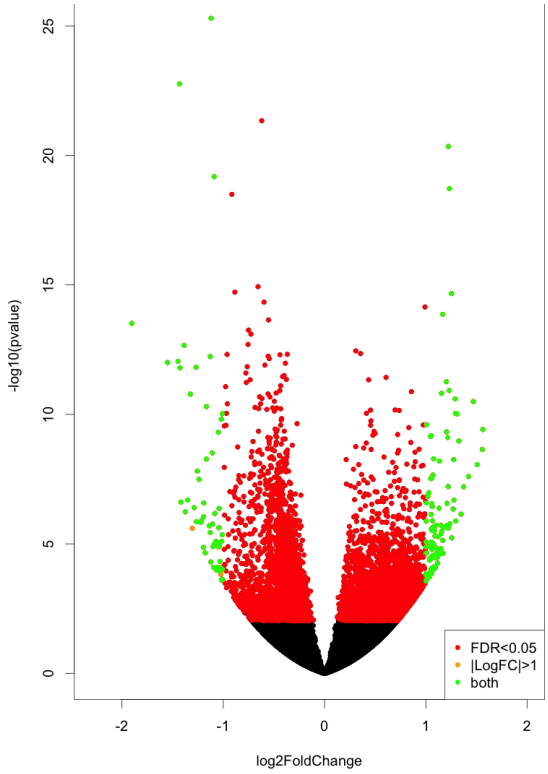



**Figure 4.1: Volcano plot showing statistically significant differentially expressed genes in HEU and HUU infants**. Significant differences are based on FDR < 0.05 and Log2FoldChange >|1|. The significantly differentially expressed genes, green in colour, are 166 in total. The red dots indicate the genes that meet the FDR threshold but not the LogFC threshold. The black dots indicate genes that do not meet the thresholds of both FDR and LogFC. The y axis represents the FDR and the x axis represents the log2-fold change.

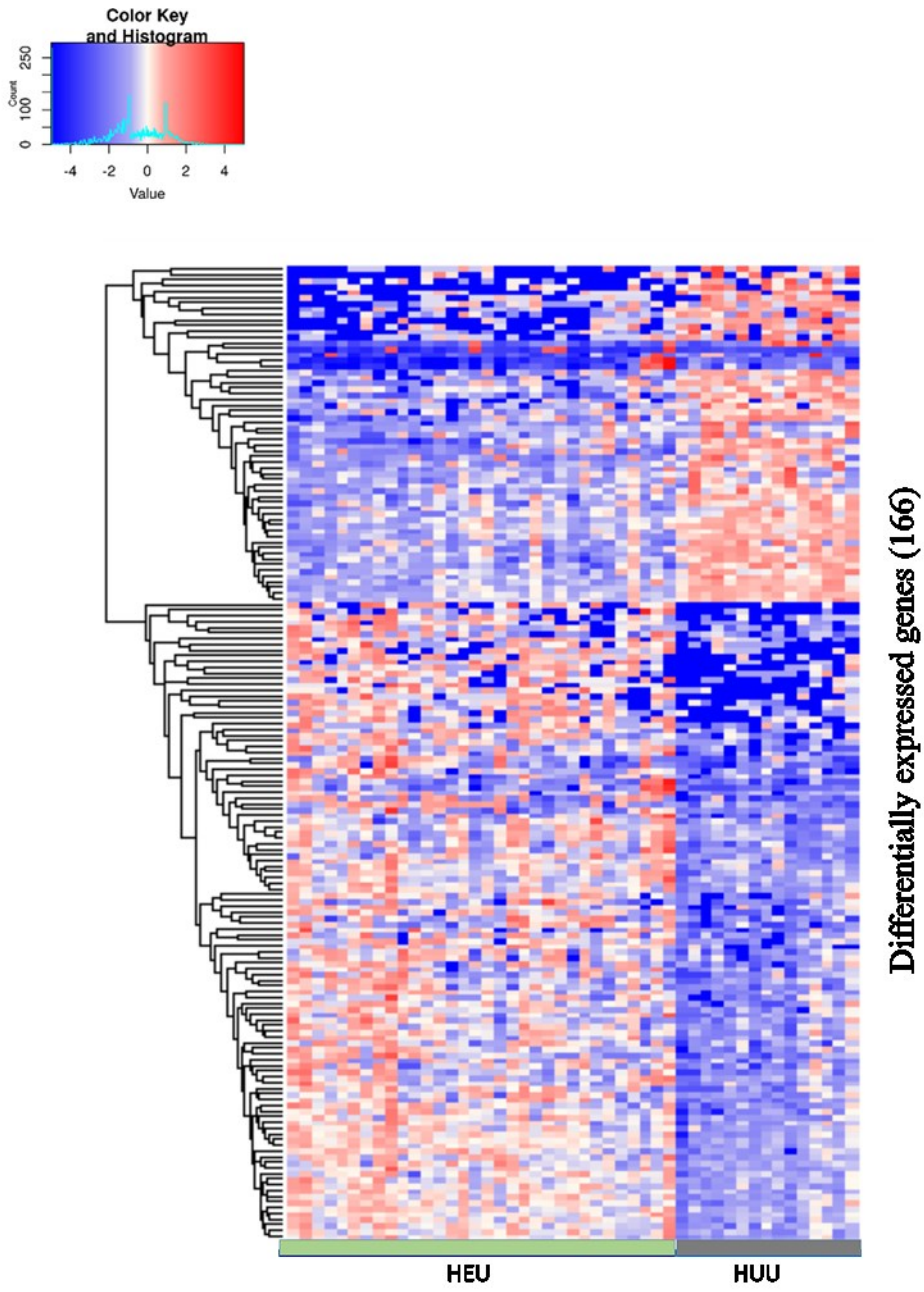



**Figure 4.2: Differentially expressed genes in HEU and HUU infants in the first 2 years of life.** Heat map of the 166 differentially expressed genes in HEU and HUU infants in the first two years of life. Padj of <0.05 and Log2FoldChange >|1| resulted in the 166 differentially expressed genes. Blue indicates downregulated genes while red indicates upregulated genes in HIV-exposed uninfected infants relative to HIV-unexposed uninfected infants. The y axis represents the genes and the x axis represents the infant groups.

Of the 166 genes, only 109 code for proteins of known function. Among the 57 genes of unknown protein function, 51 were upregulated genes. **Table 4.2** below shows the top 20 most up and down regulated genes of known protein function as per the log fold change.

**Table 4.2: Top 20 up and down regulated differentially expressed genes in HEU infants.**

| | Gene name | Symbol | L2FC | Padj |
|---|---|---|---|---|
| DOWN REGULATED GENES | Histone cluster 1 H1 family member d | HIST1H1D | -1.902599339 | 5.77E-11 |
| | Histone cluster 1 H1 family member e | HIST1H1E | -1.549489925 | 9.66E-10 |
| | Cholinergic receptor muscarinic 4 | CHRM4 | -1.445614257 | 9.03E-10 |
| | H1 histone family member X | H1FX | -1.431183991 | 2.45E-19 |
| | Transmembrane protein 119 | TMEM119 | -1.425719675 | 1.32E-09 |
| | Metallothionein 1G | MT1G | -1.417255238 | 1.97E-05 |
| | H1 histone family member 0 | H1F0 | -1.386079798 | 3.23E-10 |
| | Gamma-aminobutyric acid type A receptor delta subunit | GABRD | -1.374373732 | 4.01E-05 |
| | Early growth response 2 | EGR2 | -1.350011263 | 1.70E-05 |
| | Arrestin domain containing 3 | ARRDC3 | -1.324460409 | 9.01E-09 |
| | C-C motif chemokine ligand 3 like 3 | CCL3L3 | -1.286251212 | 2.94E-05 |
| | Zinc finger protein 844 | ZNF844 | -1.267605822 | 1.30E-09 |
| | Connective tissue growth factor | CTGF | -1.264507214 | 7.96E-05 |
| | C-C motif chemokine ligand 3 | CCL3 | -1.25479805 | 2.30E-06 |
| | Metallothionein 1E | MT1E | -1.23915972 | 4.14E-06 |
| | Histone cluster 2 H2A family member b | HIST2H2AB | -1.22204177 | 8.24E-05 |
| | Haptoglobin | HP | -1.214732381 | 7.43E-05 |
| | Insulin like 4 | INSL4 | -1.194384565 | 4.57E-04 |
| | Small nucleolar RNA U3 | U3 SnoRNA | -1.193805385 | 5.75E-05 |
| | Immunoglobulin lambda constant 7 | IGLC7 | -1.191899124 | 2.09E-05 |
| UP REGULATED GENES_20 | Forkhead box I1 | FOXI1 | 1.507376169 | 1.42E-06 |
| | Kyphoscoliosis peptidase | KY | 1.468419779 | 1.56E-08 |
| | Golgin A8 family member Q | GOLGA8Q | 1.372964011 | 6.87E-06 |
| | Major histocompatibility complex, class I, V (pseudogene) | HLA-V | 1.348117075 | 4.75E-05 |
| | Tumor necrosis factor superfamily member 14 | TNFSF14 | 1.308334009 | 3.62E-08 |
| | Leiomodin 2 | LMOD2 | 1.295044811 | 7.58E-05 |
| | Lung cancer associated transcript 1 (non-protein coding) | LUCAT1 | 1.289425197 | 1.28E-08 |
| | Cytokine inducible SH2 containing protein | CISH | 1.274929317 | 1.01E-06 |
| | Calpain small subunit 2 | CAPNS2 | 1.269688356 | 3.26E-05 |
| | Keratin 81 | KRT81 | 1.255279982 | 2.43E-04 |
| | SH3 domain containing GRB2 like 1, endophilin A2 pseudogene 1 | SH3GL1P1 | 1.253977829 | 6.09E-12 |
| | Olfactory receptor family 52 subfamily K member 2 | OR52K2 | 1.222001738 | 1.24E-04 |
| | Apolipoprotein A2 | APOA2 | 1.220798724 | 6.82E-06 |
| | Ephrin A1 | EFNA1 | 1.216223316 | 2.10E-07 |
| | G protein-coupled receptor 20 | GPR20 | 1.203480564 | 1.05E-04 |
| | Titin-cap | TCAP | 1.202260119 | 3.69E-09 |
| | Inhibitor of DNA binding 1, HLH protein | ID1 | 1.16756932 | 2.87E-04 |
| | ATP1B3 antisense RNA 1 | ATP1B3-AS1 | 1.157382512 | 1.54E-04 |
| | Coiled-coil domain containing 80 | CCDC80 | 1.154778828 | 8.70E-09 |
| | Fibronectin leucine rich transmembrane protein 1 | FLRT1 | 1.151881226 | 1.84E-04 |

The top 20 most up and down regulated genes of known protein function in HEU infants. A cut off of padj < 0.05 and Log2FoldChange >|1|.
padj defines an adjusted p value to indicate significance after multiple testing.
Log2FoldChange (L2FC) defines the effect size estimate showing the change in HEU infant samples in comparison to the HUU samples at a base 2 logarithmic scale.

## 4.4 Exposure to HIV could be associated with negative regulation of the HEU's immune system

To explore whether the observed expressional changes were associated with global biological processes, gene set enrichment analysis was performed using the Gene Ontology database. Pre-ranked 3,576 differentially expressed genes were used as input, ranked as per the 2-fold change values. Upregulated gene sets have a positive enrichment score while downregulated gene sets have a negative enrichment score. **Figure 4.3** below shows the biological processes network formed from clusters of inter-connected gene ontologies.

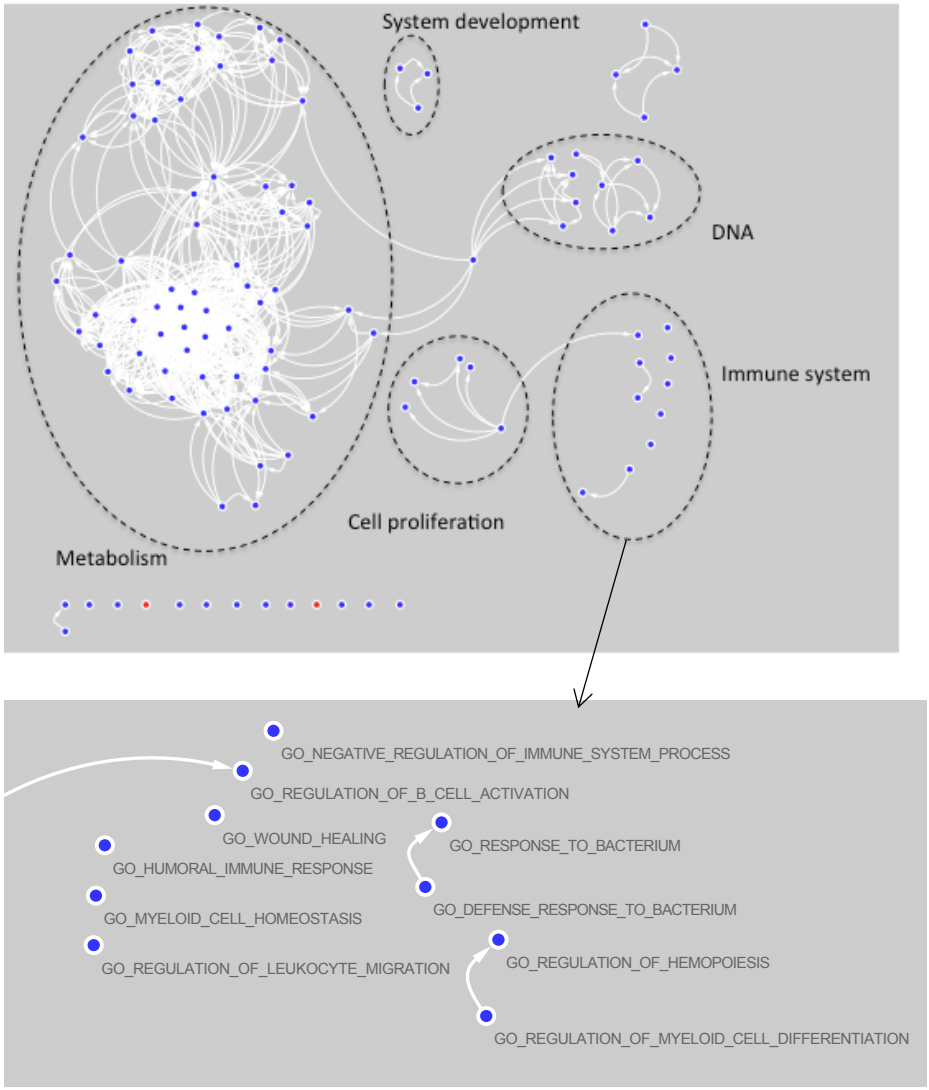



**Figure 4.3: Gene sets associated with the immune system.** The gene ontologies are formed by grouping genes that perform the same function into gene sets. The gene sets are represented by the nodes while the edges represent the GO defined relations. The blue nodes represent downregulated gene sets while the red nodes represent the upregulated gene sets.

These gene ontologies are formed through grouping genes working to perform the same function into a gene-set. Metabolism had the majority of gene ontologies but of interest to this study was the immune system.

After the threshold cut-off, majority of the gene sets were downregulated with only two sets being upregulated. Down regulated gene ontologies (**Figure 4.3**) forming part of the immune system included; negative regulation of immune system process, regulation of B cell activation, wound healing, humoral immune response, myeloid cell homeostasis, regulation of leukocyte migration, response to bacterium, regulation of myeloid cell differentiation and regulation of hemopoiesis.

The two upregulated gene sets were G Protein Coupled Receptor (GPCR) Signaling Pathway and Regulation of Lipid Biosynthetic Process. CISH, FFAR2 and HCAR3 are part of the genes that make up the GPCR Signaling Pathway and are all significantly up regulated. These genes indirectly affect the immune system hence the interest of GPCRs to the study.

### 4.5 Exposure to HIV *in utero* alters Granzyme A signaling in HEU Infants

Canonical pathways associated with the HEU infants' expression profiles resulted in the Granzyme A signalling pathway being the most affected with a p value of 4.79E-09. Granzyme A signalling relies on Histone H1, core histones and HMGB2 among others to perform its function. In the exposed uninfected infant, these genes are significantly down regulated.

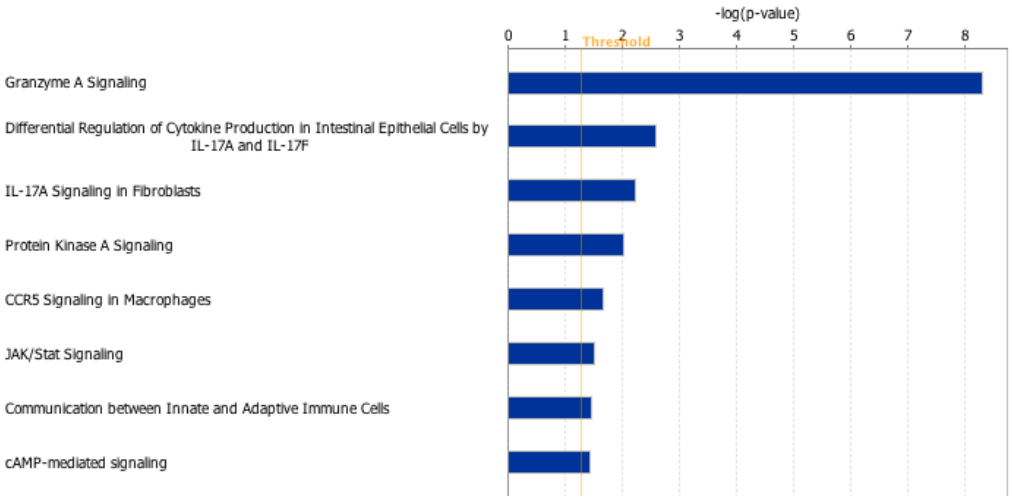



**Figure 4.4: HIV exposure alters the Granzyme A pathway significantly.** Canonical pathways associated with the 166 statistically significant differentially expressed genes were determined and a p value < 0.05 was used as a threshold cut-off. Granzyme A signalling pathway was significantly altered with a p value of 4.79E-09.

# 5.0 CHAPTER FOUR: DISCUSSION

HIV-1 *in utero* exposure has been shown to have an effect on the immune system of uninfected infants born to HIV infected women (70). These infants are expected to be healthy having escaped the deadly disease but they experience higher rates of hospitalization and mortality when compared to their unexposed uninfected counterparts. The causes of infection in both HEU and HUU infants are identical but greater severity is observed in the HEU infants (71). Here I present my analysis on the peripheral blood mononuclear cells transcriptome of HEU infants. My data confirms previous reports indicating reduced immune response in these infants (39,41,46,50,70,72–76) as well as revealing new insights on the potential pathways involved in the global dysregulation of the immune development in this group of infants. Functional analysis indicated immune functions such as defense response to bacterium, regulation of leukocyte migration, negative regulation of immune system process to be down regulated and G-Protein Coupled Receptor (GPCR) signaling pathway to be upregulated.

It is speculated that viral particles may traverse the placental membrane, most likely due to the trans-placental diffusion of HIV soluble proteins as opposed to live and replication-competent HIV. More than 20% as reviewed by Dauby and colleagues (42) of HEU infants have been shown to possess an immunological memory to HIV proteins (43). Analysis of antigen- stimulated production of interleukin-2 (IL-2) confirmed this when they observed high frequencies of HIV specific $CD4^+$ T helper cells in the HEU new-borns (39). In agreement, our data showed the upregulation of noncoding-RNA (*ncRNA*): ATPase Na+/K+ Transporting subunit Beta3- antisense RNA 1 (*ATP1B3-AS1)*, and a transmembrane protein, Tumor necrosis factor superfamily 14 (*TNFSF14*), in HEU infants. *ncRNA* function to regulate gene expression at transcriptional or post-transcriptional level(77–79) and have been shown to act as either positive or negative regulators of interferon mediated antiviral responses(80).

*ATP1B3-AS1* is an antisense transcript of *ATP1B3* gene which belongs to the family of Na+/K+ ATPase that catalyses the hydrolysis of ATP and has recently been shown to be a cofactor *BST-2* protein. *BST-2* is highly expressed on the surface of HIV infected cells and stops the release of newly produced virions. Knockdown of *ATP1B3* enhances the expression of *BST-2*, modulating the restriction of HIV-1 virions production. High expression of the antisense transcript *ATP1B3-AS1* may be an indicator of exposure to the

viral particles and the priming of BST-2 dependent anti-HIV responses in HEU (81). The up-regulation of *TNFSF14* gene which is expressed on activated T lymphocytes, preferentially $CD8^+$ (82–84),could be as a result of the increased percentage of activated $CD8^+$ lymphocytes (39,76,85,86).This increased percentage of activated $CD8^+$ lymphocytes could have contributed to the alteration of the Granzyme A pathway in the HEU infants as granzyme A is highly expressed in $CD8^+$ cytotoxic lymphocytes and natural killer cells. Granzymes are known to induce apoptosis in cells infected with viruses, intracellular bacteria and tumors. The Granzyme A opens up chromatin for the apoptotic nucleases to digest the infected cells via the substrates APE, histone H1, core histones, SET and HMGB2(87). In the HEU infant, the histone H1 and HMGB2 genes were down regulated and probably led to a defective Granzyme A signaling pathway hence reducing the ability to control other childhood infections.

We identified 166 genes with significant expression alterations. Down-regulation of genes that function in metabolism, cell proliferation and immunity was evident. Defense response to bacterium was observed to be suppressed in the HEU infant. Genes functioning in antimicrobial defense (Lipocalin 2 (*LCN2),* Cathelicidin-related antimicrobial peptides *(CAMP),* Haptoglobin *(HP),* histone family *(* Histone cluster 2 H2A family member B*(HIST2H2AB),* Histone cluster 1 H1 family member D *(HIST1H1D),* Histone cluster 1 H1 family member E (*HIST1H1E),* H1 histone family member X *(H1FX) and* H1 histone family member 0 *(H1F0*), High mobility group box 1&2 (*HMGB1* and *HMGB2))* were repressed in HEU and may contribute to the increased susceptibility to bacterial infections (88–90) partly attributed to impaired neutrophil function affecting neutrophil capacity for oxidative burst (91).

A pilot study (92) reported that although the neutrophil count of the HEU infant was normal, its functionality was impaired. On the contrary, some studies have observed reduced neutrophil count in HEU infants exposed to HIV and ARVs (86,93). It is possible that the observed down regulation of *LCN2, HP, CAMP* and the Lysine-rich Histone genes could contribute to dysregulated neutrophil function. Neutrophil's mode of action involve releasing antimicrobial peptides and nuclear constituents termed as neutrophil extracellular traps (NETs) (94,95). When histones are released into the cytoplasm by activated neutrophils they form part of the NETs contributing to the innate antimicrobial defense (96,97) through disruption of bacterial membrane, *H1* subfamily, or penetration of bacterial cell membrane and binding to bacterial DNA, *H2* subfamily. *LCN2* (neutrophil gelatinase-

associated lipocalin - NGAL) is known for its bacteriostatic role (88,98) and enhancement phagocytosis by macrophages (99). The down regulation of *Hp* gene that encodes for an acute phase bacteriostatic protein (hp) released during inflammation and injury by activated neutrophils (100) may also interfere with neutrophil function. Whilst *HMGB2* is not secreted by neutrophils, its secretion was observed when human THP-1 cell lines were stimulated with lipopolysaccharides (101) implicating its antimicrobial characteristic.

Within the GPCRs gene set, high expression of cytokine inducible SRC homology 2 (*CISH*), free fatty acid receptor2 (*FFAR2*) and hydrocarboxylic acid receptor 3 (*HCAR3)* was observed. *CISH* is expressed during the development of dendritic cells (DC) and regulates DC-mediated cytotoxic T-lymphocytes activation (102,103). Its expression is associated with negative regulation of cytokine signaling, controlling inflammation during infection, and susceptibility to infectious diseases such as bacteria and malaria (104). *FFAR2* is an important dietary nutritional sensor induced by short-chain fatty acids (105,106) and regulates macrophage and leukocyte differentiation and/or activation (107,108). Short chain fatty acids, produced by bacteria, activate the neutrophils via the *FFAR2* (109) inducing the inflammatory responses during bacterial infections, reviewed by Oliveira (110).

The up regulation of *FFAR2* in HEU infants may be an indication of neutrophil activation in HEU infants, identifying the cause of activation was outside the scope of this study. Expression of *CISH* and *FFAR2* is common during infection. The HEU infants in this study were observed to have an upregulated expression profiles of these two genes despite not having an ongoing infection. It is worth noting that the HEU infants were on daily cotrimoxazole prophylaxis and showed no signs/symptoms of infection. This suggests that the HEU infant may have an active immune system during a healthy state. This may lead to control mechanisms that lower the immunity of the infant reducing their ability to fight pathogens in case of an infection. Therefore, a bacterial infection that would otherwise be mild and not require hospitalization in an HUU infant, could be severe in an HEU infant leading to hospitalization and possibly death.

The HEU infants in this study were on cotrimoxazole prophylaxis in line with the recommended policy to protect HEU infants from *Pneumocystis jirovecii* pneumonia. Cotrimoxazole has been shown to be beneficial in reducing invasive bacterial infections (111,112), although the evidence for this is weak and current guidelines have been debated (113,114). It has also been reported to have little to no effect when it comes to mortality,

LRTIs and diarrheal infections(115–117) while others have observed a significant reduction in hospitalizations(118).

It was not possible to ascertain if Cotrimoxazole had any effect on the expression of these genes. The Mpepu study(119), carried out in non-malarial, low-breastfeeding areas, observed no clinical benefits in HEU infants on cotrimoxazole. Based on this observation, it is safe to assume that the Cotrimoxazole did not influence the expression profile of the genes associated with antimicrobial activity in the HEU infants. It may therefore be plausible that HEU infants are susceptible to microbial infections and that the innate immune system could, indeed, be compromised (89,120).

Breastfeeding was not included as a confounding factor in this study mainly because it is difficult to accurately measure mothers' breastfeeding habits as it is done at the comfort of their homes; the mother's reported information at the clinic may be totally different from her actual breastfeeding habits. Like HIV uninfected mothers, mothers with HIV infection are expected to breast feed exclusively for the first six months of the infant's life as the risk of HIV transmission from an infected mother on ART is minimal. Majority of the infected mothers are barely healthy and are undernourished hence unable to exclusively breastfeed for six months, let alone four months. On the other hand, not all healthy mothers exclusively breastfeed for 6 months. Due to the above reasons, conclusion on whether breastfeeding was a factor in the differential expression observed in the HEU infants cannot be made.

## CONCLUSIONS AND FUTURE RECOMMENDATIONS

Infections observed in the HEU infants are identical to those observed in the HUU infant as they are caused by the same pathogens. The only difference is that the HEU infant is unable to enforce an immune response that would control the infection resulting in severe infection, prolonged hospitalization and possibly death. To my knowledge, this is the first study that has used a transcriptomic approach to describe the gene expression profile of PBMCs of HEU infants. I observed changes in genes involved with innate immune responses, specifically neutrophil functions. It is possible that HEU infants have an ineffective neutrophil response predisposing them to bacterial infections and subsequent increased morbidity and mortality. It is however possible that the observed changes may be

as a result of ongoing infections as opposed to epigenetic changes and further analysis are warranted.

A molecular approach may better define the impact of HIV exposure on the infant's immune system and may help identify immunological pathways of interest. Using an integrated approach that considers phenotypic, functional and transcriptomics may help us better decipher the underlying causes of the observed increased morbidity and mortality in HEU infants. Future studies should consider the inclusion of transcriptomic analysis a powerful tool to decipher the role of HIV exposure on the infants developing immune system. This study was limited to HEU infants on cotrimoxazole and future work looking at a later time point once these children are not on active clinical follow up would give a complete picture if indeed vertical exposure has an intrinsic effect that is maintained in early childhood.

While this study gives us a hint that indeed HIV exposure may alter the developing immune responses, further analysis to whether ARV exposure and breastfeeding are a contributing factor would be important in the near future. Confirmation of the differential expression of the CISH, FFAR2, HCAR3, CAMP, HMGB2, HP and LCN2 genes in the HEU infants through RT-PCR. I also suggest a gene set enrichment re-analysis without focus on the immune system, but general biological and molecular function.

## REFERENCES


1. Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. 2011;1–22.
2. Russell GJ, Norman HD. Pathology and Therapeutics for Pharmacists. Third Edit. 2008. 982 p.
3. Maartens G, Celum C, Lewin SR. Seminar HIV infection : epidemiology , pathogenesis , treatment , and prevention. Lancet [Internet]. Elsevier Ltd; 2014;384(9939):258–71. Available from: http://dx.doi.org/10.1016/S0140-6736(14)60164-1
4. Shaw GM, Hunter E. HIV Transmission. 2012;1–23.
5. Pantaleo G, Graziosi C, Fauci AS. The immunopathogenesis of Human Immunodeficiency Virus infection. New Engl J Med [Internet]. 1993;328(27):327–35. Available from: http://www.nejm.org/doi/pdf/10.1056/NEJM199302043280508
6. Costin JM. Cytopathic Mechanisms of HIV-1. 2007;22:1–22.
7. UNAIDS. REGIONAL HIV STATISTICS — 2016. 2017.
8. UNAIDS. UNAIDS Data 2017 [Internet]. 2017. Available from: www.unaids.org/en/resources/documents/2017/2017_data_book
9. Zheng NN, Kiviat NB, Sow PS, Hawes SE, Wilson A, Critchlow CW, et al. Comparison of Human Immunodeficiency Virus ( HIV ) -Specific T-Cell Responses in HIV-1- and HIV-2-Infected Individuals in Senegal. 2004;78(24):13934–42.
10. Gue S, Popper SJ, Sarr AD, Mboup S, Essex ME, Kanki PJ. Low Plasma Human Immunodeficiency Virus Type 2 Viral Load Is Independent of Proviral Load : Low Virus Production In Vivo. 2000;74(3):1554–7.
11. Berry N, Jaffar S, Loeff MSVANDER, Ariyoshi K, Harding E, Gom PATN, et al. Low Level Viremia and High CD4 % Predict Normal Survival in a Cohort of HIV Type-2-Infected Villagers. 2002;18(16):1167–73.
12. Joseph O, Biodun O, Michael E. Pregnancy outcome among HIV positive women receiving antenatal HAART versus untreated maternal HIV infection. J Coll Physicians Surg Pakistan. 2011;21(6):356–9.
13. Slogrove AL, Cotton MF, Esser MM. Severe infections in HIV-exposed uninfected infants: Clinical evidence of immunodeficiency. J Trop Pediatr. 2009;56(2):75–81.
14. John GC, Kreiss J. Mother-to-child transmission of human immunodeficiency virus type 1. Epidemiol Rev [Internet]. 1996;18(2):149–57. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3372415&tool=pmcentrez&rendertype=abstract
15. WHO. PMTCT strategic vision 2010–2015 : preventing mother-to-child transmission of HIV to reach the UNGASS and Millennium Development Goals. World Heal Organ. 2010;1–40.
16. UNAIDS. Children and HIV. Media. 2016;296(7):2006–7.
17. European Collaborative Study. Mother-to-Child Transmission of HIV Infection in

the Era of Highly Active Antiretroviral Therapy. Clin Infect Dis. 2005;40(1):458–65.

18. NASCOP M of HK. Guidelines on Use of Antiretroviral Drugs for Treating and Preventing HIV Infection in Kenya. 2016th ed. 2016. 174 p.

19. Newell ML. Prevention of mother-to-child transmission of HIV: Challenges for the current decade. Bulletin of the World Health Organization. 2001. p. 1138–44.

20. Ruck C, Reikie BA, Marchant A, Kollmann TR. Linking Susceptibility to Infectious Diseases to Immune System Abnormalities among HIV-Exposed Uninfected Infants. 2016;7(August):1–12.

21. Sherman GG. HIV testing during the neonatal period. South Afr J HIV Med [Internet]. 2015;16(1):1–3. Available from: 10.4102/sajhivmed.v16i1.362%5Cnhttp://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=102484482&site=ehost-live

22. Singh HK, Gupte N, Kinikar A, Bharadwaj R, Sastry J, Suryavanshi N, et al. High Rates of All-cause and Gastroenteritis- related Hospitalization Morbidity and Mortality among HIV-exposed Indian Infants. BMC Infect Dis [Internet]. BioMed Central Ltd; 2011;11(1):193. Available from: http://www.biomedcentral.com/1471-2334/11/193

23. Rupérez M, Gonzalez R, Maculuve S, Quintó L, Lopez-Varela E, Joaquim OA, et al. Maternal HIV infection is an important health determinant in non HIV-infected infants. Submitted. 2017;(March).

24. Filteau S. The HIV-exposed, uninfected African child. Trop Med Int Heal. 2009;14(3):276–87.

25. Koyanagi A, Humphrey JH, Ntozini R, Nathoo K, Moulton LH, Iliff P, et al. Morbidity Among Human Immunodeficiency Virus-exposed But uninfected, Human Immunodeficiency Virus-infected, and Human Immunodeficiency Virus-unexposed Infants in Zimbabwe Before Availability of Highly Active Antiretroviral Therapy. Pediatr Infect Dis J. 2011;30(1):45–51.

26. Kuhn L, Kasonde P, Sinkala M, Kankasa C, Semrau K, Scott N, et al. Does severity of HIV disease in HIV-infected mothers affect mortality and morbidity among their uninfected infants? Clin Infect Dis [Internet]. 2005;41(11):1654–61. Available from: http://cid.oxfordjournals.org/content/41/11/1654.full.pdf

27. Taron-Brocard C, Chenadec J Le, Faye A, Dollfus C, Goetghebuer T, Gajdos V, et al. Increased risk of serious bacterial infections due to maternal immunosuppression in HIV-exposed uninfected infants in a European country. Clin Infect Dis. 2014;59(9):1332–45.

28. Mussi-Pinhata MM, Motta F, Freimanis-Hance L, de Souza R, Szyld E, Succi RCM, et al. Lower respiratory tract infections among human immunodeficiency virus-exposed, uninfected infants. Int J Infect Dis [Internet]. International Society for Infectious Diseases; 2010;14(SUPPL. 3):e176–82. Available from: http://dx.doi.org/10.1016/j.ijid.2010.01.006

29. Cohen C, Moyes J, Tempia S, Groome M, Walaza S, Pretorius M, et al. Epidemiology of Acute Lower Respiratory Tract Infection in HIV- Exposed Uninfected Infants. Pediatrics. 2016;137(4):e20153272.

30. Weinberg A, Mussi-pinhata MM, Yu Q, Cohen RA, Almeida VC, Amaral F, et al. Excess respiratory viral infections and low antibody responses among HIV-exposed , uninfected infants. 2017;(October 2016).

31. Von Mollendorf C, Von Gottberg A, Tempia S, Meiring S, De Gouveia L, Quan V, et al. Increased risk for and mortality from invasive pneumococcal disease in HIV-exposed but uninfected infants aged <1 year in South Africa, 2009-2013. Clin Infect Dis. 2015;60(9):1346–56.

32. M.S. K, K.E. W, A.P. S, C.K. C, T. A-M, S.C. B, et al. Treatment failures and excess mortality among HIV-exposed, uninfected children with pneumonia [Internet]. Journal of the Pediatric Infectious Diseases Society. 2015. p. e117–26. Available from: http://jpids.oxfordjournals.org/%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed13&NEWS=N&AN=20160166659

33. Epalza C, Goetghebuer T, Hainaut M, Prayez F, Barlow P, Dediste A, et al. High incidence of invasive group B streptococcal infections in HIV-exposed uninfected infants. Pediatrics. 2010;126(3):e631–8.

34. Zash R, Souda S, Leidner J, Ribaudo H, Binda K, Moyo S, et al. HIV-exposed children account for more than half of 24-month mortality in Botswana. BMC Pediatr [Internet]. BMC Pediatrics; 2016;16:103. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27439303%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4955224

35. Slogrove AL, Esser MM, Cotton MF, Speert DP, Kollmann TR, Singer J, et al. A Prospective Cohort Study of Common Childhood Infections in South African HIV-exposed Uninfected and HIV-unexposed Infants. Pediatr Infect Dis J [Internet]. 2017;36(2):e38–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28081048%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5242219

36. Fox MP, Brooks DR, Kuhn L, Aldrovandi G, Sinkala M, Kankasa C, et al. Role of breastfeeding cessation in mediating the relationship between maternal HIV disease stage and increased child mortality among HIV-exposed uninfected children. Int J Epidemiol. 2009;38(2):569–76.

37. Kuhn L, Sinkala M, Semrau K, Kankasa C, Kasonde P, Mwiya M, et al. Elevations in mortality associated with weaning persist into the second year of life among uninfected children born to HIV-infected mothers. Clin Infect Dis [Internet]. 2010;50(3):437–44. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2805776&tool=pmcentrez&rendertype=abstract

38. Fawzy A, Arpadi S, Kankasa C, Sinkala M, Mwiya M, Thea DM, et al. Early Weaning Increases Diarrhea Morbidity and Mortality Among Uninfected Children Born to HIV-infected Mothers in Zambia. 2011;203.

39. Clerici M, Saresella M, Colombo F, Fossati S, Sala N, Bricalli D, et al. T-lymphocyte maturation abnormalities in uninfected newborns and children with vertical exposure to HIV. Blood. 2000;96(12):3866–71.

40. Infecciosas DD, Medicina D De, Medicina EP De, Virologia L De, Medicina I De, São T De, et al. Imbalance of naive and memory T lymphocytes with sustained high

cellular activation during the first year of life from uninfected children born to HIV-1-infected mothers on HAART. 2008;1:700–8.

41. Afran L, Garcia Knight M, Nduati E, Urban BC, Heyderman RS, Rowland-Jones SL. HIV-exposed uninfected children: A growing population with a vulnerable immune system? Clin Exp Immunol. 2014;176(1):11–22.

42. Dauby N, Goetghebuer T, Kollmann TR, Levy J, Marchant A. Uninfected but not unaffected: Chronic maternal infections during pregnancy, fetal immunity, and susceptibility to postnatal infections [Internet]. The Lancet Infectious Diseases. Elsevier Ltd; 2012. p. 330–40. Available from: http://dx.doi.org/10.1016/S1473-3099(11)70341-3

43. Rowland-Jones SL, Nixon DF, Gotch F, McMichael A, Kroll JS, Hallam N, et al. HIV-specific cytotoxic T-cell activity in an HIV-exposed but uninfected infant. Lancet. 1993;341(8849):860–1.

44. Legrand FA, Nixon DF, Loo CP, Ono E, Chapman JM, Miyamoto M, et al. Strong HIV-1-specific T cell responses in HIV-1-exposed uninfected infants and neonates revealed after regulatory T cell removal. PLoS One. 2006;1(1).

45. Kidzeru EB, Hesseling AC, Passmore J-AS, Myer L, Gamieldien H, Tchakoute CT, et al. In-utero exposure to maternal HIV infection alters T-cell immune responses to vaccination in HIV-uninfected infants. AIDS [Internet]. 2014;28(January):1421–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24785950

46. Abramczuk BM, Mazzola TN, Moreno YMF, Zorzeto TQ, Quintilio W, Wolf PS, et al. Impaired humoral response to vaccines among HIV-exposed uninfected infants. Clin Vaccine Immunol. 2011;18(9):1406–9.

47. Borges-Almeida E, Milanez HMBPM, Vilela MMS, Cunha FGP, Abramczuk BM, Reis-Alves SC, et al. The impact of maternal HIV infection on cord blood lymphocyte subsets and cytokine profile in exposed non-infected newborns. BMC Infect Dis [Internet]. 2011;11:38. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3040712&tool=pmcentrez&rendertype=abstract

48. Longwe H, Phiri KS, Mbeye NM, Gondwe T, Jambo KC, Mandala WL. Proportions of CD4+, CD8+ and B cell subsets are not affected by exposure to HIV or to Cotrimoxazole prophylaxis in Malawian HIV-uninfected but exposed children. BMC Immunol [Internet]. BMC Immunology; 2015;16(1):50. Available from: http://dx.doi.org/10.1186/s12865-015-0115-y

49. Nielsen SD, Jeppesen DL, Kolte L, Clark DR, Sørensen TU, Dreves A, et al. Impaired progenitor cell function in HIV-negative infants of HIV-positive mothers results in decreased thymic output and low CD4 counts. 2016;98(2):398–405.

50. Garcia-Knight MA, Nduati E, Hassan AS, Gambo F, Odera D, Etyang TJ, et al. Altered memory T-cell responses to Bacillus Calmette-Guerin and Tetanus Toxoid vaccination and altered cytokine responses to polyclonal stimulation in HIV-exposed uninfected Kenyan infants. PLoS One. 2015;10(11):1–19.

51. Miyamoto M, Ono E, Pahwa S, Moraes-pinto MI De. Immune development in HIV-exposed uninfected children born to HIV-infected women Maristela. 2017;(July 2016):1–9.

52. Jones CE, Naidoo S, De Beer C, Esser M, Kampmann B, Hesseling AC. Maternal HIV infection and antibody responses against vaccine-preventable diseases in uninfected infants. JAMA [Internet]. 2011;305(6):576–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21304083

53. Madhi SA, Adrian P, Cotton MF, McIntyre JA, Jean-Philippe P, Meadows S, et al. Effect of HIV infection status and anti-retroviral treatment on quantitative and qualitative antibody responses to pneumococcal conjugate vaccine in infants. J Infect Dis [Internet]. 2010;202(3):355–61. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2902789&tool=pmcentrez&rendertype=abstract

54. Jones CE, Beer C De. Maternal HIV Infection and Antibody Responses Against Vaccine-Preventable Diseases in Uninfected Infants. 2011;305(6).

55. Gaensbauer JT, Rakhola JT, Onyango-Makumbi C, Mubiru M, Westcott JE, Krebs NF, et al. Impaired Haemophilus influenzae type b transplacental antibody transmission and declining antibody avidity through the first year of life represent potential vulnerabilities for HIV-exposed but-uninfected infants. Clin Vaccine Immunol. 2014;21(12):1661–7.

56. Nduati EW, Nkumama IN, Gambo FK, Muema DM, Knight MG, Hassan AS, et al. HIV-Exposed uninfected infants show robust memory B-Cell Responses in spite of a delayed accumulation of memory B cells: An observational study in the first 2 years of life. Clin Vaccine Immunol. 2016;23(7):576–85.

57. Huo Y, Patel K, Scott GB, Van Dyke RB, Siberry GK, Burchett SK, et al. Lymphocyte subsets in HIV-exposed uninfected infants and HIV-unexposed uninfected infants. J Allergy Clin Immunol. 2017;

58. Miyamoto M, Pessoa SD, Ono E, Machado DM, Salomão R, Succi RC de M, et al. Low CD4+ T-cell levels and B-cell apoptosis in vertically HIV-exposed noninfected children and adolescents. J Trop Pediatr. 2010;56(6):427–32.

59. Nduati EW, Hassan AS, Knight MG, Muema DM, Jahangir MN, Mwaringa SL, et al. Outcomes of prevention of mother to child transmission of the human immunodeficiency virus-1 in rural Kenya--a cohort study. BMC Public Health [Internet]. 2015;15:1008. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4592570&tool=pmcentrez&rendertype=abstract

60. Mofenson LM. Editorial commentary: New challenges in the elimination of pediatric HIV infection: The expanding population of HIV-exposed but uninfected children. Clinical Infectious Diseases. 2015. p. 1357–60.

61. Marquez C, Okiring J, Chamie G, Ruel TD, Achan J, Kakuru A, et al. Increased morbidity in early childhood among HIV-exposed uninfected children in uganda is associated with breastfeeding duration. J Trop Pediatr. 2014;60(6):434–41.

62. Shapiro RL, Lockman S. Mortality among HIV-Exposed Infants: The First and Final Frontier. Clin Infect Dis [Internet]. 2010;50(3):445–7. Available from: http://cid.oxfordjournals.org/lookup/doi/10.1086/649887

63. Kuhn L, Kasonde P, Sinkala M, Kankasa C, Semrau K, Scott N, et al. Does severity of HIV disease in HIV-infected mothers affect mortality and morbidity among their

uninfected infants? Clin Infect Dis [Internet]. 2005;41(11):1654–61. Available from: http://cid.oxfordjournals.org/content/41/11/1654.full.pdf

64. Legrand FA, Nixon DF, Loo CP, Ono E, Chapman JM, Miyamoto M, et al. Strong HIV-1-specific T cell responses in HIV-1-exposed uninfected infants and neonates revealed after regulatory T cell removal. PLoS One. 2006;1(1).

65. Jones CE, Naidoo S, De Beer C, Esser M, Kampmann B, Hesseling AC. Maternal HIV infection and antibody responses against vaccine-preventable diseases in uninfected infants. Jama [Internet]. 2011;305(6):576–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21304083

66. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2 : accurate alignment of transcriptomes in the presence of insertions , deletions and gene fusions. 2013;1–13.

67. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.

68. Love MI, Anders S, Huber W. Differential analysis of count data - the DESeq2 package [Internet]. Genome Biology. 2014. 550 p. Available from: http://biorxiv.org/lookup/doi/10.1101/002832%5Cnhttp://dx.doi.org/10.1186/s13059-014-0550-8

69. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol [Internet]. 2014;15(12):550. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8

70. Abu-raya B, Kollmann TR, Marchant A, Macgillivray DM, Chougnet CA. The Immune System of HIV-Exposed Uninfected Infants. 2016;7(September):1–10.

71. Slogrove AL, Goetghebuer T, Cotton MF, Singer J, Bettinger JA. Pattern of infectious morbidity in HIV-exposed uninfected infants and children. Front Immunol. 2016;7(MAY):1–8.

72. Ruck C, Reikie BA, Marchant A, Kollmann TR, Kakkar F. Linking susceptibility to infectious diseases to immune system abnormalities among HIV-exposed uninfected infants. Front Immunol. 2016;7(AUG):1–12.

73. Reikie BA, Adams RCM, Leligdowicz A, Ho K, Naidoo S, Rusk CE, et al. Altered innate immune development in HIV-exposed uninfected infants. J Acquir Immune Defic Syndr. 2014;1(3):245–55.

74. Weinberg A, Mussi-Pinhata MM, Yu Q, Cohen RA, Almeida VC, Amaral F, et al. Excess respiratory viral infections and low antibody responses among HIV-exposed, uninfected infants. Aids [Internet]. 2017;31(5):669–79. Available from: http://insights.ovid.com/crossref?an=00002030-201703130-00008

75. Kakkar F, Soudeyns H, Lamarre V, Ducruet T, Boucher M, Lapointe N. Immunological abnormalities among HIV-exposed uninfected infants in Quebec. Can J Infect Dis Med Microbiol [Internet]. 2013;24:21A. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L71971652%5Cnhttp://sfx.hul.harvard.edu/sfx_local?sid=EMBASE&issn=17129532&id=doi:&atitle=Immunological+abnormalities+among+HIV-exposed+uninfected+infants+in+Quebec&stitle=Can.+J.+Infe

76. Ono E, Nunes dos Santos AM, de Menezes Succi RC, Machado DM, de Angelis DSA, Salomão R, et al. Imbalance of naive and memory T lymphocytes with sustained high cellular activation during the first year of life from uninfected children born to HIV-1-infected mothers on HAART. Brazilian J Med Biol Res. 2008;41(8):700–8.

77. Morris K V. RNA-Directed Transcriptional Gene Silencing and Activation in Human Cells. Oligonucleotides [Internet]. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA ; 2009 Dec 23 [cited 2018 Aug 26];19(4):299–305. Available from: http://www.liebertpub.com/doi/10.1089/oli.2009.0212

78. Weinberg MS, Morris K V. Transcriptional gene silencing in humans. Nucleic Acids Res [Internet]. Oxford University Press; 2016 Aug 19 [cited 2018 Aug 26];44(14):6505–17. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw139

79. Zhang H-S, Wu T-C, Sang W-W, Ruan Z. MiR-217 is involved in Tat-induced HIV-1 long terminal repeat (LTR) transactivation by down-regulation of SIRT1. Biochim Biophys Acta - Mol Cell Res [Internet]. 2012 May [cited 2018 Aug 26];1823(5):1017–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22406815

80. Barriocanal M, Fortes P. Long Non-coding RNAs in Hepatitis C Virus-Infected Cells. Front Microbiol [Internet]. Frontiers; 2017 Sep 28 [cited 2018 Aug 26];8:1833. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01833/full

81. Nishitsuji H, Sugiyama R, Abe M, Takaku H. ATP1B3 Protein Modulates the Restriction of HIV-1 Production and Nuclear Factor κ Light Chain Enhancer of Activated B Cells (NF-κB) Activation by BST-2. J Biol Chem [Internet]. American Society for Biochemistry and Molecular Biology; 2016 Feb 26 [cited 2018 Aug 26];291(9):4754–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26694617

82. Mauri DN, Ebner R, Montgomery RI, Kochel KD, Cheung TC, Yu GL, et al. LIGHT, a new member of the TNF superfamily, and lymphotoxin α are ligands for herpesvirus entry mediator. Immunity. 1998;8(1):21–30.

83. Bhat NK, Thompson CB, Lindsten T, June CH, Fujiwara S, Koizumi S, et al. Reciprocal expression of human ETS1 and ETS2 genes during T-cell activation: regulatory role for the protooncogene ETS1. Proc Natl Acad Sci U S A [Internet]. 1990;87(10):3723–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=53975&tool=pmcentrez&rendertype=abstract

84. Morel Y, Schiano de Colella JM, Harrop J, Deen KC, Holmes SD, Wattam T a, et al. Reciprocal expression of the TNF family receptor herpes virus entry mediator and its ligand LIGHT on activated T cells: LIGHT down-regulates its own receptor. J Immunol [Internet]. 2000;165(8):4397–404. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11035077

85. Bunders M, Thorne C, Newell ML, European Collaborative Study. Maternal and infant factors and lymphocyte, CD4 and CD8 cell counts in uninfected children of HIV-1-infected mothers. AIDS [Internet]. 2005;19(10):1071–9. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/15958839

86. Le Chenadec J, Mayaux M, Guihenneuc-Jouyaux C, Blanche S. Perinatal antiretroviral treatment and hematopoiesis in HIV-uninfected infants. Aids [Internet]. 2003;17(14):2053–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14502008

87. Chowdhury D, Lieberman J. Death by a Thousand Cuts: Granzyme Pathways of Programmed Cell Death. NIH Public Access. 2009;389–420.

88. Flo T, Smith K, Sato S. Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. Nature [Internet]. 2004;432:917–21. Available from: http://www.nature.com/nature/journal/v432/n7019/abs/nature03104.html

89. Zanetti M. Cathelicidins, multifunctional peptides of the innate immunity. J Leukoc Biol [Internet]. 2003;75(1):39–48. Available from: http://www.jleukbio.org/cgi/doi/10.1189/jlb.0403147

90. Küchler R, Schroeder BO, Jaeger SU, Stange EF, Wehkamp J. Antimicrobial activity of high-mobility-group box 2: a new function to a well-known protein. Antimicrob Agents Chemother [Internet]. American Society for Microbiology Journals; 2013 Oct 1 [cited 2018 Aug 17];57(10):4782–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23877675

91. Maloupazoa Siawaya AC, Mveang-Nzoghe A, Mvoundza Ndjindji O, Mintsa Ndong A, Essone PN, Djoba Siawaya JF. Cases of Impaired Oxidative Burst in HIV-Exposed Uninfected Infants' Neutrophils-A Pilot Study. Front Immunol [Internet]. Frontiers Media SA; 2017 [cited 2018 Aug 20];8:262. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28337206

92. Siawaya ACM, Mveang-Nzoghe A, Ndjindji OM, Ndong AM, Essone PN, Siawaya JFD. Cases of impaired oxidative burst in HIV-exposed uninfected infants' neutrophils-A pilot study. Front Immunol. 2017;8(MAR):1–7.

93. Study EC. Levels and patterns of neutrophil cell counts over the first 8 years of life in children of HIV-1-infected mothers. AIDS. 2004;18(15):2009–17.

94. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil Extracellular Traps Kill Bacteria Brinkmann Science 2004.pdf. Science [Internet]. 2004;303(5663):1532–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15001782

95. Wiesner J, Vilcinskas A, Wiesner J, Vilcinskas A. Antimicrobial peptides : The ancient arm of the human immune system The ancient arm of the human immune system Antimicrobial peptides. 2016 [cited 2018 Aug 20];5594(April). Available from: http://www.tandfonline.com/action/journalInformation?journalCode=kvir20

96. Hoeksema M, Van Eijk M, Haagsman HP, Hartshorn KL. Histones as mediators of host defense, inflammation and thrombosis. Futur Microbiol [Internet]. 2016 [cited 2018 Aug 17];(3):441–53. Available from: www.futuremedicine.com

97. Rose FRAJ, Bailey K, Keyte JW, Chan WC, Greenwood D, Mahida YR. Potential role of epithelial cell-derived histone H1 proteins in innate antimicrobial defense in the human gastrointestinal tract. Infect Immun. 1998;66(7):3255–63.

98. Goetz DH, Holmes MA, Borregaard N, Bluhm ME, Raymond KN, Strong RK. The

neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. Mol Cell. 2002;10(5):1033–43.

99. Toyonaga T, Matsuura M, Mori K, Honzawa Y, Minami N, Yamada S, et al. Lipocalin 2 prevents intestinal inflammation by enhancing phagocytic bacterial clearance in macrophages. Sci Rep [Internet]. Nature Publishing Group; 2016;6(1):35014. Available from: http://www.nature.com/articles/srep35014

100. Theilgaard-Mönch K, Jacobsen LC, Nielsen MJ, Rasmussen T, Udby L, Gharib M, et al. Haptoglobin is synthesized during granulocyte differentiation, stored in specific granules, and released by neutrophils in response to activation. Blood. 2006;108(1):353–61.

101. Pusterla T, de Marchis F, Palumbo R, Bianchi ME. High mobility group B2 is secreted by myeloid cells and has mitogenic and chemoattractant activities similar to high mobility group B1. Autoimmunity [Internet]. 2009;42(4):308–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19811285

102. Miah MA, Yoon C-H, Kim J, Jang J, Seong Y-R, Bae Y-S. CISH is induced during DC development and regulates DC-mediated CTL activation. Eur J Immunol [Internet]. Wiley-Blackwell; 2012 Jan 1 [cited 2018 Aug 26];42(1):58–68. Available from: http://doi.wiley.com/10.1002/eji.201141846

103. Trengove MC, Ward AC. SOCS proteins in development and disease. Am J Clin Exp Immunol [Internet]. 2013 [cited 2018 Aug 26];2(1):1–29. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23885323

104. Khor CC, Vannberg FO, Chapman SJ, Guo H, Wong SH, Walley AJ, et al. *CISH* and Susceptibility to Infectious Diseases. N Engl J Med [Internet]. 2010;362(22):2092–101. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa0905606

105. Galvão I, Tavares LP, Corrêa RO, Fachi JL, Rocha VM, Rungue M, et al. The Metabolic Sensor GPR43 Receptor Plays a Role in the Control of Klebsiella pneumoniae Infection in the Lung. Front Immunol [Internet]. Frontiers; 2018 Feb 20 [cited 2018 Aug 26];9:142. Available from: http://journal.frontiersin.org/article/10.3389/fimmu.2018.00142/full

106. Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, Di Yu D, et al. Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. Nature [Internet]. 2009 Oct 29 [cited 2018 Aug 26];461(7268):1282–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19865172

107. Chakraborty P. G-Protein-Mediated Signaling and Its Control in Macrophages and Mammalian Cells. Crit Rev Microbiol [Internet]. Taylor & Francis; 2001 Jan 29 [cited 2018 Aug 26];27(1):1–8. Available from: http://www.tandfonline.com/doi/full/10.1080/20014091096666

108. Senga T, Iwamoto S, Yoshida T, Yokota T, Adachi K, Azuma E, et al. LSSIG is a novel murine leukocyte-specific GPCR that is induced by the activation of STAT3. Blood. 2003;101(3):1185–7.

109. Le Poul E, Loison C, Struyf S, Springael JY, Lannoy V, Decobecq ME, et al. Functional characterization of human receptors for short chain fatty acids and their role in polymorphonuclear cell activation. J Biol Chem [Internet]. 2003;278(28):25481–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12711604

110. Corrêa-Oliveira R, Fachi JL, Vieira A, Sato FT, Vinolo MAR. Regulation of immune cell function by short-chain fatty acids. Clin Transl Immunol [Internet]. 2016;5(4):e73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27195116%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4855267

111. Mulenga V, Ford D, Walker AS, Mwansa J, Sinyinza F, Lishimpi K, et al. Effect of cotrimoxazole on causes of death , hospital admissions and antibiotic use in HIV-infected children. 2007;(August 2006).

112. Bwakura-dangarembizi M, Sc M, Kendall L, Sc M, Bakeera-kitaka S, Nahirya-ntege P, et al. A Randomized Trial of Prolonged Co-trimoxazole in HIV-Infected Children in Africa. 2014;41–53.

113. Coutsoudis A, Coovadia M, Kindra G. Time for new recommendations on cotrimoxazole prophylaxis for HIV-exposed infants in developing countries ? 2010;(March):949–50.

114. Evans C, Prendergast AJ. Co-trimoxazole for HIV-exposed uninfected infants. Lancet. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY license; 2017;5:468–9.

115. Coutsoudis A, Kindra G, Esterhuizen T. Impact of cotrimoxazole prophylaxis on the health of breast-fed, HIV-exposed, HIV-negative infants in a resource-limited setting. Aids [Internet]. 2011;25(14):1797–9. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00002030-201109100-00015

116. Homsy J, Dorsey G, Arinaitwe E, Wanzira H, Kakuru A, Bigira V, et al. Protective efficacy of prolonged co-trimoxazole prophylaxis in HIV-exposed children up to age 4 years for the prevention of malaria in Uganda: A randomised controlled open-label trial. Lancet Glob Heal [Internet]. Homsy et al. Open Access article distributed under the terms of CC BY; 2014;2(12):e727–36. Available from: http://dx.doi.org/10.1016/S2214-109X(14)70329-8

117. Lockman S, Hughes M, Powis K, Ajibola G, Bennett K, Moyo S, et al. Effect of co-trimoxazole on mortality in HIV-exposed but uninfected children in Botswana (the Mpepu Study): a double-blind, randomised, placebo-controlled trial. Lancet Glob Heal [Internet]. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND license; 2017;5(5):e491–500. Available from: http://dx.doi.org/10.1016/S2214-109X(17)30143-2

118. Kourtis AP, Wiener J, Kayira D, Chasela C, Sascha R, Hyde L, et al. Health outcomes of HIV-exposed uninfected African infants Athena. AIDS. 2015;27(5):749–59.

119. Lockman S, Hughes M, Powis K, Ajibola G, Bennett K, Moyo S, et al. Effect of co-trimoxazole on mortality in HIV-exposed but uninfected children in Botswana (the Mpepu Study): a double-blind, randomised, placebo-controlled trial. Lancet Glob Heal. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND license; 2017;5(5):e491–500.

120. Flo T, Smith K, Sato S. Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. Nature. 2004;432:917–21.

# APPENDICES

## APPENDIX I: ssh Script for alignment

```
#!/bin/bash
#tools required: bowtie2, tophat, RseQC, samtools, HTSeq, R -> DESeq2 #script
echo "\n |LOADING REQUIRED SOFTWARE\n"
module unload python/3.5.0
module load python/2.7.8
module load samtools/1.3.1
module load bowtie2/2.2.5
module load htseq/0.6.1
module load perl/5.22.0
module load tophat/2.1.0
FASTQFILES="/home/local/KWTRP/zkidiavai/mapping"
q="C123_trimmed.fastq.gz C139_trimmed.fastq.gz C141_trimmed.fastq.gz
C167_trimmed.fastq.gz C210_trimmed.fastq.gz C219_trimmed.fastq.gz
C255_trimmed.fastq.gz C256_trimmed.fastq.gz C273_trimmed.fastq.gz
C275_trimmed.fastq.gz C299_trimmed.fastq.gz C331_trimmed.fastq.gz
C336_trimmed.fastq.gz C342_trimmed.fastq.gz C350_trimmed.fastq.gz
C360_trimmed.fastq.gz C367_trimmed.fastq.gz C371_trimmed.fastq.gz
C372_trimmed.fastq.gz C374_trimmed.fastq.gz C379_trimmed.fastq.gz
C380_trimmed.fastq.gz C394_trimmed.fastq.gz C421_trimmed.fastq.gz
C427_trimmed.fastq.gz C437_trimmed.fastq.gz C438_trimmed.fastq.gz
C442_trimmed.fastq.gz C445_trimmed.fastq.gz C452_trimmed.fastq.gz
C453_trimmed.fastq.gz C455_trimmed.fastq.gz N1212_2012_trimmed.fastq.gz
N1222_2013_trimmed.fastq.gz N1224_2013_trimmed.fastq.gz
N1229_2013_trimmed.fastq.gz N1235_2013_trimmed.fastq.gz
N1247_2013_trimmed.fastq.gz N1236_2013_trimmed.fastq.gz
N1237_2013_trimmed.fastq.gz N1238_2013_trimmed.fastq.gz
N1240_2013_trimmed.fastq.gz N1242_2013_trimmed.fastq.gz
N1243_2013_trimmed.fastq.gz N1244_2013_trimmed.fastq.gz
N1245_2013_trimmed.fastq.gz N1224_2012_trimmed.fastq.gz"
echo "\n |PERFORMING REFERENCE INDEXING\n"
##Bowtie
wget ftp://ftp.ensembl.org/pub/release-
86/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz
gunzip Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz
```

```
bowtie2-build -f Homo_sapiens.GRCh38.dna.primary_assembly.fa GRCh38
echo "\n |REFERENCE INDEXING DONE!\n"
echo "\n |BEGINNING READ ALIGNMENTS\n"
for i in $q; do
        i2="${i//_trimmed.fastq.gz}"
        echo "Checking if alignment is done..."
        if [ ! -f ./$i2/accepted_hits.bam ]; then
                echo "\n |Alignment not done!!!!!!!!!!!!!!"
                echo "=========================================="
                echo "\n |ALIGNING THE READS FROM $i2\n"
                wget ftp://ftp.ensembl.org/pub/release-
86/gtf/homo_sapiens/Homo_sapiens.GRCh38.86.gtf.gz
                gunzip Homo_sapiens.GRCh38.86.gtf.gz
                #server
                tophat2 -o $i2 -G Homo_sapiens.GRCh38.86.gtf  -p 30 --phred64-quals
GRCh38 $FASTQFILES/$i
                echo "\n | Alignment done...."
                echo "=========================================="
                echo "\n |Indexing of bam file\n"
                samtools index -b ./$i2/accepted_hits.bam
                echo "=========================================="
                echo "QUANTITATION OF GENE EXPRESSION"
                ##samtools
                ##HTSeq
                echo "\n|Sorting BAM\n"
                samtools sort -n ./$i2/accepted_hits.bam >
./$i2/$i2"_accepted_hits_namesorted.bam"
                echo "\n |Read counts at gene level\n"
                htseq-count -f bam --stranded=no
./$i2/$i2"_accepted_hits_namesorted.bam" ./Homo_sapiens.GRCh38.86.gtf  >
./$i2/$i2"_htseq_counts.txt"
                cp ./$i2/$i2"_htseq_counts.txt" ./$i2"_htseq_counts_raw.txt"
                echo "\n |Removing last five lines of the counts file|n"
                sed -e :a -e '$d;N;2,5ba' -e 'P;D' ./$i2"_htseq_counts_raw.txt" >
./$i2"_htseq_counts_cleaned.txt"
```

```
                echo "========================================="
        else
                echo "\n | Alignment done...."
                echo "========================================="
                echo "\n |Indexing of bam file\n"
                samtools index -b ./$i2/accepted_hits.bam
                echo "========================================="
                echo "QUANTITATION OF GENE EXPRESSION"
                ##samtools
                ##HTSeq
                echo "\n|Sorting BAM\n"
                samtools sort -n ./$i2/accepted_hits.bam >
./$i2/$i2"_accepted_hits_namesorted.bam"
                echo "\n |Read counts at gene level\n"
                htseq-count -f bam --stranded=no
./$i2/$i2"_accepted_hits_namesorted.bam" ./Homo_sapiens.GRCh38.86.gtf >
./$i2/$i2"_htseq_counts.txt"
                cp ./$i2/$i2"_htseq_counts.txt" ./$i2"_htseq_counts_raw.txt"
                echo "\n |Removing last five lines of the counts file|n"
                sed -e :a -e '$d;N;2,5ba' -e 'P;D' ./$i2"_htseq_counts_raw.txt" >
./$i2"_htseq_counts_cleaned.txt"
                echo "========================================="
        fi
        echo "\n | NEXT FILE\n"
done
echo "\n |combine count files from all samples to a table\n"
join *_htseq_counts_cleaned.txt > counts_table.txt
```

**APPENDIX II: R script for differential expression**

```
setwd("~/Desktop/Thesis /Results")
library("DESeq2")
library("ggplot2")
library("gplots")
library("RColorBrewer")
library("pheatmap")
source("~/Desktop/Thesis /Codes/Functions.R")

#INPUT DATA.....FORMATION OF A DESeqDataSetFromHTSeqCount
directory<- ("/Project/tophat2")  #location of the text files
sampleFiles <- grep("cleaned",list.files(directory),value = TRUE)
samplenames <- c("C123","C139","C141", "C167","C210","C219","C255",
                 "C256","C273","C275","C299","C331","C336","C342",
                 "C350","C360","C367","C371","C372","C374","C379",
                 "C380","C394","C421", "C427","C437","C438","C442",
                 "C445","C452","C453","C455","N1212_2012", "N1222_2013",
                 "N1224_2012","N1224_2013","N1229_2013","N1235_2013",
                 "N1236_2013","N1237_2013","N1238_2013","N1240_2013",
                 "N1242_2013","N1243_2013","N1244_2013","N1245_2013","N1247_2013")


condition <- c("exposed","exposed", "exposed", "exposed",
       "exposed","exposed","exposed","exposed","exposed", "exposed", "exposed",
       "exposed", "exposed","exposed","exposed", "exposed", "exposed",
       "exposed","exposed","exposed", "exposed", "exposed",
       "exposed","exposed","exposed", "exposed", "exposed",
       "exposed","exposed","exposed", "exposed", "exposed", "unexposed", "unexposed",
       "unexposed", "unexposed", "unexposed", "unexposed","unexposed", "unexposed",
       "unexposed","unexposed", "unexposed", "unexposed","unexposed", "unexposed",
       "unexposed")
sampleTable <- data.frame(sampleName = samplenames,
                          fileName = sampleFiles,
                          condition = condition)
dds <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable,
                                  directory = directory,
                                  design= ~ condition)
```

```
dds
dds <- dds[ rowSums(counts(dds)) > 1, ] #filters out rows with only 0 values
dds
dds$condition <- factor(dds$condition, levels=c("exposed","unexposed")) #setting
    reference factor levels
#normalised counts
dds <- estimateSizeFactors(dds)
norm<-counts(dds, normalized=TRUE)
head(norm)
write.csv(as.data.frame(norm),file="normalised results.csv")

rld <- rlog(dds, blind=FALSE)
head(assay(rld), 3)
plotPCAWithSampleNames(rld, intgroup="condition")
outliers <- c( "C299","C453","C367")
DESeq2Table.sub <- dds[, !(colnames(dds) %in% outliers)] #removing outliers
rld <- rlogTransformation(DESeq2Table.sub, blind=TRUE)

## Volcano plot
png("DE_volcanoplot.png", 1200, 1000, pointsize=20)
volcanoplot(res, lfcthresh=1, sigthresh=0.05, textcx=.8, xlim=c(-2.3, 2))
dev.off()

png("Correlation heatmap.png", 1000, 800, pointsize = 10)
sampleDists <- dist(t(assay(rld)))
sampleDistMatrix <- as.matrix(sampleDists)
rownames(sampleDistMatrix) <- paste(rld$condition, sep="-")
colnames(sampleDistMatrix) <-  paste(rld$condition, sep="-")
colors <- colorRampPalette( rev(brewer.pal(9, "YlGnBu")) )(255)
pheatmap(sampleDistMatrix,
        clustering_distance_rows=sampleDists,
        clustering_distance_cols=sampleDists,
        col=colors,
        main = "sample clustering Euclidean")
```

```
dev.off()

#Differential expression
dds <- dds[, !(colnames(dds) %in% outliers)] #removing outliers
dds
dds <- DESeq(dds)
###plotDispEsts(dds)
res <- results(dds)
res
summary(res)

#independent filtering based on the mean of normalized counts for each gene,padj=0.05
res05 <- results(dds, alpha=0.05)
summary(res05)
sum(res05$padj < 0.05, na.rm=TRUE)

#plotMA shows average of counts normalised by size factors
#Points will be colored red if the adjusted p value is less than 0.1

plotMA(res, main="Shrunken L2FC", ylim=c(-2,2)) #the shrinkage of fold changes for
      genes with low counts and high dispersion.
resMLE <- results(dds, addMLE=TRUE) #maximum likelihood estimate (MLE) for the
      log 2 fold change
plotMA(resMLE, MLE=TRUE, main="unshrunken L2FC", ylim=c(-4,4))

#significantly expressed genes
resOrdered <- res[order(res$padj),] #order by the smallest adjusted p value
resSig1 <- subset(resOrdered, padj < 0.05) #order by the smallest adjusted p value;most
      significant at 0.05
resSig2 <- subset(resOrdered, padj < 0.01) #order by the smallest adjusted p value;most
      significant at 0.01

#count the number of significantly differentially expressed genes
nrow(res[res$padj<0.05 & !is.na(res$padj) & res$log2FoldChange<=-
      1|res$log2FoldChange>=1 &  !is.na(res$log2FoldChange) ,]) # at the .05 level
```

```
nrow(res[res$padj<0.01 & !is.na(res$padj)& res$log2FoldChange<=-1
      |res$log2FoldChange>=1 &  !is.na(res$log2FoldChange),]) # at the .01 level


#exporting significantly expressed genes to csv file
write.csv(as.data.frame(res),file="differential exposedvsunexposed results no outliers.csv")
write.csv(as.data.frame(resSig1),file="significant_0.05.csv")
write.csv(as.data.frame(resSig2),file="significant_0.01.csv")
```

## APPENDIX III: Differentially expressed genes (166)

| Gene ID | Gene description | L2FC | Padj |
|---|---|---|---|
| ENSG00000006118 | transmembrane protein 132A | 1.07 | 1.41E-04 |
| ENSG00000091986 | coiled-coil domain containing 80 | 1.15 | 8.70E-09 |
| ENSG00000104918 | resistin | -1.03 | 1.78E-03 |
| ENSG00000113369 | arrestin domain containing 3 | -1.32 | 9.01E-09 |
| ENSG00000114737 | cytokine inducible SH2 containing protein | 1.27 | 1.01E-06 |
| ENSG00000115592 | protein kinase AMP-activated non-catalytic subunit gamma 3 | 1.12 | 1.31E-03 |
| ENSG00000115758 | ornithine decarboxylase 1 | -1.12 | 1.43E-21 |
| ENSG00000118523 | connective tissue growth factor | -1.26 | 7.96E-05 |
| ENSG00000120211 | insulin like 4 | -1.19 | 4.57E-04 |
| ENSG00000120436 | G protein-coupled receptor 31 | 1.00 | 3.90E-03 |
| ENSG00000122877 | early growth response 2 | -1.35 | 1.70E-05 |
| ENSG00000124216 | snail family transcriptional repressor 1 | 1.07 | 1.77E-03 |
| ENSG00000124575 | histone cluster 1 H1 family member d | -1.90 | 5.77E-11 |
| ENSG00000125144 | metallothionein 1G | -1.42 | 1.97E-05 |
| ENSG00000125735 | TNF superfamily member 14 | 1.31 | 3.62E-08 |
| ENSG00000125968 | inhibitor of DNA binding 1, HLH protein | 1.17 | 2.87E-04 |
| ENSG00000126262 | free fatty acid receptor 2 | 1.13 | 5.65E-04 |
| ENSG00000126500 | fibronectin leucine rich transmembrane protein 1 | 1.15 | 1.84E-04 |
| ENSG00000133878 | dual specificity phosphatase 26 | -1.12 | 1.21E-03 |
| ENSG00000143858 | synaptotagmin 2 | 1.05 | 4.86E-05 |
| ENSG00000148346 | lipocalin 2 | -1.04 | 4.10E-04 |
| ENSG00000153446 | chromosome 16 open reading frame 89 | 1.04 | 4.23E-04 |
| ENSG00000158874 | apolipoprotein A2 | 1.22 | 6.82E-06 |
| ENSG00000164047 | cathelicidin antimicrobial peptide | -1.05 | 4.41E-04 |
| ENSG00000164104 | high mobility group box 2 | -1.09 | 3.73E-16 |
| ENSG00000166289 | pleckstrin homology and FYVE domain containing 1 | 1.03 | 1.02E-05 |
| ENSG00000167800 | T-box 10 | 1.02 | 1.18E-03 |
| ENSG00000168269 | forkhead box I1 | 1.51 | 1.42E-06 |
| ENSG00000168298 | histone cluster 1 H1 family member e | -1.55 | 9.66E-10 |
| ENSG00000169242 | ephrin A1 | 1.22 | 2.10E-07 |
| ENSG00000169715 | metallothionein 1E | -1.24 | 4.14E-06 |
| ENSG00000170231 | fatty acid binding protein 6 | 1.07 | 9.05E-04 |
| ENSG00000170345 | Fos proto-oncogene, AP-1 transcription factor subunit | -1.06 | 3.27E-04 |
| ENSG00000170442 | keratin 86 | 1.00 | 4.12E-03 |
| ENSG00000170807 | leiomodin 2 | 1.30 | 7.58E-05 |
| ENSG00000173404 | INSM transcriptional repressor 1 | -1.05 | 1.72E-03 |
| ENSG00000173991 | titin-cap | 1.20 | 3.69E-09 |
| ENSG00000174007 | centrosomal protein 19 | 1.02 | 1.32E-05 |
| ENSG00000174469 | contactin associated protein like 2 | -1.13 | 1.69E-04 |
| ENSG00000174611 | kyphoscoliosis peptidase | 1.47 | 1.56E-08 |
| ENSG00000176641 | ring finger protein 152 | -1.14 | 1.05E-04 |
| ENSG00000178115 | golgin A8 family member Q | 1.37 | 6.87E-06 |
| ENSG00000180720 | cholinergic receptor muscarinic 4 | -1.45 | 9.03E-10 |
| ENSG00000181126 | major histocompatibility complex, class I, V (pseudogene) | 1.35 | 4.75E-05 |
| ENSG00000181778 | transmembrane protein 252 | 1.10 | 7.20E-04 |
| ENSG00000181963 | olfactory receptor family 52 subfamily K member 2 | 1.22 | 1.24E-04 |
| ENSG00000182107 | transmembrane protein 30B | -1.02 | 5.61E-08 |
| ENSG00000182782 | hydroxycarboxylic acid receptor 2 | 1.06 | 1.42E-04 |

| | | | |
|---|---|---|---|
| ENSG00000183160 | transmembrane protein 119 | -1.43 | 1.32E-09 |
| ENSG00000184260 | histone cluster 2 H2A family member c | -1.17 | 2.34E-08 |
| ENSG00000184270 | histone cluster 2 H2A family member b | -1.22 | 8.24E-05 |
| ENSG00000184897 | H1 histone family member X | -1.43 | 2.45E-19 |
| ENSG00000187730 | gamma-aminobutyric acid type A receptor delta subunit | -1.37 | 4.01E-05 |
| ENSG00000188152 | NUT family member 2G | 1.02 | 1.21E-05 |
| ENSG00000189060 | H1 histone family member 0 | -1.39 | 3.23E-10 |
| ENSG00000189120 | Sp6 transcription factor | -1.09 | 3.51E-04 |
| ENSG00000199370 | Uncharacterised | -1.19 | 5.75E-05 |
| ENSG00000203914 | heat shock protein 90 beta family member 3, pseudogene | -1.01 | 3.73E-03 |
| ENSG00000204882 | G protein-coupled receptor 20 | 1.20 | 1.05E-04 |
| ENSG00000205426 | keratin 81 | 1.26 | 2.43E-04 |
| ENSG00000205786 | long intergenic non-protein coding RNA 1531 | 1.08 | 9.72E-07 |
| ENSG00000206066 | immunoglobulin lambda like polypeptide 3, pseudogene | -1.18 | 6.65E-04 |
| ENSG00000211684 | immunoglobulin lambda joining 7 | -1.02 | 3.85E-03 |
| ENSG00000211685 | immunoglobulin lambda constant 7 | -1.19 | 2.09E-05 |
| ENSG00000212541 | RNA, U6 small nuclear 510, pseudogene | 1.00 | 2.62E-03 |
| ENSG00000213073 | Uncharacterised | 1.22 | 3.20E-17 |
| ENSG00000213613 | ribosomal protein L11 pseudogene 3 | 1.04 | 1.08E-03 |
| ENSG00000218073 | Uncharacterised | 1.06 | 3.83E-06 |
| ENSG00000218428 | NIP7, nucleolar pre-rRNA processing protein pseudogene 3 | 1.09 | 1.83E-03 |
| ENSG00000218809 | Uncharacterised | 1.23 | 7.28E-09 |
| ENSG00000222009 | BTB domain containing 19 | 1.00 | 3.10E-04 |
| ENSG00000223547 | zinc finger protein 844 | -1.27 | 1.30E-09 |
| ENSG00000223668 | eukaryotic translation elongation factor 1 alpha 1 pseudogene 24 | -1.01 | 1.96E-04 |
| ENSG00000223803 | ribosomal protein S20 pseudogene 14 | -1.03 | 1.18E-03 |
| ENSG00000223825 | DAZ associated protein 2 pseudogene 1 | -1.04 | 2.23E-03 |
| ENSG00000224950 | Uncharacterised | 1.29 | 3.49E-08 |
| ENSG00000225131 | proteasome activator subunit 2 pseudogene 2 | -1.01 | 1.96E-03 |
| ENSG00000227560 | ribosomal protein S15a pseudogene 30 | 1.04 | 2.74E-03 |
| ENSG00000228329 | long intergenic non-protein coding RNA 1890 | 1.00 | 2.09E-05 |
| ENSG00000229808 | Uncharacterised | 1.05 | 3.92E-06 |
| ENSG00000230225 | mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 5 pseudogene 14 | 1.06 | 5.22E-04 |
| ENSG00000230470 | Uncharacterised | 1.09 | 1.66E-03 |
| ENSG00000230679 | ENO1 antisense RNA 1 | -1.07 | 2.99E-04 |
| ENSG00000231964 | Uncharacterised | 1.17 | 3.00E-11 |
| ENSG00000232133 | inosine monophosphate dehydrogenase 1 pseudogene 10 | 1.09 | 8.28E-04 |
| ENSG00000232216 | immunoglobulin heavy variable 3-43 | -1.11 | 9.33E-05 |
| ENSG00000233030 | Uncharacterised | 1.06 | 8.18E-04 |
| ENSG00000233214 | Uncharacterised | 1.13 | 6.91E-04 |
| ENSG00000233360 | Uncharacterised | 1.07 | 1.17E-04 |
| ENSG00000233469 | ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 4 pseudogene 1 | 1.08 | 9.25E-04 |
| ENSG00000233602 | ERI3 intronic transcript 1 | 1.13 | 7.59E-04 |
| ENSG00000233668 | Uncharacterised | 1.06 | 1.28E-04 |
| ENSG00000234290 | Uncharacterised | 1.01 | 8.67E-08 |
| ENSG00000235328 | Uncharacterised | 1.06 | 8.47E-04 |
| ENSG00000236878 | mitochondrially encoded ATP synthase 6 pseudogene 26 | 1.07 | 2.11E-03 |
| ENSG00000238025 | zinc finger DHHC-type containing 4 pseudogene 1 | 1.15 | 4.95E-04 |
| ENSG00000238279 | Uncharacterised | 1.13 | 1.13E-06 |
| ENSG00000239350 | Uncharacterised | 1.09 | 1.73E-03 |
| ENSG00000239367 | RNA, 7SL, cytoplasmic 477, pseudogene | 1.12 | 2.21E-04 |
| ENSG00000239503 | microtubule affinity regulating kinase 2 pseudogene 8 | 1.10 | 4.71E-04 |

| | | | |
|---|---|---|---|
| ENSG00000239552 | HOXB cluster antisense RNA 2 | 1.01 | 3.09E-03 |
| ENSG00000241157 | Uncharacterised | 1.24 | 9.15E-05 |
| ENSG00000242265 | paternally expressed 10 | -1.04 | 3.14E-05 |
| ENSG00000242609 | Uncharacterised | 1.11 | 9.05E-04 |
| ENSG00000244124 | ATP1B3 antisense RNA 1 | 1.16 | 1.54E-04 |
| ENSG00000248323 | lung cancer associated transcript 1 (non-protein coding) | 1.29 | 1.28E-08 |
| ENSG00000249119 | mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 6 pseudogene 4 | -1.10 | 1.72E-03 |
| ENSG00000249396 | long intergenic non-protein coding RNA 2212 | 1.05 | 2.66E-03 |
| ENSG00000250182 | eukaryotic translation elongation factor 1 alpha 1 pseudogene 13 | 1.06 | 2.05E-03 |
| ENSG00000250254 | pituitary tumor-transforming 2 | 1.07 | 3.05E-06 |
| ENSG00000251048 | Uncharacterised | 1.04 | 3.13E-03 |
| ENSG00000251639 | Uncharacterised | 1.03 | 3.08E-03 |
| ENSG00000251867 | Uncharacterised | -1.01 | 3.56E-08 |
| ENSG00000253269 | Uncharacterised | 1.56 | 1.21E-07 |
| ENSG00000254484 | Uncharacterised | 1.07 | 1.86E-03 |
| ENSG00000254554 | Uncharacterised | 1.00 | 4.41E-04 |
| ENSG00000255398 | hydroxycarboxylic acid receptor 3 | 1.08 | 1.52E-04 |
| ENSG00000255893 | Uncharacterised | 1.13 | 3.10E-05 |
| ENSG00000256249 | Uncharacterised | 1.56 | 5.00E-07 |
| ENSG00000256393 | ribosomal protein L41 pseudogene 5 | -1.07 | 2.05E-03 |
| ENSG00000256812 | calpain small subunit 2 | 1.27 | 3.26E-05 |
| ENSG00000257017 | haptoglobin | -1.21 | 7.43E-05 |
| ENSG00000258086 | Uncharacterised | 1.42 | 3.41E-06 |
| ENSG00000258928 | Uncharacterised | 1.15 | 5.36E-04 |
| ENSG00000259374 | NADH:ubiquinone oxidoreductase subunit B4 pseudogene 11 | 1.06 | 1.80E-03 |
| ENSG00000259671 | mitochondrially encoded cytochrome b pseudogene 23 | 1.05 | 2.01E-03 |
| ENSG00000259781 | high mobility group box 1 pseudogene 6 | -1.01 | 3.63E-03 |
| ENSG00000260145 | Uncharacterised | 1.33 | 2.73E-07 |
| ENSG00000261655 | Uncharacterised | 1.04 | 9.50E-04 |
| ENSG00000262874 | chromosome 19 open reading frame 84 | 1.06 | 3.10E-04 |
| ENSG00000263606 | Uncharacterised | 1.08 | 3.58E-04 |
| ENSG00000264924 | Uncharacterised | 1.00 | 6.73E-04 |
| ENSG00000266265 | Kruppel like factor 14 | -1.07 | 3.29E-04 |
| ENSG00000266777 | SH3 domain containing GRB2 like 1, endophilin A2 pseudogene 1 | 1.25 | 6.09E-12 |
| ENSG00000267293 | Uncharacterised | 1.06 | 2.72E-05 |
| ENSG00000267325 | long intergenic non-protein coding RNA 1415 | -1.02 | 2.99E-04 |
| ENSG00000267501 | Uncharacterised | 1.09 | 2.13E-04 |
| ENSG00000267607 | Uncharacterised | 1.17 | 6.90E-04 |
| ENSG00000268355 | Uncharacterised | -1.05 | 1.43E-07 |
| ENSG00000268618 | Uncharacterised | 1.17 | 1.23E-04 |
| ENSG00000269028 | MT-RNR2 like 12 | -1.10 | 4.29E-04 |
| ENSG00000270168 | Uncharacterised | 1.03 | 3.11E-03 |
| ENSG00000270614 | Uncharacterised | 1.05 | 2.21E-03 |
| ENSG00000271840 | Uncharacterised | 1.10 | 1.01E-04 |
| ENSG00000272931 | Uncharacterised | -1.13 | 6.38E-10 |
| ENSG00000273010 | Uncharacterised | -1.08 | 4.55E-05 |
| ENSG00000273599 | Uncharacterised | 1.23 | 9.10E-16 |
| ENSG00000274156 | Uncharacterised | 1.00 | 3.90E-03 |
| ENSG00000274602 | phosphatidylinositol 4-kinase alpha pseudogene 1 | 1.09 | 1.46E-04 |
| ENSG00000275022 | microRNA 6753 | 1.06 | 2.23E-05 |
| ENSG00000275401 | Uncharacterised | 1.00 | 2.86E-03 |
| ENSG00000275494 | Uncharacterised | -1.11 | 6.30E-07 |

| | | | |
|---|---|---|---|
| ENSG00000275562 | Uncharacterised | 1.09 | 1.83E-03 |
| ENSG00000276085 | C-C motif chemokine ligand 3 like 1 | -1.29 | 2.94E-05 |
| ENSG00000276842 | Uncharacterised | 1.05 | 1.86E-07 |
| ENSG00000277595 | Uncharacterised | 1.05 | 1.94E-07 |
| ENSG00000277632 | C-C motif chemokine ligand 3 | -1.25 | 2.30E-06 |
| ENSG00000278356 | Uncharacterised | -1.04 | 1.21E-04 |
| ENSG00000278937 | Uncharacterised | 1.21 | 2.16E-05 |
| ENSG00000279187 | Uncharacterised | 1.00 | 7.60E-05 |
| ENSG00000279821 | Uncharacterised | 1.20 | 1.41E-07 |
| ENSG00000279946 | Uncharacterised | 1.15 | 1.01E-04 |
| ENSG00000280537 | Uncharacterised | 1.28 | 1.71E-05 |
| ENSG00000281708 | ERC2 intronic transcript 1 | 1.02 | 3.04E-03 |
| ENSG00000281912 | long intergenic non-protein coding RNA 1144 | -1.17 | 9.77E-07 |
| ENSG00000283341 | Uncharacterised | 1.06 | 7.43E-05 |