UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Mass spectrometry metabonomics of HIV-1 sera

by

Khanyisile Kgoadi

Submitted in partial fulfillment of the requirements for the degree:

***Magister Scientiae* Biochemistry**

In the Faculty of Natural and Agricultural Sciences

University of Pretoria

Pretoria

South Africa

30 September 2014

SUBMISSION DECLARATION

I, Khanyisile Kgoadi declare that the dissertation which I hereby submit for degree *Magister Scientiae* in the Department of Biochemistry, at the University of Pretoria is my own work and has not been previously submitted by me for a degree at this or any other tertiary institute.

Signature: Date: 30 September 2014

DEDICATION

I would like to dedicate this dissertation to my parents, Ephraim Mtshiyane Kgoadi and Donalia Kgoadi. You have been supportive not only in my studies but in all aspects of my life. Thank you for believing in me and trusting me to make the right decisions in life because you taught me how to. Words alone are not sufficient to express my gratitude for everything that you are doing for me. I am eternally grateful to God, the Almighty for blessing me with such wonderful parents. As parents you have more than proved your selflessness nature by the sacrifices you are constantly making for me and my siblings, especially making sure that our studies do not suffer. Mum and Dad, just know that I always think highly of you, I am proud of you. I am who and what I am today because of your upbringing. Forever in my heart, thank you.

I would also like to dedicate the dissertation to my uncle Chris Mthuthuzeli Ngaleka, another guardian that has been encouraging, motivating and supportive of my studies since high school until today. You always kept me going. Thank you very much *Malume* for taking it upon yourself to further my growth as an individual. Much appreciation and love in this regard and others.

To my siblings, please follow my example and also work hard to achieve your dreams. To Chabbie (Sechaba Kgoadi), my baby brother, you are such an inspiration to all of us at home. We love and treasure you so much, hope that you continue growing into a wonderful young man and pursue your dreams. You are the leading person in my life. Just the thought of you brings joy, laughter, inspiration and motivation, you make me a better person and help me reach for my goals, be happy, a proud sister and daughter. I hope you receive the best that life has to offer. Love you little guy!

UNIVERSITY OF PRETORIA

FACULTY OF NATURAL AND AGRICULTURAL SCIENCES

DEPARTMENT OF BIOCHEMISTRY

Full name: **Khanyisile Kgoadi** Student number: **26158648**

Title of the work:

Mass spectrometry metabonomics of HIV-1 sera

Declaration

1. I understand what plagiarism entails and am aware of the University's policy in this regard.
2. I declare that this **Dissertation** (e.g. essay, report, project, assignment, dissertation, thesis etc) is my own, original work. Where someone else's work was used (whether from a printed source, the internet or any other source) due acknowledgement was given and reference was made according to departmental requirements.
3. I did not make use of another student's previous work and submit it as my own.
4. I did not allow and will not allow anyone to copy my work with the intention of presenting it as his or her own work.

Signature _____ Date: **30 September 2014**

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to the following people, organizations and institutions that contributed in making this research possible and a success.

- To **GOD, the ALMIGHTY** who made all things possible by not only directing me to the University of Pretoria and under the right guidance for this specific project, but also comforting me in hard times when I felt like giving up on my goals. All these chain of events were fate. Thank you for making this work happen and a success. I did my bit and you took care of the rest, Thank you for everything.
- To my supervisor, **Prof Debra Meyer**. I did not choose your supervision; the project I chose led me to you. Through working with you, I realized I ended up in the field I have always wanted. I can confidently say I ended up with the greatest mentor to train my growth as a scientist. You have presented me with a platform to fulfill my long-term goal of making a positive difference in the lives of people affected and infected by HIV. I welcome your critique of my work and conduct as a student because I know it is aimed at training me to be a better researcher and person, thank you for being patient with me. I loved my project, and it hasn't always been smooth sailing, there were a lot of challenges. However your constant insight, critique, guidance, alternatives methods, motivation and believing in me kept me, and the project going. Under your training and supervision I have acquired the knowledge and skills that has made me into a better scientist, colleague, fellow student and person.
- To my mentor, **Dr Aurelia Williams**. Words alone are not enough to express the impact that your guidance and friendship has produced in my life, both personally and professionally. I was so lucky to be co-supervised by you in my honours project and building a strengthened professional relationship that keeps growing even today. You have been like a big sister carrying the GPS with the directions I needed to lead me to my research destination. Thank you for your teachings that I have successfully assilimated together with motivation and encouragement to reach for greater heights.
- To **Prof Paul Steenkamp**. Thank you for your teachings, patience and guidance through this work, much appreciation.
- To **Prof Francois Steffens**. Thank you for your wise statistical teachings and the time you dedicated to this study. Your welcoming nature to help and sense of urgency when you were contributing to this work was very much appreciation.

- I would like to thank all the **authors** of the paper for their input and cooperation in the successful publication that resulted from a portion of my masters' work. This is the first publication I have been involved in.
- To **Dr Edwin Ntakadzeni Madala**. Thank you for your valuable input and assistance in data processing for this study. Not forgetting the encouragement you gave me when you were my demi in the 3rd year of my undergraduate degree and warned me to pull up my socks and up my game by working harder and aiming high.
- To the **Funders**. My gratitude and appreciation to the South African National Research Foundation (NRF), the Medical Research Council (MRC), the Technology Innovation Agency (TIA) and the University of Pretoria's Faculty of Natural and Agricultural sciences for the financial assistance provided for this study and degree.
- I would like to thank the **students** and **Sister Sibongile Mdletshe** at the University of Pretoria's Student Health Services (SHS). The **nurses/doctors** and **patients** from the various hospitals where blood samples for this research was obtained. Without this starting material, this study wouldn't have been possible, much appreciation.
- To my **Colleagues.** Thank you for providing me with a support system and encouragement throughout this journey, especially **Sindisiwe Nondaba**; girl, you have been more than helpful and resourceful in the completion of this project. Thank you for everything!
- To my **family and friends.** I apologize for constantly crying to you about the hardships experienced through this degree. Thank you for your unconditional and ongoing support. I really appreciate it.

PREFACE

Part of the research presented in this thesis has been published in a peer-reviewed journal and presented at both a national and international conferences. A copy of the published manuscript in its submitted word version is provided in section 4.1.

Publication

Published manuscript

- Aurelia Williams*, Khanyisile Kgoadi*, Francois Steffens, Paul Steenkamp and Debra Meyer (2014) UPLC-MS Reveals Perturbed Metabolites in HIV-Infected Sera. Current Metabolomics. Issue 1. Volume 2. 37-52.
  DOI: 10.2174/2213235X02666140214201251

(* joint first authorship)

Conferences

Portions of this work were presented at:

- The HIV DART conference held in San Diego (California, USA) which took place on the 4-7th December 2012. Poster Presentation by Khanyisile Kgoadi, Aurelia Williams, Paul Steenkamp, Francois Steffens and Debra Meyer (2012) entitled: Tandem Mass Spectrometry for the Identification of Biomarkers of HIV-1 Infection.
- The 6th SA AIDS conference held in Durban (South Africa) which took place on the 18-21st of June 2013. Poster Presentation by Khanyisile Kgoadi, Aurelia Williams, Paul Steenkamp, Francois Steffens and Debra Meyer (2013) entitled: Tandem Mass Spectrometry for the Identification of Biomarkers of HIV-1 Infection.

Awards

- I was granted a scholarship from the organizing committee of the 6th SA AIDS conference for attending the conference in 2013.

SUMMARY

Mass Spectrometry metabonomics of HIV-1 sera

By

Khanyisile Kgoadi

Supervisor: **Prof Debra Meyer**

Department: **Biochemistry**

Degree: **MSc Biochemistry**

Background: Metabolic complications resulting from the human immunodeficiency virus type 1 (HIV-1) and the acquired immunodeficiency syndrome (AIDS) are as common as the immune system disruption caused by the virus, but not as well known. Highly active antiretroviral therapy (HAART) used to treat HIV-1 infection exacerbates the effects HIV-1 has on the host's metabolism. Common metabolic complications such as insulin resistance, lipodystrophy, lactic acidosis and others contribute to morbidity and mortality during HIV/AIDS. The detection of HIV-1 related metabolic biomarkers assists in diagnosing and monitoring metabolic complications, however, limitations of the conventional methodologies used for detecting these molecules caused a paucity of data on HIV-related metabolic indicators. Metabonomics, the ability to measure multiple metabolites simultaneasously, shows promise in distinguishing HIV-1 negative and positive patients through nuclear magnetic resonance (NMR) and vibrational spectroscopy as well as mass spectrometry (MS) profiles of various biofluids. The objective of this study was to determine the abilities of ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) in the identification of serum metabolites associated with HIV and/or HAART. The study was also aimed at determining whether data from a less sensitive technique, Fourier transform infrared (FTIR) spectroscopy would lead to comparable conclusions as those derived using UPLC-MS data.

Methods: Sera were collected from three experimental groups; HIV negative (n=32), HIV positive (n=29) and HIV positive patients receiving HAART (n=34). Metabolites were extracted using a conventional approach of cold methanol extraction as well as the Ostro^{TM} plate extraction technology which involved filtration by positive pressure. The filtrate was analysed in the negative and positive electrospray ionization (ESI) modes of UPLC-MS. Serum samples were also dried overnight and analysed using FTIR. Data processing and chemometric analysis was carried out using the SPSS 19.0 and MassLynx v4.1 software

packages. Following extensive statistical evaluation of data, bioinformatics approaches that assisted with metabolite identification were conducted.

Results: The combination of Ostro™ plates and UPLC-MS produced high resolution chromatograms that showed visible differences among the serum samples of HIV negative, HIV positive and HIV positive patients receiving HAART. Linear discriminant analysis (LDA) classified experimental groups into the correct categories with great accuracy (>88%), using potential biomarkers responsible for the observed group variations. Principal component analysis (PCA) showed clear separations as well as some overlap among the three experimental groups. Orthogonal projections to latent structures-discriminant analysis (OPLS-DA) showed clear differences between two classes of samples at a time and potential biomarkers were selected from accompanying S-Plots. Hundred and twelve distinct group distinguishing metabolites detected from both ESI positive and ESI negative modes were significantly altered ($p<0.05$). HIV and/or HAART altered metabolites of energy, neuronal and mitochondrial processes were identified and were evident in the amino acid, carbohydrate, lipid and nucleoside/nucleotide metabolic products being detected. Antiviral drugs [mostly nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs)], anti-inflammatory and anti-cancer drugs, nutrient supplements and other drugs associated with the treatment of HIV conditions also contributed to class variations. FTIR generated metabolic patterns that separated the three experimental groups on an LDA scatter plot which achieved correct (>87%) classification accuracy. The significantly altered patterns indicated hydroxyl and alkene group vibrations and these groups were present in the metabolites identified by MS.

Conclusion: Ostro™ plates and UPLC-MS successfully purified, detected and identified sera metabolites distinguishing HIV and/or HAART patients. The different statistical analysis methods applied in this study were in agreement and the OPLS-DA statistical tool complemented the sensitivity of UPLC-MS for the detected distinguishing metabolites. The approach employed here delivered promising findings for use in the discovery of metabolic biomarkers. Distinguishing metabolites identified could be traced to HIV-infection and/or treatment. Findings from this study corroborated with others which showed that NRTIs remain a challenge in the era of HAART toxicities, especially their dominant effect on mitochondrial dysfunction. This work therefore suggests the use of UPLC-MS in HIV disease diagnosis, prognosis, monitoring of treatment success or failure and the ability to link

treatment to metabolic complications. Even though FTIR is less sensitive than UPLC-MS, it was successful in detecting metabolic patterns that corresponded to some metabolites detected by UPLC-MS. This suggests that this easier to perform technique also has potential clinical application in monitoring HIV/AIDS.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

LIST OF TABLES

LIST OF IMPORTANT ABBREVIATIONS

2D	Two-dimensional space
3D	Three-dimensional space
3TC	Lamivudine
ABC	Abacavir
ADA	Adenosine deaminase
AIDS	Acquired immunodeficiency syndrome
ANOVA	Analysis of variance
ARVs	Antiretrovirals
ATP	Adenosine triphophate
ATR	Attenuated total reflectance
AZT	Zidovudine
BPI	Base peak ion
C	Carbon
CAD	Coronary artery disease
d4T	Stavudine
d-CTP	Deoxycytidine triphosphate
DHEAS	Dehydroepiandrosterone sulphate
DModX	Distance to model in x-space
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
EFV	Efavirenz
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionization
F	Flourine
FDA	Food and Drug Administration

FMN	Flavin monnucleotide
$FMNH_2$	Reduced flavin mononucleotide
FTIR	Fourier transform infrared spectroscopy
GC-MS	Gas chromatography-mass spectrometry
GSH	Glutathione
GTP	Guanosine triphosphate
H	Hydrogen
H^1 NMR	Proton nuclear magnetic resonance spectroscopy
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HIV-1	Human immunodeficiency virus type 1
HMDB	Human metabolome database
HPLC	High performance liquid chromatography
IR	Insulin resistance
ITP	Inosine triphosphate
LC	Liquid chromatography
LC-MS	Liquid chromatography- mass spectrometry
LC-MS/MS	Liquid chromatography tandem masss spectrometry
LDA	Linear discriminant analysis
LLE	Liquid-liquid extraction
MS	Mass spectrometry
N	Nitrogen
MS/MS	Tandem mass spectrometry
NIAID	National Institute of Allergy and Infectious Diseases
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance spectroscopy

NNRTIs	Non-nucleoside reverse transcriptase inhibitors
NRTIs	Nucleoside/nucleotide reverse transcriptase inhibitors
NVP	Nevirapine
O	Oxygen
OPLS-DA	Orthogonal projections to latent structures-discriminant analysis
P	Phosphate
PCA	Principal Component Analysis
PDA	Photoiodide array
PNP	Purine nucleotide phosphorylase
Q-TOF	Quadrupole time-of-flight
RNA	Ribonucleic acid
SPE	Solid phase extraction
SPSS	Statistical Package for the Social Sciences
TOF	Time-of-flight
UPLC	Ultra performance liquid chromatography
UPLC-MS	Ultra performance liquid chromatography-mass spectrometry
VIP	Variable importance projection
XS	Extended Statistics
XIC	Extracted chromatogram

CHAPTER 1: INTRODUCTION

1) Introduction

HIV-1 suppresses the immune system by decreasing CD4 cell count which contributes to an increased viral load, which eventually causes AIDS (Watstein and Jovanovic, 2003; Libman and Makadon, 2007). Since its discovery in 1981, HIV/AIDS has claimed the lives of >25 million people. The global prevalence of HIV infection was estimated to be 34 million in 2011 (UNAIDS, 2012) and sub-Saharan Africa is the most affected region in the world with about 70% of the total prevalence. HIV positive women in sub-Saharan Africa exceed the global percentage of positive women by nine percent (UNAIDS, 2012). South Africa is home to 12 % of the global population but has the highest HIV infections in the world, with an HIV prevalence of 5.26 million in 2013 (Statistics South Africa, 2013). Of the 5.26 million HIV positive patients, about 49% are receiving HIV treatment (Lopez Gonzalez, 2013).

The introduction of HAART in 1996 made it possible to delay the onset of AIDS. In undeveloped countries access to treatment still poses a huge challenge due to the lower levels of income and the cost of treatment. Commonly used HAART regimens in sub-Saharan Africa are nucleoside/nucleosttide reverse transcriptase inhibitors (NRTIs), non-nucleotide reverse transcriptase inhibitors (NNRTIs) and protease inhibitors (PIs). HAART regimens target the HIV life cycle at different stages to decrease and possibly prevent replication, which allows survival of the host CD4 cells (Brik & Wong, 2003). This lower viral load allows the weakened immune system to recover, however, the virus drugs do however disrupts host metabolism and HAART treatment augments this effect on the metabolic system in an attempt to restore homeostasis. HIV alters the overall chemical reactions occurring in the host's biological system by disrupting metabolic pathways of molecules such as carbohydrates, amino acids, lipids and nucleotides. HAART toxicities persist and there are ongoing studies to modify and improve the treatment regimens for fewer or no side-effects.

The HIV targets the host metabolic pathways during viral replication, hence disrupting metabolism, which leads to metabolic complications (Chen et al., 2012). Metabolic complications include; insulin resistance, lipodystrophy, lactic acidosis and may result in death (Pinti et al., 2006). Mitochondrial dysfunction is the most documented and studied HIV related metabolic complication and known to a side effect to NRTIs (Carr et al., 2002;

Apostolova & et al, 2006, 2011; Pinti et al., 2011; Williams et al., 2012). During HIV infection, mitochondrial damage occurs as viral ribonucleic acids (RNA) accumulates in the mitochondria instead of the cytoplasm or the nucleus (Somasundaran et al., 1994; Pinti et al., 2011).

Metabolic dysfunctions like opportunistic infections have been implicated in HIV/AIDS complications, leading to death. Because HIV induced metabolic complications are also augmented by HAART, it makes sense to investigate ways to address metabolic disruptions starting with quicker detection of markers of metabolic disorders for better management of HIV/AIDS. Limitations of conventional methods sensitivity and specificity used for assessing HIV-related metabolism problems include lower detection of a single metabolite per assay using enzyme-linked immonosorbent assay (ELISA) (Holvoet, 2008). Recently, high-throughput, high resolution and high sensitivity analytical tools such as nuclear magnetic resonance (NMR) spectroscopy, gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS), Fourier transform Infrared (FTIR) spectroscopy that are employed in metabonomics studies have been successful in discovering disease-related metabolic biomarkers and distinguishing healthy compared to diseased samples. Nicholson et al (1999) defined metabonomics, as "the quantitative measurement of multiparametric responses of a living organism to pathophysiological stimuli or genetic modification."

Metabonomics investigations of HIV/AIDS biofluids has been successfully performed with NMR, GC-MS, UPLC-MS and IR and some of which are detailed in section 2.6 (Hewer et al., 2006; Philippeos et al., 2009; Ghannoum et al., 2011, Hollenbaugh et al., 2011; Williams et al., 2012). Sample preparations for use in metabonomics are normally laborious and time-consuming (Luque-Garcia & Neubert, 2007). For example, sample preparation for GC-MS requires chemical derivatization prior to analysis (Wu et al., 2009) and the derivitization process results in sample loss and increased sample variability. Other metabolite properties like volatility or masking of metabolites by other molecules make it necessary to optimize sample preparation methods. Tavazzi et al (2005) has shown that minimum sample preparation is favorable for reliable analytical data. Currently no study has utilized serum samples to study HIV and/or HAART metabolites that are detectable by UPLC-MS.

With the objective of this study being to determine the abilities of UPLC-MS to identify HIV and/or HAART associated metabolites, serum metabolic profiles of HIV negative and HIV positive patients (receiving and not receiving HAART) were investigated. Analysis was conducted in both ESI- and ESI+ mode to cover a broader dynamic range of detectable compounds. Some compounds ionize better in one mode, while some ionize in both modes and others have specific preferential ionization modes. Villas-Boas et al (2007) reported that amino acids, amines, sugars, and nucleotide bases are detected in the positive ionization mode and organic acids are detected in the negative ionization mode. This study followed an untargeted approach in an attempt to develop a global serum metabolic profile of HIV negative, HIV positive and HIV positive patients receiving HAART. UPLC-MS samples were analyzed in two batches, where the first batch was used for the development of sample preparation protocol with minimal sample manipulation. The first sample batch was also used to optimize the UPLC-MS protocol for sera sample analysis. An easy to use, fast, and reproducible solid phase extraction (SPE) protocol in the form of Ostro^{TM} plates was used for the production of reliable UPLC-MS test material. The collected data was published. The second batch of data investigated multivariate statistical analysis as a tool to enable specific assignments of significantly affected metabolites to HIV, HAART or both. Both data sets resulted in chromatograms of greater resolving power with narrow, distinct peaks that were free of matrix effects. These chromatograms showed visible differences of the metabolic profiles of the three experimental groups (HIV negative, HIV positive and HAART).

Although MS gives vastly more superior biofluid metabonomics data than that collected with vibrational spectroscopy, the later method was also investigated because it is simpler to perform. In metabonomics, the use of multi-platforms for metabolome coverage is recommended because of advantages and limitations of different analytical techniques (Dunn et al., 2012; Tugizima et al., 2014). An objective of this work was also to determine how comparable the data from two instruments would be. Serum samples that were analyzed using FTIR showed differentiable metabolic profiles of the three experimental groups through LDA and overlaid spectra. However, only two spectral regions (910 cm^{-1} and 3915 cm^{-1}) were significantly altered ($p<0.05$); that was the OH and $RCH=CH_2$, regions which corresponded to the hydroxyl and alkene vibrational groups. These vibrational bands correspond to functional groups of amino acids, saccharides and lipids (Bahmani et al., 2009).

Chromatograms showed visible differences among the three experimental groups that were amplified using chemometrics. LDA achieved great classification accuracy (>88%) of HIV negative, HIV positive and HIV positive patients receiving HAART treatment, while PCA showed both separations and overlaps of some samples from the three experimental groups. Mean plots of LDA distinguishing metabolites showed changes in metabolite levels among the three experimental groups, while OPLS-DA identified potential biomarkers between two groups. For statistical analysis employing different statistical packages or methods, multivariate analysis was consistent with univariate analysis.

Detected and putatively identified HIV and/or HAART altered metabolites affected energy metabolism, mitochondrial metabolism and neuronal metabolism. This study showed enhanced mitochondrial damage, from increased accumulation of nucleosides/nucleotides in the mitochondria, disrupted electron transport chain, to decreased levels in metabolites responsible for maintaining membrane integrity that were mainly caused by NRTIs metabolites and derivatives. Other detected drug metabolites consisted mostly of; anti-HIV drugs, anti-inflammatory agents, anti-oxidants, anti-cancer, nutrient supplements etc. These drugs are used for the treatment of HIV and HIV-associated conditions such as cancer.

Findings of this study show the advantageous use of UPLC-MS biofluid metabonomics for detection of metabolic biomarkers involved in metabolic complications. The potential of this approach could achieve an overall goal of developing metabonomics as additional prognostic tool for the management of HIV/AIDS. The discovery of disease biomarkers can aid in the improvement of therapies because this approach offers better monitoring of treatment effects.

The organization of this dissertation is as follows; the next chapter is the literature review (chapter 2) that contains background on HIV/AIDS, HAART, metabolic complications, metabonomics, chemometrics and a brief summary of some metabonomics studies applied to HIV/AIDS. This is followed by the rationale of this study, then hypothesis, objectives and aims. Chapter 3 contains the materials and methods used to conduct this study. Next chapter is the results chapter (chapter 4) that contains the published manuscript of the work conducted on the first batch of serum samples analyzed using UPLC-MS (4.1) and presents the results of the second batch data of UPLC-MS and FTIR analysis (4.2 and 4.3 respectively). The discussion of section 4.2 and 4.3 results follows in chapter 5, and then followed by the overall discussion of section 4.1 and 4.2 findings. The overall conclusion of

the entire study is presented in chapter 6, as are the future perspectives. A comprehensive list of references is also provided followed by an appendix, which contains all supporting information and tables/graphs.

CHAPTER 2: LITERATURE REVIEW

2.1) Brief introduction to HIV-1 and AIDS

The human immunodeficiency virus type 1 (HIV-1) was identified as a lentivirus belonging to the retroviridae family of viruses in 1981. Its 10kb genome consists of single stranded ribonucleic acid (RNA) (Watstein & Jovanovic, 2003; Sierra et al., 2005; Levy, 2007; Ramaih, 2008). HIV-1 is mainly transmitted through the contact or exchange of bodily fluids (Watstein & Jovanovic, 2003; Ramaih, 2008; Libman & Makadon, 2009). Once the host is infected, HIV-1 targets and destroys the white blood cells of the immune system (Watstein & Jovanovic, 2003; Levy, 2007). The presences of antibodies to HIV-1 in the blood as well as CD4 cell levels of less than 14% are indicative of HIV-1 infection (Watstein & Jovanovic, 2003). When HIV-1 replicates in the human host there is an increase in viral load and a reduction in the number of CD4 cells thus paving a way for opportunistic infections such as tuberculosis, pneumonia and cancers such as kaposi's sarcoma (Levy, 2007; Ramaih, 2008). If left untreated; HIV-1 infection ultimately causes the acquired immunodeficiency syndrome (AIDS), a disorder resulting from the suppression of the immune system when the viral load is high and the CD4 cell count is below 200 cells/ µl of total blood volume (Watstein & Jovanovic, 2003; Libman & Makadon., 2009). The spherically shaped HIV-1 virion is surrounded by a viral envelope that consists of lipids and proteins (Hirsch & Curran, 1990; Sierra et al., 2005). The virion has glycoproteins (gp41 and gp120) that it uses to attach to the CD4 receptors on the host cell, when the virus enters the host cell, it uses three enzymes (reverse transcriptase, integrase and protease) as well as host machinery to integrate into the host cell genome and make multiple copies of itself (Levy, 2007; Sierra et al., 2005; Ramaih, 2008). This is illustrated in figure 2.1 that shows the HIV-1 cell cycle.

There is currently no known vaccine or cure, only treatment in the form of antiretroviral therapy that decreases mortality and slows down disease progression (Ramaih, 2008; Pirrone et al., 2011). AIDS-related deaths dropped by 18 % in 2010 since the mid-2000s, this decrease is attributed to the increase in the accessibility of antiretroviral treatment in under developed, and developing countries (UNAIDS, 2012). The most effective way to reduce HIV-related deaths is the use of combinational antiretroviral therapy also called highly active antiretroviral therapy (HAART) which was introduced in 1996 **(**Murphy et al., 2001; Pirrone et al., 2011; Apostolova et al., 2011).

2.2) Highly active antiretroviral therapy

HAART functions to restore the immune system that is suppressed by HIV/AIDS (Jevtovic et al., 2009) and its effects include; 1) an ability to lower viral load which allows an increase in CD4 T cell count to above 200 cells per cubic meter which prevents the development of opportunistic infections, 2) increasing the life expectancy of HIV/AIDS patients and 3) prolonging the development of HIV to AIDS (Pinti et al., 2006; Ramaih, 2008; Jevtovic et al., 2009). Drug toxicity is a long-term adverse effect of HAART usage and has proved to be inevitable in most HIV positive patients (Jevtovic et al., 2008 and 2009). Drug toxicities occur because of patients not wanting to tolerate the use of HAART for life, drug-drug interactions usually in combination with tuberculosis drugs (Subbaraman et al., 2007), host genetics (Hetherington et al., 2001) etc. Antiviral drug toxicities cause some of the following conditions; dementia, anemia, depression, intestinal pseudo obstruction, myopathy, elevated serum bilirubin levels, pancreatitis, hypothyroidism, diabetes mellitus and gastrointestinal complaints (Pinti et al., 2006).

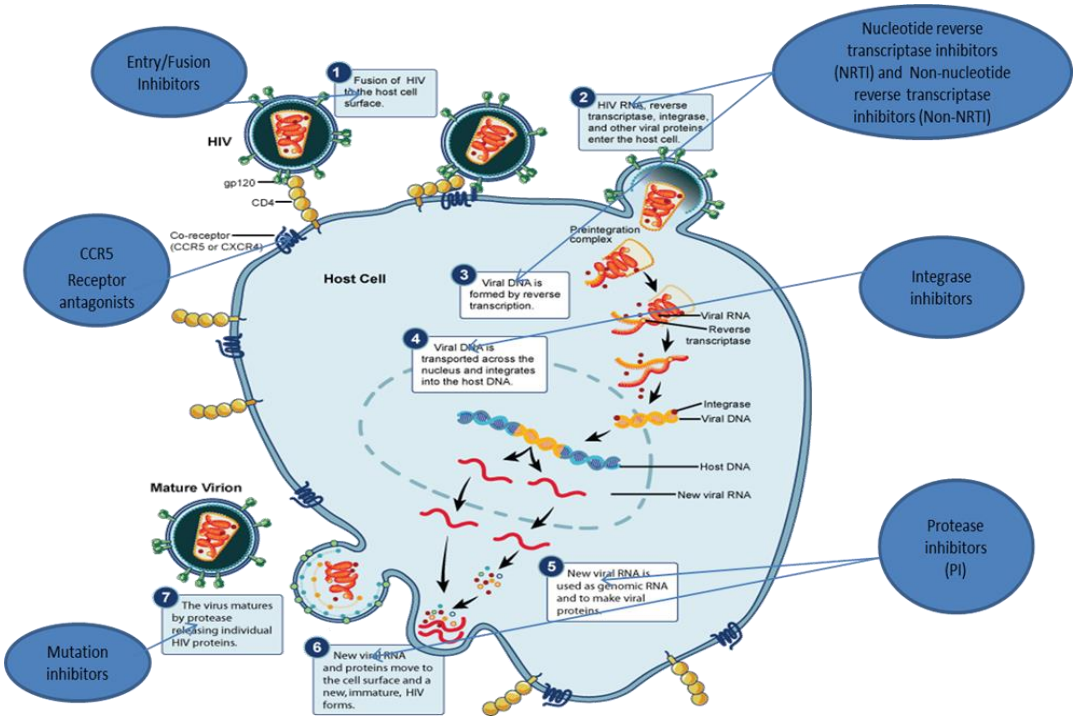


Figure 2.1: Illustration of the HIV replication cycle showing the different stages, from attachment to the host cell using cell host machinery, budding and maturation for making multiple copies of itself (Courtesy: NIAID, 2014). Indicated by arrows are the various HAART regimens that target the different stages of the life-cycle for termination of HIV replication.

2.2.1) Reverse Transcriptase Inhibitors

Although a number of possible drug targets are present in the HIV life-cycle, reverse transcriptase which converts single stranded viral RNA into a complementary double stranded deoxyribonucleic acid (DNA) (Pirrone et al., 2011). Reverse transcriptase inhibitors were the first point of interest in preventing viral replication because it is imperative to retrovirus replication and is absent from the HIV-1 receptive host cells (Prasad & Goff, 1990; Basavapathruni & Anderson, 2007). Telomerase are a specialized type of reverse transciptase that lengthens the telomeres (Gomez et al., 2012). Telomeres are a decisive factor for cell replication because telomeres function as a cell division clock (Vaziri & Benchimol, 1998). This makes telomerase a drug target for therapeutics, zidovudine is an example of an anti-HIV drug that inhibits telomerase (Gomez et al., 2012). Combination therapy (HAART) normally administers a combination of two NRTIs and one protease inhibitor (Drake, 2000).

2.2.1.1) Nucleotide reverse transcriptase Inhibitors

Nucleotide reverse transcriptase Inhibitors (NRTIs) are administered as prodrugs which are activated to metabolically active triphosphates through phosphorylation inside the host cells (Apostolova et al., 2011). These drugs discontinue viral DNA replication because as nucleoside analogues that lack the 3'-OH group, the drugs effects are exerted when reverse transcriptase integrates them into the proviral DNA (Apostolova et al., 2011). NRTIs inhibit polymerase DNA by the following mechanisms; competitive inhibition with natural nucleotides, mitochondrial DNA chain terminators, inhibition of the exonucleolytic proof reading of DNA polymerase gamma (Pol γ) and resistance of exonucleolytic clearance by DNA Pol γ activity (due to 3'OH-group absence) (Lewis et al., 2003). Examples of NRTIs include; zidovudine (AZT), stavudine (d4T), didanosine, abacavir (ABC), lamivudine (3TC) and emtricitabine (Lai et al., 2008). Conditions that can arise as a result of using NRTIs include hyperlactemia, lipodystrophy, lactic acidosis, insulin resistance and diabetes, pancreatitis, hepatosteatosis, myopathy, peripheral neuropathy and bone marrow suppression (Kakuda, 2000).

The synthesis of essential mitochondrial electron transport chain proteins is interfered with when mitochondrial DNA replication is terminated (Apostolova et al., 2011). Myopathy was the first mitochondrial toxicity side effect reported to have resulted from the use of zidovudine and it was one of the first drugs approved by the Food and Drug Administration (FDA) in 1987 for treatment of HIV-1 infection (Basavapathruni & Anderson, 2007; Pirrone et al., 2011).

2.2.1.2) Non-nucleotide reverse transcriptase inhibitors

Non-nucleotide reverse transcriptase inhibitors (NNRTIs) inhibit HIV-1 reverse transcriptase activity by binding at an alternative site that is not an active site, therefore not competing with deoxynucleotide triphosphates (dNTPs) for inhibition (Basavapathruni & Anderson, 2007). Nevirapine was the first approved NNRTI drug and like other NNRTIs, namely, [1-(2-2-hydroxyethoxymethyl)-6-(phenylthio)thymine and tetrahydroimidazo[4,5,1-jkj][1.4]benzodiazepine-2(1H)-one and thione, nevarapine was able to distinguish different HIV strains (Basavapathruni & Anderson, 2007). Examples of NNRTIs include; nevirapine (NVP), efavirenz (EFV), rilpivirine, etravirine (ETR) and delavirdine (Drake, 2000; Haubrich et al., 2009).

2.2.2) Protease inhibitors

Proteases hydrolyze pepetide bonds and perform important functions in vast biological processes. Protease inhibitors (PIs) prevent HIV-1 from reaching maturation to prevent further infection of the host cells by the virus (Karacostas et al., 1989; Flexner, 1998). This is achieved by preventing the cleavage of the gag and gag-pol polyproteins by the HIV protease to stop maturation of the virus, however; PIs do not affect or make changes to host cells that already contain proviral DNA (Flexner, 1998). The HIV protease is an important protein required by the virus to complete virus assembly, reach maturity, and continue to infect new host cells (Brik & Wong, 2003). Not only do PIs have anti-HIV activity, they also have anticancer properties (Bernstein & Dennis, 2008). Examples of PIs include liponavir, nelfinavir, indavir, saquivir, atanavir, darunavir, tipanavir and ritinovir (Flexner, 1998; Llibre, 2009; See et al, 2011). Most HAART toxicities like lipodystrophy have been attributed to the use of PIs.

Because of the high prevalence of HIV-1 infections and the increasing availability of antiretroviral therapy, HIV/AIDS has now become a chronic disease and various other conditions arise because of treatment toxicities. HIV-1 not only impairs the immune system it also disrupts the metabolism causing metabolic complications (Couture, 2005). These metabolic complications are becoming problematic to the host. To contribute to this problem, HAART exacerbates these metabolic complications.

2.3) HIV and/or HAART Metabolic Complications

Metabolic disorders are conditions resulting from the disruption of an organism's internal environment (Villas-Boas et al., 2007). The metabolism being the chemical reactions (anabolic and catabolic) that occur in biological systems can be affected by external stimuli or genetic modification (Voet et al., 2008). Metabolic disorders can be inherited or acquired and sometimes both. Inherited metabolic disorders are called inborn errors of metabolism and include mitochondrial diseases, lactic acidosis, glycogen storage diseases etc (Villa-Boas et al., 2007). As mentioned is section 2.2; HAART can cause mitochondrial diseases and this is an example of an acquired metabolic complication. Most metabolic disorders are untreatable, while expensive therapy can be used in some cases (Villas-Boas et al., 2007). Early diagnosis and initiation of therapy are highly recommended (Villas-Boas et al., 2007).

HIV-1 affects the metabolism of the host (Couture 2005; Hewer et al., 2005; Rose et al, 2008; Jevtovic et al., 2009). In addition, HAART administered to HIV positive patients increases the occurrence of metabolic complications (Brennan-Benson, 2009). The toxicity of the HAART regimens give rise to metabolic abnormalities or disruptions termed metabolic syndrome (Jevtovic et al., 2009; Brennan-Benson, 2009). Carr et al (2003) used "metabolic syndrome" to refer to the cluster of metabolic abnormalities characterized by dyslipidemia and insulin resistance resulting from lipodystrophy. The metabolic syndrome results in hypertension, hypertriglyceridemia, adiposity (mainly in HIV positive patients), low high-density lipoprotein (L-HDL) cholesterol etc (Ford et al., 2002). The extent to which the metabolic syndrome occurs differs according to gender, age, genetic markup and the type of HAART drugs received by an HIV-1 positive patient (Jevtovic et al., 2009) as well as how long an individual has been receiving HAART (Jevtovic et al., 2009).

Many deaths in HIV-1 positive patients are due to metabolic and cardiovascular complications (Rose et al, 2008; Brennan-Benson, 2009). The cardiovascular complications caused by HIV-1 and HAART are dilated cardiomyopathy, pericardial effusion and pulmonary artery hypertension (Brennan-Benson, 2009). Osteopaenia and osteoporosis also have higher possibilities of developing because of HIV-1 and HAART (Brennan-Benson, 2009). Some metabolic complications are associated with HIV-1 infection and/or HAART (Table 2.1) are elaborated on and ways in which they can be determined follow.

2.3.1) Insulin Resistance

Insulin resistance (IR) results from a decreased sensitivity of insulin to peripheral tissues as higher levels of insulin accumulate in the blood but normal physiological insulin requirements are not met (Lebovitz, 2001; Tungsiripat-Gerber & Aberg, 2004; Brennan-Benson, 2009). IR leads to type 2 diabetes mellitus and was one of the first metabolic complications identified as being caused by HIV is associated with the use of HAART (mainly PIs) (Carr et al., 1998; Couture, 2005; Brennan-Bensen 2009; Jevtovic et al., 2009). IR and diabetes are examples of acquired/induced metabolic disorders that are mostly prominent in adults as a result of carbohydrate metabolism disruption (Vilas-Boas et al., 2007).

About 8-10% of type 2 diabetes mellitus patients are at risk of developing cardiovascular complications (Jevtovic et al., 2008). Impairment of glucose mostly occurs in male patients, AIDS patients prior to HAART and in older patients (Jevtovic et al., 2009). HIV positive patients who are at a higher risk of developing diabetes and IR are those patients co-infected with hepatitis C (Badiou et al., 2008). IR is directly caused by NRTI or indirectly caused by PI (Badiou et al., 2008) and can be reduced by changing lifestyle and the HAART regimens (Brennan-Benson, 2009).

2.3.2) Lipodystrophy

Lipodystrophy refers to the redistribution of adipose tissue that occurs in about 2-84 % HIV positive patients (Safrin et al., 1999). Lipodystrophy can be lipoatrophy, lipohypertrophy or both conditions (Jevtovic et al., 2008). Lipohypertrophy is characterized by the accumulation of visceral fat or dorsocervical fat (buffalo hump), the occurrence of obesity and increased breast size, while lipoatrophy refers to subcutaneous and buttock fat wasting and fat loss in the face, arms and legs (Tungsiripat-Gerber & Aberg, 2004; Couture, 2005 Jevtovic et al, 2008; Brennan-Benson, 2009). Buffalo hump and the accumulation of abdominal fat are statistically proven as morphological changes that are apparent in one-third of females who are HIV positive and taking HAART (Tungsiripat-Gerber & Aberg, 2004). Insulin resistance and hyperlipidemia are also characteristic of lipodystrophy (Viraben et al., 1998; Carr et al., 1998).

Lipodystrophy risk factors include; patients who start using HAART when they already have AIDS, patients on ongoing NRTIs and HIV positive females over the age of 40 and NNRTIs decreases the chances of developing lipodystrophy (Jevtovic et al., 2009). The chances of

developing lipodystrophy are increased by 100 % in ten years of using HAART and also increased by 15 % when co-infected with hepatitis C virus (Jevtovic et al., 2008). Patients who begin taking HAART when they have AIDS have a 95 % probability (p>0, 01) of developing lipodystrophy. About fourty-percent of patients on HAART (mostly protease inhibitors) suffer from impaired adipocyte differentiation (Couture, 2005).

In 2004, there was no known therapeutic treatment for lipodystrophy (Tungsiripat-Gerber & Aberg, 2004). Presently LDS can be treated by statins and fibbrates (lipid reducing drugs), exercise, change in diet, cosmetic surgery, hormone therapy or discontinuation of specific HAART drugs (Robinson, 2004; Tungsiripat-Gerber & Aberg, 2004).

2.3.3) Hyperlipidemia and Dyslipidemia

The functions of lipids include; regulating cell proliferation, apoptosis, tissue repair, blood clotting and inflammation (Psychogios et al., 2011). HIV and HAART disrupt the lipoprotein metabolic pathway by increasing low-density lipoprotein (LDL) levels and slightly lowering the high-density lipoprotein (HDL) levels, thus increasing the development of cardiovascular complications and metabolic complications (Couture 2005; Rose et al., 2008; Brennan-Benson, 2009).

Coronary artery disease (CAD) and dyslipidemia are caused by increased levels of triglycerides, total and LDL cholesterol and diabetes (Brennan-Benson, 2009; Rose et al., 2008). Coronary artery disease increases in patients who already have cardiovascular risk factors (Jevtovic et al., 2009). Dyslipidemia occurs mainly as a result of PIs (Jevtovic et al., 2008; Rose et al., 2008). Seventy percent of hyperlipidemia patients are at risk of developing coronary artery disease (Jevtovic et al., 2008). Hyperlipidemia is more prevalent in HIV positive males, especially those using PI and NNRTIs and can be reduced by taking a combination of HAART regimens consisting of two NRTIs and NNRTIs (Jevtovic et al., 2008).

2.3.4) Atherosclerosis

The metabolic syndrome induced by HIV and/or HAART causes the distortion of lipids that end up serving as direct markers of atherosclerosis (Brennan-Benson, 2009). An increase in non-HDL cholesterol is a direct result of a disrupted lipoprotein metabolic pathway that increases the LDL or delays clearance (Badiou et al., 2008). HIV also acts as a pro-

inflammatory agent which increases levels of plasma pro-inflammatory cytokines causing atherosclerosis (Couture, 2005; Brennan-Benson, 2009).

2.3.5) Lactic Acidosis

Lactic acidosis is a life threatening disease, which develops following nucleoside dysfunction (Tungsiripat-Gerber & Aberg, 2004). Lactic acidosis is associated with hyperlactemia (elevated lactic acid) and occurs mainly in HIV positive women, especially pregnant women (Tungsiripat-Gerber & Aberg, 2004). It is commonly caused by stavudine, didanosine and zalcibine (Tungsiripat-Gerber & Aberg, 2004). Hyperlactemia and lactic acidosis are condition characterized by decreased blood pH and elevated plasma lactate caused by NRTIs and death results from high levels of lactic acid (Brennan-Benson, 2009). For treatment of lactic acidosis the use of NRTI must be stopped (Tungsiripat-Gerber & Aberg, 2004).

Table 2.1: HIV metabolic complications resulting from HAART toxicities.

Metabolic complication	HAART regimen/s or class	References
Lipodystrophy	Stavudine and zidovudine Nelfinavir Atanavir Efavirenz and PIs NRTIs	Bogner et al., 2001; van Griensven et al., 2007 Dube et al., 2006; Haubrich et al., 2009; White, 2001 McComsey et al., 2011 Brown, 2008 Gkrania-Klotsas & Klotsas,
Insulin resistance and type 2 diabetes	Zidovidine/lamivudine Stavudine Indinavir, liponavir/ritonavir, didanosine PIs	Blumer et al., 2008 Fleischman et al., 2007 Hester et al., 2012 Schambelan et al., 2002; Grunfeld & Tien 2003
Dyslipidemia	Stavudine, zidovudine, abacavir, efavirenz and PIs Ritonavir Nevirapine Tenofovir PIs	Dube et al., 2003 Estrada & Portilla, 2011 van Leth et al., 2004 Tungsiripat et al., 2010 Mary-Krause et al., 2003
Lactic acidosis	Stavudine, didanosine, zalcitabine Zidovudine, lamivudine Stavudine NRTIs	Arenas-Pinto et al., 2003 John et al., 2001; Tungsiripat & Aberg et al., 2004 McComsy et al., 2004 Tungsiripat-Gerger & Aberg, 2004
Mitochondrial damage	NRTIs	Kontorlnis & Dieterich, 2003; Carr et al., 2002; Mussini et al., 2005; Apostolova et al., 2011; Pinti et al., 2011
Neurocognitive impairment	NNRIs Stavudine, lamivudine, nevirapine	Marra et al., 2009 Valcour et al., 2009
Pancreatitis	Stavudine, didanosine	Kline et al., 1998; Moore et al., 2001; McComsey et al., 2004
Myopathy and cardiomyopathy	zidovudine	Lewis et al., 2000; Pinti et al., 2006
Hematological toxicity	zidovudine	Wilde et al., 1993
Neuropathy	Didanosine, stavudine, zalcitabine	Moore et al., 2000
Hepatotoxicity	Stavudine and didanosine	Gisolf et al., 2000
Vitamin D deficiency	Efavirenz and PIs	Holick, 2007

2.4) Measurement of metabolic complications

Having defined some of the metabolic complications associated with HIV and the use of HAART. The only way of determining these metabolic complications is by detecting and measuring levels of metabolites. Both conventional and metabonomics methodologies have been applied for the measurement of these metabolic complications. The two methods are explained below.

2.4.1) Conventional methods

Colorimetric assays are normally employed for the identification and quantification of metabolites, however these tests are point analysis tests because they can only detect or monitor one metabolite at a time (Matsumoto & Kuhara, 1996). Many of the characterized metabolic complications in the host can be measured using different biochemical assays like competitive enzyme-linked immunosorbent assays (ELISAs) which are commercially available (Holvoet, 2008). Holvoet et al (1998) correlated the increased levels of oxidized LDL (detected in plasma using 4E6 ELISA) as possible markers for the degree of coronary artery disease. Competitive ELISA depends on highly specific antibodies like DLH3, which requires the isolation of LDL from plasma which prevents mass analysis of plasma samples (Holvoet, 2008). Other methods include; enzymatic colorimetric assays like the glucose tolerance tests and fasting total cholesterol test (Samaras et al., 2007; Hattingh et al., 2009; Ayala et al., 2010), radioimmunoassays for glucose and insulin (Hewer et al., 2006) and more expensive methods like dual-energy X-ray absorptiometry and computed tomography scanning (Hewer et al., 2006; Williams et al., 2012).

These limitations have resulted in the necessity for new techniques to detect and quantify metabolic complications. Metabolic complications are not always a result of a single altered metabolite but rather a combination of sometimes simultaneously altered pairs or groups of metabolites to evoke a reaction and disrupt the metabolism (Villas-Boas et al., 2007). Methods that are able to simultaneously measure a number of metabolites at the same time are required. Metabonomics, a relatively new “omics” complies with such requirements and is already making progress in measuring changes in multiple metabolites in one experiment.

2.4.2) Metabonomics

Metabonomics is defined as 'the quantitative measurement of dynamic multiparametric pathophysiological responses of biological systems to stimuli or genetic modification' (Nicholson et al., 1999). Metabonomics and metabolomics are words used interchangeably (Gika et al., 2007; Axelson, 2010; Beger & Colatsky, 2012) because they both conduct studies of the metabolic changes of the living organisms (Axelson, 2010). Metabolomics is defined by Fiehn (2001) as "the comprehensive and quantitative analysis of all metabolites in biological systems". Alexson (2010) stated that metabolomics is mostly applied to microbial and plant studies. Metabonomics and metabolomics form an integral part of systems biology (Lin et al., 2011; Dunn et al., 2011) because they focus on performing untargeted studies of the metabolome. Most metabonomics studies follow a holistic approach that is normally employed in systems biology. In this approach, individual components of a biological system are studied as a whole and not in isolation like the reductionist principle (Kell, 2007). Metabolic network analysis that is normally conducted in metabonomics studies represents a holistic research approach because the networks are connecting systems of individual components (Kell, 2007).

Metabonomics's which is known to refelect the living systems biological status in metabolic profiles of biofluids (Wu et al., 2009) has vast applications that include; drug toxicity studies (such as HAART), disease diagnosis, analysis of genetic disorders, biomarker identification, metabolic state monitoring, and organ transplantation (Nicholson et al., 2002; Jiye et al., 2005; Villas-Boas et al., 2007; Lin et al., 2011; Dunn et al., 2011). It is increasingly being applied in research such as the discovery and development of medicines because it offers discrimination between experimental groups and allows for high throughput analysis (Wilson et al, 2005; Zelena et al., 2009). An unbiased and global classification of samples to attain discriminating metabolites as a result of stimuli is reffered to as metabolic fingerprinting. An illustration of a typical workflow that is followed when conducting metabonomics studies is shown in figure 2.2.

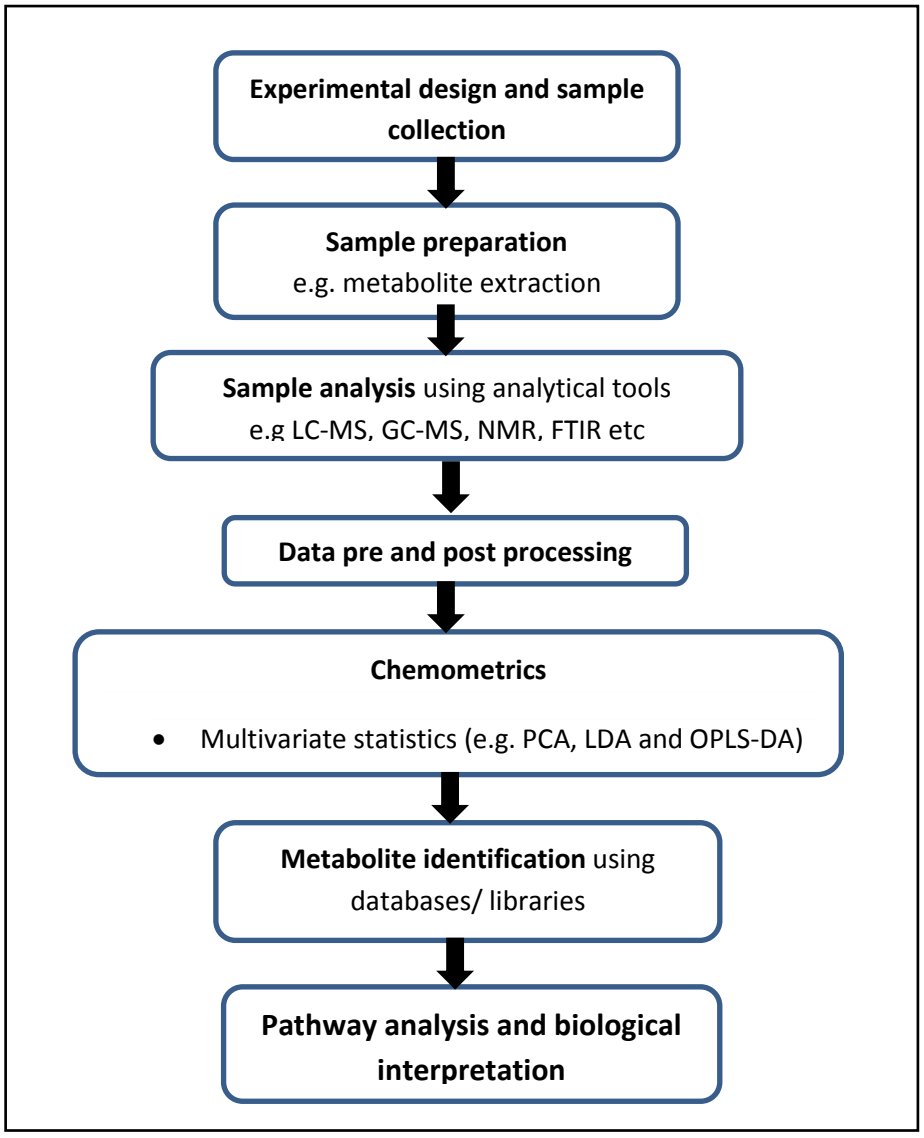


Figure 2.2: An illustration showing a typical metabonomics workflow

Disadvantages of metabonomics are apparent when working with unknown analytes because diagnostic differences go undetected due to factors like instability of analytical systems, chemical noise, an individual's metabolic make-up etc, due to no prior knowledge being known about the significant analytes before analysis (Gika et al., 2007). These factors hinder knowledge about the sample stability (Gika et al., 2008), however; this is the host's first line of response to stimuli or genetic modification.

Given that metabonomics gives a snapshot of the metabolome phenotype, it can be used to produce metabolic profiles of mainly biofluids (serum/plasma, urine, saliva etc) and tissue (van den Berg et al., 2006; Gika et al., 2007) because it is the closest omics to an organism's phenotype (Villas-Boas et al., 2007). Metabolic phenotypes are representations of end-points

of protein and gene expression shown by metabolites levels and physiological changes (Beger & Colatsky, 2012). Metabolomics is of extreme value because it lies at the top of the pyramid of life, which has proteomics and genomics in the lower hierarchy and focuses on the identification and quantification of the metabolome, which is closest to the phenotype (Dettmer et al., 2007; Hollywood et al., 2006; Villas-Boas et al., 2007; Psychogios et al., 2011).

2.4.2.1) The Metabolome

The metabolome, being the total low molecular weight metabolites of biological systems that are less than 1500Da (Wu et al., 2009; Psychogioes et al., 2011) gives an indication of an organism's phenotype and physiology and this information proves useful in health-related applications (Hollywood et al., 2006). These low molecular weight compounds include lipids, amino acids, nucleotides, signaling molecules and hormones and they indicate what has occurred and is still occurring in the organism (Dettmer et al., 2007; Tyagi et al., 2010). The metabolome is affected by gender, age, race, diet, genetic factors, lifestyle and environment and can be used as an alert of the body's condition (Jiye et al., 2005; Villas-Boas et al., 2007; Zelena et al., 2009). The mammalian metabolome is not defined because it has complex properties (Wu et al., 2009), and is therefore determined by the how it is measured and the location of measurement (Wishart et al., 2007). Mammalians have a larger exogenous metabolome and a smaller endogenous metabolome because unlike microbes or plants, mammals need to synthesize other metabolites that are required to keep them alive from sources in the external environment (Villas-Boas et al., 2007). The metabolome depends on metabolic network properties and some of the low molecular weight compounds are metabolites (Kell, 2007).

Metabolites are intermediates and products of cellular regulatory processes of the metabolism and their levels determine the biological systems phenotypic response to the environmental or genetic changes (Bruce et al., 2009). These micromolecules that are building blocks of biological components (e.g proteins) are required for growth, maintenance and homeostatic functions of the cell (Dettmer et al., 2007; Dunn et al., 2011). Metabolites can be endogenous (synthesized by the host genome) or exogenous also called xenobiotics (drug metabolites) because of being synthesized from food, foreign chemicals or host specific microbes (Villas-Boas et al., 2007; Beger & Colatsky, 2012). Human beings mostly contain primary metabolites compared secondary metabolites. Factors that affect the concentration of

metabolites include diet, age, health, gender, time of day and genetic background (Wishart et al., 2007).

With metabolites being end products of many large molecule interactions (enzymatic reactions) in an organism's body, human beings contain more than 2100 endogenous metabolites and discoveries are increasing in numbers (Kell, 2007; Villas-Boas et al., 2007; Bruce et al., 2009). Metabolites are still being discovered and not all presently detected human metabolites are identified or quantified resulting in some being absent in databases/libraries. This series of chemical reactions that occur in the cell are called metabolic pathways and each principal chemical is modified by a series of reactions as shown in figure 2.3 that is a representative illustration of human metabolic pathways. With the metabolism representing all the chemical reactions that occur in a living organism (Voet et al., 2008), its products and intermediates (termed metabolites) are present in all types of samples derived from the host organism.

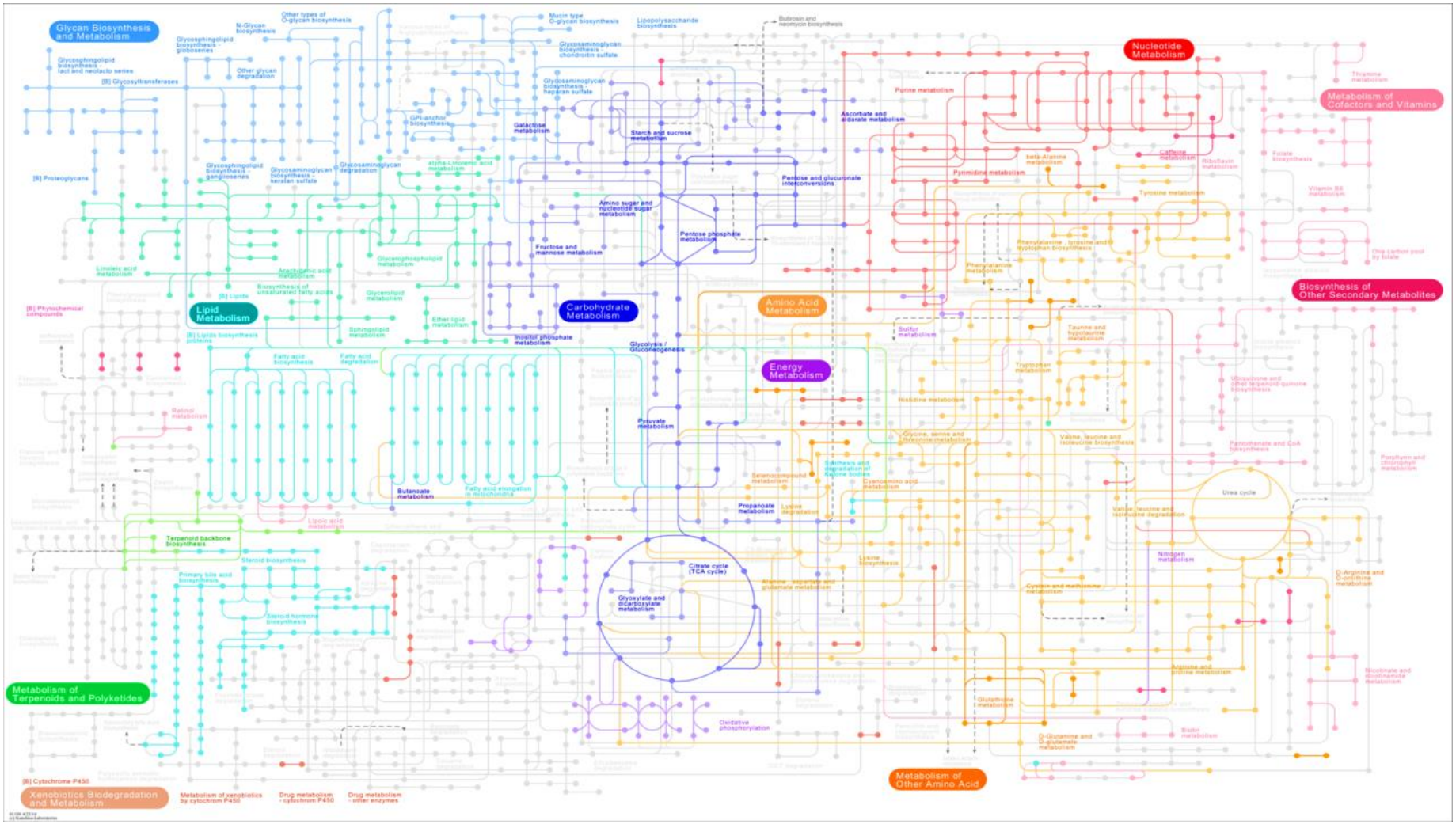


Figure 2.3: The metabolic pathways of *homo sapiens* showing an overview of the metabolism. Complexity of the interconnected and interrelated molecules that are involved in biochemical and physiological processes depicted in this diagram. Figure downloaded from: http://www.kegg.jp/kegg-bin/show_pathway?map01100. Link Active 08 May 2014.

2.4.2.2) Metabonomics samples and sample preparation

Metabonomics studies require optimizations for sample preparation, chromatographic conditions (Nordstrom et al., 2010) and ionization (Abad-Garcia et al., 2008) to achieve a greater coverage of detectable compounds. This section will only highlight the types of samples used for MS metabonomics studies and preparation procedures such as metabolite extractions that are required prior to analysis using metabonomics techniques.

2.4.2.2a) Samples type

Depending on the study aims, the criteria for deciding which sample to use for a specific study includes; the questions being asked, metabolites of interest and instrument sensitivity (Villas-Boas et al., 2007). Sample stability, suitable storage conditions and time are of great importance in metabonomics studies to ensure validity of results (Gika et al., 2007). In order to produce valid data, samples need to be analyzed a number of times either for collection of more information or to identify key biomarker molecules (Gika et al., 2007). Lower temperatures generally minimize biological sample degradation and desirable extractions are performed at 4° C for potential metabolites whilst -80° C is generally utilized for storage purposes (Villas-Boas et al., 2007; Michopoulos et al., 2009). Because each cell, tissue, organ or biofluid produces its own uniquely different metabolites (Villas-Boas et al., 2007), sample collection can be easy or a challenge depending on the metabolites of interest.

Biofluid collection is less invasive (than for example tissue collection) and analytical technologies analyze biofluids easier (Villas-Boas et al., 2007) even though tissues can work as well and are in some cases the only samples to answer a specific question. Different types of biofluids; namely blood, saliva, cerebrospinal fluid and urine produce varying numbers of metabolites (Villas-Boas et al., 2007; Psychogios et al., 2011). All the processes happening in organs are normally reflected in blood making it easy to detect metabolite disturbances in blood even though it is a challenge to track these metabolite alterations back to their specific organ of origin (Villas-Boas et al., 2007).

Serum is the yellow liquid component derived from the clotting of blood in the absence of an anti-coagulant and primarily consists of water, blood proteins, inorganic electrolytes and low molecular weight compounds (Villas-Boas et al., 2007; Luque-Garcia & Neubert, 2007; Psychogios et al., 2011). Serum transports dissolved gases, nutrients, metabolic wastes, restricts fluid loss at injury site, regulates the pH and ion composition of intestinal fluids,

stabilizes body temperature, and defends against pathogens and toxins (Psychogios et al., 2011).

Metabolic profiling became a routine component of medical practice because metabolite alterations could be associated with medical conditions (Bahadori & Mohammadi, 2005). Not all areas of the metabolism are richly represented during profiling but metabolites useful in disease diagnosis can usually be observed (Kell, 2007). Analytical chemistry has enabled the quantification of biofluid compounds because biofluids are easily processed and analyzed with metabonomics technologies.

2.4.2.2b) Sample preparation

Sample preparation and pretreatment are of extreme importance in metabonomic studies (Thiel, 1999; Luque-Garcia & Neubert, 2007). It is required before analytical instrument analysis to allow removal of unwanted compounds to improve the identification and quantification of compounds of interest (Juhascik & Jenkins, 2009; Zwir-Ferenc & Bizuik et al., 2006). Sample preparation is of extreme importance for metabolite extraction and must be reproducible (Sheikh et al., 2011; Lee et al., 2011) and it must be reproducible, robust, non-destructive and free of artifacts (Axelson et al, 2010).

Sample preparation can be time consuming and is the point where most mistakes occur (Luque-Garcia & Neubert, 2007). For the production of analytical data that is reliable, reducing the sample manipulation to as little as possible is necessary in order to achieve reasonable sample recovery for complex data collection (Tavazzi et al., 2005). Most of the metabolites are bound to proteins and lipoproteins that are in circulation and proteins cause a challenge in metabolomic analysis (Villas-Boas et al., 2007). Metabonomics studies require removal of proteins from blood and its derivatives to simplify chromatographic separation, MS analysis and NMR data collection (Villas-Boas et al., 2007). It is important to standardize sample preparation procedures for serum/plasma profiling in order to obtain reliable biomarkers (Luque-Garcia & Neubert, 2007). Serum samples normally contain little contaminants but metabolites can be bound to proteins that cause signal suppression in LC-MS experiments (Wu et al., 2006; Juhascik & Jenkins, 2009). Metabolite extraction techniques for serum must have an increased degree of protein removal and must increase the number of detectable low molecular weight compounds (Sheikh et al., 2011). Want et al (2006) has shown that the use of organic solvents for metabolite extraction is efficient for

protein removal and effective for detecting a broader range of metabolite features in serum samples.

The following two compound extraction techniques are used in metabonomics studies; Liquid-Liquid extraction (LLE) and solid phase extraction (SPE). LLE uses two liquid components to extract the compound of interest from two layers (from an aqueous layer into an organic layer) that form following a mixture of solvents (Juhascik & Jenkins, 2009). Methanol, ethanol and acetonitrile were shown to be three of the most effective organic solvents for metabolite extraction because the chromatograms obtained using these solvents produced a large number of detectable metabolite features compared to other methods (Sheikh et al., 2011), see figure 2.4. Want et al (2006) showed that the use of solvents for extraction is effective in detecting a broader coverage of metabolite features in serum samples and is efficient in removing proteins.

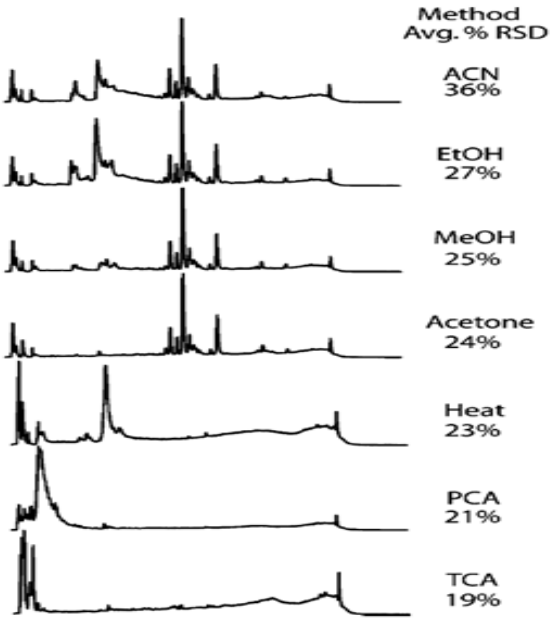


Figure 2.4: LC-MS chromatograms of serum samples showing metabolites extracted using different methods [acetonitrile (AcN), ethanol (EtOH), methanol (MeOH), acetone, heat, perchloric acid and trichloroacetic acid (TCA)] (Want et al., 2006).

LLE uses too much volume of organic solvents and this method ends up requiring too much material that is a challenge to dispose of, not to mention that an expert is required for separating the immiscible layers that form (Juhascik & Jenkins, 2009). Separating these two layers is also time consuming (Zwir-Ferenc & Bizuik, 2006). On the other hand, SPE is quick

and extracts an increasing number of metabolites compared to LLE (Zwir-Ferenc & Bizuik, 2006). SPE uses a liquid and solid phase to extract the analyte of interest and also removes the compounds in a complex mixture that cause interference in detecting the analyte of interest (Zwir-Ferenc & Bizuik, 2006). It normally uses stationary cartridges that are used in liquid chromatography (LC) columns and it is a very selective method (Zwir-Ferenc & Bizuik, 2006). SPE methods were largely used for drug metabolism and a need for an increase in throughput and efficiency lead to the development of 96-well format SPE plates (Venn et al., 2005). The 96-well format SPE plates use low volumes and the extraction step is quick. These SPE plates complement the qualities (sensitivity and selectivity) of the high throughput techniques by increasing the number of samples that can be analysed simultaneously at a time (Venn et al., 2005). More details on SPE are provided in section 4.1, which is a manuscript by Williams, Kgoadi et al (2014) that described an optimized SPE method that demonstrated an easy to use minimal sample preparation method for the removal of proteins and phospholipids using Ostro^{TM} plates. This method produced matrix free chromatograms and better distinguished metabolites among the experimental groups.

2.4.2.5) Metabonomics Techniques

Analytical tools used for conducting metabonomics studies are required to be unbiased, robust, cover a wide range, be accurate and of high sensitivity, reproducibility and specificity (Tavazzi et al., 2005; Lu et al., 2007; Jevtovic et al., 2009). These tools are usually high-throughput techniques that result in the generation of massive amounts of data. The techniques used can either be applied to targeted or untargeted analysis. The reductionist approach fails to account for existence and functions of other molecules while holistic approaches are unbiased in diagnosing and classifying a disease (Nicholson et al., 1999, 2002; Kell, 2007). Targeted analysis detects and quantifies specific metabolites of interest in a sample and increases selectivity and sensitivity of MS methods whereas untargeted analysis broadens the range of detectable metabolites and is accompanied by an increased amount of data analysis (Lee et al., 2011). Unknown compounds are analyzed using untargeted analysis. Detection of a broader range of compounds is an advantage as indicated by Chen et al (2007) who showed the discovery of potential biomarkers resulting from disease and drug induced metabolite changes. With targeted analysis it is easier to select an instrument that works best in detecting the metabolites of interest for a specific study by using literature as a guide. Even though untargeted analysis detects diverse compounds, the types or number of detectable

metabolites in complex mixtures is also instrument dependent. No single analytical platform can cover the entire metabolome (Dettmer et al., 2007) and this is why when conducting metabonomics studies, the use of other validation methodologies is required. Commonly used metabonomics are mass spectrometry chromatographic techniques and vibrational spectroscopy. The dominantly applied techniques are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS).

2.4.2.6) Spectroscopic techniques

Given the fact that biological materials have the ability to absorb and transmit light over varying frequency ranges, it is no surprise that spectroscopic techniques are used to conduct metabonomics studies. The two major classes of spectroscopic techniques employed include nuclear magnetic resonance (NMR) and vibrational spectroscopy. Vibrational spectroscopy refers to Infrared and Raman spectroscopy, however; Fourier transforms infrared (FTIR) spectroscopy being the commonly used technique (Faolain et al., 2005; Theodoridis et al., 2008). Both techniques use simple sample preparation methods and samples are obtained noninvasively from biological material (Harz et al., 2009).

2.4.2.6a) Nuclear magnetic resonance

The commonly used form of NMR is proton NMR (^{1}H NMR) (Bruce et al., 2009; Gika et al., 2007). NMR has been widely applied in metabonomics studies for distinguishing diseased patients from healthy controls. NMR is a robust, reproducible, and superior technique with high sample throughput that analyzes most abundant molecules in biological materials (Bruce et al., 2009). These qualities and others such as; High information content from its spectra (Lu et al., 2007), its quantitative nature for several molecules (Serkova & Niemann, 2006; Lu et al., 2007), ability to provide chemical nature of atoms or molecules and compound structural analysis, which aids in identification, makes it the most prominent method of choice in spectroscopy. It requires simple sample preparation (Serkova & Niemann., 2006). Disadvantages of NMR include lower sensitivity than MS and narrow dynamic range of nonselective detectable compounds (Serkova & Niemann, 2006; Want et al., 2006) and the need for ion suppression (Wu et al., 2008). It is however a recommendable technique (just like LC-MS) for analysis of blood samples (Daykin et al., 2002). The combinational use of NMR and MS is recommended in metabonomics studies because of their complementary nature (Serkova & Niemann, 2006).

2.4.2.6b) Fourier Transform infrared spectroscopy

In infrared spectroscopy, vibrational and rotational status of materials become different when light is absorbed in the infrared region (Ahmand & Tullin, n.d). FTIR is a qualitative and quantitative technique that finds many applications in various fields of science that require structural analysis (Baravkar & Kale, 2011). This technique characterizes substances proteins and other substances by determining information about the functional groups and bonds of the samples (Jackson & Mantschi, 1999; Dritsa, 2012). Vibrational bonds that are strong in the infrared region spectrum are dipole moments (Harz et al., 2009), with lipids having the strongest molecular vibrations followed by hydroxyl groups, amine stretching vibrations and carbonyl groups (Harz et al., 2009). Output spectra must be normalized and standardized (Petibois et al., 1999) prior to statistical analysis for data mining. Even though FTIR is a cost effective, high-throughput, automated, fast technique normally recommended for first step use in screening analysis (Wharfe et al., 2010), simple, repeoducible, no sample preparation (Dritsa, 2012), it is a less sensitive technique than NMR regardless of its non-destructive nature (Nicolet, 2001; Dritsa, 2012). FTIR is easier to perform than NMR or MS and has been used for diagnosis of metabolic disorders using serumprofiles (Shaw & Mantsch, 2000) and is implicated as the potential diagnostic tool for cancer (Pijanka et al., 2010). FTIR has been successful in distinguishing diseased and healthy samples and a recent study by Sheng et al (2013) showed serum samples of leukemia patients were distinguishable from healthy subjects.

2.4.2.7) Mass spectrometry

MS separates and analyses samples based on mass-to-charge (m/z) ratios to determine the masses of the molecules and produces data in the form of mass spectra of ion intensity as a function of m/z ratio (Thiel, 1999; Ho et al., 2003; Boyer, 2006; Axelson, 2010). The mass spectrometer consists of three components, namely; (1) an ion source, which ionizes the sample into a gaseous ion, which is electrochemically directed into (2) an analyzer, which analyzes and separates gaseous samples based on the mass-to-charge ratio and finally, (3) a detector that detects and records ions (Thiel 1999; Boyer, 2006; Axelson, 2010).

Samples can be introduced by electrospray ionization (ESI), thermospray, ionspray, desorption techniques or particle beam interface (Boyer, 2006). The sample introduction in ESI occurs as follows; the mobile phase solution is sprayed from a metal needle tip held at several kilovolts, and the resulting charged droplets are evaporated producing ions that are

directed into the vacuum of the analyzer through an orifice for separation (Boyer, 2006; Thiel, 1999). The sensitivity of the mass spectrometer can be improved by cleaning the ion source (Michopoulos et al., 2009). This soft ionization method offers high sensitively at low flow rates and is able to determine both high and low molecular weight compounds (Kebarle & Tang, 1993; Want et al., 2005). ESI provides enhanced quantification and reproducibity (Want et al., 2005). ESI MS is applied in clinical laboratories because it is a sensitive, reliable and robust technique that studies molecules that are thermally liable and non-volatile (Ho et al., 2003). Clinical applications include screening of inborn errors of metabolism and screening of purine and pyrimidine metabolism (Ho et al., 2003). Amino acids, amines, sugars and nucleotide bases are normally detected in the ESI positive mode while organic acids are detected in the ESI negative mode (Villas-Boas et al., 2007). Since this is an untargeted metabonomic study aimed at detecting all possible metabolites from the complex serum sample mixtures, both ionization modes of ESI (negative and positive mode) were used because they offer a greater coverage of metabolites that have different ionization properties.

Available mass analyzers include time-of-flight (TOF), quadrupoles, ion trap and fourier transform analyzers (Thiel, 1999; Boyer, 2006). TOF offers quick separation of biological mixtures and measures a dynamic range of high masses of biological molecules (Guilhaus, 1995; Want et al., 2005; Lacorte & Fernandez-Alba, 2006) and is one of the high resolution mass analyzers that can simultaneously determine exact masses and allows for calculation of definitive molecular formulas (Brown et al., 2005; Want et al., 2005). A TOF analyzer is able to establish a middle way between improved sensitivity and speed (Guilhaus, 1995).

Multiple analyzers can be used in conjunction as this method is reffred to as tandem mass spectrometry (MS/MS) and the analyzers can be hybrid (combination of different analyzers) or not. Selection of analyzers is based on the distinct features that each analyzer offers such as resolution, quantification or identification (Lacorte & Fernandez-Alba, 2006). Quadrupoles are mostly used in hybrid systems because of being easy to use and caliberate (Lacorte & Barcelo, 1996). If quantification is of interest then the triple quadrupole (QQQ) MS is the instrument of choice because it offers less resolution but greater quantification. QQQ also offers high selectivity that makes the identification process easier but structure determination is limited (Lacorte & Fernandez-Alba, 2006). If high resolution is of interest then a quadrupole-TOF (QTOF) MS can be used for qualitative studies like this study. Other hybrid

systems are beyond the scope of this study. The major drawback of these systems is that a lot of peaks are produced because of sensitivity loss (Steen et al., 2001).

During mass spectrometry metabonomics, complex samples are analyzed with the separation power of chromatography, namely gas chromatography (GC) and liquid chromatography (LC) (Tyagi et al., 2010). Most frequently used chromatographic sytems that are coupled to mass spectrometry are gas chromatography-mass spectrometry (GC-MS) and liquid chromatography mass spectrometry (LC-MS) (Luque-Garcia & Neubert, 2007; Gika et al., 2007; Bruce et al., 2009).

2.4.2.7a) GC-MS and LC-MS systems

GC-MS has a high resolution and reproducibility but only analyzes volatile compounds; non-volatile compounds must be chemically derivatized before analysis (Villas-Boas et al., 2007; Wu et al., 2009). It is mostly used as a line analyses (targeted analysis) test because it can characterize a specific group of compounds (Wishart, 2006) and is optimal for organic acid profiling even though it can also analyze nucleic acids, amino acids, sugars etc. (Kuhara, 2005; Villas-Boas et al., 2007). On the other hand, LC-MS analyzes non-volatile and thermally unstable metabolites and does not require chemical derivatization of the sample prior to analytical analysis (Dunn et al., 2005; Villas-Boas et al., 2007; Wu et al., 2009) and as a result, the use of LC-MS in metabolomics studies is expanding.

GC-MS and LC-MS techniques have identified 300 metabolites in the human plasma metabolome but the information is not grouped into one specific scientific resource material (Bruce et al., 2009). Like GC-MS, LC-MS identifies molecules based on comparison to elution time and molecular weight of known reference compounds in libraries (Wishart, 2006); however, unlike GC-MS, LC-MS has limited standard spectral libraries and databases for the identification of metabolites (Psychogios et al., 2011). LC-MS total ion current resolves 20-30 peaks from various parent ions with masses of 50-80 atomic mass units, in addition fragmentation of parent ions increases the possibility of identifying different metabolites (Villas-Boas et al., 2007). The most recent advances in metabolomic analytical tools, like LC/MS/MS technology have made it possible to simultaneously measure cellular metabolic intermediates thus allowing the determination of steady state metabolic outcomes (Munger et al., 2006, 2008).

The column chromatography of the LC-MS system for separation of complex mixtures uses specific column types. C18 reversed-phase columns are normally utilized for LC-MS metabolite separation with volatile carrier solvents (water, acetonitrile or methanol) because of their high resolution for hydrophobic metabolites (Wishart, 2007). Mixed mode metabonomics columns are recommended for separation of sugars, hydrophilic amino acids and nucleosides because C18 columns do not offer good separation for hydrophilic metabolites (Dunn et al., 2005; Wilson et al., 2005; Wishart, 2006). The use of a combination of different columns, reduced column diameter and particle size improves the quality, resolution and sensitivity of the compounds being separated (Wilson et al., 2005; Wishart, 2007). The above mentioned columns reduce the diffusive band broadening that result in the increase of signal-to-noise ratio (Wishart, 2007).

2.4.2.7b) UPLC-MS system

A newer form of LC, ultra-performance liquid chromatography (UPLC), applies the same principles as high performance liquid chromatography (HPLC) but is faster than HPLC, has a higher resolution and has larger separation efficiencies especially for those peaks that elute late (Lambert et al., 1997; Swartz, 2005; Novakova etal., 2005; Theodoridis et al., 2008). UPLC utilizes columns with smaller particle size, has a greater resolution than HPLC (Wishart, 2007). This chromatographic system is normally connected to Acquity UPLC column containing X-Terra sorbent of second generation to resist backpressures (Novakova et al., 2006) and these columns are available in C18, Shield RP18, C8 and Phenyl stationary phases. These columns use ethysiloxane/silica hybrid (BEH) structure which offers the column stability at high pressure and broader pH range (Novakova et al., 2006). UPLC has fast injection cycles due to low injection volumes and has temperature control (4-40̊ C) (Novakova et al., 2006). Coupled to MS, UPLC-MS is a mixture of a 1.7µm reversed-phase packing material and chromatographic column working at a 6000-15000 psi pressure range that analyzes complex biofluids and offers quantitative reproducibility (Lu et al., 2008). This technique is a more robust and reproducible method that detects a high number of peaks with increased resolution in a shorter time period (Zelena et al., 2009). The shorter run times reduce the solvent consumption (Novakova et al., 2006). Illustrated in the figure 2.5 are the principles on which HPLC-MS system operates.

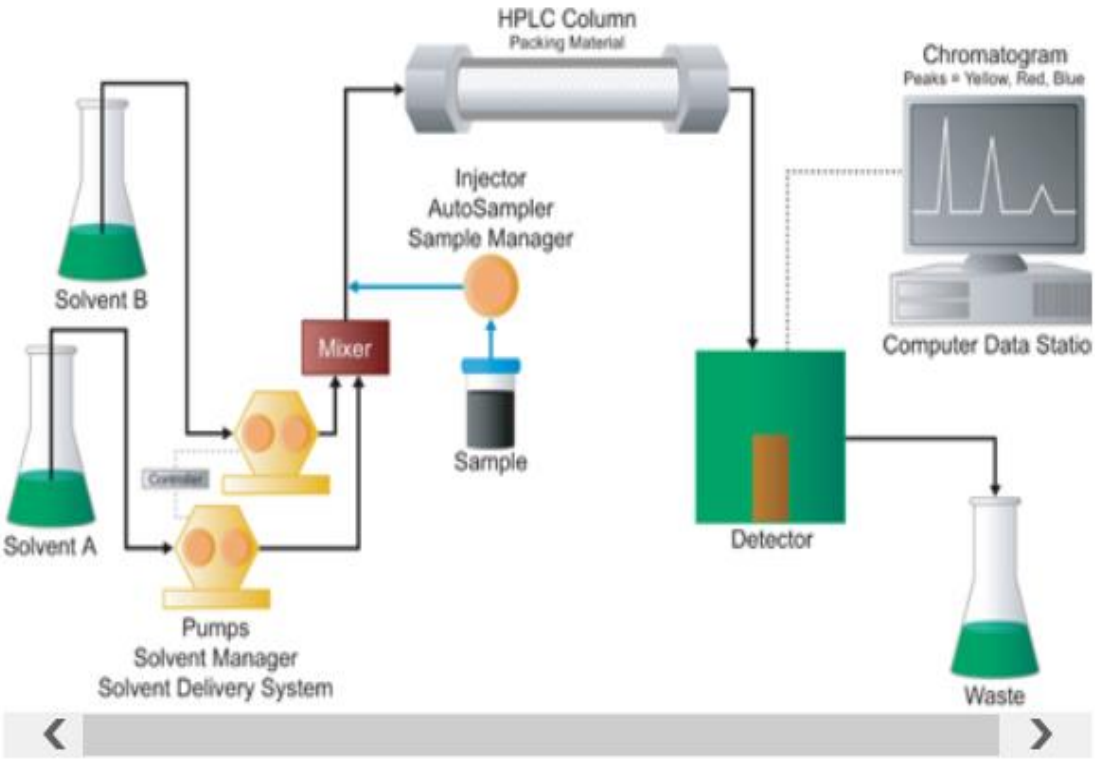


Figure 2.5: Typical illustration of an HPLC-MS system. Two solvents are pumped into a mixer, when the sample is added its complex mixtures are separated in the column and the metabolites exit the column. These metabolites are ionized and separated based on their mass-to-charge ratio, then detected and produce a signal in the form of a spectra. Figure downloaded from: http://www.waters.com/waters/en_US/Identifying-and-Quantitating-Compounds/nav.htm?cid=10049064. Link Active 13 May 2014.

The disadvantages of UPLC include low peak area repeatability values in analyses, high working pressure, and thermal effects that cause solute decomposition (Novakova et al, 2006). Lastly, minor changes in the instrument parameters result in major changes in the spectra thus making it a challenge to maintain stable signal intensity because UPLC-MS can be unstable (Lu et al., 2007). UPLC-MS metabonomics is easily performed on biofluids such as urine, saliva and cerebrospinal fluid (Psychogios et al., 2011) and has been used in the identification and quantification of low molecular weight molecules (Psychogios et al., 2011).

A combination of feature selection techniques and new methods are required for unrestricted classification of biological fluids in metabonomics studies (Wu et al., 2009). Metabolites can be validated by point analysis assays that detect and monitor only one metabolite or by using pure standards of the potential metabolites and analyzing them using metabonomic

instrumentations to verify their retention time, fragmentation and/or adduct-forming patterns (Kristensen et al., 2012). Typical base peak ion (BPI) chromatograms of serum samples or any other sample can be obtained through analysis in either positive or negative electrospray ionization (ESI- or ESI+) mode of LC-MS. An example of serum sample BPI chromatograms analyzed in ESI+ and ESI- mode of UPLC-MS is shown below in figure 2.6. Analysis of these output results (chromatograms and spectra) can be conducted using various statistical methods. To identify metabolites with LC-MS, peaks have to be identified and quantified followed by the analysis of the spectra (for metabolic fingerprinting) through the use of chemometric or multivariate statistical methods (Villas-Boas et al., 2007).

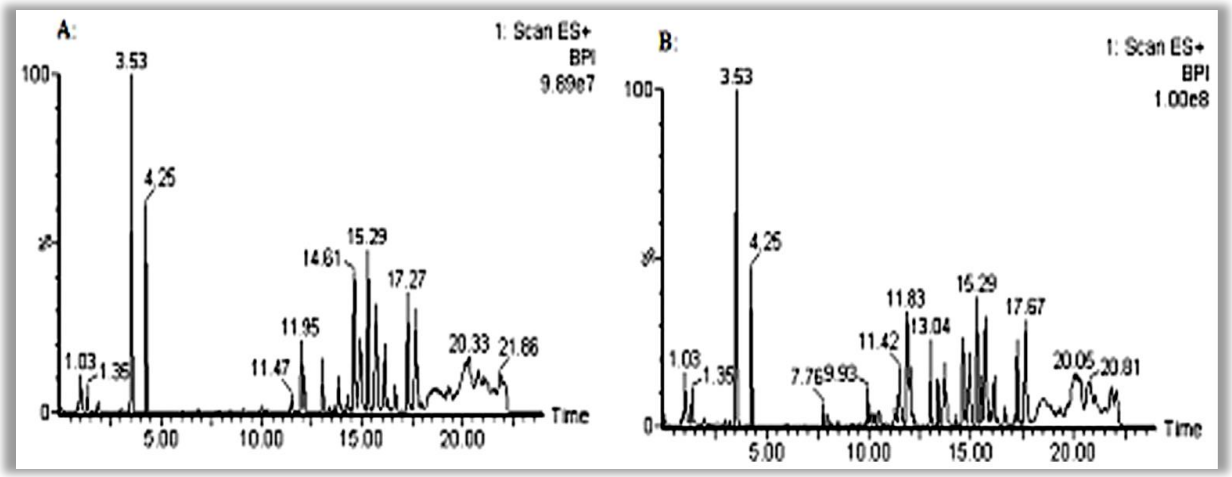


Figure 2.6: Typical BPI ion chromatograms serum samples of samples analyzed in the positive and negative ESI modes of UPLC-MS (Peng et al., 2014).

Massive datasets are generated in metabonomics studies because complex sample mixtures of living sytems and multivariate statistics are needed for identification of metabolites/biomarkers (Gika et al., 2007; Theodoridis et al., 2008).

2.5) Chemometrics and Bioinformatics

Bioinformatics is a practical science that uses computational techniques with biology for functions such as database design, data mining, explanation of data clustering etc for collection and analysis of complex biological samples (Luscombe et al., 2001). It is aimed at developing software tools that can be utilized to generate important biological information. Many software tools are available for biological data manipulation and more are still being developed, however; these tools do not always equate to improved analysis of biological data.

It is a fact that metabonomics studies rely on bioinformatics for data processing and analysis. Software tools like MassLynx v4.1 and Statistical Package for the Social Sciences (SPSS) 19.0 used in this study for data processing proved valuable. Bioinformatics was applied by using databases for linkage, from which metabolites were identified for biological interpretation.

Prior to the application of statistics, data pretreatment is usually performed to assist with normalization and standardization of the data. The pre-treatment method chosen is dependent on the hypothesis of the study, characteristics of the data sets and analytical tools utilized (Eriksson et al., 2006; van den Berg et al., 2006). These data pretreatment methods include class I (centering), class II (auto scaling, range scaling, pareto scaling, vast scaling and level scaling) and class III (log transformations and transformation (Eriksson et al., 2006; van den Berg et al., 2006). Pareto scaling (of importance to the study presented here) is aimed at reducing relatively large values and still partially maintaining the intact data set using the square root of the standard deviation. This scaling method results in the reduction of large folds of data without rendering the data dimensionless (van den Berg et al., 2006). By using Pareto scaling, the measurements remain closer to the original value compared to using autoscaling (Eriksson et al., 2006; van den Berg et al, 2006) however, Pareto scaling is sensitive towards fold changes (van den Berg et al., 2006). Transformation, which was also used in this study, ensures the decorrelation of the given signal and the containment of energy in a small number of components (Wang, 2007). After data pre-treatments then statistics can be conducted.

Metabolomic studies require large sample sizes, whether longitudinally or cross-sectionally in order to perform statistics that can confidently report on metabolite levels, trends or responses and controls or references are to be utilized all the time (Villas-Boas et al., 2007). This massive data generated by metabonomics studies is analyzed using chemometrics. According to Wold (1995)" chemometrics is the chemical discipline that uses mathematical and statistical methods, (a) to design or select optimal measurement procedures and experiments, and (b) to provide maximum chemical information by analyzing chemical data." The chemometrics methods or multivariate methods produce output in a form of projected spectra or chromatograms (Villas-Boas et al., 2007; Lin et al., 2011). New chemometric methods are now focusing on looking at all global molecules or metabolites when conducting statistics (Villas-Boas et al., 2007).

This UPLC-MS study employed chemometrics because the majority of LC-MS studies rely on chemometric methods for comparisons between healthy and diseased samples. This is because identification and quantification of metabolites is limited by the availability of few libraries (Idborg-Bjorkman et al., 2005; Plumb et al., 2005; Wilson et al., 2005). The identification process of compounds based on MS data is very time consuming and riddledwith pitfalls. The effectiveness of using chemometrics instead of libraries is largely dependent on controlled sample collection, sample preparation and the comparisons made to the subjects at hand (Villas-Boas et al., 2007).

Even though metabonomics studies mainly use multivariate analysis, it's recommended that both univariate and multivariate statistics be used for analyzing data. This is because univariate analysis like analysis of variance (ANOVA) measures one parameter at a time while multivariate analysis measures different parameters simultaneously. Multivariate analysis is favored by metabonomics studies because the induced effect or cause of a metabolic complication is not always a result of a change in one metabolite at a time, rather a synergy of simultaneously affected metabolites which may be altered in different manners to produce an effect. Many multivariate techniques exist for the efficient projection of data, this includes; linear discriminant analysis (LDA), principal component analysis (PCA) and orthogonal projection to latent structures discriminant analysis (OPLS-DA) (Gika et al., 2007; Stenlund et al., 2008).

2.5.1) Analysis of variance

Analysis of variance (ANOVA) is a univariate statistical method that indicates significant differences between means of two groups of data (Bartlett et al., 2000; Axelson, 2010) and is normally referred to as one way ANOVA. According to Altman (1991) ANOVA's variance operates on the principle that involves the separation of the total dataset's variability into components resulting from various sources of variation. It evaluates the hypothesis based on the fact that all samples have an equal mean against means differing (Axelson, 2010). The basis of the analysis is that the populations from which the samples are derived are normally distributed containing the same standard deviation (Altman, 1991). ANOVA analysis of ^{1}H NMR based metabonomics data proved that the serum spectra of HIV positive or AIDS

patients were separable from that of healthy HIV negative individuals (Hewer et al., 2006). Multivariate ANOVA can be done and it is called MANOVA (Tabachnick et al., 1996).

2.5.2) Linear discriminant analysis and mean plots

Linear discriminant analysis (LDA) is a supervised multivariate method that reduces dimensionality of data without rendering the data useless because class discriminating information is highly preserved (Lu et al., 2005; Axelson, 2010). It finds a linear transformation of the predictors to provide an accurate discrimination that predictors cannot provide separately (Axelson, 2010). LDA simultaneously maximizes the between class clustering while minimizing within class variation (Axelson, 2010; Chen et al., 2012). It discriminates different groups and classifies samples into correct categories that were not labelled into their classes (Mika et al., 1999). Disadvantages of LDA include; not using higher order information because only the first and second order statistics are applied and the dimensionality of the projected space cannot be more than the number of the data classes (Chen et al., 2012). An illustration of how classes separate on an LDA scatter plot is shown in figure 2.7 below.

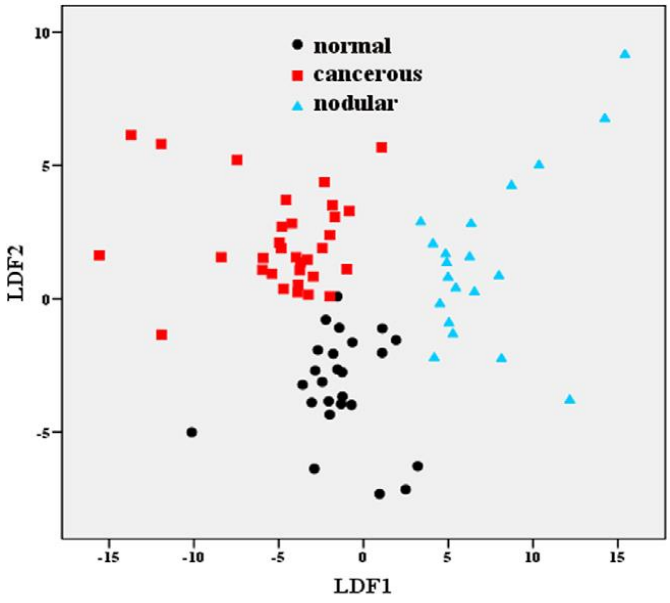


Figure 2.7: LDA scatter plot of tissue spectra showing the three experimental groups (normal, cancerous and nodular) which are separated from each other with some within group variations and minor overlaps (Li et al., 2014).

The LDA scatter plots show the greatest variation that can be observed among the experimental groups, and this variation is brought about by variables that cause maximum between group separations and minimal within group variations. These variables do not necessarily have to be significantly altered to cause such separations and their relative levels can be measured using mean plots.

Relative levels of the variables that are responsible for the clustering observed on the LDA scatter plots can be shown and measured using mean plots. These mean plots evaluate whether the mean differs between various groups of data and can evaluate the change in the mean over time in ungrouped cases of data (Croakin & Tobias, 2013). Even though mean plots are normally used in combination with standard deviations; the mean plot is known to monitor shifts in location of a variable that occur between or among groups (Croakin & Tobias, 2013). Relative magnitudes of change of a variable among groups can be measured and the plot can also provide clear patterns of the variable among the different groups (Croakin & Tobias, 2013).

2.5.3) Principal Component Analysis

PCA is an unsupervised multivariate statistical data tool (Eriksson et al., 2006; Lendl & Karlberg, 2005) that gives an overview or systematic view of the grouping trends and allows for easier inspection of outliers (Eriksson et al., 2006; Wu et al., 2009; Ilin & Raiko, 2010). It is an unsupervised statistical method because it does not require prior knowledge to be known about the subjects or samples before classification, therefore making it a neutral method (Gika et al., 2007). This scale dependent tool is aimed at explaining variation in datasets by using the least number of components (van der Berg et al., 2006) and two principal components normally define a PCA model (Eriksson et al., 2006). It classifies observations/scores (e.g samples) and the clustering of the samples is determined by the variables, which can be of spectral (NMR or infrared) or chromatographic origin such as GC or LC (Eriksson et al., 2006). PCA has interpreted NMR, MS, and FTIR spectral profiles of different samples (blood, urine or tissue) into specific categories, conditions or disease states (Holmes et al., 2000; Smith & Baert, 2003; Wilson et al., 2005). This pattern recognition technique has succeeded in identifying inborn errors of the metabolism (IEM) and monitoring of drug toxicities (Wishart, 2006). PCA is also used for process monitoring and calibration

of analytical analysis (Shaffer, 2002). An illustration of a PCA scores plot is shown in figure 2.8.

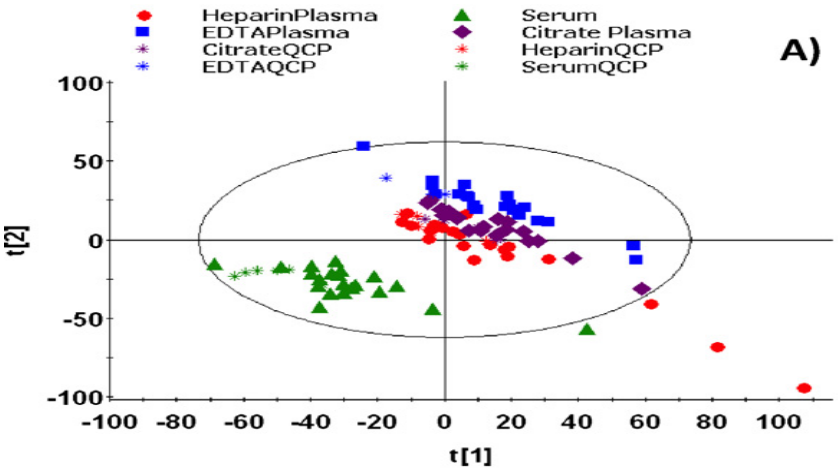
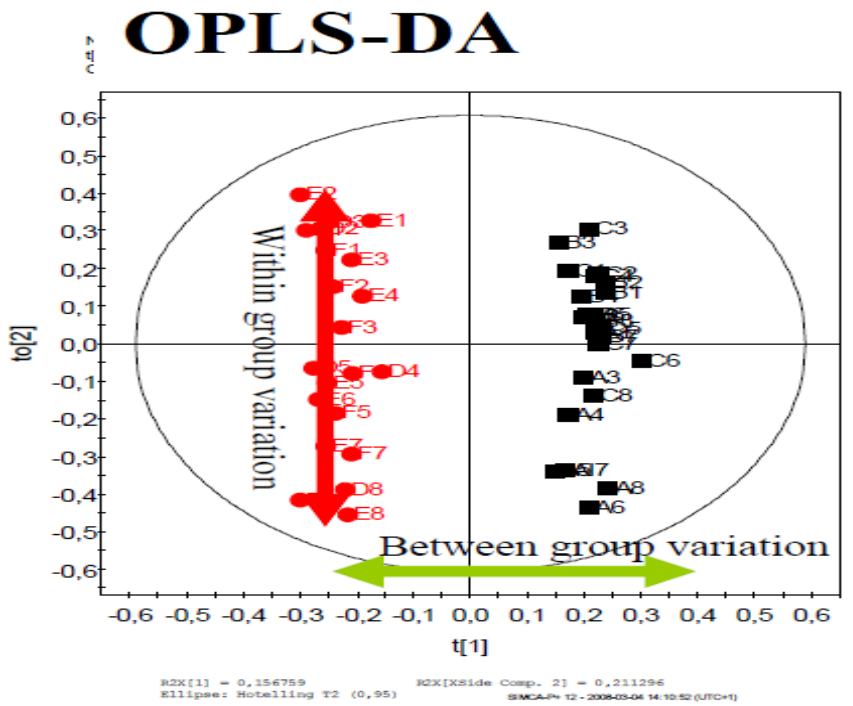


Figure 2.8: PCA scores plot showing pareto scaled data of serum samples, plasma samples and their quality control (QC) samples that were analyzed in ESI positive mode of UPLC-QTOF-MS. Plasma and serum samples distinguishable from each other (Barri & Dragsted, 2013).

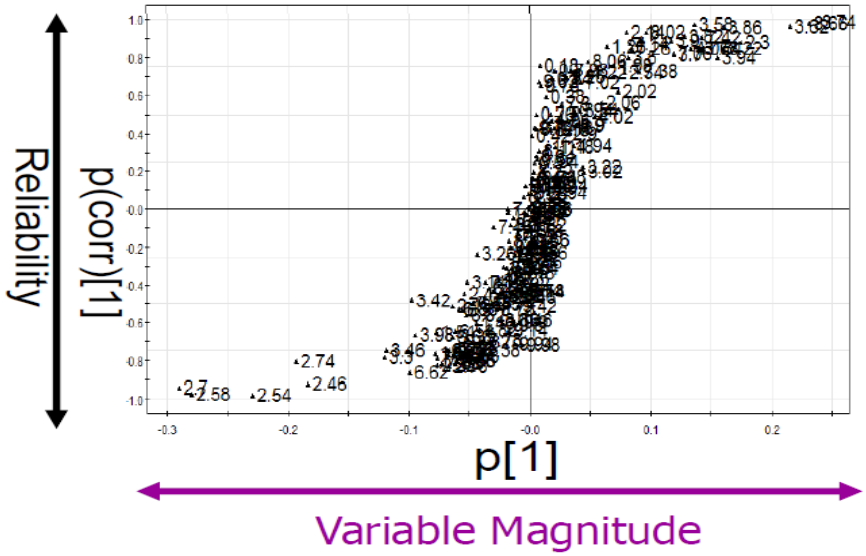
2.5.4) Orthogonal projection to latent structures discriminant analysis

OPLS-DA is a supervised modeling technique (Lendl & Karlberg, 2005) used for discrimination (Stenlund et al., 2008; Westerhuis et al., 2010). It facilitates the separation and interpretation of different types of variations in two distinct samples (Stenlund et al., 2008) because it performs paired analysis (Westerhuis et al., 2010). It is supervised because it requires prior knowledge of the sample before classification and maybe regarded as biased (Lendl & Karlberg, 2005). It consists of a single component that that predicts class difference and variation is determined by the second orthogonal component relative to the first principal component (Westerhuis et al., 2010). It is a very versatile and usable technique in biological and medical research and separates predictive from non-predictive variations (Stenlund et al., 2008). Even though OPLS-DA can discriminant more than two classes of samples, it works best with two classes because it makes it easier to interpret data (Wiklund et al., 2008; Westerhuis et al., 2010). The class discriminating information between the two groups is shown on the left and right side of the scores plot along the first principal component, while the within group variations are shown up and down along the second principal component (Stenlund et al., 2008).

An OPLS-DA scores plot is normally accompanied by an S-Plot that generally facilitates biomarker discovery. This plot determines variable contributions compared confidence, from this information, the magnitude and reliability of varying variable can be determined (Wiklund, 2008). The variables or coefficients with large changes between groups are found at the extreme ends of the "S" shape and the most abundant or common changes are concentrated closer to the center of the plot ("S" shape) (Wiklund, 2008). Variables at the extreme ends of the ''S'' shape represent potential biomarkers (Wiklund, 2008). A representative OPLS-DA scores plot and accompanying S-Plot with indications showing the basis of separation as determined among samples is illustrated in figure 2.9. OPLS-DA has separated metabolic profiles of diseased compared to samples (Qi et al., 2012; Zhang et al.,2013) OPLS-DA has successfully differentiated metabolic profiles of HIV positive patients suffering from bronchoalvoelar lavage fluid from HIV negative subjects (Cribbs et al., 2014).



A



B

Figure 2.9: (A) An OPLS-DA scores plot showing between and within class variations and (B) accompanying S-plot containing the variables contributing to the separation observed on the OPLS-DA scores plot (Wiklund, 2008). Potential biomarkers at extreme ends of the S shape.

2.6) Metabonomics Applied to HIV/AIDS

+Most HIV/AIDS metabonomics studies focus on the discovery of metabolic biomarkers that differentiate HIV/AIDS patients (receiving or not receiving treatment) from HIV negative subjects. A number of molecules that have shown biomarker characteristics for diagnosis, prognosis, and monitoring of HIV/AIDS metabolic complications include disrupted levels of sugars, amino acids, nucleotides, and lipids. Despite the discovery of these biomarker molecules, no agreed on HIV global metabolic profile exists. There is a need for the development of a global metabolic profile for the better management of HIV/AIDS and its therapy related metabolic complications. This gap is one of the reasons why there is no significant improvement in the management of HIV/AIDS induced metabolic complications. This study attempts to contribute in filling that gap. Success of a project like this is also dependent on work that other metabonomics studies on different human samples delivered towards achieving a common goal of discovering HIV/AIDS biomarkers.

Spectroscopic HIV/AIDS metabonomics studies will be discussed first that include FTIR and NMR and will be followed by MS-based studies. Spectroscopic work by Bahmani et al (2009) demonstrated that components of plasma samples belonging to HIV positive patients not receiving treatment were different from HIV negative subjects. This was demonstrated using Vis-near infrared spectroscopy which produced discriminating spectra of the two groups showing that the HIV positive spectral patterns were specific. In a diagnostic study by Sitole et al (2014), HIV negative subjects were successfully distinguished from HIV positive patients using FTIR. Among HIV/AIDS metabonomics studies, a number of reviews of HIV and/or microorganism metabolism, which includes viruses, have recently been published. A review by Pendyala et al (2007) highlighted the identification of reliable HIV or central nervous system disease biomarkers in order to gain a better insight into the mechanisms of neuropathogenesis, and ways to monitor HIV treatment benefits. Another review by Feeney et al (2011) on insulin resistance in HIV infection, suggested that IR screening as valuable and that newer anti-HIV agents should be monitored closely in order to pick-up on insulin and glucose homeostasis disruption. Finally, a review by Sitole et al (2013) showed that metabonomic analysis (targeted or untargeted) of HIV-infected biofluids using mass spectrometry and/or NMR can differentiate HIV experimental groups based on multiple metabolites and metabolite groups.

The majority of HIV/AIDS metabonomics data has been collected using NMR and MS, starting with a pioneering paper from this lab in 2006. In an NMR study from our lab include, Hewer et al (2006) suggested that HIV antiviral therapy (ARVs) use can be monitored using ^{1}H NMR spectroscopy after demonstrating that serum metabolic profiles of HIV negative and HIV positive (receiving and not receiving ARVs) patients were distinguishable using this technique. Next Phillipeos et al (2009) demonstrated in a comparative study that ^{1}H NMR spectroscopy distinguished HIV negative subjects from HIV positive patients and presented a new chemometrics methodology for characterizing experimental groups. This statistics approach is called multinomial linear regression and proved effective in the biological classification of HIV serum samples. Similar NMR work conducted by Munshi et al (2013) showed differentially regulated metabolites in biofluids distinguishing HIV negative subjects from HIV positive patients. As mentioned earlier, the use of different samples to discover HIV/AIDS biomarkers is recommended highly. Maher et al (2011) achieved a metabolic correlation among tissue samples, blood plasma and cerebrospinal fluid biofluids of HIV positive patients who were receiving HAART and were prone to neuropsychological dysfunction. Authors showed that the statistical integration of ^{1}H NMR spectroscopy data generated from biofluids for the extraction of biological information regarding the distribution of metabolites within an organism could be obtained.

Some MS-based work includes; Williams et al (2012) from our group, who identified biomarkers detectable by mass spectrometry metabolomics as indicators of mitochondrial dysfunction caused by HIV infection. Possible application of these biomarkers in the monitoring of HIV disease progression and the response to HIV treatment were considered. Williams et al (2012) used a qualitative GC-MS study conducted on serum samples of HIV negative and HIV positive patients that were not on treatment and the results of this study yielded significantly altered metabolites that were linked to mitochondrial dysfunction. Cassol et al (2013) found that lipid alterations of HIV positive patients receiving PIs were linked to markers of inflammation, microbial translocation, and hepatic function. The observations were made during a global plasma study using ultra high GC-MS, LC-MS, LC-MS/MS. In an LC-MS study, Hollenbaugh et al (2011) validated the effective use of HPLC-MS/MS technology for the monitoring of metabolic alterations caused by HIV-1 infection. In this study, it was determined that different cell types exhibited different metabolic results during the production of HIV, an observation already predicted by Villas-Boas et al (2007) on metabonomic samples in general. Ghannoum et al (2011) showed that GC-MS, UPLC-MS

and LC-MS/MS metabolomics revealed differential levels of oral metabolites in saliva from HIV positive patients compared to HIV negative subjects and this lead to the discovery and identification of potential HIV oral metabolites that may have novel diagnostic application. Some LC-MS studies selected to only use one ESI mode (Hollenbaugh et al., 2011) whilst others used both the ESI- and ESI+ mode (Ghannoum et al., 2011). There is currently no known knowledge of LC-MS HIV/AIDS metabonomics work that has been conducted on serum samples and using both ESI modes for a broader coverage of the metabolome. The study presented here attempts to achieve a global serum metabolic profile using UPLC-MS.

There is still need for further studies to cover a broader range of biomarkers for HIV/AIDS so that it can lead to the successful development of the HIV/AIDS and HAART global metabolic profiles. In this study, human blood samples were collected, sera prepared and metabolites extracted then analyzed using UPLC-MS. Statistical analysis was conducted to identify metabolites that were significantly altered by HIV and/or HAART and linked to respective metabolic pathways.

2.7) Rationale

Most investigations for the management of HIV/AIDS have been focused on the immune system biomarkers (e.g CD4 count and viral load). The discovery of metabolic biomarkers has been studied to a lesser extent due to hindering factors such as the methodologies in use for measuring metabolic complications. Unlike conventional methods of measuring metabolic complications, metabonomics combined with the application of multivariate statistical data analysis is a promising tool for biomarker discovery. This approach accommodates the investigation of the broad effects of HIV/AIDS on the metabolism and identified biomarkers have the potential to aid in the diagnosis, prognosis and monitoring of metabolic complications for the management of HI/AIDS.

2.8) Null Hypothesis

UPLC-MS cannot detect differences induced by HIV and/or HAART serum metabolites.

2.9) Hypothesis

HIV and/or HAART induced metabolic changes are detectable in sera through UPLC-MS metabonomics and can lead to the identification of biomarkers that may aid in the differentiation of HIV negative subjects and HIV positive (receiving and not receiving HAART) patients.

2.10) Overall Objective

The overall objective was to determine whether UPLC-MS would identify metabolites (in serum samples) and metabolic pathways associated with HIV and HAART. Because metabonomics methodologies in general are aimed at contributing to improving HIV/AIDS management, it is possible that the sensitivity and specificity of MS provides advantages to meet this requirement.

2.11) Aims

- To develop a serum sample metabolite extraction protocol for UPLC-MS analysis.
- To identify sera metabolites that are significantly affected by HIV and/or HAART using UPLC-MS.
- To use FTIR to detect similar metabolic patterns as those detected by UPLC-MS because FTIR is easier to perform even though it is a less sensitive technique.
- To link metabolic changes to specific metabolic pathways (biological interpretation using relevant databases).

The next chapter, (Chapter 3, Materials and Methods) explains the details of the experimental methodologies (including all materials) that were used to achieve the above mentioned aims.

CHAPTER 3: MATERIALS AND METHODS

Two sets of experiments were performed in this study; the first being UPLC-MS experiments to detect metabolic changes of metabolites significantly affected by HIV and/or HAART and secondly, FTIR experiments to see if the latter technique could lead to a conclusion similar to those made by a more sensitive approach. Even though FTIR is a less sensitive technique than UPLC-MS, it can provide a metabolic fingerprint of a sample. The rest of the work for this project was computer-based; extensive chemometrics and bioinformatics.

3.1) Ethics Approval

Ethics approval for the use of human blood in this research was obtained from the Faculties of Natural and Agricultural Sciences and Health Sciences of the University of Pretoria with protocol numbers EC 080506-19 and 163/2008, respectively.

3.2) Sample Collection and Sera Preparation

Blood samples of HIV negative (n=32), HIV+ (n=29) and HIV+HAART+ (n=34) donors were collected respectively, giving a total of 95 samples. Venous blood was collected in non ethylenediaminetetraacetic acid (EDTA) vacutainers (Greiner Bio-One GmbH, Kremsmünster). Blood was collected from these study groups following written informed consent. Healthy (HIV-free) control donors were recruited at the Hatfield campus of the University of Pretoria and had no known metabolic or other medical condition at the time of blood collection. The HIV negative status of these samples was confirmed with VISITECT ® HIV 1/2 rapid tests (Omega Diagnostics Limited, Scotland, UK). This rapid test utilizes the principle of immunochromatography, which is a quantitative immune assay that functions by binding of an antigen to an antibody to form a reaction (Zuk et al, 1985).

HIV treated and untreated donors were obtained from the King's Hope Development Foundation and Mooiplaas Clinics in Diepsloot, Johannesburg with the exception of one sample collected from the Pretoria Academic Hospital in Pretoria. Serum was removed by centrifugation at 1610 × g for 10 minutes (Allegra 25R Beckman Coulter, USA) and stored at -70 °C until use. The experimental groups were fairly well-matched in terms of gender but not age. The HIV negative group was younger than both the HIV+ and HIV+HAART+ groups, this turned out to not be a confounding factor for the research (see appendix, figure

A8). Patient samples were selected randomly to test how well the approach would perform in situations where patient visits are usually not by appointment. The participating HIV+ donors had not been diagnosed as having AIDS and patients on treatment were also on different HAART combinations. Different HAART regimens were not part of the exclusion criteria because this global study was not targeted at any specific HAART regimens. Detailed demographic information of the participating donors is supplied in Table A1 (in the appendix).

3.3) UPLC-MS methodologies

UPLC-MS analysis was conducted on two batches of data that consisted of different patients per batch that totalled up to 90 patients [HIV- (n=31), HIV+ (n=29) and HIV+HAART+ (n=30)]. The first batch contained 16 HIV negative samples, 13 HIV positive untreated samples and 15 HIV positive samples receiving HAART treatment while the second batch contained 15 HIV negative samples, 16 HIV positive untreated samples and 15 HIV positive samples on HAART treatment. The first batch of samples were analysed as part of a metabolite extraction method development strategy. The second batch of samples were analysed using the newly optimized method and advanced chemometrics applied for data processing. The UPLC-MS workflow followed in this study is illustrated in figure 3.1.

3.3.1) Chemicals and Solvents

All chemicals for UPLC-MS experiments were of ultra-pure LC-MS grade. Formic acid was purchased from Fluka (Steinheim, Germany) while acetonitrile was purchased from Romil Ltd (Cambridge, United Kingdom). Ultra-pure water was generated by a Millipore Elix 5 RO system and Millipore Advantage Milli-Q system (Millipore SAS, Molsheim, France).

3.3.2) UPLC-MS Method Development and Optimization

The first aspect considered was the type of samples to be analyzed. Being blood-derived, one would expect polar to intermediate polarity compounds to be predominantly present in sera. Due to the nature of the samples, a C18 UPLC column served as a good starting point but a C18 column with polar embedded functionality or special end capping could enhance the separation of blood sample extracts. To enhance the possibility of separating all compounds to be ionized in the ion source of the mass spectrometer, the maximum column length of 150 mm was used for all column evaluations. The columns evaluated were the BEH C18 column as well as the T3 C15 column from Waters. Although the BEH C18 column chemistry gave

good retention of the more polar compounds, it could not equal the superior retention and resolution observed on the T3 C15 column. The T3 150mm UPLC column was therefore used in all further experimental work.

The second aspect was to optimise the ionization settings of the mass spectrometer. Quality control (QC) samples which comprised a pool of all samples used in the study (Bijlsma et al., 2006) or a pool of representative samples from each condition being investigated (Dunn et al., 2011), HIV negative, HIV positive and positive HAART were prepared. The QC sample was infused into the inlet probe while scanning the TOF analyzer in both ESIPos and ESINeg ionization modes. Mass ions were observed in both modes and resulted in the optimisation of both modes with regards to capillary voltage, cone voltage and extractor voltage. The trap voltage was set at 4 V to ensure that no unnecessary fragmentation would occur during ion transmission. The settings as detailed in section 3.3.2 were found to provide the best ionization of the most compounds in both polarity modes and were used in all experimental work.

The third aspect evaluated was the stability of the method. QC samples were prepared and used to monitor the stability of sample runs during UPLC-MS analysis. A pooled sample was analyzed by doing ten consecutive injections on the same day and repeating the same analysis on the next day and comparing the two data sets. It was found that the UPLC column required three conditioning injections prior to starting the main analysis run. The results for the inter-day and intra-day experiments compared very well and ensured that extended sample runs would be possible even if the whole analysis run would span over several days. To minimise the effect of subtle changes in the separation over long analysis runs, all the samples were randomised in the sample table prior to analysis. Blanks were included to assess carryover and exclude possible contaminating ion features.

3.3.3) Metabolite extraction from sera using cold methanol extraction

Cold methanol extraction was utilized to facilitate serum metabolite extraction (5 from each experimental group). Serum samples were removed from -70°C storage and thawed on ice for approximately 60 min to avoid loss of volatiles. Fifty microliters of each serum sample was transferred into 15 individual 2 ml eppedorf tubes. Hundred and fifty microliters of 80% methanol (ice cold) was added to each serum samples in the individual eppendorf tubes,

placed on ice, then vortexed for 5 min at room temperature (RT) and centrifuged at 16100 xg for 10 min at 4°C (Eppendorf 5415R, Eppendorf, Germany). Following centrifugation, 160 ul of each supernatant was transferred to new 2ml eppendorfs and kept on ice. The pellets were resuspended in 200 µl of 80% methanol (ice cold), vortexed for 5 min at RT and then centrifuged at 16100 xg for 10 min at 4°C. Again, 160 µl of the supernatant was transferred to the 2 ml eppendorf tubes that were kept on ice. This resuspension step was repeated twice leading to a total collected supernatant volume of 480 µl in each tube. The supernatants were dried by being exposed to nitrogen gas for approximately 1 hour. Fifty microliters of 50% cold methanol was added to the individually dried residues, then vortexed for 5 min at RT, and centrifuged at 2100 xg at 10°C for 20 min. This procedure was also performed on a pooled sample that contained a mixture of all 15 samples combined. The samples were aliquoted into volumes of 200 µl in 1.5 ml Eppendorf tubes and stored at -70°C until UPLC-MS analysis.

The cold methanol extraction method (LLE) did not work well for metabolite extraction. Contamination by surfactants and phospholipids that were observed with the LLE method masked metabolites. A new method that uses inert plastic and was able to remove phospholipids and proteins was required. Ostro plates™ plates (a SPE) offered these qualities and metabolite extraction was optimized using these plates.

3.3.4) Optimized Metabolite Extraction from Sera using Ostro™ Plates

Ostro™ sample preparation plates (Waters Corporation, Milford, USA) and filtration through the Waters positive pressure-96 processor (Waters Corporation, Milford, USA) were utilized to facilitate serum metabolite extraction. Serum samples were thawed on ice prior to extraction to avoid loss of volatiles. Briefly, the Ostro™ plate was keyed onto a 96-well collection plate. One hundred microliters of sera, blanks and QC samples were randomly loaded onto the Ostro™ plates in duplicate, with a few exceptions as explained below which subsequently influenced the number of cases analysed in the positive (ESI+) and negative (ESI-) electrospray ionization mode respectively. In batch one, due to insufficient volumes, not all samples were sufficient for duplicate analyses (see materials and methods under section 4.1 for batch one sample layout), whereas batch 2 only had one HIV positive sample that was insufficient for duplication and was replaced by a new sample in the duplication well. Three hundred microliters of freshly prepared 1% formic acid in acetonitrile was

forcefully added to the plate wells and mixed through aspiration. The plate assembly containing sample and solvent was filtered under positive pressure at 80 psi (551.58 kPa). The collected filtrates in the collection plate were transferred on ice to the UPLC-MS instrument for subsequent analysis.

3.3.5) UPLC-TOF MS Analysis

All chromatographic and mass spectrometric analysis was performed at the Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa. A Waters Acquity UPLC coupled in tandem to a Waters photodiode array (PDA) detector and a SYNAPT G1 HDMS mass spectrometer was used to generate accurate mass data. Chromatographic separation was done utilizing a Waters HSS T3 column (150 mm x 2.1 mm, 1.8 µm) thermostated at 60 °C. A binary solvent mixture was used consisting of water (Eluent A) containing 10 mM formic acid (pH 2.3) and acetonitrile (Eluent B). Gradient elution was performed at a flow rate of 0.4 $ml.min^{-1}$. The initial conditions were 95% of A for 0.1 minute with a gradient profile 7 to 40% of A at 8 minutes, followed by a gradient profile 5 to 10% of A at 14 minutes and kept constant for 1 minute. The system was re-equilibrated using the initial conditions with a total runtime of 20 minutes. The injection volume was set at 3 µl with triplicate injections acquired per sample. The PDA detector was scanned between 200 and 500 nm (1.2 nm resolution) and collected 20 spectra per second.

The SYNAPT G1 mass spectrometer was used in V-optics and operated in electrospray mode. Leucine enkephalin (50 pg/mL) was used as reference calibrant to obtain typical mass accuracies between 1 and 3 mDalton. Although positive ionization is commonly used and more compounds are ionized in this state, the mass spectrometer was operated in both negative and positive mode of electrospray ionization to enable greater coverage of the metabolome for untargeted analysis. The capillary voltage was set at 2.0 kV, the sampling cone at 30 V and the extraction cone at 3 V. The micro channel plate detector voltage was set at 1950 V. The scan time was 0.1 seconds covering the 100 to 1000 Dalton (Da) mass range. The source temperature was 120 °C and the desolvation temperature set at 450 °C. Nitrogen was used as the nebulization gas at a flow rate of 800 L/h. MassLynx V4.1 (SCN 704) software was used to control the hyphenated system and for data manipulation.

3.3.6) Data Processing and Statistical Analysis

Data mining was done by processing the raw ESI- and ESI+ data files with the MarkerLynx Extended Statistics (XS) processing method of the MassLynx software. During this time, the MarkerLynx XS V4.1, SCN 704 software performed peak integration and alignment as well as background noise subtraction yielding data in a suitable format for statistical analysis. Further data analysis (such as statistical analysis) was not performed on samples resulting from cold methanol extraction, only Ostro^{TM} plate samples were subjected to further analyses. The following method processing parameters were used; retention window of 0.5-17.00 min, a mass range and tolerance of 100-1000 and 0.01 Da, extracted chromatogram (XIC) window 0.02, peak to peak baseline was 0.00, mass window was 0.05, noise elimination level was 6, marker to intensity threshold was 30 for ESI+ data and 40 for ESI- data (for both data batches), no de-isotope of the data was applied and smoothing was applied for both ionization modes.

For both datasets (batch one and two), the data was exported into Excel® and further analyzed using IBM SPSS (version 19.0). With the assistance of the Department of statistics at the University of Pretoria, multivariate statistical analysis in the form linear discriminant analysis was performed to classify experimental groups and identify metabolites contributing to group variances. In the case of the LDA based on individual metabolites, stepwise discriminant analysis was used for correct classification of sample groups. LDA was performed as previously reported by Hewer et al (2006) and Phillipeos et al (2009). The significance of these molecules were also verified through univariate analysis in the form of ANOVA.

The following was also applied to the second batch of data for advanced chemometrics. The data matrix was analysed using MarkerLynx XS EZInfo 2.0 software (Waters Corp., Milford, USA) for advanced chemometrics (PCA, OPLS-DA, trend plots etc). The following template was applied to the data; pareto-scaled, all automatically-transformed and cross validation was allowed by choosing the number of maximum components to be negative one. This was accompanied by additional noise filtering after template application to strengthen the signal response compared to noise. The mass-to-charge ratio and retention feature and ion intensity of the samples were used to create an S-plot from OPLSA-DA. The metabolites at the extreme ends of the S-shape (head or tail) were considered significant for the group separation observed on the OPLS-DA scores plot and selected as potential biomarkers.

Additional data projections and models (indicated in the results section) were also performed to extract more information from this dataset.

3.3.7) Metabolite Identification

For putative identification, an automated metabolite identification method was used because it is recommended for untargeted metabonomics/metabolomics studies that deal with complex mixtures in combination with expert knowledge (Schymanski & Neuman, 2013). This automated approach that is being used is not very accurate, however it is supported by using reliable biological databases. Formic acid and sodium adducts were taken into consideration becuase these adducts were the main possibilities in this study and detectectable by looking for a M+22 or M+46/M-H+46.

Briefly, the identification process was carried out as follows; selected loadings data of potential biomarkers from the S-plots that accompanied OPLS-DA scores plot in the EZinfo statistical package were transferred to MarkerLynx XS results bar for metabolite identification. These loadings contained metabolite features of m/z ratio and retention time. The mass search window was set at 0.1 and maximum hits chosen was 10 to search for MH+ or MH- ions data. Some of the databases/libraries used to conduct mass spectral searches and/or structure for metabolite identification include ChemSpider, BioCyc, MassBank, National Institute of Standards and Technology (NIST), Human Metabolome Database (HMDB), Golm Metabolome Databse, KEGG, National Institute of Allergy and Infectious Diseases (NIAID), PubChem, ChEBI, ChEMBL etc. The molecular formula of the analytes were obtained using MassLynx software. Searches were restricted to C, H, N, O, S, CL, F and P elementary compositions. Significantly altered metabolites were linked to metabolic pathways for biological interpretation using the database for linkage. UPLC-MS Workflow is shown in figure 3.

3.4) Attenuated total reflectance-FTIR (Mid-ATR-FTIR) Methodology

Sample count for FTIR analysis was 40 and consisted of HIV- (n=13), HIV+ (n=13) and HIV+HAART+ (n=14). These samples consisted of mostly batch two UPLC-MS samples and five new samples. Prior to FTIR analysis, one of the HIV+ samples (collected from a different clinic than the rest of the positive samples) did not dry overnight. Because it failed to dry overnight like the other samples, it was not analyzed. The FTIR workflow followed in

this study is illustrated in figure 3.2. Healthy (HIV-free) control donors were recruited at the Hatfield campus of the University of Pretoria and HIV+ and HIV+HAART+ donors were obtained from the King's Hope Development Foundation, Pretoria academic hospital and Mooiplaas Clinics in Diepsloot, Johannesburg. Serum was removed after centrifugation at 1610 × g for 10 minutes (Allegra 25R Beckman Coulter, USA) and stored at -70 °C until required for analysis.

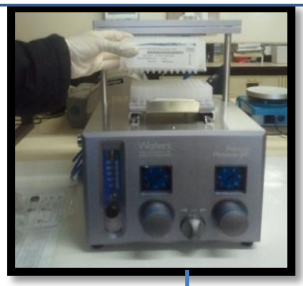
Sera samples were thawed at room temperature then heat inactivated at 56°C in a water bath. A volume of 5-10 µl of each serum sample was transferred onto separate IR-reflective glass slides (Kevley Technologies, Ohio, USA), and then left to dry overnight in order to form a homogenous film. The FTIR analysis was conducted at the Physics department of the University of Pretoria. On the day of analysis, IR spectra was acquired in absorbance mode (Mid-IR source 4000-400 cm^{-1}) using Bruker FTIR V70x spectrometer (Bruker Optik GmbH, Karlsruhe, Germany) with an ATR attachment operated by OPUS 6.5 software. A diamond single reflection ATR accessory was used to procure the mid-IR of the serum spectrums.

The atmospheric correction background spectrum was collected after every 10 samples. The ATR diamond crystal was cleaned with ethanol and dried with paper towel every time a new serum sample was analyzed. Spectra were obtained at 4 cm^{-1} resolution and 32 scans were performed per sample. Rubberband baseline correction, vector normalization and smoothing (17 points) were carried out in OPUS software. Statistical analysis in the form of LDA and ANOVA were conducted as previously discussed in the second paragraph of section 3.3.6. The data was also analysed using the MarkerLynx XS EZInfo 2.0 software to generate a PCA scores plot. The results section highlights the findings.

Ethics Approval

Sample collection and Sera preparation (n=90)
HIV- (n=31), HIV+ (n=29), HIV+HAART+ (n=30)

SPE: Ostro™ plates metabolite extraction (n=90)
Batch 1: HIV- (n=16), HIV+ (n=13), HIV+HAART+ (n=15)
Batch 2: HIV- (n=15), HIV+ (n=16), HIV+HAART+ (n=15)



LLE: Cold methanol metabolite extraction (n=15)
HIV- (n=5), HIV+ (n=5), HIV+HAART+ (n=5)

UPLC-MS Analysis (ESI- and ESI+ mode)

Data Pre- and Post-Processing (MarkerLynx v4.1)
Peak integration, peak alignment, and background noise subtraction

MarkerLynx XS EZInfo 2.0
PCA, OPLS-DA, trend plots etc

Chemometrics (statistical analysis)

IBM SPSS v19.0
LDA, mean plots and ANOVA

Bioinformatics (Metabolite I.D)
Databae/library searches for metabolite Identities (HMDB, Golm, BioCyc etc)

Metabolic pathways for biological interpretation

Figure 3.1: Summary of UPLC-MS workflow for this study.

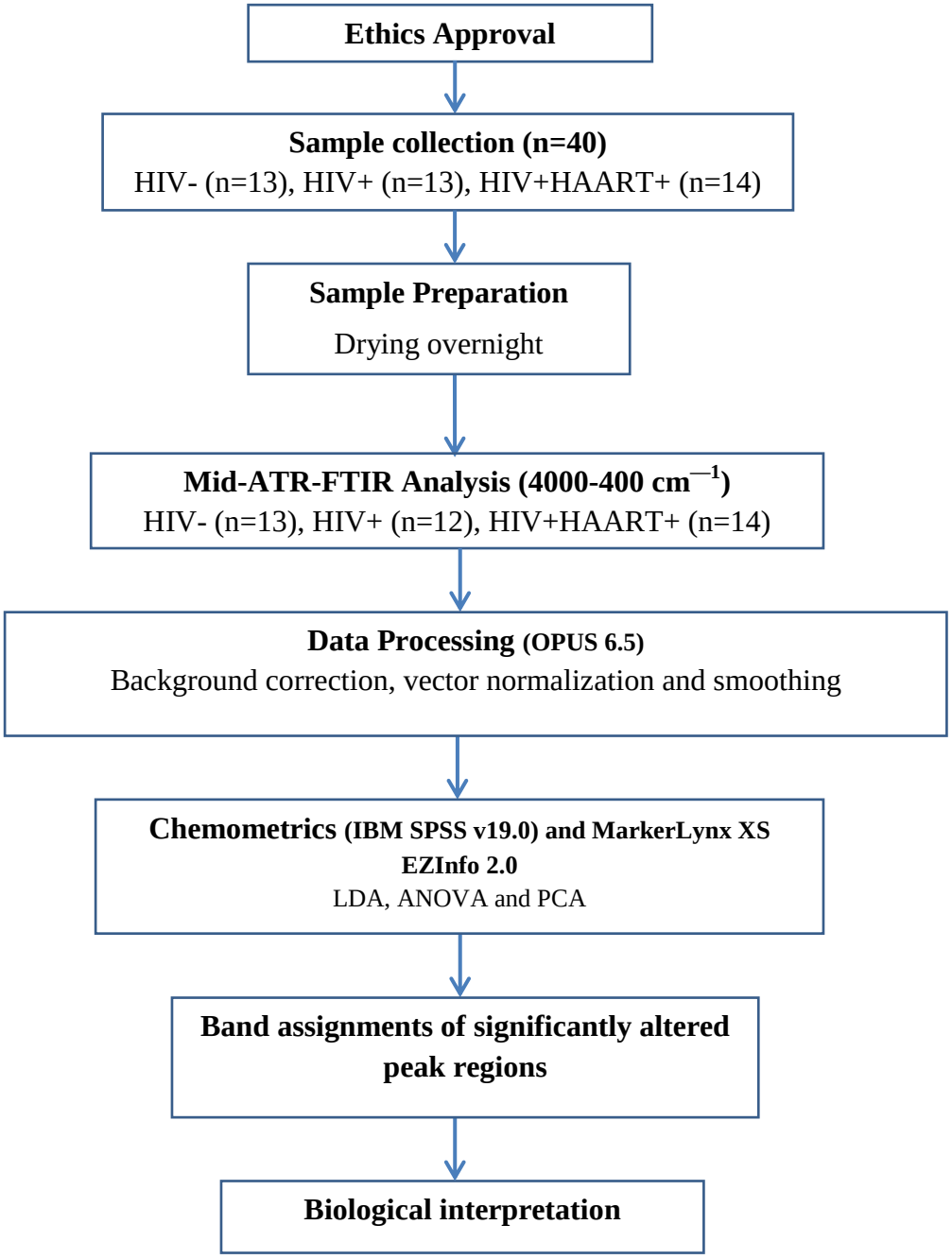


Figure 3.2: Summary of Mid-ATR-FTIR workflow for this study.

CHAPTER 4: RESULTS

Explaining chapter outline

The first aim of this project was to develop a serum metabolite extraction protocol which led to a comparison of cold methanol LLE and Ostro™ plate SPE approaches followed by UPLC-MS analysis. UPLC-MS data collected from Ostro™ plate prepared samples was published in "Current Metabolomics" as a dual first author paper between myself and Aurelia Williams. This article is presented in section 4.1. Ostro™ plates solved the problems encountered with the traditional cold methanol LLE to produce clean, narrow and distinct chromatographic features of metabolites. This proof of concept article demonstrated that solid phase extraction using Ostro™ plates led to the production of matrix free chromatograms that enabled the detection of metabolic alterations caused by HIV and/or HAART.

A second batch of samples was used to assess the reproducibility of the Ostro™ plate approach. The next dataset was also subjected to additional and new statistics tools (section 4.2). The additional advanced chemometrics used in this section (PCA, OPLS-DA, S-Plots etc) included statistical methods recommended for metabonomics studies because of being more informative. All these findings are presented in section 4.2.

Distinguishing metabolites and their relative levels were simultaneously determined for the three experimental groups by using statistical models (LDA and mean plots) that looked at sample groups altogether (HIV-, HIV+ and HIV+HAART+) and ANOVA was used to calculate the significance of the altered metabolites. Specific assignment of distinguishing metabolites responsible for separation of only two experimental groups was not achieved by these statistical methods. Therefore, the use of further statistical analysis models was required for conducting paired analysis of the sample groups for the purpose of specifically allocating the distinguishing metabolites to either HIV or HAART.

The third section of the results (section 4.3) presents FTIR data of the three experimental groups. FTIR was used because it is an easy analysis method that can also produce fingerprints of biofluids. FTIR analysis was conducted using most of the samples in batch two of the UPLC-MS analysis with an additional five new samples. It is not possible to do a direct comparison between UPLC-MS and FTIR because the former technique detects

individual metabolites while the latter technique produces a vibrational spectral biofluid profile. It is possible though to determine whether the data sets generated by the two approaches, identifies similar molecules as being influenced by HIV and/or HAART.

Section 4.1: Published manuscript

Title: UPLC-MS Metabonomics Reveals Perturbed Metabolites in HIV-Infected Sera

Running title: MS Metabonomics of HIV-Infected Sera

Aurelia Williams[1], Khanyisile Kgoadi[1], Francois Steffens[2], Paul Steenkamp[3], Debra Meyer[1]*

Biochemistry Department, University of Pretoria, Pretoria 0002, South Africa

[2] Department of Statistics, University of Pretoria, Pretoria 0002, South Africa

[3] Natural Products Research Group, Biosciences Unit, CSIR, Pretoria 001, South Africa

*Address correspondence to this author at the Biochemistry Department, University of Pretoria, Pretoria 0002, South Africa

Tel: +27 12 420 2300; Fax: +27 12 362 5302; Email: debra.meyer@up.ac.za

Correspondent footnote: Address correspondence to this author at the Biochemistry Department, University of Pretoria, Pretoria 0002, South Africa; Tel: +27 12 420 2300; Fax: +27 12 362 5302; Email: debra.meyer@up.ac.za

Keywords: Biofluid • Biomarker • HAART • HIV • LDA • Metabonomics • Ostro™ plates • UPLC-MS

Abstract

Immune responses to infection by the human immunodeficiency virus (HIV) and the use of highly active antiretroviral therapy (HAART) to treat HIV infection, contributes to metabolic irregularities in the host. Current methods for the extraction and identification of metabolites in biofluids generally make use of laborious, time-consuming protocols. Here, 96-well Ostro™ plates and filtration under positive pressure was used to facilitate the simultaneous, reproducible extraction of metabolites from multiple serum samples which were then analyzed by ultra-performance liquid chromatography mass spectrometry (UPLC-MS). The easy to use solid phase extraction (SPE) protocol eliminated numerous potential contaminants while the UPLC-MS detection of metabolites produced visibly different chromatograms for HIV negative (n=16), HIV+ (n=13) and HIV+HAART+ (n=15) serum samples. Linear discriminant analysis (LDA) amplified these differences, classified the groups with 100% accuracy and identified biomarkers explaining the greatest variances between the groups. The 21 metabolites altered by HIV and/or HAART primarily represented those linked to lipid and energy pathways which is where known metabolic changes associated with HIV infection occur. This work demonstrated for the first time that Ostro™ plates and UPLC-MS metabonomics was able to successfully identify distinct differences between the experimental groups and detected metabolites related to HAART and other drugs used in the treatment of HIV-associated conditions. The findings of this approach suggests a possible role for this methodology in disease prognosis as well as in the monitoring of treatment success or failure and linking treatment to metabolic complications.

Keywords: Biofluid • Biomarker • HAART • HIV • LDA • Metabonomics • Ostro™ plates • UPLC-MS

INTRODUCTION

During HIV infection, host metabolism is influenced among other things by the virus [1-3], immune responses [4-6] and treatment [7, 8]. HAART, primarily used to restore HIV-induced effects, prolongs the life-span of infected individuals [8] and has subsequently resulted in HIV infection becoming a chronic condition, especially in developed countries. As a result, there has been an increase in the incidence of chronically infected individuals presenting with HIV and treatement-induced metabolic disturbances. Metabolic dysfunction during HIV infection is not a singular event so much so that a combination of metabolic changes can be induced at one time. This combination of metabolic disorders are routinely defined as metabolic syndrome in the relevant literature where the terminology refers to for example insulin resistance, dyslipidaemia, lipodystrophy, and so on, sometimes occurring in concert [9-11].

Metabolites associated with HIV and/or HAART-induced metabolic disturbances include glucose, other carbohydrates and lipids. Alterations in the metabolism of carbohydrates (e.g glucose, [12]) and lipids [13, 14] have been identified very early in AIDS research as key metabolites affected during HIV infection. These early observations were however, made using laborious conventional rather than metabonomics methodologies [15].

Because metabolites are (i) the end products of all enzymatic/chemical reactions occurring in the body, (ii) altered by infection and treatment, (iii) detectable through analytical techniques and (iv) informative regarding an individual's physiological state [16], these molecules are being revisited for their role as biomarkers of HIV infection (see studies highlighted in [15]. To date, accurate and reliable markers for the diagnosis of HIV infection, monitoring of disease progression and assessing the success of therapeutic interventions are still lacking [17]. This is because many molecules that are influenced by HIV infection fail to meet the basic requirements of a biomarker (as defined by Kanekar, 2010 [18] as well as Beger and Colatsky, 2012 [19]). In addition; the mechanisms of HIV infection, the functioning of the metabolic and immune systems as well as the roles of various molecules and cells (such as antibodies, CD4 T cells, *etc*) during infection, are not fully understood [20] and as such hampers the characterization of biomarkers. CD4 count and viral load are two parameters currently used in the clinic for the prognosis and management of HIV/AIDS but are not always reliable (e.g changes in CD4 count due to other infections, collection time *etc*.)

supporting the need to develop new detection methodologies and novel biomarkers for the characterization of HIV infection and treatment success/failure.

Traditional biochemical assays used to measure metabolic changes have the disadvantage of detecting one analyte at a time. Studies measuring several metabolic changes through the use of multiparametric analytical tools such as UPLC-MS and the application of metabonomics methodologies have already demonstrated the advantage of this holistic approach to better understanding the effects of HIV infection [21, 22]. UPLC-MS has numerous advantages over conventional biochemical analysis of metabolites including data collected in relatively short time frames, low sample volumes, large separation efficiencies and most importantly higher resolutions [23-25]. Although analytical instrumentation such as UPLC-MS can detect low molecular weight molecules effeciently, sample preparation for these type of experiments is of extreme importance if accurate clinical information is to be obtained [16]. In previous MS studies where metabolic changes were measured; tedious, time-consuming extraction protocols were used (see Figure 1 and associated references). These protocols comprise several extraction steps which increases the variability of the data, lowering both the validity and the reliability thereof. To minimize these issues, the Metabolomics Standards Initiative (MSI) published guidelines to help standardize metabonomics sample handling, data analysis and data storage protocols [26].

As a contribution to the MSI, in this article, an easy, rapid yet reproducible method for extracting metabolites from HIV-infected sera through the use of Ostro™ plates for sample analysis in UPLC-MS metabonomics is reported (also outlined in Figure 1). Although Ostro™ and other 96-well SPE plates, under vacuum filtration, have been applied to non-HIV studies [27-30], biofluid metabonomics experiments [31] as well as to plasma samples to determine the concentration of antiretroviral drugs [32, 33], such analysis has not been applied to HIV-infected biofluids for the identification of virus and/or HAART-influenced metabolic differences between groups. Neither has the extraction of metabolites been facilitated through filtration under positive pressure. The advantage of filtration under positive pressure is that it allows for the uniform, reproducible recovery of a diverse group of analytes and an improved flow for viscous samples. Most plate formats incorporate several steps (plate conditioning followed by sample loading and well washing) [28, 32-34 and Figure 1] before the final sample is eluted while Ostro™ plates use a one step precipitation and extraction process. Sample preparation, handling time and long extraction steps are eliminated which minimize sample degradation, as reported by others [31]. Less sample

handling means less variation in the data and less time spent on method development. Ostro™ plate procedures are generally also less destructive to metabolites [30], the plastic used in the manufacture of these plates is inert minimizing leaching of plasticizers. The most novel aspect of these plates are their ability to remove phospholipids from samples. Phospholipids contribute to matrix effects and can accumulate on the chromatographic column ultimately resulting in variable analyte signals. Phospholipids are also known to cause ion suppression during MS analysis, masking analyte detection [35].

The aim of this work was to apply an easy to perform SPE protocol (using Ostro™ plates) to better prepare serum samples for improved UPLC-MS biofluid metabonomics data collection. It was important to demonstrate that the protocols (for sample extraction and data collection) could be applied to randomly collected serum samples (to show that the method applied was able to detect metabolic changes in any patient irrespective of disease status or other possibly confounding factors like age, lifestyle etc) because metabonomics is ultimately aimed at clinical application. In the data presented here, clear differences were observed between the metabolic profiles of the three study groups prior to the use of statistics. These differences were amplified following multivariate analysis. LDA classified the groups as HIV negative, HIV+ and HIV+HAART+ with 100% accuracy. The variances between the groups were mainly explained by molecules involved in lipid and energy metabolism as well as metabolites (e.g. nucleotides) influenced by treatment of the patients. This is the first time, using Ostro™ plates and a metabonomics approach that the metabolites of drugs used to treat HIV-associated conditions are detected in infected biofluids such as serum.

MATERIALS AND METHODS

Experimental Design and Sample Preparation

A global analysis was done on the sera of HIV-seronegative (HIV negative), HIV-seropositive (HIV+) and HIV+ individuals undergoing treatment (HIV+HAART+). Blood samples of HIV negative, HIV+ and HIV+HAART donors (n = 16, 13 and 15 respectively) were obtained from The Fountain of Hope Clinic in Pretoria as well as the King's Hope Development Foundation and Mooiplaas Clinics in Diepsloot, Johannesburg. Patient samples were selected randomly to test how well the approach would perform in situations where patient visits are not always by appointment. Blood was collected from these study groups following written informed consent. The participating HIV+ donors had not been diagnosed as having AIDS with the exception of two patients who presented with CD4 counts of less

than 200 cells/mm^3 after the start of the study. Patients on treatment were also prescribed different HAART combinations. Viral load, measured by the COBAS AmpliPrep/COBAS TaqMan HIV-1 Test (Roche Molecular Systems, Pleasanton, CA) was not part of the inclusion criteria because practices at the source clinics were such that the viral load of individuals was determined only when they were tested for HIV infection for the first time and then immediately before treatment was initiated. Healthy control donors were recruited at the Hatfield campus of the University of Pretoria and had no known metabolic or other medical condition at the time of blood collection. The HIV negative status of these samples was confirmed with VISITECT ® HIV 1/2 rapid tests (Omega Diagnostics Limited, Scotland, UK). The negative controls and HIV+ samples were fairly well-matched in terms of gender but not age (HIV negative group was younger than HIV+ groups). Detailed demographic information of the participating donors is supplied in Table 1.

Venous blood was collected in non-EDTA vacutainers (Greiner Bio-One GmbH, Kremsmünster). Serum was removed after centrifugation at 1610 × g for 10 minutes and stored at -70 °C until use. The time from venous puncture to storage of the sera was approximately 4 hours.

Chemicals and Solvents

All chemicals for UPLC-MS experiments were of ultra-pure LC-MS grade. Formic acid was purchased from Fluka (Steinheim, Germany) while acetonitrile was purchased from Romil Ltd (Cambridge, United Kingdom). Ultra-pure water was generated by a Millipore Elix 5 RO system and Millipore Advantage Milli-Q system (Millipore SAS, Molsheim, France).

UPLC-MS Method Development and Optimization

The first aspect considered was the type of samples to be analyzed. Being blood-derived, one would expect polar to intermediate polarity compounds to be predominantly present in sera. Due to the nature of the samples, a C18 UPLC column served as a good starting point but a C18 column with polar embedded functionality or special end capping could enhance the separation of blood sample extracts. To enhance the possibility of separating all compounds ionized in the source of the mass spectrometer, the maximum column length of 150 mm was used for all column evaluations. The columns evaluated were the BEH C18 column as well as the T3 C15 column from Waters. Although the BEH C18 column chemistry gave good retention of the more polar compounds, it could not equal the

superior retention and resolution observed on the T3 C15 column. The T3 150mm UPLC column was therefore used in all further experimental work.

The second aspect was to optimize the ionization settings of the mass spectrometer. Quality control (QC) samples which comprised a pool of all samples used in the study [36] or a pool of representative samples from each condition being investigated (e.g. HIV negative, HIV+ and HIV+HAART+; [35]) were prepared. The QC sample was infused into the inlet probe while scanning the TOF analyzer in both ESIPos and ESINeg ionization modes. Mass ions were observed in both modes and resulted in the optimization of both modes in regards to capillary voltage, cone voltage and extractor voltage. The trap voltage was set at 4 V to ensure that no unnecessary fragmentation would occur during ion transmission. The settings as detailed in the experimental section were found to provide the best ionization of the most compounds in both polarity modes and were used in all experimental work.

The third aspect evaluated was the stability of the method. QC samples were prepared and used to monitor the stability of sample runs during UPLC-MS analysis. A pooled sample was analyzed by doing ten consecutive injections on the same day and repeating the same analysis on the next day and comparing the two data sets. It was found that the UPLC column required three conditioning injections prior to starting the main analysis run. The results for the inter- day and intra-day experiments compared very well and ensured that extended sample runs would be possible even if the whole analysis run would span over several days. To minimize the effect of subtle changes in the separation over long analysis runs, all the samples were randomized in the sample table prior to analysis. Blanks were included to assess carryover and exclude possible contaminating ion features.

Optimized Metabolite Extraction from Sera using Ostro™ Plates

Ostro™ sample preparation plates (Waters Corporation, Milford, USA) and filtration through the Waters positive pressure-96 processor (Waters Corporation, Milford, USA) was utilized to facilitate serum metabolite extraction. Serum samples were thawed on ice prior to extraction to avoid loss of volatiles. Briefly, the Ostro™ plate was keyed onto a 96-well collection plate. One hundred microliters of sera, blanks and QC samples were randomly loaded onto the Ostro™ plate in duplicate, with a few exceptions as explained below which subsequently influenced the number of cases analyzed in electrospray ionization positive (ESI+) and negative (ESI-) mode respectively. Three hundred microliters of freshly prepared 1 % formic acid in acetonitrile was forcefully added to the plate wells and mixed through

aspiration. The plate assembly containing sample and solvent was filtered under positive pressure at 414 kPa (60 psi). The collected filtrates in the collection plate were transferred on ice to the UPLC-MS facility for subsequent analysis.

The top-half of the 96-well collection plate was analyzed in ESI+ mode and the bottom half in ESI- mode. Thirteen **HIV+** samples were loaded onto the top half of the plate. Due to limited sample volumes, only 11 cases were duplicated in the lower half of the plate for analysis in ESI- mode. Fifteen **HIV negative** samples were loaded onto the top half of the plate and duplicated in the bottom half too. Due to insufficient volumes of HIV+ samples, two wells in the bottom half of the plate were vacant and an additional HIV negative sample loaded, in duplicate. The total number of HIV negative samples was therefore 16, with 15 cases analyzed in the ESI+ mode and 17 (16^{th} sample duplicated) analyzed in ESI- mode. Fifteen **HIV+HAART+** samples were analyzed in ESI+ and ESI- mode respectively. Due to the exclusion of an outlier only 14 HIV+HAART+ cases were considered for further data processing in ESI+ mode.

UPLC-TOF MS Analysis

All chromatographic and mass spectrometric analysis was performed at the Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa. A Waters Acquity UPLC coupled in tandem to a Waters photodiode array (PDA) detector and a SYNAPT G1 HDMS mass spectrometer was used to generate accurate mass data. Chromatographic separation was done utilizing a Waters HSS T3 column (150 mm x 2.1 mm, 1.8 µm) thermostated at 60 °C. A binary solvent mixture was used consisting of water (Eluent A) containing 10 mM formic acid (pH 2.3) and acetonitrile (Eluent B). Gradient elution was performed at a flow rate of 0.4 $ml.min^{-1}$. The initial conditions were 95% of A for 0.1 minute with a gradient profile 7 to 40% of A at 8 minutes, followed by a gradient profile 5 to 10% of A at 14 minutes and kept constant for 1 minute. The system was re-equilibrated using the initial conditions with a total runtime of 20 minutes. The injection volume was set at 3 µl with triplicate injections acquired per sample. The PDA detector was scanned between 200 and 500 nm (1.2 nm resolution) and collected 20 spectra per second.

The SYNAPT G1 mass spectrometer was used in V-optics and operated in electrospray mode. Leucine enkephalin (50 pg/mL) was used as reference calibrant to obtain typical mass accuracies between 1 and 3 mDalton. Although positive ionization is commonly used and more compounds are ionized in this state [37], the mass spectrometer was operated

in both negative and positive mode of electrospray ionization to enable greater coverage of the metabolome for untargeted analysis. The capillary voltage was set at 2.0 kV, the sampling cone at 30 V and the extraction cone at 4 V. The micro channel plate detector voltage was set at 1950 V. The scan time was 0.1 seconds covering the 100 to 1000 Dalton (Da) mass range. The source temperature was 120 °C and the desolvation temperature set at 450 °C. Nitrogen was used as the nebulization gas at a flow rate of 800 L/h. MassLynx V4.1 (SCN 704) software was used to control the hyphenated system and for data manipulation.

Data Processing and Statistical Analysis

Data mining was done by processing the raw ESI- and ESI+ data files with the MarkerLynx Extended Statistics (XS) processing method of the MassLynx software under the following criteria i.e. a retention window of 0.5-17.00 min, a mass range and tolerance of 100-1000 and 0.01 Da respectively. During this time the MarkerLynx XS V4.1, SCN 704 software performed peak integration and alignment as well as background noise subtraction yielding data in a suitable format for statistical analysis.

An additional peak threshold and noise filter was applied to the data within MarkerLynx. The data was exported into Excel® and further analyzed using IBM SPSS (version 19.0). Multivariate statistical analysis in the form of linear discriminant analysis was performed to classify experimental groups and identify metabolites contributing to group variances. LDA is a statistical method that uses a linear combination of variables to maximize between-class variation compared to within-class variation, to bring about an improved separation of experimental groups. Principal component analysis (PCA) was also performed (see supplementary information, Figure S1) but LDA was found to be more appropriate in this case. The method used in the case of PCA was to compute principal components and then to do an LDA on the principal components. This was done because the method is commonly seen in the literature. In the case of LDA based on PCA, all the metabolites (n=1976 for ESI neg mode and n=2511 for ESI+ mode) are involved in the definition of each principal component. In the case of the LDA based on individual metabolites, stepwise discriminant analysis was used. Eleven and 10 metabolites selected by the stepwise procedure produced a 100% correct classification for ESI neg and ESI+ modes respectively. The model was not corrected for confounding factors, an aspect addressed under the discussion. The significance of these molecules were also verified through univariate analysis in the form of analysis of variance (ANOVA).

Metabolite Identification

Mass spectral searches conducted using metabolite and/or structure databases and libraries [BioCyc, ChemSpider, National Institute of Standards and Technology (NIST), Human Metabolome Database (HMDB), Golm, Lipid Maps, Molecular Modeling Database (MMDB), *etc*] were used to confirm the identities of significantly altered metabolites.The molecular formula of the analytes were obtained using MassLynx software. Searches were restricted to C, H, N, O, S elementary composition.

Ethics Approval

Ethics approval for the use of human blood in this research was obtained from the Faculties of Natural and Agricultural Sciences and Health Sciences of the University of Pretoria with protocol numbers EC 080506-19 and 163/2008, respectively.

RESULTS

Experimental Design and Sample Selection

To study metabolic complications, metabolites are often extracted from affected tissues and biofluids using laborious protocols which incorporate numerous handling steps (Figure 1). As a result, variability is introduced into the data [38]. Utilizing a straightforward, easy to use extraction protocol with minimal handling steps should minimize the chances of obtaining erroneous data. A one-step protein precipitation and extraction protocol is presented here for use in UPLC-MS metabonomics experiments aimed at evaluating HIV and HAART-induced metabolic changes. Metabonomics using Ostro^{TM} plates and UPLC-MS is also presented to show the unique detection of treatment-associated metabolites in infected biofluid.

The samples used in this study were well-matched in terms of gender and CD4 counts (Table 1). In the case of the treatment-naive HIV+ group, CD4 counts were relatively high and representative of an asymptomatic cohort. The HIV-infected cases were generally older and presented with varying viral loads. The possible influences of these confounding factors are elaborated on in the discussion.

Metabonomic Profiling

UPLC-MS base peak ion (BPI) chromatograms of HIV negative, HIV+ and HIV+HAART+ sera are shown in Figure 2. Distinct, narrow chromatographic peaks suggest

the successful detection of metabolites as well as the removal of phospholipids which previously masked the detection of metabolites when another extraction protocol was followed (see supplementary Figure S2). The chromatograms show the sera of the three groups exhibiting different metabolic profiles, notable by differences in their peak numbers and intensities (Figure 2, also refer to representative circles). It is also evident (Figure 2 [ii] and [iii]) that the metabolism of the uninfected participants and asymptomatic HIV patients are visually more similar and perhaps indicative of the health status of the infected patients. The metabolism of the patients receiving HAART was most affected as seen by the increase in the intensitiy of the peaks at 6.65 and 12.95 minutes respectively (Figure 2 [i]).

Following integration and alignment of the MS data, LDA was applied to classify the groups as HIV negative, HIV+ and HIV+HAART+ and also to identify variables accounting for group differences. This statistical tool classified all cases correctly achieving 100% accuracy (Table 2) pre- and post cross-validation. The number of participants was too small for a full cross-validation by means of a hold-out sample. Instead, the method applied was leave-one-out-at-a-time. Metabonomics data is multivariate and should ideally be analyzed using multivariate statistical approaches to increase the power of the analysis and avoid erroraneous conclusions [39]. ANOVA, applied to determine how well this statistics tool complemented mutlivariate stastics findings, detected many more metabolites that were significantly different between the groups than did LDA. Of the 21 important metabolites identified through LDA, 13 (62%) had p values < 0.05 (see Table 3 [i] and [ii]) indicating that ANOVA, in terms of the number of significantly altered molecules identified, complimented LDA findings to an extent.

Scatter plots (Figure 3) show maximum between-group separation, suggesting distinct metabolic profiles for each group. The metabolism of the treatment-experienced group was most affected as seen by how far this group is positioned from the HIV negative and HIV+ groups (Figure 3) with the greatest separation achieved along the first discriminant. In Figure 2 there are visual similarities between the metabolism of the HIV negative and HIV+ group which indicate that the infected individuals to still be relatively healthy. This was further amplified in the scatter plots (Figure 3, along the first discriminant). If group separation is evaluated along the second discriminant, it appears that the metabolic profile of the HIV+HAART+ group shifts toward that of the HIV negative group. Within-group metabolic differences were minimal and representative of a homogeneous group of

samples. Reliability and validity of the MS runs were also confirmed through the tight clustering of QC samples and tight clustering of technical repeats (finally averaged to allow for LDA).

Following LDA, important variables (n=21) contributing to the clustering of the experimental groups in ESI negative and ESI+ mode were identified (Table 3 [i] and [ii]). Serum contains 90% water and mainly has polar metabolites [40]. In this study, each ionization mode resulted in the detection of unique metabolites. In the ESI- mode (Table 3 [i]) metabolites involved in lipid and/or membrane biosynthesis as well as nucleotide metabolism were mainly identified. Metabolites related to the treatment of HIV-associated conditions as well as amino acid and protein metabolism were mainly identified in ESI+ mode (Table 3[ii]). If the assumption is that the metabolite levels in uninfected sera are normal then the 21 distintinguishing metabolites showed either infection or infection and HAART as being responsible for the shifts/changes in the levels of these molecules (Figure 4).

DISCUSSION

Experimental Design and Sample Selection

During random sample selection, the influence of possible confounding factors must be considered, because these factors may account for metabolic differences and influence the clustering of samples. For our dataset, age, gender and the varying viral loads of patients did not appear to be confounding since samples grouped according to these parameters failed to cluster together. When the possible confounder, gender, was included as a variable in the LDA model (see supplementary Figure S3), it did not affect the data. Similarly, when age was applied to the model, it did not play a significant role in the results. In a non-HIV biofluid metabonomics study, Tulipani *et al* (2013) [31] concluded that solvent extraction in combination with Ostro^{TM} plates, as used in our study too, allowed for subtle metabolic changes between groups, independent of confounding factors, to be observed. A truly randomized sample set should have included the same age across experimental groups. However, being at the mercy of the volunteer approach, we were limited to using only patients who agreed to involvement and it turns out that younger individuals totally sure of their status only volunteered when negative. As mentioned, the LDA model not responding to gender or age gave the confidence to continue the study. In addition, the fact that infected

individuals with overlapping ages (Table 2, mean age of HIV+ and HIV+HAART groups) separated based on metabolite identities strongly supports the validity of the collected data.

Metabonomics Profiling, Perturbed Metabolites and their Biological Significance

The body largely depends on oxidation processes for its energy supply. This occurs mainly through the breakdown of fats, carbohydrates and amino acids in mitochondria and peroxisomes. HIV, the immune response to infection and treatment disrupts host metabolism. In our findings, treated patients were most differentiated from the HIV negative and HIV+ groups (Figure 2 and Figure 3) suggesting HAART to drastically affect host metabolism. The alteration of the host metabolism by the HIV therapeutics is in agreement with the literature [2, 8]. There was also a shift of the HIV+HAART+ group toward the HIV negative group (Figure 3, along the 2nd discriminant). Whether or not the shift implies successful therapeutic intervention needs to be confirmed by specific analysis. It is important to note that Figure 2 provides representative chromatograms for each experimental group while Figure 3 shows responses of every individual used in the study. The data presented here indicates that HIV and/or HAART interferes mainly with lipid, carbohydrate and therefore energy metabolism which ultimately translates to alterations in cell signalling and other membrane associated reactions. In addition, metabolites associated with the treatment of HIV-associated conditions were also observed. HIV-induced disruption of fatty acid/lipid metabolism which is usually associated with lipodystrophy (fat redistribution) is well-documented in the literature [8, 11-14] as is the disruption of carbohydrate and energy metabolism [22, 41-43].

To get a better idea of the contribution of HIV and/or treatment to the measured metabolic alterations, changes in the levels of the respective metabolites were evaluated through mean plots (Figure 4). Of the 21 metabolites identified through LDA as contributing to group differences; 3 (14.3%) were altered due to HIV, 5 (23.8%) were altered due to HAART while 13 (61.9%) were altered as a result of infection and treatment. The plots confirmed infection and treatment to augment metabolic alterations and also reflects the strong influence of HAART on the metabolic profile of the cohort under investigation. Upon reviewing patient records, it was found that of the HAART combinations given to patients, a large percentage of individuals were on combinations of nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs, Table 1). These regimens are known for causing mitochondrial dysfunction, lipodystrophy and lactic acidosis [44-46]. HIV-infected cells and lipolysis, which is a common feature of infected

individuals, require increased amounts of energy and therefore functional mitochondria for this purpose. Studies have shown early on that HIV-infected cells have higher rates of resting energy expenditure [41, 42] which is in keeping with the high energy demands of infected cells. Mitochondria drive metabolic processes [47] and together with the Krebs cycle serve as the "hub" of cellular metabolism. The human immunodeficiency virus obtains energy mainly by altering host metabolic processes by impacting for example, mitochondria [43]. If compromised, as occurs during HIV infection, energy processes are affected such that adenosine triphosphate (ATP) is depleted and glycolysis activated [48]. This disruption in energy metabolism is evident by the alteration in carbohydrate metabolism detected in our study. Because of the role of this organelle in producing ATP, energy metabolism is ultimately affected. Several metabolites, for example, cytidine triphosphate (Table 3 [i]) and 1-Hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycerol 3-diphosphate (Table 3 [ii]) were linked to changes in energy metabolism (glycolysis, energy storage and use) in this study. Hollenbaugh *et al* (2011) [22] too showed an alteration in glycolysis following the *in vitro* HIV infection of primary CD4 cells and a long-lived macrophage cell line. The increase in glycolysis ultimately results in the increased production of lactic acid, which led to lactic acidosis, another common feature associated with NRTI use [46]. Synonomous with mitochondrial dysfunction, apoptosis inducing factors and proteolytic caspases are released causing cellular damage. Lipolysis is also decreased contributing to the developmnet of lipodystrophy [44]. The use of NRTIs and NNRTIs is thus reflected in our study through changes in lipid, carbohydrate and energy metabolism. The treatment-induced metabolic changes observed here are usually only observed as severe clinical manifestations (very visible/unsightly fat deposits) but was detected here by LC-MS metabonomics in the absence of such clinical changes. This is largely attributed to the sensitiviy of the MS instrument. Detecting these subtle changes which may not yet be visible to the eye, provides reason to further develop metabonomics for disease diagnosis, monitoring adherence and linking specific drug classes to particular metabolic complications. Alterations in fatty acid/lipid metabolism are known to place the affected individuals at risk of cardiovascular diseases making these metabonomics findings highly relevant for the management of HIV/AIDS.

A closer evaluation of the mean plots in Figure 4 could add to our understanding of the mechanisms of HIV and/or HAART-induced metabolic complications. Since most of the metabolic changes involved lipid changes and were as a result of HAART as well as the synergy between infection and treatment, the changes may suggest a mechanism used by the

virus to ensure its survival depending on whether the levels of these molecules are increased or decreased. For example, the virus depends on the host's metabolic machinery to survive. It can therefore induce an increase in metabolites involved in lipid and membrane biosynthesis (e.g. V0274MZ213.9635 in Figure 4 [i], identified as Methyl dodecanoate in Table 3 [i]) as well as cell formation ensuring it has a host in which to replicate. Alternatively, the increase in cell turnover may be a host response to infection in order to produce more cells to ward off the virus. A decrease in the levels of this metabolite following the administration of HAART suggests an opposing effect by treatment to restore the level of the metabolite to that of uninfected controls. Whether the shift implies successful therapeutic intervention would require further investigation. Similarly, the remaining plots can be evaluated in this manner but quantitative analysis would be more informative.

Metabolites Indicative of Treatment

Of the metabolites detected, there were also indications of changes in nucleotide, amino acid and protein metabolism. Nucleotides serve as the building blocks for deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) synthesis. That nucleotide metabolism is affected in the host is in agreement with the life cycle of the virus i.e. HIV integrating its genetic information into that of the host. Nucleotides serve as energy carriers, cell signalling molecules and cofactors to various reactions. A disruption in the metabolism of these molecules thus adds to the disruption in cell signalling already induced by HIV and/or its treatment. Nucleotide metabolism in the context of HIV infection is important because HIV cannot encode nucleotide-metabolizing enzymes and it's intracellular replication requires a continous supply of deoxynucleotidetriphosphate (dNTPs, [49]).

Proteins serve important biological roles including catalyzing metabolic reactions, assisting with DNA replication, initiating responses to stimuli and transporting molecules from one location to another. An alteration to amino acid and protein metabolism coincides with the change in nucleotide metabolism which is required for the replication of genetic information. It also confirms a disruption in cell signaling processes highlighted previously to occurr as a result of changes in lipid and membrane biosynthetic processes.

Anemia is a documented complication of HIV infection associated with decreased patient survival [50]. We identified metabolites potentially related to the treatment of anemia and sickle cell disease suggesting the occurrence of these conditions in patients. As shown in table 3, we detect for example, Leucovorin [metabolite 4 in table 3(i)], which is used in the

treatment of anemias and hematological complications. Anemia has been identified as a frequent complication of HIV infection and hemoglobin was also shown by Obirikorang and Yeboah, 2009 [50] to be a predictor for the progression of HIV/AIDS. Leucovorin was detected in the ESI- mode. After tracing the variable (in this case V1367MZ458.1292) in the dataset using individual metabolite trend plots prepared in MassLynx v4.1 software, this variable was found to be mostly present in treated patients allowing us to infer the presence of a specific condition and its treatment in those patients. Looking at the particular patient's files where the variable was present, we were able to confirm that the patient was receiving treatment for anemia. We detect breakdown products of hemoglobin which serve as markers for urinary tract function. Hemoglobin breakdown products such as globin are known to be converted to amino acids which participate in metabolism or the formation of new proteins. It is quite interesting to observe how metabolites of blood-related conditions coincide with protein metabolism changes noted in the preceding paragraph.

Psoriasis is a skin condition associated amongst others, with HIV infection. This condition is associated with cardiovascular disease, a common complication of HIV-infected and treated individuals [51]. In this study we detect DHNA, an anthraquinone used to treat psoriasis. Other therapeutics given to the HIV+ individuals of this study include for example antifungal agents such as clotrimazole. Immunocompromised patients are burdened with opportunistic infections making the detection of metabolites which are linked to these types of therapeutics valid. We also detected metabolites suggesting the use of anti-inflammatory, anti-pyretic and anti-rheumatic agents used to treat common HIV-related conditions. HIV-infected individuals are known to be under oxidative stress [52]. We detected 5'deoxyadenosine which is a marker of oxidative damage. Taxifolin and chlorogenic acid have antioxidant properties and were also detected implying the presence of over the counter anti-oxidants because they are plant derived compounds that could be green tea. Because EGCg was detected in patients, but not caffeine, the source might be decaffeinated tea extract and some yew tree extract. In a gas chromatography mass spectrometry (GC-MS) metabonomics study investigating HIV-induced mitochondrial dysfunction in asymptomatic HIV-infected patients, Williams *et al* (2012) [3] detected two metabolites linked to "self-treatment". GC-MS detects volatile compounds [53]. Most metabolites in nature are polar and non-volatile [53] as is the metabolites of sera [40]. Because GC requires samples to be volatile, the process of volatilization could have contributed to the failure to detect more "self treatments" by the technique. Inferring the treatment of HIV-related complications was made

here after tracking the specific metabolite back to the actual sample in the dataset and reviewing the patient's status i.e. HIV negative, HIV+ and/or HIV+HAART+. Where treatment related metabolites were detected, the patients were treatment-experienced. Metabonomics therfore has the potential to provide real phenotypic patient information.

These findings imply that SPE in the form of Ostro^{TM} plates is recommended for biofluid metabolome extraction because it eliminates large molecules (proteins and phospholipids) that have been shown to mask metabolites that are analyzed using UPLC-MS. Ultra performance LC-MS proved ideal for biofluid metabonomics because of its increased sensitivity and resolution to detect subtle metabolic disturbances. Based on data presented here the approach could find use in biomarker identification and therefore disease diagnosis and prognosis as well as HIV/AIDS management i.e. monitoring adherence to treatment, treatment success/failure and linking specific drugs to specific metabolic complications.

CONCLUSION

Shown here for the first time is an easy SPE protocol which allowed for the collection of reproducible, reliable UPLC-MS data which in turn distinguished HIV-influenced compared to therapy-induced metabolic changes. UPLC-MS detected changes in lipid and energy metabolism in negative ionization mode whilst metabolites indicative of the treatment of HIV-associated conditions were mainly evident in positive ionization mode. The biological role of the metabolites causing group differentiation coincides well with that reported in the literature. The mean plots already gave an indication of an increase or decrease in metabolite levels while chemometrics confirmed the visual observations. The quantification of metabolite changes as a necessary next step will be simplified by either a targeted analysis or a decision on the specific metabolite or groups of metabolites to investigate in order to identify useful standards and/or molecules for sample spiking. Quantification should contribute to the clarification of the exact HIV-related mechanisms involved in metabolite increases or decreases. This study presented an easy sample preparation protocol designed for minimal sample handling that favoured quality metabonomics data collection, which is important for clinically relevant biomarker identification. The end use of mass spectrometric metabonomics should eventually be a clinical application, detecting biomarkers of value in diagnostics, prognostics or disease management. The detection of metabolites related to therapy implies a strong role for MS-metabonomics in HIV/AIDS management by monitoring the success/failure of specific treatment regimens, patient adherence and linking

specific treatment regimens to particular metabolic complications. The ability to infer additional information from the data suggests an added advantage to metabonomics, in particular the ability to screen randomly and then go back to the sample group and determine whether the conclusions/inferences correlate to the patient's reality.

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
ATP	Adenosine Triphosphate
BPI	Base Peak Ion
CSIR	Council for Scientific and Industrial Research
C	Carbon
Da	Dalton
DHNA	1,4-Dihydroxy-2-naphthoic acid
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
ESI-	Electrospray ionization, negative mode
ESI+	Electrospray ionization, positive mode
H	Hydrogen
HIV	Human Immunodeficiency Virus
HIV+	HIV positive
HAART	Highly Active Antiretroviral Therapy
HMDB	Human Metabolome Database
LDA	Linear Discriminant Analysis
MMDB	Molecular Modeling Database
MRC	Medical Research Council
MS	Mass Spectrometry
MSI	Metabolomics Standards Initiative
N	Nitrogen

NIST	National Institute of Standards and Technology
NRF	National Research Foundation
NRTIs	Nucleoside/Nucleotide Reverse Transcriptase Inhibitors
NNRTIs	Non- Nucleoside/Nucleotide Reverse Transcriptase Inhibitors
O	Oxygen
PCA	Principal component analysis
PDA	Photodiode Array
QC	Quality Control
RNA	Ribonucleic acid
S	Sulfur
SPE	Solid Phase Extraction
TIA	Technology Innovation Agency
UPLC-MS	Ultra Performance Liquid Chromatography Mass Spectrometry
XS	Extended Statistics

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by grants from the Medical Research Council (MRC), Technology Innovation Agency (TIA) and the National Research Foundation (NRF) of South Africa.

Aurelia Williams performed the study, analyzed and interpreted the data as well as wrote the article. Khanyisile Kgoadi performed the study, analyzed and interpreted the data as well as wrote the article. Francois Steffens analyzed the data and wrote the article. Paul

Steenkamp collected and analyzed data as well as wrote the article. Debra Meyer was responsible for the design of the project, interpreted the data and wrote the article. All individuals listed as authors therefore contributed substantially to the design, performance, analysis, and reporting of the work.

REFERENCES

[1] Couture, P. HIV & Metabolism:What Are the Complications? *The Canadian Journal of CME.*, **2005**, 65-67.

[2] Brennan-Benson, P. Metabolic and cardiovascular complications of HIV. *Medicine.*, **2009**, *37*, 348-351.

[3] Williams, A.; Koekemoer, G.; Lindeque, Z.; Reinecke, C.; Meyer, D. Qualitative serum organic acid profiles of HIV-infected individuals not on antiretroviral treatment. *Metabolomics.*, **2012**, 8, 804-818.

[4] Beisel, W. R. Interrelated changes in host metabolism during generalized infectious illness. *Am. J. Clin. Nutr.*, **1972**, *25*, 1254-1260.

[5] Wolowczuk, I.; Verwaerde, C.; Viltart, O.; Delanoye, A.; Delacre, M.; Pot, B.; Grangette, C. Feeding our immune system: impact on metabolism. *Clin. Dev. Immunol.*, **2008**, *2008*, 639803.

[6] Mathis, D.; Shoelson, S. E. Immunometabolism: an emerging frontier. *Nat. Rev. Immunol.*, **2011**, *11*, 81-83.

[7] Hewer, R.; Vorster, J.; Steffens, F. E.; Meyer, D. Applying biofluid 1H NMR-based metabonomic techniques to distinguish between HIV-1 positive/AIDS patients on antiretroviral treatment and HIV-1 negative individuals. *J. Pharm. Biomed. Anal.*, **2006**, *41*, 1442-1446.

[8] Jevtović, D.; Dragović, G.; Salemović, D.; Ranin, J.; Djurković-Djaković, O. The metabolic syndrome, an epidemic among HIV-infected patients on HAART. *Biomedicine & Pharmacotherapy.*, **2009**, *63*, 337-342.

[9] Powderly, W. Metabolic Complications Associated with HIV Infection and Antiretroviral Therapy. *AIDS Patient Care STDS.*, **2004**, *18*, 431-435.

[10] Slama, L.; Le Camus, C.; Serfaty, L.; Pialoux, G.; Capeau, J.; Gharakhanian, S. Metabolic disorders and chronic viral disease: The case of HIV and HCV. *Diabetes Metab.*, **2009**, *35*, 1-11.

[11] Alberti, K. G.; Zimmet, P.; Shaw, J. Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabet. Med.*, **2006**, *23*, 469-480.

[12] Pascal, S.; Resnick, L.; Barker, W. W.; Loewenstein, D.; Yoshii, F.; Chang, J.; Boothe, T.; Sheldon, J.; Duara, R. Metabolic Asymmetries in Asymptomatic HIV-1

Seropositive Subjects: Relationship to Disease Onset and MRI Findings. *J. Nucl. Med.*, **1991**, *32*, 1725-1729.

[13] Grunfeld, C.; Kotler, D. P.; Hamadeh, R.; Tierney, A.; Wang, J.; Pierson, R. N. Hypertriglyceridemia in the acquired immunodeficiency syndrome. *Am. J. Med.*, **1989**, *86*, 27-31.

[14] Grunfeld, C.; Pang, M.; Doerrler, W.; Shigenaga, J. K.; Jensen, P.; Feingold, K. R. Lipids, lipoproteins, triglyceride clearance, and cytokines in human immunodeficiency virus infection and the acquired immunodeficiency syndrome. *Journal of Clinical Endocrinology & Metabolism.*, **1992**, *74*, 1045-1052.

[15] Sitole, L. J.; Williams, A. A.; Meyer, D. Metabonomic analysis of HIV-infected biofluids. *Mol. Biosyst.*, **2013**, *9*, 18-28.

[16] Villas-Boas, S.G.; Roessner, U.; Hansen, M.A.E.; Smadsgaard, J.; Nielson, J. Metabolomics in Humans and other mammals. In: Metabolome Analysis: An introduction.; John Wiley & Sons, **2007**; 253-288.

[17] Zhang, M.; Versalovic, J. HIV update. Diagnostic tests and markers of disease progression and response to therapy. *Am. J. Clin. Pathol.*, **2002**, *118 Suppl*, S26-32.

[18] Kanekar, A. Biomarkers predicting progression of human immunodeficiency virus-related disease. *J. Clin. Med. Res.*, **2010**, *2*, 55-61.

[19] Beger, R.; Colatsky, T. Metabolomics data and the biomarker qualification process. *Metabolomics.*, **2012**, *8*, 2-7.

[20] Keane, N.; John, M. Biomarkers of natural and vaccine immunity against HIV. *Biomark Med.*, **2011**, *5*, 113-116.

[21] Ghannoum, M. A.; Mukherjee, P. K.; Jurevic, R. J.; Retuerto, M.; Brown, R. E.; Sikaroodi, M.; Webster-Cyriaque, J.; Gillevet, P. M. Metabolomics Reveals Differential Levels of Oral Metabolites in HIV-Infected Patients: Toward Novel Diagnostic Targets. *OMICS: A Journal of Integrative Biology.*, **2011**, 17, 5-15.

[22] Hollenbaugh, J. A.; Munger, J.; Kim, B. Metabolite profiles of human immunodeficiency virus infected CD4+ T cells and macrophages using LC–MS/MS analysis. *Virology.*, **2011**, *415*, 153-159.

[23] Schwartz, S. J. How Can the Metabolomic Response to Lycopene (Exposures, Durations, Intracellular Concentrations) in Humans Be Adequately Evaluated? *The Journal of Nutrition.*, **2005**, *135*, 2040S-2041S.

[24] Nováková, L.; Solichová, D.; Solich, P. Advantages of ultra performance liquid chromatography over high-performance liquid chromatography: Comparison of

different analytical approaches during analysis of diclofenac gel. *Journal of Separation Science.*, **2006**, *29*, 2433-2443.

[25] Zelena, E.; Dunn, W. B.; Broadhurst, D.; Francis-McIntyre, S.; Carroll, K. M.; Begley, P.; Oâ€™Hagan, S.; Knowles, J. D.; Halsall, A.; Wilson, I. D.; Kell, D. B. Development of a Robust and Repeatable UPLCâ˜'MS Method for the Long-Term Metabolomic Study of Human Serum. *Anal. Chem.*, **2009**, *81*, 1357-1364.

[26] Fiehn, O.; Robertson, D.; Griffin, J.; van, d. W.; Nikolau, B.; Morrison, N.; Sumner, L.; Goodacre, R.; Hardy, N.; Taylor, C.; Fostel, J.; Kristal, B.; Kaddurah-Daouk, R.; Mendes, P.; van Ommen, B.; Lindon, J.; Sansone, S. The metabolomics standards initiative (MSI). *Metabolomics.*, **2007**, *3*, 175-178.

[27] Allanson, J. P.; Biddlecombe, R. A.; Jones, A. E.; Pleasance, S. The Use of Automated Solid Phase Extraction in the '96 well' Format for High Throughput Bioanalysis using Liquid Chromatography Coupled to Tandem Mass Spectrometry. *Rapid Communications in Mass Spectrometry.*, **1996**, *10*, 811-816.

[28] Simpson, H.; Berthemy, A.; Buhrman, D.; Burton, R.; Newton, J.; Kealy, M.; Wells, D.; Wu, D. High throughput liquid chromatography/mass spectrometry bioanalysis using 96-well disk solid phase extraction plate for the sample preparation. *Rapid Commun. Mass Spectrom.*, **1998**, *12*, 75-82.

[29] Zimmer, D.; Pickard, V.; Czembor, W.; Muller, C. Comparison of turbulent-flow chromatography with automated solid-phase extraction in 96-well plates and liquid-liquid extraction used as plasma sample preparation techniques for liquid chromatography-tandem mass spectrometry. *J. Chromatogr., A* **1999**, *854*, 23-35.

[30] Michopoulos, F.; Lai, L.; Gika, H.; Theodoridis, G.; Wilson, I. UPLC-MS-based analysis of human plasma for metabonomics using solvent precipitation or solid phase extraction. *J. Proteome Res.*, **2009**, *8*, 2114-2121.

[31] Tulipani, S.; Llorach, R.; Urpi-Sarda, M.; Andres-Lacueva, C. Comparative analysis of sample preparation methods to handle the complexity of the blood fluid metabolome: when less is more. *Anal. Chem.*, **2013**, *85*, 341-348.

[32] Rose, M. J.; Merschman, S. A.; Eisenhandler, R.; Woolf, E. J.; Yeh, K. C.; Lin, L.; Fang, W.; Hsieh, J.; Braun, M. P.; Gatto, G. J.; Matuszewski, B. K. High-throughput simultaneous determination of the HIV protease inhibitors indinavir and L-756423 in human plasma using semi-automated 96-well solid phase extraction and LC-MS/MS. *J. Pharm. Biomed. Anal.*, **2000**, *24*, 291-305.

[33] Schuster, A.; Burzawa, S.; Jemal, M.; Loizillon, E.; Couerbe, P.; Whigan, D. Quantitative determination of the HIV protease inhibitor atazanavir (BMS-232632) in human plasma by liquid chromatography-tandem mass spectrometry following automated solid-phase extraction. *J. Chromatogr. B. Analyt Technol. Biomed. Life. Sci.*, **2003**, *788*, 377-386.

[34] Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K. Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protocols.*, **2010**, *5*, 1005-1018.

[35] Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protocols.*, **2011**, *6*, 1060-1083.

[36] Bijlsma, S.; Bobeldijk, I.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; van Ommen, B.; Smilde, A. K. Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. *Anal. Chem.*, **2006**, *78*, 567-574.

[37] Lin, L.; Yu, Q.; Yan, X.; Hang, W.; Zheng, J.; Xing, J.; Huang, B. Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabonomics? A serum metabonomic study of kidney cancer. *Analyst.*, **2010**, *135*, 2970-2978.

[38] Luque-Garcia, J.; Neubert, T. A. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. *Journal of Chromatography A; Advances in Sample Preparation Part II.*, **2007**, *1153*, 259-276.

[39] Want, E.; Masson, P. Processing and analysis of GC/LC-MS-based metabolomics data. *Methods Mol. Biol.*, **2011**, *708*, 277-298.

[40] Lin, L.; Huang, Z.; Gao, Y.; Yan, X.; Xing, J.; Hang, W. LC-MS based serum metabonomic analysis for renal cell carcinoma diagnosis, staging, and biomarker discovery. *J. Proteome Res.*, **2011**, *10*, 1396-1405.

[41] Hommes, M. J. T.; Romijn, J. A.; Godfried, M. H.; Eeftinck Schattenkerk, J. K. M.; Buurman, W. A.; Endert, E.; Sauerwein, H. P. Increased resting energy expenditure in human immunodeficiency virus-infected men. *Metab. Clin. Exp.*, **1990**, *39*, 1186-1190.

[42] Hommes, M.; Romijn, J.; Endert, E.; Sauerwein, H. Resting energy expenditure and substrate oxidation in human immunodeficiency virus (HIV)-infected asymptomatic

men: HIV affects host metabolism in the early asymptomatic stage. *The American Journal of Clinical Nutrition.*, **1991**, *54*, 311-315.

[43] Macho, A.; Castedo, M.; Marchetti, P.; Aguilar, J. J.; Decaudin, D.; Zamzami, N.; Girard, P. M.; Uriel, J.; Kroemer, G. Mitochondrial dysfunctions in circulating T lymphocytes from human immunodeficiency virus-1 carriers. *Blood.*, **1995**, *86*, 2481-2487.

[44] Kakuda, T. N.; Brundage, R. C.; Anderson, P. L.; Fletcher, C. V. Nucleoside reverse transcriptase inhibitor-induced mitochondrial toxicity as an etiology for lipodystrophy. *AIDS.*, **1999**, *13*, 2311-2312.

[45] Kakuda, T. N. Pharmacology of nucleoside and nucleotide reverse transcriptase inhibitor-induced mitochondrial toxicity. *Clin. Ther.*, **2000**, *22*, 685-708.

[46] Sweet, D.E. Perspective Metabolic Complications of Antiretroviral Therapy. International AIDS Society: *Topics in HIV Medicine.*, **2005**, 13, 70-74.

[47] Shedlock, D. J.; Hwang, D.; Choo, A. Y.; Chung, C. W.; Muthumani, K.; Weiner, D. B. HIV-1 viral genes and mitochondrial apoptosis. *Apoptosis.*, **2008**, *13*, 1088-1099.

[48] Hofhaus, G.; Johns, D. R.; Hurko, O.; Attardi, G.; Chomyn, A. Respiration and growth defects in transmitochondrial cell lines carrying the 11778 mutation associated with Leber's hereditary optic neuropathy. *J Biol Chem.*, **1996**, 271, 13155–13161.

[49] Balzarini, J. Effect of antimetabolite drugs of nucleotide metabolism on the anti-human immunodeficiency virus activity of nucleoside reverse transcriptase inhibitors. *Pharmacol. Ther.*, **2000**, *87*, 175-187.

[50] Obirikorang, C.; Yeboah, F. Blood haemoglobin measurement as a predictive indicator for the progression of HIV/AIDS in resource-limited setting. *J. Biomed. Sci.*, **2009**, *16*, 102-108.

[51] Ghazizadeh, R.; Tosa, M.; Ghazizadeh, M. Clinical improvement in psoriasis with treatment of associated hyperlipidemia. *Am. J. Med. Sci.*, **2011**, *341*, 394-398.

[52] Pace, G. W.; Leaf, C. D. The role of oxidative stress in HIV disease. *Free Radic. Biol. Med.*, **1995**, *19*, 523-528.

[53] Pasikanti, K. K.; Ho, P. C.; Chan, E. C. Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *J. Chromatogr. B. Analyt Technol. Biomed. Life. Sci.*, **2008**, *871*, 202-211.

[54] Beets, C.; Huang, J.; Madala, N.; Dubery, I. Activation of camalexin biosynthesis in Arabidopsis thaliana in response to perception of bacterial lipopolysaccharides: a gene-to-metabolite study. *Planta.*, **2012**, *236*, 261-272.

[55] Zhang, J.; Bowers, J.; Liu, L.; Wei, S.; Gowda, G. A.; Hammoud, Z.; Raftery, D. Esophageal cancer metabolite biomarkers detected by LC-MS and NMR methods. *PLoS One.*, **2012**, *7*, e30181.

[56] Dietmair, S.; Timmins, N. E.; Gray, P. P.; Nielsen, L. K.; Kromer, J. O. Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. *Anal. Biochem.*, **2010**, *404*, 155-164.

[57] Sellick, C. A.; Hansen, R.; Stephens, G. M.; Goodacre, R.; Dickson, A. J. Metabolite extraction from suspension-cultured mammalian cells for global metabolite profiling. *Nat. Protoc.*, **2011**, *6*, 1241-1249.

[58] Lu, X.; Xu, G. LC-MS Metabonomics Methodology in Biomarker Discovery. In: Biomarker Methods in Drug Discovery and Development. Springer, **2008**; pp291-315.

[59] Masson, P.; Alves, A. C.; Ebbels, T. M.; Nicholson, J. K.; Want, E. J. Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS. *Anal. Chem.*, **2010**, *82*, 7779-7786.

[60] Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G. Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal. Chem.*, **2006**, *78*, 743-752.

[61] Lin, C.; Wu, H.; Tjeerdema, R.; Viant, M. Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. *Metabolomics*., **2007**, *3*, 55-67.

[62] Wheaton, J.P.; Chambers, E.E.; Fountain, K.J. Improving the Cleanliness of DBS Extracts using the Ostro 96-well Sample Preparation Plate and Single Step Method. *Waters Corporation*. **2011,** Waters Application Note #., 720004047EN, 1-6.

[63] Ritchie, M.; Mal, M.; Wong, S. Extraction of Phospholipids from Plasma using Ostro Sample Preparation. *Waters Corporation*. **2012,** Waters Application Note #., 720004201EN, 1-5.

[64] Hoffman, B. Evaluation of a new sample preparation approach for quantitative bioanalytical LC–MS. *Bioanalysis*., **2010,** 2, 1931-1942.

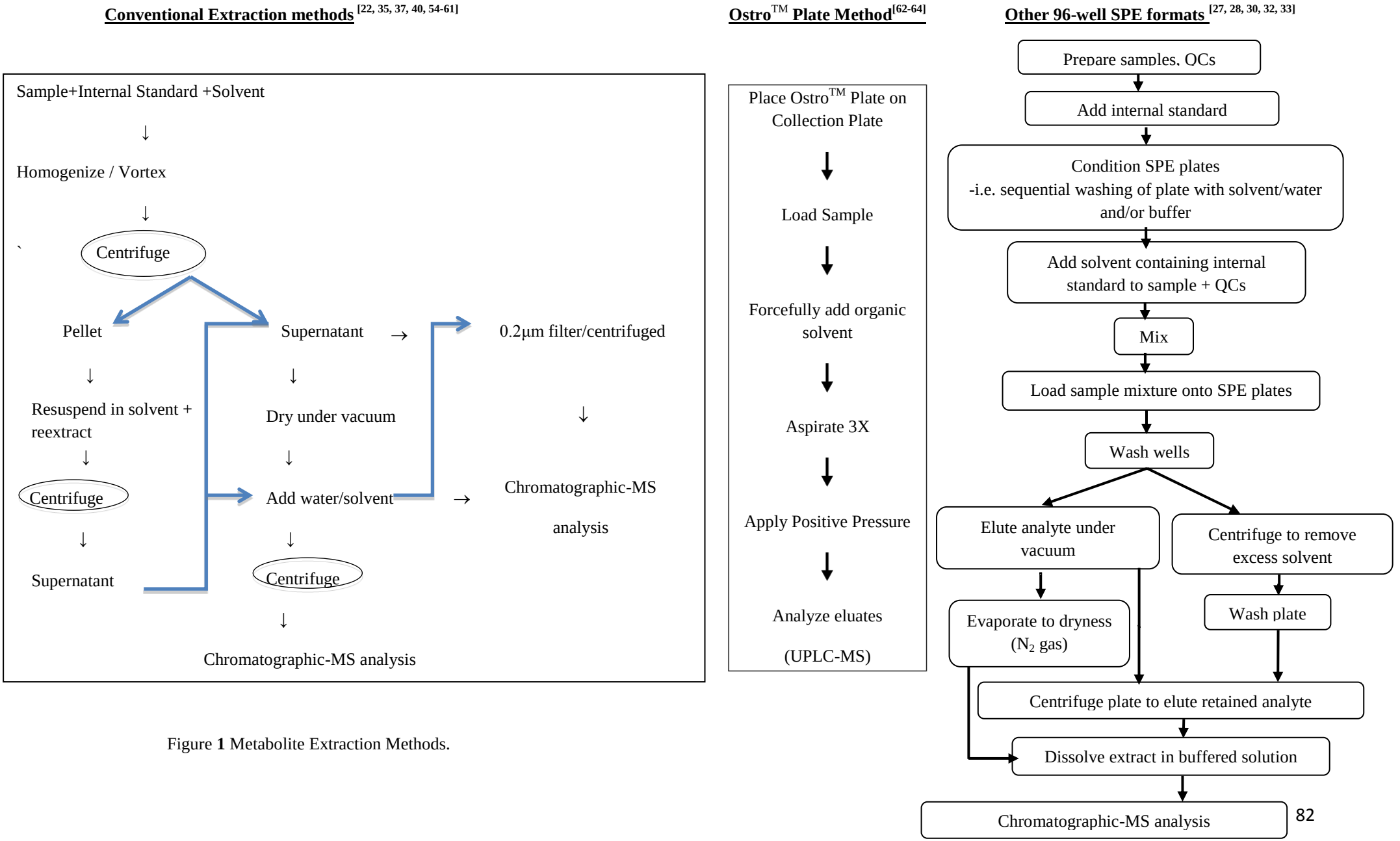


Figure **1** Metabolite Extraction Methods.

Table **1** Characteristics of the participating donors.

	HIV Status		
Characteristic	HIV negative (n=16)	HIV+ (n=13)	HIV+HAART+ (n=15)
% Females	75.0	69.2	80.0
% Males	25.0	30.8	20.0
Mean Age ± SD (years)	26.7 ± 6.1	38.5± 8.9	43.2 ± 11.3
Mean CD4 count (cells/µl blood, range)	Not Determined	549.3 (100 - 2380)	468.9 (204 - 728)
Viral Load (copies/ml, range)	Not Determined	149557 (23-350 430)	1352 (43-3624)
HAART combinations			
Unknown	-	-	5 (33.33%)
NRTIs and NNRTIs	-	-	10 (66.66%)
P-values comparing differences in age among groups			
Group comparisons	HIV negative and HIV+	HIV negative and HIV+HAART+	HIV+ and HIV+HAART+
p-value	0.000228534	1.8539E-05	0.24206586

i

ii

iii

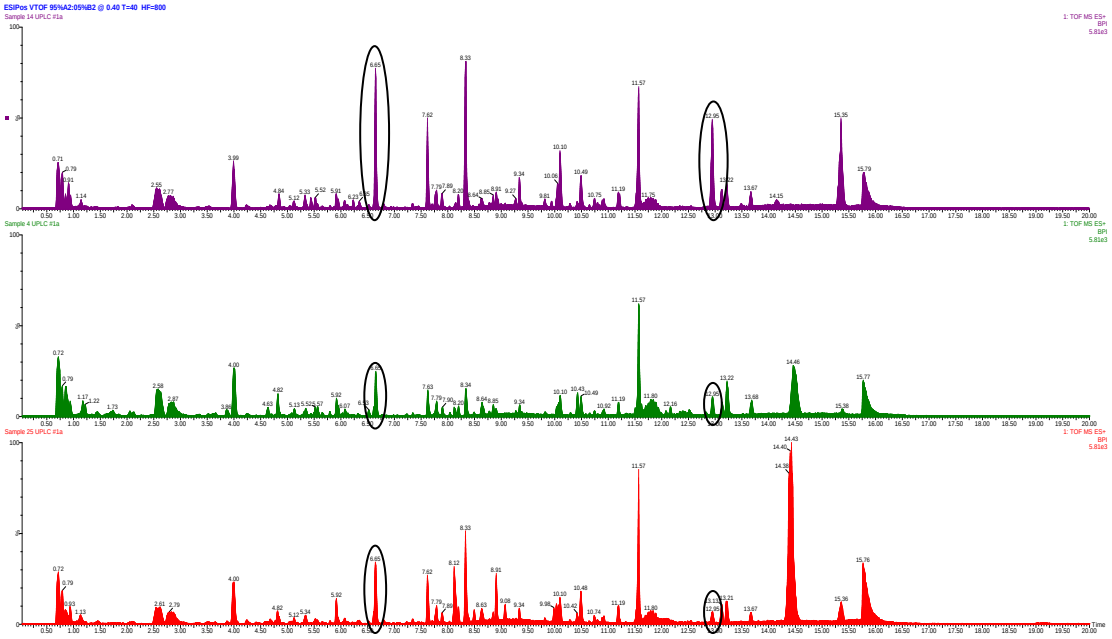


Figure **2** Representative stacked base peak ion (BPI) chromatograms of (i) HIV+HAART+, (ii) HIV+ and (iii) HIV negative sera samples analyzed in ESI+ mode of UPLC-MS. Differences in the three representative groups' metabolites are visible through visual inspection (see representative circles). The use of Ostro™ plates minimize plasticizers from leaching and yielded improved quality chromatograms compared to Figure S2 where plasticizers and phospholipids were present.

Table **2** Classification of experimental cases as HIV negative, HIV+ or HIV+HAART+ in ESI- and ESI+ mode, respectively, using stepwise LDA. This statistical tool uses a linear combination of variables to maximize the separation between experimental groups. In ESI- mode shown in [i], 17 HIV negative, 15 HIV+HAART+ and 11 HIV+ cases were analyzed and correctly classified with 100% accuracy (shaded areas) into their respective groups. Likewise, 15 HIV negative, 14 HIV+HAART+ and 13 HIV+ cases were analyzed and correctly classified with 100% accuracy (shaded areas) in ESI+ mode [ii]. An explanation for the number of cases analyzed is provided under the "**Optimized Metabolite Extraction from Sera using Ostro™ Plates"** section.

i. ESI- Classification Results[a]

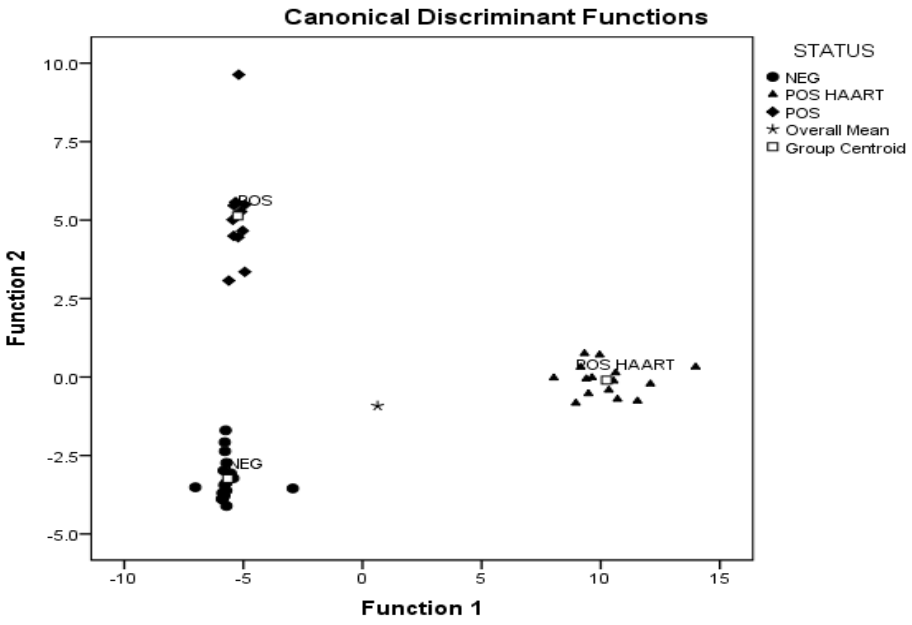
		STATUS	Predicted Group Membership			Total
			NEG	POS HAART	POS	
Original	Count	NEG	17	0	0	17
		POS HAART	0	15	0	15
		POS	0	0	11	11
	%	NEG	100.0	.0	.0	100.0
		POS HAART	.0	100.0	.0	100.0
		POS	.0	.0	100.0	100.0

a. 100.0% of original grouped cases correctly classified.

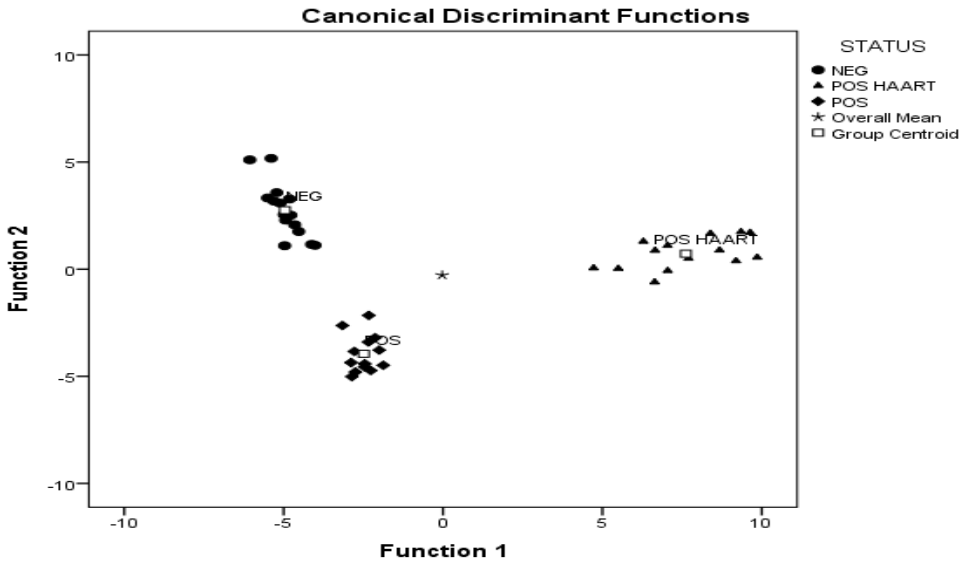
ii. ESI+ Classification Results[a]

		STATUS	Predicted Group Membership			Total
			NEG	POS HAART	POS	
Original	Count	NEG	15	0	0	15
		POS HAART	0	14	0	14
		POS	0	0	13	13
	%	NEG	100.0	.0	.0	100.0
		POS HAART	.0	100.0	.0	100.0
		POS	.0	.0	100.0	100.0

a. 100.0% of original grouped cases correctly classified.



i. ESI-



ii. ESI+

Figure **3** Scatter plot of the discriminant functions for the HIV negative, HIV+ and HIV+HAART+ groups. Here the summed responses of every individual used in the study, is reflected. Shown in [i] and [ii] are the ESI- and ESI+ data respectively showing the HIV+HAART group to be most differentiated from the HIV negative and HIV+ groups respectively.

Table **3** Metabolites from the ESI- [i] and ESI+ [ii] mode data identified through LDA as causing group variances. The differences in group means of these metabolites were also calculated through ANOVA.

i. ESI- mode

	LDA variable (m/z)	**Retention Time**	**p-value (ANOVA)**	**Metabolite ID as per database/library searches**	**Common name**	**Compound Class**	**Biological Role**
1.	V0274MZ213.9635	7.8494	0.000	Methyl laurate	Methyl dodecanoate	Lipids -Fatty acid esters	Cell signalling Fuel and energy source Fuel and energy storage Membrane integrity/stability
2.	V0822MZ343.2067	9.4372	0.020	1-[(2-Chlorophenyl)(diphenyl)methyl]-1H-imidazole	Clotrimazole	-Aromatic heterpolycyclic compounds -Diphenyl methanes	Antifungal medication Used to treat sickle cell disease
3.	V1105MZ411.782	9.4796	0.016	(3beta,24E)-Stigmasta-5,24(28)-dien-3-ol	Fucosterol	Lipids -Steroids and steroid derivatives	Cell signalling Fuel and energy source Fuel and energy storage

							Membrane integrity/stability
4.	V1276MZ458.1392	5.073	0.016	N-(4-{[(2-Amino-5-methyl-4-oxo-1,4,5,6,7,8-hexahydro-6-pteridinyl)methyl]amino}benzoyl)-D-glutamic acid	Leucovorin	Pteridines and derivatives	Treatment of anemias and hematological complications linked to folic acid deficiency
5.	V1367MZ480.0522	8.1732	0.000	Methyl hentriacontanoate	No information		
6.	V1376MZ481.0554	8.1757	0.000	Cytidine 5'-(tetrahydrogen triphosphate)	Cytidine triphosphate	Nucleosides, nucleotides and analogues -Pyrimidine nucleotides	Cell signalling Energy source Membrane component Phospholipid synthesis Membrane integrity/stability Pyrimidine metabolism Transcription/translation Amino sugar metabolism

7.	V0398MZ251.0939	6.3542	0.079	2'-Deoxyadenosine	5'-Deoxy-adenosine	Nucleosides, nucleotides and analogues -Purine nucleotides	Component of DNA Marker for oxidative damage
8.	V0806MZ339.2316	13.3905	0.057	2,6-Di-O-phosphonohex-2-ulofuranose	No information		
9.	V0857MZ353.2123	11.9254	0.056	(1S,3R,4R,5R)-3-{[3-(3,4-Dihydroxyphenyl)acryloyl]oxy}-1,4,5-trihydroxycyclohexanecarboxylic acid	**Chlorogenic acid	Aromatic homonocyclic compounds -Cyclic alcohols and derivatives	Biosynthetic intermediate in lignin (antioxidant) synthesis Limits release of glucose into bloodstream after a meal Antihypertensive effects to minimize strokes and cardiovascular mortalities
10.	V0641MZ305.0963	5.9993	0.403	2-(3,4,5-Trihydroxyphenyl)-3,5,7-chromanetriol	**(-) - Epigallocatechin	Flavonoid	Nutriceutical with anti-cancer, anti-tumor and anti-inflammatory activity
11.	V1192MZ437.0222	6.6879	0.538	-	Nanocosanoic acid	Lipids -Fatty acid	Cell signalling Fuel and energy storage Fuel and energy source

							Membrane integrity/stability

**Metabolites coming from traditional medicine and not diet because they were high in the positive group and metabolite 10 was lowin the HIV negative group. Chlorogenic acid might be from self treatment over the counter anti-oxidant products.

i. ESI + mode

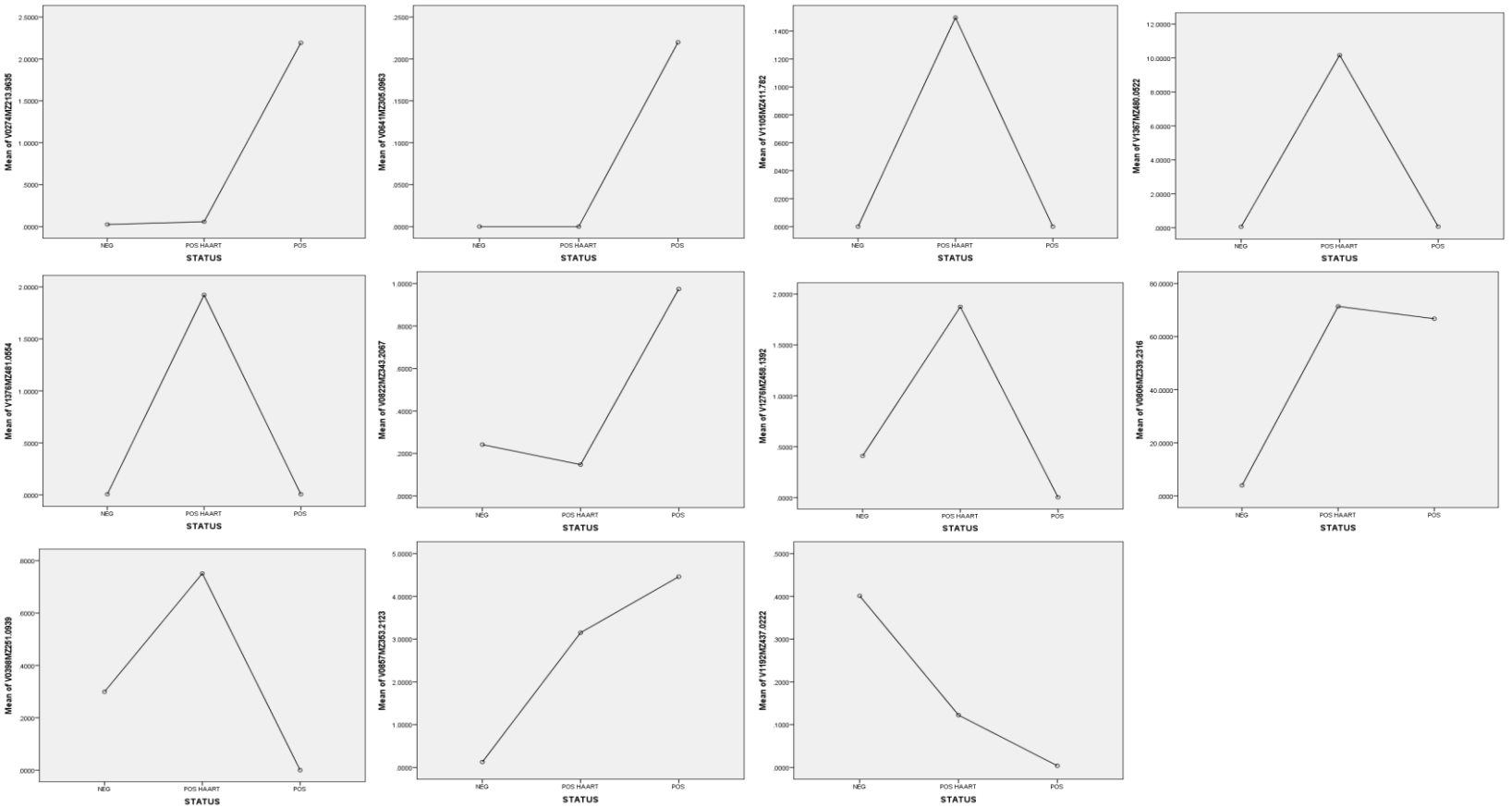
	LDA variable (m/z)	**Retention Time**	**p-value (ANOVA)**	**Metabolite ID as per database/library searches**	**Common name**	**Compound Class**	**Biological Role**
12.	V1658MZ306.0295	8.1739	0.000	(2R,3R)-2-(3,4-Dihydroxyphenyl)-3,5,7-trihydroxy-2,3-dihydro-4H-chromen-4-one	Taxifolin	Flavonoid	Antiviral and antibacterial activity Inhibits tumor development in animals Inactivation of cytotoxic substances Anti-diabetes effect Prevents osmotic pressure in hyperlipidermia *etc* Radical scavenger
13.	V1496MZ283.1159	5.0752	0.005	2-Amino-9-pentofuranosyl-3,9-dihydro-6H-purin-6-one	No information		

14.	V1544MZ288.0421	9.9799	0.006	2-(Hydroxymethyl)phenyl hexopyranoside	Salicin	Carbohydrates and carbohydrate conjugates -Glycosyl compounds	Anti-inflammatory Antipyretic Antirheumatic Glycolysis/ Gluconeogenesis Biosynthesis of phenylpropanoids Phosphotransferase system
15.	V2199MZ433.366	14.3574	0.000	3,5-Diiodo-L-tyrosine	L-Diiodotyrosine	Amino acids,peptides and derivatives	Amino acid and protein biosynthesis Hyperthyroidism (biosynthesis of thyroid hormones) Chronic kidney/ renal disease
16.	V0179MZ564.4313	14.5494	0.000	Tetracontane	n-Tetracontane	No information	
17.	V1007MZ205.1216	10.6055	0.001	1,4-Dihydroxy-2-naphthoic acid	DHNA	No information	Precursor of vitamin K in bacteria Stimulates bifidobacteria growth, improves intestinal conditions in the human

							intestine
18.	V0542MZ675.813	5.3449	0.015	(2R)-1-(Palmitoyloxy)-3-(phosphonooxy)-2-propanyl (9Z)-9-octadecnoate	1-Hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycerol 3-diphosphate PPA(16:0/18:1(9Z))	No information	Cell Signalling Fuel and energy source Fuel and energy storage Phoshpholipid biosynthesis Membrane component Membrane integrity/ stability
19.	V2473MZ516.3383	7.3946	Not available*	2-{[(3alpha,5beta,6beta,7beta,8xi,9xi,14xi)-3,6,7-Trihydroxy-24-oxocholan-24-yl]amino}ethanesulfonic acid	No information		
20.	V1387MZ268.3745	6.6322	0.650	9-Pentofuranosyl-9H-purin-6-amine	No information		

21.	V0287MZ595.2893	11.738	0.616	3-[2-[(E)-[3-(2-carboxyethyl)-5-[(3-ethyl-4-methyl-5-oxo-pyrrolidin-2-yl)methyl]-4-methyl-pyrrol-2-ylidene]methyl]-5-[(4-ethyl-3-methyl-5-oxo-pyrrolidin-2-yl)methyl]-4-methyl-1H-pyrrol-3-yl]propanoic acid	L-Urobilin or Stercobilin	Aromatic heteropolycyclic compounds -Tetrapyrroles and derivatives	Product of heme/hemoglobin degradation Indicator of urinary tract function	*Metabolite below detection hindering p-value calculation

i. ESI-



ii. ESI+

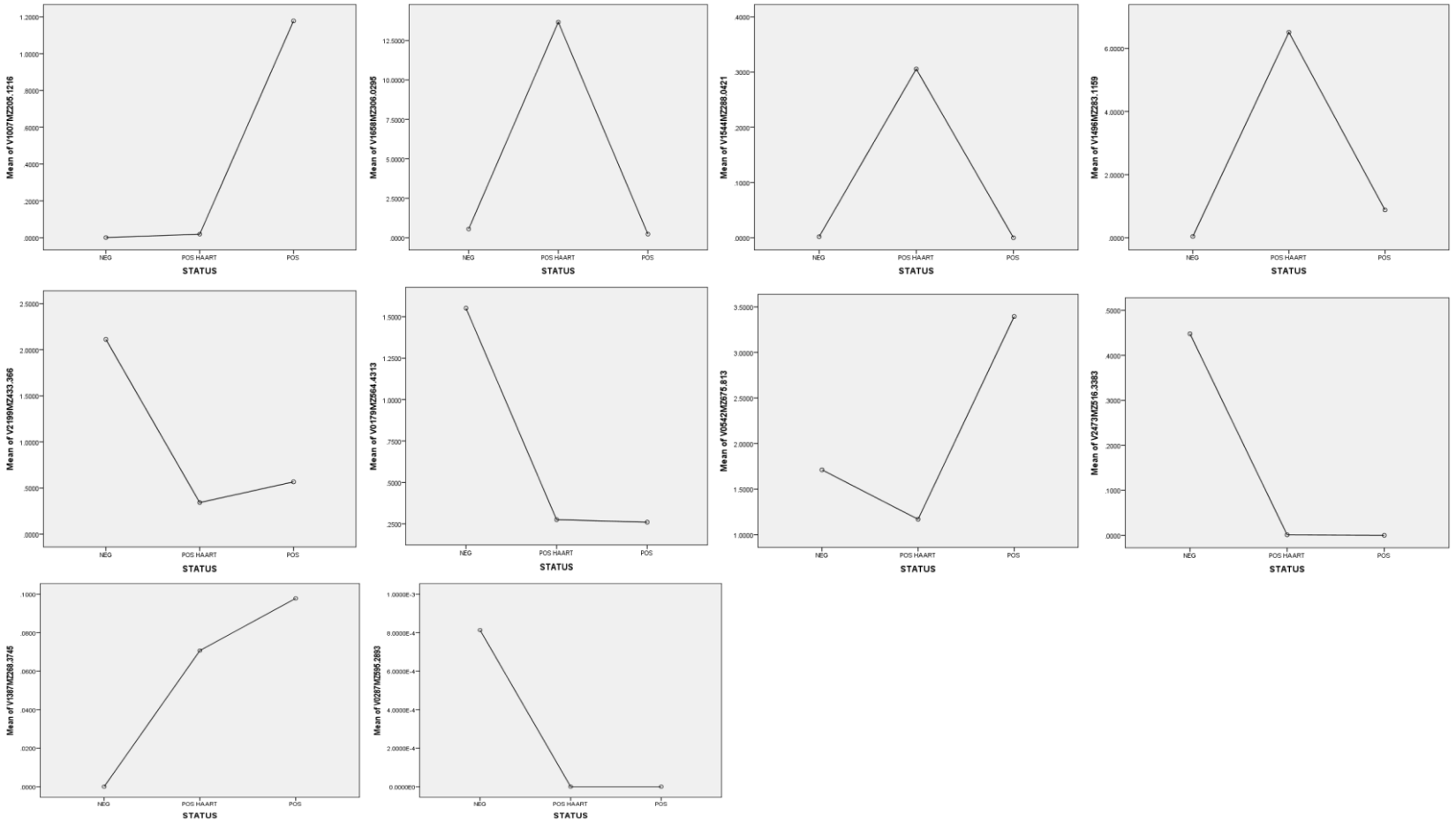


Figure **4** Changes in the mean levels of metabolites extracted from the sera of HIV negative, HIV+ and HIV+HAART+ individuals. Metabolites detected in (i) ESI negative and (ii) ESI+ mode are shown. Mean plots are arranged from left to right as influenced by HIV, HAART and/or a combination of HIV and treatment. Shifts in the levels of the variables among the three groups are visible.

Supportive/Supplementary Material

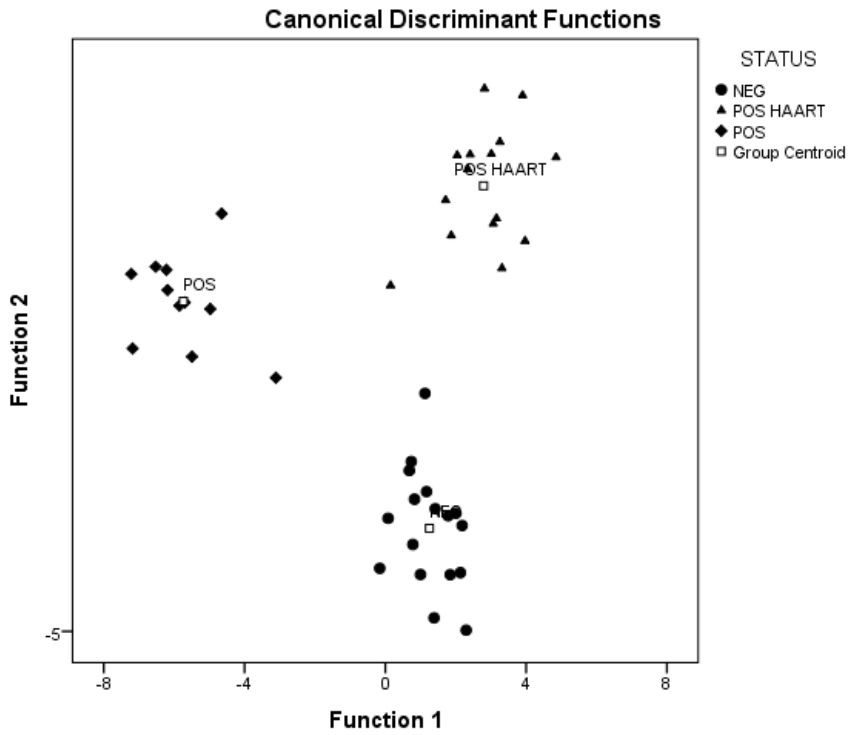


Figure S1 Principal component analysis (PCA) scores plot of HIV negative, HIV+ and HIV+HAART+ groups that were detected using UPLC-MS. The three experimental groups separated from each other, indicative of clear differentiation among them. The HIV+ group differentiated from the HIV negative and treated patients along the first discriminant while the HIV+ and treatment-experienced patients differentiated from HIV negative samples along the second discriminant.

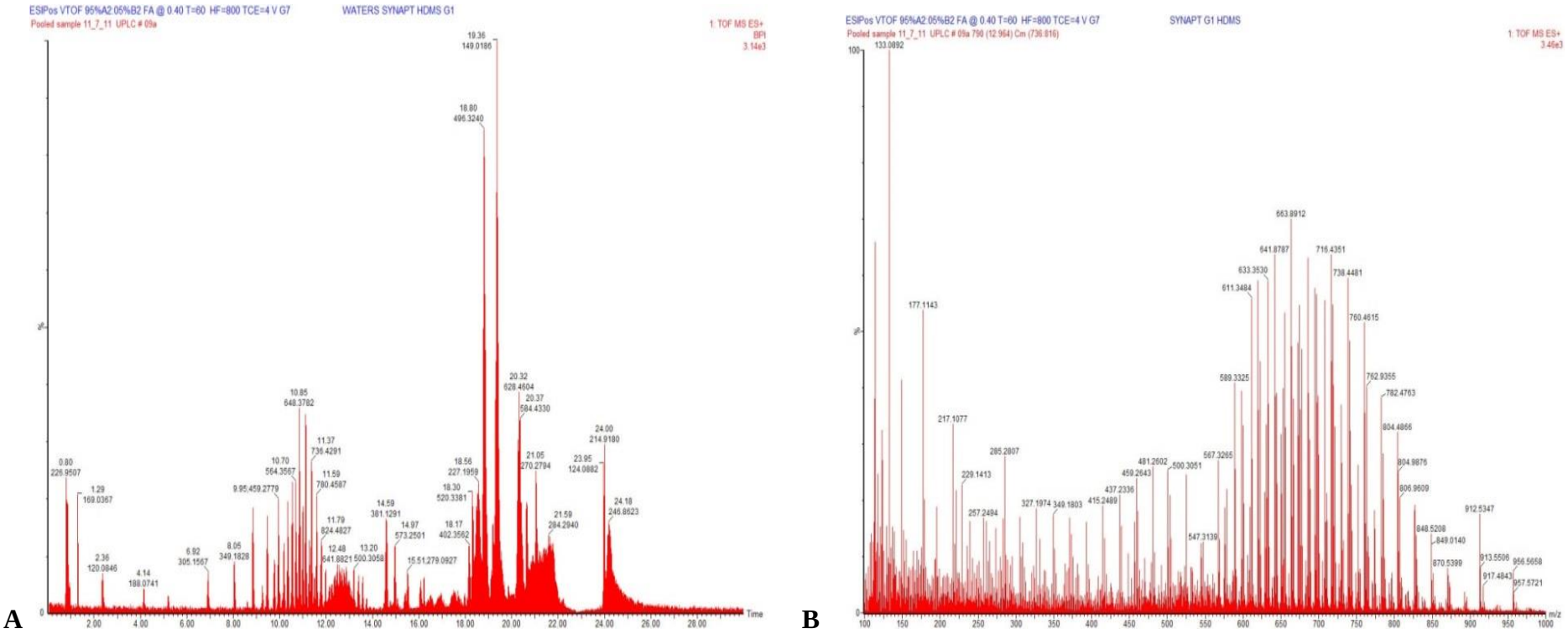


Figure **S2**: ESI positive BPI chromatogram (A) of pooled HIV sera after a multi-step solvent extraction protocol. In B, the presence of plasticizers and phospholipids are visible. The large amount of chromatographic peaks that originated from plasticizers made detecting and selecting metabolomics-related compounds very difficult. This is not surprising as these molecules are known to mask metabolites and suppress ionization of target compounds. Using the Ostro™ plates eliminated both classes of compounds resulting in improved chromatography and mass spectral quality (shown in Figure **2** of the main manuscript).

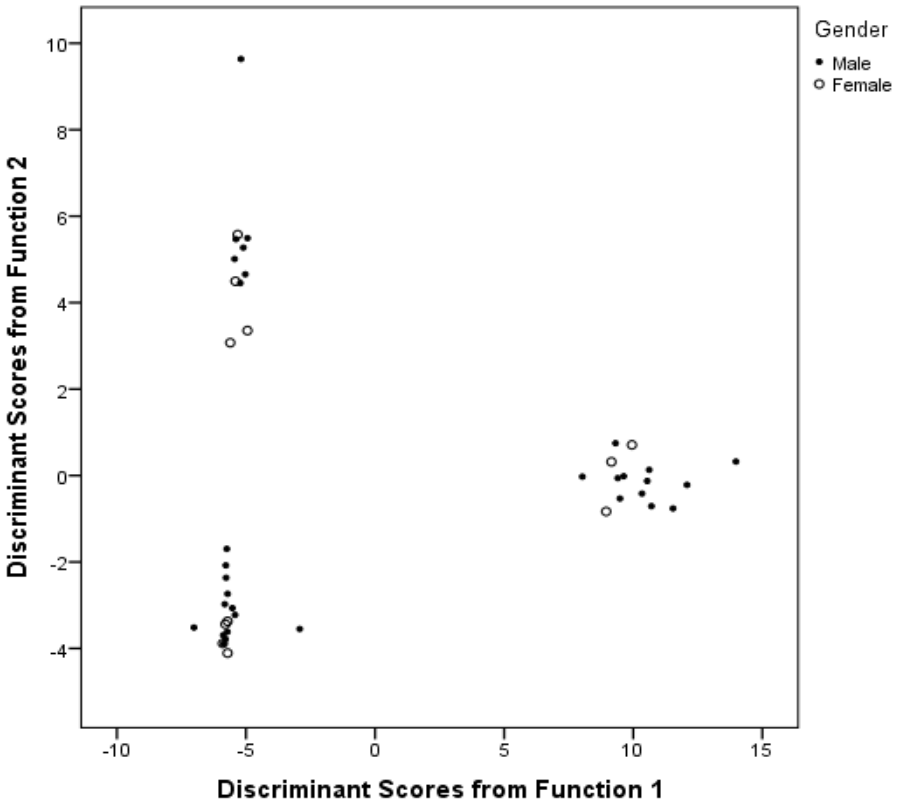


Figure **S3.** LDA scatter plot of HIV negative, HIV+ and HIV+HAART+ groups with the inclusion of gender as a variable. The scatter plot shows no clear separation between the genders. The means of the groups were not significant. This failure to cluster according to gender suggests that the separation of the groups is mainly as a result of HIV infection and/or treatment and not due to confounders like gender. Similar to the "gender-LDA" plot, age when entered as a variable did not play a significant role in the results either.

Section 4.2: UPLC-MS and advanced chemometrics

4.2.1) UPLC-MS Chromatographic Output

Serum samples of HIV negative subject (NEG), HIV positive (POS) and HIV positive patients receiving treatment (POS HAART) were analyzed in both ESI negative and ESI positive mode of UPLC-MS to obtain global metabolic profiles of HIV. The representative stacked BPI chromatograms shown in figure 4.1 were closely matched in terms of age, gender and race to avoid the influence of these potentially confounding factors. This was done to show comparable differences among the three experimental samples. There were no broad peaks because Ostro^{TM} plates used in the metabolite extraction process removed compounds that cause matrix effects. This extraction method eliminated metabolite masking and matrix effects. Figure 4.1 clearly indicates matrix-free, distinct, narrower and higher resolution peaks that resulted from using this sample preparation method. Chromatographic differences of the three experimental samples were visible in changes in peak elution times (brown diamonds), peak patterns (yellow diamonds) and peak intensities (blue diamonds). Peaks eluted at the time period of 4.00 minutes showed increased intensities in the POS samples in comparison to the NEG and POS HAART samples. The peaks that eluted at a time range between 0.5 minutes and 1.00 minute drastically differ in the POS sample in comparison to the NEG and POS HAART samples. The POS HAART sample appears to have slightly shifted peaks, such as the high intensity peak observed at 6.62 minutes which differs from the low intensity peaks eluted at 6.65 minutes in the NEG and POS serum samples.

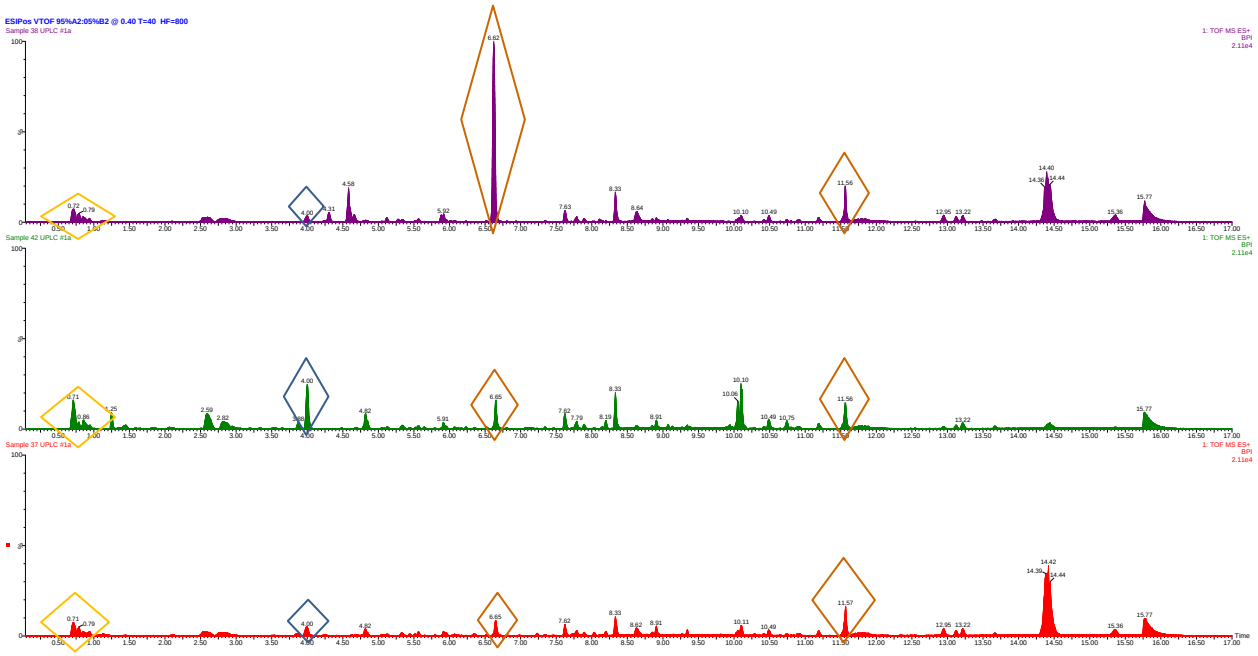


Figure 4.1: Representative stacked base peak ion (BPI) chromatograms of NEG (red), POS (green) and POS HAART (purple) serum samples detected in ESI+ mode of UPLC-MS. Differences in affected metabolites were clearly visible by visual inspection (colored diamonds). The POS HAART sample presented as the most altered chromatographic profile in comparison to the NEG sample. The POS sample showing mainly changes in peak intensities when compared to NEG sample.

Visual inspection of chromatograms is not ideal for data mining in high throughput studies such as metabonomics. This study being holistic in nature looked at the metabolic profiles of all the detectable compounds, therefore visual differences of raw data cannot serve as hypothesis driven results from which conclusions can be drawn but merely serve as a good starting point for data mining. Data is usually processed to reduce the details to the main points and to derive meaningful information from data. For data processing, bioinformatics tools provided a platform for optimizing parameters for biological data analysis. Notably, visible changes in chromatograms are not always representative of the cause or effect induced among the experimental groups, therefore, chemometrics were conducted to determine the significance of the affected metabolites.

4.2.2) Data Processing using MassLynx v4.1 software

Raw data containing metabolite features (retention time and mass-to-charge ratio) was processed using the Masslynx v4.1 software. Following processing, ESI positive mode detected 1605 variables from 186 observations and ESI negative mode data detected 1079 variables from 185 observations. Processing method parameters were set as mentioned in

section 3.3.6. Additional noise filtering was conducted to aid the assessment between background noise and signal response.

4.2.3) Statistical Analysis using SPSS software v19.0

Processed data files were exported to the SPSS software version 19.0. The following multivariate and univariate statistics were performed; LDA, ANOVA and Mean plots. LDA uses linear combinations of features from variables that will achieve the most separation among the groups. Variables, coefficients and metabolites are used in this context interchangeably referring to the same thing. LDA scatter plots showed a clear separation among the three HIV sera groups (negative, positive and HAART). As seen in figure 4.2(A), ESI negative mode serum samples obtained maximum between group separation and minimum within group separation and achieved 100 % correct classification [Table 1 (i)] following leave-one-out-at-a-time cross validation. An extremely good classification of 88.9 % [Table 1 (ii)] was achieved for ESI positive serum samples shown in figure 4.2(B) with five cases being misclassified following cross validation as can be seen in the separation of the serum samples which show partially maximized within group separation that resulted in some groups clustering closer to other groups.

Table 4.1: LDA classification results of HIV serum samples analyzed in (i) ESI negative and (ii) ESI positive mode of UPLC-MS. ESI negative mode data achieved 100% accuracy and ESI positive mode achieved 88.9% classification accuracy following cross validation.

i. ESI- mode Classification Results[a,c]

Group			Predicted Group Membership			Total
			Negative	HAART	Positive	
Original	Count	Negative	15	0	0	15
		HAART	0	15	0	15
		Positive	0	0	15	15
	%	Negative	100.0	0.0	0.0	100.0
		HAART	0.0	100.0	0.0	100.0
		Positive	0.0	0.0	100.0	100.0
Cross-validated[b]	Count	Negative	15	0	0	15
		HAART	0	15	0	15
		Positive	0	0	15	15
	%	Negative	100.0	0.0	0.0	100.0
		HAART	0.0	100.0	0.0	100.0
		Positive	0.0	0.0	100.0	100.0

a. 100.0% of original grouped cases correctly classified.

b. Cross validation is done for all cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 100.0% of cross-validated grouped cases correctly classified.

ii. ESI+ mode Classification Results[a,c]

Group			Predicted Group Membership			Total
			Negative	HAART	Positive	
Original	Count	Negative	15	0	0	15
		HAART	0	15	0	15
		Positive	0	0	15	15
	%	Negative	100.0	0.0	0.0	100.0
		HAART	0.0	100.0	0.0	100.0
		Positive	0.0	0.0	100.0	100.0
Cross-validated[b]	Count	Negative	15	0	0	15
		HAART	2	13	0	15
		Positive	3	0	12	15
	%	Negative	100.0	0.0	0.0	100.0
		HAART	13.3	86.7	0.0	100.0
		Positive	20.0	0.0	80.0	100.0

a. 100.0% of original grouped cases correctly classified.

b. Cross validation is done for all cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 88.9% of cross-validated grouped cases correctly classified.

Figure 4.2 showed that along the first discriminant function, the HAART treatment group was the most differentiated group as it scattered furthest away from both the HIV negative and positive groups. This shows that the metabolism of the treatment group was the most affected. The metabolism of the positive group could be considered relatively healthy because the HIV negative and positive groups clustered along the same side. Evaluation of the treatment group along the second discriminant function showed that HAART was grouped between the negative and positive groups. This could imply that the therapy might be succeeding in exerting its desirable effects of restoring the metabolic system; however, the metabolic profile is still being disrupted, as treatment is known to do.

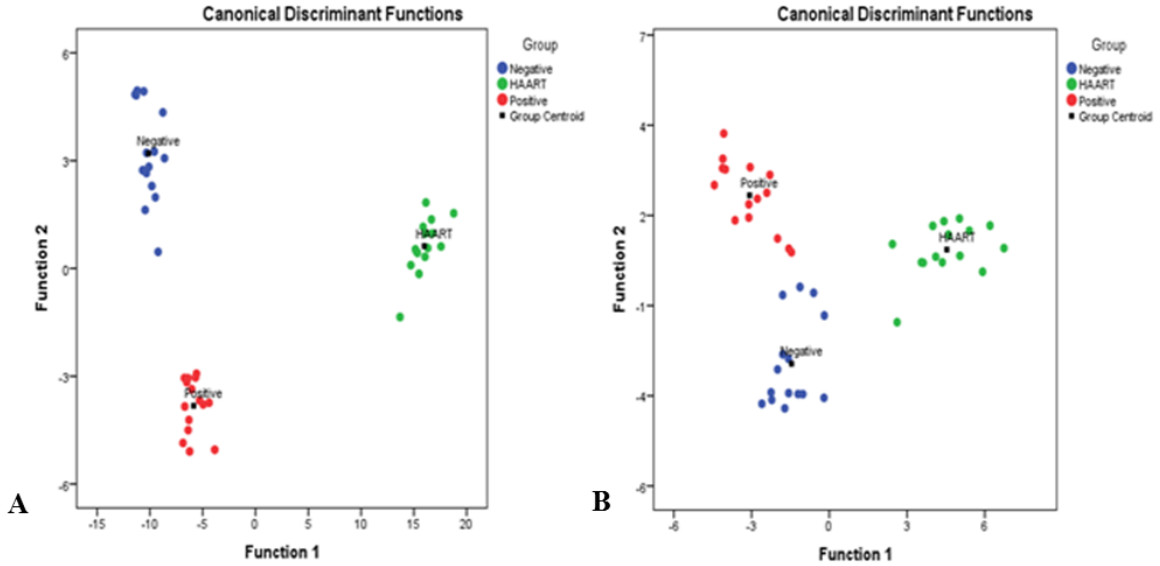


Figure 4.2: LDA scatter plots of the three respective HIV serum sample groups (negative, positive and HAART) analyzed in (A) ESI- mode and (B) ESI+ mode of UPLC-MS. Both modes achieved good separation and group classification. In both figures, the HAART group seems to be the most differentiated group from HIV negative and positive groups along the first conical discriminant functions. The HIV negative and positive samples grouped on the same side along the first discriminant function. Along the second discriminant function, the HAART group appears between the negative and positive group, showing some admirable HAART restoring effects.

Twenty-five metabolites were detected in both ionization modes following LDA as contributing to the group separations, 14 in negative mode and 11 in the positive mode and were significant ($p<0.05$) as calculated by ANOVA. LDA metabolites were identified using the MarkerLynx XS software because this software is linked to the relevant databases. Significance of these metabolites together with their characteristics and functions are listed in table 4.1 and 4.2.

Table 4.2: Identification characteristics of significantly altered LDA sera metabolites detected in ESI negative mode of UPLC-MS and their functions or pathways involved in. These metabolites caused separation of the HIV negative, positive and HAART groups as observed in LDA scatter plots.

Variable features	Function		p-value	Variable ID	Metabolic functions/pathways
	1	2			
Z0089RT1443MZ180841	.736	.039	0.035	Divanadium pentaoxide	N/A
Z0199RT959MZ265967	-5.631	-2.135	0.001	Sulfafurazole	Anti-infective
Z0405RT845MZ371176	-.459	.557	0.005	Lysine,N6-Biotinyl	Biosynthesis and degradation of Biotin
Z0560RT1257MZ441358	-1.141	1.164	0.000	Obtusifoliol	Intermediate in the biosynthesis of cholesterol
Z0602RT528MZ457136	5.280	1.309	0.001	Reduced flavin mononucleotide ($FMNH_2$)	Active form of riboflavin
Z0621RT83MZ465248	-.520	.642	0.001	Clobetasol propionate	N/A
Z0665RT103MZ480308	1.310	-.296	0.044	(2R)-3-(Hexadecyloxy)-2-hydroxypropyl 2-(trimethylammonio)ethyl phosphate	N/A
Z0693RT851MZ49403	2.870	.601	0.002	Dihydroneopterin triphosphate	Intermediate in folate biosynthesis
Z0720RT814MZ507051	-5.191	-.696	0.000	Inosine triphosphate (ITP)	Intermediate in purine metabolism
Z0764RT726MZ522042	14.596	3.517	0.000	Guanosine triphosphate (GTP)	Intermediate of folate metabolism
Z0839RT79MZ56732	.663	.647	0.046	Protoporphyrinogen IX	Intermediate in heme biosynthesis Component of porphyrin and metabolism
Z0852RT1323MZ587055	-1.720	-.467	0.003	Kolaflavanone	Antiviral, anti-inflammatory and anti-diabetic
Z0862RT1268MZ593478	.872	.119	0.041	1-Hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycerol	Cell signaling, fuel and energy source/storage and membrane component involved in integrity/stability
Z0906RT864MZ66117	.079	-.626	0.029	Godaversetamide	N/A

Table 4.3: Identification characteristics of significantly altered LDA sera metabolites detected in ESI positive mode of UPLC-MS and their functions or pathways involved in. These metabolites caused separation of the HIV negative, positive and HAART groups as observed in LDA scatter plots.

Variable features	Function		p-value	Variable ID	Metabolic functions/pathways
	1	2			
Z0152RT086MZ162113	-.403	.429	0.018	Nicotine imine	An intermediate in conversion of nicotine to cotinine and a waste product
Z0233RT482MZ196092	.518	-.228	0.034	6-hydroxymethyl-7, 8-dihydro-pterin	Folate biosynthesis pathways
Z0634RT72MZ310202	-.519	.577	0.001	Sialic acid	Amino acid metabolism
Z0730RT1443MZ33915	-.522	.708	0.016	5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide	Purine metabolism
Z1000RT172MZ459113	.736	.101	0.029	Reduced FMN	Active form of riboflavin
Z1005RT512MZ460155	1.751	-.274	0.001	5-Methyltetrahydrofolate	Active form of folic acid, folate biosynthesis
Z1144RT812MZ511067	1.600	-.015	0.000	Lufenuron	Insect developer inhibitor
Z1163RT1213MZ521421	-.241	-.561	0.001	1-oleoyl-3-α-D-galactosyl-sn-glycerol	N/A
Z1270RT792MZ601265	-.127	-.813	0.000	Torcetrapib	Anti-cholesterol drug
Z1455RT556MZ750386	-.613	.031	0.008	3'-N-debenzoyltaxol	Anticancer drug
Z1499RT588MZ811376	-.337	.699	0.002	Vinblastine	Anti-tumor drug

Following determination of the metabolites causing the observed LDA separations among the three HIV serum sample groups, relative levels of the group distinguishing metabolites were determined using mean plots. Mean plots show shifts in location of a specific metabolite among the three experimental groups, increases and decreases in the level of the metabolite among the various groups give an indication of the factor/s (HIV and/or HAART) causing metabolic profile alteration.

Mean plots of all 25 metabolites are shown in the figure A1 (see appendix). Mean plots of these metabolites depicted changes in the metabolites that contributed to the group differences observed through LDA scatter plots. Of the 25 LDA metabolites, 16 (64.00%) of the observable alterations in the metabolites were caused by the combination of HIV and HAART, while 9 (36.00%) were due to HAART alteration and HIV infection appears to have not solely contributed to any alteration in the levels of the metabolites. Mean plots show a trend that the combination of HIV and HAART is the major cause of shifts in location of most metabolites compared to the HIV negative group.

Shown below in figure 4.3 are mean plots of (A) ESI+ metabolite, reduced flavin mononucleotide ($FMNH_2$) and (B) ESI- metabolite, sialic acid, they are two representative types of metabolites normally detected from each ionization mode. $FMNH_2$ appearing to be increased in the HAART group as compared to the decreased levels seen in the HIV negative and positive group. Sialic acid levels were high in both the HIV positive and HAART group and low in the HIV negative group. $FMNH_2$ levels were drastically altered by HAART and the sialic acid level showed that a combination of HIV and HAART caused metabolic alteration.

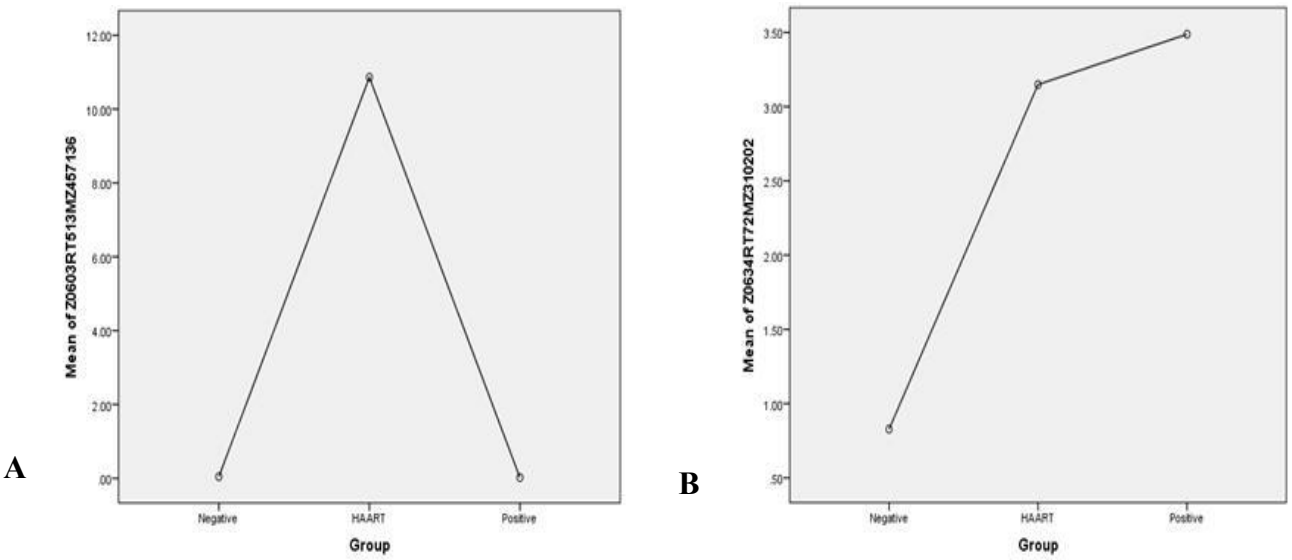


Figure 4.3: Mean plots of metabolites detected in HIV serum samples analyzed in (A) ESI+ and (B) ESI- mode of UPLC-MS. A) $FMNH_2$ showing higher (12-fold) levels in the HAART group as compared to the low levels observed in the HIV negative and positive group. HAART causing most disruption of the metabolism. B) Sialic acid showing higher levels in both the HIV positive group and HAART group compared to the negative group, most increase seen in the HIV positive group. Combination of HIV and HAART altering sialic acid levels.

4.2.4) Advanced chemometrics using MarkerLynx XS, EZinfo 2.0 software

LDA, mean plots and ANOVA were performed by a statistician at the department of Statistics of the University of Pretoria. The additional statistics that follow were necessary for the determination of metabolites that were unique for differentiating only two classes of samples (e.g. only NEG and POS differentiating metabolites). This enables the specific assignment of class separating metabolites to either HIV or HAART.

The MassLynx v4.1 software was purchased and employed in this study because it was highly recommended by LC-MS and MS experts for the handling of LC-MS data. Statistical analysis was conducted using the MarkerLynx extended statistics (XS), the EZinfo 2.0 statistical package. PCA, OPLS-DA's and accompanying S-plots together with other statistical models or projections were used for data mining of observations (samples) and variables (metabolites). Basic training for this software was provided, however not enough literature was available on the use of this software for global metabonomics studies. As a result; how the software works, how to use the software, optimization of parameters that work best for this dataset, the kind of information that can be determined and validation of models became part of this masters research project.

EZinfo 2.0 statistical package contains models or projections that allow for greater in-depth data mining. Information that could be uncovered includes; an overview of the grouping trends without supervision using PCA, supervised between and within group variations using OPLS-DA, determination of distinguishing metabolite's covariance and correlation using S-Plots, detection of outliers using Hoteling's T2 ellipse (95% confidence interval) and the distance to model statistical tool, determination of levels of metabolites and the factor of change between groups using variable average plots, scaling of the variables of importance of the distinguishing metabolites using VIP plots, determination of the levels of metabolites in individual samples using trend plots etc. These are some of the reasons additional statistics had to be conducted in order to obtain more information from the data.

4.2.4a) Principal component analysis results

To get an overview of the grouping trends, PCA uses a few principal components that project spectral data and give a view of raw data. As can be seen in figure A2 and A3 in the appendix, which shows the PCA scores plots of the study samples before and after data pre-treatment, showing that normalization improved the separation of HIV serum samples into their experimental groups. PCA data was normalized using the following template settings; pareto scaled, automatic, transform all if any transformed and maximum number of components was -1 to allow cross validation. Some outliers were visible outside the Hotelling's T2 ellipse (95% confidence interval) and were removed. An outlier is a sample that behaves largely different from the other samples, thus making it dissimilar from the rest of the other samples and it can either be interesting or distorting to the data. The three PCA outliers were removed because they appeared to be distorting the data.

Further data mining from this point onwards was performed from the normalized PCA scores plots (containing all groups) that had some outliers removed. These PCA scores plot are shown in figure 4.4 and 4.5 for ESI- and ESI+ mode data respectively. PCA scores plot contain several principal components that project the data differently. The combination of two components of any of several components is selected based on which pair provides a better separation. Usually the first two components are selected because they show the highest variation in the data as compared to the other additional components. Model validation parameters of PCA scores plot for metabonomics studies are usually acceptable or

good when they have goodness of fit [R^2X (cum)] values of ≥0.5 and higher and model predictability [Q^2(cum)] values that are above 0.5 and closer to 1.

PCA scores plot of the ESI negative mode shown in figure 4.4 only shows 22.00% of the variation as explained by the first two principal components of a twenty component model. The highest variation (13.81%) is explained by the first principal component and 8.19% of the data variation is explained by the second principal component. Some overlapping of the three experimental groups (NEG, POS and POS HAART) was observed along the second principal component. Approximately 70% of the POS HAART sera samples grouped further away from the NEG and POS samples as seen along the first principal component. Model validation parameters were R^2X (cum) (0.591) and Q^2 (cum) (0.390). Quality control (pooled) samples clustered together towards the origin as expected. Outliers visible outside the Hotelling's T2 ellipse and some POS HAART samples appeared to be outliers. This may represent patients that are recalcitrant to HAART and alternative treatment may be required. The NEG and POS groups appear to have a mostly similar metabolic profile because of the overlapping samples along the first and second principal component. It is clear from the scores plot that the POS HAART group contains the most altered metabolic profile as compared to the NEG and POS group.

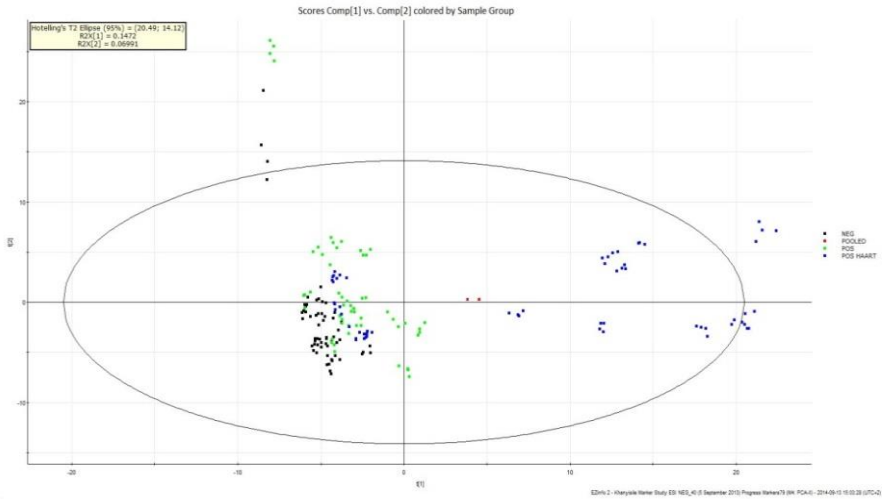


Figure 4.4: PCA scores plot of HIV serum samples (including pooled samples) analyzed in ESI negative mode of UPLC-MS. An overview of the grouping trends visible by visual inspection. Pooled samples clustered together and most sample groups achieved good separation. Overlapping of the three experimental groups observed along the second principal component. Approximately 70% of the POS HAART samples clustered completely in isolation from other groups samples along the first principal component. POS HAART outliers also visible outside the Hotelling's T2 ellipse of the (95% confidence interval). The treatment group metabolic profile mostly altered. Model validation parameters were R^2X (cum) (0.547) and Q^2 (cum) (0.371).

In figure 4.5, the three experimental groups separated from each other well while some serum samples overlapped with other groups. The quality control samples grouped extremely closer to the origin and clustered together as expected. Model validation were R^2X (cum) (0.607) and Q^2 (cum) (0.425). Only 24.29% of the variation as explained by the first two principal components of a nineteen component model is shown. Highest variation in the data (16.78%) is explained by the first principal component and 7.51% of the data is explained by the second principal component. The NEG group and POS group dominantly showing a clear separation along the first principal component with a few samples overlapping. The POS HAART samples showing overlaps with both the NEG and POS group along the first and second principal components. The scores plot shows that the metabolic profile of the NEG and POS groups are mostly different while the POS HAART group contains serum samples that resemble metabolic profiles of both the NEG and POS group.

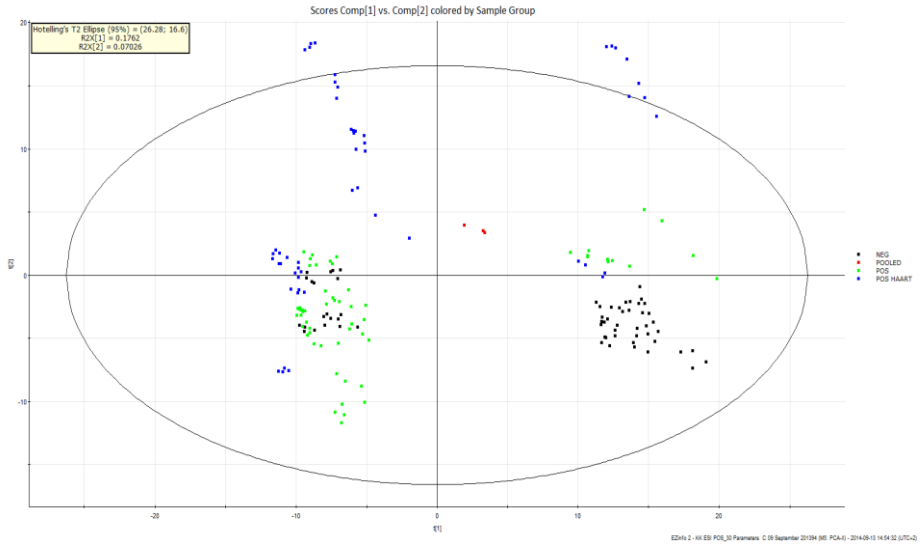


Figure 4.5: PCA scores plot of HIV serum samples (including pooled samples) analyzed in ESI positive mode of UPLC-MS. An overview of the grouping trends visible by visual inspection. Pooled samples clustered together and extremely closer to the origin. The three experimental group samples overlapped with each other along the second principal component. The first principal component responsible for most observed separation between the NEG and POS group with some samples overlapping between the two groups. Samples of the POS HAART group overlapped with samples of the NEG and POS groups. Model validation parameters were R^2X (cum) (0.539) and Q^2 (cum) (0.421). The NEG and POS samples mostly separable, showing that metabolic alteration is caused by the combination of HIV and HAART.

The five previously misclassified LDA cases of the ESI+ data [shown in table 4.1(ii)] corresponded highly with five of the six ESI+ PCA scores plot samples that were overlapped

among other groups. Bearing in mind that LDA uses sample averages and PCA using multiple injections of a single sample to classify samples. LDA cross validation on ESI positive data predicted three of the sera samples that were initially categorized as having HIV infection to actually group with the negative individuals, and the figure 4.5 confirms that two of these three samples predicted by LDA clustered with negative samples. And again, two of the misclassified LDA treatment cases that were predicted as belonging to the negative group are shown in figure 4.5 to cluster with the negative samples and figure 4.5 shows an additional treatment sample observation set that also clustered with the negative group but not predicted by LDA as a misclassified. LDA scatter plots and PCA scores plots showing a high degree of complementary findings.

4.2.4b) Orthogonal projections to latent structures and S-Plot results

Following PCA, a supervised method in the form of OPLS-DA was performed in order to determine the distinguishing metabolites between the primary two classes of samples. It offers maximum separation projection and interpretation is easier with only two groups (Stenlund et al, 2008). Interpretation of data is advantageous because OPLS-DA shows the class discriminating variables. Group differences are necessary for the development of the global metabolic profile so that distinguishing metabolites can specifically be attributed to either HIV or HAART. OPLS-DA was necessary for explaining detailed between and within group differences of two groups of HIV serum samples. OPLS-DA shows the identity of the serum samples that overlap between the two groups and the identities of the samples forming within group clusters, this aids the process of finding out why some observations are behaving the way they do. This scores plot is accompanied by an S-Plot that contains the variables that contribute to the separation that is observed on the OPLS-DA scores plot, their correlation and their levels of confidence. Unlike mean plots used earlier that simultaneously determined the metabolite levels of a single metabolite in the entire three experimental groups relative to each other, S-Plots are restricted to comparing only two groups. This is an advantage because some metabolites are present only in one or two sample groups and not the entire three experimental groups. It is a challenge to represent the alterations/presence of these metabolites if the mean plots show the levels of such metabolites in the entire three experimental groups. The variables at the extreme ends of the S shape are normally selected because they are variables with the highest confidence levels for the observed correlation.

OPLS-DA score plots are usually presented in two dimensions (2D), however they can also be presented in three dimension (3D) depending on the dataset because data can be viewed using a combination of different principal components. If the 2D scores plot does not show a clear heterogeneity in the separation of the samples that is expected and known to exist then 3D OPLS-DA scores plot can be used. The 3D scores plot provide a better view of the data at different angles to enable the better visual inspection of the separation observed. This results in more information being deducted from the scores plot about the clustering of the samples. As mentioned earlier, the first two principal components are normally chosen to represent the data, not because they always show the better separation of samples but because they show the highest variation in the dataset. In this study, the 2D OPLS-DA scores plots were not sufficient for visually representing the heterogeneity of the samples that were clearly separable as can be seen in figure 4.6 scores plot.

The three 2D OPLS-DA scores plots and their accompanying S-Plots of the ESI negative data are shown in figure A4 (see appendix). The interest of this study was to determine the metabolites that were affected by HIV and/or HAART; this can be achieved by accounting for most of the observed variations between the groups which are shown in these plots by the first and second principal component. Figure A5 in the appendix shows only two of the three OPLS-DA combinations of the experimental groups. The missing third pair combination (NEG and HAART) OPLS-DA is represented in this section as a representative of the analysis that was conducted on each pair.

Figure 4.6 shows the OPLS-DA scores plot of group differences between NEG and POS HAART samples. This is shown by the separation observed along the first principal component projection (left and right side clusters) and within group differences evident along the second principal component projection (top and down clusters) of an eight principal component model. Model validation was R^2Y (cum) (>0.999) and Q^2 (cum) (0.982). Tight clustering of samples that appeared to show a linear line was a result of detected outliers seen lying outside the Hotelling's T2 ellipse (the 95% confidence interval) in figure 4.6 (C and D). According to EZInfo 2.0 software package; Hotelling's T2 test detects strong outliers, which are extreme observations as it measures how far away the observation is from the center of the model. While moderate outliers are detected by the distance to model in X-space (DModx) statistical tool which shows how poorly summarized other observations are. It is the

summary of the noise in the observation, the larger the noise, the more dissimilar the observation is from the others. Moderate outliers are not always upsetting to the model.

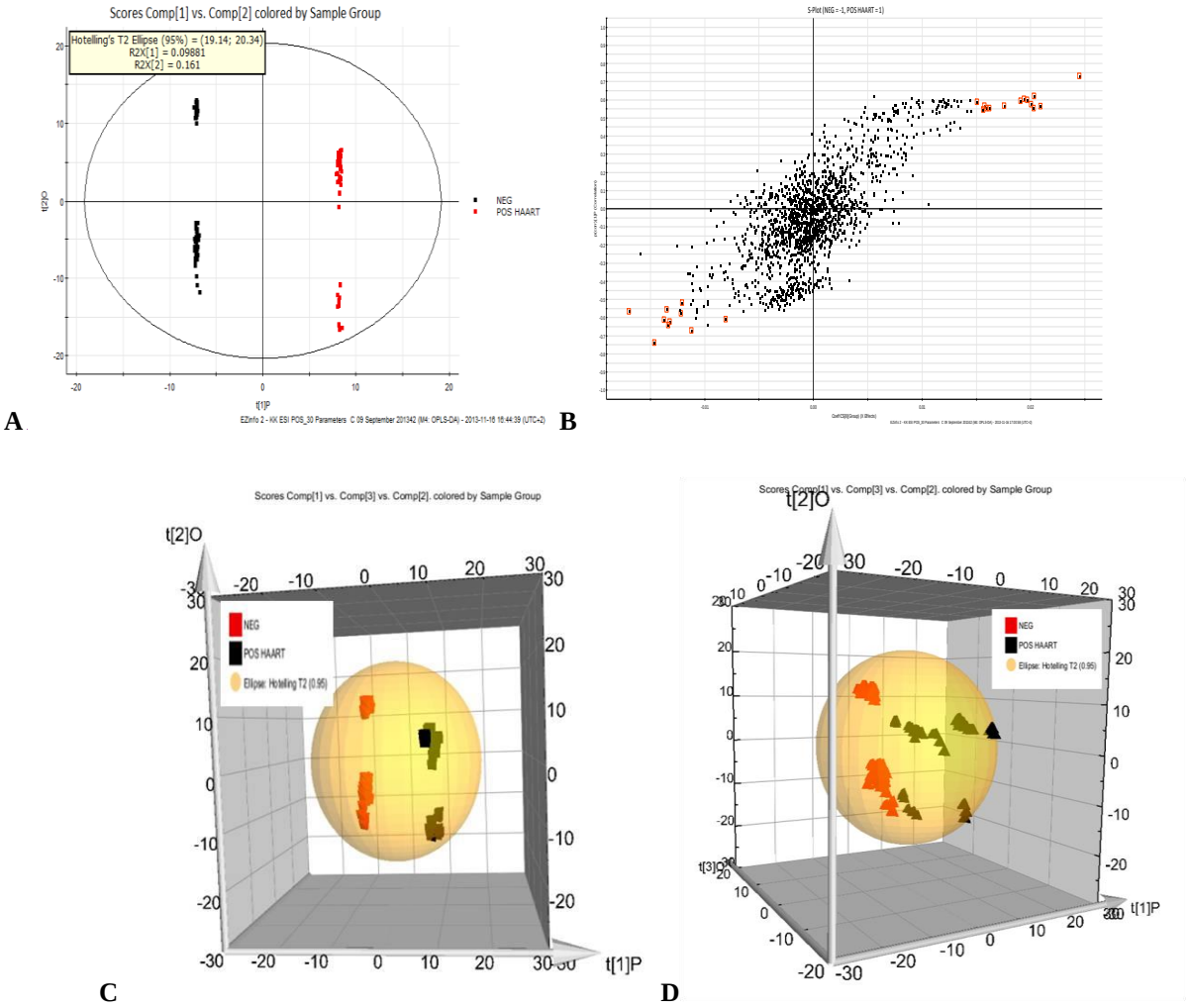


Figure 4.6: (A) 2D OPLS-DA scores plot and accompanying (B) S-plot showing metabolites contributing to the group differences of HIV sera samples analyzed in ESI positive mode of UPLC-MS. NEG and POS HAART group showing between and within group differences with observations appearing to be within the Hotelling's T2 ellipse (95% confidence interval). Scores plot showing 9.88% and 16.10% of the variation as explained by the first and second principal component respectively, of an eight-component model. Model validation was R^2Y (cum) (>0.999) and Q^2 (cum) (0.982). Selected metabolites (potential biomarkers) that were significantly altered ($p<0.05$) shown by red squares on the S-plot extremes. C and D show 3D OPLS-DA scores plots. (C) NEG and POS HAART group separable as viewed from the 3D angle that is depicted in the 2D scores plot, showing tight clustering between and within group differences. (D) NEG and POS HAART group showing a good between and within group separation when viewed at a different angle in a 3D scores plot. Outliers from the POS HAART group outside the Hotelling's T2 ellipse. Heterogeneity of samples clearly visible.

Figure A6 (see appendix) shows the two types of the outlier models (Hotelling's T2 ellipse and DModX) before and after the removal of outliers. After the removal of some outliers, the Hotelling's T2 ellipse falls away from the 3D OPLS-DA scores plot (figure 4.7) and 2D heterogeneity of serum samples cluster is improved. Between and within group heterogeneity of the sera samples now more visible in both the 2D and 3D scores as is normally expected with human samples. The total number of principal components was reduced to two after the outliers were removed for all OPLS-DA scores plots in this study. Model validation parameters of the PCA and OPLS-DA scores plots that are indicative of the goodness of fit values and predictability of the models are provided in table 4.4.

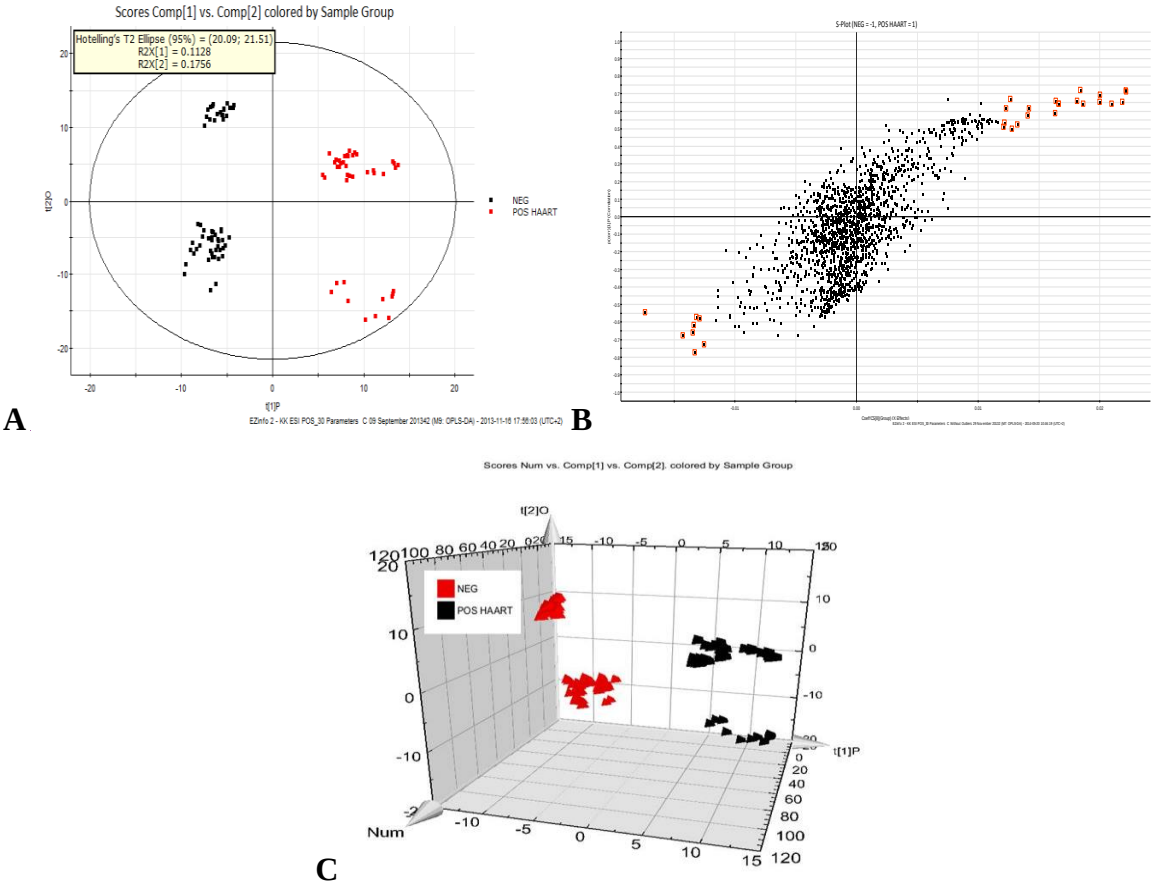


Figure 4.7: (A) 2D OPLS-DA scores plot and accompanying (B) S-plot showing group differences of HIV sera samples analyzed in ESI positive mode of UPLC-MS and metabolites contributing to the separation observed between the NEG and POS HAART group. Observations appearing to be within the Hotelling's T2 ellipse (95% confidence interval) with heterogeneity of samples clearly visible. Scores plot showing 11.28% and 17.56% of the variation as explained by the first and second principal component respectively of a two component model. Model validation was R^2Y (cum) (>0.999) and Q^2 (cum) (0.917). Selected metabolites (potential biomarkers) that were significant ($p<0.05$) shown by red squares on the S-plot extremes. (C) 3D OPLS-DA scores plots. NEG and POSHAART group clearly showing between and within group differences, heterogeneity of clustered samples visible. Removal of outliers made the Hotelling's T2 ellipse fall away.

Table 4.4: Model validation parameters for PCA and OPLS-DA based on ESI negative and ESI positive data of HIV serum samples analyzed using UPLC- MS.

Ionization mode	**Group comparisons**	**Model**	**Number of principal components**		**R^2X(cum)**		**R^2Y(cum)**		**Q^2(cum)**	
Outliers			With	Without	With	Without	With	Without	With	Without
ESI negative	All groups	PCA	21	20	0.604	0.591	-	-	0.412	0.390
	NEG and POS	OPLS-DA	5	2	-	-	0.990	0.990	0.924	0.856
	NEG and POS HAART	OPLS-DA	11	2	-	-	>0.999	>0.999	0.983	0.983
	POS and POS HAART	OPLS-DA	12	2	-	-	>0.999	>0.999	0.952	0.852
ESI positive	All groups	PCA	20	19	0.630	0.607	-	-	0.459	0.460
	NEG and POS	OPLS-DA	8	2	-	-	0.999	0.999	0.907	0.787
	NEG and POS HAART	OPLS-DA	8	2	-	-	>0.999	>0.999	0.982	0.917
	POS and POS HAART	OPLS-DA	16	2	-	-	1.000	1.000	0.976	0.893

Relative quantification of the OPLS-DA metabolites was calculated using variable average statistical tool. Shown in figure 4.8 is a variable average plot of the NEG and POS HAART group differentiating metabolites. The relative levels varied between the two groups. The metabolites that appear to have lower than detectable levels in one group and visible levels in the other group are unique metabolites that differentiate the two groups. They are only found in one group when compared to the other group. These are metabolites indicative of HAART induced metabolic complications. The importance of these metabolites is shown in the variable importance projection (VIP) plots in figure 4.9. VIP defines the importance parameter of each variable and the average value of the squared VIP is one (Axelson, 2010). Majority of the group distinghuishing metabolite had VIP values greater than one (>1) which correspond to informative regions. Both figure 4.8 and 4.9 show how adenosine levels drastically differ between the two groups. HAART treatment is attributed to the high increase observed between the two groups. These plots add more information that support the major metabolic alterations caused by HAART in this study subjects.

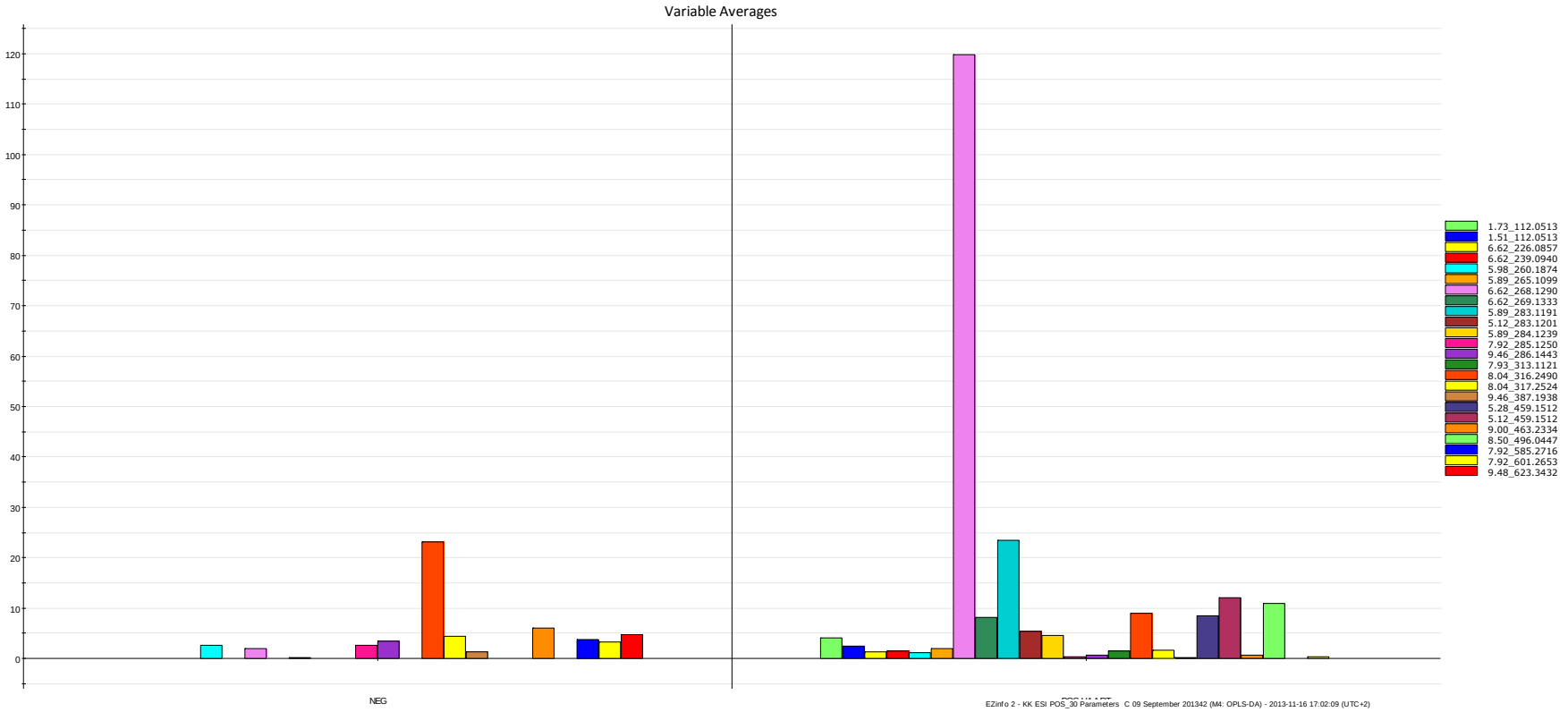


Figure 4.8: Variable averages of group differentiating (NEG and POS HAART) HIV serum metabolites detected in ESI positive mode of UPLC-MS. Relative levels of these metabolites varied between the two groups. The metabolite indicated by a pink colored bar eluted at time 6.62 with m/z ratio of 268.1290 highly increased in the POS HAART group and compared to low levels in the NEG group. This metabolite, adenosine has an extremely high factor of change between the two groups. The POS HAART group contained more metabolites that differentiate it from the NEG group.

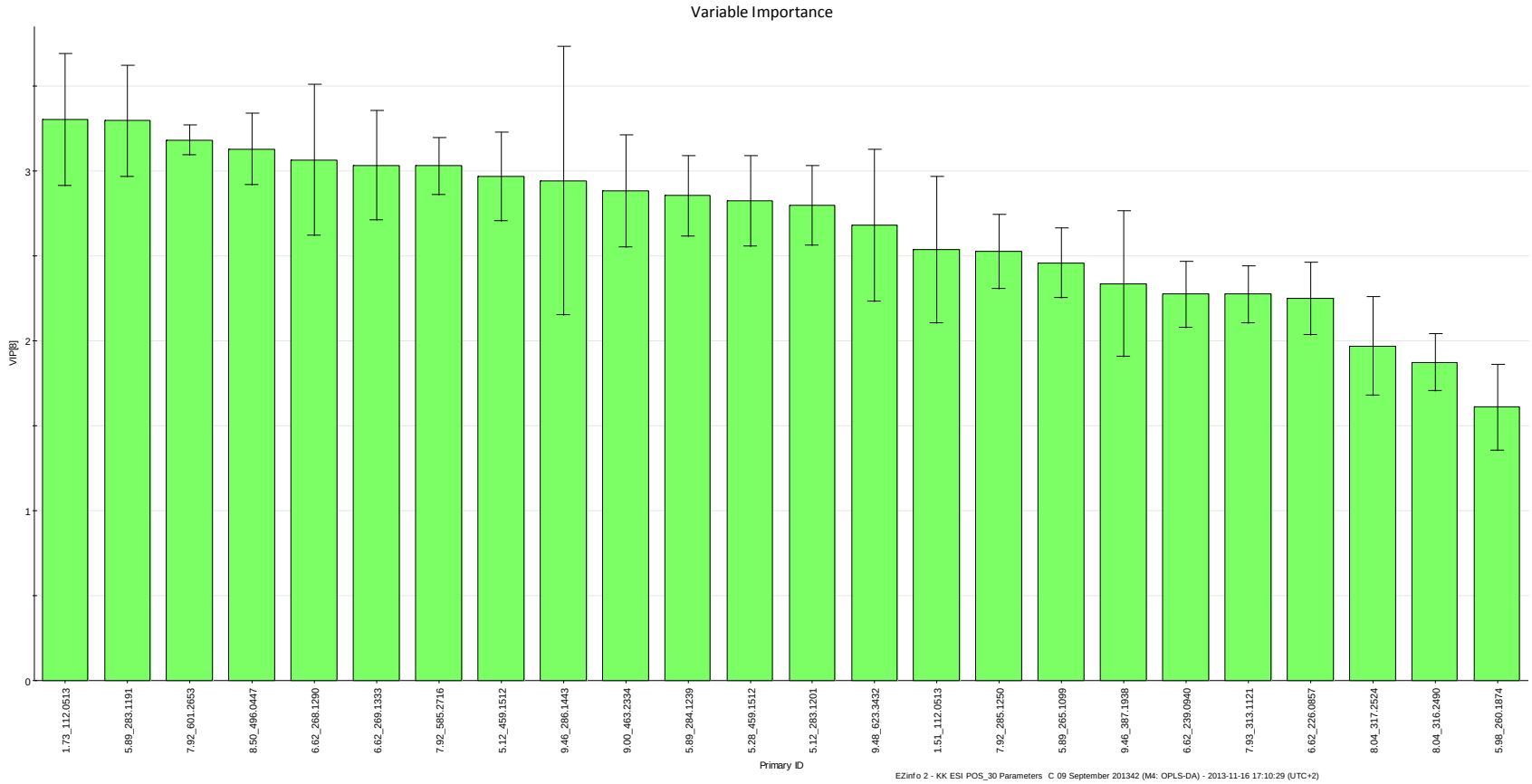


Figure 4.9: Variable importance projection (VIP) plot of group differentiating (NEG and POS HAART) HIV serum metabolites detected in ESI positive mode of UPLC-MS. All metabolite VIP values greater than one. Some metabolites such as Adenosine (6.62_268.1290), showing large error bars between the means of the two groups. This is clearly indicative of how largely this metabolite differs between the groups. Again, this is indicative of how HAART treatment drastically alters the metabolic profile.

The significance of these altered metabolites was confirmed by calculated ANOVA values ($p<0.05$). Table 4.5 (4.5.1-4.5.6) lists the putatively identified metabolites, the identification characteristics and relative quantification levels before and after outlier removal. Of the 116 OPLS-DA significant metabolites that were selected (see table 4.5), only 79 were distinct because; with some metabolites, the same metabolite would be elute at different times and other metabolites were common for discriminating two different pairs (e.g. a metabolite distinguishing the NEG and POS group would also be distinguishing the POS and POS HAART group), these were the overlapping metabolites among different groups and finally metabolites (like alpha-ribazole) were detected in both the ESI negative and ESI positive mode. These 79 (38 in ESI- mode and 41 in ESI+ mode) metabolites are listed in Table 4.6 (4.6.1-4.6.4) together with their biological classification and functions or pathways they're involved in, listed in the order of the respective pair they distinguish. Forty percent (10/25) of the metabolites detected by step-wise LDA were also detected by OPLS-DA as differentiating metabolites. A total of 94 metabolites were detected as different in batch two as determined by LDA and OPLS-DA.

Combined with the manuscript's LDA metabolites (21 from batch one), which had two metabolites that were similarily detected in batch two through OPLS-DA, it adds up to 112 (94 + 18) distinct metabolites from both batches. Overall, the statistical models used in this study clearly showed that treatment caused most of the alterations observed in the metabolic profiles of the study subjects.

Table 4.5: Identification characteristics, relative levels and factors of change of significantly altered sera metabolites detected in (A) ESI positive mode and (B) ESI negative mode of UPLC-MS.

Table 4.5.1A: HIV Negative and HIV Positive group differentiating metabolites (19 metabolites)

Mass-to-charge ratio	**Retention Time**	**Metabolite ID**	**Common Name**	**Elementary composition**	**p-value**	**Relative levels**		**Factor of Change**	
						Negative group	Positive group	With Outliers	Without Outliers
120.0809	5.6431	Threonine	Threonine	$C_4H_9NO_3$	0.003	High	Low	2.00	2.17
120.0814	5.9149	Threonine	Threonine	$C_4H_9NO_3$	0.007	High	Low	1.52	1.59
138.0668	4.8192	3-Pyridinylacetic acid	3-Pyridylacetic acid	$C_7H_7NO_2$	0.032	High	Low	3.17	8.60
181.0730	3.8716	2,5-Dichloro-2,5-cyclohexadiene-1,4-diol	N/A	$C_6H_6Cl_2O_2$	0.018	High	Low	2.02	3.45
195.0883	4.8212	Hex-2-ulosonic acid	2-keto-D-gluconic acid	$C_6H_{10}O_7$	0.029	High	Low	2.95	6.80
196.0923	4.8189	2-Amino-6-(hydroxymethyl)-7,8-dihydro-4(1H)-pteridinone	6-hydroxymethyl-7,8-dihydro-pterin	$C_7H_9N_5O_2$	0.034	High	Low	3.14	9.40
251.1298	15.5841	gamma-Glutamylcysteine	L-γ-glutamylcysteine	$C_8H_{14}N_2O_5S$	0.002	High	Low	3.83	4.70
279.1716	5.6467	5,6-Dimethyl-1-pentofuranosyl-1H-benzimidazole	Alpha-Ribazole	$C_{14}H_{18}N_2O_4$	0.003	High	Low	1.94	2.04
286.1442	9.3273	4,4'-(1H-1,2,4-Triazol-1-ylmethylene)dibenzonitrile	Letrozole	$C_{17}H_{11}N_5$	0.016	High	Low	2.90	4.50

286.1443	9.4641	4,4'-(1H-1,2,4-Triazol-1-ylmethylene)dibenzonitrile	Letrozole	$C_{17}H_{11}N_5$	0.000	High	Low	5.40	13.70
286.2019	6.6968	4,4'-(1H-1,2,4-Triazol-1-ylmethylene)dibenzonitrile	Letrozole	$C_{17}H_{11}N_5$	0.010	Low	High	2.80	2.0
287.2039	6.6950	3-(Hydroxymethyl)phenyl hexopyranoside	Salicin	$C_{13}H_{18}O_7$	0.004	Low	High	4.3	3.9
303.2329	11.0808	2-[5-(Nonyloxy)-1H-indol-3-yl]ethanamine	5-nonyloxytryptamine	$C_{19}H_{30}N_2O$	0.022	Low	High	5.20	5.25
310.2011	7.0137	5-Acetamido-3,5-dideoxy-6-(1,2,3-trihydroxypropyl)hex-2-ulopyranosonic acid	Sialic acid	$C_{11}H_{19}NO_9$	0.048	Low	High	5.24	5.27
310.2015	7.2046	5-Acetamido-3,5-dideoxy-6-(1,2,3-trihydroxypropyl)hex-2-ulopyranosonic acid	Sialic acid	$C_{11}H_{19}NO_9$	0.001	Low	High	4.30	4.06
313.1550	5.9168	12,22-Dihydro-2H,3H-porphine	Porphyrin-ring	$C_{20}H_{16}N_4$	0.008	High	Low	1.51	1.59
373.2746	8.8619	3-[2-Acetyl-4-(butyrylamino)phenoxy]-2-hydroxy-N-isopropyl-1-propanaminium chloride	Acebutolol hydrochoride	$C_{18}H_{29}CLN_2O_4$	0.042	Low	High	7.70	17.90

463.2334	9.001	4-{[(11beta)-11,17-Dihydroxy-3,20-dioxopregn-4-en-21-yl]oxy}-4-oxobutanoic acid	Cortisol succinate	$C_{25}H_{34}O_8$	0.007	High	Low	6.30	5.72
585.2716	7.9164	3-[2-[[3-(2-carboxyethyl)-4-methyl-5-[(4-methyl-5-oxo-3-vinyl-pyrrol-2-ylidene)methyl]-1H-pyrrol-2-yl]methyl]-4-methyl-5-[(3-methyl-5-oxo-4-vinyl-pyrrol-2-ylidene)methyl]-1H-pyrrol-3-yl]propanoic acid	Bilirubin	$C_{33}H_{36}N_{34}O_6$	0.002	High	Low	15.3	92.75
623.3432	9.4758	Rodiasine	Rodiasine	$C_{38}H_4N_2O_6$	0.040	High	Low	248.20	208.36
633.2542	8.6737	N-[4-(4-Morpholinylcarbonyl)benzoyl]-L-valyl-N-[(3S)-5,5,6,6,6-pentafluoro-2-methyl-4-oxo-3-hexanyl]-L-prolinamide	N/A	$C_{33}H_{37}F_5N_4O_6$	0.005	Low	High	4.00	4.00

Table 4.5.2A: HIV Negative and POS HAART group differentiating metabolites (28 metabolites)

Mass-to-charge ratio	Retention Time	Metabolite ID	Common Name	Elementary Composition	p-value	Relative levels		Factor of Change	
						Negative group	POS HAART group	With Outliers	Without Outliers
112.0513	1.5116	6-Amino-2(1H)-pyrimidinone	Cytosine	$C_4H_{11}NO_5$	0.002	Low	High	480.90	262.20
112.0513	1.7251	6-Amino-2(1H)-pyrimidinone	Cytosine	$C_4H_{11}NO_5$	0.002	Low	High	365.30	302.30
226.0857	6.6230	9-[(2-Hydroxyethoxy)methyl]-2-imino-3,9-dihydro-2H-purin-6-ol	Aciclovir	$C_8H_{11}N_5O_3$	0.001	Low	High	453.70	474.66
239.0940	6.6214	2,3-Dihydroxy-3-(1H-imidazol-5-yl)propyl dihydrogen phosphate	D-erythro-imidazole-glycerol-phosphate	$C_6H_{11}N_2O_6P$	0.001	Low	High	-3.47^{8}	-8.20^{7}
260.1874	5.9814	N~4~-(6-Methoxy-8-quinolinyl)-1,4-pentanediamine	Primaquine	$C_{15}H_{21}N_3O$	0.005	High	Low	2.23	1.92
265.1099	5.8863	3,8-Dimethyl-7-decen-1-yl dihydrogen phosphate	Dolichyl/dolichol phosphate	$C_{12}H_{25}O_4P$	0.001	Low	High	-5.67^{8}	N/A
268.1290	6.6208	9-Pentofuranosyl-9H-purin-6-amine	Adenosine	$C_{10}H_{13}N_50_4$	0.001	Low	High	58.10	61.80

269.1333	6.6212	9-Pentofuranosyl-3,9-dihydro-6H-purin-6-one	Inosine	$C_{10}H_{12}N_4O_5$	0.001	Low	High	89.70	93.90
283.1191	5.8869	3,6,9,12,15-Pentaoxaheptadecane-1,17-diol	Hexaethyleneglycol	$C_{12}H_{26}O_7$	0.001	Low	High	108.70	118.80
283.1201	5.1219	3,6,9,12,15-Pentaoxaheptadecane-1,17-diol	Hexaethyleneglycol	$C_{12}H_{26}O_7$	0.001	Low	High	621.90	668.65
284.1239	5.8871	2-Amino-9-pentofuranosyl-3,9-dihydro-6H-purin-6-one	Guanosine	$C_{10}H_{13}N_5O_5$	0.001	Low	High	427.30	464.11
285.1250	7.9207	3-Pentofuranosyl-3,4,7,8-tetrahydroimidazo[4,5-d][1,3]diazepin-8-ol	Coformycin	$C_{11}H_{16}N_4O_5$	0.000	High	Low	6.10	6.10
286.1443	9.4641	4,4'-(1H-1,2,4-Triazol-1-ylmethylene)dibenzonitrile	Letrozole	$C_{17}H_{11}N_5$	0.000	High	Low	4.90	3.70
287.2018	9.001	3-(Hydroxymethyl)phenyl hexopyranoside	Salicin	$C_{13}H_{18}O_7$	0.006	Low	High	10.50	23.50
303.2329	11.0808	2-[5-(Nonyloxy)-1H-indol-3-yl]ethanamine	5-nonyloxytrptamine	$C_{19}H_{30}N_2O$	0.022	Low	High	6.50	7.60
313.1121	7.9274	12,22-Dihydro-2H,3H-porphine	Porphyrin-ring	$C_{20}H_{16}N_4$	0.001	Low	High	593.90	544.69
316.2490	8.0365	Cephalotaxine	Cephalotaxine	$C_{18}H_{21}NO_4$	0.001	High	Low	2.58	2.41
317.2524	8.0361	3-Hydroxypregn-5-en-20-one	Pregnenolone	$C_{21}H_{32}O_2$	0.001	High	Low	2.73	2.51

387.1938	9.4627	2-Amino-6-({4,5-dihydroxy-3-oxo-6-[(phosphonooxy)methyl]tetrahydro-2H-pyran-2-yl}amino)hexanoic acid (non-preferred name)	Fructoselysine-6-phosphate	$C_{12}H_{23}N_2O_{10}P$	0.000	High	Low	8.50	8.90
459.1512	5.1199	1-Deoxy-1-(7,8-dimethyl-2,4-dioxo-1,3,4,5-tetrahydrobenzo[g]pteridin-10(2H)-yl)-5-O-phosphonopentitol	FMNH2/ FMNH	$C_{17}H_{23}N_4O_9P$	0.001	Low	High	1587.30	1583.91
459.1512	5.2754	1-Deoxy-1-(7,8-dimethyl-2,4-dioxo-1,3,4,5-tetrahydrobenzo[g]pteridin-10(2H)-yl)-5-O-phosphonopentitol	FMNH2/ FMNH	$C_{17}H_{23}N_4O_9P$	0.001	Low	High	524.00	592.36
463.2334	9.0010	4-{[(11beta)-11,17-Dihydroxy-3,20-dioxopregn-4-en-21-yl]oxy}-4-oxobutanoic acid	a) Cortisol succinate b) Hydrocortisone succinate	$C_{25}H_{34}O_8$	0.007	High	Low	9.40	16.40
480.067	8.5001	9-(2-Carboxy-4-thiocyanatophenyl)-6-(dimethylamino)-N,N-dimethyl-3H-xanthen-3-iminium chloride	Tetramethylrodamine thiocynate	$C_{25}H_{22}CLN_3OS$	0.001	Low	High	398.50	414.40

482.0694	8.1549	4-Amino-1-[2-deoxy-5-O-(hydroxy{[hydroxy(phosphonooxy)phosphoryl]oxy}phosphoryl)pentofuranosyl]-5-methyl-2(1H)-pyrimidinone	5-methyl-dCTP	$C_{10}H_{18}N_3O_3P_3$	0.001	Low	High	526.10	515.00
496.0447	8.5001	Triphosphoric acid, mono[3-(2-amino-1,4,7,8-tetrahydro-4-oxo-6-pteridinyl)-2,3-dihydroxypropyl] ester	Dihydronepterin triphosphate	$C_9H_{16}N_5O_{13}P_3$	0.008	Low	High	197.90	113.22
509.0668	8.1308	9-[5-O-(Hydroxy{[hydroxy(phosphonooxy)phosphoryl]oxy}phosphoryl)pentofuranosyl]-3,9-dihydro-6H-purin-6-one	Inosine triphosphate (ITP)	$C_{10}H_{15}N_4O_{14}P_3$	0.001	Low	High	987.30	279.70
521.408	12.1312	(2S)-3-(alpha-D-Galactopyranosyloxy)-2-hydroxypropyl stearate	1-oleoyl-3-α-D-galactosyl-sn-glycerol	$C_{27}H_{52}O_9$	0.001	High	Low	6.40	9.60
585.2716	7.9164	3-[2-[[3-(2-carboxyethyl)-4-methyl-5-[(4-methyl-5-oxo-3-vinyl-pyrrol-2-ylidene)methyl]-1H-pyrrol-2-yl]methyl]-4-methyl-5-[(3-methyl-5-oxo-4-vinyl-pyrrol-2-ylidene)methyl]-1H-pyrrol-3-yl]propanoic acid	Bilirubin	$C_{33}H_{36}N_4O_6$	0.002	High	Low	1435.00	1076.20

601.2653	7.9186	Ethyl (2R,4S)-4-{[3,5-bis(trifluoromethyl)benzyl](methoxycarbonyl)amino}-2-ethyl-6-(trifluoromethyl)-3,4-dihydro-1(2H)-quinolinecarboxylate	Torcetrapib	$C_{26}H_{25}F_9N_2O_4$	0.000	High	Low	10.50	9.70
623.3432	9.4758	Rodiasine	Rodiasine	$C_{38}H_4N_2O_6$	0.040	High	Low	311.10	233.34

Table 4.5.3A: HIV Positive and POS HAART group differentiating metabolites (20 metabolites)

Mass-to-charge ratio	Retention Time	Metabolite ID	Common Name	Elementary Composition	p-value	Relative levels		Factor of Change	
						Positive group	POS HAART group	With Outliers	Without Outliers
112.0513	1.5116	6-Amino-2(1H)-pyrimidinone	Cytosine	$C_4H_{11}NO_5$	0.002	Low	High	-9.96[8]	-2.87[8]
112.0513	1.7251	6-Amino-2(1H)-pyrimidinone	Cytosine	$C_4H_{11}NO_5$	0.002	Low	High	-2.67[8]	-1.07[8]
160.0768	2.0525	1H-Indol-3-ylacetaldehyde	a) Tryptaldehyde b) Indole acetaldehyde	$C_{10}H_9NO$	0.009	High	Low	2.02	2.21
196.0923	4.8189	2-Amino-6-(hydroxymethyl)-7,8-dihydro-4(3H)-pteridinone	6-hydroxymethyl-7,8-dihydropterin	$C_7H_9N_5O_2$	0.034	Low	High	2.51	5.40
226.0857	6.6230	4-Amino-3-[(1-carboxyvinyl)oxy]-1,5-cyclohexadiene-1-carboxylic acid	a) ADC b) 4-amino-4-deoxychorismate	$C_{10}H_{11}NO_5$	0.001	Low	High	197.48	198.70
230.0608	1.7179	5-O-Phosphonopentofuranosylamine	5-Phosphoribosylamine	$C_5H_{12}NO_7P$	0.015	High	Low	3.18[7]	5.18[7]
230.0612	1.5081	5-O-Phosphonopentofuranosylamine	5-Phosphoribosylamine	$C_5H_{12}NO_7P$	0.016	High	Low	1.39[7]	1.5[7]
239.0940	6.6214	2,3-Dihydroxy-3-(1H-imidazol-5-yl)propyl dihydrogen phosphate	D-erythro-imidazole-glycerol-phosphate	$C_6H_9N_2O_6P$	0.001	Low	High	407.60	410.00
265.1099	5.8863	3,8-Dimethyl-7-decen-1-yl dihydrogen phosphate	Dolichyl/dolichol phosphate	$C_{12}H_{25}O_4P$	0.001	Low	High	170.90	172.00

267.1245	6.6241	2-{4-[2-Hydroxy-3-(isopropylamino)propoxy]phenyl}acetamide	a) Atenolol b) Tenormin	$C_{14}H_{22}N_2O_3$	0.001	Low	High	39.40	36.24
267.3901	6.6198	1-Nonadecene	1-nonadecene	$C_{19}H_{38}$	0.001	Low	High	141.54	142.39
268.1290	6.6208	9-Pentofuranosyl-9H-purin-6-amin	Adenosine	$C_{10}H_{13}N_50_4$	0.001	Low	High	37.10	40.30
268.2557	6.6208	1-(Isopropylamino)-3-[4-(2methoxyethyl)phenoxy-2-propanol	Metoprolol	$C_{15}H_{25}NO_3$	0.003	Low	High	639.38	643.19
269.0372	6.6193	9-Pentofuranosyl-3,9-dihydro-6H-purin-6-one	Inosine	$C_{10}H_{12}N_4O_5$	0.004	Low	High	1.47^7	1.60^7
269.1333	6.6212	9-Pentofuranosyl-3,9-dihydro-6H-purin-6-one	Inosine	$C_{10}H_{12}N_4O_5$	0.001	Low	High	46.5	50.30
283.1191	5.8869	3,6,9,12,15-Pentaoxaheptadecane-1,17-diol	Hexaethyleneglycol	$C_{12}H_{26}O_7$	0.001	Low	High	53.90	58.90
283.1201	5.1219	3,6,9,12,15-Pentaoxaheptadecane-1,17-diol	Hexaethyleneglycol	$C_{12}H_{26}O_7$	0.001	Low	High	415.40	417.80
284.1239	5.8871	2-Amino-9-pentofuranosyl-3,9-dihydro-6H-purin-6-one	Guanosine	$C_{10}H_{13}N_5O_5$	0.001	Low	High	141.90	142.80
286.2019	6.6968	4,4'-(1H-1,2,4-Triazol-1-ylmethylene)dibenzonitrile	Letrozole	$C_{17}H_{11}N_5$	0.010	High	Low	1.60	1.49
313.1121	7.9274	12,22-Dihydro-2H,3H-porphine	Porphyrin-ring	$C_{20}H_{16}N_4$	0.001	Low	High	211.40	212.70
317.2524	8.0361	3-Hydroxypregn-5-en-20-one	Pregnenolone	$C_{21}H_{32}O_2$	0.001	High	Low	1.26	1.33
459.1512	5.1199	1-Deoxy-1-(7,8-dimethyl-2,4-dioxo-1,3,4,5-tetrahydrobenzo[g]pteridin-10(2H)-yl)-5-O-phosphonopentitol	$FMNH_2$	$C_{17}H_{23}N_4O_9P$	0.001	Low	High	497.20	500.20

459.1512	5.2754	1-Deoxy-1-(7,8-dimethyl-2,4-dioxo-1,3,4,5-tetrahydrobenzo[g]pteridin-10(2H)-yl)-5-O-phosphonopentitol	$FMNH_2$	$C_{17}H_{23}N_4O_9P$	0.001	Low	High	370.90	373.10
496.0447	8.5001	Dihydronepterin triphosphate	Dihydroneopterin triphosphate	$C_9H_{16}N_5O_{13}P_3$	0.008	Low	High	125.40	123.20

Table 4.5.4B: HIV Negative and HIV Positive group differentiating metabolites (16 metabolites)

Mass-to-charge ratio	Retention Time	Metabolite ID	Common Name	Elementary Composition	P-value	Relative levels		Factor of Change	
						Negative group	Positive group	With Outliers	Without Outliers
179.0717	8.4975	Hexopyranose	Alpha-D-Glucose	$C_6H_{12}O_6$	0.000	High	Low	8.90	6.04
199.0058	6.4418	4,6-Dioxo-2-octenedioic acid	Fumarylacetoacetate	$C_8H_8O_6$	0.012	Low	High	6.10	6.08
277.1560	5.6543	5,6-Dimethyl-1-pentofuranosyl-1H-benzimidazol	Alpha-Ribazole	$C_{30}H_{50}O_2$	0.000	High	Low	2.44	2.08
441.3580	12.5704	Lanosta-8,24-diene-3,30-diol	a) Obtusifoliol	$C_{30}H_{50}O_2$	0.000	High	Low	8.10	7.60
457.3530	12.4958	(1R,3aS,5aR,5bR,7aR,9S,11aR,11bR,13aR,13bS)-1-Acetyl-9-hydroxy-5a,5b,8,8,11a-pentamethylicosahydro-3aH-cyclopenta[a]chrysene-3a-carboxylic acid	a) Triterpenoid acid b) Platanic acid	$C_{29}H_{46}O_4$	0.000	High	Low	5.40	4.70
465.3588	12.7820	(3beta)-Cholest-5-en-3-yl hydrogen sulfate	Cholesterol sulfate	$C_{27}H_{46}O_4S$	0.000	High	Low	35.80	30.50
467.3740	14.4183	3,11-Dioxoolean-12-en-30-oic acid	3-Oxoglycyrrhetinate	$C_{30}H_{44}O_4$	0.000	High	Low	2.42	2.17
481.3523	11.9215	Decyl 4-O-alpha-D-glucopyranosyl-beta-D-glucopyranoside	Decyl-beta-D-maltopyranoside	$C_{22}H_{42}O_{11}$	0.000	High	Low	3.20	2.62

485.3843	12.8337	(3beta,16alpha)-3,16-Dihydroxy-23-oxoolean-12-en-28-oic acid	a)Quallaja Sapogenin b)Quallaic acid	$C_{30}H_{46}O_5$	0.000	High	Low	4.40	3.90
489.3556	14.4205	(3beta,6alpha,9beta,16beta,20R,24S)-20,24-Epoxy-9,19-cyclolanostane-3,6,16,25-tetrol	Cycloastragenol	$C_{30}H_{50}O_5$	0.000	High	Low	3.32	2.96
507.2237	9.0133	(6S)-1-[4-(Dimethylamino)-3-methylbenzyl]-5-(diphenylacetyl)-4,5,6,7-tetrahydro-1H-imidazo[4,5-c]pyridine-6-carboxylic acid	PD 123319	$C_{31}H_{32}N_4O_3$	0.003	High	Low	6.40	7.10
529.2680	529.2680	Methyl (2S,3R,4S)-2-(beta-D-glucopyranosyloxy)-4-[(1S)-2,3,4,9-tetrahydro-1H-beta-carbolin-1-ylmethyl]-3-vinyl-3,4-dihydro-2H-pyran-5-carboxylate	Strictosidine	$C_{27}H_{34}N_2O_9$	0.005	Low	High	10.20	12.19
539.4312	14.6135	7,7',8,8'-Tetrahydro-psi,psi-carotene	Zeta-carotene	$C_{40}H_{60}$	0.000	High	Low	5.70	10.10
539.4302	14.9804	7,7',8,8'-Tetrahydro-psi,psi-carotene	Zeta-carotene	$C_{40}H_{60}$	0.000	High	Low	10.60	5.00
540.4349	14.6128	17-[(1E)-1-(1-Acetyl-5-methyl-2-piperidinylidene)ethyl]androstane-3,16-diyl diacetate	Triacetyl pseudotomatidine	$C_{33}H_{51}NO_5$	0.000	High	Low	8.30	7.40

559.3984	14.4275	(1Z)-N-(5-{[(4Z)-4-({5-[Acetyl(hydroxy)amino]pentyl}imino)-4-hydroxybutanoyl](hydroxy)amino}pentyl)-4-[(5-aminopentyl)(hydroxy)amino]-4-oxobutanimidic acid	Deferoxamine	$C_{25}H_{48}N_6O_8$	0.000	High	Low	3.93	3.69
621.3264	9.4923	Rodiasine	Rodiasine	$C_{38}H_{48}N_2O_6$	0.034	High	Low	134.80	209.40

Table 4.5.5B: HIV Negative and POS HAART group differentiating metabolites (20 metabolites)

Mass-to-charge ratio	Retention Time	Metabolite ID	Common Name	Elementary Composition	p-value	Relative levels		Factor of Change	
						Negative group	POS HAART group	With Outliers	Without Outliers
293.1758	9.4423	5-Methyl-2-[(5-methyl-1H-imidazol-4-yl)methyl]-2,3,4,5-tetrahydro-1H-pyrido[4,3-b]indol-1-one	Alosetron	$C_{17}H_{18}O_{40}$	0.025	Low	High	3.50	2.72
319.2277	11.0968	2-(3,7-Dimethyl-2,6-octadien-1-yl)-5,6-dimethoxy-3-methyl-1,4-benzenediol	Ubiquinol-2	$C_{19}H_{28}O_4$	0.012	Low	High	4.49	5.7
320.2319	11.0976	2-(Acetoxymethyl)-4-(2-amino-9H-purin-9-yl)butyl acetate	Famciclovir	$C_{14}H_{19}N_50_4$	0.013	Low	High	7.34	7.40
369.1735	8.2065	17-Oxoandrostan-3-yl hydrogen sulfate	DHEA sulphate	$C_{19}H_{30}O_5S$	0.000	High	Low	9.80	8.20
369.1740	8.4428	17-Oxoandrostan-3-yl hydrogen sulfate	DHEA sulphate	$C_{19}H_{30}O_5S$	0.041	High	Low	10.10	8.60
370.1775	8.2062	2-[4-(1,2-Diphenyl-1-buten-1-yl)phenoxy]-N,N-dimethylethanamine	Tamoxifen	$C_{26}H_{29}NO$	0.002	High	Low	24.50	20.60
370.1783	8.4424	2-[4-(1,2-Diphenyl-1-buten-1-yl)phenoxy]-N,N-dimethylethanamine	Tamoxifen	$C_{26}H_{29}NO$	0.001	High	Low	15.60	12.80

385.1685	6.2906	2-Amino-6-({4,5-dihydroxy-3-oxo-6-[(phosphonooxy)methyl]tetrahydro-2H-pyran-2-yl}amino)hexanoic acid (non-preferred name)	Fructoselysine-6-phosphate	$C_{12}H_{23}N_2O_{16}P$	0.011	Low	High	7.70	7.50
441.3580	12.5704	Lanosta-8,24-diene-3,30-diol	a) Obtusifoliol	$C_{30}H_{50}O_2$	0.000	High	Low	8.00	7.90
457.1357	5.2842	1-Deoxy-1-(7,8-dimethyl-2,4-dioxo-1,3,4,5-tetrahydrobenzo[g]pteridin-10(2H)-yl)-5-O-phosphonopentitol	a)Reduced FMN b)$FMNH_2$	$C_7H_{23}N_4O_9P$	0.001	Low	High	1663.10	1972.60
457.1360	5.1289	1-Deoxy-1-(7,8-dimethyl-2,4-dioxo-1,3,4,5-tetrahydrobenzo[g]pteridin-10(2H)-yl)-5-O-phosphonopentitol	a)Reduced FMN b)$FMNH_2$	$C_7H_{23}N_4O_9P$	0.001	Low	High	259.70	308.00
457.3530	12.4958	(1R,3aS,5aR,5bR,7aR,9S,11aR,11bR,13aR,13bS)-1-Acetyl-9-hydroxy-5a,5b,8,8,11a-pentamethylicosahydro-3aH-cyclopenta[a]chrysene-3a-carboxylic acid	Platanic acid	$C_{29}H_{46}O_4$	0.000	High	Low	4.40	4.30
463.3411	14.0821	(2beta,3beta,5beta,22R)-2,3,14,22,25-Pentahydroxycholest-7-en-6-one	Ecdysone	$C_{27}H_{44}O_6$	0.000	High	Low	5.05	6.60
465.2483	8.2957	(11beta,16beta)-21-Chloro-9-fluoro-11-hydroxy-16-methyl-3,20-dioxopregna-1,4-dien-17-yl propionate	Clobetasol propionate	$C_{25}H_{32}ClFO_5$	0.001	High	Low	8.40	7.20
467.3740	14.4183	3,11-Dioxoolean-12-en-30-oic acid	3-Oxoglycyrrhetinate	$C_{30}H_{44}O_4$	0.000	High	Low	2.95	2.94

485.3843	12.8337	(3beta,16alpha)-3,16-Dihydroxy-23-oxoolean-12-en-28-oic acid	a) Quillaja Sapogenin b) Quallaic acid	$C_{30}H_{46}O_5$	0.000	High	Low	5.80	5.90
507.2237	9.0133	(6S)-1-[4-(Dimethylamino)-3-methylbenzyl]-5-(diphenylacetyl)-4,5,6,7-tetrahydro-1H-imidazo[4,5-c]pyridine-6-carboxylic acid	PD123319	$C_{31}H_{32}N_4O_5$	0.003	High	Low	11.20	10.80
508.2265	9.0148	[6-Hydroxy-2-(4-hydroxyphenyl)-1-benzothiophen-3-yl]{4-[2-(1-piperidinyl)ethoxy]phenyl}methanone hydrochloride (1:1)	Roloxifene hydrochloride	$C_{28}H_{28}CLNO_4S$	0.005	High	Low	13.00	13.20
513.2720	8.0129	4'-[(1,7'-Dimethyl-2'-propyl-1H,3'H-2,5'-bibenzimidazol-3'-yl)methyl]-2-biphenylcarboxylic acid	a) Telmisartan b) Micardis	$C_{33}H_{30}N_4O_2$	0.020	High	Low		
538.4200	14.4252	N-[(2S,3R,4E)-1,3-Dihydroxy-15-methyl-4-hexadecen-2-yl]-2-hydroxyhexadecanamide	N/A	$C_{33}H_{65}N0_4$		High	Low	4.37	4.40
539.4302	14.9804	7,7',8,8'-Tetrahydro-psi,psi-carotene	*Zeta-carotene	$C_{40}H_{60}$	0.000	High	Low	8.80	9.10
539.4312	14.6135	7,7',8,8'-Tetrahydro-psi,psi-carotene	Zeta-carotene	$C_{40}H_{60}$	0.000	High	Low	6.50	6.90
540.4349	14.6128	17-[(1E)-1-(1-Acetyl-5-methyl-2-piperidinylidene)ethyl]androstane-3,16-diyl diacetate	Triacetyl pseudotomitidine	$C_{33}H_{51}NO_5$	0.000	High	Low	8.70	9.80

*Generally unusual to detect zeta carotene with ESI because of lack of oxygen, sulphur and nitrogen. Probably detected due to adduct formation.

Table 4.5.6B: HIV Positive and POS HAART group differentiating metabolites (13 metabolites)

Mass-to-charge ratio	Retention Time	Metabolite ID	Common Name	Elementary Composition	p-value	Relative levels		Factor of Change	
						Positive group	Positive HAART group	With Outliers	Without Outliers
311.1399	5.9219	12,22-Dihydro-2H,3H-porphine	Porphyrin-ring	$C_{20}H_{16}N_4$	0.002	Low	High	1.37	1.83
312.1437	5.9220	N-(1-Amino-2-formamidoethylidene)-5-O-phosphonopentofuranosylamine	5'-phosphoribosyl-a-formylglycine	$C_8H_{11}N_3O_8P$	0.002	Low	High	1.45	1.83
369.1740	8.4428	17-Oxoandrostan-3-yl hydrogen sulfate	DHEA sulphate	$C_{19}H_{30}O_5S$	0.000	High	Low	3.90	7.00
385.1685	6.2906	2-Amino-6-({4,5-dihydroxy-3-oxo-6-[(phosphonooxy)methyl]tetrahydro-2H-pyran-2-yl}amino)hexanoic acid	Fructoselysine-6-phosphate	$C_{12}H_{23}N_2O_{16}P$	0.011	Low	High	3.78	1.77
386.1713	6.2884	N-(1-Benzyl-2-methyl-3-pyrrolidinyl)-5-chloro-2-methoxy-4-(methylamino) benzamide	Nemonapride	$C_{21}H_{26}ClN3O_2$	0.022	Low	High	6.24	2.06
413.1998	7.4572	1-[2-(Diphenylmethoxy)ethyl]-4-(3-phenylpropyl) piperazine	GBR12935	$C_{28}H_{34}N_2O$	0.003	High	Low	3.10	2.88
414.2039	7.4594	2-(Hydroxymethyl)-4-(1-hydroxy-2-{[6-(4-phenylbutoxy)hexyl]amino}ethyl)phenol	Salmeterol	$C_{25}H_{37}NO_4$	0.004	High	Low	5.80	5.20
457.1357	5.2842	1-Deoxy-1-(7,8-dimethyl-2,4-dioxo-1,3,4,5-tetrahydrobenzo[g]pteridin-10(2H)-yl)-5-O-phosphonopentitol	a)Reduced FMN b)$FMNH_2$	$C_7H_{23}N_4O_9P$	0.001	Low	High	711.70	1319.80

457.1360	5.1289	1-Deoxy-1-(7,8-dimethyl-2,4-dioxo-1,3,4,5-tetrahydrobenzo[g]pteridin-10(2H)-yl)-5-O-phosphonopentitol	a)Reduced FMN b)$FMNH_2$	$C_7H_{23}N_4O_9P$	0.001	Low	High	503.90	932.10
601.7728	5.3437	5-Bromo-2-[9-chloro-3-(sulfooxy)naphtho[1,2-b]thiophen-2-yl]-1H-indol-3-yl hydrogen sulfate	SALOR2	$C_{20}H_{11}BrCN_8S$	0.016	High	Low	3.40	2.70
602.2746	5.3465	2,5,8,11,14,17,20,23,26-Nonaoxaoctacosan-28-yl 4-(butylamino)benzoate	Benzonatate	$C_{30}H_{53}NO_{11}$	0.028	High	Low	4.00	2.90
731.3121	5.4467	(1R,2R,1'R,2'R)-2,2'-{[(2S,3R,4R,5S)-3,4-Dimethyltetrahydrofuran-2,5-diyl]bis[(2-methoxy-4,1-phenylene)oxy]}bis[1-(3,4-dimethoxyphenyl)-1-propanol]	Mannasantin A	$C_{42}H_{52}O_{11}$	0.021	High	Low	3.90	2.40
732.3136	5.4459	N-{(2S,5S,8S,13S,16R,17S)-5-[(2S)-2-Butanyl]-8-isobutyl-13-isopropyl-2,10,10,16-tetramethyl-3,6,9,11,14,18-hexaoxo-1,4,12,15-tetraoxa-7-azacyclooctadecan-17-yl}-3-formamido-2-hydroxybenzamide	Cyclodepsipeptide 2	$C_{36}H_{51}N_3O_{13}$	0.032	High	Low	4.90	3.70
807.3127	5.5678	12,12,13,13,14,14,15,15,16,16,17,17,18,18,19,19,19-Heptadecafluorononadecyl di-4-morpholinylphosphinate	11-(perfluorocytyl)undecyl dimorpholinophosphirate	$C_{27}H_{38}F_{17}N_2O_4P$	0.021	High	Low	4.20	5.04

Table 4.6.1: Biological classification of HIV sera metabolites found to be common among compared pairs of experimental groups detected in ESI positive mode of UPLC-MS together with their functions or pathway the compounds are involved in.

Group Difference Comparisons set A: **1) NEG and POS & 2) NEG and POS HAART**		
Metabolite Category	**Metabolite name**	**Functions/ Pathways**
1) Carbohydrates and Carbohydrate Conjugates, Glycosyl Compounds	Salicin/ salicoside	Anti-inflammatory agent
2) Aromatic Heteropolycyclic Compounds, Indoles	5-nonyloxytryptamine	High-Affinity 5-HT1D.beta. Serotonin Receptor Agonist
3) Aromatic Heteropolycyclic Compounds, Tetrapyrroles and Derivatives	Bilirubin	Component of Porphyrin and chlorophyll metabolism
4) cyclic bisbenzylisoquinoline Alkaloids	Rodiasine/ Antioquine	Anti-cancer Calcium channel blockers Anti-parasitic drugs
Group Difference Comparisons set B: **1) NEG and POS & 2) POS and POS HAART**		
5) N/A	6-hydroxymethyl-7,8-dihydro-pterin	Folate biosynthesis pathway and metabolic pathways
Group Difference Comparisons set C: **1) NEG and POS HAART & 2) POS and POS HAART**		
6) Aromatic Heteromonocyclic Compounds, Diazines	Cytosine	Pyrimidine metabolism
7) Aromatic Heteromonocyclic Compounds, Azoles	D-erythro-imidazole-glycerol-phosphate	Intermediate in histidine metabolism Biosynthesis of amino acids and alkaloids from histidine and purine

8) Nucleosides, Nucleotides, and Analogues, Purine Nucleosides and Analogues	Adenosine	Neurotransmitter and potent vasodilator Class V antiarrhythmic agent Plays a role in energy transfer and signal transduction as cyclic adenosine Component of methionine metabolism, purine metabolism and seleno amino acid metabolism
9) Nucleosides, Nucleotides, and Analogues, Purine Nucleosides and Analogues	Inosine	Component of purine metabolism
10) N/A	Hexaethyleneglycol or PEG 300	Plasticizer Lubricant Adhesive
11) Lipids, Steroids and Steroid Derivatives	Pregnenolone	Cell signaling Fuel and energy source and storage Membrane integrity/stability Component of C21-Steroid hormone metabolism
12) Aromatic Heteropolycyclic Compounds, Pteridines and Derivatives	FMNH2	Substrate for FMN reductase
13) Nucleosides, Nucleotides, and Analogues, Purine Nucleosides and Analogues	Guanosine	Neuroprotective and neuromodulator roles in the central nervous system
14) Aromatic Heteropolycyclic Compounds, Pteridines and Derivatives	7,8-Dihydrohepterin triphosphate	Component of folate biosynthesis
Group Difference Comparisons set D: **1) NEG and POS & 2) NEG and POS HAART & 3) POS and POS HAART**		
15) Aromatic Heteropolycyclic Compounds, Diphenylmethanes	Letrozole/Famara	Oral non-steroidal aromatose inhibitor
16) N/A	Porphyrin-ring	N/A

Table 4.6.2: Biological classification of HIV sera group distinguishing metabolites detected in ESI positive mode of UPLC-MS together with their functions or pathways the compounds are involved in.

NEG and POS group differentiating metabolites		
Metabolite Category	**Metabolite name**	**Functions/ Pathways**
17) Aromatic Heteromonocyclic Compounds, Pyridines and Derivatives	3-Pyridylacetic acid	Protein synthesis Amino acid biosynthesis Nicotene breakdown product Part of nicotine degradation pathway
18) Carbohydrates and Carbohydrate Conjugates, Monosaccharaides	2-keto-D-gluconate	Food additive Acidity regulator Strong chelating agent
19) Amino Acids, Peptides, and Analogues, Amino Acids and Derivatives	Threonine	Component of Aminoacyl-tRNA biosynthesis Component of Glycine, serine and threonine metabolism Essential amino acids
20) Amino Acids, Peptides, and Analogues, Amino Acids and Derivatives	γ-Glutamyl-cysteine	Component of glutamate metabolism Component of Glutathione metabolism
21) Nucleosides, Nucleotides, and Analogues, Benzimidazole Nucleosides and Nucleotides	Alpha-Ribazole	Intermediate in riboflavin metabolism
22) Carbohydrates and Carbohydrate Conjugates, Sugar Acids and Derivatives	Sialic acid	Amino acid metabolism
23) Aromatic Homomonocyclic Compounds, Anilides	Acebutolol hydrochloride	Adrenergic beta-Antagonists Anti-Arrhythmia Agents Anti-Hypertensive Agents

24) N/A	2,5-dichloro-2,5-cyclohexadiene-1,4-diol	Chlorocyclohexane and chlorobenzene degradation Metabolic pathways Microbial metabolism in diverse environments
25) Proline derivative	CHEBI: 45517	N/A
NEG and POS HAART group differentiating metabolites		
26) Aromatic Heteropolycyclic Compounds, Imidazopyrimidines	Aciclovir	Antiviral agents Nucleosides and Nucleotides Guanosine analogue acting as an antimetabolite
27) Aromatic Heteropolycyclic Compounds, Quinolines and Derivatives	Primaquine	Antimalarial agents Antiprotozoal agents
28) N/A	Dolichyl/dolichol phosphate	N/A
29) Carbohydrates, *N*-Glycosyl	Coformycin	Nucleoside antibiotic Adenosine deaminase inhibitor
30) Ester derivative, Alkaloids	Cephalotaxine	Anti-tumor
31) Carbohydrates and Carbohydrate Conjugates, Monosaccharides	Fructoselysine-6-phosphate	Component of Aminosugars metabolism Component of Fructose and mannose metabolism Component of Galactose metabolism Component of Glutamate metabolism Component of Starch and sucrose metabolism
32) Hydrocortisone analogues & derivatives	Cortisol succinate	Anti-inflammatory agent Targets HIV protease
33) N/A	Tetrathylrodamine thiocynate	N/A
34) Nucleosides, Nucleotides, and Analogues, Pyrimidine Nucleotides	5-methyl-dCTP	Used in the synthesis of DNA

35) Nucleosides, Nucleotides, and Analogues, Purine Nucleotides	Inosine triphosphate	Intermediate in the purine metabolism pathway
36) N/A	1-oleoyl-3-α-D-galactosyl-sn-glycerol	N/A
37) N/A	Torcetrapib	CETP Inhibitor Anti-cholesterenic drug
POS and POS HAART group differentiating metabolites		
38) Aromatic Heteropolycyclic Compounds, Indoles	Tryptaldehyde	Tryptophan degradation metabolism
39) N/A	ADC/ 4-amino-4-deoxychorismate	N/A
40) Carbohydrates and Carbohydrate Conjugates, Amino Sugars	5-Phosphoribosylamine	Substrate for Aminophosphoribosyltransferase and Trifunctional purine biosynthetic protein adenosine-3
41) Aromatic Homomonocyclic Compounds, Ethers	Atenolol/ Tenormin	beta-1 selective or cardioselective drug
42) N/A	1-nonedecene	N/A
43) N/A	Metroprolol	Selective β_1 receptor blocker

Table 4.6.3: Biological classification of HIV sera metabolites found to be common among compared pairs of experimental groups detected in ESI negative mode of UPLC-MS together with their functions or pathways the compounds are involved in.

Group Difference Comparisons set A: **1) NEG and POS & 2) NEG and POS HAART**		
Metabolite Category	**Metabolite name**	**Functions/ Pathways**
44) Lipids, Prenol Lipids	Obtusifoliol	Intermediate in the biosynthesis of cholesterol Cell signaling Fuel and energy source and storage Hormones, Membrane component Membrane intergrity/stability
45) N/A	Platanic acid	Anti-HIV agent
46) Lipids, Prenol lipids	3-Oxyglycyrrheritinate	N/A
47) Terpenoids, Pentacyclic triterpenoids	Quallaja Sapogenin/ Quallaic acid	Anti-inflammatory agent
48) Lipids, Prenol lipids	Zeta-carotene	Anti-oxidant
49) N/A	Triacetyl pseudotomatidine	N/A
50) N/A	PD 123319	Angiotensin II receptor antagonist
None of the metabolites overlapped between the two groupings; **1) NEG and POS and 2) POS and POS HAART,** set B is absent		
Group Difference Comparisons set C: **1) NEG and POS HAART & POS and POS HAART**		
51) Lipids, Steroids and Steroid Derivatives	DHEA Sulphate	Cell signaling Fuel and energy source and storage Membrane integrity/stability Waste products
52) Carbohydrate derivative, sugar phosphate	Fructoselysine-6-phosphate	N/A
53) Aromatic Heteropolycyclic Compounds, Pteridines and Derivatives	$FMNH_2$	Substrate for FMN reductase Riboflavin metabolism

Table 4.6.4: Biological classification of HIV group distinguishing metabolites detected in ESI positive mode of UPLC-MS together with their functions or pathways the compounds are involved in.

NEG and POS group differentiating metabolites		
Metabolite number	**Metabolite name**	**Functions/ Pathways**
54) Carbohydrates and Carbohydrate Conjugates, monosaccharaides	Alpha-D-glucose	Component of galactose metabolism Component of nucleotide sugars metabolism Component of starch and sucrose metabolism Component of streptomycin biosynthesis Primary source of energy for living organisms Therapeutically in fluid and nutrient replacement
55) Organic Acids and Derivatives, Keto-Acids and Derivatives	Fumarylacetoactate	Component of Tyrosine and Phenylalanine metabolism
56) Nucleosides, Nucleotides, and Analogues, Benzimidazole Nucleosides and Nucleotides	Alpha Ribazole	Intermediate in riboflavin metabolism
57) Lipids, Steroids and Steroid Derivatives	Cholesterol sulphate	Cell signaling Fuel and energy storage Fuel or energy source Hormones, Membrane component Membrane integrity/stability Anti-viral agent
58) N/A	Decyl-beta-D-maltopyranoside	N/A
59) N/A	Cycloastragenol	Anti-neoplastic activity
60) N/A	Strictosidine	N/A

61) Organic Acids and Derivatives, Carboxylic Acids and Derivatives	Deferoxamine	Chelating agent Iron chelating agents Siderophores Drug
NEG and POS HAART group differentiating metabolites		
62) Aromatic Heteropolycyclic Compounds, Indoles	Alosetron	Gastrointestinal agents Serotonin antagonists Drug
63) N/A	Ubiquinol-2	Anti-oxidant
64) Aromatic Heteropolycyclic Compounds, Imidazopyrimidines	Famciclovir/ Famvir	Guanosine analogue antiviral drug for various herpes virus infection Antiviral agents Nucleosides and nucleotides Prodrugs Anti-microbial agent DNA polymerase inhibitor
65) Aromatic Homomonocyclic Compounds, Stilbenes	Tamoxifen	Antineoplastic agents, Hormonal Bone density conservation agents Estrogen antagonists Selective estrogen receptor modulators
66) N/A	Ecelysone	N/A
67) N/A	Clobetasol propionate/ Clobex	N/A
68) N/A	Roloxifene hydrochloride	N/A
69) Aromatic Homomonocyclic Compounds, Benzimidazoles	Telmisartan	Angiotensin II type 2 receptor blockers Angiotensin-converting enzyme inhibitor Used in the management of hypertension

70) N/A	Tetramethylrodamine thiocynate	N/A
POS and POS HAART group differentiating metabolites		
71) N/A	5-phosphoribosyl-α-formylglycine	N/A
72) N/A	Nemonapride	N/A
73) N/A	GBR 12935	N/A
74) Aromatic Homomonocyclic Compounds, Phenethylamines	Salmeterol	Adrenergic beta-agonists Bronchodilator agents Sympathomimetic
75) N/A	SALOR 2	N/A
76) Aromatic Homomonocyclic Compounds, Benzoic Acid and Derivatives	Benzonatate	Antitussives
77) N/A	Mannasantin A	N/A
78) N/A	Cycloelepsipeptide 2	N/A
79) N/A	11-(perfluorocytyl) undecyl dimorpholino phosphirate	N/A

With the detection of the above metabolites, it was possible to identify samples that contained them. This was conducted using trend plots, which are plots that enable the tracking of metabolites of interest back to the sample or patient exhibiting this metabolite. This produces a more pronounced image of how a metabolite is behaving among the samples of different groups. As seen in figure 4.10 the trend plot shows the levels of adenosine between the NEG samples and POS HAART samples. Adenosine levels increased in about 46% of the patients on treatment and the NEG samples all containing low levels of this metabolite. Because now it is known which serum samples contain high levels of adenosine, then this can be traced back to the patient's chromatograms to aid in the identification of this metabolite by extracting its spectrum. Figure 4.11 shows a patient's chromatogram and the extracted spectra of adenosine as guided by the trend plot and metabolite features. Only after the spectrum is extracted can elementary analysis be conducted to confirm that this is indeed the identified adenosine. The trend plot is a useful tool that aids in metabolite identification.

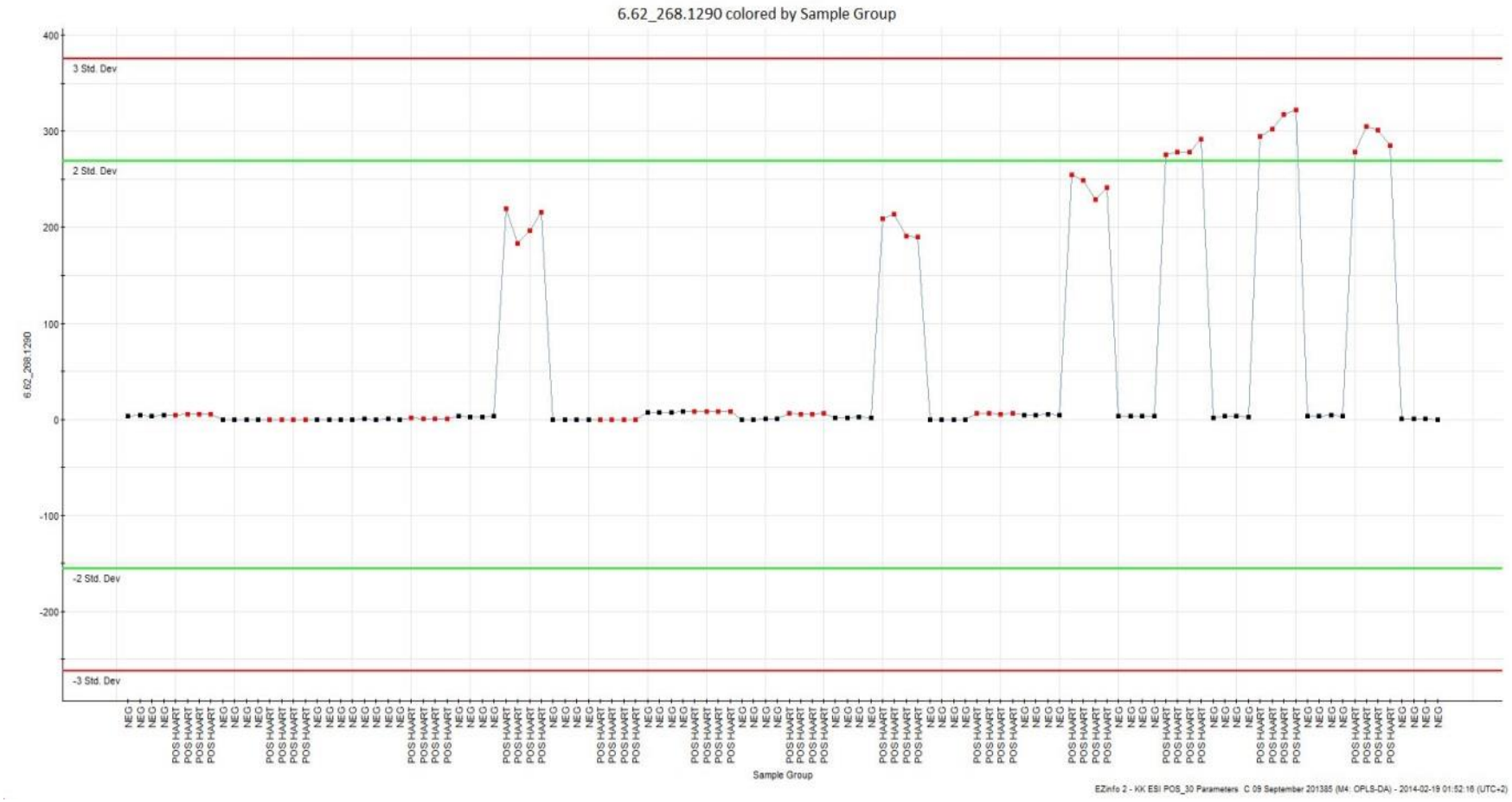


Figure 4.10: Trend plot of an HIV serum metabolite [6.62_268.1290 (Adenosine)] detected in ESI+ mode of UPLC-TOF-MS. Metabolite appearing to be increased in about 46 % of the POS HAART group. This is indicative of the alteration that treatment causes on the metabolic profile. This is an example of a trend plot used to trace the samples that contain the levels of interest in the study subjects.

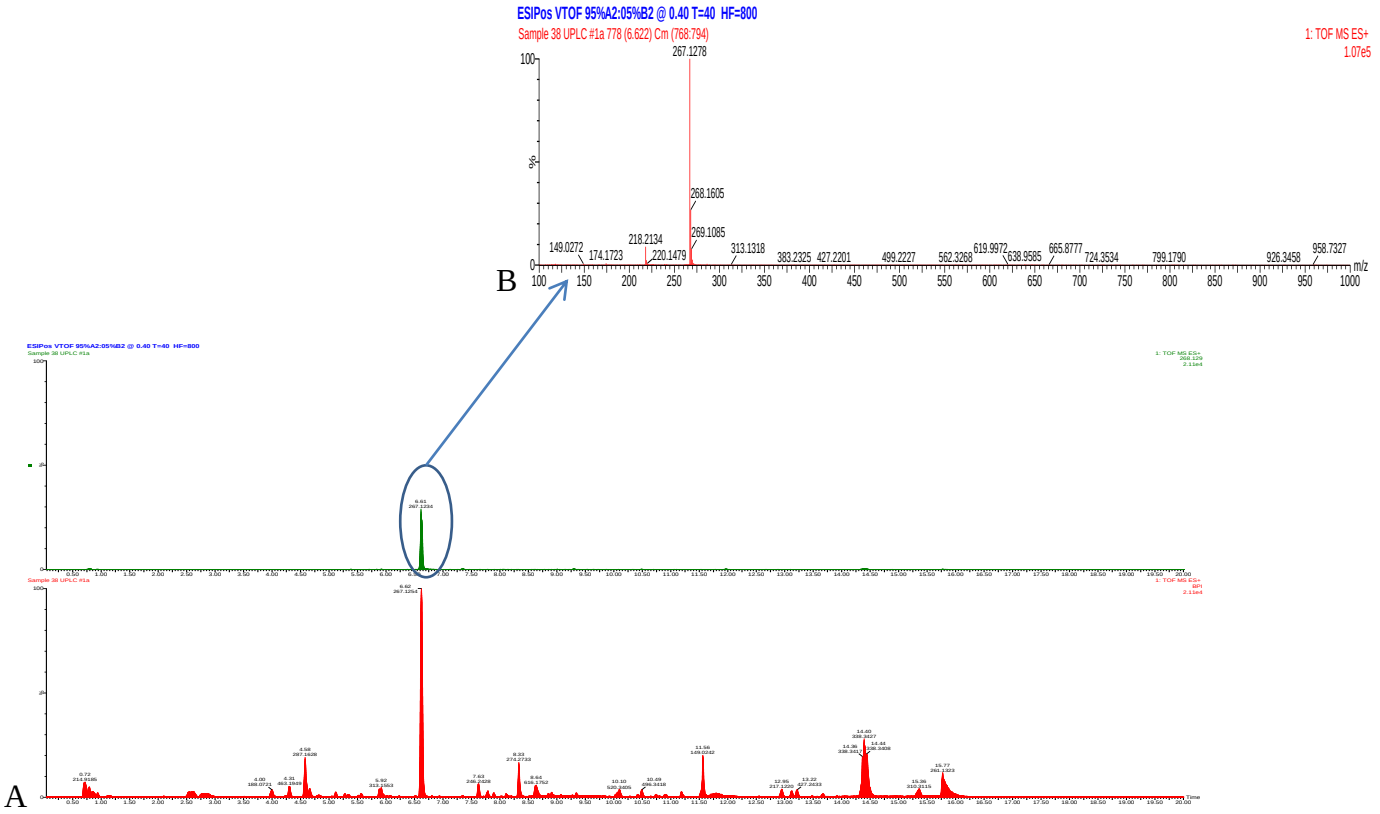


Figure 4.11: (A) BPI chromatogram of adenosine in an HIV serum sample analyzed in ESI+ mode of UPLC-TOF-MS showing the BPI chromatogram of adenosine. The patient was on treatment and contained high levels of adenosine as traced by the metabolite trend plot. (B) The extracted adenosine spectrum from the chromatograms.

4.3) FTIR data

Samples that were analyzed using FTIR were used for a comparative investigation to determine whether a less sensitive technique than UPLC-MS could discern similar metabolic patterns as UPLC-MS. Figure 4.12 shows the overlaid FTIR spectra of the three experimental groups (HIV-, HIV+HAART- and HIV+HAART+). Observable differences among the three representative serum samples were evident through visual inspection. The treatment group appears to have the most up-regulated spectrum compared to the HIV negative and HIV positive sample, especially at the peak region between 1500-1800 cm^{-1}. This shows how HAART drastically altered the metabolic profile.

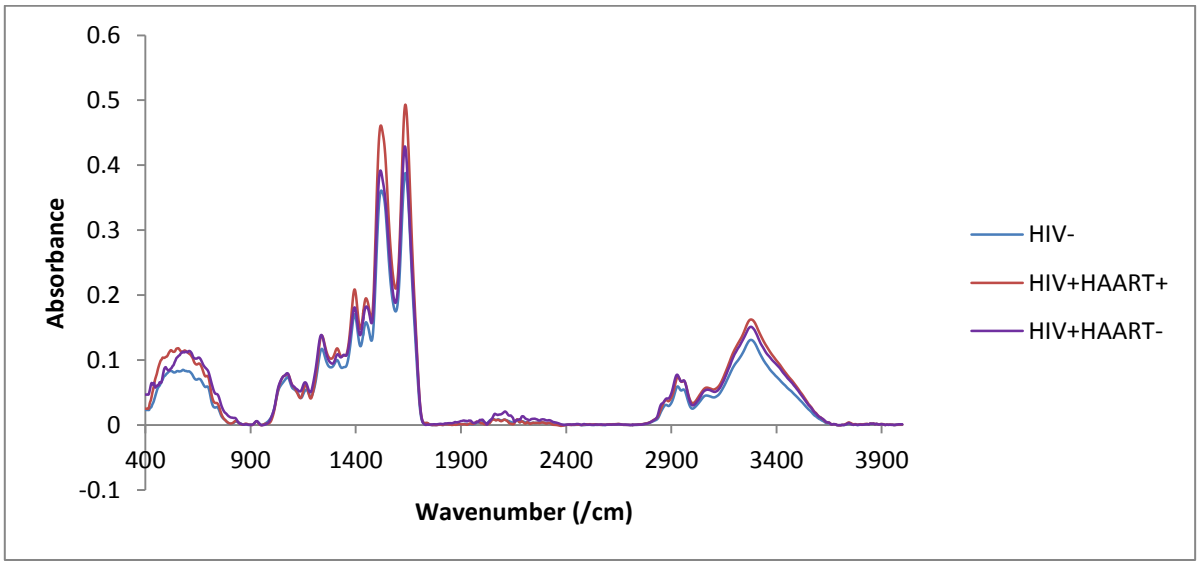


Figure 4.12: Overlaid FTIR spectra of HIV-, HIV+ and HIV+HAART+ sera samples. Spectra clearly showed slight differences among three experimental groups. The treatment group appears to contain an up-regulated profile.

The differences observed in figure 4.12, were further amplified using LDA shown in figure 4.13. The three groups were clearly separable from each other, however, within group variations appear maximized. Of the ten peak regions that were altered and responsible for the variations observed in this LDA scatter plot, only two peaks regions were significantly different, namely; region (a) 910 cm^{-1} ($p=0.000$) and (b) 3915 cm^{-1} ($p=0.018$) which corresponded to the alkene and hydroxyl functional groups, respectively because they absorb strongly in these regions (Argarwal, 2010; Ramachandran & Velraj, 2011).

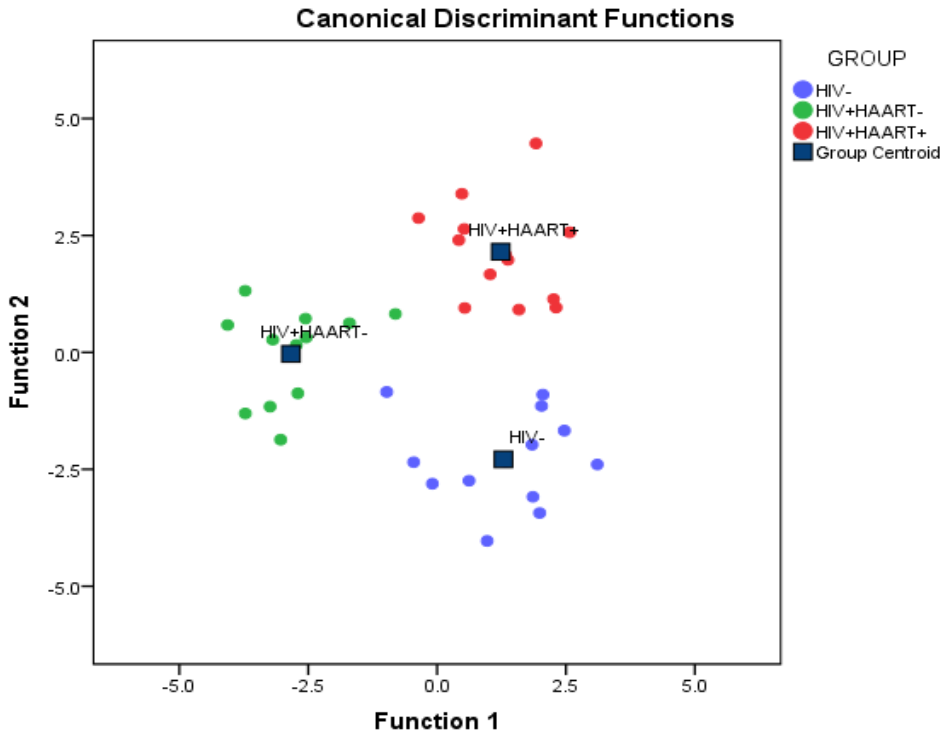


Figure 4.13: LDA scatter plot of HIV sera samples analyzed using FTIR showing the separation among the three experimental groups (HIV-, HIV+ and HIV+HAART+). Sera samples show within and between group variation. Cross validation correctly grouped 87.2% cases.

This FTIR data was converted to a suitable readable file (.csv file) and transported to MarkerLynx XS EZInfo 2.0 for PCA analysis. The PCA showed poor separation of the three groups (see figure A7 in appendix) compared to the LDA, so as a result the information from this PCA scores plot was not considered for further analysis.

This brings to an end the results chapter where all the experimental findings have been presented. The next chapter that follows will attempt to explain the implications of these results. The biological and toxicological effects will also be highlighted in the discussion section that follows.

CHAPTER 5: DISCUSSION

This chapter discusses the findings of the second batch of UPLC-MS data (5.1), then highlights the findings of both the first and second batch of UPLC-MS data (5.2) and finally discusses the FTIR data (5.3). Batch one details are comprehensively discussed and concluded upon in the published manuscript.

5.1.1) Chromatographic Output and Interpretation

Given that HPLC has already proved to be a valid tool for the diagnosis of inborn errors of metabolism and its newer form, UPLC offers improved qualities that were detailed in section 2.5.3.2c (Tavazzi et al., 2005; Swartz et al., 2005; Novakova et al., 2006), some of these specifications (e.g. sensitivity, increased resolution, reproducibility etc) combined with those of MS were noticed in the chromatograms obtained in this study. Depending on the sample of interest to be analyzed with UPLC-MS, it is of extreme importance to determine and eliminate possible interfering compounds because of the high sensitivity of this analytical tool. Narrow and distinct chromatographic peaks free of contaminants like phospholipids and plasticizers were obtained by a fast, easy, and reproducible SPE protocol that used Ostro^{TM} plates for sample preparation for UPLC-MS analysis (Williams, Kgoadi et al., 2014). A previous study by Tavazzi et al (2005) has shown that minimal sample preparation is favorable for reliable analytical data and Ostro^{TM} plates sample preparation meet these requirements.

Figure 4.1 showed changes in peak intensities, peak elution times, and peak patterns that were clearly indicative of differences among the three sample groups. A single peak is supposed to represent a single compound, however it is not always the case. Observed changes could be explained as follows; (1) increases and decreases in peak intensities could be representative of up-regulation and down-regulation of metabolites, (2) shifts in peak retention times or could be indicative of other metabolite/s that were detected only in one specific sample group in comparison to other sample groups, and (3) peak patterns could be representative of the increased number of peaks (compounds) that can be detected by UPLC-MS. Wilson et al (2005) showed that individual peaks have a number of parent ions with a range of molecular weights. Aberg et al (2009) also showed that the same metabolite detected in different samples does not always appear the same.

The metabolites appearing at different retention times (e.g letrozole) could mean they are isomers of each other but the only way this can be determined is after definitive identification when chemical structure elucidation is performed. One of the challenges in the identification of MS metabolites is the identification of nominal and monoisotopic masses with different chemical structures and single metabolites are detected as multiple derived species (Dunn et al., 2012).

It was clear from figure 4.1 that the HAART samples appeared to have the most altered metabolic profile compared to the NEG and POS serum samples. Intensities of the metabolites in the HIV positive patients were decreased as compared to HIV negative subjects. HAART appears to be reversing the effects of HIV infection on the metabolism as can be seen by the increased metabolite intensities in the HAART patient that were decreased by HIV, however HAART is also altering other metabolites as can be seen in metabolites that are drastically increased by HAART. Again, proving what literature has already shown that HAART enhances the metabolic complications caused by HIV infection. These changes were amplified using statistical analysis to identify the significance of the affected metabolites.

5.1.2) Biological Interpretation of statistical patterns/trends

LDA as explained in section 2.5.2 was the starting point for statistical analysis in this study to show between and within class variations of the three experimental groups because it proved valuable in other studies in our group (Hewer et al., 2006; Phillippeos et al., 2009; Williams, Kgoadi et al., 2014). The observed LDA patterns in figure 4.2 imply that therapy may be succeeding in initiating the desirable effects of restoring the immune system. The metabolism was however, not entirely restored, this was evident in the HAART group as it clustered furthest away from the HIV negative and positive serum samples.

PCA scores plots of ESI negative mode data and ESI positive mode data respectively (figure 4.4 and 4.5) showed clustering of pooled samples closer together and closer to the origin. This is representative of the stability of the UPLC-MS analysis runs, rendering this data worth mining because it was of good quality. Figure 4.4 of the ESI negative mode data showed that some of the HIV negative and positive samples overlapped while the majority of the HAART serum samples clustered furthest away from the other sample groups. This could possibly indicate that the positive group is asymptomatic and possibly still classified as

healthy according to the immune system. Villas-Boas et al (2007) reported organic acids as being detected in the ESI negative mode and Williams et al (2012) profiled organic acids using GC-MS showing that organic acids were altered during HIV-infection, most of which were increased in concentration. Findings of this study demonstrated that HAART was the major contributor of changes in organic acids; and that these molecules were markers of mitochondrial dysfunction. Organic acids were detected in both modes in this study because different compounds are not restricted to a specific ionization mode.

Interestingly, figure 4.5 showing the ESI positive mode data showed that majority of the treatment group overlapped with more than half of the positive group and greater than half of the HIV negative samples clustered separately. This suggests that the metabolic profile of some patients on HAART treatment could be responding successfully to treatment while other treated patients may be experiencing HAART toxicities, in this case disrupted metabolic profiles. Clearly, the degree at which HAART affects the different classes of molecules that are affected by HIV differs. Organic acids appear to be the affected molecules.

Three serum samples were dropped from the study following PCA because they were detected as outliers when they clustered outside the Hotelling's T2 ellipse (95% confidence interval) and disrupted the data. One of the dropped sample contained high levels of ethylenediaminetetraacetic acid (EDTA) and was the only sample that contained EDTA. This chelating agent (EDTA) that prevents blood clotting is normally present in plasma samples not serum samples. PCA scores plot showed that plasma and serum samples cluster far apart (Barri & Dragsted, 2013), this scores plot was shown in figure 2.8. The levels of the detected metabolites can be traced back to the specific patients. This is very useful in selecting patients that display metabolic changes that are of interest so that further studies can be conducted on such serum samples to answer further research questions that arise. For instance, metabolites that exhibit anti-inflammatory or anti-oxidant activities can provide direction about which samples to conduct biochemical assays of interest on, because the researchers will already have an idea of what results are expected from patients that exhibit these metabolites.

The OPLS-DA scores showed tight clustering of between and within sample separations. These scores plot patterns are normally expected from uniform samples such as plants (Villas-Boas et al., 2007) not heterogeneous human samples like the randomly selected HIV serum samples used in this study. Patients used in this study were definitely not uniform, they

varied because of factors like treatment. Ideally, what is best is less within group variation and larger between group variation. With the POS HAART group there was a larger within group variation than within group variation, it is expected because the patients are receiving different HAART regimens and this affects their metabolism differently. The ESI positive data showed greater variations in the NEG group as compared to the POS and POS HAART groups while the ESI negative data showed less within variation in the NEG groups when compared to other groups.

Data reproducibility of mammalian samples is a challenge because sample homogeneity is normally affected by sampling time, age, gender etc (Bollard et al., 2001; Robosky et al., 2005). This UPLC-MS study has compensated for confounding factors by providing two biological repeats and 3-4 technical replicates through sample analysis as was suggested by Villas-Boas et al (2007). Gender and race were not confounding factors because the samples failed to group together according to gender or race (See appendix, figure A7) on the PCA score plot. Age was also not a confounding factor (appendix, figure A8). Variability in metabolite levels, responses or trends among humans is difficult to eliminate and still a challenge to overcome because unlike animals which can be controlled to eliminate the unwanted variables, humans have the freedom to control their metabolic status resulting in their intrinsic metabolome being highly variable (Kaiser et al., 2005; Villas-Boas et al., 2007).

5.1.3) Energy Metabolism metabolites

Glucose and lipids are usually metabolized as the first source of energy by the body. Proteins are broken down as a source of energy only when glucose and lipid reserves are depleted. Some detected lipid metabolites such as pregnenolone, Dehydroepiandrosterone sulphate (DHEAS), cholesterol sulphate and obtusifoliol are energy sources and storage metabolites that were decreased by HIV infection and most declines were attributed to the treatment group as compared to HIV negative subjects. Cholesterol has important regulatory functions (Strott & Higashi, 2003). Alpha-D-glucose, which is the primary source of energy in living organisms (Wishart et al., 2013) was also decreased in the HIV positive patients. Phosphoribosylamine, an amino sugar was also decreased by HAART treatment. Treatment has already been shown to cause energy deficiency (Kohler & Lewis, 2007), a phenomenon that has been observed in this study by the decrease in lipid levels as a result of HAART.

Total energy expenditure in HIV positive patients is decreased during weight loss (Macallan et al., 1995), these patients could already be suffering from lipoatrophy. Energy is trapped in the nutrient molecules (Macallan et al., 1995), and most energy molecules detected in this study are clearly reduced by HIV infection and treatment, therefore displaying a sign of energy deficiency. HAART treatment also increased the levels of some carbohydrates such as salicin and fructoselysine-6-phosphate while HIV infection increased the levels of sialic acid. This may be indicating that the patients are either already suffering from insulin resistance or their risk of developing this disease was increased or that is why they are desensitized from glucose levels. NRTI treatment, which were administered to most patients in this study are known to cause cellular damage in the form of energy deficiency (which is already observed here) and mitochondrial dysfunction (as was shown by Kohler & Lewis et al., 2007).

Energy in the form of ATP is generated by oxidative phosphorylation through the electron transport chain in the mitochondrion (Voet et al., 2008). This energy is synthesized when the proton gradient is formed by the pumping of the electrons outside the mitochondria into the matrix (Voet et al., 2008). The electrons normally pass through electron transport complexes (Complex I, II, III and IV) before ATP synthesis. This study detected drastically high levels of reduced FMN and ubiquinol (reduced coenzyme Q) in the patients on HAART treatment compared to the negative subjects and positive patients, these two metabolites form part of complex I and II. FMN is a biologically active form of riboflavin (vitamin B2) and is involved in important functions of aerobic cells because of its involvement in the redox reactions (Powers, 2002). Clearly these patients were administered vitamin B2 supplementation together with HAART treatment to help assist with restoration of their energy metabolism, however this attempt was not successful because ATP was not detected in this study. This could mean the patients energy levels were depleted because insufficient ATP was being synthesized (see figure 5.1A for ATP energy synthesis). Even though $FMNH_2$ and ubiquinol saved their function of collecting electrons in the electron transport chain (see figure 5.1B), oxidative phosphorylation did not occur resulting in no ATP synthesis. Bearing in mind that oxidative phosphorylation requires an intact mitochondrial membrane to transport protons out of the mitochondria and into the matrix in order to form a measurable electrochemical gradient (Voet et al., 2008), the membrane integrity in these study patients seemed to have been damaged. The levels of metabolites, mostly lipids that play a role in membrane integrity and stability were decreased by HIV infection and HAART

treatment leading to a compromised mitochondrial membrane. This in turn inhibited the synthesis of ATP and it explains why ATP was not detected in this study. It could be speculated that energy is being generated anaerobically and that pyruvate is being converted to lactate leading to lactic acidosis but because no lactate was detected, this argument does not have astrong basis.

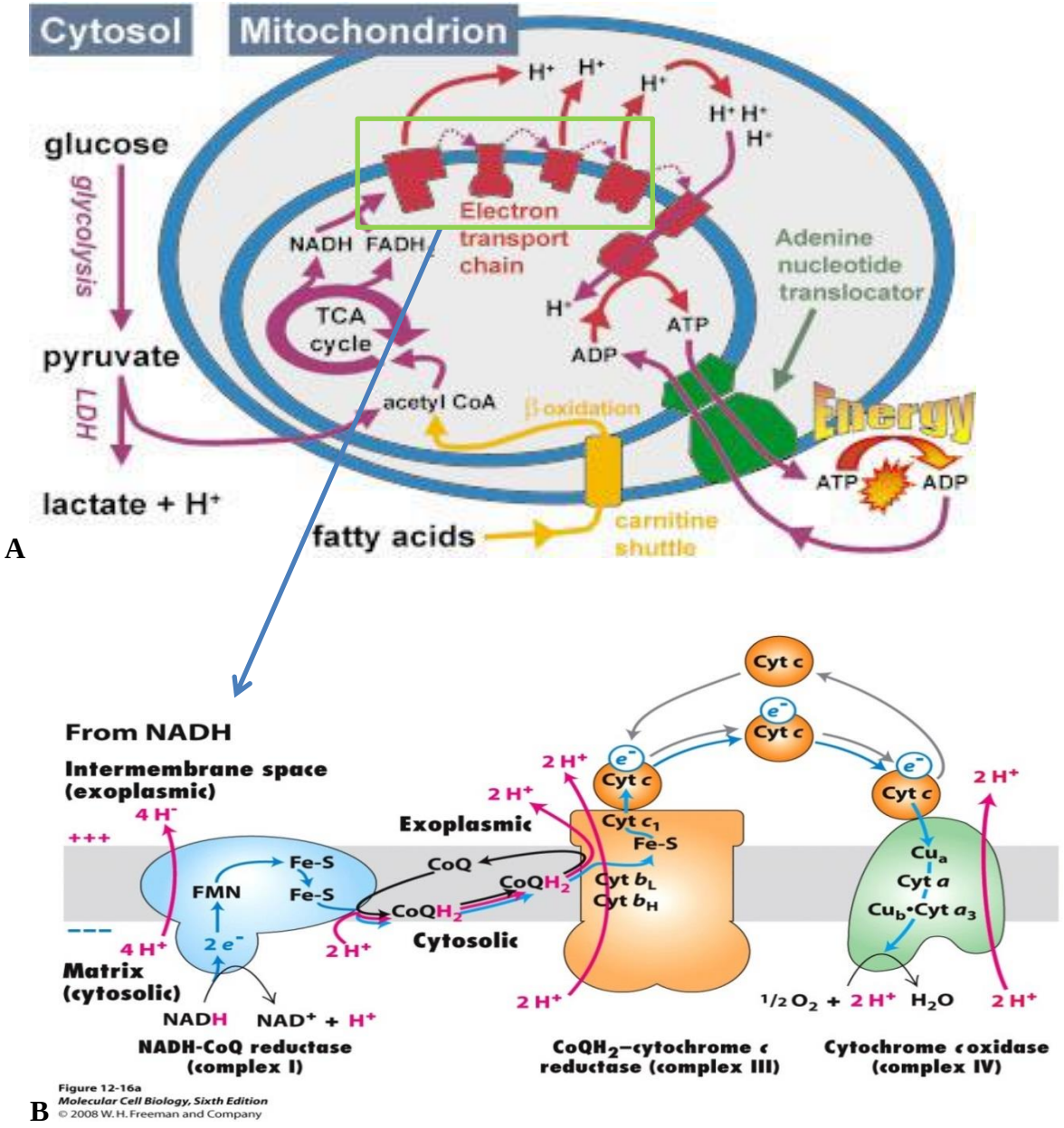


Figure 5.1: (A) ATP energy generation of the mitochondria through oxidative phosphorylation, also showing the anaerobic fate of pyruvate that result in production of lactate. (B) Electron transport chain showing the flow of electrons among the protein complexes and the reduced FMN in Complex I and reduced coenzyme Q as they transport electrons out of the mitochondria and into the matrix to power oxidative phosphorylation (White, 2001; Lodish et al., 2008).

5.1.4) HIV-1 Altered amino acids

Most of the essential and non-essential amino acids have been shown to be decreased in HIV/AIDS patients as compared to HIV negative subjects (Laurichesse et al., 1998). Threonine, an essential and rate limiting amino acid for protein synthesis in HIV/AIDS patients (Laurischesse et al., 1998) was detected in low levels in the HIV positive patients as compared to the HIV negative subjects. The relative small factor of change of threonine corresponds with Hortin et al (1994) who showed HIV infection does not drastically decrease concentrations of essential amino acids. Threonine's important function in the body is its role in protein synthesis (Stoll, 2006).

Another depleted metabolite due to HIV infection that was L-Gamma-glutamyl-cysteine, a substrate for γ-glutamylcyclotransferase and glutathione (GSH) synthatase (Misra & Griffith, 1998). Even though it is a substrate for two enzymes, it favors GSH synthetase in cells because of the higher affinity and activity that GSH synthatase has. This enzyme catalyzes the ATP-dependent ligation of L-glutamate to L-cysteine to produce L-γ-glutamyl-L-cysteine (Griffifth & Mulcahy, 1998), a metabolite that protects GSH from hydrolysis by intracellular peptidases. (Wu et al., 2004; Misra & Griffith, 1998). The synthesis of L-γ-glutamyl-L-cysteine is the first step for the biosynthesis of GSH and it is rate a limiting step (Meister, 1991; Griffith & Mulcahy, 1999). Figure 5.2 illustrates the metabolic pathways involved in the synthesis of GSH and its utilization in animals, highlighted also is the detected metabolite in this study, the intermediate γ-Glutamyl-Cysteine. GSH was not detected in this study, this could be attributed to; low levels of this rate limiting metabolite, failure of low levels of L-γ-glutamyl-L-cysteine to protect GSH from hydrolysis (Wu et al., 2004) , or absence of glycine as a required substrate to react with L-γ-glutamyl-L-cysteine to produce GSH. Threonine, present in low levels due to HIV-infection is also a component of glycine metabolism (HMDB, 2012). GSH functions as a defense against reactive oxidative species and electrophiles, nutrient metabolism and a cellular events regulator (Wu et al., 2004; Misra and Griffith, 1998).

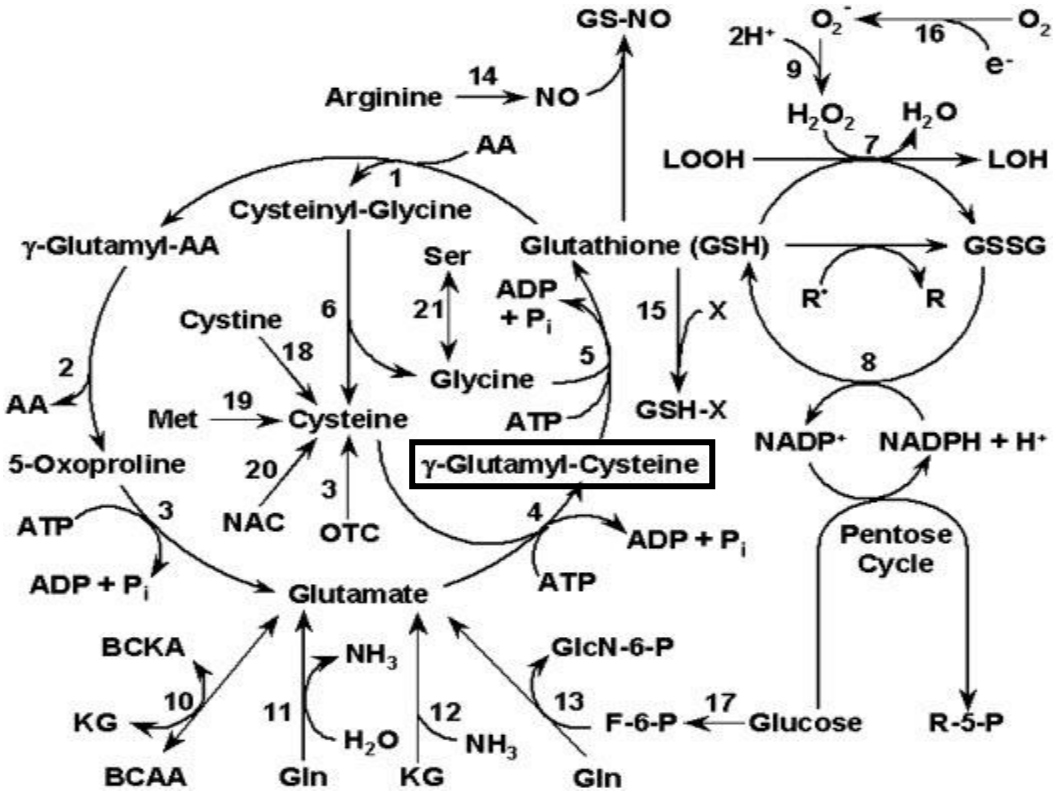


Figure 5.2: Illustration of metabolic pathways involved in the synthesis of glutathione and usage in animals. Highlighted is γ-Glutamyl-Cysteine, a glutathione metabolite detected in this study (Wu et al., 2004).

GSH suppresses HIV replication (Kalebic et al., 1991) and compounds that boost GSH concentrations are being taken into consideration for HIV treatment/therapy (Hortin et al., 1994). A GSH prodrug (glutamylcysteine ethyl ester) was shown to possess novel HIV inhibition properties and this drug passed through the permeable cell membrane and was intracellulary metabolized to GSH through γ-glutamyl-cysteine (Kubota et al., 1998). These levels of amino acids that are depleted due to the human immunodeficiency virus serve as an indicator of how the hosts metabolism is disrupted resulting in failure of the host to maintain homeostasis. HIV positive patients in this study could be suffering from glutathione deficiency and this deficiency contributes to the development of metabolic complications. Deficiency of GSH adds to oxidative stress and participates in the pathogenesis of many pathophysiological conditions like HIV/AIDS, diabetes, cancer etc (Wu et al., 2004). The life span of the HIV patients who are not receiving HAART in this study could be shortened if no intervations are made to increase their levels of GSH because Herzenberg et al (1997) showed that GSH deficiency decreases the survival of HIV positive patients. These depleted levels of amino acids could be used as potential metabolic biomarker candidates of HIV infection because literature has already associated their decreased levels with HIV-1.

Sialic acid levels were low in the HIV negative subjects compared to the HAART treatment patients, this metabolite is involved in amino acid metabolism. Increased sialic acid concentrations have been attributed to ageing (Lindberg et al., 1991). This is expected because HIV infection has been associated with accelerating ageing in positive patients.

5.1.5) Nucleotide, nucleoside and analogues

5.1.5.1) Nucleoside metabolites: Adenosine and Inosine biological significance

During the process of inflammation and infection, adenosine levels are increased and in HIV infection, adenosine concentrations were shown to be high as a result of the degradation of cAMP by the virus (Lakshmi et al., 2013). Adenosine produces its anti-inflammatory properties at the sites of inflammation, where it is normally released (Nakav et al., 2008). Adenosine deaminase (ADA) is an enzyme that catalyzes the conversion of adenosine and 2'deoxyadenosine to inosine and deoxyinosine, these products are toxic to the cells (Cordero et al., 2001; Lakshmi et al., 2013). ADA receptors are dependently activated by local adenosine concentrations and build-up of these toxic metabolites cause severe immunodeficiency as ADA activity is inhibited (Zavialov et al., 2010: Lakshmi et al., 2013). ADA activity is also inhibited by the lymphokine interferon gamma (Lakshmi et al., 2013). This study detected high levels of adenosine in the HIV positive and HAART patients compared to the HIV negative patients, implying that ADA activity was inhibited because the immune system of these patients is already compromised by HIV infection. The observed high levels of adenosine in the treatment group may imply that ADA activity was low in the treatment group compared to the positive group as was shown by Lakshmi and colleagues (2013) who used serum and plasma samples in their study. However, higher levels of adenosine detected in the treatment group compared to the HIV negative group imply that ADA activity was higher in the negative group than the treatment group, which is in contrast with what Lakshmi et al (2013) observed.

Even though high levels of adenosine were present in serum, high levels of the inosine were also detected in HIV positive and HAART patients when compared to the HIV negative subjects indicative of partial inhibition of adenosine ADA activity. Adenosine is a regulatory molecule used in coronary artery disease and supraventricular tachycardia (Das, 2003; Zavialov et al., 2010) while inosine is a drug metabolite that causes purine nucleoside

phophorylase deficiency when present in high levels and also causes coronary artery disease and degenerative disc disease (Wishart et al., 2013). However, inosine was shown to decrease mortality and suppress pro-inflammatory production of cytokines (Das, 2003). Coronary blood flow during myocardial metabolism is mediated by adenosine and inosine, both molecules were termed cardiac nucleosides because they serve as sensitive markers of myocardial supply/demand imbalance and playing a role in coronary artery disease (Fazekas et al., 1999). This indicates that the HIV positive and HAART patients in this study are exposed to a higher risk of developing or already suffer from coronary artery disease. These patients also contain possible risk factors of cardiovascular complications because Funaya et al (1997) correlated the severity of chronic heart failure to increased adenosine levels in systematic blood. Lakshmi et al (2013) suggested that ADA activity can be used for assessment of HIV disease progression and the status of its immune system implications and can be used as a guide on when to initiate antiretroviral treatment. This study suggests that ADA activity may be assessed indirectly by measuring the relative levels of its substrate (adenosine) or degradation product (inosine) to aid in the management of HIV/AIDS and its treatment.

5.1.5.2) Nucleotide reverse transcriptase Inhibitor metabolites and mitochondrial dysfunction

This study showed that most alteration in the metabolic profile of HIV positive patients was caused by treatment with reverse transcriptase inhibitors, with the major regimens causing alteration being NRTIs. These regimens are known to cause mitochondrial toxicity. The factors influencing mitochondrial toxicity resulting from NRTIs include; high concentration levels of NRTIs above the threshold required for exerting their effect at a subcellular level, the cellular nucleoside kinases ability to make nucleoside triphosphates and functional mitochondria threshold before depletion occurs (Apostolova et al., 2006, 2011). Pol γ amino acid residues are suggested to be responsible for the selection of dinucleotide triphosphates (dTNP), ultimately NRTIs (Lim et al., 2003). The conversion of NRTIs to triphosphate from monophosphate forms occurs at a reduced speed, resulting in high levels of nucleotides accumulating in the mitochondrial matrix and causing possible unspecific mitochondrial DNA synthesis inhibition (Lim et al., 2003). Pharmacogenomics have also linked NRTIs toxicities to specific genetic polymorphisms (Yamanaka et al., 2007). Inosine triphosphate (ITP), guanosine triphosphate (GTP) and 5-methyl-dicytidine triphospahte (5-methyl-dCTP) showed accumulation in the mitochondrial matrix because the HAART patients contained higher levels of these metabolites compared to the HIV negative subjects. The HIV positive

group also had high levels of GTP. This was a clear indication of the inhibitory actions of NRTI triphosphates on polymerase DNA. NRTIs are mainly responsible for lipoatrophy and induce apoptosis and induce mitochondrial depletion that ultimately leads to mitochondrial dysfunction (Apostolova et al., 2011; Pinti et al., 2011). These patients may be suffering from lipodystrophy or their risk of developing lipoatrophy are increased by HAART drugs. Mitochondrial function is reversible because cessation or suspension of using NRTIs results in increased mitochondrial DNA and improved mitochondrial framework (Mussini et al., 2005). This is why it is strongly encouraged to monitor metabolic complications resulting from HIV infection or its treatment (HAART) because changing the class of HAART regimens administered to patients can help prevent/combat the development or decrease the degree of metabolic complications such as mitochondrial dysfunction.

5.1.5.3) Mitochondrial Dysfunction

Some of the energy metabolites (pregnolone, obtusifoliol, DHEAS and cholesterol sulphate) also play a role in membrane integrity and stability (Wishart et al., 2013). A decrease in these metabolites translates to changes in the mitochondrial membrane potential. Mitochondrial alteration occurs resulting in electron leakage, production of reactive oxygen species and decreased ATP synthesis (Kohler & Lewis, 2007). The first evident sign of NRTI toxicity is energy deprivation resulting from the depletion of mitochondrial DNA (Kohler & Lewis, 2007). The mitochondrion takes part in anabolic pathways of fatty acids, cholesterol, bile acids, haem and urea (Apostolova et al., 2006, 2011).

5.1.5.4) Biological significance of nucleosides/nucleotides metabolites

Inosine concentrations were shown to be increased in Caravan diseased patients in an HPLC study for chemical diagnosis of inborn errors of metabolism; this study also detected cytosine and guanosine metabolites (Tavazzi et al., 2005). Other detected metabolites included; cytosine, alpha ribazole and coformycin. Chantin et al (1996) showed that HPLC is an accurate and useful tool for detecting purine metabolism abnormalities, and that high concentrations of inosine and guanosine detected in plasma and urine resulted from purine nucleoside phosphorylase (PNP) deficiency. High concentrations of some nucleotide triphosphates cause inhibition of ribonucleotide reductase and DNA replication in ADA and PNP deficient individuals resulting in immunodeficiency (Chantin et al., 1996). The patients are already immunocompromised due to HIV infection. This study shows that the advanced form of HPLC; UPLC, when used in combination with MS is a valuable tool for detecting

nucleobases, nucleosides/nucleotides involved metabolic complications arising from the disruption of the host metabolism due to stimuli such as disease and therapy.

5.1.6) Anti-HIV and anti-viral drug metabolites

Detected anti-viral drugs and their nucleoside analogues include; platanic acid, cortisol succinate, acyclovir, famciclovir and kolaflavanone. Guanosine analogues of AZT, namely; acyclovir and famciclovir (Brown et al., 2002) are mostly used for the treatment of herpes. These two metabolites were detected in higher levels in the HAART patients with acyclovir having the highest factor of change between the HIV negative subjects and HAART patients. Such high levels are expected because the clearance of zidovudine is very inefficient in comparison to other NRTI drugs because dideoxynucleotides have the capacity to act as efficient substrates of Pol γ as much as the natural deoxynucleotides (Lim & Copeland, 2001; Lim et al., 2003). Levels of platanic acid were surprisingly high in HIV negative subjects in comparison to the HIV positive and HAART patients. Platanic acid and cortisol succinate target HIV and protease (NIAID, 2013). Because these two metabolites are both anti-HIV and antiviral drugs, you would expect higher levels to be detected in patients receiving treatment for HIV and other viruses. It is possible that some of the HIV negative subjects were being administered this drug for unknown viral infections (unknown to this study). Kolaflavanone, a biflavonoid with properties such as antiviral, anti-diabetic, anti-inflammatory abilities etc is used in African traditional medicines (Iwu & Wootton, 2002). This study showed that not only is chromatography (HPLC) a validated tool for separation of AZT and acyclovir as shown by Brown et al (2002) but that when used in its advanced form (UPLC) in combination with mass spectrometry (UPLC-MS), it is a valuable analytical tool for the separation and detection of HAART metabolites.

5.1.7) Anti-inflammatory metabolites in HIV-1

Inflammation is a common occurrence during HIV infection and is decreased by treatment; however, treatment also contributes to the high levels of inflammation that still persist after its administration (Neuhaus et al., 2010). Inflammation is associated with abnormalities that arise from the disruption of lipids in cardiovascular diseases and has been known to contribute to morbidity and mortality rates (Baker et al., 2011). The inflammation caused by HIV causes mitochondrial dysfunctions such as neuropathy and Type II diabetes as a result of the soluble markers of inflammation (Lichtenstein et al., 2005; Brown & Neher., 2010). With

HIV being an inflammatory disease, most antiviral drugs also contain anti-inflammatory properties (e.g cortisol succinate detected in this study) and this study detected anti-inflammatory metabolites that included salicin and quallic acid. Salicin levels were higher in the HIV positive and HAART patients compared to HIV negative subjects. Quallic acid levels were lower in the HIV positive and HAART patients. This suggests that the patients were already experiencing inflammation because they were treated with anti-inflammatory drugs. The decrease of quallic acid in patients could mean that this metabolite has exerted the desired function of decreasing inflammation and did not accumulate in the blood stream. However; salicin accumulated in the blood stream, because of its increased half-life or bioavailability. The presence of anti-inflammatory drugs in these patients could be an indication of experiencing or risk of developing HIV metabolic complications. These metabolic complications include lipodystrophy, hyperlipidermia, dyslipidermia or coronary artery disease because inflammation disrupts lipid metabolism. Adenosine, a nucleoside with anti-inflammatory properties was detected in high levels in the HAART patients.

5.1.8) Metabolites indicative of cancer treatment

The progression of HIV to AIDS leads to AIDS-related cancers such as kaposi sacrcoma, non-Hodgkin lymphoma and cervical cancer (Grulich et al., 2007). Detected cancer treatment drug metabolites included; rodiasine, cephalotaxine, vinblastine, letrozole, mannasantin A tamoxifen and 3'-N-debenzoyltaxol. Only vinblastine, used to treat lymphomas, Kaposi sarcoma, breast cancer etc (Wishart et al., 2013) was detected in high levels in the HIV positive patients as compared to the HIV negative subjects. Surprisingly, the rest of the anti-cancer metabolites were detected in higher levels in the HIV negative subjects compared to the HIV positive patients receiving and not receiving HAART. It would be expected for HIV positive patients and patients receiving HAART to contain higher levels of anti-cancer drugs because it is therapy for their HIV-associated cancers. These compounds are also used to treat cancers that are not HIV-associated or the mentioned cancers when occurring in the absence of HIV infection. The risks of developing cancer are increased in HIV positive patients because the immune system is compromised (Grulich et al., 2007). Khandekar et al (2011) showed that the risk of developing various cancers is also increased by alterations that occur in the metabolism of organisms.

The possible source of these anti-cancer metabolites that were high in HIV negative subjects could be dietary supplements or botanical drug products. It is common cultural practice in South Africa for people to use herbal medicines from traditional doctors or elderly family members to meet nutritional needs of the body or treat diseases while most westernized people take dietary supplements. Feng et al (2011) highlighted that the FDA product policy regards products that are derived from herbal medicines to be dietary supplements and botanical drug products. A review by Feng et al (2011) highlighted that herbal medicines used daily may contain a single pure compound or fractions or extracts that have anticancer properties. Manassatin A, one of the detected metabolites in this study was listed by Feng et al (2012) as one of the active constituents together with Eepimanassantin A as an anticancer agent that carries out its mechanism of action by causing cell apoptosis. These two molecules were patented in 2004.

5.1.9) Neurocognitive metabolites associated with HIV-1

DHEAS and pregnenolone were decreased in the HAART patients compared to the HIV negative and positive patients. DHEAS and pregnenolone are circulating neurosteroids which exhibit neuroprotective functions against apoptosis, oxidative damage, and neurite growth (Rietsner et al., 2010). These metabolites are energy sources and storage molecules that also play a role in membrane integrity and stability (Wishart et al., 2013). Pregnenolone mainly protects against neuronal apoptosis; pregnenolone together with DHEAS are broadly produced in the brain and reductions in these metabolites exposes neurons to excitotoxicity that results in HIV-associated dementia which can be fatal (Charalampopoulos et al., 2006; Leskiewicz et al., 2008). HIV-associated dementia, mild neurocognitive disorder and asymptomatic neurocognitive impairment are the three categories of neurocognitive complications of HIV-1, termed HIV-associated neurocognitive disorders (Antinori 2007; Kaul, 2009).

A decrease in factor or change of DHEAS between HIV negative subjects and the HAART patients was higher than between HIV positive and HAART patients. DHEAS concentrations are decreased during HIV infection and aging (Wisniewski et al., 1993) and low levels of DHEAS have been associated with neurocognitive impairment (Silver et al., 2005). DHEAS, being an androgen secreted by the adrenal glands has a protective function in the pathogenesis of diseases such as coronary artery disease, diabetes mellitus, atherosclerosis etc

(Khaw, 1996). Kandathil et al (2005) has already shown that DHEAS measurement assays (ELISA) can be used for monitoring the success of HIV treatment (Kandathil et al., 2005). Some of the listed diseases resulting from decreased DHEAS levels have already been identified metabolic complications arising from HIV infection.

This study showed that even HAART used to treat HIV infection decreased the levels of DHEAS and pregnonelone, this is an expected outcome because HAART toxicities are known to aggravate HIV metabolic complications. This corresponds to the drastically altered metabolic profiles observed in the HAART patients. Mild neurocognitive disorder is more prevalent in patients receiving HAART (Kaul, 2009), which suggests that the HAART patients in this study are at a higher risk of developing dementia or mild neurocognitive disorder. The higher factor of change observed in the HAART patients could also be indicative of treatment doing little to improve the neurocognitive impairment caused by HIV-infection as shown by Robinson-Papp & Simpson (2009).

5.1.10) HIV-1 vitamin related metabolites

The following four metabolites metabolites; 6-hydroxymethyl-7, 8-dihydro-pterin, alpha-ribazole, 7,8-Dihydrohepterin triphosphate and 5-methyltetrahydrofolate play a role in the synthesis of vitamins B2 and B9. High levels of 6-hydroxymethyl-7, 8-dihydro-pterin were present in HIV negative subjects compared to the HIV positive patients and the HAART patients had higher levels of this metabolite when compared to the HIV positive patients. There is clearly a decrease in the amount of folate biosynthesis in the HIV positive patients, meaning there is a low serum folate level due to HIV-infection, a scenario that was observed by Castro & Goldani (2009). High levels of 7,8-dihydrohepterin triphosphate were found in the HAART patients compared to the HIV negative and positive patients indicative of vitamin B supplements. Both metabolites are involved in the folate (vitamin B9) biosynthesis pathway. 5-methyltetrahydrofolate is an active derivative of folic acid. Metabolic complications that manifest as a result of HIV-infection are possible causes of anemia because folate deficiency is one of the identified factors that interfere with the production of blood cells (Castro & Goldani, 2009). Anemia, being the common hematological sign of HIV and AIDS, serves as an indicator of premature death in HIV positive patients (Moore et al., 1999; Claster, 2002). The HAART patients had high levels of folate because the three folate metabolites were increased. This implies that appropriate treatment (inclusive of dietary

supplementation) was given to the HIV positive patients to reduce their risk of developing anemia. The value of multivitamin supplementation provision to HIV positive patients as suggested by Claster (2002) is also evident in this study.

The levels of alpha-ribazole, which was detected in both ESI negative and ESI positive mode were higher in the HIV negative subjects compared to the HIV positive patients. This serves as another confirmation that our HIV negative subjects were relatively healthy because high levels of this riboflavin (vitamin B2) intermediate implies that vitamin B2 is being produced. The enzyme that catalyzes the final conversion of this intermediate to the final product (vitamin B2) might be exhibiting low activity. It is normal practice to administer HIV positive patients with vitamin B supplements to improve their health and survival. Five-methyltetrahydrofolate was one of the active vitamin B metabolites that were responsible for distinguishing the three experimental groups on an LDA scatter plot.

5.1.11) Other metabolites and their biological significance

High levels of bilirubin were observed in the HIV negative subjects in comparison to HIV positive patients and low concentrations of bilirubin were previously shown to block the entry of infectious viral particles (McPHEE et al., 1996). This could imply that bilirubin is inhibiting the entry of the human immunodeficiency virus in the HIV positive patients. Bile pigment products are normally increased in HIV infection. The human metabolome database (Wishart et al., 2013) stated the following about bilirubin; it is a heme degradation product, whose levels are elevated in certain diseases, however there are no normal levels of bilirubin because levels are indicative of a balance between production and excretion. The HIV negative subjects contained high levels of an anti-malarial drug, primaquine (KEGG, 2012) in comparison to the HIV positive patients. Primaquine acts an antiprotozoa and antimicrobial agent (NIAID, 2012). In most sub-Saharan African regions, antimalarial medicines are over the counter medicines. It has been shown to be a challenge to obtain pharmacokinetic data on primaquine because studies cannot be properly controlled due to antimalarial drugs being over the counter drugs (Fehintola et al., 2011). Some of the HIV negative subjects were taking or took antimalarial drugs prior to their blood donation to this study. HIV negative blood was collected during the time that students were returning from their university recess break, suggesting that some travelled to other countries during their recess break, and were

possibly on antimalarial drugs if travelled to affected regions. Other detected metabolites and their functions are listed in table 4.6 (4.6.1-4.6.4).

5.2) UPLC-MS findings for batch one and batch two

Both batches of data produced chromatographic outputs of similar quality that showed visible differences among the experimental groups, this demonstrated the reproducibity of the Ostro^{TM} plate SPE method. LDA achieved 100% classification accuracy for batch one data and batch two ESI- data, however, ESI+ data of batch two only achieved 88,9% classification because of samples that were outliers. The 21 metabolites responsible for the LDA separation observed for batch one data comprised of both significant and non-significantly altered metabolites, while batch two LDA metabolites (25) comprised entirely of significantly altered metabolites. With both batches, it was clear that HAART patients had the most altered metabolic profiles. The ANOVA looks at each variable on its own, while LDA is a multivariate procedure looking at the joint distribution of several variables. A variable that is not significant on its own, may be significant if taken jointly with other related variables. LDA metabolites from both batches consisted mostly of drugs used for treating viral infections and cancer and also vitamin supplements.

The additional statistical tools applied to the second batch data resulted in the generation of more informative data. The EZInfo 2.0 proved to be a better statistical analysis package tool for extracting information from untargeted high throughput metabonomics studies. OPLS-DA detected a high number of metabolites (all significantly altered) compared to LDA (an equivalent of PCA) and was able to assign group-distinguishing metabolites to either HIV or HAART. Both bacthes detected biochemical molecules (e.g carbohydrates, lipids nucleotides) and and drug related metabolites, however, batch two OPLS-DA metabolites were more specifically related to HIV and HAART associated metabolic complications.

The increased resting energy expenditure associated with HIV positive cells results in the depletion of the host's mitochondrial energy (Hommes et al., 1990, 1991). The evidence of the mitochondrial energy depletion by HIV and/or HAART is evident in batch two findings that show decreased levels of sugars, lipids and other molecules required to drive mitochondrial energy synthesis. Comparing the findings of both batches, the advanced statistical analysis applied to batch two data yielded more metabolites and provided more information about the study subjects than batch one, which makes it a better method to use.

The observed changes in the detected metabolites (e.g carbohydrates, amino acids, and lipids) appeared to be caused by HIV and/or HAART and not merely a general disease response. Schambelan et al (2002) recommended the regular assessment and monitoring of changes in molecules such as glucose and lipids in HIV positive patients as indicators of HIV metabolic complications and the use of these molecule changes to optimize the ratio of HAART benefits. The observed metabolite changes in this study mostly agree with other HIV/AIDS studies. Due to the absence of information on clinical symptoms for most of the study subjects, conlusions concerning the findings of this study were collaborated by literature.

5.3) FTIR data implications

FTIR picked up similar metabolic patterns as those observed by UPLC-MS that showed spectral separation of the three experimental groups. The ten variable regions that were responsible for the clustering observed on the LDA scatter plot did not entirely consist of significant vibrational regions. It should be noted that LDA does not necessarily depend on significant variables for separation of experimental groups; it also uses non-significant variables just so as to achieve the best separation. The vibrational regions that were significantly different among the three serum samples groups corresponded to the alkene (=C-H) and hydroxyl (O-H) functional groups.

Alkenes ($RCH=CH_2$) are known to absorb strongly at the frequency of 910-920 cm^{-1} (Argarwal, 2010) and this study observed significant vibration at 915 cm^{-1}. Carbon hydrogen bending is strongest in the 800-1000 cm^{-1} region of alkenes (Argarwal, 2010). Aliphatic absorption is also strongest at vibrational frequencies higher than 900 cm^{-1} (Argarwal, 2010). The second absorption region that was significantly different was the region at 3915 cm^{-1} which corresponded to the hydroxyl (O-H) stretching vibration, also known as the alcohol group (Ramachandran & Velraj, 2011). These functional groups occur as functional groups of amino acids, saccharides and lipids (Bahmani et al., 2009). The mentioned molecules (e.g amino acids and sugars) were detected by UPLC-MS, this is indicative that FTIR can complement metabolic patterns of certain metabolites detected using UPLC-MS. Bahmani et al (2009) already showed that the hydroxyl vibrational group indicates specific patterns in HIV positive plasma samples.

CHAPTER 6: CONCLUSION

The Ostro™ plates sample preparation proved optimal for UPLC-MS biofluid metabonomics as an easy to use method involving minimal sample handling (Wiliams, Kgoadi et al., 2014) and also that it is a reproducible SPE extraction method. In this study, UPLC-MS proved to be a valuable tool in detecting metabolic changes in serum samples. Statistical models or projections applied in this study were mostly complementary to each other regardless of the different source or software package. Advanced chemometrics in the form of PCA and OPLS-DA that are already recommended for metabonomics studies were found to be highly complementary to the increased sensitivity of UPLC-MS derived data in detecting distinguishing metabolites. The combination of UPLC-MS and chemometrics was successful in detecting and separating metabolites that differed among HIV negative, HIV positive and HIV positive patients on HAART treatment. The null hypothesis is therefore rejected because qualities such as high sensitivity and high resolution of UPLC-MS made it possible to discover potential HIV and therapy (HAART) biomarkers, with relative levels that corresponded highly to available literature. FTIR was able to detect metabolic patterns that showed separation among the three experimental groups.

Untargeted analysis was shown to be ideal for detecting a broader range of metabolites significantly affected by HIV and/or HAART, and detecting treatments for other diseases (such as cancer, anemia, malaria etc), as well as HIV-associated diseases. The metabolism of molecules such as lipids, sugars, amino acids and nucleotides was shown to be significantly disrupted by HIV and its treatment. NRTIs are still a challenge in this HAART era because of their mitochondrial toxicities and recent literature keeps highlighting this occurrence. The mitochondria still remains the most affected organelle by NTRIs. Currently more work is being conducted in improving the toxicities associated with HAART. Using UPLC-MS to monitor treatment success and failure so that patients can be switched to different regimens that are better suited for their metabolic make-up could be developed into a supporting tool for HIV/AIDS management.

CHAPTER 7: FUTURE PERSPECTIVES

Given that UPLC-MS was proven succesful for detecting metabolic changes, more samples need to be analyzed to select prominent metabolites that are potential candidate biomarkers. Sample collection can be improved by collecting sufficient demographic information inclusive of clinical symptoms. Demographics such as CD4 cell counts, viral load, other treatments already being received by study subjects, smoking etc could be very useful in follow up studies because UPLC-MS detected a broader range of compounds that are a result of some non-HIV treatments and indicative of other diseases. Improving sample collection will lead to increased confidence in concluding study findings because metabonomics data can be compared to already known or detectable clinical manifestations. An example is the patient whose supplementation for anemia was successfully linked to the patient's record.

This studies metabolites were putatively identified with the identification confidence level of three, following which definitive identification of prominent metabolites using UPLC-MS will be performed. The next step in identification is to use MSMS data to confirm the fragmentation pattern and linking the proposed structures to the mass spectral data. The third step would be to buy CRM standards of those compounds tentatively identified and confirming their retention time and mass spectral data. This will increase the identification confidence level to one from the current level three. Further validation methods such as NMR can be used to confirm the identity of potential biomarkers. NMR has already proved successful in detecting and quantifying HIV and treatment distinguishing metabolites as evident in literature. Because NMR detects the most abundant compounds in a sample, other validation methods will be needed for unknown molecules; e.g. once a potential metabolite is identified as a useful biomarker then a concentration standard of that molecule can be analyzed using UPLC-MS to confirm its identity, retention time and any other validating information. Quantification of metabolites must also be conducted.

Further biochemical characterization assays for potential biomarkers are required. This can include; lactate dehydrogenase assay which can classify the patients according to the severity of the disease stage (HIV/AIDS), antioxidant assays like ORAC that give an indication of the patients ability to scavenge free radicals which are known to be increased during HIV-infection because this study did demonstrate that the mitochondrial membrane was compromised. Also, ELISA for detecting and quantifying molecules such as DHEAS, and

cytokine studies (using flow cytometry) which can determine the degree of inflammation that is occurring as a result of HIV/AIDS. The rationale behind the above mentioned assays or techniques is because of the properties that some of the metabolites displayed during this study.

As highlighted above, the ultimate goal is to develop metabonomics as an additional tool for HIV/AIDS management; this can be achieved by discovering metabolic biomarkers. For example, metabolites that display some biomarker qualities were reported on here but further validation studies are needed. Detection, definitive identification and quantification of potential biomarkers can finally lead to the development of HIV/AIDS metabonomics database.

CHPATER 8: REFERENCES

Aberg KM, Alm E and Torgrip RJO (2009) The correspondence problem for metabonomics datasets. *Analytical and bioanalytical chemistry*. 394 (1), 151–62.

Altman DG (1991) Practical Statistics for Medical Research. First Edition London: Chapman & Hall.218.

Amand LE and Tullin CJ The Theory Behind FTIR analysis. *Dep. of Energy Conversion, Chalmers University of Technology, Sweden*. (0), 1–15.

Antinori A, Arendt G, Becker JT, Brew BJ, Byrd DA, Cherner M, Clifford DB, Cinque P, Epstein LG, Goodkin K, Gisslen M, Grant I, Heaton RK, Joseph J, Marder K, Marra CM, McArthur JC, Nunn M, Price RW, Pulliam L, Robertson KR, Sacktor N, Valcour V and Wojna VE (2007) Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*. 69 (18), 1789–99.

Apostolova N, Blas-García A and Esplugues J V (2011) Mitochondrial interference by anti-HIV drugs: mechanisms beyond Pol-γ inhibition. *Trends in pharmacological sciences*. 32 (12), 715–25.

Apostolova N, Esplugues J V, Farmacología D, Medicina F De and Valencia U De (2006) Mitochondrial DNA Replication in Health and Disease. In: *DNA replication-Current advances*. InTech. 167–182.

Arena-Pinto A, Grant AD, Edwards S and Weller IVD (2003) Lactic acidosis in HIV infected patients: a systematic review of published cases. *Sexually transmittede Infections*. 79 (2), 340–344.

Argarwal RK (2010) Kirshna's Engeneering Chemistry: (U.P.) (Theory and Practicals). Eleventh edition. KRISHNA Prakashan Media (P) Ltd: Meerut. 352-354.

Axelson DE (2010) Data Preprocessing for Chemometric and Metabonomic Analysis. MRi_Consulting, 2-3,34,64-65,122-128,252,272-277,294.

Ayala JE, Samuel VT, Morton GJ, Obici S, Croniger CM, Shulman GI, Wasserman DH and McGuinness OP (2010) Standard operating procedures for describing and performing metabolic tests of glucose homeostasis in mice. *Disease models & mechanisms*. 3 (9-10), 525–34.

Badiou S, Thiebaut R, Aurillac-Lavignolle V, Dabis F, Laporte F, Cristol JP and Mercie P (2008) Association of non-HDL cholesterol with subclinical atherosclerosis in HIV-positive patients. *The Journal of infection*. 57 (1), 47–54.

Bahmani MK, Khosravi A, Miri R, Iwabu Y, Ikuta K and Sakudo A (2009) Spectroscopic characterization of human immunodeficiency virus type-1-infected plasma by principal component analysis and soft independent modeling of class analogy of visible and near-infrared spectra. *Molecular Medicine Reports*. (24), 805–809.

Baker RG, Hayden MS and Ghosh S (2011) NF-κB, inflammation, and metabolic disease. *Cell metabolism*. 13 (1), 11–22.

Baravkar A. and Kale R (2011) FT-IR SPECTROSCOPY : PRINCIPLE , TECHNIQUE AND MATHEMATICS. *Pharmaceutical Analysis*. 2 (1), 513–519.

Barri T and Dragsted LO (2013) UPLC-ESI-QTOF/MS and multivariate data analysis for blood plasma and serum metabolomics: effect of experimental artefacts and anticoagulant. *Analytica chimica acta*. 768, 118–28.

Bartlett HP, Simonite V, Westcott E and Taylor HR (2000) A comparison of the nursing competence of graduates and diplomates from UK nursing programmes. *Journal of clinical nursing*. 9, 369–381.

Basavapathruni A and Anderson KS (2007) Reverse transcription of the HIV-1 pandemic. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 21 (14), 3795–808.

Beger RD and Colatsky T (2012) Metabolomics data and the biomarker qualification process. *Metabolomics*. 8 (1), 2–7.

Bernstein WB, & Dennis PA (2009) Repositioning HIV Protease Inhibitors as Cancer Therapeutics. *Current Opinion in HIV and AIDS*. 3(6), 666–675. doi:10.1097/COH.0b013e328313915d.Repositioning

Blümer RM, van Vonderen MG, Sutinen J, Hassink E, Ackermans M, van Agtmael M a, Yki-Jarvinen H, Danner S a, Reiss P and Sauerwein HP (2008) Zidovudine/lamivudine contributes to insulin resistance within 3 months of starting combination antiretroviral therapy. *AIDS (London, England)*. 22 (2), 227–36.

Bogner J., Vielhauser V, Beckmann RA, Michl G, Wille L, Salzberger B and Goebel F-D (2001) Stavudine Verses Zidovudine and the Development of Lipodystrophy. *Journal of acquired immune deficiency syndromes*. 27, 237–244.

Bollard ME, Holmes E, Lindon JC, Mitchell SC, Branstetter D, Zhang W and Nicholson JK (2001) Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution (1)H NMR spectroscopy of urine and pattern recognition. *Analytical biochemistry*. 295 (2), 194–202.

Brennan-Benson P (2009) Metabolic and cardiovascular complications of HIV. *Medicine*. 37 (7), 348–351. Brik A and Wong C-H (2003) HIV-1 protease: mechanism and drug discovery. *Organic & biomolecular chemistry*. 1 (1), 5–14.

Brown GC and Neher JJ (2010) Inflammatory neurodegeneration and mechanisms of microglial killing of neurons. *Molecular neurobiology*. 41 (2-3), 242–7.

Brown SC, Kruppa G and Dasseux J-L (2005) Metabolomics applications of FT-ICR mass spectrometry. *Mass spectrometry reviews*. 24 (2), 223–31.

Brown SD, Bartlett MG, White CA, Brown SD, Bartlett MG and White CA (2003) Pharmacokinetics of Intravenous Acyclovir , Zidovudine , and Acyclovir-Zidovudine in Pregnant Rats Pharmacokinetics of Intravenous Acyclovir , Zidovudine , and Acyclovir-Zidovudine in Pregnant Rats. *Antimicrobial agents and chemotherapy*. 47 (3).

Brown SD, White C a. and Bartlett MG (2002) Hplc Determination of Acyclovir and Zidovudine in Maternal Plasma, Amniotic Fluid, Fetal, and Placental Tissues Using Ultra-Violet Detection. *Journal of Liquid Chromatography & Related Technologies*. 25 (18), 2857–2871.

Brown K, Tompkins EM and White INH (2006) Applications of accelerator mass spectrometry for pharmacological and toxicological research. *Mass spectrometry reviews*. 25 (1), 127–45.

Brown TT (2008) Approach to the human immunodeficiency virus-infected patient with lipodystrophy. *The Journal of clinical endocrinology and metabolism*. 93 (8), 2937–45.

Bruce SJ, Tavazzi I, Parisod V, Rezzi S, Kochhar S and Guy P a (2009) Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. *Analytical chemistry*. 81 (9), 3285–96.

Carr A, Miller J, Law M and Cooper DA (2000) A syndrome of lipoatrophy, lactic acidaemia and liver dysfunction associated with HIV nucleoside analogue therapy: contribution to protease inhibitor-related lipodystrophy syndrome. *AIDS (London, England)*. 14 (3), F25–32.

Carr a, Samaras K, Burton S, Law M, Freund J, Chisholm DJ and Cooper D a (1998) A syndrome of peripheral lipodystrophy, hyperlipidaemia and insulin resistance in patients receiving HIV protease inhibitors. *AIDS (London, England)*. 12 (7), F51–8.

Carr A and Cooper DA (2010) Adverse drug reactions Adverse effects of antiretroviral therapy. *Adverse Drug Reactions*. 356, 1423–1430.

Carr MC (2003) The emergence of the metabolic syndrome with menopause. *The Journal of clinical endocrinology and metabolism*. 88 (6), 2404–11.

Cassol E, Misra V, Holman A, Kamat A, Morgello S and Gabuzda D (2013) Plasma metabolomics identifies lipid abnormalities linked to markers of inflammation, microbial translocation, and hepatic function in HIV patients receiving protease inhibitors. *BMC infectious diseases*. 13, 203.

Castro L and Goldani LZ (2009) Iron, folate and vitamin B12 parameters in HIV-1 infected patients with anaemia in southern Brazil. *Tropical doctor*. 39 (2), 83–5.

Chantin C, Bonin B, Boulieu R and Bory C (1996) Liquid-chromatography study of purine metabolism abnormalities in purine nucleoside phosphorylase deficiency. *Clinical Chemistry*. 42 (2), 326–328.

Charalampopoulos I, Alexaki V-I, Tsatsanis C, Minas V, Dermitzaki E, Lasaridis I, Vardouli L, Stournaras C, Margioris AN, Castanas E and Gravanis A (2006) Neurosteroids as endogenous inhibitors of neuronal cell apoptosis in aging. *Annals of the New York Academy of Sciences*. 1088, 139–52.

Chen C and Gonzalez FJ (2007) LC-MS-BASED METABOLOMICS IN DRUG METABOLISM. *Drug Metabolism*. 39, 581–597.

Chen D, Misra A and Garg A (2002) Clinical review 153: Lipodystrophy in human immunodeficiency virus-infected patients. *The Journal of clinical endocrinology and metabolism*. 87 (11), 4845–56.

Chen K-C, Wang T-Y and Chan C (2012) Associations between HIV and human pathways revealed by protein-protein interactions and correlated gene expression profiles. *PloS one*. 7 (3), e34240.

Chen M, Carson W, Calderbank R and Carin L (2012) Communications Inspired Linear Discriminant Analysis. *29th International Conference on machine Learning. Edinburgh. Scotland. UK*.

Chen Q, Park H-C, Goligorsky MS, Chander P, Fischer SM and Gross SS (2012) Untargeted plasma metabolite profiling reveals the broad systemic consequences of xanthine oxidoreductase inactivation in mice. *PloS one*. 7 (6), e37149.

Claster S (2002) Biology of anemia, differential diagnosis, and treatment options in human immunodeficiency virus infection. *The Journal of infectious diseases*. 185 Suppl , S105–9.

De Clercq E (1998) The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. *Antiviral research*. 38 (3), 153–79.

Cordero O., Salgado F., Fernadez-Alonso C., Herrera C, Lluis C, Franco R and Nogueira M (2001) Cytokines regulate membrane adenosine deaminase on human activated lymphocytes anchor adenosine deaminase (ADA) on the T cell plasma membranes , and the different effect of CD26 indicated that cytokines regulate the trans-. *Journal of Leukocyte Biology*. 70, 920–930.

Couture P (2005) HIV & Metabolism : What Are the Complications ? *The Canadian Journal of CME*. (August), 65–67.

Croarkin C and Tobias P (2001) Mean Plot. In: *NIST/SEMATECH Engineering Statistics Handbook*. 1.3.3.

Das N. (2003) Adenosine in Sepsis and Septic Shock. *Advances in Sepsis*. 3 (2), 56–61.

Daykin C a, Foxall PJD, Connor SC, Lindon JC and Nicholson JK (2002) The comparison of plasma deproteinization methods for the detection of low-molecular-weight metabolites by (1)H nuclear magnetic resonance spectroscopy. *Analytical biochemistry*. 304 (2), 220–30.

Dettmer K, Aronov PA and Hammock BD (2007) MASS SPECTROMETRY-BASED METABOLOMICS. *Mass spectrometry reviews*. 26, 51–78.

Drake SM (2000) JAC Leading article NNRTIs — a new class of drugs for HIV. *Journal of Antimicrobial Chemotherapy*. 45, 417–420.

Dritsa V (2012) FT-IR Spectroscopy in Medicine. In: P. T. Theophile ed. *Infrared Spectroscopy-Life and Biomedical Sciences*. InTech. 271–288.

Dubé MP, Stein JH, Aberg J a, Fichtenbaum CJ, Gerber JG, Tashima KT, Henry WK, Currier JS, Sprecher D and Glesby MJ (2003) Guidelines for the evaluation and management of dyslipidemia in human immunodeficiency virus (HIV)-infected adults receiving antiretroviral therapy: recommendations of the HIV Medical Association of the Infectious Disease Society of America and the Adult . *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 37 (5), 613–27.

Dube MP (2006) HIV-Associated Lipoatrophy : What Are the Kinder , Gentler Agents ? *HIV/AIDS*. 42, 281–282.

Dunn WB, Bailey NJC and Johnson HE (2005) Measuring the metabolome: current analytical technologies. *The Analyst*. 130 (5), 606–25.

Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles JD, Halsall A, Haselden JN, Nicholls AW, Wilson ID, Kell DB and Goodacre R (2011) Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature protocols*. 6 (7), 1060–83.

Dunn WB, Erban A, Weber RJM, Creek DJ, Brown M, Breitling R, Hankemeier T, Goodacre R, Neumann S, Kopka J and Viant MR (2012) Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*. 9 (S1), 44–66.

Eriksson L, Antti H, Gottfries J, Holmes E, Johansson E, Lindgren F, Long I, Lundstedt T, Trygg J and Wold S (2004) Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). *Analytical and bioanalytical chemistry*. 380 (3), 419–29.

Eriksson L, Johansson E, Kettaneh-Wold N, Trygg C, Wikström C and Wold S (2006) PCA. In: *Multi- and Megavariate Data Analysis Part 1, Basic Principles and Applications*. John Wiley and Sons, Ltd. 39–62.

Estrada V and Portilla J (2011) Dyslipidemia related to antiretroviral therapy. *AIDS reviews*. 13 (1), 49–56.

Faoláin EO, Hunter MB, Byrne JM, Kelehan P, Lambkin H a, Byrne HJ and Lyng FM (2005) Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*. 53 (1), 121–9.

Fazekas L, Horkay F, Kekesi V, Huszar E, Fazekas R, Juhkz-nagy A and Naszlady A (1999) Pi1 soo24-3205(!b)oo331-8. *Life Sciences*. 65 (10), 1005–1012.

Feeney ER and Mallon PWG (2011) Insulin resistance in treated HIV infection. *Best practice & research. Clinical endocrinology & metabolism*. 25 (3), 443–58.

Fehintola FA, Akinyinka OO, Adewole IF, Maponga CC, Ma Q and Morse GD (2011) Drug Interactions in the Treatment and Chemoprophylaxis of Malaria and HIV Infected Individuals in Sub Saharan Africa. *Current Drug metabolism*. 12 (1), 51–56.

Feng Y, Wang N, Zhu M, Feng Y, Li H and Tsao S (2011) Recent progress on anticancer candidates in patents of herbal medicinal products. *Recent patents on food, nutrition & agriculture*. 3 (1), 30–48.

Fleischman A, Johnsen S, Systrom DM, Hrovat M, Farrar CT, Frontera W, Fitch K, Thomas BJ, Torriani M, Côté HCF and Grinspoon SK (2007) Effects of a nucleoside reverse transcriptase inhibitor, stavudine, on glucose disposal and mitochondrial function in muscle of healthy adults. *American journal of physiology. Endocrinology and metabolism*. 292 (6), E1666–73.

Flexner C (1998) HIV-PROTEASE INHIBITORS. *The New England Journal of Medicine*. 338 (18), 1281–1292.

Ford E (2005) Risks for All-Cause Mortality , Cardiovascular Disease , and Diabetes. *Diabetes Care*. 28, 1769–1778.

Ford ES, Giles WH and Dietz WH (2002) Prevalence of the Metabolic Syndrome Among US Adults. *JAMA : the journal of the American Medical Association*. 287 (3), 356–359.

Funaya H, Kitakaze M, Node K, Minamino T, Komamura K and Hori M (1997) Plasma adenosine levels increase in patients with chronic heart failure. *Circulation*. 95 (6), 1363–5.

Lindberg GL, Rastam B, Gullberg GA, Eklund ST (1991) Serum sialic acid concentration and smoking : a population based study lecturer in medical statistics Surveillance of antibiotic resistance in clinical isolates of Neisseria gonorrhoeae. *Bill Moyers Journal*. 303 (November), 1306–1307.

Gerschenson M and Brinkman K (2004) Mitochondrial dysfunction in AIDS and its treatment. *Mitochondrion*. 4 (5-6), 763–77.

Ghannoum MA, Mukherjee PK, Jurevic RJ, Retuerto M, Brown RE, Sikaroodi M, Webster-Cyriaque J and Gillevet PM (2013) Metabolomics reveals differential levels of oral metabolites in HIV-infected patients: toward novel diagnostic targets. *Omics : a journal of integrative biology*. 17 (1), 5–15.

Gika HG, Theodoridis GA, Wingate JE, & Wilson, ID (2007) Within-Day Reproducibility of an HPLC-MS-Based Method for Metabonomic Analysis: Application to Human Urine. Journal of Proteome Research, doi: 10.1021/pr070183p

Gika HG, Theodoridis G a and Wilson ID (2008) Liquid chromatography and ultra-performance liquid chromatography-mass spectrometry fingerprinting of human urine: sample stability under different handling and storage conditions for metabonomics studies. *Journal of chromatography. A*. 1189 (1-2), 314–22.

Gisolf EH, Dreezen C, Danner S a, Weel JL and Weverling GJ (2000) Risk factors for hepatotoxicity in HIV-1-infected patients receiving ritonavir and saquinavir with or

without stavudine. Prometheus Study Group. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 31 (5), 1234–9.

Gkrania-Klotsas E and Klotsas A-E (2007a) HIV and HIV treatment: effects on fats, glucose and lipids. *British medical bulletin*. 84, 49–68.

Gkrania-Klotsas E and Klotsas A-E (2007b) HIV and HIV treatment: effects on fats, glucose and lipids. *British medical bulletin*. 84, 49–68.

Gomez DE, Armando RG and Alonso DF (2012) AZT as a telomerase inhibitor. *Frontiers in oncology*. 2 (September), 113.

Griffith, O. W & Mulcahy, R. T. (1999).The enzymes of glutathione synthesis: gamma-glutamylcysteine synthetase. Advances in Enzymology and Related Areas of Molecular Biology, 73, 209-267.

Grulich AE, van Leeuwen MT, Falster MO and Vajdic CM (2007) Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. *Lancet*. 370 (9581), 59–67.

Grunfeld C and Tien P (2003) Difficulties in understanding the metabolic complications of acquired immune deficiency syndrome. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 37 Suppl 2 (DK 57508), S43–6.

Harz M, Rösch P and Popp J (2009) Vibrational spectroscopy--a powerful tool for the rapid identification of microbial cells at the single-cell level. *Cytometry. Part A : the journal of the International Society for Analytical Cytology*. 75 (2), 104–13.

Hattingh Z, Walsch C, Veldman F and Bester C (2009) The metabolic profiles of HIV-infected and non-infected women in Mangaung , South Africa. *South African Clinical Journal of Nutrition*. 22 (1), 23–28.

Haubrich RH, Riddler SA, Dirienzo AG, Komarow L, Haas DW, Mellors JW, Havlir D V, Clinical A and Group T (2010) Metabolic Outcomes in a Randomized Trial of Nucleoside, Nonnucleoside and Protese Inhibitor-Sparing Regimens for Initial HIV Treatment. *AIDS*. 23 (9), 1109–1118.

Herzenberg L a, De Rosa SC, Dubs JG, Roederer M, Anderson MT, Ela SW and Deresinski SC (1997) Glutathione deficiency is associated with impaired survival in HIV disease. *Proceedings of the National Academy of Sciences of the United States of America*. 94 (5), 1967–72.

Hester EK (2012) HIV medications: an update and review of metabolic complications. *Nutrition in clinical practice : official publication of the American Society for Parenteral and Enteral Nutrition*. 27 (1), 51–64.

Hetherington S, McGuirk S, Powell G, Cutrell a, Naderer O, Spreen B, Lafon S, Pearce G and Steel H (2001) Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. *Clinical therapeutics*. 23 (10), 1603–14.

Hewer R, Vorster J, Steffens FE and Meyer D (2006) Applying biofluid 1H NMR-based metabonomic techniques to distinguish between HIV-1 positive/AIDS patients on antiretroviral treatment and HIV-1 negative individuals. *Journal of pharmaceutical and biomedical analysis*. 41 (4), 1442–6.

Holick MF (2007) Vitamin D deficiency. *The New England journal of medicine*. 357 (3), 266–81. 266–81. doi:10.1056/NEJMra070553

Hollenbaugh J a, Munger J and Kim B (2011) Metabolite profiles of human immunodeficiency virus infected CD4+ T cells and macrophages using LC-MS/MS analysis. *Virology*. 415 (2), 153–9.

Hollywood K, Brison DR and Goodacre R (2006) Metabolomics: current technologies and future trends. *Proteomics*. 6 (17), 4716–23.

Holmes E, Nicholls a W, Lindon JC, Connor SC, Connelly JC, Haselden JN, Damment SJ, Spraul M, Neidig P and Nicholson JK (2000) Chemometric models for toxicity classification based on NMR spectra of biofluids. *Chemical research in toxicology*. 13 (6), 471–8.

Holvoet P (2008) Oxidized LDL and the metabolic syndrome. *Future Lipidology*. 3 (6), 637–649.

Holvoet P, Stassen J-M, Van Cleemput J, Collen D and Vanhaecke J (1998) Oxidized Low Density Lipoproteins in Patients With Transplant-Associated Coronary Artery Disease. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 18 (1), 100–107.

Hommes MJT, Romijn JA, Godfried MH, Eeftinck Schattenkerk JKM, Buurman WA, Endert E, Sauerwein HP (1990) Increased resting energy expenditure in human immunodeficiency virus-infected men. *Metabolism-Clinical and Experimental journal*. 39, 1186-1190.

Hommes M, Romijn J, Endert E, Sauerwein H (1991) Resting energy expenditure and substrate oxidation in human immunodeficiency virus (HIV)-infected asymptomatic

men: HIV affects host metabolism in the early asymptomatic stage. *The American Journal of Clinical Nutrition*. 54, 311-315.

Hortin GL, Landt M and Powderly WG (1994) Changes in plasma amino acid concentrations in response to HIV-1 infection. *Clinical chemistry*. 40 (5), 785–9.

Idborg H, Zamani L, Edlund P-O, Schuppe-Koistinen I and Jacobsson SP (2005) Metabolic fingerprinting of rat urine by LC/MS Part 2. Data pretreatment methods for handling of complex data. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*. 828 (1-2), 14–20.

Ilin A and Raiko T (2010) Practical Approaches to Principal Component Analysis in the Presence of Missing Values. *Journal of Machine Learning Research*. 11, 1957–2000.

Jackson M and Mantsch HH (1995) The use and misuse of FTIR spectroscopy in the determination of protein structure. *Critical reviews in biochemistry and molecular biology*. 30 (2), 95–120.

Jevtović D, Ranin J, Salemović D, Pesić I, Dragović G, Zerjav S and Djurković-Djaković O (2008) The prevalence and risk of hepatitis flares in a Serbian cohort of HIV and HCV co-infected patients treated with HAART. *Biomedicine & pharmacotherapy = Biomédecine & pharmacothérapie*. 62 (1), 21–5.

Jevtović DJ, Dragović G, Salemović D, Ranin J and Djurković-Djaković O (2009) The metabolic syndrome, an epidemic among HIV-infected patients on HAART. *Biomedicine & pharmacotherapy = Biomédecine & pharmacothérapie*. 63 (5), 337–42.

Jiye A, Trygg J, Gullberg J, Johansson AI, Jonsson P, Antti H, Marklund SL and Moritz T (2005) Extraction and GC/MS analysis of the human blood plasma metabolome. *Analytical chemistry*. 77 (24), 8086–94.

John M, Moore CB, James IR, Nolan D, Upton RP, McKinnon EJ and Mallal S a (2001) Chronic hyperlactatemia in HIV-infected patients taking antiretroviral therapy. *AIDS (London, England)*. 15 (6), 717–23.

Juhascik MP and Jenkins AJ (2009) Comparison of liquid/liquid and solid-phase extraction for alkaline drugs. *Journal of chromatographic science*. 47 (7), 553–7.

Kaiser LG, Schuff N, Cashdollar N and Weiner MW (2005) Age-related glutamate and glutamine concentration changes in normal human brain: 1H MR spectroscopy study at 4 T. *Neurobiology of aging*. 26 (5), 665–72.

Kakuda TN (2000) Pharmacology of nucleoside and nucleotide reverse transcriptase inhibitor-induced mitochondrial toxicity. *Clinical therapeutics*. 22 (6), 685–708.

Kalebic T, Kinter a, Poli G, Anderson ME, Meister a and Fauci a S (1991) Suppression of human immunodeficiency virus expression in chronically infected monocytic cells by glutathione, glutathione ester, and N-acetylcysteine. *Proceedings of the National Academy of Sciences of the United States of America*. 88 (3), 986–90.

Kandathil AJ, Kannangai R, David S, Selvakumar R, Job V, Abraham OC and Sridharan G (2005) Human Immunodeficiency Virus Infection and Levels of Dehydroepiandrosterone Sulfate in Plasma among Indians. *Clinical and Diagnostic laboratory immunology*. 12 (9), 1117–1118.

Kanekar A (2010) Biomarkers predicting progression of human immunodeficiency virus-related disease. *Journal of clinical medicine research*. 2 (2), 55–61.

Karacostas V, Nagashima K, Gonde MA and Moss B (1989) Human immunodeficiency virus-like particles produced by a vaccinia virusexpression vector. *Proceedings of the National Academy of Sciences of the United States of America*. 86 (November), 8964–8967.

Kaul M (2009) HIV-1 associated dimentia. *Current Opinion in Neurology*. 22 (3), 315–320.doi:10.1097/WCO.0b013e328329cf3c.HIV-1

Kebarle P and Tang L (1993) FROM IONS IN SOLUTION TO IONS IN THE GAS PHASE: The Mechanism of Electrospray Mass Spectrometry. *Analytical chemistry*. 65 (22), 972–986.

Kell DB (2007) Metabolomic biomarkers: search, discovery and validation. *Expert review of molecular diagnostics*. 7 (4), 329–33. doi:10.1586/14737159.7.4.329

Kanehisa M and Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 28 (1), 27–30. Available at: http://www.genome.jp/kegg/kegg1.html

Khandekar, M. J., Cohen, P. & Spiegelman, (2011). Molecular machanisms of cancer development in obesity. *Nature Reviews cancer,11,* 886-895.

Khaw, K. T. (1996). Dehydroepiandrosterone, dehydroepiandosterone sulphate and cardiovascular disease. *Journal of Endocrinology,150,* 3 Suppl. S149-S153.

Kline M., Van Dyke RB, Lindsey J., Gwynne M, Culnane M, McKinney R., Nichols JS, Mitchell WG, Yogev R, Hutcheon N and Team TACTG 240 (1998) A Randomized Comparative Trial of Stavudine (d4T) Versus Zidovudine (ZDV, AZT) in Children With Human Immunodeficiency Virus Infection. *Pediatrics*. 101 (2), 214–220. doi:10.1542/peds.101.2.214

Kohler JJ and Lewis W (2007) Review Article A Brief Overview of Mechanisms of Mitochondrial Toxicity From NRTIs. *Enviromental and Molecular Mutagenesis*. 48, 166–172.

Kontorinis N and Dieterich D (2010) Hepatotoxicity of Antiretroviral Therapy © Permanyer Publications 2010 © Permanyer Publications 2010. , 36–43.

Kristensen M, Engelsen SB and Dragsted LO (2012) LC–MS metabolomics top-down approach reveals new exposure and effect biomarkers of apple and apple-pectin intake. *Metabolomics*. 8 (1), 64–73.

Kubota S, Shetty S, Zhang H and Pomerantz RJ (1998) Novel Inhibitory Effects of γ -Glutamylcysteine Ethyl Ester against Human Immunodeficiency Virus Type 1 Production and Propagation Novel Inhibitory Effects of γ -Glutamylcysteine Ethyl Ester against Human Immunodeficiency Virus Type 1 Production and Propa. *Antimicrobial Agents and Chemotherapy*. 42 (5), 1200–1206.

Kuhara T (2005) Gas chromatographic-mass spectrometric urinary metabolome analysis to study mutations of inborn errors of metabolism. *Mass spectrometry reviews*. 24 (6), 814–27. 27. doi:10.1002/mas.20038

Lacorete S and Fernandez-alba AR (2006) TIME OF FLIGHT MASS SPECTROMETRY APPLIED TO THE LIQUID CHROMATOGRAPHIC ANALYSIS OF PESTICIDES. *Mass Spectrometry Reviews*. 25, 866–880. doi:10.1002/mas

Lacorte S and Barceló D (1996) Determination of parts per trillion levels of organophosphorus pesticides in groundwater by automated on-line liquid-solid extraction followed by liquid chromatography/atmospheric pressure chemical ionization mass spectrometry using positive and negative . *Analytical chemistry*. 68 (15), 2464–70.

Lai J, Wang J and Cai Z (2008) Nucleoside reverse transcriptase inhibitors and their phosphorylated metabolites in human immunodeficiency virus-infected human matrices. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*. 868 (1-2), 1–12.

Lakshmi LJ, Zephy D and Bhaskar MV (2013) Adenosine deaminase activity in hiv positive cases. *International Journal of Bioassays*. 02 (12), 1553–1556.

Lambert P-H, Boutin J a., Bertin S, Fauchère J-L and Volland J-P (1997) Evaluation of high performance liquid chromatography/electrospray mass spectrometry with selected ion monitoring for the analysis of large synthetic combinatorial peptide libraries. *Rapid Communications in Mass Spectrometry*. 11 (18), 1971–1976.

Lambert WE, Van Bocxlaer JF and De Leenheer a P (1997) Potential of high-performance liquid chromatography with photodiode array detection in forensic toxicology. *Journal of chromatography. B, Biomedical sciences and applications*. 689 (1), 45–53.

Laurichesse H, Tauveron I, Gourdon F, Cormerais L, Champredon C, Charrier S, Lamain S, Laveran H, Thieblot P, Beytout J and Grizard J (1998) Human Nutrition and Metabolism Threonine and Methionine Are Limiting Amino Acids for Protein Synthesis in Patients with AIDS 1 , 2 , 3. *The Journal of Nutrition*. (April), 1342–1348.

Lebovitz HE (2001) Insulin resistance: definition and consequences. *Experimental and clinical endocrinology & diabetes*. 109 Suppl , S135–48.

Lendl B and Karlberg B (2005) Advancing from unsupervised, single variable-based to supervised, multivariate-based methods: A challenge for qualitative analysis. *rends in Analytical Chemistry*. 24 (6), 488–492.

Leskiewicz M, Jantas D, Budziszewska B and Lason W (2008) Excitatory neurosteroids attenuate apoptotic and excitotoxic cell death in primary cortical neurons. *Journal of physiology and pharmacology : an official journal of the Polish Physiological Society*. 59 (3), 457–75.

Lewis BP, Shih I, Jones-Rhoades MW, Bartel DP and Burge CB (2003) Prediction of mammalian microRNA targets. *Cell*. 115 (7), 787–98.

Lewis W, Grupp IL, Grupp G, Hoit B, Morris R, Samarel a M, Bruggeman L and Klotman P (2000) Cardiac dysfunction occurs in the HIV-1 transgenic mouse treated with zidovudine. *Laboratory investigation; a journal of technical methods and pathology*. 80 (2), 187–97.

Li Z, Li C, Lin D, Huang Z, Pan J, Chen G, Lin J, Liu N, Yu Y, Feng S and Chen R (2014) Surface-enhanced Raman spectroscopy for differentiation between benign and malignant thyroid tissues. *Laser Physics Letters*. 11 (4), 045602.

Libman, H. and Makadon, H. J. (2007) HIV, Third Edition. American College of Physicians Press: USA. 1-6.

Lichtenstein KA, Armon C, Baron A, Moorman AC, Wood KC and Holmberg SD (2005) Modification of the incidence of drug-associated symmetrical peripheral neuropathy by host and disease factors in the HIV outpatient study cohort. *Clinical infectious diseases*. 40 (1), 148–57.

Llibre JM (2009) First-Line Boosted Protease Inhibitor-Based Regimens in Treatment-Naive HIV-1-Infected Ptaients- Making a Good Thing Better. *AIDS Review*. 11, 215–22.

Lim SE and Copeland WC (2001) Differential incorporation and removal of antiviral deoxynucleotides by human DNA polymerase gamma. *The Journal of biological chemistry*. 276 (26), 23616–23.

Lim SE, Ponamarev M V., Longley MJ and Copeland WC (2003) Structural Determinants in Human DNA Polymerase γ Account for Mitochondrial Toxicity from Nucleoside Analogs. *Journal of Molecular Biology*. 329 (1), 45–57.

Lin L, Huang Z, Gao Y, Yan X, Xing J and Hang W (2011) LC-MS based serum metabonomic analysis for renal cell carcinoma diagnosis, staging, and biomarker discovery. *Journal of proteome research*. 10 (3), 1396–405.

Lodish A, Berk A, Kaiser CA, Kriger M, Scott, Bretscher A, Hidde P and Matsudaira P (2008) Molecular Cell Biology, Sixth Edition. W.H Freeman & Company: Iowa

Lopez Gonzalez, L. (2014) The South African Heath News Service. Retrieved from http://www.health-e.org.za/?s

Lu J, Plataniotis KN and Venetsanopoulos a. N (2005) Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. *Pattern Recognition Letters*. 26 (2), 181–191.

Lu X, Zhao X, Bai C, Zhao C, Lu G and Xu G (2008) LC-MS-based metabonomics analysis. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*. 866 (1-2), 64–76.

Luque-Garcia JL and Neubert T a (2007) Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. *Journal of chromatography. A*. 1153 (1-2), 259–76.

Luscombe NM, Greenbaum D and Gerstein M (2001) What is bioinformatics? A proposed definition and overview of the field. *Methods of information in medicine*. 40 (4), 346–58.

Lv H, Hung CS, Chaturvedi KS, Hooton TM and Henderson JP (2011) Development of an integrated metabolomic profiling approach for infectious diseases research. *The Analyst*. 136 (22), 4752–63.

Macallan DC, Noble C, Baldwin C, Jebb S a, Prentice a M, Coward W a, Sawyer MB, McManus TJ and Griffin GE (1995) Energy expenditure and wasting in human immunodeficiency virus infection. *The New England journal of medicine*. 333 (2), 83–8.

Maher AD, Cysique L a, Brew BJ and Rae CD (2011) Statistical integration of 1H NMR and MRS data from different biofluids and tissues enhances recovery of biological

information from individuals with HIV-1 infection. *Journal of proteome research*. 10 (4), 1737–45.

Marra F and Bertolani C (2009) Adipokines in liver diseases. *Hepatology (Baltimore, Md.)*. 50 (3), 957–69.

Mary-Krause M, Cotte L, Simon A, Partisani M and Costagliola D (2003) Increased risk of myocardial infarction with duration of protease inhibitor therapy in HIV-infected men. *AIDS (London, England)*. 17 (17), 2479–86.

Matsumoto I and Kuhara T (1997) A NEW CHEMICAL DIAGNOSTIC METHOD FOR INBORN ERRORS OF METABOLISM BY MASS SPECTROMETRY-RAPID, PRACTICAL, ANDSIMULTANEOUS URINARY METABOLITES ANALYSIS. *Mass spectrometry reviews*. 15 (June 1996), 43–57.

McComsey GA, Kitch D, Sax PE, Tebas P, Tierney C, Jahed NC, Myers L, Melbourne K, Ha B and Daar ES (2011) Peripheral and central fat changes in subjects randomized to abacavir-lamivudine or tenofovir-emtricitabine with atazanavir-ritonavir or efavirenz: ACTG Study A5224s. *Clinical infectious diseases*. 53 (2), 185–96.

McComsey G a and Leonard E (2004) Metabolic complications of HIV therapy in children. *AIDS (London, England)*. 18 (13), 1753–68.

McPhee F, Caldera PS, Bemis GW, McDonagh a F, Kuntz ID and Craik CS (1996) Bile pigments as HIV-1 protease inhibitors and their effects on HIV-1 viral maturation and infectivity in vitro. *The Biochemical journal*. 320 (Pt 2, 681–6.

Meister a (1991) Glutathione deficiency produced by inhibition of its synthesis, and its reversal; applications in research and therapy. *Pharmacology & therapeutics*. 51 (2), 155–94.

Michopoulos F, Lai L, Gika H, Theodordis G and Wilson I (2009) UPLC-MS-based analysis of human plasma for metabonomics using solvent precipitation or solid phase extraction. *Journal of proteome research*. 8 (4), 2114–21.

Misra I and Griffith OW (1998) Expression and Purification of Human gamma-Glutamylcysteine Synthetase. *Protein Expression and Purification*. 13, 268–276.

Moore RD (1999) Human immunodeficiency virus infection, anemia, and survival. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 29 (1), 44–9.

Moore RD, Wong WM, Keruly JC and McArthur JC (2000) Incidence of neuropathy in HIV-infected patients on monotherapy versus those on combination therapy with didanosine, stavudine and hydroxyurea. *AIDS*. 14 (3), 273–8.

Munger J, Bajad SU, Coller H a, Shenk T and Rabinowitz JD (2006) Dynamics of the cellular metabolome during human cytomegalovirus infection. *PLoS pathogens*. 2 (12), e132.

Munger J, Bennett BD, Parikh A, Feng X, Rabitz HA, Shenk T and Rabinowitz JD (2010) Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nature Biotechnology*. 26 (10), 1179–1186.

Munshi SU, Rewari BB, Bhavesh NS and Jameel S (2013) Nuclear magnetic resonance based profiling of biofluids reveals metabolic dysregulation in HIV-infected persons and those on anti-retroviral therapy. *PloS one*. 8 (5), e64298.

Murphy EL, Collier AC, Kalish LA, Assmann SF, Para MF, Flanigan TP, Kumar PN, Mintz L, Wallach FR and Nemo GJ (2001) Article Highly Active Antiretroviral Therapy Decreases Mortality and Morbidity in Patients with Advanced HIV Disease. *Annals of Internal Medicine*. 135, 17–26.

Mussini C, Pinti M, Bugarini R, Borghi V, Nasi M, Nemes E, Troiano L, Guaraldi G, Bedini A, Sabin C, Esposito R and Cossarizza A (2005) Effect of treatment interruption monitored by CD4 cell count on mitochondrial DNA content in HIV-infected patients: a prospective study. *AIDS (London, England)*. 19 (15), 1627–33.

Nakav S, Chaimovitz C, Sufaro Y, Lewis EC, Shaked G, Czeiger D, Zlotnik M and Douvdevani A (2008) Anti-inflammatory preconditioning by agonists of adenosine A1 receptor. *PloS one*. 3 (5), e2107.

Neuhaus J, Jacobs DR, Baker J V, Calmy A, Duprez D, La Rosa A, Kuller LH, Pett SL, Ristola M, Ross MJ, Shlipak MG, Tracy R and Neaton JD (2010) Markers of inflammation, coagulation, and renal function are elevated in adults with HIV infection. *The Journal of infectious diseases*. 201 (12), 1788–95.

NIAID (2014) The HIV Life Cycle. Available at http://www.niaid.nih.gov/topics/HIVAIDS/UnderstandingBiology/pages/hivreplicationcycle.aspx

Nicholson JK, Lindon JC and Holmes E (1999) "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica; the fate of foreign compounds in biological systems*. 29 (11), 1181–9.

Nicholson JK, Connelly J, Lindon JC, & Holmes E (2002) Metabonomics: a platform for studying drug toxicity and gene function. *Neture Reviews Drug Discovery*,1(2), 153-161.

Nicolet T and All C (2001) Introduction to Fourier Transform Infrared Spectrometry. , 1–8.

Nordström A and Lewensohn R (2010) Metabolomics: moving to the clinic. *Journal of neuroimmune pharmacology : the official journal of the Society on NeuroImmune Pharmacology*. 5 (1), 4–17.

Nováková L, Matysová L and Solich P (2006) Advantages of application of UPLC in pharmaceutical analysis. *Talanta*. 68 (3), 908–18.

Pendyala G, Want EJ, Webb W, Siuzdak G and Fox HS (2007) Biomarkers for neuroAIDS: the widening scope of metabolomics. *Journal of neuroimmune pharmacology*. 2 (1), 72–80.

Peng J, Cai S, Wang L, Zhao N, Zhang T, Chen Z and Meng F (2014) A metabonomic analysis of serum from rats treated with ricinine using ultra performance liquid chromatography coupled to mass spectrometry. *PloS one*. 9 (3), e90416. doi:10.1371/journal.pone.0090416

Petibois C, Rigalleau V, Melin a M, Perromat a, Cazorla G, Gin H and Déléris G (1999) Determination of glucose in dried serum samples by Fourier-transform infrared spectroscopy. *Clinical chemistry*. 45 (9), 1530–5.

Philippeos C, Steffens FE and Meyer D (2009) Comparative 1H NMR-based metabonomic analysis of HIV-1 sera. *Journal of biomolecular NMR*. 44 (3), 127–37.

Pijanka J, Sockalingum GD, Kohler A, Yang Y, Draux F, Parkes G, Lam K-P, Collins D, Dumas P, Sandt C, van Pittius DG, Douce G, Manfait M, Untereiner V and Sulé-Suso J (2010) Synchrotron-based FTIR spectra of stained single cells. Towards a clinical application in pathology. *Laboratory investigation; a journal of technical methods and pathology*. 90 (5), 797–807.

Pinti M, Nasi M, Gibellini L, Roat E, De Biasi S, Bertoncelli L and Cossarizza A (2010) The Role of Mitochondria in HIV Infection and Its Treatment. *Journal of Experimental & Clinical Medicine*. 2 (4), 145–155.

Pinti M, Salomoni P and Cossarizza A (2006) Anti-HIV drugs and the mitochondria. *Biochimica et biophysica acta*. 1757 (5-6), 700–7.

Pirrone V, Thakkar N, Jacobson JM, Wigdahl B and Krebs FC (2011) Combinatorial approaches to the prevention and treatment of HIV-1 infection. *Antimicrobial agents and chemotherapy*. 55 (5), 1831–42.

Plumb RS, Granger JH, Stumpf CL, Johnson K a, Smith BW, Gaulitz S, Wilson ID and Castro-Perez J (2005) A rapid screening approach to metabonomics using UPLC and oa-TOF mass spectrometry: application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice. *The Analyst*. 130 (6), 844–9.

Powers HJ (2003) Riboflavin (vitamin B-2) and health. *The American journal of clinical nutrition*. 77 (6), 1352–60.

Prasad VR and Goff SP (1990) Structure-function studies of HIV reverse transcriptase. *Annals of the New York Academy of Sciences*. 616, 11–21.

Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, Young N, Xia J, Knox C, Dong E, Huang P, Hollander Z, Pedersen TL, Smith SR, Bamforth F, Greiner R, McManus B, Newman JW, Goodfriend T and Wishart DS (2011) The Human Serum Metabolome D. Flower ed. *PLoS ONE*. 6 (2), e16957.

Qi S, Tu Z, Peng W, Wang L, Ou-yang X, Cai A and Dai Y (2012) H NMR-based serum metabolic profiling in compensated and decompensated cirrhosis. *World Journal of Gastroenterology*. 18 (3), 285–290.

Ramachandran S and Velraj G (2012) Ft-ir, ft-raman spectral analysis and density functional theory calculations studies of 3-chloro-2-nitrobenzyl alcohol. *Romanian Journal of Physics*. 57 (7-8), 1128–1137.

Ramaih, S. (2008). HIV/AIDS: Health Solutions. Sterling Publishers: New Delhi. 1-17.

Ritsner MS (2010) Pregnenolone, dehydroepiandrosterone, and schizophrenia: alterations and clinical trials. *CNS neuroscience & therapeutics*. 16 (1), 32–44.

Robinson FP (2004) HIV lipodystrophy syndrome: a primer. *The Journal of the Association of Nurses in AIDS Care : JANAC*. 15 (1), 15–29.

Robinson-Papp J and Simpson DM (2009) Neuromuscular diseases associated with HIV-1 infection. *Muscle & nerve*. 40 (6), 1043–53.

Robosky LC, Wells DF, Egnash L a, Manning ML, Reily MD and Robertson DG (2005) Metabonomic identification of two distinct phenotypes in Sprague-Dawley (Crl:CD(SD)) rats. *Toxicological sciences : an official journal of the Society of Toxicology*. 87 (1), 277–84.

Rose H, Hoy J, Woolley I, Tchoua U, Bukrinsky M, Dart A and Sviridov D (2008) HIV infection and high density lipoprotein metabolism. *Atherosclerosis*. 199 (1), 79–86.

Safrin S and Grunfeld C (1999) Fat distribution and metabolic changes in patients with HIV infection. *AIDS (London, England)*. 13 (18), 2493–505.

Samaras K, Wand H, Law M, Emery S, Cooper D and Carr A (2007) Prevalence of metabolic syndrome in HIV-infected patients receiving highly active antiretroviral therapy using International Diabetes Foundation and Adult Treatment Panel III criteria: associations with insulin resistance, disturbed body fat compartmental. *Diabetes care*. 30 (1), 113–9.

Schambelan M, Benson C a, Carr A, Currier JS, Dubé MP, Gerber JG, Grinspoon SK, Grunfeld C, Kotler DP, Mulligan K, Powderly WG and Saag MS (2002) Management of metabolic complications associated with antiretroviral therapy for HIV-1 infection: recommendations of an International AIDS Society-USA panel. *Journal of acquired immune deficiency syndromes (1999)*. 31 (3), 257–75.

Schymanski EL and Neumann S (2013) CASMI: And the Winner is . . . *Metabolites*. 3 (2), 412–39.

See AP, Zeng J, Tran PT and Lim M (2011) Acute toxicity of second generation HIV protease-inhibitors in combination with radiotherapy: a retrospective case series. *Radiation oncology (London, England)*. 6 (1), 25.

Serkova NJ and Niemann C (2006) Pattern recognition and biomarker validation using quantitative ^{1}H-NMR-based metabonomics. *Expert Review of Molecular Diagnostics.* 6 (5), 717-731.

Shaffer R (2002) Book Review: MULTI- AND MEGA DATA ANALYSIS, PRINCIPLES AND APPLICATIONS,I. *Journal of Chemometrics*. 16, 261–262.

Shaw RA and Mantsch HH (2000) Infrared Spectroscopy in Clinical and Diagnostic Analysis. In: *Encyclopedia of Analytical Chemistry*. Chichester: John Wiley and Sons, Ltd. 1–19.

Sheikh KD, Khanna S, Byers SW, Fornace A and Cheema AK (2011) Small molecule metabolite extraction strategy for improving LC/MS detection of cancer cell metabolome. *Journal of biomolecular techniques : JBT*. 22 (1), 1–4.

Sheng D, Liu X, Li W, Wang Y, Chen X and Wang X (2013) Distinction of leukemia patients' and healthy persons' serum using FTIR spectroscopy. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy*. 101, 228–32.

Sierra M, Thomson MM, Ríos M, Casado G, Castro RO, Delgado E, Echevarría G, Muñoz M, Colomina J, Carmona R, Vega Y, Parga EV, Medrano L, Pérez-Alvarez L, Contreras

G and Nájera R (2005) The analysis of near full-length genome sequences of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12_BF. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 5 (3), 209–17.

Sierra S, Kupfer B and Kaiser R (2005) Basics of the virology of HIV-1 and its replication. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 34 (4), 233–44.

Silver H, Knoll G, Isakov V, Goodman C and Finkelstein Y (2005) Blood DHEAS concentrations correlate with cognitive function in chronic schizophrenia patients: a pilot study. *Journal of psychiatric research*. 39 (6), 569–75.

Sitole L, Steffens F, Krüger TPJ and Meyer D (2014) Mid-ATR-FTIR Spectroscopic Profiling of HIV/AIDS Sera for Novel Systems Diagnostics in Global Health. *Omics : a journal of integrative biology*. 18 (00), 1–11.

Sitole LJ, Williams A a and Meyer D (2013) Metabonomic analysis of HIV-infected biofluids. *Molecular bioSystems*. 9 (1), 18–28.

Smith ICP and Baert R (2003) Medical diagnosis by high resolution NMR of human specimens. *IUBMB life*. 55 (4-5), 273–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12880209 (accessed 26/06/14).

Somasundaran M, Zapp ML, Beattie LK, Pang L, Byron KS, Bassell GJ, Sullivan JL and Singer RH (1994) Localization of HIV RNA in mitochondria of infected cells: potential role in cytopathogenicity. *The Journal of cell biology*. 126 (6), 1353–60.

Statistics South Africa (2013) Statistical release Mid-year population estimates. . (July). Available at: http://beta2.statssa.gov.za/publications/P0302/P03022013.pdf.

Steen RJ, Bobeldijk I and Brinkman U a (2001) Screening for transformation products of pesticides using tandem mass spectrometric scan modes. *Journal of chromatography. A*. 915 (1-2), 129–37.

Stenlund H, Gorzsás A, Persson P, Sundberg B and Trygg J (2008) Orthogonal projections to latent structures discriminant analysis modeling on in situ FT-IR spectral imaging of liver tissue for identifying sources of variability. *Analytical chemistry*. 80 (18), 6898–6906. .

Stoll B (2006) Intestinal Uptake and Metabolism of Threonine : Nutritional Impact. *Advances in Pork Production*. 17, 257–263.

Strott C a and Higashi Y (2003) Cholesterol sulfate in human physiology: what's it all about? *Journal of lipid research*. 44 (7), 1268–78.

Subbaraman R, Chaguturu SK, Mayer KH, Flanigan TP and Kumarasamy N (2007a) Adverse effects of highly active antiretroviral therapy in developing countries. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 45 (8), 1093–101.

Subbaraman R, Chaguturu SK, Mayer KH, Flanigan TP and Kumarasamy N (2007) Adverse effects of highly active antiretroviral therapy in developing countries. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 45 (8), 1093–101.

Swartz ME (2005) Ultra Performance Liquid Chromatography (UPLC): An Introduction. *SEPARATION SCIENCE REDEFINED*. (May), 8–14.

Tabachnick BG and Fidell LS (1996) Using Multivariate Statistics. Harper Collins College Publishers: New York.

Tavazzi B, Lazzarino G, Leone P, Amorini AM, Bellia F, Janson CG, Di Pietro V, Ceccarelli L, Donzelli S, Francis JS and Giardina B (2005) Simultaneous high performance liquid chromatographic separation of purines, pyrimidines, N-acetylated amino acids, and dicarboxylic acids for the chemical diagnosis of inborn errors of metabolism. *Clinical biochemistry*. 38 (11), 997–1008.

Thiel T (1999) Science in the Real World: Microbes in Action. St. Louis. 1-6, 95-101.

Theodoridis G, Gika HG and Wilson ID (2008) LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. *TrAC Trends in Analytical Chemistry*. 27 (3), 251–260.

Trygg J, Holmes E and Lundstedt T (2007) Chemometrics in metabonomics. *Journal of proteome research*. 6 (2), 469–79.

Tugizimana F, Steenkamp P a, Piater L a and Dubery I a (2014) Multi-platform metabolomic analyses of ergosterol-induced dynamic changes in Nicotiana tabacum cells. *PloS one*. 9 (1), e87846.

Tulipani S, Llorach R, Urpi-Sarda M and Andres-Lacueva C (2013) Comparative analysis of sample preparation methods to handle the complexity of the blood fluid metabolome: when less is more. *Analytical chemistry*. 85 (1), 341–8.

Tungsisripat-Gerber M and Aberg J (2004) Gender-Specific Issues in the Metabolic Complications of HIV Disease and Its Treatment. In: *Principles of Gender-Specific Medicine*. Elsevier. 931–937.

Tungsiripat M, Kitch D, Glesby MJ, Samir K, Mellors JW, Moran L, Jones L, Alston- B, Rooney JF, Aberg JA, Clinic C and Western C (2011) A Pilot to determine the Impact on Dyslipidemia of Adding Tenofovir to Stable Background Antiretroviral Therapy: ACTG 5206. *AIDS*. 24 (11), 1781–1784.

Tyagi S, Singh U, Kalra T and Munjal K (2010) APPLICATIONS OF METABOLOMICS - A SYSTEMATIC STUDY OF THE UNIQUE CHEMICAL FINGERPRINTS : AN OVERVIEW. *International journal of Pharmaceutical Sciences Review and Research*. 3 (1), 83–86.

UNAIDS World AIDS Report. (2011). How to get to zero: Faster, Smarter, Better Available from March 2012 from http//: www.unaids.org

Valcour V, Yeh T, Bartt R, Clifford D, Gerschenson M, Evans SR, Cohen BA, Ebenezer GJ, Hauer P and Millar L (2010) NIH Public Access. *HIV Medicine*. 10 (2), 103–110.

Van den Berg, R. a, Hoefsloot, H. C. J., Westerhuis, J. a, Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, *7*, 142. doi:10.1186/1471-2164-7-142.

Van Griensven J, De Naeyer L, Mushi T, Ubarijoro S, Gashumba D, Gazille C and Zachariah R (2007) High prevalence of lipoatrophy among patients on stavudine-containing first-line antiretroviral therapy regimens in Rwanda. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 101 (8), 793–8.

Van Leth F, Phanuphak P, Stroes E, Gazzard B, Cahn P, Raffi F, Wood R, Bloch M, Katlama C, Kastelein JJP, Schechter M, Murphy RL, Horban A, Hall DB, Lange JM a and Reiss P (2004) Nevirapine and efavirenz elicit different changes in lipid profiles in antiretroviral-therapy-naive patients infected with HIV-1. *PLoS medicine*. 1 (1), e19.

Venn RF, Merson J, Cole S and Macrae P (2005) 96-Well solid-phase extraction: a brief history of its development. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*. 817 (1), 77–80.

Villas-Boas S, Roessner J, Hansen M, Smegsgaard J and Nielson J (2007) *METABOLOMICS IN HUMANS AND OTHER MAMMALS*. New Jersey: John Wiley and Sons, Inc

Voet DJ, Voet JG, & Pratt CW (2008) Principles of Biochemistry. Third Edition. John Wiley and Sons, Inc. Asia 246-249, 600-625.

Viraben R and Aquilina C (1998) Indinavir-associated lipodystrophy. *AIDS (London, England)*. 12 (6), F37–9.

Wang P, Tang H, Fitzgibbon MP, McIntosh M, Coram M, Zhang H, Yi E and Aebersold R (2007) A statistical method for chromatographic alignment of LC-MS data. *Biostatistics (Oxford, England)*. 8 (2), 357–67.

Want EJ, Maille GO, Smith CA, Brandon TR, Uritboonthai W, Qin C, Trauger SA and Siuzdak G (2006) Clustering , and Protein Extraction for Serum Profiling with Mass Spectrometry. *Analytical chemistry*. 78 (3), 743–752.

Want EJ, Cravatt BF and Siuzdak G (2005) The expanding role of mass spectrometry in metabolite profiling and characterization. *Chembiochem : a European journal of chemical biology*. 6 (11), 1941–51.

Waters (2014) HPLC-High Performance Liquid Chromatography: Identifying and Quantifying Compounds. Available at: http://www.waters.com/waters/en_US/Identifying-and-Quantitating-Compounds/nav.htm?cid=10049064

Watstein SB, and Jovanonic J (2003). Statistical handbook on infectious diseases. Human Immunodeficiency virus (HIV) and Acquired Immunodeficiency virus (AIDS). Greenwood Publishing group, Inc. United States of America. 47-54.

Westerhuis J a, van Velzen EJJ, Hoefsloot HCJ and Smilde AK (2010) Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics*. 6 (1), 119–128. doi:10.1007/s11306-009-0185-z

Wharfe ES, Jarvis RM, Winder CL, Whiteley AS and Goodacre R (2010) Fourier transform infrared spectroscopy as a metabolite fingerprinting tool for monitoring the phenotypic changes in complex bacterial communities capable of degrading phenol. *Environmental microbiology*. 12 (12), 3253–63.

White AJ (2001) Mitochondrial toxicity and HIV therapy Review. *Sexually Transmitted Infections Journal*. 77 (3), 158–173.

Wiederin J, Rozek W, Duan F and Ciborowski P (2009) Biomarkers of HIV-1 associated dementia: proteomic investigation of sera. *Proteome science*. 7, 8.

Wiklund S (2008) Multivariate Data Analysis for Omics. OMICS, AN MKS Company, 1–228.

Wilde MI and Langtry HD(1993) Zidovudine. An Update of its pharmacodynamic and pharmacokinetic properties, and therapeutic efficacy. *Drugs.* 43 (3) 515-578.

Williams A, Kgoadi K, Steffens F, Steenkamp P and Meyer D (2014) UPLC-MS Metabonomics Reveals Perturbed Metabolites in HIV-Infected Sera. *Current Metabolomics*. 2 (1), 37–52.

Williams A, Koekemoer G, Lindeque Z, Reinecke C and Meyer D (2011) Qualitative serum organic acid profiles of HIV-infected individuals not on antiretroviral treatment. *Metabolomics*. 8 (5), 804–818.

Wilson ID, Plumb R, Granger J, Major H, Williams R and Lenz EM (2005) HPLC-MS-based methods for the study of metabonomics. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*. 817 (1), 67–76.

Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R and Scalbert A (2013) HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic acids research*. 41 (Database issue), D801–7.

Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz J a, Lim E, Sobsey C A, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, Xiong Y, Clive D, Greiner R, Nazyrova A, Shaykhutdinov R, Li L, Vogel HJ and Forsythe I (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic acids research*. 37 (Database issue), D603–10.

Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly M-A, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ and Querengesser L (2007) HMDB: the Human Metabolome Database. *Nucleic acids research*. 35 (Database issue), D521–6.

Wisniewski, T., Hilton, C. W., Morse, E. V. & Svec, F. (1993). The relationship of serum DHEA-S and cortisol levels to measures of immune function in human immunodeficiency virus related illness. *AMJ Med S*ci, 305, 79-83.

Wold S (1995) Chemometrics; what do we mean with it and what do we do from it, Chemometrics and Inelligent Laboratory Systems. 30. 109-115.

Worley B and Powers R (2013) Multivariate Analysis in Metabolomics. 92–107.

Wu G, Fang Y, Yang S, Lupton JR and Turner ND (2004) Glutathione Metabolism and Its Implications for Health 1. *The American journal of clinical nutrition*. (December 2003), 489–492.

Wu H, Southam AD, Hines A and Viant MR (2008) High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Analytical biochemistry*. 372 (2), 204–12.

Wu Z, Huang Z, Lehmann R, Zhao C and Xu G (2009) The Application of Chromatography-Mass Spectrometry: Methods to Metabonomics. *Chromatographia*. 69 (S1), 23–32.

Yamanaka H, Gatanaga H, Kosalaraksa P, Matsuoka-Aizawa S, Takahashi T, Kimura S and Oka S (2007) Novel mutation of human DNA polymerase gamma associated with mitochondrial toxicity induced by anti-HIV treatment. *The Journal of infectious diseases*. 195 (10), 1419–25.

Zavialov A V, Gracia E, Glaichenhaus N, Franco R, Zavialov A V and Lauvau G (2010) Human adenosine deaminase 2 induces differentiation of monocytes into macrophages and stimulates proliferation of T helper cells and macrophages. *Journal of leukocyte biology*. 88 (2), 279–90.

Zelena E, Dunn WB, Broadhurst D, Francis-McIntyre S, Carroll KM, Begley P, O'Hagan S, Knowles JD, Halsall A, Wilson ID and Kell DB (2009) Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Analytical chemistry*. 81 (4), 1357–64.

Zhang Z-W, Zhang X-F, Feng J, Yang Y-H, Wang C-C, Feng J-C and Liu S (2013) Formal synthesis of cephalotaxine. *The Journal of organic chemistry*. 78 (2), 786–90.

Zwir-Ferenc and Biziuk M (2006) Solid Phase Extraction Technique – Trends , Opportunities and Applications. *Polish journal of Enviromental Studies*. 15 (5), 677–690.

CHAPTER 9: APPENDIX

This section consists of the extra work that was conducted to finalize the data presented in the dissertation. It provides supporting information and materials (in the form of tables/ graphs) of the UPLC-MS second batch data and FTIR data.

Table A1: Patient demographics of study subjects used for UPLC-MS (batch two) and FTIR analysis together with their administered HAART regimens.

Sample Status	Sample ID	Age	Gender	Race	HAART Regimens
HIV Negative	N1	22	F	W	N/A
	N2	22	F	B	N/A
	N3	24	M	W	N/A
	N4	22	F	W	N/A
	N5	23	M	B	N/A
	N6	22	M	B	N/A
	N7	21	F	B	N/A
	N8	21	F	B	N/A
	N9	27	F	C	N/A
	N10	24	F	B	N/A
	N11	22	F	I	N/A
	N12	26	F	B	N/A
	N13	20	M	B	N/A
	N14	18	F	B	N/A
	N15	20	F	B	N/A
	N16	26	F	W	N/A
HIV Positive	P1	27	F	N/A	None
	P2	26	F	N/A	None
	P3	23	F	B	None
	P4	41	F	B	None
	P5	35	M	N/A	None
	P6	46	F	N/A	None

	P7	36	F	B	None
	P8	30	F	N/A	None
	P9	26	M	C	None
	P10	28	F	B	None
	P11	36	M	B	None
	P12	50	F	B	None
	P13	53	F	N/D	None
	P14	25	F	B	None
	P15	27	M	N/A	None
	P16	52	M	N/A	None
HIV Positive HAART	PH1	36	F	B	AZT 3TC EFV
	PH2	33	F	B	d4T 3TC NVP
	PH3	38	F	B	3TC d4T EFV
	PH4	N/D	F	B	3TC d4T EFV
	PH5	51	F	B	AZT 3TC NVP
	PH6	38	M	B	d4T 3TC EFV
	PH7	39	F	B	AZT 3TC EFV
	PH8	33	F	B	ARVs
	PH9	43	F	B	AZT 3TC EFV
	PH10	27	F	B	ABC 3TC NVP
	PH11	34	F	B	d4T 3TC NVP
	PH12	40	F	B	3TC d4T EFV
	PH13	34	F	B	3TC d4T NVP
	PH14	53	F	B	d4T 3TC EFV
	PH15	35	F	B	d4T 3TC EFV
	PH16	50	M	B	d4T 3TC EFV
	PH16	32	F	B	3TC d4T NVP
	PH17	44	F	B	AZT 3TC EFV
	PH18	36	F	B	AZT 3TC EFV
	PH19	46	M	B	d4T 3TC EFV

*Clear rows represent samples analyzed only using UPLC-MS, light grey shaded rows represent samples that were analyzed for both UPLC-MS and FTIR and dark grey rows represent samples analyzed using only FTIR.

Table A2: Patients demographics statistics for gender and age UPLC-MS (batch two) and FTIR samples

	HIV negative		HIV positive		HIV positive on treatment	
	UPLC-MS n=15	FTIR n=13	UPLC-MS n=16	FTIR n=12	UPLC-MS n=15	FTIR n=14
% Male	20.00	23.08	25.00	23.08	20.00	14.29
% Female	80.00	76.92	75.00	76.92	80.00	78.57
Mean Age ± SD (years)	22.67 ± 2.41	22.23 ± 2.49	34.00 ± 9.31	35.31 ± 9.60	39.36 ± 6.88	38.85 ± 7.12

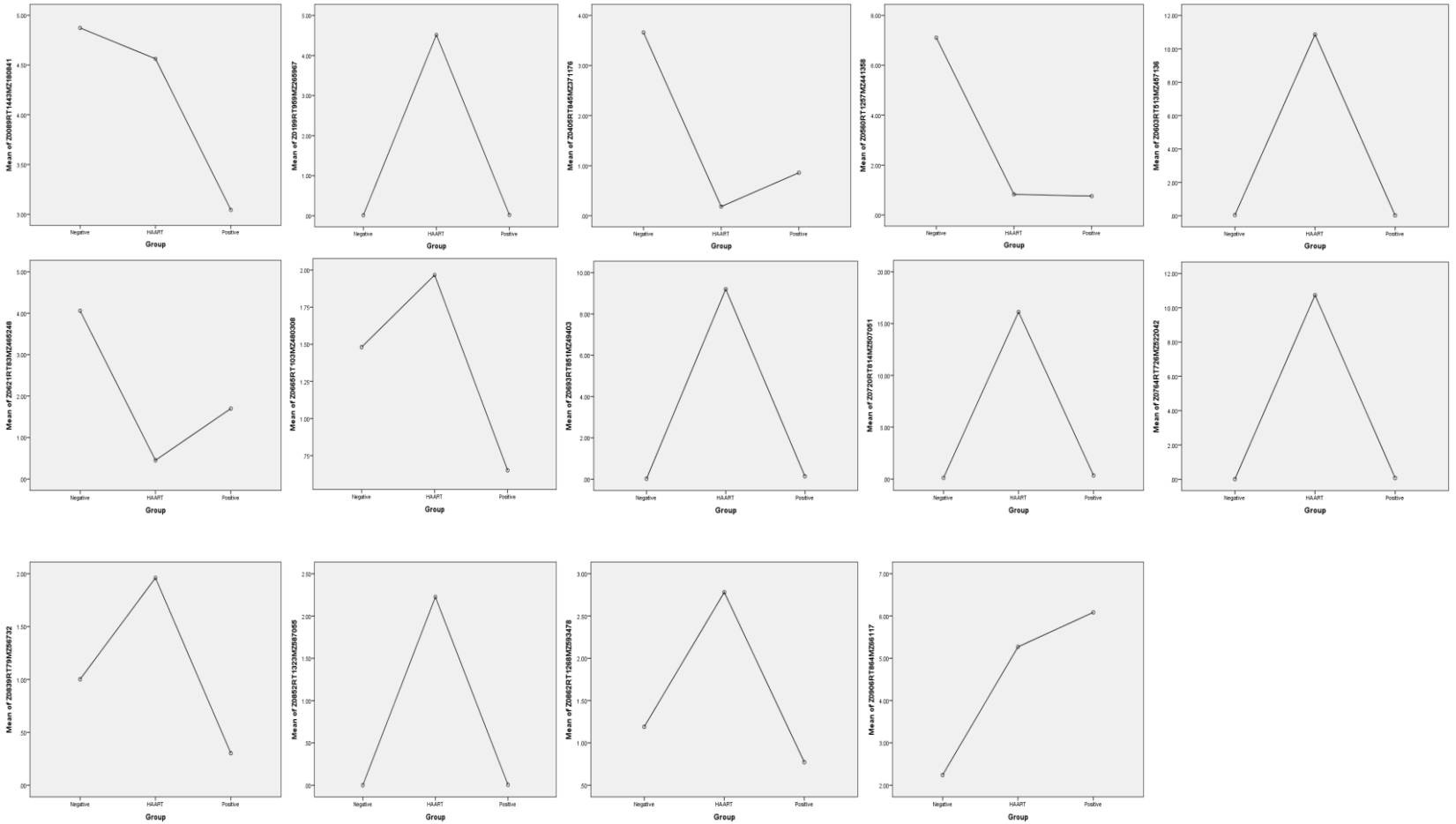


Figure A1: (i) Mean plots of the standardized conical discriminant function coefficients/metabolites that resulted from HIV serum samples analyzed in ESI- mode of UPLC-TOF-MS. These variables were responsible for the separation observed in the LDA scatter plots among the three respective HIV experimental groups (negative, positive and HAART). Shifts in location of the variables among the three groups clearly visible as variable levels differ among the groups. About 57, 10 % of the variables show a common trend, a combination of HIV and HAART appears to cause major shifts in location of the metabolites when compared to the HIV negative group.

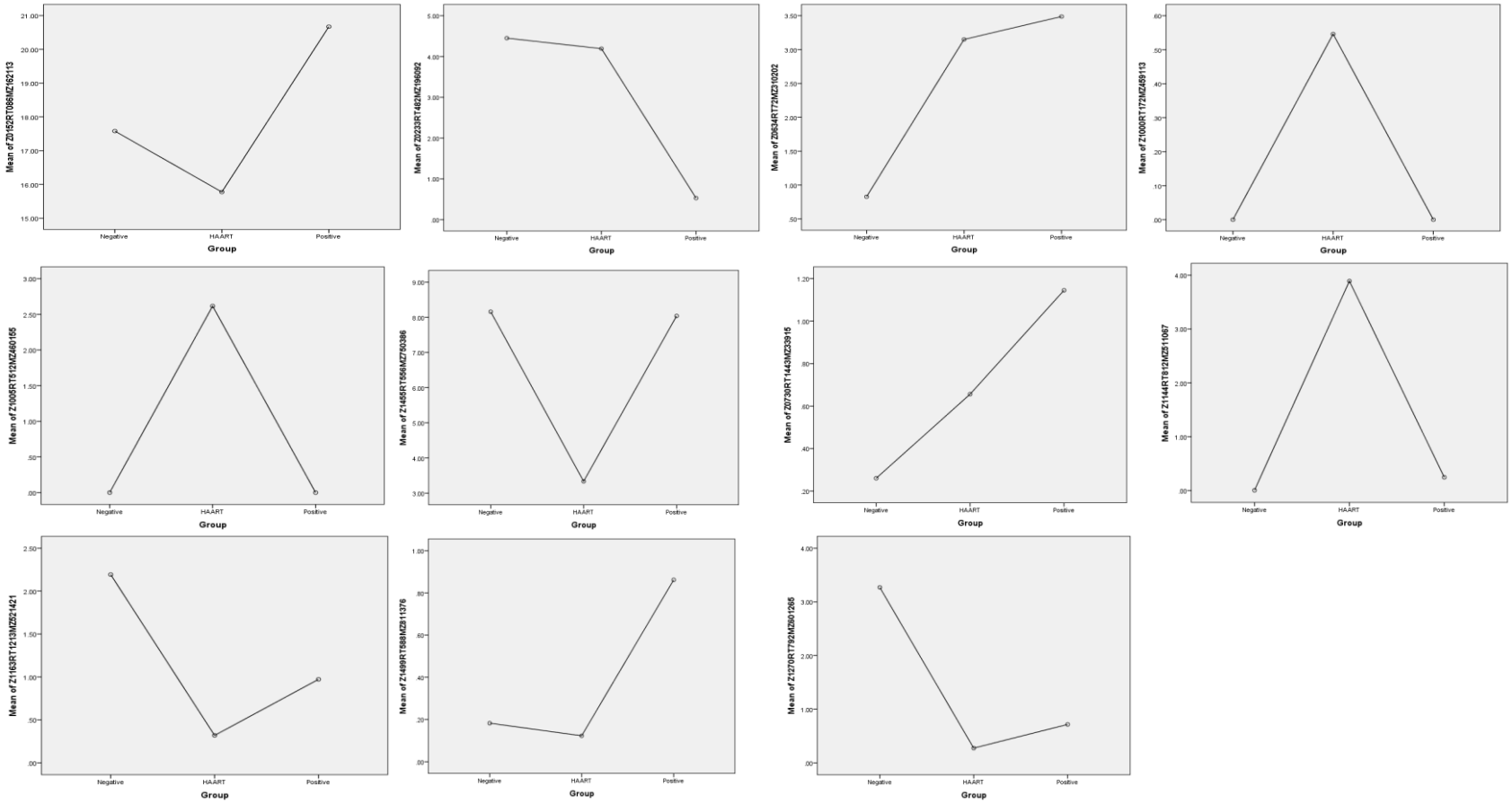


Figure A1: (ii) Mean plots of the standardized conical discriminant function coefficients/metabolites that resulted from HIV serum samples analyzed in ESI+ mode of UPLC-TOF-MS. These variables were responsible for the separation observed in the LDA scatter plots among the three respective HIV experimental groups (negative, positive and HAART). Shifts in location of the variables among the three groups clearly visible as variable levels differ among the groups. About 72.73 % of the variables show a common trend, a combination of HIV and HAART appears to cause major shifts in location in comparison to the negative group.

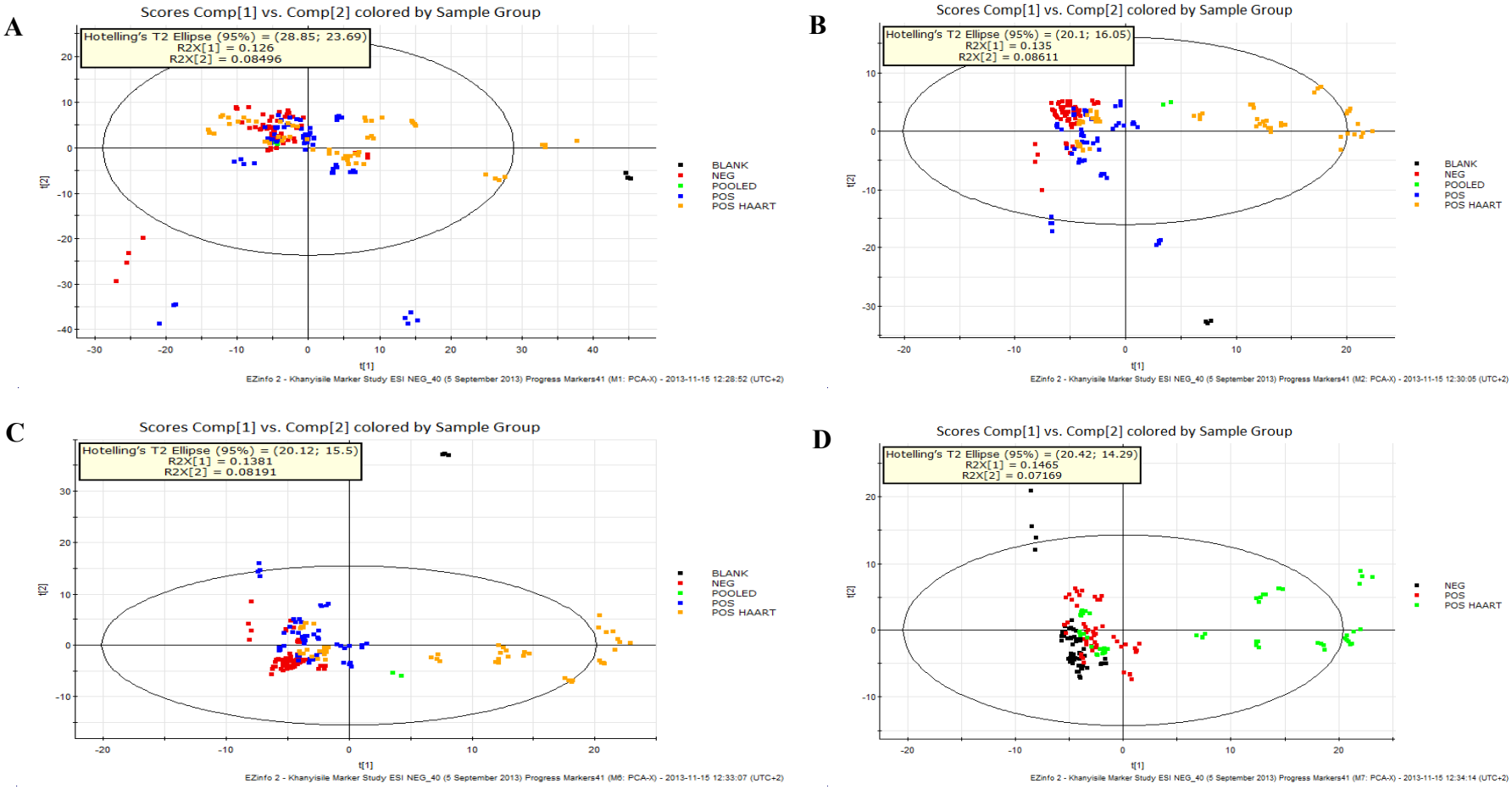


Figure A2: PCA scores plots of HIV serum samples analyzed in ESI negative mode of UPLC-MS. (A) PCA scores plot of all groups prior to normalization and (B) PCA scores plot of all groups after normalization that also show outliers outside the Hotellings T2 ellipse (95% confidence interval). Normalization improved the separation of samples and quality control (pooled) samples clustered together towards the origin. (C) PCA scores plot after normalization and removal of some outliers. (D) PCA scores of the three experimental groups (NEG, POS and POS HAART). In each scores plot, HIV serum sample groups separated from each other while some samples overlapped among different groups. POS HAART samples clustered in isolation, showing most metabolic alteration caused by treatment.

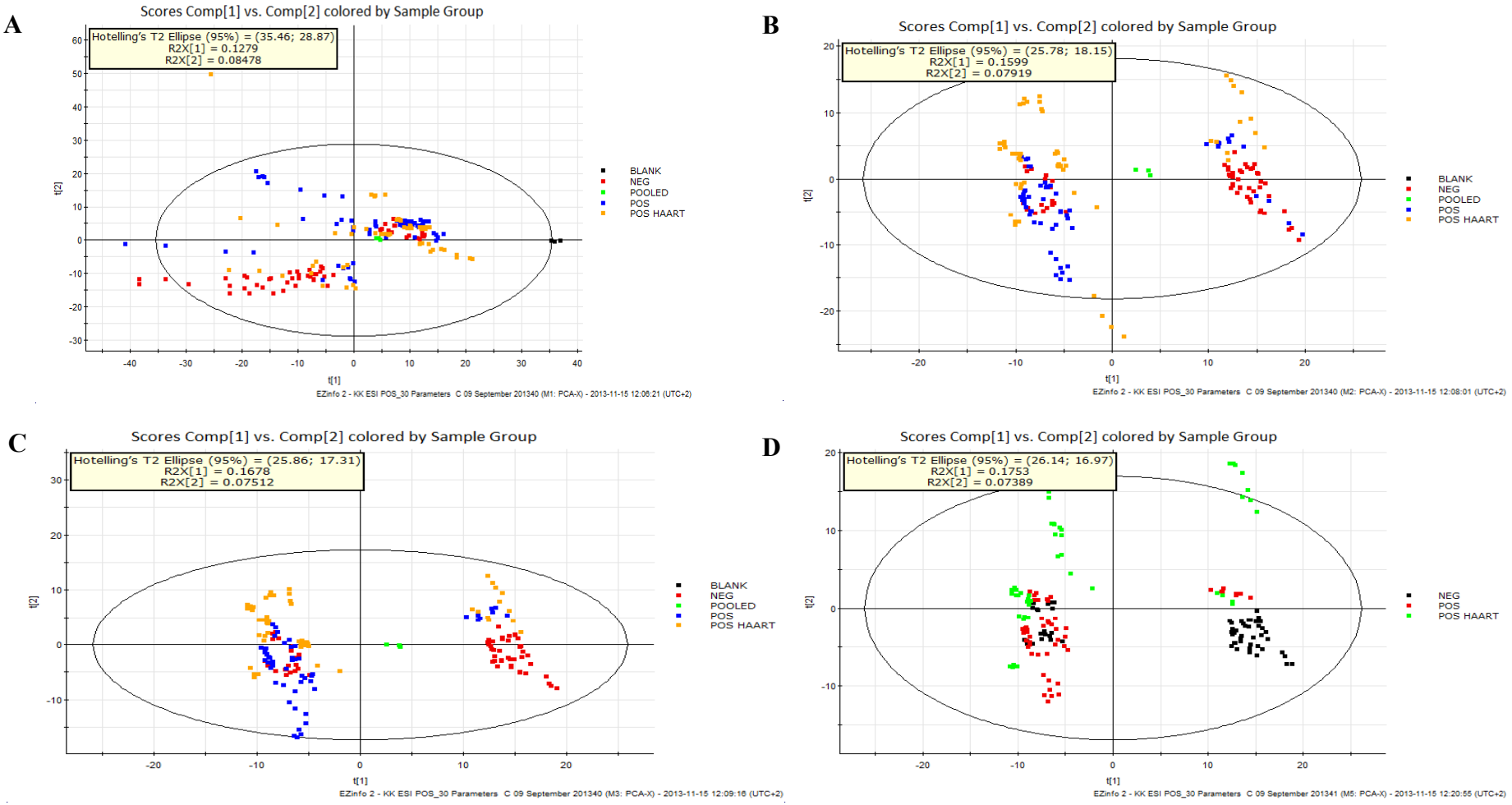


Figure A3: PCA scores plots for HIV serum samples analyzed in ESI positive mode of UPLC-TOF-MS. (A) PCA scores plot of all groups prior to normalization and (B) PCA scores plot of all groups after normalization and also showing outliers outside the Hotelling's T2 ellipse (95% confidence interval). Normalization improved the separation of samples and quality control (pooled) samples clustered closer together and closer to the origin. (C) PCA scores plot after normalization and removal of some outliers. (D) PCA scores of the three experimental groups (NEG, POS and POS HAART). Sample groups separated from each other while some samples overlapped among different groups. Most POS HAART samples clustering further away from the NEG and POS samples, showing metabolic alteration cause by treatment.

The following OPLS-DA scores plots show all the three possible combinations that the three experimental groups can pair with among each other. All selected potential biomarkers were significantly different ($p<0.05$) between the two compared groups are highlighted in red squares on the extreme ends of the S-plot. Figure A4 (i) shows a clear separation between the serum samples of the NEG and POS group, within group differences also evident. Heterogeneity of the samples visible in the two component scores plot of five-component model that shows 13.55% of the explained variation in the data. The POS group containing an outlier sample set outside the Hotelling's T2 ellipse. In figure A4 (ii) A clear separation of the NEG and POS HAART group achieved and within group separation of the POS HAART group evident. Tight clustering of the samples that appear as linear lines observed in this 2D OPLS-DA scores plot. The first two principal components explaining 22.02% variation of the data in this eleven principal component model. Same tight clustering also observed in figure A4 (iii) which shows a clear separation between the serum samples of the POS and POS HAART group and within group separation present in both groups. Only 20.11% of the explained variation is shown by the first two principal components of a twelve principal component model.

Shown in figure A5 are two of the three 2D OPLS-DA scores plots of serum samples analyzed in ESI positive mode of UPLC-MS. (i) shows 22.43% percent of the variation as explained by the first two principal components of an eight component model. Clear separation of the NEG and POS group observed showing within and between group differences. One observation set an outlier as clustered outside the 95% confidence interval. (ii) Scores plot showing 20.50 % of the variation as explained by the first two principal components of a nineteen component model. The POS and POS HAART group separable and observable sub-groups within a both groups. Tightly clustered observations appeared as straight lines because of outliers. These straight lines making the samples appear homogenous when they are not.

Figure A4 and figure A5 clustering appearing as though the serum samples of the compared groups are homogeneous because of the first two principal components chosen for viewing the data. As mentioned earlier, the first two principal components were chosen for viewing the data because they show the highest explained variation in the data.

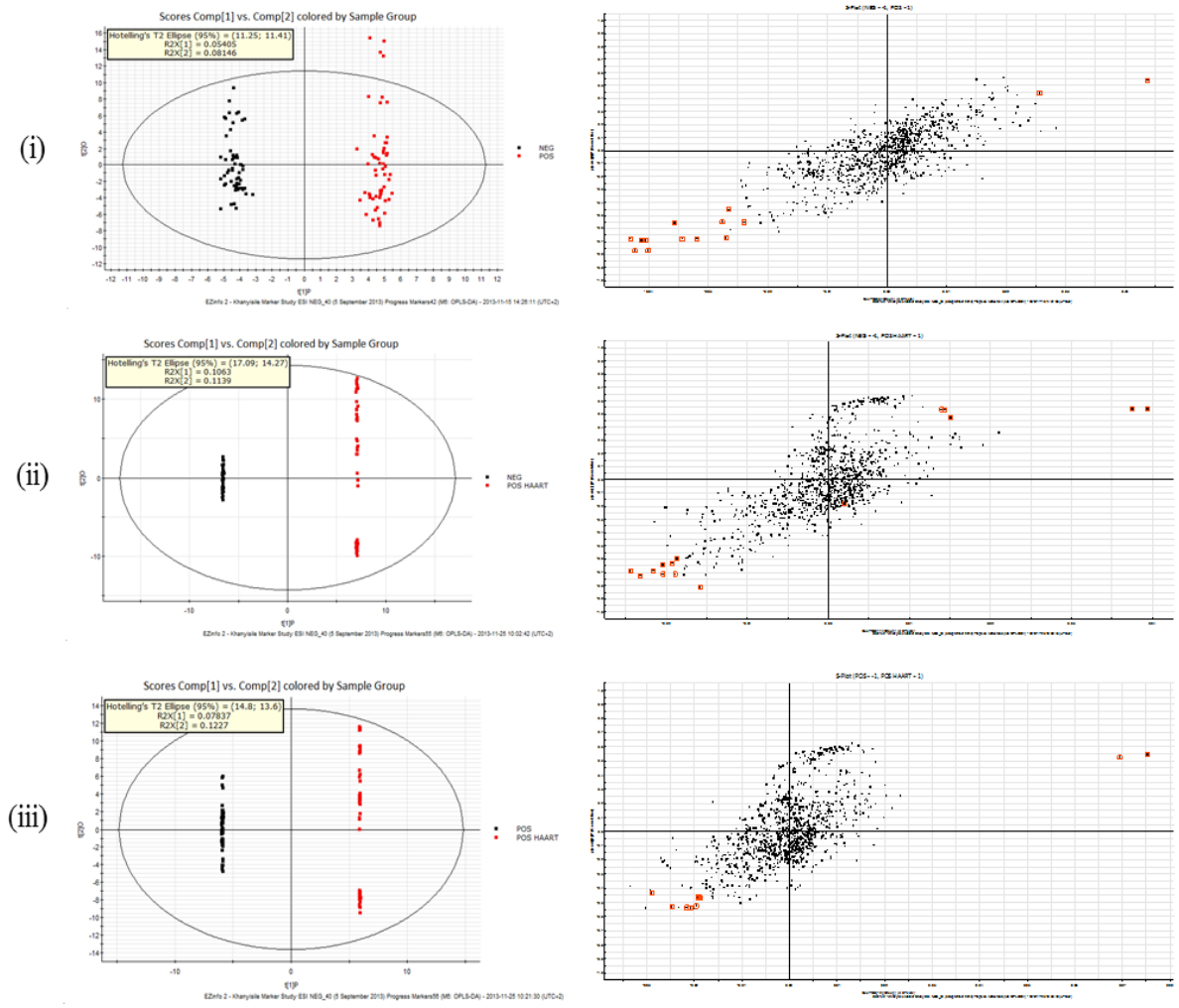


Figure A4: 2D OPLS-DA scores plots and their accompanying S-plots showing group differences of HIV serum samples that were analyzed in ESI negative mode of UPLC-MS. Selected metabolites (potential biomarkers) that were significantly different ($p<0.05$) shown by red squares on the S-plot extremes. (i) Scores plot showing a clear separation observed between the NEG and POS group and subtle within group variation of both groups evident. Heterogeneity of samples visible. One observation set (outlier) outside the Hotelling's T2 ellipse (95% confidence interval) shown in the POS group. Explained data variation (13.55%) shown by the first two principal components of a five principal component model. (ii) Scores plot showing 22.02% of the variation that is explained by the first two principal components of an eleven component model. Tight clustering and separation between the NEG and POS HAART group evident NEG group not showing distinct within group variation, whereas the POS HAART group having different sub-group separations within the same sample group. (iii) Scores plot showing 20.11% of the variation that is explained by the first two principal components of a twelve component model. Between and within group differences evident in the separation of the POS and/or POS HAART groups that show a tight linear clustering of samples.

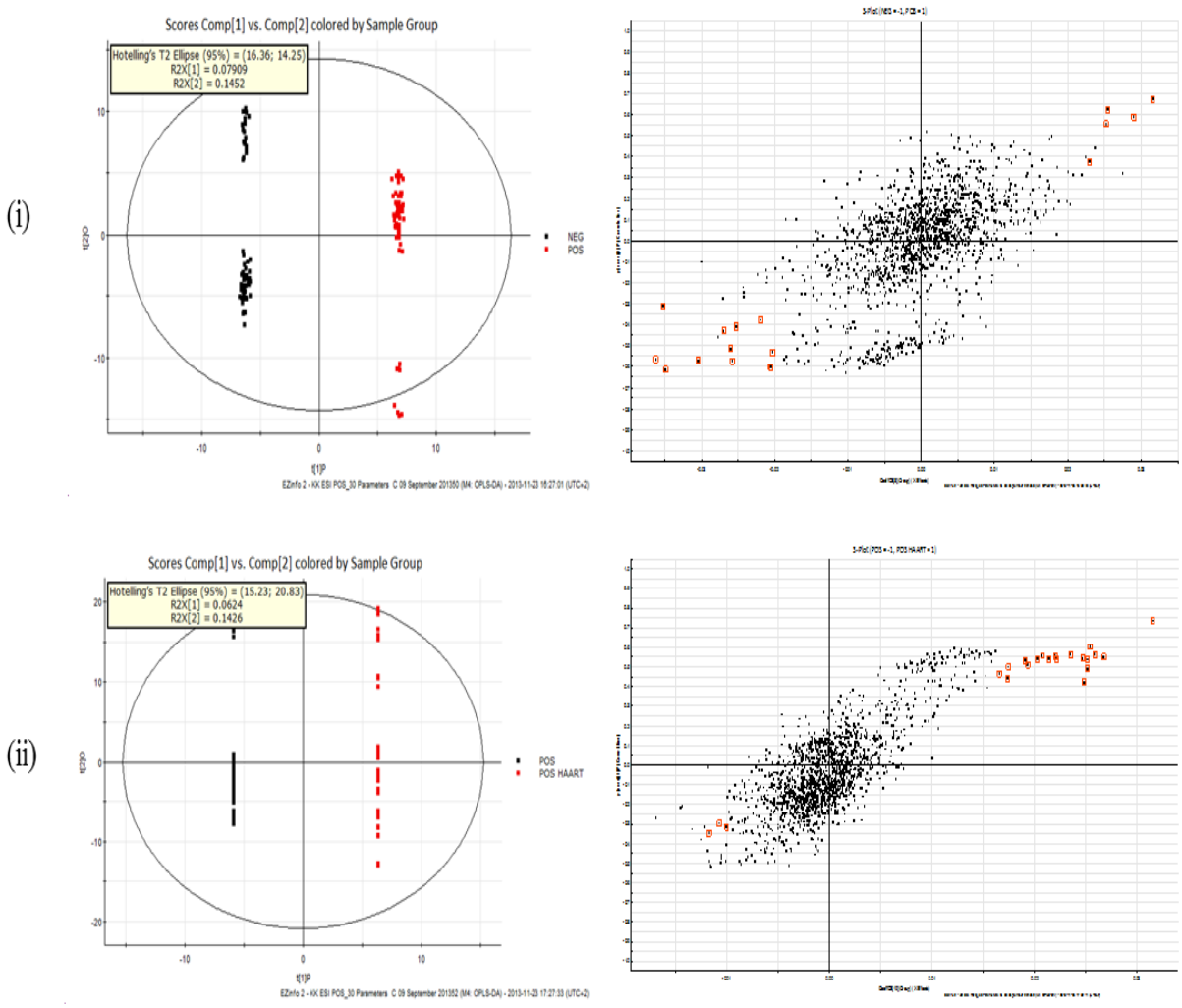
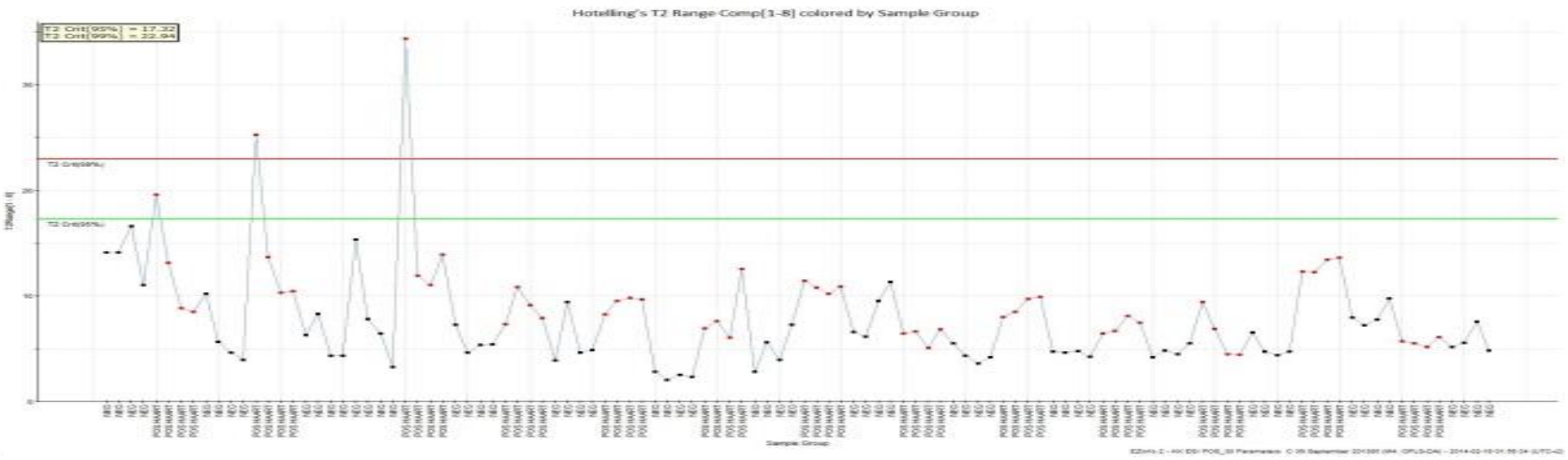
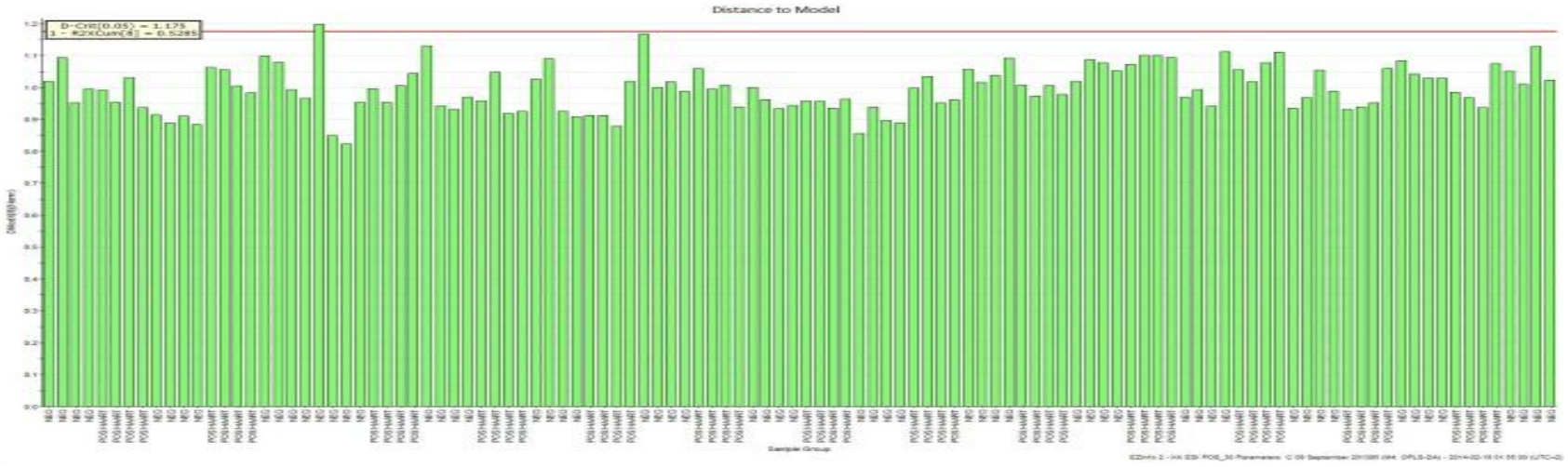


Figure A5: 2D OPLS-DA scores plots and their accompanying S-plots showing group differences of HIV serum samples that were analyzed in ESI positive mode of UPLC-MS. Selected metabolites (potential biomarkers) that were significantly different ($p<0.05$) shown by red squares on the S-plot extremes. (i) Scores plot showing 22.43% of the variation that is explained by the first two principal components of an eight principal component model. NEG and POS group showing between and within group differences with one observation set (outlier) outside the Hotelling's T2 ellipse (95% confidence interval). (ii) Scores plot showing 20.50% of the variation that is explained by the first two principal components of a nineteen principal component model. POS and POS HAART group showing between and within group differences, observations appearing as straight lines due to outliers, even though it is a heterogeneous sample dataset that is not expected to show linear clustering of samples.

Figure A6 [(i) and (ii)] clearly shows both types of these outliers were apparent in the 2D and 3D OPLS-DA scores plot of the NEG and POS HAART group samples while figure A6 [(iii) and (iv)] shows these plots after the removal of outliers. Most of these outliers could not be eliminated based on their biological significance because not enough demographic information was available on these study samples, as a result the removal of these outliers was guided by the statistical models. These tight clustering corresponded well with the observed high model validation values for goodness of fit (R^2X) and model predictability (Q^2) (see table 4.4), these table shows model validation parameters before and after the removal of outliers.

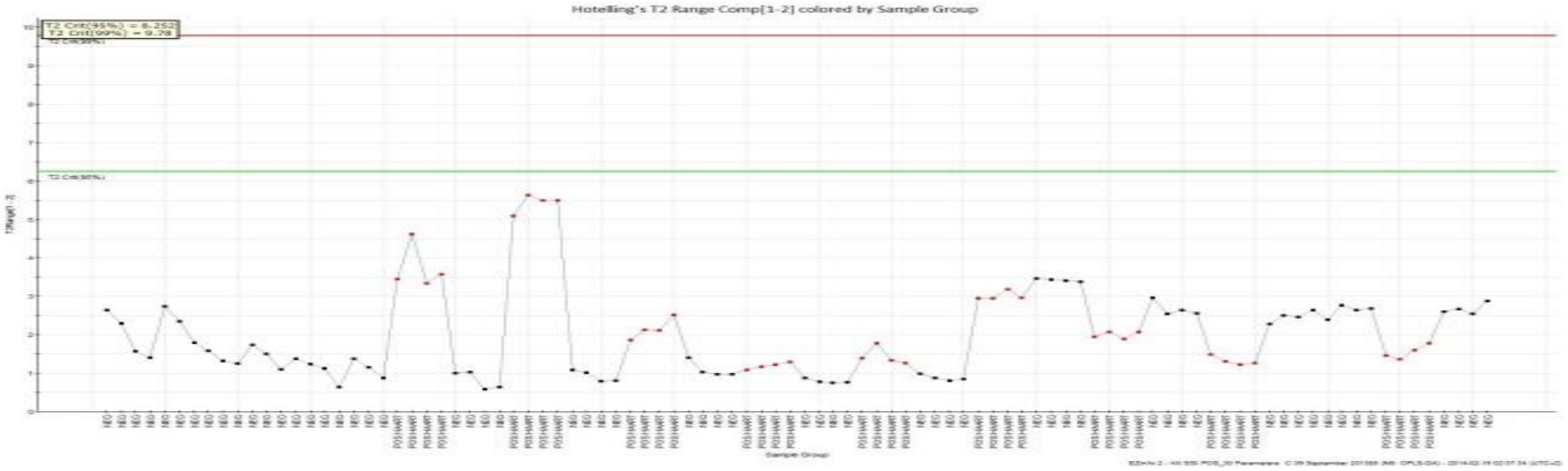


(A)

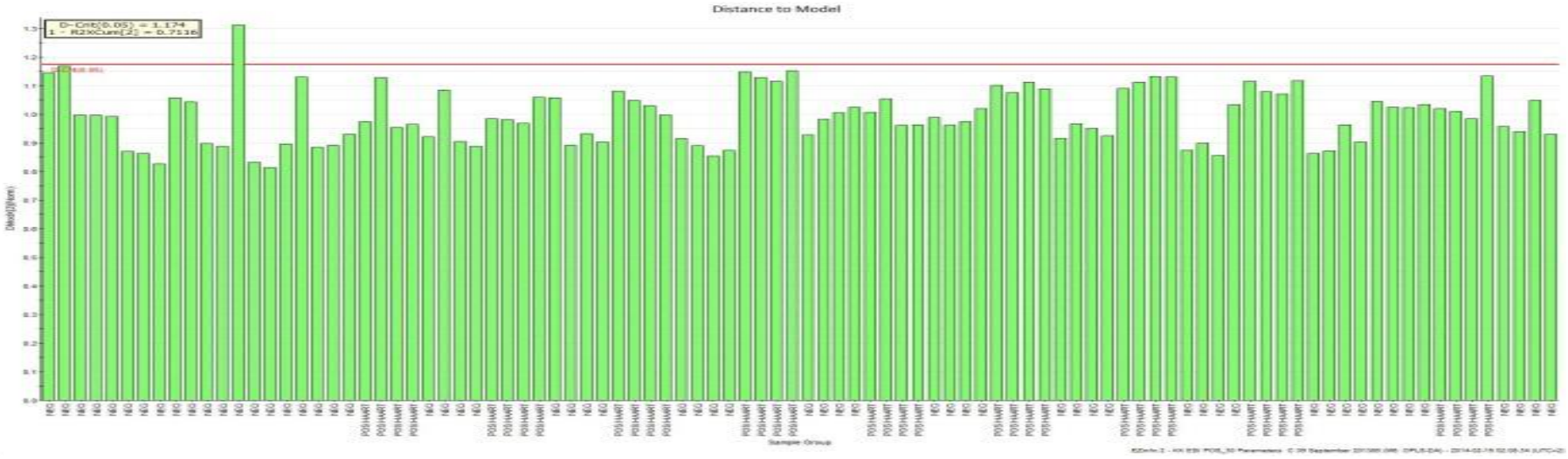


(B)

Figure A6: (A) Hotelling's T2 Range showing outliers grouping outside the 95% confidence interval and (B) Distance to model score plot showing outliers outside the 95% confidence interval and the D-Crit value of HIV sera samples analyzed in the ESI+ mode of UPLC-MS. Group differences between the NEG group and POS HAART group.



(C)



(D)

Figure A6: (C) Hotelling's T2 Range and (D) Distance to model score plot after removal of outliers. Clearly showing that no observations lay outside or above the 95% confidence interval, however the moderate outliers are still present that appear above the D-Crit value.

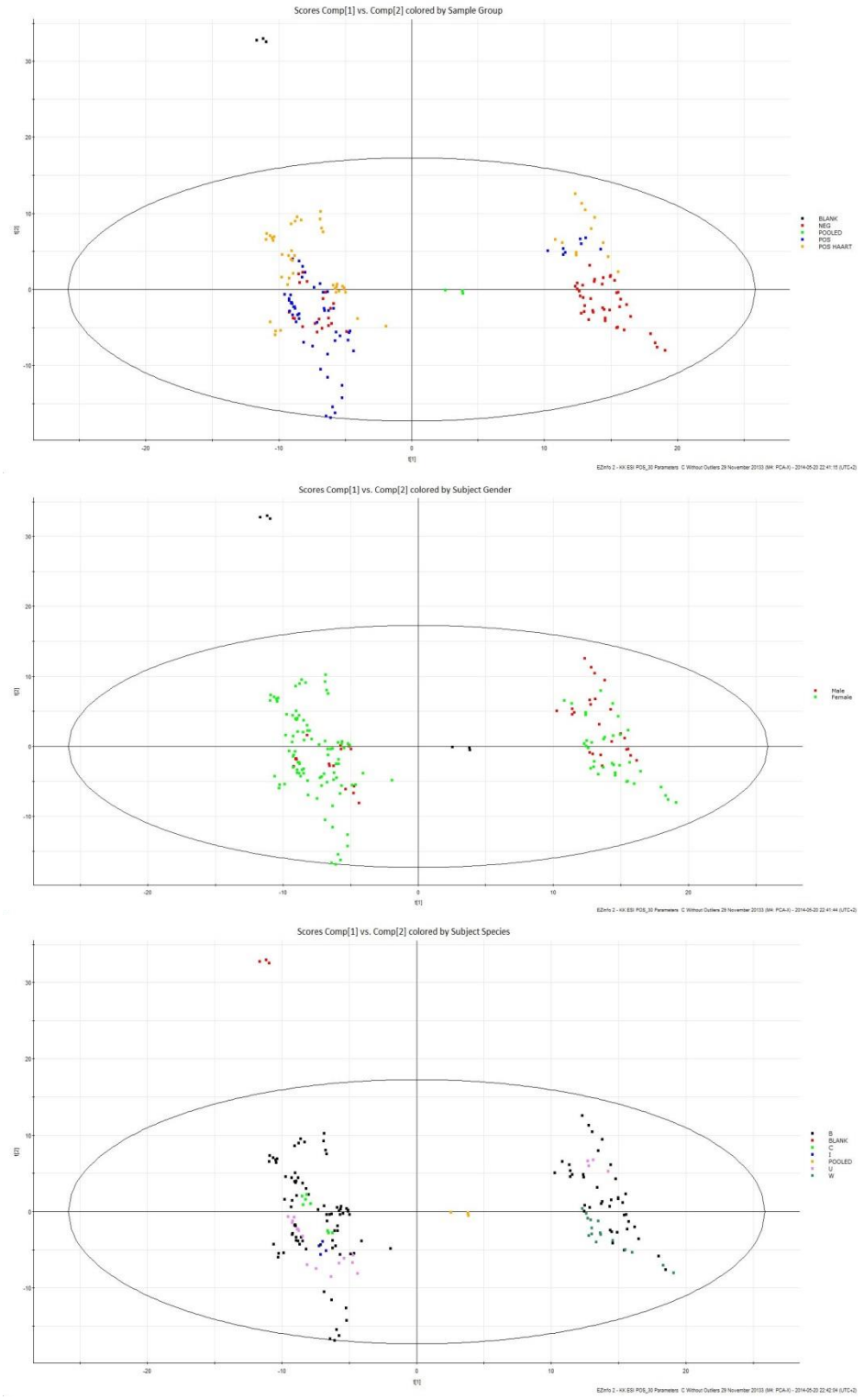


Figure A7: PCA scores plots of HIV serum samples analyzed in ESI+ mode of UPLC-TOF-MS showing the samples failing to group according to gender and race, proving that these factors are not confounding factors in this study.

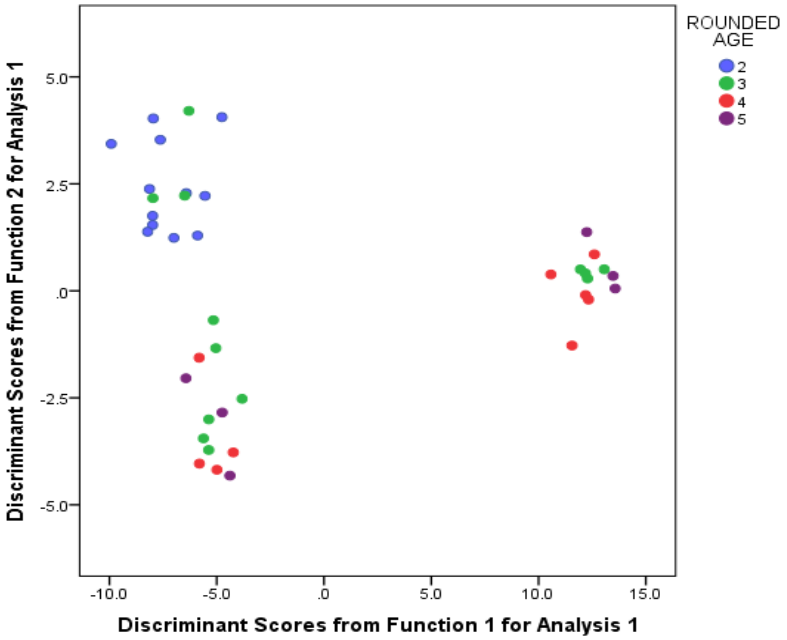


Figure A8: LDA scatter plot of HIV serum samples analyzed in ESI+ mode of UPLC-TOF-MS showing the samples failing to group according to age, proving that this factor is not a confounding factor in this study. Age was rounded off to the nearest multiple of 10.

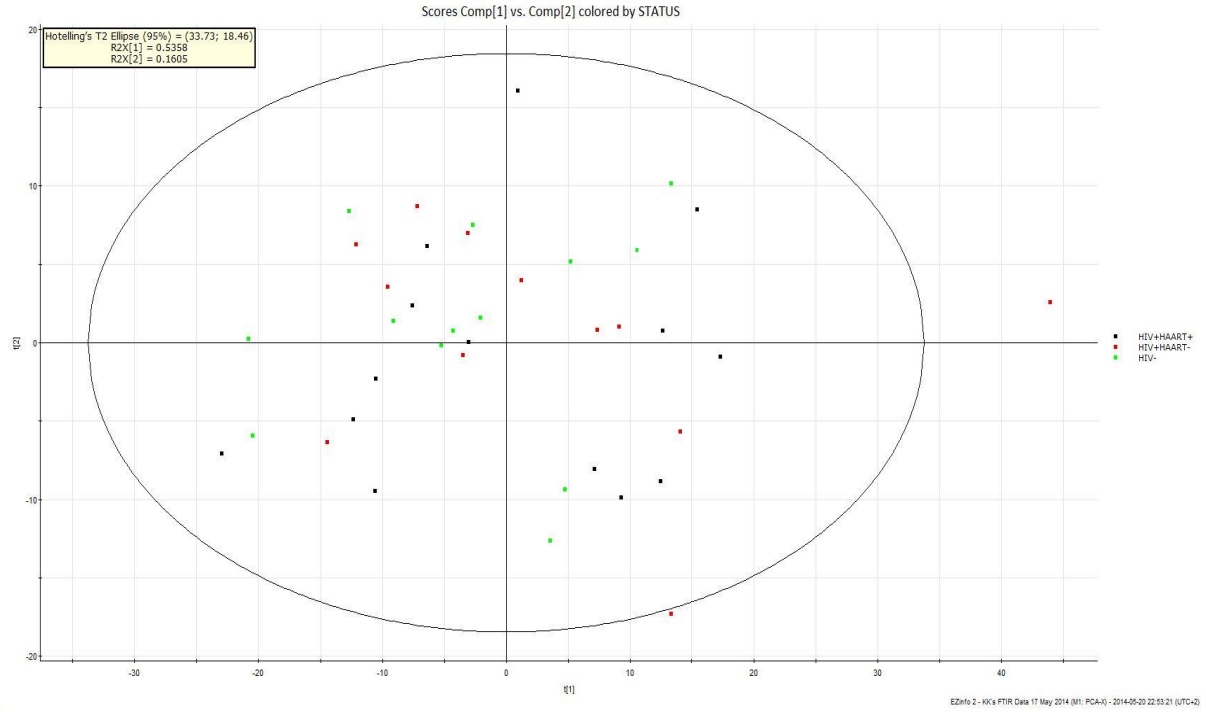


Figure A9: PCA scores plot of HIV serum samples analyzed using Mid-ART-FTIR. Samples from three experimental groups (HIV-, HIV+HAART- and HIV+HAART+) not clearly distinguished, clustering everywhere forming overlaps. Poor separation of the three groups observed.