From the Department of Laboratory Medicine
Karolinska Institutet, Stockholm, Sweden

AN INTEGRATIVE SYSTEMS BIOLOGY STUDY TO UNDERSTAND IMMUNE AGING IN PEOPLE LIVING WITH HIV

Flora Mikaeloff

Karolinska Institutet

Stockholm 2023

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023

ISBN 978-91-8017-002-4
Cover illustration: The illustration was done by Lae-Young Kim. Ngoc Boi Vo helped in the discussion for the choice of the design.

An integrative systems biology study to understand immune aging in people living with HIV
Thesis for Doctoral Degree (Ph.D.)

By

Flora Mikaeloff

The thesis will be defended in public at Karolinska Institute, Campus Flemingsberg, Alfred Nobels allé 8, lecture hall 4Y, June 2nd 2023 at 10.00 am

Principal Supervisor:
Docent Ujjwal Neogi
Karolinska Institute
Department of Laboratory Medicine
Division of Clinical Microbiology

Co-supervisor(s):
Rui Benfeitas
Karolinska Institute
Department of Laboratory Medicine
Division of Clinical Microbiology

Docent Erin Gabriel
Karolinska Institute
Department of Medical Epidemiology and Biostatistics

Opponent:
Professor Thomas Sauter
University of Luxembourg
Department of Life Sciences and Medicine

Examination Board:
Professor Kristina Broliden
Karolinska Institute
Department of Medicine
Division of Infectious Diseases

Docent Juan Du
Karolinska Institute
Department of Microbiology, Tumor and Cell Biology

Docent Peter Spégel
Lund University
Department of Chemistry

“There was a time when I thought a great deal about the axolotls. I went to see them at the aquarium at the Jardin des Plantes and stayed for hours watching them, observing their immobility; their faint movements. Now I am an axolotl. ”

Julio Cortazar

Abstract

Antiretroviral therapy (ART) reduces viral replication, restores T helper cells and improves the survival of people living with HIV (PWH), transforming a life-threatening disease into a manageable chronic infection. Nevertheless, PWH under ART shows aging-related diseases such as bone abnormalities, non-HIV-associated cancers, and cardiovascular and neurocognitive diseases. The complex immune metabolic dysregulation leading to these comorbidities is called immune aging. The main question raised by my thesis was, what are the complex mechanisms responsible for immune aging in HIV? Using advanced system biology and machine learning tools, I used multi-omics-based patient stratification to identify biologic perturbations associated with immune aging in PWH.

First, we investigated PWH with Metabolic Syndrome (MetS), a relatively common aging-related disease in HIV-1. In **paper I**, we identified the dysregulation of glutamate metabolism in PWH with MetS using plasma metabolomics and measure of cell transporters by flow cytrometry. Then, we investigated the mechanisms of differing PWH on long-term successful ART from HIV-negative controls (HC). In **paper II**, we identified the dysregulation of amino acids and, more specifically, glutaminolysis (i.e., lysis of glutamine to glutamate) in PWH compared to HC using metabolomics in two independent cohorts to avoid the potential cohort biases. We identified five neurosteroids to be lower in PWH and potentially create neurological impairments in PWH. The glutaminolysis inhibition in chronically infected HIV-1 promonocytic (U1) cells induced apoptosis and latency reversal which could clear HIV reservoirs.

The first two papers universally clarified our knowledge about dysregulated metabolic traits following a prolonged ART in PWH. However, we observed heterogeneity among the clinically defined PWH. Therefore, we focused more on the multi-omics data-driven approaches to stratify the at-risk group who were either dysregulated metabolically at-risk PWH (**paper III**) or immunometabolic at-risk group (**paper IV**) and clarified the biological aging process by measuring transcriptomics age (**paper V**).

In **paper III**, we found three groups of PWH based on multi-omics integration of lipidomics, metabolomics, and microbiome. The severe at-risk metabolic complications showed increased weight-related comorbidities and di- and triglycerides compared to the other clusters. At-risk and HC-like groups displayed similar metabolic profiles but were different from HC. An increase in *Prevotella* was linked to the overrepresentation of men having sex with men (MSM) in the at-risk group. The microbiome-associated metabolites (MAM) appeared dysregulated in all HIV groups compared to controls. We improved this clustering by adding transcriptomics and proteomics data for a refined immunometabolic at-risk-related clustering in PWH. In **paper IV**, immune-driven HC-like and at-risk groups were clustered based on metabolomics, transcriptomics, and proteomics. Several biomarkers from central carbon metabolism (CCM) and senescence-associated proteins

were linked to the at-risk phenotype based on random forest, structural causal modeling, and co-expression networks. Senescent protein changes were associated with a deficiency in macrophage function based on single-cell data, cell profiling, flow cytometry, and proteomics from macrophage data and *in vitro* validation. We also developed personalized and group-level genome-scale metabolic models (GSMM) and confirmed the implication of metabolites from CCM and polyamides in at-risk phenotypes. Finally, we investigated the accelerated aging process (AAP) in PWH. In **paper V**, we calculated the biological age of PWH using transcriptomics data and grouped patients into aging groups; The decelerated aging process (DAP) group was linked with higher age, European origin, and a higher proportion of tenofovir disoproxil fumarate /alafenamide (TDF/TAF). AAP had a downregulation of metabolic pathways and an upregulation of inflammatory pathways.

In conclusion, my thesis identifies underlying mechanisms of immune aging using system biology tools in three independent cohorts of PWH for mechanistic studies and to improve their care and achieve healthy aging.

List of scientific papers

I. Gelpi M*, **Mikaeloff F***, Knudsen AD, Benfeitas R, Krishnan S, Svenssson Akusjärvi S, Høgh J, Murray DD, Ullum H, Neogi U†, Nielsen SD. (2021) The central role of the glutamate metabolism in long-term antiretroviral treated HIV-infected individuals with metabolic syndrome. ***Aging (Albany NY).*** 2021 Oct 11;13(19):22732-22751. (Equal contribution)

II. **Mikaeloff F**, Svensson Akusjärvi S, Ikomey GM, Krishnan S, Sperk M, Gupta S, Magdaleno GDV, Escós A, Lyonga E, Okomo MC, Tagne CT, Babu H, Lorson CL, Végvári Á, Banerjea AC, Kele J, Hanna LE, Singh K, de Magalhães JP, Benfeitas R, Neogi U†. (2022) Trans cohort metabolic reprogramming towards glutaminolysis in long-term successfully treated HIV-infection. ***Communications Biology.*** 2022 Jan 11;5(1):27.

III. **Mikaeloff F**†, Gelpi M, Benfeitas R, Knudsen AD, Vestad B, Høgh J, Hov JR, Benfield T, Murray D, Giske CG, Mardinoglu A, Trøseid M, Nielsen SD, Neogi U† (2023) Network-based multi-omics integration reveals metabolic at-risk profile within treated HIV-infection. ***Elife.*** 2023 Feb 16;12:e82785.

IV. **Mikaeloff F**, Gelpi M, Escós A, Olofsson A, Svensson Akusjärvi S, Schuster S, Nikouyan N, Knudsen AD, Vestad B, Høgh J, Hov JR, Benfield T, Murray D, Trøseid M, Gupta S, Pawar V, Benfeitas R, Vegvary A, O'Mahony L, Savai R, Björkström N, Lourda M, de Magalhães JP, Mardinoglu A, Weiß S, Karlsson A, Nielsen SD, Neogi U†. Integrative systems analysis-based risk stratification for metabolic complications in well-treated HIV-infected individuals (Manuscript)

V. **Mikaeloff F**†, Gelpi M, Escos A, Knudsen AD, Høgh J, Benfield T, de Magalhães JP, Nielsen SD, Neogi U†. Transcriptomics age acceleration in prolonged treated HIV infection (Manuscript)

List of scientific papers not included in the thesis

I. Appelberg S, Gupta S, Svensson Akusjärvi S, Ambikan AT, **Mikaeloff F**, Saccon E, Végvári Á, Benfeitas R, Sperk M, Ståhlberg M, Krishnan S, Singh K, Penninger JM, Mirazimi A, Neogi U. Dysregulation in Akt/mTOR/HIF-1 signaling identified by proteo-transcriptomics of SARS-CoV-2 infected cells, ***Emerg Microbes Infect.*** 2020 Dec;9(1):1748-1760.

II. Sperk M, van Domselaar R, Rodriguez JE, **Mikaeloff F**, Sá Vinhas B, Saccon E, Sönnerborg A, Singh K, Gupta S, Végvári Á, Neogi U, Utility of Proteomics in Emerging and Re-Emerging Infectious Diseases Caused by RNA Viruses, ***J Proteome Res.*** 2020 Nov 6;19(11):4259-4274.

III. Saccon E, Chen X, **Mikaeloff F**, Rodriguez JE, Szekely L, Vinhas BS, Krishnan S, Byrareddy SN, Frisan T, Végvári Á, Mirazimi A, Neogi U, Gupta S, Cell-type-resolved quantitative proteomics map of interferon response against SARS-CoV-2, ***iScience***. 2021 May 21;24(5):102420.

IV. Chen X, Saccon E, Appelberg KS, **Mikaeloff F**, Rodriguez JE, Vinhas BS, Frisan T, Végvári Á, Mirazimi A, Neogi U, Gupta S, Type-I interferon signatures in SARS-CoV-2 infected Huh7 cells, ***Cell Death Discov.*** 2021 May 18;7(1):114.

V. Appelberg S, John L, Pardi N, Végvári Á, Bereczky S, Ahlén G, Monteil V, Abdurahman S, **Mikaeloff F**, Beattie M, Tam Y, Sällberg M, Neogi U, Weissman D, Mirazimi A, Nucleoside-Modified mRNA Vaccines Protect IFNAR(-/-) Mice against Crimean-Congo Hemorrhagic Fever Virus Infection, ***J Virol.*** 2022 Feb 9;96(3):e0156821.

VI. Saccon E, **Mikaeloff F**, Figueras Ivern P, Végvári Á, Sönnerborg A, Neogi U, van Domselaar R, Cytotoxic Lymphocytes Target HIV-1 Gag Through Granzyme M-Mediated Cleavage, ***Front Immunol.*** 2021 Apr 19;12:669347.

VII. Krishnan S, Nordqvist H, Ambikan AT, Gupta S, Sperk M, Svensson-Akusjärvi S, **Mikaeloff F**, Benfeitas R, Saccon E, Ponnan SM, Rodriguez JE, Nikouyan N, Odeh A, Ahlén G, Asghar M, Sällberg M, Vesterbacka J, Nowak P, Végvári Á, Sönnerborg A, Treutiger CJ, Neogi U, Metabolic Perturbation Associated With COVID-19 Disease Severity and SARS-CoV-2 Replication, ***Mol Cell Proteomics.*** 2021;20:100159.

VIII. Albrich WC, Ghosh TS, Ahearn-Ford S, **Mikaeloff F**, Lunjani N, Forde B, Suh N, Kleger GR, Pietsch U, Frischknecht M, Garzoni C, Forlenza R, Horgan M, Sadlier C, Negro TR, Pugin J, Wozniak H, Cerny A, Neogi U, O'Toole PW, O'Mahony L, A high-risk gut microbiota configuration associates with fatal hyperinflammatory immune and metabolic responses to SARS-CoV-2, ***Gut Microbes.*** 2022 Jan-Dec;14(1):2073131.

IX. Akusjärvi SS, Ambikan AT, Krishnan S, Gupta S, Sperk M, Végvári Á, **Mikaeloff F**, Healy K, Vesterbacka J, Nowak P, Sönnerborg A, Neogi U, Integrative proteo-transcriptomic and immunophenotyping signatures of HIV-1 elite control

phenotype: A cross-talk between glycolysis and HIF signaling, ***iScience***. 2021 Dec 10;25(1):103607.

X. Zhou X, **Mikaeloff F**, Curbo S, Zhao Q, Kuiper R, Végvári Á, Neogi U, Karlsson A, Coordinated pyruvate kinase activity is crucial for metabolic adaptation and cell survival during mitochondrial dysfunction, ***Hum Mol Genet.*** 2021 Oct 13;30(21):2012-2026.

XI. Sperk M, **Mikaeloff F**, Svensson-Akusjärvi S, Krishnan S, Ponnan SM, Ambikan AT, Nowak P, Sönnerborg A, Neogi U, Distinct lipid profile, low-level inflammation, and increased antioxidant defense signature in HIV-1 elite control status, ***iScience.*** 2021 Jan 28;24(2):102111.

XII. Saccon E, Bandera A, Sciumè M, **Mikaeloff F**, Lashari AA, Aliberti S, Sachs MC, Billi F, Blasi F, Gabriel EE, Costantino G, De Roberto P, Krishnan S, Gori A, Peyvandi F, Scudeller L, Canetta C, Lorson CL, Valenti L, Singh K, Baldini L, Fracchiolla NS, Distinct Metabolic Profile Associated with a Fatal Outcome in COVID-19 Patients during the Early Epidemic in Italy, Microbiol Spectr. 2021 Oct 31;9(2):e0054921.

XIII. Sperk M, Ambikan AT, Ray S, Singh K, **Mikaeloff F**, Diez RC, Narayanan A, Vesterbacka J, Nowak P, Sönnerborg A, Neogi U, Fecal Metabolome Signature in the HIV-1 Elite Control Phenotype: Enrichment of Dipeptides Acts as an HIV-1 Antagonist but a Prevotella Agonist, ***J Virol.*** 2021 Aug 25;95(18):e0047921.

XIV. Tiwari R, Mishra AR, **Mikaeloff F**, Gupta S, Mirazimi A, Byrareddy SN, Neogi U, Nayak D., In silico and in vitro studies reveal complement system drives coagulation cascade in SARS-CoV-2 pathogenesis, ***Comput Struct Biotechnol J.*** 2020;18:3734-3744.

XV. Svensson Akusjärvi S, Krishnan S, Ambikan AT, **Mikaeloff F**, Munusamy Ponnan S, Vesterbacka J, Lourda M, Nowak P, Sönnerborg A, Neogi U, Role of myeloid cells in system-level immunometabolic dysregulation during prolonged successful HIV-1 treatment, ***AIDS.*** 2023 Feb 9.

XVI. Humbert M, Olofsson A, Wullimann D, Niessl J, Hodcroft EB, Cai C, Gao Y, Sohlberg E, Dyrdak R, **Mikaeloff F**, Neogi U, Albert J, Malmberg KJ, Lund-Johansen F, Aleman S, Björkhem-Bergman L, Jenmalm MC, Ljunggren HG, Buggert M, Karlsson AC, Functional SARS-CoV-2 cross-reactive CD4(+) T cells established in early childhood decline with age, ***Proc Natl Acad Sci U S A.*** 2023 Mar 21;120(12):e2220320120.

Contents

List of abbreviations

16S rRNA	16 sequencing rRNA
2-DG	2-Deoxyglucose
ABC	Abacavir
AIDS	Acquired immunodeficiency syndrome
ALAT	Alanine aminotransferase
ART	Antiretroviral therapy
BBNs	Bayesian Belief Networks
BINGO	Biological Networks Gene Ontology tool
BMI	Body mass index
CCM	Central carbon metabolism
CER	Ceramides
COBRA	COnstraint-Based Reconstruction and Analysis
COCOMO	Copenhagen Comorbidity in HIVinfection cohort
CVDs	Cardiovascular diseases
DAG	Diglycerides
DGE	Differential expression analysis
DIABLO	Latent variable approaches for Omics studies
DON	Glutamine antagonist 6-diazo-5-oxo-L-norleucine
FACS	Fluorescence-activated cell sorting
FBA	Flux balance analysis
FDR	False discovery rate
GSEA	Gene Set Enrichment Analysis
GSMM	Genome-scale metabolic models
HAND	HIV-associated neurocognitive disorders
HC	Healthy controls
HIV	Human immunodeficiency virus
INSTI	Integrase strand transfer inhibitors
IPA	Ingenuity Pathway Analysis software
IRIS	Immune reconstitution inflammatory syndrome
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS/MS	Liquid Chromatography-Mass Spectrometry

LEfSe	LDA Effect Size
limma	Linear models for microarray data
LMICS	Low- and middle-income countries
MAM	Microbiome associated metabolites
MetS	Metabolic syndrome
MOFA	Multi-Omics Factor Analysis
mRNA	Messenger ribonucleic acid
MSigDB	Molecular Signatures Database
MSM	Men having sex with men
mtDNA	Mitochondrial DNA
NGS	Next-generation sequencing
NNRTI	Non-nucleoside reverse transcriptase inhibitors
NRTI	Nucleoside/nucleotide reverse transcriptase inhibitors
OXPHOS	Oxidative phosphorylation
PBMCs	Peripheral blood mononuclear cells
PCA	Principal component analysis
PI	Protease inhibitors
PLS-DA	Partial least squaresdiscriminant analysis
PWH	People living with HIV
RAVEN	Reconstruction, Analysis and Visualization of Metabolic Networks
RF	Random forest
RNA-seq	RNA-sequencing
ROS	Reactive oxygen species
SASP	Senescence-associated secretory phenotype
SAT	Subcutaneous adipose tissue
SNF	Similarity network fusion
TAF	Tenofovir Alafenamide
TAG	Triglycerides
TDF	Tenofovir disoproxil fumarate
UNAIDS	United Nations Programme on HIV/AIDS
VAT	Visceral adipose tissue
VP	Viremic progressor

1 Introduction

1.1 Living with HIV

1.1.1 Epidemiology of HIV

Since 1981 and the Human immunodeficiency virus (HIV) identification, the HIV crisis has been a significant global public health issue. The HIV-1 infection leads to immune system impairment by infection and suppression of T helper cells[1]. If not treated over time, an acquired immunodeficiency syndrome (AIDS) will develop, leaving the body vulnerable to life-threatening opportunistic infections. In total, more than 80 million people were infected with the virus, and half have died from AIDS-related causes. In 2021, 38 million people globally were living with HIV, and 28.7 million people were under antiretroviral therapy (ART) (https://www.unaids.org/). ART reduces viral replication, increases CD4+ T cell counts, and highly improves the survival of infected patients, making HIV a manageable disease[2]. Nevertheless, a better investigation of the virus mechanisms and the development of new treatments and vaccines are necessary to control the pandemic.

1.1.2 No HIV cure or vaccine is available

The development of HIV vaccines has been highly challenging due to early virus integration into the host genome and the difficulty of targeting the fusion peptide of native HIV-1 envelope glycoprotein because of its variability, compactness, and glycosylation. From 1987 to 2013, six HIV vaccine candidates failed, and one had low effect[3]. The most recent trial, the Mosaic trial, went until phase three but was stopped due to a lack of efficiency[4] (https://www.mosaicostudy.com/). On the other hand, a recent project, the Antibody-Mediated Prevention trial, has started testing at which levels broadly neutralizing antibodies (bnAbs) protect from HIV. The bnAbs were shown to protect from HIV in robust animal models[5].

1.1.3 Antiretroviral therapy: saving and cost to the body

The principal treatment for HIV infection is ART. ART leads to viral suppression and tremendous expansion in life span. Treatment reduces virus load but fails its elimination, meaning patients must take the medication daily for the rest of their lives. People on ART take a combination of three HIV medicines (also called HIV regimens) from at least two different HIV drug classes (https://hivinfo.nih.gov/). The preferred ART regimens are usually two nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) (abacavir (ABC)/lamivudine or tenofovir alafenamide (TAF) /emtricitabine or tenofovir disoproxil fumarate (TDF)/emtricitabine) with one drug from a different type including booster protease inhibitors (PI) or integrase strand transfer inhibitors (INSTI), or non-nucleoside reverse transcriptase inhibitors (NNRTI)[6]. Food and drug administration-approved drugs

and their target are presented in figure 1. The Joint United Nations Programme on HIV/AIDS (UNAIDS) planed to have 90% of seropositive patients knowing their status in 2020, 90% of these patients should have access to sustained ART, and 90% of patients under drugs must present viral suppression. In December 2020, UNAIDS proposed a new objective for 2025, the 95-95-95, representing 95 % of patients knowing their status, 95 % on therapy, and 95 % with suppressed viral loads[7].

Good adherence to ART is crucial for efficient viral suppression and the absence of drug resistance but leads to consequent side effects. ART drugs display a large panel of side effects that have been carefully studied and lowered with the newer regimen Patients are screened for mutations and a history of comorbidities before starting ART and monitored during the treatment. The common side effects of ART drugs are short-term gastrointestinal effects or fatigue and, in the long term, vary between classes of drugs, drugs, and patients' genetic backgrounds. Complications (Figure 1) might lead to severe cardiovascular, kidney, and bone diseases[6], decreasing the quality of life.

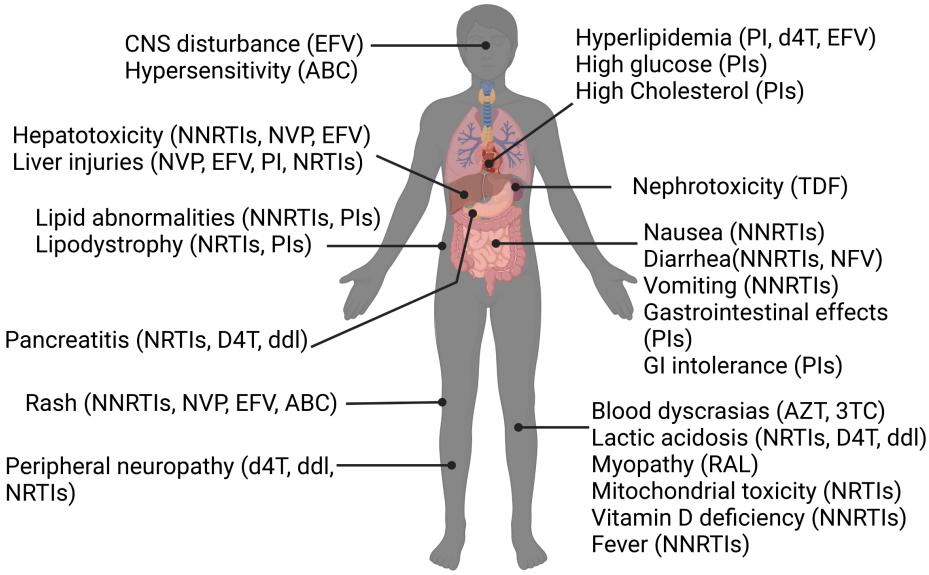


Figure 1: ART's most common side effects per drug regimen and for specific drugs. Created with biorender. Adapted from Reust et al[6], https://www.hivaidsclinichyderabad.com/ and Montessori et al[8].

Regarding toxicity, old ART regimens (NRTIs, PI) have more incidence of insulin resistance and body composition changes, while new treatments (tenofovir, abacabir) affect glucose and fat metabolism[9]. These effects and virus induced chronic inflammation and immune activation lead to immuno-metabolic disorders, including metabolic syndrome. In low-income and middle-income countries, tenofovir+efavirenz with lamivudine or emtricitabine is the first-line regimen due to its availability and high adherence[10]. However, some patients switch from PIs and NRTIS to integrase inhibitors[11].

1.1.4 Aging with HIV

Aging is described as the decreased function of organs and biologic systems with time[8]. In humans, aging is already a risk factor for many diseases, such as cardiovascular diseases (CVDs), cancers, arthritis, cataracts, osteoporosis, and neurological diseases. The probability of developing aging-related diseases increases exponentially with age[12]. Interestingly, not all patients are developing aging-related comorbidities at the same age, indicating that the aging process is not uniform or constant and highly influenced by genetics and personal lifestyle.

The protective effect of the immune system declines with age. Older people display increased and more severe infections, lower immune surveillance, and decreased response to vaccines. This process, called immune aging or immuno-senescence, involves all the immune cell types[13]. Senescence provokes cell cycle arrest in response to various stressors and is a major biological element of aging. Cell senescence and innate immunity dysregulation lead to monocyte/macrophage lineage cell activation and chronic low-grade inflammation defined as inflamm-aging[14] by the increased production of proinflammatory cytokines. Alterations in mitochondria could also be part of the developing the immune-aging process. Mitochondrial DNA (mtDNA) depletion and reactive oxygen species (ROS) production from mitochondria activate the immune response as circulating mtDNA correlates with serum inflammatory markers[15].

ART has drastically reduced the morbidity of HIV, and consequently, the number of elderly patients living with HIV has significantly increased[16]. Nevertheless, it has been shown in several studies that HIV patients under long-term successful ART are suffering from accentuated or accelerated aging effects. Age-related disorders occur in young HIV patients compared to the controls. It includes metabolic disorders, non-AIDS-related cancers, CVDs, liver-kidney diseases, neurocognitive disorders, bones implications, and frailty[1,17–19]. These diseases are suspected to depend on multiple factors, including chronic viral activity, ART toxicity[20,21] (Figure 1), and thymopoiesis impairment[13]. Individual risk factors such as smoking, weight, sex[22], co-infections[23], and late starting of cART observed in Sweden[24], have been proven to have a crucial role in the mosaicism of a patient's phenotype. The geriatric facilities have implemented screening for multi-morbidities, and regular follow-up has been implemented by the geriatric facilities[16].

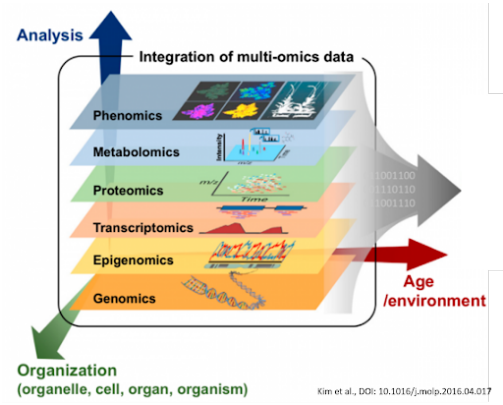
Regarding the immune system, clear parallels were observed between aging and HIV. First, telomere lengths in CD8 T and B cells were shorted in PWH. More, memory CD8 T cells were reported in HIV and elders. Also, a reduction of CD4 and CD8 naïve T cells' frequency, indicating decreased thymus function, has been reported in PWH on ART. Defects in the activation of naïve CD8 were also observed. A decrease in B cells and peripheral DC were observed in PWH under ART. Finally, if monocyte types are similar in

PWH under ART than in HC, PWH display activated cells (expressing CD163 and CXCL10) and proinflammatory cytokines that could lead to inflammatory age-related comorbidities. Finally, HIV patients display impaired hematopoiesis and poorer response to the influenza vaccine despite effective viral suppression[13].

1.2 OMICs Analysis

1.2.1 The Omics revolution

Omics includes biology fields informally finishing in -omics (Figure 2). It is related to the term -ome, which describes the totality of a biological system. It includes genomics, transcriptomics, epigenomics, proteomics, metabolomics, microbiome, or fluidomics[25]. The Omics field is recent. Omics are universally used to study human diseases in case-control studies[26].



Omics analyses are applied to identify or quantify molecules from a sample material to understand complex biological systems better and develop therapeutic drugs later on[25]. The development of omics was made possible by advancing high-throughput technologies as next-generation sequencing (NGS)[25] and Liquid Chromatography-Mass Spectrometry (LC-MS/MS)[27].

Figure 2: 3D Multi-omics layers. Figure from Kim et al, 2016[28]

1.2.2 Multi-omics integration

Omics integration aims to answer a biological question by finding biological patterns not apparent in single omics. If single omics gave results about one layer, the multi-omics models allow retrieving the interactions between molecules and generally the layers of omics. Before integration, the normalization of individual omics before global scaling is necessary as data comes from different platforms[26]. Omics studies in patients are challenging due to patient variability, cost of replicates, ethical issues, and difficulty in sample collection. Biological candidates identified in omics studies need to be validated using more efficient methods such as long-term large-scale trials[29]. Analysis should also differentiate the cause of disease from consequences that can lead to the selection of the wrong targets for treatment.

Using previous knowledge, omics type, and number of samples are essential for selecting the proper integration tool[26]. First, we chose matrix factorization-based tools

because they are straightforward, unsupervised, and manage the integration of continuous, binary, and count data[30]. Then, the network analysis is a flexible method that allows extracting "modules" of molecules that can be further analyzed together in their association to phenotype[27]. Finally, we will overview genome-scale metabolic models as integrative tools that give a mechanistic explanation of the associations observed[31].

1.2.3 Matrix factorization-based tool and MOFA

The simplest approach for omics integration is called matrix factorization, a method also used for principal component analysis (PCA). The matrix factorization focuses on projecting into dimension-reduced space of the variation among data. An example of a tool based on matrix factorization is the Multi-Omics Factor Analysis (MOFA)[32], which allows unsupervised integration of heterogeneous data. MOFA is described as a statistical framework adapted to omics. The input files are data matrices with different omics data. The authors recommend removing any technical variability source using size factor normalization and variance stabilization for RNA-sequencing (RNA-seq)[33]. MOFA works with data from the same samples but can also handle partially overlapping samples. MOFA handles non-overlapping features, non-normal distribution, and missing values. MOFA gives a low-dimensional data representation of latent factors capturing significant variation sources, decomposed into factors (samples and factors) and weight matrices (features vs. factors for each omics layer). MOFA can compare factors between datasets and find common traits between omics layers. It presents as a compelling method as it allows for integrating all types of omics and has already been extensively employed in other contexts[30].

1.2.4 Network analysis

Network analyses tend to represent biological systems by sets of nodes connected through edges, thus permitting the topological analysis of the network. This means that one can present the interactions between the nodes in a straightforward way and be used to provide biological context. Here, nodes represent molecules, and edges drawn between nodes imply a relationship between two molecules (correlations, directed effect, indirect effect, etc.). Many network methods exist with supervised and unsupervised methods[34]. Here, we present methods we applied,, including co-expression analysis, similarity network fusion, and genome-scale metabolic models.

1.2.4.1 Graph-based approaches for network analysis

Networks can be analyzed using graph-based approaches. Only concepts used in the thesis will be presented, and many alternative measures are described in the literature[35,36]. The node connectivity measure is particularly important to determine highly interconnected nodes or hubs[37]. A hub is described as a node with many edges in a central localization in the network[35]. The centrality defines important structural actors in the network[38]. Metrics to measure centrality can be degree, degree centrality,

normalized degree, or global centrality, among other measures. The degree considers the number of immediate connections called first neighbours and is considered a local measure. Degree centrality assigns a score based on degree to a node. The normalized degree is the degree divided by the maximum degree. Global centrality represents centrality taking into consideration the full network indicating the finding of hub nodes in a network. The betweenness centrality represents the number of shortest path (minimized number of edges separating two nodes) passing by a node[38]. The closeness centrality is the distance between a node and each other node and represented as the global measure of degree (Figure 3).

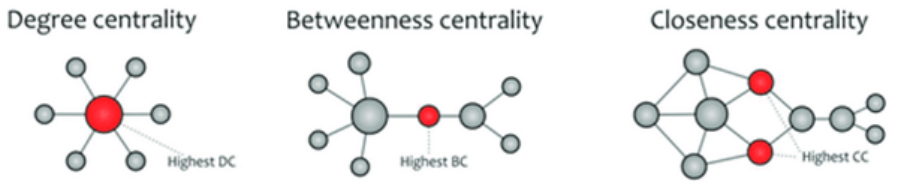


Figure 3: Examples of measures of network centrality. From Farahani et al[39].

1.2.4.2 Association analysis and community detection

Association analysis or co-expression analysis is a highly relevant tool for omics integration because they reveal co-expression patterns of features[40]. Many methodologies exist for community detection[41]. The most common is the weighted gene co-expression network analysis[42]. Here we present two algorithms: Leiden and Louvain. As described previously, proper normalization and scaling are crucial to compare all omics data in the same network. To overpass a binary connection (0 = not connected, 1 = connected), connection weight is added to each edge, for instance, based on the correlation coefficient. Clustering algorithms are applied to find highly interconnected molecules poorly connected with the rest of the network resulting in communities[43]. Louvain algorithm starts by defining each node as its community. Then nodes are moved from one community to another to find partitions. Then, the aggregate network is created. Then, the nodes are changed in the network until the quality cannot improve through multiple iterations. The Leiden algorithm begins with the local moving of nodes. Only nodes whose neighbors have changed are visited; to change them to a different community. Then the partition is refined several times, and the network is aggregated[44]. Pathway analysis can be performed on communities if we assume that correlated molecules can reflect a logical biological signal.

1.2.4.3 Similarity network fusion

To identify omics-driven clusters of patients, the similarity network fusion (SNF) analysis reduces high dimensional data without having the inconvenience of mixing data types. Patient-by-patient similarity matrices and similarity networks for single omic are first

calculated. Then, individual networks are fused by iteratively updating each network with information from the other networks until convergence to the final network[45]. A strength of this approach is that merging dimensional networks is less dependent on the biases in individual omics. SNF gives a confusion matrix of the similarities between the final network and the individual networks for each omic type and identifies which layers impact the patients clustering more. Though SNF has not been employed in HIV, it has been used for patient grouping in several fields. Similarity has been applied primarily in cancer research to identify cancer subtypes[45] and still nowadays for cancer subtyping[46] and prognostic groups[47]. SNF has also been used to make groups in patients suffering from respiratory diseases[48] and to study the influence of diet[49]. Then, statistics can be done between clusters to identify molecules and differing mechanisms.

1.2.4.4 Genome-scale metabolic models

Genome-scale metabolic models (GSMM) are another network-based approach used to characterize the known metabolic networks systematically. They are used in data integration for information about a system and predictive power. Integration for GSMM can be done using tINIT tool[50] based on omics, where features are considered into not expressed, low, medium, and high expression. The model is then used to identify those essential reactions necessary for a model to be biologically feasible. tINIT can also implement loops and blocks the production of some metabolites[50]. Then, optimization is performed using flux balance analysis (FBA) which calculates the optimal flux according to the objective function[51]. Metabolism involves important biochemical reactions involving the transformation of metabolites catalysed by enzymes[52], where GSMM comprehensively considers all reference human metabolic networks.

GSMM can be built using quantitative metabolomics, fluxomics, and enzyme data (quantitative proteomics or transcriptomics). GSMM assumptions are that the system is in steady-state, which presents an advantage because it allows for computational identification of optimal flux distributions without requiring detailed knowledge of enzymatic kinetics. However, this approximation comes at the cost of assuming that our system is in a constant and stable state and is strongly dependent on the imposed constraints. Constraints such as genetic, physicochemical, and reaction directionality are integrated into the model. The tool GIMME builds reactions based on metabolomics data, and then transcriptomics data are added to adjust the flux values. Refinement can be done on the created model with a gap-filling procedure also implemented into COnstraint-Based Reconstruction and Analysis (COBRA) and Reconstruction, Analysis, and Visualization of Metabolic Networks (RAVEN) to reduce errors due to the missing knowledge about biochemical reactions. Reactions linked to human organelles can also be implemented (Figure 4).

GSMMs have been extensively employed to study metabolism shifts associated with human diseases. Recon 1 Human GSMM was created in 2007 and progressively improved over years[53], including tissue-specific reactions and model refinement. Several models were created for different tissues[52]. For the studies on human metabolism, the generic human GSMM Human-GEM was recently curated (http://www.metabolicatlas.com/) based on Recon 1 and other sources such as HMR2 and Recon 3D[52]. This model comprises 3625 genes, 13417 reactions, and 4164 metabolites. Different reference models must be created based on each organism and cellular conditions. New methods of collecting information about biochemical kinetics increase the accuracy and predictive power of genome-scale models[54,55].

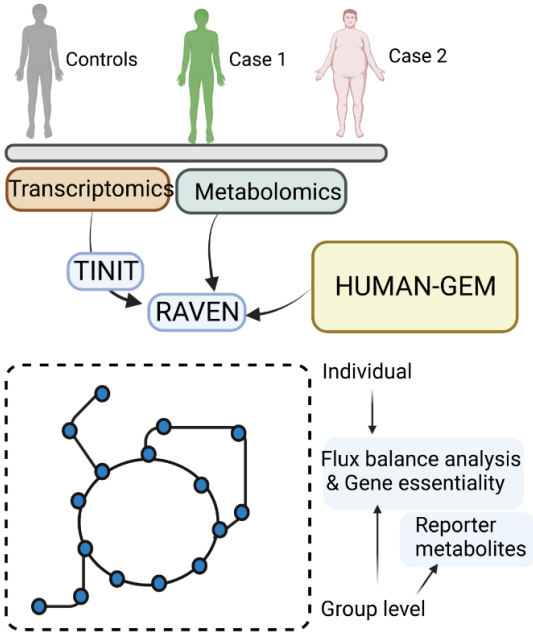


Figure 4: Workflow of Genome-scale metabolic model using the RAVEN toolbox

1.2.5 Omics studies in HIV

To identify the biological mechanisms associated with PWH status, omics studies are highly relevant. Most studies investigate biomarkers in a relatively low number of patients by comparing PWH on ART with HC or viremic progressor (VP)[56–61] or differences in HIV-infected cell lines[62–66]. The most recent works actively study co-infections[67], effects of treatments[57,68], and comorbidities in HIV[58,69]. Examples of recent single omics studies are presented in Table 1. Most studies display a low number of patients, a short ART duration, and no omics integration, which we addressed in our study.

Table 1 : Single omics studies HIV (transcriptomics, proteomics, metabolomics and microbiome)

Omics (Method)	Patients / Models	Results	Reference
Microbiome (16S rRNA)	10 HC 15 HIV+ EFV 15 HIV+ PI Groups comparisons	Lower alpha diversity in PWH. Markers of disturbance gut homeostasis in HIV + PI	Pinto-Cardoso et al[68]
Microbiome (16S rRNA)	35 PWH before and after therapy (24h) Diversity and differential abundance	PWH with low CD4 counts had higher oral bacterial richness and diversity. Species associated with periodontal disease increased with ART	Presti et al[56]
Metabolomics Lipidomics	217 HC 218 PWH before and after therapy PLS-DA, lasso, correlations	PWH displayed higher TAG and lipogenesis. PWH on ART show high inflammation linked to lipid metabolites modification after ART start	Jao et al[57]
Metabolomics	87 HC 87 PWH (3 years) 148 PWH + non-communicable diseases	Glycerophospholipid, glutamine and glutamate metabolism are disrupted in PWH with non-communicable diseases	Ding et al[58]
Proteomics Metabolomics	Review - 13 studies PWH with HAND in different tissues	Identification of myoinositol (metabolic) and SOD-1, gelsolin, afamin, sphingomyelin, and ceramide (proteomic) marker of HAND	Williams et al[69]
Transcriptomics (single-cell RNA-seq)	9 VP 8 PWHART (< 1 year) 4 HC peripheral T cells	Decrease in naive T cell and increase in inflammation partially restored in ART	Wang et al[59]
Transcriptomic, proteomics phosphoproteomics	SupT1 CD4+ T cells with or without HIV (five-time points) a Gaussian mixed-effects model	Host responses of PWH involved in transduction, metabolic pathways, cell signaling, and immune regulation	Golumbeanu et al[62]
Proteomics (SWATH mass spectrometry)	Jurkat cell lines U937 Group Comparisons and correlations	During latent HIV, the two pathways CypB and CD147 as well as intracellular processing and translocation factors are disregulated	Belshan et al[63]

Previous work from the lab investigated omics differences in PWH but in a smaller cohort and mostly as case-control studies (Table 2).

Table 2: Recent Omics studies from the lab

Omics (Method)	Patients / Models	Results	Reference
Metabolomics Proteomics	22 each HC and PWH Correlations, PLS-DA, RF	Essential amino acids differ between HC and PWH. Markers of inflammatory and neurological diseases dysregulated in HIV	Babu et al[70]
Metabolomics Proteomics	22 HC, 29 $PLWH_{ART}$ and 11 VP Correlations, co-expression analysis, UMAP	High glutamate, lactate, and pyruvate and inflammatory markers in PWH. Role of Myeloid cell populations	Akuskarvi et al[71]
Transcriptomics Metabolomics	19 PWH each (ART, EC and VP GSSM	Flux balance analysis identified dysregulated glycolysis	Ambikan et al[72]

Multi-omics studies in HIV are still limited in the literature. Multi-omics integrations in PWH related to meta-analysis and case-control studies are reported in Table 3.

Table 3 : Multi-omics integration studies in HIV

Omics (Method)	Patients / Models	Results	Reference
Transcriptomics, proteomics, epigenomics	Public databases Co-expression, networks analysis, Pathway analysis	Identifying interactions, pathways, proteins, and protein-protein interactions linked to HIV infection	Ivanov et al[73]
Microbiome, methylome transcriptomics,	34 PLWH and 42 HC DIABLO	PWH have lower bacterial diversity and correlation of TRNAU1AP, two CpGs, and two ASVs, linked to oxidative stress in airway epithelial cells	Jude et al[74]
Metabolomics proteomics transcriptomics	12 PWH 8 PWH + IRIS (2 month before and 12 months after ART) MOFA	IRIS was associated to oxidative stress, tryptophan pathway, lipid-mediated signaling, inflammation and cell type-specific immune activation.	Pei et al[75]

Meta-analysis can be efficient in studying HIV. The strengths of a recent network-based analysis[73] were the use of several methods, including computational and manual omics. They also identified the overlap between many studies and found consensus mechanisms of HIV infections. They calculated a risk score for each pair of comorbidities. The weaknesses are that they had no control over data production and handling, which can induce bias. More, databases are not always displaying patients' characteristics which could also induce co-founding effects (origin, age, gender). Related to this point, some categories of patients are over-represented in databases (Europeans, men having sex with men (MSM)), which were not corrected in this analysis.

Cases-controls studies are also relevant using omics integration. A study comparison of PWH to PWH with Immune reconstitution inflammatory syndrome (IRIS) was done[75]. Comparisons were made at single omics and using MOFA for multi-omics integration. Despite a low number of patients and short ART time, the study was robust in identifying the mechanism underlying IRIS. Another study integration was performed in the airway epithelial cells of PWH, and HC controls using Data Integration Analysis for Biomarker discovery using Latent variable approaches for Omics studies (DIABLO), another integration tool[74]. One limitation is that DIABLO does not work with all omics types, and integration of count data is still in development. Secondly, authors could have validated their findings in other cohorts. Again, patient information and clinical parameters are not included in the analysis.

In conclusion, multi-omics integration is a powerful asset for understanding complex mechanisms of HIV in patients and laboratory models. Regardless of the cost of omics and the difficulty of working with HIV samples, developing more extensive studies with precise methodologies is underway.

2 Research aims

The general aim of the Ph.D. was to apply multi-omics and system biology tools to understand HIV-associated immune aging. My omics of interest were transcriptomics, proteomics, metabolomics, lipidomics, and microbiome data. I aimed to find biomarkers, pathways, and underlying mechanisms related to metabolism and immune aging in PWH following long-term ART therapy using omics, advanced statistics, multi-omics integration and network approaches.

Specific aims

Paper I: To investigate metabolic perturbation in successfully long-term well-treated PWH with metabolic syndrome (MetS) compared to PWH without, which is a major comorbidity in PWH.

Paper II: To characterize the metabolomics profile from PWH on ART and lifestyle and gender-matched HIV-negative controls from two low- and middle-income countries (LMICS), Cameroon and Indian cohorts.

Paper III: To stratify PWH with prolonged suppressive therapy based on plasma metabolomics/lipidomics and fecal microbiota into risk groups of metabolic complications. Also, we tried to identify clinical and biological factors differing from these clusters using advanced bioinformatics tools.

Paper IV: To identify data-driven patient stratification using a robust system biology integration tool on transcriptomics, proteomics, and metabolomics data on PWH with prolonged suppressive therapy. We tried to identify cellular, immuno-metabolism mechanisms differing risk-clusters by applying advanced statistics, genome-scale metabolic models, and in vitro cell models.

Paper V: To investigate the biological aging process using transcriptomics age estimator in a large cohort of PWH on prolonged successful ART and identify PWH with accelerated aging and underlying mechanisms.

3 Materials and methods

3.1 Ethical considerations

Ethical clearances covered all patients' material. Ethical clearance for the Copenhagen Comorbidity in HIV infection cohort (COCOMO) study **(papers I, III, IV, and V)** was approved by the Regional Ethics Committee of Copenhagen, Denmark (COCOMO: H- 524 15017350) and Etikprövningsmyndigheten, Sweden (Dnr: 2022-01353-01). The Cameroonian National Ethics Committee approved the study performed in Cameroon (**paper II**) for Human Research, Cameroon (N2019/08-198-CE/ CNERSH/SP). The study conducted in India **(paper II)** was approved by the Institutional Ethics Committee of the National Institute for Research in Tuberculosis (NIRT IEC No: 2015023) and the Institutional Review Board Committee of Government Hospital for Thoracic Medicine (GHTM-27102015) Chennai, India. The Ethical approval (Etik- prövningsmyndigheten, Sweden) was waived (Dnr: 2019-05086). All study participants wrote an informed consent before inclusion.

3.2 Samples collection and pre-processing

For **papers, I, III, IV,** and **V**, patients were part of the COCOMO cohort initiated in 2015 (N = 1,089)[76]. PWH from Copenhagen area were included at the Copenhagen Rigshospitalet and Hvidovre Hospital. The cohort of healthy controls (HC)[77] was age-and sex-matched with the PLWH. All patients completed a REDCap electronic data capture questionnaire regarding more than one hundred clinical parameters, including demographic, lifestyle, self-reported symptoms, and diseases. Venous blood samples were collected from non-fasting patients. The whole blood, buffy coat, and plasma serum were stored. Peripheral blood mononuclear cells (PBMCs) were extracted from the whole blood. The patients collected stool samples themselves. In the Cameroon project (**paper II**), HC and PWH were gender- age- and body mass index (BMI)- matched. The samples from plasm and whole blood were retrieved from Yaoundé University Teaching Hospital, Cameroon. In the Indian study (**paper II**), PWH with successful long-term ART (> five years) were selected from patients at the Government Hospital for Thoracic Medicine (N = 553), Chennai, India. HC matched for age, sex, and lifestyle with PWH were recruited in Chennai, India. A non-fasting blood sample was taken from each patient. Samples were sent to Karolinska Institutet and omics generation companies. Samples used in different papers for different methodologies are indicated in Table 4.

Table 4: Summary of samples and omics technologies per paper

Paper	Cohort	PWH	VP	HC	Methods
Paper I	COCOMO	200	0	20	Metabolomics
Paper II	Cameroon India	50 41	25 20	50 30	Metabolomics (targeted and untargeted) Proteomics (LC-MS/MS)
Paper III	COCOMO	97	0	18	Metabolomics, Lipidomics and 16S Microbiome
Paper IV	COCOMO	158	0	18	Metabolomics, Transcriptomics Proteomics (Olink, LC-MS/MS)
Paper V	COCOMO	178	0	18	Transcriptomics

3.3 Cell lines

For **paper II**, cell lines Jurkat (Leukemic T lymphocyte) and U937 (lymphomatic pro-monocytic cell line) and their HIV latency cell models J-Lat 10 (NIH HIV reagent program) and U1 were provided by Helena Jernberg Wiklund, Uppsala University. For **papers II** and **IV**, PBMCs were isolated from the whole blood of the donor samples obtained from Karolinska University Hospital.

3.4 OMICS analysis

Omics is the characterization of the global set of biological molecules at a specific time point[78]. All omics analysis included dimensionality reduction (PCA) to detect outliers, biasing sources, and potential clustering patterns among patients, followed by filtering of empty or missing data features and normalization. Finally, comparing features between groups was usually done using statistics adapted to each data. All analyses were regularly saved and shared on GitHub (https://github.com/). Each data type and its characteristics are presented in figure 5 and below.

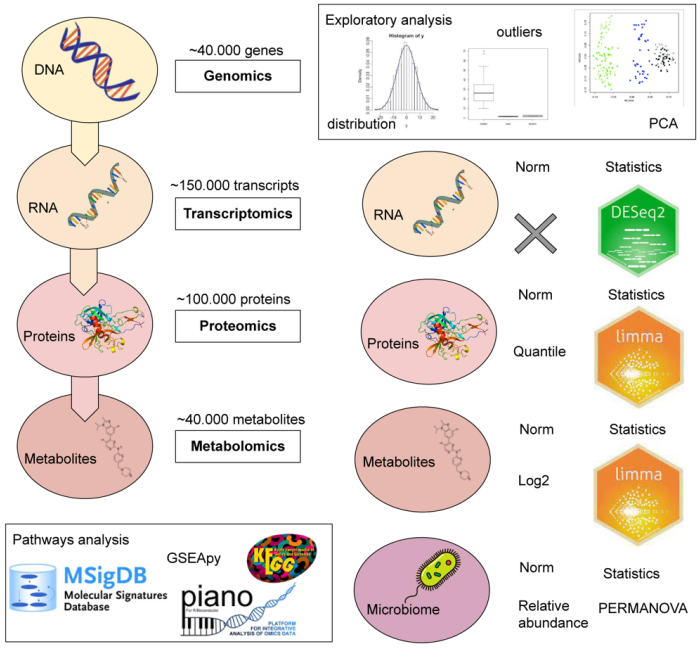


Figure 5: From genes to metabolites, linked omics technologies and pipelines used in the thesis

3.4.1 Transcriptomics

Transcriptome is the complete set of transcripts and their quantity in a cell for a specific physiological condition. RNA-seq has the advantage of giving both quantitative and qualitative gene expression and has become a favored technology for transcriptome-wide gene expression analysis. RNA-seq is still limited for short sequences due to non-accurate read alignment and sequencing errors[79]. Measuring messenger ribonucleic acid (mRNA) is also not always well correlated to the level of the corresponding proteins since the stability of mRNA and proteins differs. For **papers IV** and **V,** RNA-seq was done on PBMCs samples using Illumina TruSeq Stranded mRNA (Illumina, USA) with Poly-A selection followed by sequencing on NovaSeq6000 and processed by nf-core/rnaseq17 by the National Genomics Infrastructure (NGI), SciLifeLab. We used transcripts per million as standard normalization and DESeq2[80] on unnormalized raw counts for differential expression analysis (DGE).

3.4.2 Proteomics

Proteomics analysis represents the quantitative measurement of proteins from a sample[81]. Targeted proteomics is the most sensitive and is used to detect and quantify a particular protein or peptide in a complex mixture of proteins. In opposition, untargeted proteomics is less sensitive but allows qualitatively and quantitatively measuring the proteome without targeting a specific protein. Still, compared to transcriptomics, the number of proteins detected is limited, around 10 000 proteins for LC-MS/MS, and not representative of the whole proteome, which contains between 80,000 and 400,000 proteins[82]. Moreover, proteins are subject to post-translational modifications such as phosphorylation, glycosylation, lipidation, and cleavage of peptide bonds. In **papers II** and **IV**, LC-MS/MS proteomics was done by the Proteomics Biomedicum, Karolinska Institute, Solna. After evaluation of several normalization methods using NormalizerDE[83], we applied quantile normalization to the raw data. In **paper IV**, a targeted protein set was extracted using a proximity extension assay by Olink Bioscience AB, Uppsala, Sweden. We performed differential abundance analysis using linear models for microarray data (limma)[84] .

3.4.3 Metabolomics

Metabolomics describes changes in metabolites, the intermediate end products of metabolism. Metabolites are the most representative of the entire organismal state compared to transcriptomics or proteomics[85]. As for proteomics, metabolites are quantified using untargeted or targeted mass spectrometry technologies. Metabolites have been shown to affect the phenotype directly[86] and disease status[87,88]. The reproducibility of metabolomics studies is still challenging. It has been shown that food habits, gut microbiome composition, smoking or alcohol have crucial impacts on metabolites [89,90]. Moreover, between 30 and 50 % of metabolites are influenced by

genetic background[91]. In papers, **I, II, III, and IV**, untargeted metabolomics were done using the Metabolon HD4TM Discovery platform (Metabolon Inc, Morrisville, NC 27560, USA) using ultrahigh-performance liquid chromatography/mass spectrometry/mass spectrometry. In **paper II**, amino acids detection (targeted metabolomics) using LC-MS/MS method was performed at the Swedish Metabolomics Centre (Umeå, Sweden). For analysis, data were log2 transformed and groups compared using limma[84].

3.4.4 Lipidomics

Metabolomics is the study of water-soluble metabolites (ex: sugar, amino acids, nucleotides), lipidomics measures the lipid molecular species using similar technologies[92]. The main limitation of lipidomics is mostly the difficult quantification of low-abundance lipid species[93]. Lipidomics was done through the Complex Lipid Panel™ technique (Metabolon Inc, Morrisville, NC 27560, 433 USA) for **paper III**. Analysis pipeline was similar to the metabolomics pipeline.

3.4.5 Microbiome

Microbiota is the ensemble of bacteria, archaea, protists, fungi, and fungi communities found in an organism. Microbiome analysis represents the study of these microorganisms as communities[25]. In the study of the association between disease and microbiome, important individual variations, including lifestyle, diet, medication, and physiology, must be taken into consideration[94]. Microbiome data is highly influenced by technical variation and contaminants. The 16 sequencing rRNA (16S rRNA) targets highly variable genome regions of bacteria to determine microbial phylogeny. In **paper III**, we used 16S rRNA done at the Norwegian Sequencing Centre in Oslo. The sequencing was performed with the Illumina MiSeq platform and v3 kit (Illumina). The microbiome analysis consisted of calculating alpha and beta diversity and using the permutational multivariate analysis of variance test[95] to compare microbial communities between groups.

3.5 Bioinformatics tools

Several tools were applied in this thesis to better extract underlying mechanisms from complex biological systems. Tools used in the thesis are indicated in Figure 6.

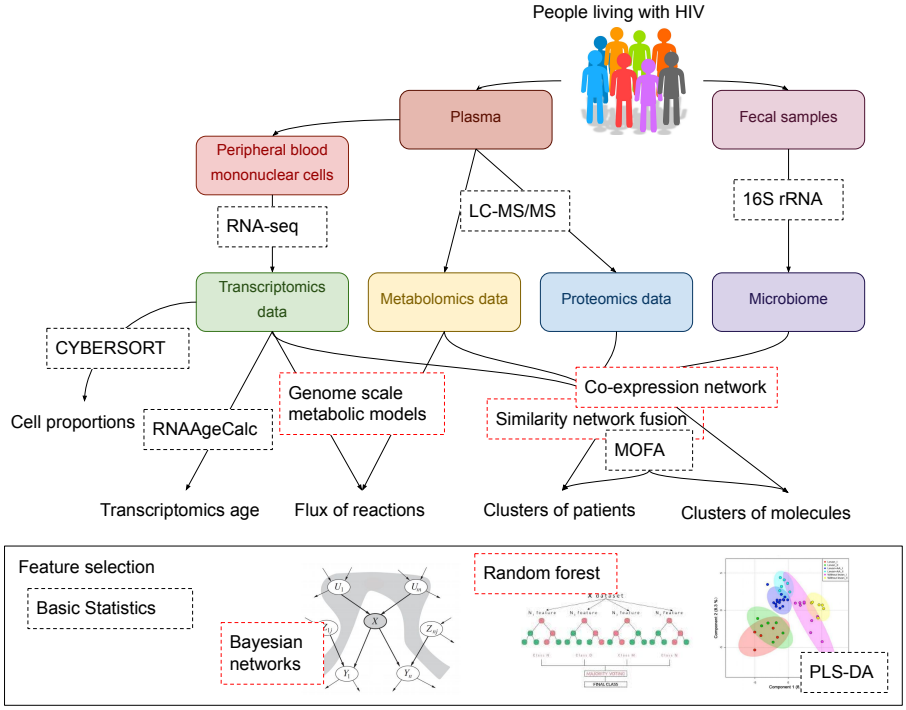


Figure 6: Bioinformatics tool used for the different studies. Tools are indicated in dotted boxes. The red color indicates system biology tools.

3.5.1 Feature selection

Feature selection reduces the number of variables used as input for a model. It is particularly used in omics data to reduce the "curse of dimensionality," where the number of features is much higher than the number of observations. Variables are ranked by a score specific to each method. In **papers I and II,** we used Partial least squares-discriminant analysis (PLS-DA)[96], a supervised dimensionality reduction tool that tries to have a higher variance between a dependent variable (ex: two conditions) and the independent variables. Then in **papers, I, II, and IV**, the Boruta algorithm, a wrapper built around the machine learning algorithm random forest (RF) was also applied[97]. RF merges the multiple decision tree's output into final results and can be used for regression and classification models[98]. The Bayesian Belief Networks (BBNs) used in **paper IV** is a network-based approach that identifies driver genes with the most decisive influence on a consensus and refined network structure. This network is obtained after the generation of many random trees from the hill-climbing algorithm[99].

3.5.2 Pathway analysis

Pathway analysis or functional enrichment analysis allows the identification of larger biological themes from a list of features from a comparison (usually fold change and p

values are obtained from statistics test)[100]. In this thesis, we used Ingenuity Pathway Analysis software (IPA)[101](**paper I** and **II**), Gene Set Enrichment Analysis (GSEA) in Python (gseapy)[102] (**papers II** and **IV**), piano[103] (**paper IV** and **V**), and Biological Networks Gene Ontology tool (BINGO)[104] (**paper IV**)(Figure 5).

3.5.3 Sample clustering and multi-omics integration

We used ConsensusClusterplus to identify clusters of patients by repeated subsampling and clustering steps[105] (**paper II**). Omics integration aims to answer a biological question by finding biological patterns not apparent in single omics. Indeed, Single omics analysis can lead to incomplete results. SNF[45] (**paper III and IV**) and its derived program netDx[106] (**paper IV**) were used to integrate data by building sample similarity matrices and similarity networks for each data type and then fusing them into a consensus network. In **paper III**, MOFA[30] gave a low-dimensional representation of latent factors capturing most sources of variability in the omics. Weights associated with each feature were extracted.

3.5.4 Other network analysis

3.5.4.1 Co-expression analysis

Co-expression analysis[107] is mainly used for detecting highly intercorrelated modules of features and was used in **papers I to IV**. Pairwise correlations were performed using Spearman correlations and filtered based on the false discovery rate (FDR). We performed co-expression analysis using igraph[108] and Leiden[44] algorithms for community detection. Pathway analysis and centrality analysis were performed for each community.

3.5.4.2 Genome scale metabolic models

Metabolism involves important biochemical reactions involving the transformation of metabolites catalyzed by enzymes. GSMM is a reconstruction of the metabolic reactions of an organism[52]. The generic human GSMM Human-GEM was modified using transcriptomics and metabolomics data using the RAVEN toolbox[50]. Then, the optimization was performed using FBA, which finds an optimal flux distribution given a pre-decided cellular objective[51] (**paper IV**).

3.5.5 Visualization and databases

Most plots were made with ggplot2[109] in R (**paper I to V**). Plots were designed and colored for the best readability of the results. Networks were made with Cytoscape[110], which is a tool for network analysis (**paper I to IV**). Plots were also done in python (**paper II**).
The three databases used for this thesis were Kyoto Encyclopedia of Genes and Genomes (KEGG)[111](https://www.genome.jp/kegg/) (**papers I, II, IV and V**), the Molecular Signatures Database (MSigDB)[112] (https://www.gsea-msigdb.org/gsea/msigdb/) (**paper IV**) (Figure 1) and STITCH[113] (http://stitch.embl.de/) (**paper I**). KEGG is a database based on high-throughput technologies containing several biological systems information and functions

including for example pathways, complete genomes and orthologs. MSigDB contains annotated gene sets for humans and mice. It was made with the GSEA software. STITCH is a database of interactions between proteins and chemicals from genomic context predictions, high-throughput lab experiments and databases.

3.5.6 Others

3.5.6.1 Cell profiling

Cell profiling was performed to determine the pourcentage and abundance of each immune cell type in samples of mixed cell population (here PBMCs) using the online server CIBERSORTx[114].

3.5.6.2 Effect sizes and regression models

The effect size measures the magnitude of the relationship between two variables over significance[115]. In R, we used the R package effects to calculate effect sizes (**paper II** and **paper IV**). Galaxy web application LDA Effect Size (LEfSe)[116] was applied to find cluster-specific microbes (**paper III**). Regression models measure the association between a variable (called dependant variable) and one or more independent variables[117]. We used linear regression to correct features for clinical parameters (**paper I, paper II**) and logistic regression to find clinical variables differing between two groups (**paper V**). Transcriptomics age was calculated as described in **paper V**.

3.5.6.3 Single-cell transcriptomics analysis

Single-cell RNA seq is the study of the transcriptome at the single-cell level. This allows the extraction of information at the individual cell level compared to previous bulk analysis. All analyses were done using the R package Seurat[118] (**paper IV**).

3.6 Analytical methods

Wet lab experts from the group or collaboration performed all analytical methods.

3.6.1 Flow cytometry

All analyses were performed using Fluorescence-activated cell sorting (FACS) symphony (BD Bioscience) and FlowJo. FACS[119] measured essential metabolic transporters with central carbon metabolism in T cells and monocytes (**paper I**). In **paper II**, we measured ROS production and latency reactivation in HIV cell models. In **paper IV**, we also used FACS to measure surface receptors, functional markers, activation from latency markers and exhaustion markers.

3.6.2 T-cell and monocytes functionality assay

In **paper IV**, PBMCs from donors were incubated with patients' plasma (conditions are separated), simulated with the CEF Control Peptide Pool or unstimulated, and

supplemented with DNase I recombinant and antiviral inhibitors were added before sample collection. Samples were stained for surface CCR7, surface receptors, and aqua viability stain for t-cell assay. For monocytes assay, PBMCs were incubated with plasma from patients and simulated with lipopolysaccharide or unstimulated.

3.7 Experimental assays

All experimental assays were performed by wet lab experts from the group or collaboration.

3.7.1 Cell treatments with chemicals

In **paper II**, U1 and U937 were treated with glutamine antagonist 6-diazo-5-oxo-L-norleucine (DON), prostratin and 2-Deoxyglucose (2-DG). Mitochondria were extracted from these cells, and proteins from the oxidative phosphorylation (OXPHOS) pathway were measured using the total OXPHOS Human WB antibody cocktail (Abcam). Proteins were compared using statistics. Also, the cytotoxicity of the two ART regimens, still in these cells, was quantified using AlmarBlue assay (Invitrogen). Glucose-GloTM Assay, Lactate-GloTM Assay, and Glutamate-GloTM Assay (Promega) were used to measure the intracellular concentration of glucose, lactate, and glutamate in these conditions. In **paper IV**, monocytes from donor PBMCs were treated with spermine, spermine, or both or untreated. Differentiation to macrophages was induced using LPS, and after 48h, samples were sent for proteomics.

4 Results and Discussion

Immune aging in HIV is a complex process driven by several factors. The thesis aimed to understand this phenomenon better using different cohorts, multi-omics, and system biology tools (Figure 7). In **paper I**, we compared metabolomics profiles of PWH with or without metabolic syndrome, a common aging-related comorbidity in PWH. Then in **paper II**, we extracted metabolites differing from HC and PWH in two independent cohorts to identify long-term metabolic changes due to long-term ART. We divided patients into risk groups based on multi-omics profiles in **papers III** and **IV**. In **paper V**, we investigated the transcriptomics accelerated aging and potentials associated factors in PWH.

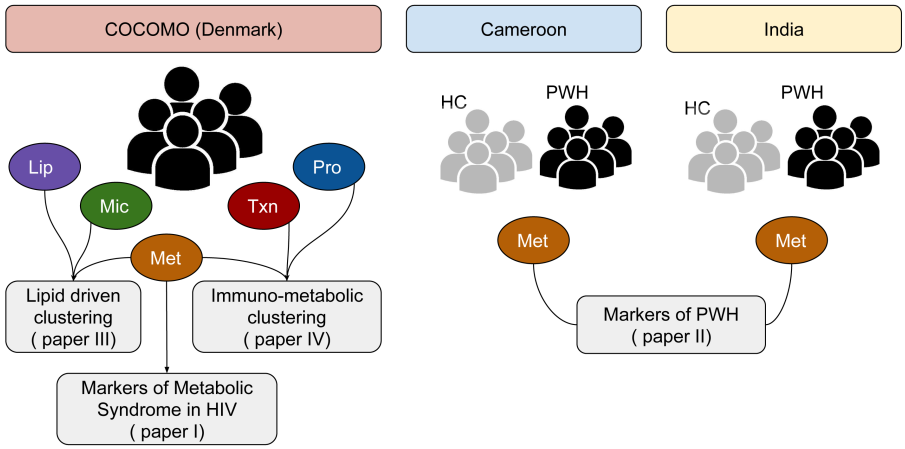


Figure 7: Summary aims and patients' materials of the thesis (papers I to IV)

4.1 Metabolic Perturbations in HIV

Since the start of ART, PWH on ART has been reported to display disrupted metabolism especially dysregulated glucose and lipid metabolism[120]. In our Danish cohort, we identified a cluster of PWH with highly upregulated lipids, including triglycerides (TAG), diglycerides (DAG), ceramides (CER), and presenting weight-related comorbidities including high BMI, high visceral adipose tissue (VAT) , high subcutaneous adipose tissue (SAT), and high proportion of metabolic syndromes (MetS) (**paper III**) associated with dyslipidemia. Dyslipidemia has already been associated with cardiovascular diseases and mortality in naive[121] and ART patients[60] in coherence with our results. Nevertheless, the lipid profile cannot alone explain the metabolic perturbations in HIV (**paper III**). In this thesis, we have explored the metabolic shifts linked to treatment regimens and latent reservoirs in patients in different cohorts and the consequences of these perturbations.

4.1.1 Amino acids dysregulation in HIV

In our cohorts under long-term successful ART, amino acid metabolism was disrupted and associated with weight-related comorbidities (**papers I to IV**). Previously, amino acids and fatty acid metabolism were found dysregulated in PWH before or after short-term treatment[122,123] and associated with late immune recovery[122] or HIV-associated neurocognitive disorders (HAND)[124]. The increase of several amino acids (43 % of total dysregulation) involved in methionine, valine and tyrosine degradation, as well as the sirtuin signaling pathway was observed in PWH with MetS compared to PWH without Mets with similar ART duration and CD4 count (**paper I**, Figure 8).

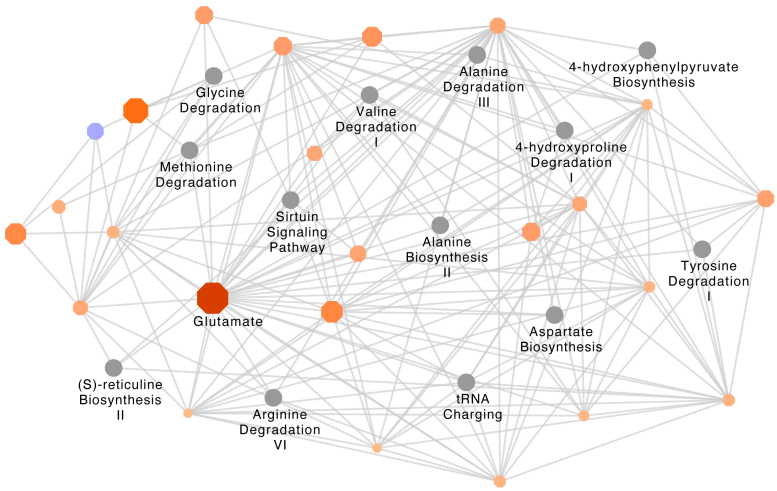


Figure 8: Network of the top significant pathways and their associated metabolites differing from PWH and PWH+MetS. Node color and size are proportional to the log2foldchange between PWH and PWH+MetS. Red indicates increase in PWH with MetS and blue decrease,

The amino acids dysregulation in PWH compared to HC was observed in three independent cohorts from Denmark, Cameroon, and India based on untargeted[70] and targeted metabolomics with noticeable patients' heterogeneity (**paper II**, **paper III**). Among PWH, the regulation of amino acid reactions was even more complex. While patients with a high fat profile had the same metabolic profile as HC-like fat-based profile patients (**paper III**), patients with dysregulated immune-metabolism profiles displayed amino acid dysregulation (**paper IV**). The amino acids pathway is a crucial pathway dysregulated with HIV and metabolic disorders, as observed in several cohorts in PWH on long-term ART.

4.1.2 Glutaminolysis disruption in PWH

Among metabolic pathways, glycolysis and glutaminolysis are essential for HIV replication[122]. The production of ATP through glycolysis is essential to produce new viruses and maintain infectivity[125]. Glutamate has also been shown to be preferentially used in macrophages of PWH[126]. Glutamine is converted to glutamate in the mitochondria and

then converted to alpha-ketoglutarate, which is used in the citric acid cycle which will be used to produce biosynthetic precursors[127]. Moreover, late immune recovery following cART results from high glutamate levels[122].

In PWH with MetS, we observed a drastic increase in glutamate and several metabolites from glutamate metabolism (**paper I,** Figure 8). The association between MetS and glutamate has been shown in non-infected individuals[128]. Glutamate was increased in PWH compared to HC in the Cameroon cohort (**paper II**) and in the targeted metabolomics of the Indian cohort as well as in the risk groups compared to HC-like groups in the Danish cohort (**papers III and IV**) (Figure 9). This showed a significant increase of glutamate with HIV, MetS, and severity and a potential shift in metabolism toward glutaminolysis. However, the shift in glutaminolysis did not cause accelerated biological aging in PWH (**paper V**), as no differences were observed between the groups.

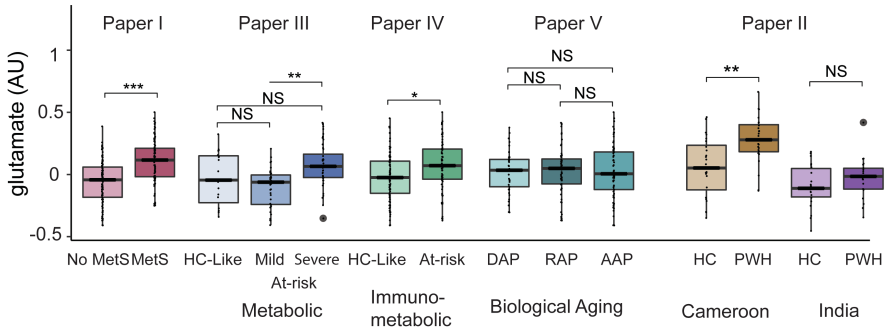


Figure 9: Glutamate levels in different papers and patient groups. Cohorts for Cameroon and India are indicated in parenthesis. Other samples are from the Denmark cohort.

We inhibited glutaminolysis in HIV cell models to see the effects of latent reservoirs. During ART, HIV remains quiescent in latent reservoirs of lymphocytic and monocytic cell lines. A potential cure for HIV is the activation and consequent clearing of this latent reservoirs[129]. In latency cell models J-lat and U1, we blocked glycolysis (2-DG) and glutaminolysis (DON). The treatment with DON activated the latent reservoirs in U1 but not J-Lat cells (**paper II**). The treatment with 2-DG did not activate latency reversal. Proteomics analysis on U1 treated with DON revealed proteins involved in glycolysis, TCA cycle, sulfur metabolism, amino acids, and OXPHOS. OXPHOS complex I to IV proteins were lower in U1+DON compared to U1 in proteomics and western blot quantification (**paper II**). No difference was observed in ROS, which indicates unbalanced redox homeostasis. Then, we tested the influence of glutaminolysis inhibition during cART for latency reversal but did not see a difference between the ART regimens. DON decreased glutamate and lactate and increased glucose in all ART treatments (**paper II**).

In conclusion, living with HIV increases OXPHOS and glycolysis and in glutamate/glutamine as an energy source. Inhibition of glutaminolysis could solve latency reversal in monocytic cell lines.

4.1.3 Glutamate-induced neurotoxicity

The higher abundance of glutamate is known to be harmful to the brain. Glutamate is the major excitatory neurotransmitter involved in many brain functions, such as learning, memory, and pain[130]. It has been shown that glutamate-induced neurotoxicity can be induced by glutamatergic system dysregulation, resulting in neurodegenerative disorders[131]. On the other hand, lactate and pyruvate have been shown in animal models to decrease glutamate-induced neurotoxicity[132].

In **paper II,** we identified five neurosteroids differing from HC and PWH, potential markers of neurocognitive impairments. Both cohorts showed a lower abundance of these biomarkers in PWH except for methionine sulfone, which displayed a higher abundance in PWH. A decrease in neurosteroids has been associated with depression previously[133]. We believe that patients from this study are displaying glutamate toxicity and potential future cognitive impairments. The increase in lactate and pyruvate observed in PWH with MetS (**paper I**) could result from the increase of glutamate and therefore have a protective effect and avoid neurological impairment in these patients, which was not observed in the Cameroon and India cohorts (**paper II,** Figure 10).

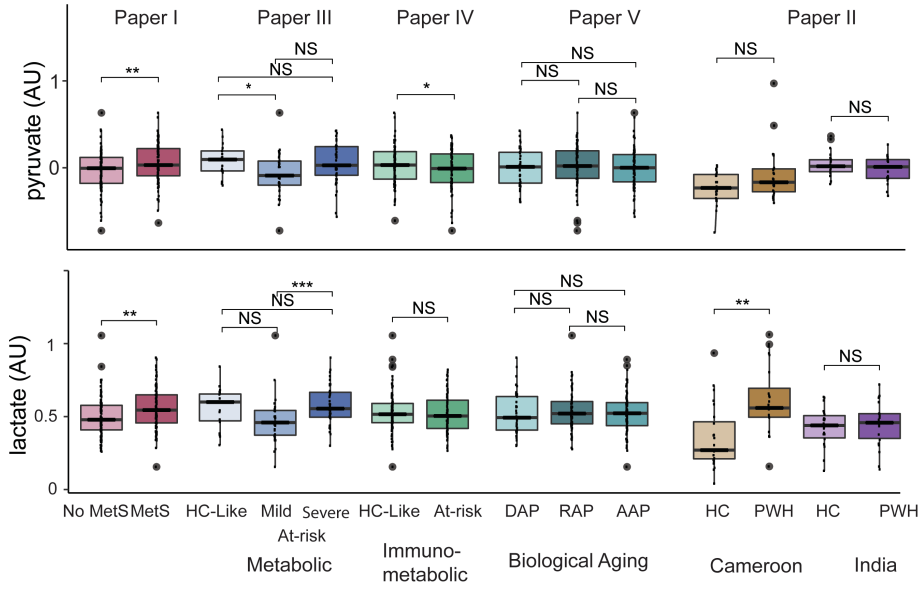


Figure 10: Pyruvate and lactate levels in different papers and cohorts.

4.1.4 Central carbon metabolism

We identified a general dysregulation of the central carbon metabolism (CCM) in HIV (**papers I, III, IV**). The CCM includes the TCA cycle, glutamate, glycolysis, and pyruvate metabolism[134]. In addition to disrupted glutamate metabolism associated with MetS in PWH, other metabolites from the CCM were increased in the blood of PWH with MetS, including pyruvate, lactate, and α-ketoglutarate levels but not glucose (**paper I**, Figure 8). This indicates the release of glycolytic and TCA metabolites in the plasma. Pyruvate is

usually converted to acetyl-coA but, with disrupted metabolism, mainly changes to lactate[127].

The major transporters of CCM, Glut1 (glucose), MCT-1 (pyruvate and lactate), and xCT (exchange glutamate/cystine) were measured in T cells and different monocytes populations of PWH with MetS and without MetS. MCT-1 was higher in PWH with MetS compared to PWH in all cell types. xCt was higher in monocytes and CD4 T cells but not CD8 T cells showing a critical transport of metabolites from CCM in monocytic cell lines (**paper I**). Patients with immunometabolic at-risk profile presented a metabolic dysregulation of the central carbon metabolism driven by senescence associated proteins compared to those with the at-risk profile. Also, using GSMM and FBA, we identified 64 flux specific to the at-risk group, including transport reactions, pentose phosphate pathway, and fatty acid oxidation, which are part of central carbon metabolism (**paper IV**). In addition to glutaminolysis, the whole CCM is dysregulated in PWH and an essential marker of severity in HIV.

4.1.5 Microbiome associated metabolites

We observed difficulty in proving the importance of the microbiome on the severity in PWH but saw an evident dysregulation of microbiome-associated metabolites (MAM) (**paper III**). The gut microorganisms process elements from diet and endogenous compounds into MAM. These have been associated with diseases, including metabolic disorders[135]. In our **paper III**, we showed that HIV patients display an essential difference in the profile of MAM, especially indole derivatives, compared to controls. The variability between PWH clusters was more complex but deserved further attention (**papers III and IV**). Also, MAM was highly intercorrelated in PWH, particularly secondary bile acid metabolites (**paper III**). Most of the MAM differing between clusters were linked to the lifestyle and medication of patients (**paper III**, Figure 11).

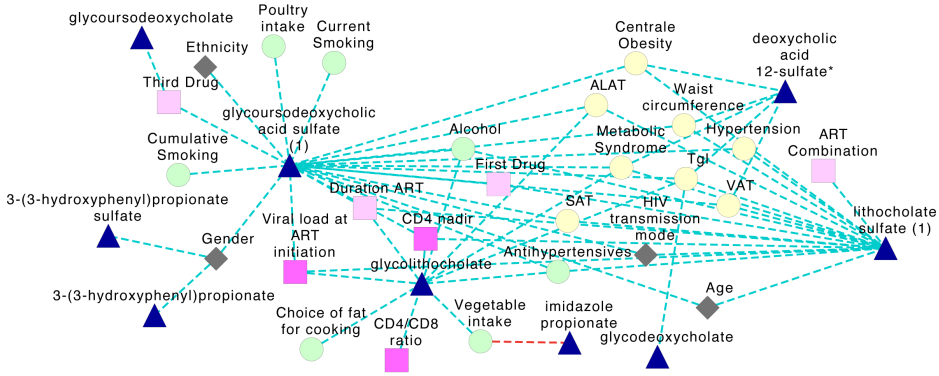


Figure 11: Microbiome-associated metabolites are highly associated with clinical and life parameters. Bluedotted lines indicates negative and red positive association. Metabolites are represented as blue triangles, food and life-style habits as green circles, weight related comorbidities as yellow circles, ART related parameters as pink squares and other paramters as grey diamonds.

In the groups driven by immune metabolic profile, three crucial MAM (serotonin, kynurenate, quinolinate) were decreased in the HC-like compared to the risk group. Taurine, hydroxybutyrate, and acetoacetate were more abundant and linked with liver injury, which was confirmed by the high alanine aminotransferase (ALAT) level in the at-risk group (**paper IV**).

In conclusion, metabolic perturbation in HIV involves lipids, amino acids, particularly glutaminolysis, central carbon metabolism, and microbiome-associated metabolites. This dysregulation is responsible for aging-related comorbidities, including metabolic disorders and neurological diseases. Glutaminolysis inhibition could be a potential HIV remission strategy targeting the latent reservoirs.

4.2 Immuno-Metabolism with HIV

Despite ART treatment, many PWHs suffer from chronic inflammation and immune dysregulation linked to aging and age-related comorbidities[136]**.** We showed that PWH suffering from transcriptomics aging have increased inflammation (**paper V**). More treatment of PBMCs with at-risk plasma activated the inflammatory response according to proteomics analysis (**paper IV**). My thesis aims to understand better immune cell activation and immune system dysregulation in PWH.

4.2.1 Immune cells impairment in HIV

HIV infection induces the death of T-cell helper and the progressive decline of immune function. After ART, the CD4 count increases, and the CD4/CD8 ratio indicates disease activity. In **paper IV**, we showed that at-risk patients had a lower CD4/CD8 ratio in at-risk compared to HC-like, which could be due to higher HIV activity or other comorbidities[137]. Interestingly, there were no differences in markers of inhibitory checkpoints, transcription, function, and phenotype on memory CD4+ and CD8+ T cells exposed to HC-like and at-risk plasma, indicating no change in T cell activity *in vitro*. Moreover, a higher proportion of CD8 T cells and a lower of CD4 T cells were observed using cell profiling from transcriptomics between HC-like and at-risk but not monocytes.

On the other hand, monocytes are used as a latent reservoir by the virus and present activation with the virus[138] . We saw an increase of CCR2 and CX3CR1 in monocytes treated with at-risk plasma compared to HC-like treated cells indicating their activation, while the high CD86 and low CD38 and PDL1 indicate impaired macrophage function due to monocyte exhaustion (**paper IV**, Figure 12).

This was confirmed by pathway analysis in the proteomics data, where we observed increase of inflammation and down-regulation of monocyte migration, activation, and differentiation. Interestingly, monocyte lineages expressed senescence-associated proteins differing significantly between HC-like and at-risk (**paper IV**).

These results indicate no modifications and successful recovery of T-cell function with ART but a potential disruption of monocyte function by exhaustion which could explain the risk profile in HIV.

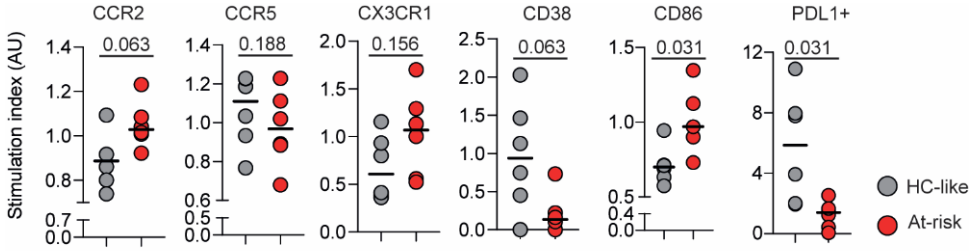


Figure 12: Dotplot showing the expression of phenotypic and functional markers on monocytes

4.2.2 Polyamines

Polyamines are also dysregulated in our studies (**paper IV**). Polyamines are low molecular weight compounds with at least two amino groups. They have an influence on several metabolic processes, including inflammation and immunity[139]**.** The most important polyamines of the cells are spermine, spermidine and putrescine. Spermidine is produced from spermine by the enzyme spermine oxidase. In PWH, polyamines could have a role or be a biomarker of neurologic disorders[140] and T cell dysfunction[141]. An increase in spermine and spermidine which could correlate with the disease severity was seen in the at-risk compared to HC-like patients,. Moreover, we wanted to investigate the role of high polyamides on monocyte function by treating monocytes from donors with spermine and spermidine and simulating with LPS (**paper IV**). The proteomics analysis revealed that in spermidine-treated cells, proteins regulating metabolic, pro-inflammatory, and chemokine signaling were more abundant compared to controls, and proteins involved in monocyte functions were downregulated (**paper IV**, Figure 13).

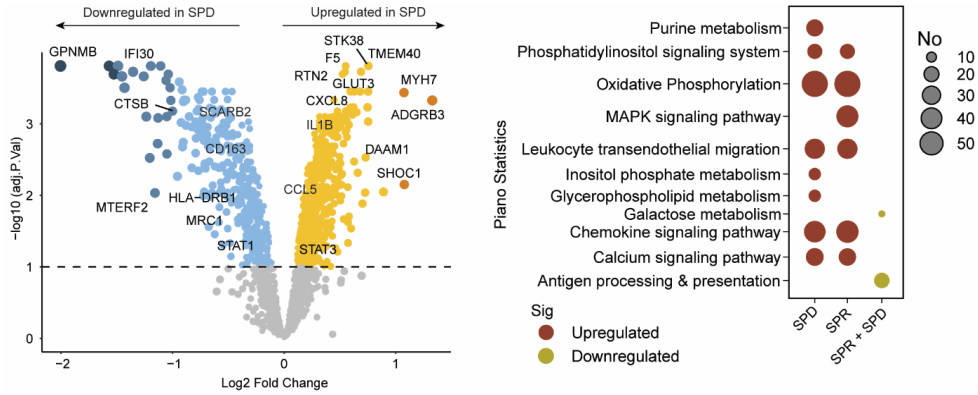


Figure 13: Volcano plot (right) and pathway analysis (left) from comparison of proteins between PBMcs vs PBMCs treated with spermidine

4.3 Multi-omics data analysis

Omics integration aims to answer a biological question by finding biological patterns which are not apparent in single omics. In my thesis, we explored the information given by different omics and combinations of omics.

4.3.1 Single omics and combination of methods

We used ConsensusClusterplus[105] for single omics analysis because of the resampling, tunning parameters, quality metrics, and functionality to identify the optimal clustering (**paper II**). To identify even a lower number of biomarkers while reducing the risk of false positives, we used a combination of three totally different methodologies of feature selection in papers **I, II and IV**.

It has been shown that biological pathways can be modeled by causal dependency structures[142,143]. First step was to select features (here genes) strongly differing two conditions; we used an RF classifier. Then, we built a directed acyclic draph using a naïve Bayes algorithm[99] which modelized the proteome interactions of interest. We disrupted the network by removing nodes randomly and identified five driver genes with a major impact on the network and "driving" the pathology. We found that the set of genes was relevant biologically: they were all involved in stress-related dysregulated metabolism (**paper IV**), showing the robustness of the approach. Indeed, Single omics analysis can lead to incomplete results, as in the Cameroon India study (**paper II**). Multiple omics (paper III, paper IV) were a tremendous asset.

4.3.2 Multi-omics Integration tools choices

For multi-omics integration, we have used tools able to handle different types of data: continuous (proteomics, metabolomics, lipidomics, clinical), categorical (clinical), and counts (16S, transcriptomics).

4.3.2.1 Comparison of SNF and MOFA

Several advanced tools are available for patient clustering[144,145]. In **papers III and IV,** we grouped patients using network analysis SNF[45], which fused individual networks into a summary of all information. The original SNF algorithm returns patients labeled based on the most optimal clustering based on matrices of data (usually a large number of features) but if the addition of clinical data is mathematically possible, it was not implemented. Then, we used netDx[106], a tool built on SNF which integrates the clinical parameters to each cluster (**paper IV**). On the other hand, this functionality was already implemented in MOFA[30]. Factors were correlated with clinical parameters by converting categorical variables into numeric variables. Also, in MOFA but not in SNF, each feature had a "weight" which could be interpreted as the importance of the feature in the global model or in the factor. We used them to extract the most informative biomarkers related to clusters (**paper III**). Both MOFA and SNF provide the importance of each omic layer on the final

integration model (**papers III and IV**). MOFA managed the integration of different data types but was not able at all to exploit microbiome data while SNF could (**paper III**). MOFA has been used to integrate microbiomes with fungal data on the same scale[146].The main difficulty in our project was integrating 16S data with a low number of operational taxonomic unit (OTUs) (N = 241) and presenting several missing data with intercorrelated mass spectrometry data (metabolomics and lipidomics data) (**paper III**). A solution could be to reduce the number of features challenging lipidomics because they were highly inter-correlated. Another possibility would be to include whole genome sequencing data[147]. In general, SNF performed better than MOFA because it could combine the distances from each data more efficiently. A further analysis would be to perform a consensus SNF.

4.3.2.2 Strengths and limitations of co-expression networks

Large-scale biological networks are complex structures and cannot always be efficiently described using global distribution. Still, they can present patterns of interconnections[148]. An efficient way to analyze networks is to select local modules with individual topology and biological functions[36]. Specific interactions within modules and between modules can be investigated with much more efficiency than with a global network. In co-expression networks, the weight of the edge separating two nodes is their coefficient of correlation. The advantages of co-expression networks are that we can integrate many data types, reduce the number of nodes by selecting the most correlated features, use statistics and graph measures on much smaller and coherent modules, and identify consequently independent biological processes.

To identify groups of co-expressed features, we made association analysis and community detection using the Leiden algorithm (**paper I, II, III,** and **IV**). We mainly used connectivity (represented by the degree) at the node and community levels. We calculated the centrality for each community to determine the most central community with biological importance in the network. We used differential abundance analysis and pathway analysis to determine the biological processes in the communities. For example, in **paper II**, the most central community was majority lipids, but the neurosteroid biomarkers were in another community with most metabolites downregulated.

The limitations of the co-expression network were mainly linked to the omics type, the relevance of correlations, optimal cut-offs, and time consumption. First, omics produced from the same batch tend to correlate together. While combining different omics, we saw an omics-type effect in **paper III,** with lipids, metabolites, and microbes segregated in separated communities (Figure 14) but not in **paper IV**. Secondly, correlated features do not always mean they are on the same pathway. Third, the correlation cutoff is arbitrary and can lead to different networks and communities. Then, the more we had omics, the more the computation time can be important for large networks. We performed a consensus co-expression network in paper III to reduce this time.

An advance for community detection would be more biological information about nodes as observed in GSMMs (described further). The whole field of network biology could need to develop temporal and spatial dimensions of the networks.

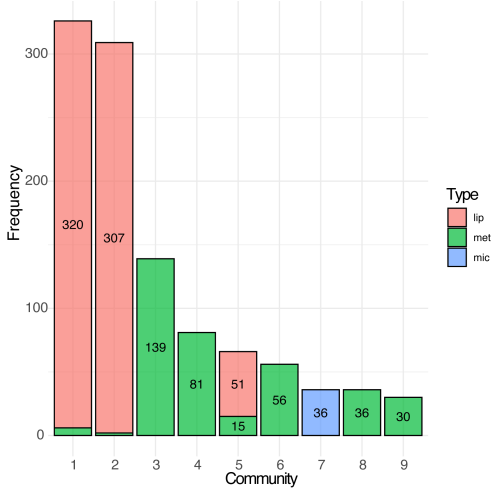


Figure 14: Omics distribution per community in paper III

4.3.2.3 Advantages and limitations of Genome-scale metabolic models

GSMM integrates omics into fluxes of metabolic reactions that we can quantify[149]. Models are submitted to constraints defined mathematically as optimizing the objective function here, the cell growth. Compared to classic network approaches, which are primarily unsupervised, the main strength of GSMM is to be able to use existing models. They summarize all the information about the organism metabolism especially biochemical knowledge, including kinetics constraints, fluxes, directions and reversibility of fluxes, catalyzing enzymes, and reaction rates manually curated and validated[150]. We were able to add our constraints. Transcriptomics data allows the application of regulatory constraints to the model[151], while adding metabolomics data constrains the metabolic rates across the membrane[152]. Models can also be disrupted by removing genes or metabolites or lowering metabolite concentration (ex: O2). We integrated transcriptomics into genome-scale metabolic models and performed flux balance analysis (FBA) to obtain flux and metabolic pathways most indicative of the metabolic state (**paper IV**). In a heterogeneous population with metabolic disruptions, performing a single patient genome-scale helped us to include each patient's difference and characterize the spectrum of metabolic changes[153]. A clear increase in metabolic fluxes in almost all the at-risk patients matched with the grouped GSMM (**paper IV**).

Still, GSMM has several limitations. First, we are making strong assumptions about the kinetics of the model. Then GSMM was designed to model the growth of one cell, which we changed to a whole organism while we know that cells can be at a different stage. Also, some fluxes or data do not have sufficient annotations. We performed gene essentiality

and identified an essential gene in most at-risk patients with unknown functions. Then, each constructed model needs to be validated and tested in the lab. Also, alternative modeling techniques could be an asset to define better fluxes and metabolic perturbations, including modified FBA, analysis, or deletion strain phenotypes[151]. For example, parsimonious enzyme usage FBA classifies genes based on the importance of optimal growth[153].

On the other hand, flux variability analysis is applied to find several optimal flux distributions solving the problem under different optimal functions. It is useful to investigate suboptimal growth or alternative optima[151,154]. Also, machine learning approaches for the reconstruction of GSSM have been investigated and could be applied in our projects[155].

In conclusion, we selected SNF, MOFA, GSSMs, and co-expression networks for their properties and addressed their limitations using other tools or laboratory experiments. We could have applied DIABLO, which is also strong in multi-omics analaysis[156].

4.4 Phenotypic analysis

Patients' phenotype was a constant driver of the omics-based clusters' qualification into at-risk or HC-like groups. More, in the COCOMO cohort, the abundance and quality of the clinical data allowed us to provide a detailed and precise definition of patients and groups of patients and investigate the comorbidities, risk factors, interactions, and co-founding effects (**papers I, III,** and **IV**). The phenotypes of patients are presented in Figure 15.

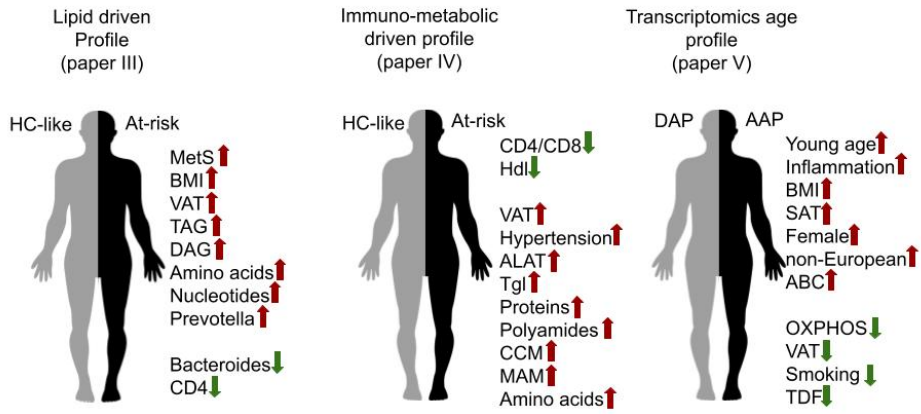


Figure 15: Patient phenotypes summary for each clustering. Red arrows are representing an increase and green a decrease.

4.4.1 Metabolic syndrome and weight related comorbidities

Our patients were selected to study metabolic syndrome, so we have multiple clinical data to investigate these comorbidities (**paper I, III,** and **IV**). The definition of metabolic syndrome describes the combination of cardiovascular comorbidities, including insulin

resistance, hypertension, obesity, and dyslipidemia. Patients with metabolic syndrome are more at risk of developing CVDs and type II diabetes[157].

In our **papers III and IV**, we decided that the definition of metabolic syndrome was not representative of the patient's heterogeneity in terms of weight-related comorbidities. In **paper III**, we used lipidomics data to separate patients and saw that severe at-risk patients displayed high BMI, SAT, VAT, hypertension, and obesity (Figure 15). In **paper IV**, the at-risk group displayed a higher SAT but not VAT and no difference in BMI. Interestingly, VAT is supposed to be more pathogenic than SAT[158]. Still, they had a higher level of TAG, high-density lipoprotein, central obesity and cytosolic blood pressure compared to the HC-like group indicating a less severe metabolic profile than patients from **paper III** but still, apparent immunometabolic perturbations identified by low CD4/CD8 and high ALAT which is indicating kidney disfunction[159] (Figure 15).

On the other hand, fat distribution is influenced by gender. It has been reported that men have more VAT giving the apple shape, and female SAT, giving the pear shape coherent to our data[160]. Unfortunately, the cohort contains primarily males (90 %). Still, in our cohort (**paper V**), the accelerated aging group presented a higher SAT and a lower VAT than the other aging groups. This could be explained by the presence of almost all females (17/19) in the accelerated aging group (Figure 15).

4.4.2 Comparison of clusters identified in III, IV and V

Different data inputs and patients gave us clusters with clinical profiles (Figure 15). We should investigate these differences to successfully define the heterogeneity in PWH.

Patients from the same cohort (Denmark) were clustered according to data (lipidomics, metabolomics and microbiome in **paper III**, proteomics, transcriptomics and metabolomics in **paper IV**) and transcriptomics age (**paper V**). First, all patients were not available for all types of data (Figure 16A). In **paper III**, the clustering was mostly driven by lipids while in **paper IV,** was driven by proteomics profile. More, we concluded that patients grouping in **paper III** was due to lipid and metabolic profile while clustering in **paper IV** was representing the immuno-metabolic perturbations and immune dysregulation. Still, low overlap was observed between clusters (Figure 16B) and risk-groups (Figure 16C). Only nine patients were found in the risk group in all studies. Interestingly patients from an immune dysregulated profile (at-risk in **paper IV**) displayed a higher transcriptomics age than HC-like but no difference in delta age, showing a potential dysregulated aging in this population.

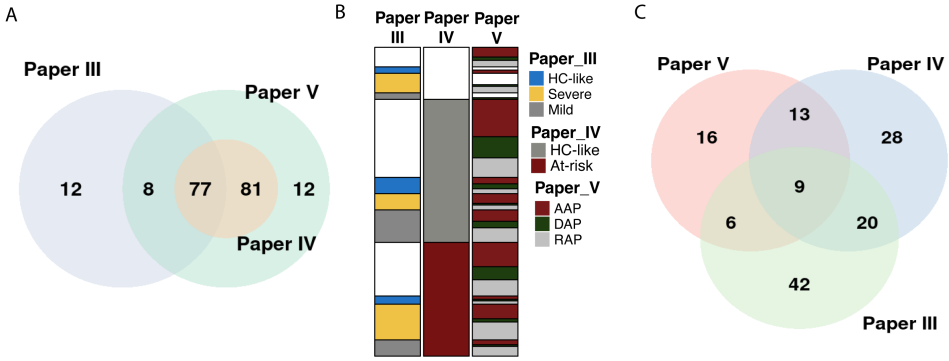


Figure 16: a) Overlap of patients between different studies b) Overlap of clusters c) Overlap of risk cluster between studies.

4.4.3 Interactions and confounding effects in omics studies

In epidemiology, we differentiate mediation, joint effect, and co-funding effect. The mediation effect is observed when the association between independent variables on the dependent variable operates through one or several mediators. In interaction, there is a joint action between the two variables where the effect of one variable depends on the other[161]. Conversely, the confounding effect represents a misleading association and must be corrected in the model to avoid bias when interpreting the results[162].

Different confounding effects were addressed in our papers. First, they were identified by comparing all clinical parameters from the study groups using basic statistics (Mann-Whitney U test and Chi-square Test). In **paper I**, we performed a univariate linear regression between metabolites and metabolic syndrome status corrected for BMI and treatment regimen. The aim was to identify biomarkers of metabolic syndrome in HIV, not simply an obesity marker. Biomarkers were robust to the correction except for one. In **paper II**, we corrected for exercise while comparing HC and PWH in the Cameroon cohort. In **paper III**, the correction was applied for HIV transmission mode and CD4 count and finally, in **paper IV,** for ethnicity and beef intake. For **papers IV and III** where we compare HIV infected patients, we could see a clear difference in results between the corrected and not corrected models. In **paper V**, we tried to identify interactions and mediation effects on the association of clinical parameters with accelerated aging. Still, nothing was significant, probably due to the insufficient number of patients from the cohort.

4.4.4 Influence of Microbiome in HIV

In this thesis, we found that the microbiome had a common effect on fat metabolism (**paper III**), and no effect on immune metabolism (**paper IV**). More, in **paper III**, the severe at-risk group displayed high levels of *Prevotella* and low levels of *Bacteroides,* which have been linked to MSM behaviors in several studies[163,164]. This indicates that for this group, the confounding effects of sexual orientation on the association between microbiome and the high lipid profile and potential weight-related comorbidities in HIV. Surprinsingly, the

mild at-risk PWH displayed similar microbiome composition as the severe at-risk group. No significant difference in the pourcentage of MSM between the mild and the HC-like cluster was observed. Still, HC-like group had a totally different microbiome profile. An investigation in a larger cohort with a larger proportion of women and heterosexual patients could be conducted to identify a potential link between metabolic perturbation and microbiome in HIV.

4.5 Aging Processes in HIV

All patients in this thesis have been in long-term therapy and represent an aging population living with the virus. They are experiencing a mix of natural aging processes with immune aging due to latent HIV and drug toxicity[165].

4.5.1 Increase senescence in part of HIV patients

Senescent cells stop multiplying and release molecules, including pro-inflammatory cytokines, proteases, immune modulators, growth factors, and, described as Senescence-associated secretory phenotype (SASP)[166]. Cell senescence leads to the activation of monocyte–macrophage lineage cells and chronic low-grade inflammation[14]. An increase in senescence has been involved in aging-related diseases[167] and HIV[168]. Senescent proteins were more abundant in the at-risk group (**paper IV,** Figure 17).

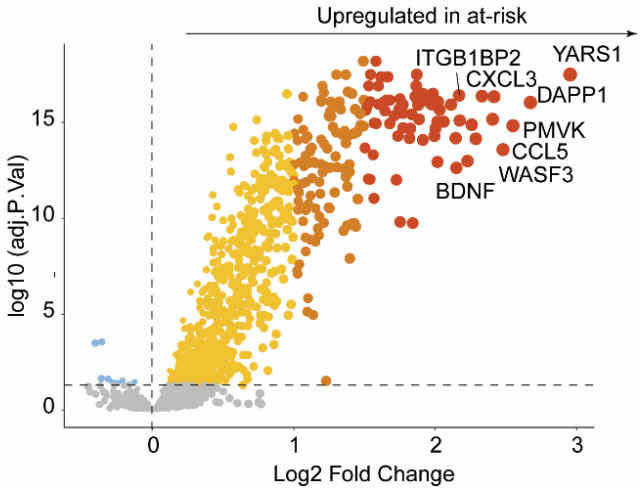


Figure 17: Volcano plot based representing differential abundance between HC-like and at-risk clusters

Moreover, they were mainly in the most central community of the co-expression network, indicating an essential modulatory role in the metabolic dysregulation in PWH. Moreover, these senescent proteins differing risk groups were detected in the myeloid lineage cells [classical monocytes, nonclassical monocytes, and dendritic cells] (**paper IV**).

4.5.2 Accelerated aging with HIV

In opposition to chronological age, biological aging is described as the underlying aging mechanisms in the body and can be measured using telomeres length, methylation, or transcriptomics. Accelerated aging, when biological age is higher than chronological age, has been associated with disease[169]. PWH with metabolic syndrome displayed pathways linked to aging, including the sirtuin signaling pathway (**paper I**). To better understand the biological aging in PWH, we benchmarked biological clocks trained on transcriptomics data from HC patients (**paper V**) by applying it to our patients. We identified that accelerated aging was influenced by a complex interaction of ART regimen, gender, ethnicity and smoking (**paper V**, Figure 18).

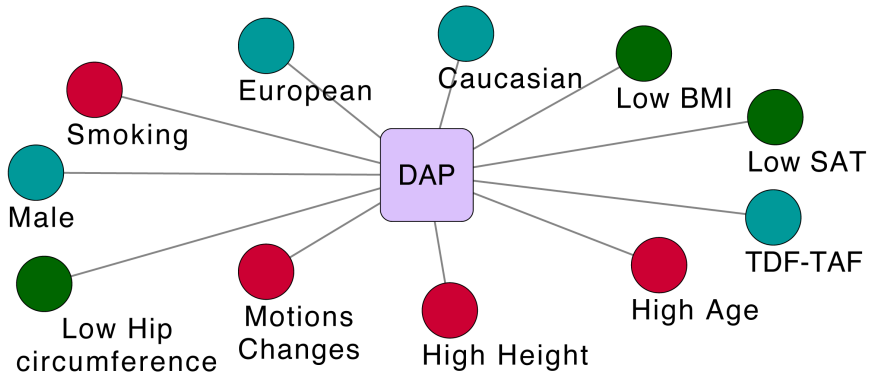


Figure 18: Summary figure of important factors linked to decreasing aging (DAP) in PWH. Red represents increase and green decrease. Blue indicates categorical parameters.

PWH from the COCOMO cohort with highly abundant senescent proteins display an at-risk profile, while low senescence represented HC-like. More, accelerated aging was present in some patients, which does not necessarily overlap with the at-risk immune profile.

5 Conclusions and perspectives

The development of age-related diseases in PWH has been reported and documented, but the immune-aging mechanisms linked to HIV are still challenging for scientists in the field. The recent omics revolution has been crucial for understanding many diseases but is not fully developed in PWH due to the complexity of samples handling. In this thesis, we have applied and tested advanced bioinformatics tools and system biology to understand better immune aging in HIV.

With this thesis, we could have a better understanding of the mechanisms of immune aging in HIV. A comparison of patients from three independent cohorts indicated a global increase of amino acids and particularly glutamine and glutamate metabolism in PWH compared to HC (**paper II**), even more, accentuated with metabolic syndrome (**paper I**) and severity (**paper III** and **paper IV**) (Figure 9). Increased glutamate can lead to glutamate-induced cytotoxicity in the brain and potential neurocognitive diseases (**paper II**). A decrease in pyruvate and lactate could be a protective mechanism in the Denmark cohort compared to Cameroon and Indian cohorts (Figure 10). An increase of lipids, especially tri-and diglycerides, was observed in PWH highly correlated with at-risk severity profile (**paper III**) and weight-related comorbidities (Figure 19).

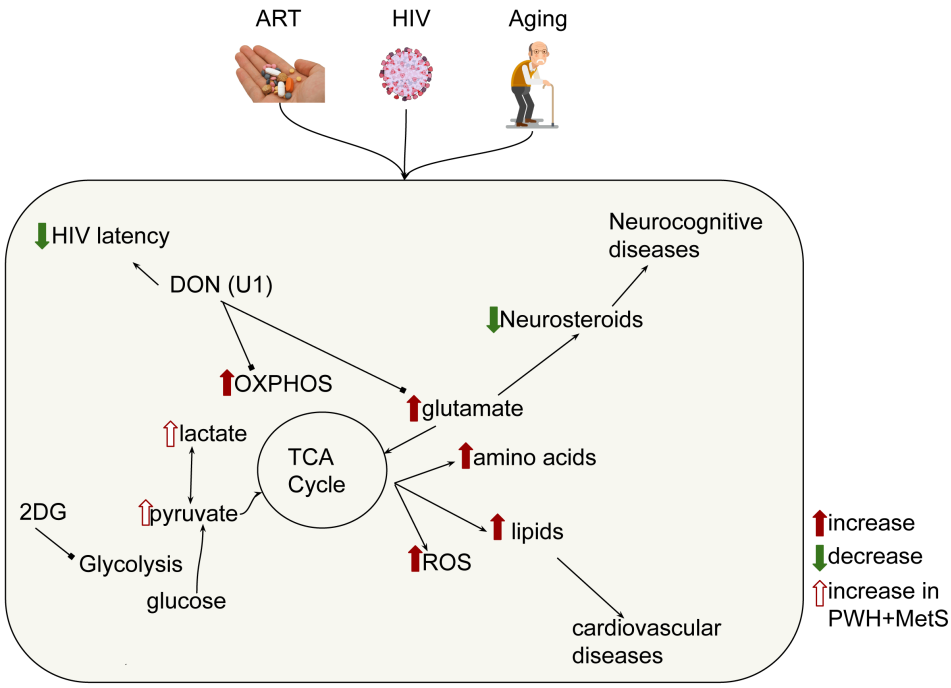


Figure 19: Summary mechanisms investigated in the thesis

Other immune-aging drivers are the potential monocyte exhaustion, senescence protein production, and increased inflammatory markers linked to polyamides and ROS production (**paper IV**). Increased transcriptomics aging is associated with the general process (**paper V**), but more investigations are needed.

The principal limitations of our studies were linked to the patients' heterogeneity, the study type and the number of participants, and the bias of omics technologies.
First, patients are highly heterogeneous compared to cells or animal models. People display different genetic backgrounds, lifestyle habits and age[170]. Especially when studying aging, the aging process is highly dependent on an individual, but also in the case of PWH, ART regimen (**paper V**). If COCOMO patients reported numerous clinical parameters, Cameroon and India's cohort had a limited number. Clinical parameters significantly impact our analysis, such as in **paper V,** where several dependent variables were highly correlated. Still, we had a consequent number of missing values in the COCOMO database. Also, exercise differs between HC and PWH on ART, which is known to influence the number of immune cells and was corrected in the model. Also, we had an under-representation of females though 54% of PWH worldwide are women (https://www.unaids.org/, 2021). Patients from several origins (Denmark, India, Cameroon) were included to try to address the differences linked to origin and ethnicity. Still, more cohorts, especially from LMICS, will be necessary to validate our findings. Finally, the definition of metabolic syndrome (**paper I**) did not fully represent the metabolic perturbations spectrum observed in PWH. This limitation was addressed in **papers III and IV** by removing this definition and considering all weight-related parameters. Also, the characteristics of clusters that we observed in **papers III and IV** need to be validated in other cohorts.
Secondly, producing omics samples, especially for HIV samples, is costly and challenging[171]. Also, the curse of dimensionality (low number of samples and high number of features) particularly impacts multi-omics settings. We did only cross-sectional studies instead of longitudinal studies for this thesis. The Cameroon and Indian cohorts were relatively small, which limited our analysis. On the other hand, the COCOMO cohort is one of the largest cohorts reported, which was a strength of our study. The large size allowed us to investigate differences among PWH, which was impossible in small cohorts due to statistical limitations. Also, in the COCOMO study, only 20 uninfected individuals were included and were not matched to the PWH. We did not include them in all the group comparisons, but they were kept as a reference. Still, HC clustering gave us relevant information in microbiome analysis (**paper III**) and parameterizing the transcriptomics aging clock (**paper V**). Regarding this project, knowing if people are true, HC is always challenging as we do not have high numbers. For example, HC displayed higher BMI than HC-like PWH patients.

Finally, my main work was manage limitations of omics analysis. Most challenges of omics per type were described in the introduction. In general, omics data can be influenced by extraction time, sample processing, state of metabolism, and food intake. Also, technical differences between the omics can arise, such as feature stability or coverage[172]. For example, we used 16S rRNA for microbiome analysis which has a lot of missing data for the genus. Perhaps doing the whole genome would have given more information. Nevertheless, we used microbiome-associated metabolites to measure microbiome impact. Regarding computational tools, still, for metabolomics and proteomics, there is no clear pipeline as in transcriptomics or 16S microbiome. In the bioinformatics analysis, normalization and choice of the statistics had an impact on the results. Doing statistics, we also had the risk of false positives that we controlled with the FDR. Also, results from bioinformatics were validated in the lab using different cell models, western blot, and FACS. Still, more lab experiments would be relevant to confirm our theories and validate the theories we are developing.

Further work is necessary to understand immune aging in HIV and potential therapeutic strategies.

One of the remaining limitations in the field is the number of samples and omics data available. A recent study with many patients (N>1500) and omics types is currently designed to identify biomarkers and pathways of non-AIDS comorbidities by comparing PWH to HC[173]. The clinical parameters are well-detailed, including follow-up, psychological evaluation, and extreme phenotypes. Patients have been in therapy for more than ten years, but the overrepresentation of European, middle-aged men is still observed. Also, omics technologies are improving in precision. We observed the Olink proteomics (**paper IV**) had many proteins (N=3000), was more precise than mass spectrometry, and was highly informative for the study. Moreover, since I started my thesis, we have observed a fantastic development of single-cell and omics technologies, including transcriptomics, proteomics, and metabolomics at the single-cell level[174]. Developing single cells models of PWH could quantify and compare immune cell types and their interactions. With our data, we performed digital cell quantification in bulk PBMCs. We identified potential proportions of cell types but found differences from expected cell proportions. Also, in the project, we mostly worked on blood samples with molecules from different tissues (for example, metabolites). The development of omics technologies in tissues has been done[175] and would be highly relevant, for example, in fat tissue or tissue organoids, to understand the mechanisms of immune aging.

Potential cures developed from this thesis could be the inhibition of glutaminolysis first in animal models and potentially in patients to induce latency reversal. Based on cluster analysis and aging groups, patients at risk should benefit from a personalized follow-up and adapted lifestyle to not develop aging-related comorbidities. Also, from the aging

study, TDF/TAF drugs were associated with normal and decreasing aging compared to ABC and could be preferentially given to patients.

In general, we improved the understanding of immune aging in PWH and developed combinations of bioinformatics and system biology tools which could be relevant for further research to improve the care of PWH.

6 Acknowledgments

My thesis was an exciting and enriching journey where I learned a lot, struggled, developed myself, and, most importantly, confirmed my passion for bioinformatics. Shifting from genetics to bioinformatics was challenging but brought me so much and now I really know that I have amazing tools to pursue my career. I would not have done so much without the people around me, so I would like to thank all of you for the support and strength you gave me.

First, I would like to thank my main supervisor **Ujjwal Neogi**. Thank you for having trust in me and giving me this amazing opportunity to learn with you during this Ph.D. During the whole Ph.D., you showed me the right directions and was present at each step of the way. Your guidance and help have been invaluable. I think you are a brilliant scientist and I know you will go far and make amazing discoveries. I really admire your capacity to take all the challenges to yourself, your flexibility and your innovative ideas.

I also would like to thank my second supervisor **Rui Benfeitas** who showed me the best way to do bioinformatics and always guided me, corrected me, and taught me to be a strong and independent bioinformatician with much pedagogy.

Then I want to thank my third supervisor **Erin Gabriel** who helped me understand statistics and has been an important source of inspiration for me.

Then I would like to thank the people and former people from the group who I enjoyed the company and learned so much from. I would like to thank **Soham** for explaining to me immunology concepts. I would like to thank **Xi** for being such a nice support in our parallel PhD roads. I had so much fun climbing with you. I would like to thank **Sara, Alejandra, Sabrina** and **Negin** for doing experiments for the papers and so many nice discussions. **Ale**, it was really nice to do the conference (and the paella/ beach) at Valencia together. I want to thank **Shambhu, Duncan, Ashok, Haleh, Maike, Anoop, Shubha, Beatriz, and Stefanit** for the nice times over the years.
I want to thank also my colleagues at KI. Particularly I would like to thank **Marion**, who is a real friend now! Can't wait to go for the next hike. **Ashwathy,** it was super fun to learn bioinformatics side to side and start with you my journey at KI. I would like to thank **Robert** for the beers and crispy gossip**, Rafa** so enthusiastic in the climbing gym**, Laura, Elisa, Rekeya, Xiaoshan** and **Bai,** for the collaborations, fun moments, and nice conversations. I also miss former members, including **Noelia, Hissa, Lydia, and Alicia**. A big thanks to the girls in the office who made last year so fun **Anna, Lisa, Francesca** and **Giota**. Thank you **Luca** for the last minute proof reading of the thesis and your constant enthousiam.

I would like to thank the administration team at Karolinska including **Marita** and **Lisa** who are doing incredible work for us. Thank you, **Ann**, also for being patient and not stopping my doctoral studies when I was late (more than once) with my documents.
I would like to thank our collaborators from Copenhagen, especially the people I have communicated with **Susanne Dam Poulsen** and **Marco Gelpi** for their trust and collaboration.
I want to thank **Joao Pedro de Magalhaes** and people from the former Institute of Ageing & Chronic Disease at Liverpool for welcoming me into the lab and sharing with me the secrets of aging! Special thanks to **Cyril, Anais** and **Gustavo** who were so friendly. Still, I cannot skip lunch like the whole of you.

Then, I would like to thank my amazing friends here in Sweden. I learned about so many cultures and finally found out we are not so different: **Sara** I miss you and believe in you so much**, Valeria** so sweet and strong cant wait to go exploring the forest and camping with you**, Minh Duc** my partner in crime no Swedish song can resit us**, Guillaume** my chess partner**, Oskar** from the black parade you were amazing support during this PhD. Je veux aussi remercier mes amis géniaux de France! Laura, tu es si géniale et je suis si chanceuse de t'avoir comme meilleure amie. Je nous souhaite les plus belles tournées concerts de metal et destinations exotiques. Mes super copines du Magistère **Noémie, Fouley et Pauline** qui sont parties tout autour du monde et que j'ai hate de retrouver. Mes petits FABBLE, hâte pour ceux qui finissent leurs études, hate de faire des mega picnics avec vous **Boi, Lae-young, Lucy, Ashraf, Baptiste et Elisa**. Un merci special a Lae-young qui m'a sortie une magnifique couverture. Mes copines de soirée **Gwendoline, Lyna** et **Sophie** je pense très fort a vous. Grosse pensée aussi a **Florence** qui crossfit fort et **Daniel** et les baleines.

Fredrik, jag är så glad att vi är tillsammans! Du har varit ett fantastiskt stöd, alltid snäll och förstående. Du visade mig hur en riktig och cool svensk person är och introducerade mig för alla coola rutiner. Tack för att du lyssnade och var entusiastisk när jag visade dig mina figurer och min kod. Tack till **Ann, Magnus, Sofia** och resten av din familj för att ni välkomnade mig och fick mig att känna mig som en i familjen.

Finalement, je tenais à remercier ma famille exceptionnelle qui m'a toujours soutenue dans tous mes projects. Vous êtes loin mais je pense fort a vous. Je voulais remercier mes parents **Yann** et **Marie-Odile** qui sont en plus d'être des scientifiques émerites sont des parents formidables et attentionés. Merci d'être venus me voir et de m'avoir écoutée et motivée pendant toute cette thèse. Merci à mes fréros **Bastien** et **Thomas** qui ont aussi bravé le froid suédois avec moi pour voir les aurores boréales ! Je pense bien fort à vous. Merci à mes axolotls Hondra et Neowise qui se sont montrés joyeux et non-demandants même avec des soins approximatifs durant cette thèse. On les félicite pour leurs 85 enfants!

7 References

1 Lint, C. V. *et al.* HIV-1 transcription and latency : an update. 1-38 (2013).

2 Yoshimura, K. Current status of HIV/AIDS in the ART era. *J Infect Chemother* **23**, 12-16, doi:10.1016/j.jiac.2016.10.002 (2017).

3 Haynes, B. F. New approaches to HIV vaccine development. *Curr Opin Immunol* **35**, 39-47, doi:10.1016/j.coi.2015.05.007 (2015).

4 Rosengarten, J. F., Schatz, S., Wolf, T., Barbe, S. & Stitz, J. Components of a HIV-1 vaccine mediate virus-like particle (VLP)-formation and display of envelope proteins exposing broadly neutralizing epitopes. *Virology* **568**, 41-48, doi:10.1016/j.virol.2022.01.008 (2022).

5 Burton, D. R. Advancing an HIV vaccine; advancing vaccinology. *Nat Rev Immunol* **19**, 77-78, doi:10.1038/s41577-018-0103-6 (2019).

6 Reust, C. E. Common adverse effects of antiretroviral therapy for HIV disease. *Am Fam Physician* **83**, 1443-1451 (2011).

7 Frescura, L. *et al.* Achieving the 95 95 95 targets for all: A pathway to ending AIDS. *PLoS One* **17**, e0272405, doi:10.1371/journal.pone.0272405 (2022).

8 Montessori, V., Press, N., Harris, M., Akagi, L. & Montaner, J. S. Adverse effects of antiretroviral therapy for HIV infection. *CMAJ* **170**, 229-238 (2004).

9 Koethe, J. R. *et al.* HIV and antiretroviral therapy-related fat alterations. *Nat Rev Dis Primers* **6**, 48, doi:10.1038/s41572-020-0181-1 (2020).

10 Ford, N. *et al.* The WHO public health approach to HIV treatment and care: looking back and looking ahead. *Lancet Infect Dis* **18**, e76-e86, doi:10.1016/S1473-3099(17)30482-6 (2018).

11 Naito, T. *et al.* Analysis of antiretroviral therapy switch rate and switching pattern for people living with HIV from a national database in Japan. *Sci Rep* **12**, 1732, doi:10.1038/s41598-022-05816-5 (2022).

12 Belikov, A. V. Age-related diseases as vicious cycles. *Ageing Res Rev* **49**, 11-26, doi:10.1016/j.arr.2018.11.002 (2019).

13 Appay, V., , Sauce, D. & Vol. 8 529-538 (Taylor & Francis, 2017).

14 Plowden, J., Renshaw-Hoelscher, M., Engleman, C., Katz, J. & Sambhara, S. Innate immunity in aging: impact on macrophage function. *Aging Cell* **3**, 161-167, doi:10.1111/j.1474-9728.2004.00102.x (2004).

15 Jang, J. Y., Blum, A., Liu, J. & Finkel, T. The role of mitochondria in aging. *J Clin Invest* **128**, 3662-3670, doi:10.1172/JCI120842 (2018).

16 Guaraldi, G., Milic, J. & Mussini, C. Aging with HIV. *Curr HIV/AIDS Rep* **16**, 475-481, doi:10.1007/s11904-019-00464-3 (2019).

17 Guaraldi, G. *et al.* Aging with HIV vs. HIV seroconversion at older age: a diverse population with distinct comorbidity profiles. *PLoS One* **10**, e0118531, doi:10.1371/journal.pone.0118531 (2015).

18 Meir-Shafrir, K., , Pollack, S. & Vol. 3 e0025 (2012).

19 Onen, N. F. *et al.* Aging and HIV infection: a comparison between older HIV-infected persons and the general population. *HIV Clin Trials* **11**, 100-109, doi:10.1310/hct1102-100 (2010).

20 Badley, A. D. & In vitro and in vivo effects of HIV protease inhibitors on apoptosis. **12**, 924-931, doi:10.1038/sj.cdd.4401580 (2005).

21 Payne, B. A. *et al.* Mitochondrial aging is accelerated by anti-retroviral therapy through the clonal expansion of mtDNA mutations. *Nat Genet* **43**, 806-810, doi:10.1038/ng.863 (2011).

22 Pelchen-Matthews, A. *et al.* Aging and the evolution of comorbidities among HIV-positive individuals in a European cohort. *AIDS* **32**, 2405-2416, doi:10.1097/QAD.0000000000001967 (2018).

23 Brunner, S. *et al.* Persistent viral infections and immune aging. **10**, 362-369, doi:10.1016/j.arr.2010.08.003 (2011).

24 Lercher, A. *et al.* Type I Interferon Signaling Disrupts the Hepatic Urea Cycle and Alters Systemic Metabolism to Suppress T Cell Function. **0**, 1-14, doi:10.1016/j.immuni.2019.10.014 (2019).

25 Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biol* **18**, 83, doi:10.1186/s13059-017-1215-1 (2017).

26 Misra, B. B. *et al.* Integrated omics: Tools, advances and future approaches. **62**, R21-R45, doi:10.1530/JME-18-0055 (2019).

27 Huang, S. *et al.* More is better: Recent progress in multi-omics data integration methods. **8**, 1-12, doi:10.3389/fgene.2017.00084 (2017).

28 Kim, J. *et al.* Toward Systems Understanding of Leaf Senescence: An Integrated Multi-Omics Perspective on Leaf Senescence Research. **9**, 813-825, doi:10.1016/j.molp.2016.04.017 (2016).

29 Ioannidis, J. P. A., , Khoury, M. J. & Improving validation practices in "omics" research. **334**, 1230-1232, doi:10.1126/science.1211811 (2011).

30 Argelaguet, R. *et al.* Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol* **14**, e8124, doi:10.15252/msb.20178124 (2018).

31 Angione, C. Human Systems Biology and Metabolic Modelling: A Review-From Disease Metabolism to Precision Medicine. *Biomed Res Int* **2019**, 8304260, doi:10.1155/2019/8304260 (2019).

32 Asparouhov, T. & Muthén, B. Multiple-Group Factor Analysis Alignment. *Structural Equation Modeling: A Multidisciplinary Journal* **21**, 495-508, doi:10.1080/10705511.2014.919210 (2014).

33 Mark D Robinson1, Alicia Oshlack1*. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* (2010).

34 Gosak, M. *et al.* Network science of biological systems at different scales: A review. *Phys Life Rev* **24**, 118-135, doi:10.1016/j.plrev.2017.11.003 (2018).

35 Aittokallio, T. & Schwikowski, B. Graph-based methods for analysing networks in cell biology. *Brief Bioinform* **7**, 243-255, doi:10.1093/bib/bbl022 (2006).

36 Lima-Mendez, G. & van Helden, J. The powerful law of the power law and other myths in network biology. *Mol Biosyst* **5**, 1482-1493, doi:10.1039/b908681a (2009).

37 Zhang, B., , Horvath, S. & A general framework for weighted gene co-expression network analysis. **4**, doi:10.2202/1544-6115.1128 (2005).

38 Bloch, F., , Jackson, M. O. & Centrality Measures in Networks. doi:10.2139/ssrn.2749124 (2016).

39 Farahani, F. V., Karwowski, W. & Lighthall, N. R. Application of Graph Theory for Identifying Connectivity Patterns in Human Brain Networks: A Systematic Review. *Front Neurosci* **13**, 585, doi:10.3389/fnins.2019.00585 (2019).

40 Li, J. *et al.* Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. *Sci Rep* **8**, 622, doi:10.1038/s41598-017-18705-z (2018).

41 Rahiminejad, S., Maurya, M. R. & Subramaniam, S. Topological and functional comparison of community detection algorithms in biological networks. *BMC Bioinformatics* **20**, 212, doi:10.1186/s12859-019-2746-0 (2019).

42 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

43 Hu, J. X., Thomas, C. E. & Brunak, S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* **17**, 615-629, doi:10.1038/nrg.2016.87 (2016).

44 Traag, V. A. *et al.* From Louvain to Leiden: guaranteeing well-connected communities. **9**, doi:10.1038/s41598-019-41695-z (2019).

45 Wang, B. *et al.* Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* **11**, 333-337, doi:10.1038/nmeth.2810 (2014).

46 Chierici, M. *et al.* Integrative Network Fusion: A Multi-Omics Approach in Molecular Profiling. *Front Oncol* **10**, 1065, doi:10.3389/fonc.2020.01065 (2020).

47 Wang, T. H. *et al.* Biomarker Identification through Multiomics Data Analysis of Prostate Cancer Prognostication Using a Deep Learning Model and Similarity Network Fusion. *Cancers (Basel)* **13**, doi:10.3390/cancers13112528 (2021).

48 Narayana, J. K., Mac Aogain, M., Ali, N., Tsaneva-Atanasova, K. & Chotirmall, S. H. Similarity network fusion for the integration of multi-omics and microbiomes in respiratory disease. *Eur Respir J* **58**, doi:10.1183/13993003.01016-2021 (2021).

49 Burton-Pimentel, K. J. *et al.* Discriminating Dietary Responses by Combining Transcriptomics and Metabolomics Data in Nutrition Intervention Studies. *Mol Nutr Food Res* **65**, e2000647, doi:10.1002/mnfr.202000647 (2021).

50 Agren, R. *et al.* The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. *PLoS Comput Biol* **9**, e1002980, doi:10.1371/journal.pcbi.1002980 (2013).

51 Phelan, V. V. & Feature-Based Molecular Networking for Metabolite Annotation. **2104**, 227-243, doi:10.1007/978-1-0716-0239-3_13 (2020).

52 Ryu, J. Y., Kim, H. U. & Lee, S. Y. Reconstruction of genome-scale human metabolic models using omics data. *Integr Biol (Camb)* **7**, 859-868, doi:10.1039/c5ib00002e (2015).

53 Rolfsson, O. *et al.* The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions. 1-16 (2011).

54 Ramon, C., Gollub, M. G. & Stelling, J. Integrating -omics data into genome-scale metabolic network models: principles and challenges. *Essays Biochem* **62**, 563-574, doi:10.1042/EBC20180011 (2018).

55 Monsalvo, M. *et al.* CD4/CD8 ratio improvement in HIV-1-infected patients receiving dual antiretroviral treatment. *Int J STD AIDS* **30**, 656-662, doi:10.1177/0956462419834129 (2019).

56 Presti, R. M. *et al.* Alterations in the oral microbiome in HIV-infected participants after antiretroviral therapy administration are influenced by immune status. *AIDS* **32**, 1279-1287, doi:10.1097/QAD.0000000000001811 (2018).

57 Jao, J. *et al.* Distinct Lipidomic Signatures in People Living With HIV: Combined Analysis of ACTG 5260s and MACS/WIHS. *J Clin Endocrinol Metab* **107**, 119-135, doi:10.1210/clinem/dgab663 (2022).

58 Ding, Y. *et al.* Comprehensive metabolomics profiling reveals common metabolic alterations underlying the four major non-communicable diseases in treated HIV infection. *EBioMedicine* **71**, 103548, doi:10.1016/j.ebiom.2021.103548 (2021).

59 Wang, X. M. *et al.* Global transcriptomic characterization of T cells in individuals with chronic HIV-1 infection. *Cell Discov* **8**, 29, doi:10.1038/s41421-021-00367-x (2022).

60 Bowman, E. & Funderburg, N. T. Lipidome Abnormalities and Cardiovascular Disease Risk in HIV Infection. *Curr HIV/AIDS Rep* **16**, 214-223, doi:10.1007/s11904-019-00442-9 (2019).

61 Peltenburg, N. C. *et al.* Vol. 8 1-11 (2018).

62 Golumbeanu, M. *et al.* Proteo-Transcriptomic Dynamics of Cellular Response to HIV-1 Infection. *Sci Rep* **9**, 213, doi:10.1038/s41598-018-36135-3 (2019).

63 Belshan, M. *et al.* Discovery of candidate HIV-1 latency biomarkers using an OMICs approach. *Virology* **558**, 86-95, doi:10.1016/j.virol.2021.03.003 (2021).

64 Barclay, R. A. *et al.* An Omics Approach to Extracellular Vesicles from HIV-1 Infected Cells. *Cells* **8**, doi:10.3390/cells8080787 (2019).

65 Chan, E. Y. *et al.* Quantitative analysis of human immunodeficiency virus type 1-infected CD4+ cell proteome: dysregulated cell cycle progression and nuclear transport coincide with robust virus production. *J Virol* **81**, 7571-7583, doi:10.1128/JVI.00288-07 (2007).

66 Chang, S. T. *et al.* Next-generation sequencing reveals HIV-1-mediated suppression of T cell activation and RNA processing and regulation of noncoding RNA expression in a CD4+ T cell line. *mBio* **2**, doi:10.1128/mBio.00134-11 (2011).

67 Krishnan, S. *et al.* Integrative Multi-Omics Reveals Serum Markers of Tuberculosis in Advanced HIV. *Front Immunol* **12**, 676980, doi:10.3389/fimmu.2021.676980 (2021).

68 Pinto-Cardoso, S. *et al.* Fecal Bacterial Communities in treated HIV infected individuals on two antiretroviral regimens. *Sci Rep* **7**, 43741, doi:10.1038/srep43741 (2017).

69 Williams, M. E., Naude, P. J. W. & van der Westhuizen, F. H. Proteomics and metabolomics of HIV-associated neurocognitive disorders: A systematic review. *J Neurochem* **157**, 429-449, doi:10.1111/jnc.15295 (2021).

70 Babu, H. *et al.* Plasma Metabolic Signature and Abnormalities in HIV-Infected Individuals on Long-Term Successful Antiretroviral Therapy. *Metabolites* **9**, doi:10.3390/metabo9100210 (2019).

71 Svensson Akusjärvi, S. *et al.* Role of myeloid cells in system-level immunometabolic dysregulation during prolonged successful HIV-1 treatment. *AIDS*, doi:10.1097/QAD.0000000000003512 (2023).

72 Ambikan, A. T. *et al.* Genome-scale metabolic models for natural and long-term drug-induced viral control in HIV infection. *Life Sci Alliance* **5**, doi:10.26508/lsa.202201405 (2022).

73 Ivanov, S. *et al.* Vol. 11 1-21 (2020).

74 Jude, M. S. *et al.* Multi 'Omics Profiling of the HIV Airway Epithelium: Integration of the Microbiome, Methylome, and Transcriptome. **12274272**, A2940-A2940, doi:10.1164/ajrccm-conference.2020.201.1_meetingabstracts.a2940 (2020).

75 Pei, L. *et al.* Plasma Metabolomics Reveals Dysregulated Metabolic Signatures in HIV-Associated Immune Reconstitution Inflammatory Syndrome. *Front Immunol* **12**, 693074, doi:10.3389/fimmu.2021.693074 (2021).

76 Ronit, A. *et al.* Copenhagen comorbidity in HIV infection (COCOMO) study: a study protocol for a longitudinal, non-interventional assessment of non-AIDS comorbidity in HIV infection in Denmark. **16**, 713, doi:10.1186/s12879-016-2026-9 (2016).

77 Mortensen, M. B., Afzal, S., Nordestgaard, B. G. & Falk, E. The high-density lipoprotein-adjusted SCORE model worsens SCORE-based risk classification in a contemporary population of 30,824 Europeans: the Copenhagen General Population Study. *Eur Heart J* **36**, 2446-2453, doi:10.1093/eurheartj/ehv251 (2015).

78 Groen, N. *et al.* Stepping into the omics era: Opportunities and challenges for biomaterials science and engineering. *Acta Biomater* **34**, 133-142, doi:10.1016/j.actbio.2016.02.015 (2016).

79 Williams, C. R. *et al.* Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA-Seq. **18**, doi:10.1186/s12859-016-1457-z (2017).

80 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

81 Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587, doi:10.1038/nature13319 (2014).

82 Calderon-Celis, F., Encinar, J. R. & Sanz-Medel, A. Standardization approaches in absolute quantitative proteomics with mass spectrometry. *Mass Spectrom Rev* **37**, 715-737, doi:10.1002/mas.21542 (2018).

83 Willforss, J., Chawade, A. & Levander, F. NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. *J Proteome Res* **18**, 732-740, doi:10.1021/acs.jproteome.8b00523 (2019).

84 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).

85 Shin, S. Y. *et al.* An atlas of genetic influences on human blood metabolites. **46**, 543-550, doi:10.1038/ng.2982 (2014).

86 Guijas, C., Montenegro-Burke, J. R., Warth, B., Spilker, M. E. & Siuzdak, G. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat Biotechnol* **36**, 316-320, doi:10.1038/nbt.4101 (2018).

87 Milburn, M. V., , Lawton, K. A. & Application of metabolomics to diagnosis of insulin resistance. **64**, 291-305, doi:10.1146/annurev-med-061511-134747 (2013).

88 Kennedy, A. D. *et al.* Metabolomics in the clinic: A review of the shared and unique features of untargeted metabolomics for clinical research and clinical testing. **53**, 1143-1154, doi:10.1002/jms.4292 (2018).

89 Becker, S., Kortz, L., Helmschrodt, C., Thiery, J. & Ceglarek, U. LC-MS-based metabolomics in the clinical laboratory. *J Chromatogr B Analyt Technol Biomed Life Sci* **883-884**, 68-75, doi:10.1016/j.jchromb.2011.10.018 (2012).

90 Bar, N. *et al.* A reference map of potential determinants for the human serum metabolome. *Nature* **588**, 135-140, doi:10.1038/s41586-020-2896-2 (2020).

91 Yet, I. *et al.* Genetic Influences on Metabolite Levels: A Comparison across Metabolomic Platforms. *PLoS One* **11**, e0153672, doi:10.1371/journal.pone.0153672 (2016).

92 Zhao, Y. Y., Vaziri, N. D. & Lin, R. C. Lipidomics: new insight into kidney disease. *Adv Clin Chem* **68**, 153-175, doi:10.1016/bs.acc.2014.11.002 (2015).

93 Hu, C., Duan, Q. & Han, X. Strategies to Improve/Eliminate the Limitations in Shotgun Lipidomics. *Proteomics* **20**, e1900070, doi:10.1002/pmic.201900070 (2020).

94 Knight, R. *et al.* Best practices for analysing microbiomes. *Nat Rev Microbiol* **16**, 410-422, doi:10.1038/s41579-018-0029-9 (2018).

95 Anderson, M. J. in *Wiley StatsRef: Statistics Reference Online* 1-15 (2017).

96 Lee, L. C., Liong, C. Y. & Jemain, A. A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst* **143**, 3526-3539, doi:10.1039/c8an00599k (2018).

97 Kursa, M. B., , Rudnicki, W. R. & Feature Selection with the Boruta Package **36**, doi:10.18637/jss.v036.i11 (2015).

98 Díaz-Uriarte, R., , Alvarez de Andrés, S. & Gene selection and classification of microarray data using random forest. **7**, 3, doi:10.1186/1471-2105-7-3 (2006).

99 Carapito, R. *et al.* Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. **14**, 1-21, doi:10.1126/scitranslmed.abj7521 (2022).

100 Garcia-Campos, M. A., Espinal-Enriquez, J. & Hernandez-Lemus, E. Pathway Analysis: State of the Art. *Front Physiol* **6**, 383, doi:10.3389/fphys.2015.00383 (2015).

101 Krämer, A. *et al.* Causal analysis approaches in ingenuity pathway analysis. **30**, 523-530, doi:10.1093/bioinformatics/btt703 (2014).

102 Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, doi:10.1093/bioinformatics/btac757 (2023).

103 Varemo, L., Nielsen, J. & Nookaew, I. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res* **41**, 4378-4391, doi:10.1093/nar/gkt111 (2013).

104 Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448-3449, doi:10.1093/bioinformatics/bti551 (2005).

105 Wilkerson, M. D., , Hayes, D. N. & ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. **26**, 1572-1573, doi:10.1093/bioinformatics/btq170 (2010).

106 Pai, S. *et al.* netDx: interpretable patient classification using integrated patient similarity networks. **15**, 1-11, doi:10.15252/msb.20188497 (2019).

107 van Dam, S. *et al.* Gene co-expression analysis for functional classification and gene-disease predictions. **19**, 575-592, doi:10.1093/bib/bbw139 (2018).

108 The Igraph Software Package for Complex Network Research (Budapest, Hungary, 2005).

109 ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York, 2016).

110 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

111 Kanehisa, M. *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **34**, D354-357, doi:10.1093/nar/gkj102 (2006).

112 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740, doi:10.1093/bioinformatics/btr260 (2011).

113 Kuhn, M., von Mering, C., Campillos, M., Jensen, L. J. & Bork, P. STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res* **36**, D684-688, doi:10.1093/nar/gkm795 (2008).

114 Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. **12**, 453-457, doi:10.1038/nmeth.3337 (2015).

115 Ialongo, C. Understanding the effect size and its measures. *Biochem Med (Zagreb)* **26**, 150-163, doi:10.11613/BM.2016.015 (2016).

116 Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60, doi:10.1186/gb-2011-12-6-r60 (2011).

117 *Cancer Epidemiology*. (2009).

118 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

119 Alvarez, D. F., Helm, K., Degregori, J., Roederer, M. & Majka, S. Publishing flow cytometry data. *Am J Physiol Lung Cell Mol Physiol* **298**, L127-130, doi:10.1152/ajplung.00313.2009 (2010).

120 Willig, A. L., , Overton, E. T. & Vol. 13 289-296 (Current HIV/AIDS Reports, 2016).

121 Wang, Q. *et al.* Lipids profile among ART-naive HIV infected patients and men who have sex with men in China: a case control study. *Lipids Health Dis* **15**, 149, doi:10.1186/s12944-016-0297-1 (2016).

122 Rosado-Sanchez, I. *et al.* Glutaminolysis and lipoproteins are key factors in late immune recovery in successfully treated HIV-infected patients. *Clin Sci (Lond)* **133**, 997-1010, doi:10.1042/CS20190111 (2019).

123 Munshi, S. U., Rewari, B. B., Bhavesh, N. S. & Jameel, S. Nuclear magnetic resonance based profiling of biofluids reveals metabolic dysregulation in HIV-infected persons and those on anti-retroviral therapy. *PLoS One* **8**, e64298, doi:10.1371/journal.pone.0064298 (2013).

124 Nedelcovych, M. T. *et al.* Glutamine Antagonist JHU083 Normalizes Aberrant Glutamate Production and Cognitive Deficits in the EcoHIV Murine Model of HIV-Associated Neurocognitive Disorders. *J Neuroimmune Pharmacol* **14**, 391-400, doi:10.1007/s11481-019-09859-w (2019).

125 Kishimoto, N. *et al.* Glucose-dependent aerobic glycolysis contributes to recruiting viral components into HIV-1 particles to maintain infectivity. *Biochem Biophys Res Commun* **549**, 187-193, doi:10.1016/j.bbrc.2021.02.071 (2021).

126 Castellano, P., Prevedel, L., Valdebenito, S. & Eugenin, E. A. HIV infection and latency induce a unique metabolic signature in human macrophages. *Sci Rep* **9**, 3941, doi:10.1038/s41598-019-39898-5 (2019).

127 Akins, N. S., Nielson, T. C. & Le, H. V. Inhibition of Glycolysis and Glutaminolysis: An Emerging Drug Discovery Approach to Combat Cancer. *Curr Top Med Chem* **18**, 494-504, doi:10.2174/1568026618666180523111351 (2018).

128 Shim, K., Gulhar, R. & Jialal, I. Exploratory metabolomics of nascent metabolic syndrome. *J Diabetes Complications* **33**, 212-216, doi:10.1016/j.jdiacomp.2018.12.002 (2019).

129 Sengupta, S. & Siliciano, R. F. Targeting the Latent Reservoir for HIV-1. *Immunity* **48**, 872-895, doi:10.1016/j.immuni.2018.04.030 (2018).

130 Zoia, C. *et al.* Glutamate transporters in platelets: EAAT1 decrease in aging and in Alzheimer's disease. *Neurobiology of Aging* **25**, 149-157, doi:10.1016/s0197-4580(03)00085-x (2004).

131 Gorska, A. M. & Eugenin, E. A. The Glutamate System as a Crucial Regulator of CNS Toxicity and Survival of HIV Reservoirs. *Front Cell Infect Microbiol* **10**, 261, doi:10.3389/fcimb.2020.00261 (2020).

132 Ros, J., Pecinska, N., Alessandri, B., Landolt, H. & Fillenz, M. Lactate reduces glutamate-induced neurotoxicity in rat cortex. *J Neurosci Res* **66**, 790-794, doi:10.1002/jnr.10043 (2001).

133 Mukerji, S. S. *et al.* Low Neuroactive Steroids Identifies a Biological Subtype of Depression in Adults with Human Immunodeficiency Virus on Suppressive Antiretroviral Therapy. *J Infect Dis* **223**, 1601-1611, doi:10.1093/infdis/jiaa104 (2021).

134 Noor, E., Eden, E., Milo, R. & Alon, U. Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Mol Cell* **39**, 809-820, doi:10.1016/j.molcel.2010.08.031 (2010).

135 Agus, A., Clement, K. & Sokol, H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut* **70**, 1174-1182, doi:10.1136/gutjnl-2020-323071 (2021).

136 Hoenigl, M., Kessler, H. H. & Gianella, S. Editorial: HIV-Associated Immune Activation and Persistent Inflammation. *Front Immunol* **10**, 2858, doi:10.3389/fimmu.2019.02858 (2019).

137 Gojak, R. *et al.* CD4/CD8 Ratio as a Predictor for the Occurrence of Metabolic Syndrome in HIV / AIDS Patients During 6 Months of cART Therapy. *J Med Biochem* **38**, 489-495, doi:10.2478/jomb-2018-0049 (2019).

138 Wong, M. E. *et al.* The HIV reservoir in monocytes and macrophages. **10**, doi:10.3389/fimmu.2019.01435 (2019).

139 Chia, T. Y., Zolp, A. & Miska, J. Polyamine Immunometabolism: Central Regulators of Inflammation, Cancer and Autoimmunity. *Cells* **11**, doi:10.3390/cells11050896 (2022).

140 Merali, S. *et al.* Polyamines: Predictive Biomarker for HIV-Associated Neurocognitive Disorders. *J AIDS Clin Res* **5**, 1000312, doi:10.4172/2155-6113.1000312 (2014).

141 Mahalingam, S. S. *et al.* Polyamine metabolism impacts T cell dysfunction in the oral mucosa of people living with HIV. *Nat Commun* **14**, 399, doi:10.1038/s41467-023-36163-2 (2023).

142 Carapito, R. *et al.* Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. **14**, 1-21, doi:10.1126/scitranslmed.abj7521 (2022).

143 Ricard, N. *et al.* Endothelial ERK1/2 signaling maintains integrity of the quiescent endothelium. *J Exp Med* **216**, 1874-1890, doi:10.1084/jem.20182151 (2019).

144 Chauvel, C., Novoloaca, A., Veyre, P., Reynier, F. & Becker, J. Evaluation of integrative clustering methods for the analysis of multi-omics data. *Brief Bioinform* **21**, 541-552, doi:10.1093/bib/bbz015 (2020).

145 Huang, S., Chaudhary, K. & Garmire, L. X. More is better: Recent progress in multi-omics data integration methods. **8**, 1-12, doi:10.3389/fgene.2017.00084 (2017).

146 Haak, B. W. *et al.* Integrative Transkingdom Analysis of the Gut Microbiome in Antibiotic Perturbation and Critical Illness. *mSystems* **6**, doi:10.1128/mSystems.01148-20 (2021).

147 Ranjan, R., Rani, A., Metwally, A., McGee, H. S. & Perkins, D. L. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun* **469**, 967-977, doi:10.1016/j.bbrc.2015.12.083 (2016).

148 Horvath, S. & Dong, J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol* **4**, e1000117, doi:10.1371/journal.pcbi.1000117 (2008).

149 Fang, X. *et al.* Reconstructing organisms in silico: genome-scale models and their emerging applications. 1-13, doi:10.1038/s41579-020-00440-4 (2020).

150 Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol* **20**, 121, doi:10.1186/s13059-019-1730-3 (2019).

151 Terzer, M., Maynard, N. D., Covert, M. W. & Stelling, J. Genome-scale metabolic networks. *Wiley Interdiscip Rev Syst Biol Med* **1**, 285-297, doi:10.1002/wsbm.37 (2009).

152 Hattwell, J. P. N., Hastings, J., Casanueva, O., Schirra, H. J. & Witting, M. Using Genome-Scale Metabolic Networks for Analysis, Visualization, and Integration of Targeted Metabolomics Data. *Methods Mol Biol* **2104**, 361-386, doi:10.1007/978-1-0716-0239-3_18 (2020).

153 Lewis, J. E., Forshaw, T. E., Boothman, D. A., Furdui, C. M. & Kemp, M. L. Personalized Genome-Scale Metabolic Models Identify Targets of Redox Metabolism in Radiation-Resistant Tumors. *Cell Syst* **12**, 68-81.e11, doi:10.1016/j.cels.2020.12.001 (2021).

154 Kenefake, D., Armingol, E., Lewis, N. E. & Pistikopoulos, E. N. An improved algorithm for flux variability analysis. *BMC Bioinformatics* **23**, 550, doi:10.1186/s12859-022-05089-9 (2022).

155 Zampieri, G., Vijayakumar, S., Yaneske, E. & Angione, C. Machine and deep learning meet genome-scale metabolic modeling. *PLoS Comput Biol* **15**, e1007084, doi:10.1371/journal.pcbi.1007084 (2019).

156 Singh, A. *et al.* DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* **35**, 3055-3062, doi:10.1093/bioinformatics/bty1054 (2019).

157 Huang, P. L. A comprehensive definition for metabolic syndrome. *Dis Model Mech* **2**, 231-237, doi:10.1242/dmm.001180 (2009).

158 Neeland, I. J. *et al.* Associations of visceral and abdominal subcutaneous adipose tissue with markers of cardiac and metabolic risk in obese adults. *Obesity (Silver Spring)* **21**, E439-447, doi:10.1002/oby.20135 (2013).

159 Ochiai, H. *et al.* Elevated alanine aminotransferase and low aspartate aminotransferase/alanine aminotransferase ratio are associated with chronic kidney disease among middle-aged women: a cross-sectional study. *BMC Nephrol* **21**, 471, doi:10.1186/s12882-020-02144-6 (2020).

160 Chang, E., Varghese, M. & Singer, K. Gender and Sex Differences in Adipose Tissue. *Curr Diab Rep* **18**, 69, doi:10.1007/s11892-018-1031-3 (2018).

161 Corraini, P., Olsen, M., Pedersen, L., Dekkers, O. M. & Vandenbroucke, J. P. Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. *Clin Epidemiol* **9**, 331-338, doi:10.2147/CLEP.S129728 (2017).

162 Skelly, A. C., Dettori, J. R. & Brodt, E. D. Assessing bias: the importance of considering confounding. *Evid Based Spine Care J* **3**, 9-12, doi:10.1055/s-0031-1298595 (2012).

163 Tuddenham, S. A. *et al.* The Impact of Human Immunodeficiency Virus Infection on Gut Microbiota alpha-Diversity: An Individual-level Meta-analysis. *Clin Infect Dis* **70**, 615-627, doi:10.1093/cid/ciz258 (2020).

164 Zhou, J. *et al.* Gut Microbiome Changes Associated With HIV Infection and Sexual Orientation. *Front Cell Infect Microbiol* **10**, 434, doi:10.3389/fcimb.2020.00434 (2020).

165 Lv, T., Cao, W. & Li, T. HIV-Related Immune Activation and Inflammation: Current Understanding and Strategies. *J Immunol Res* **2021**, 7316456, doi:10.1155/2021/7316456 (2021).

166 Kumari, R. & Jat, P. Mechanisms of Cellular Senescence: Cell Cycle Arrest and Senescence Associated Secretory Phenotype. *Front Cell Dev Biol* **9**, 645593, doi:10.3389/fcell.2021.645593 (2021).

167 McHugh, D. & Gil, J. Senescence and aging: Causes, consequences, and therapeutic avenues. *J Cell Biol* **217**, 65-77, doi:10.1083/jcb.201708092 (2018).

168 Cohen, J. & Torres, C. HIV-associated cellular senescence: A contributor to accelerated aging. *Ageing Res Rev* **36**, 117-124, doi:10.1016/j.arr.2016.12.004 (2017).

169 Wu, J. W. *et al.* Biological age in healthy elderly predicts aging-related diseases including dementia. *Sci Rep* **11**, 15929, doi:10.1038/s41598-021-95425-5 (2021).

170 Franzago, M., Santurbano, D., Vitacolonna, E. & Stuppia, L. Genes and Diet in the Prevention of Chronic Diseases in Future Generations. *Int J Mol Sci* **21**, doi:10.3390/ijms21072633 (2020).

171 De La Torre-Tarazona, E., Ayala-Suarez, R., Diez-Fuertes, F. & Alcami, J. Omic Technologies in HIV: Searching Transcriptional Signatures Involved in Long-Term Non-Progressor and HIV Controller Phenotypes. *Front Immunol* **13**, 926499, doi:10.3389/fimmu.2022.926499 (2022).

172 Lay, J. O., Liyanage, R., Borgmann, S. & Wilkins, C. L. Problems with the "omics". *TrAC Trends in Analytical Chemistry* **25**, 1046-1056, doi:10.1016/j.trac.2006.10.007 (2006).

173 Vos, W. *et al.* The 2000HIV study: Design, multi-omics methods and participant characteristics. *Front Immunol* **13**, 982746, doi:10.3389/fimmu.2022.982746 (2022).

174 Hu, Y. *et al.* Single Cell Multi-Omics Technology: Methodology and Application. *Front Cell Dev Biol* **6**, 28, doi:10.3389/fcell.2018.00028 (2018).

175 Assum, I. *et al.* Tissue-specific multi-omics analysis of atrial fibrillation. *Nat Commun* **13**, 441, doi:10.1038/s41467-022-27953-1 (2022).