Network-based multi-omics integration reveals metabolic risk profile within treated HIV-infection

Flora Mikaeloff[1*], Marco Gelpi[2], Rui Benfeitas[3], Andreas D. Knudsen[2], Beate Vestad[4], Julie Høgh[2], Johannes R. Hov[4,5], Thomas Benfield[6], Daniel Murray[7], Christian G Giske[8], Adil Mardinoglu[9,10], Marius Trøseid[4,11], Susanne D. Nielsen[2]†, Ujjwal Neogi[1,]†*

[1]The Systems Virology Lab, Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden.

[2]Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark,

[3]National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, S-10691 Stockholm, Sweden.

[4]Research Institute of Internal Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway

[5]Norwegian PSC Research Center, Oslo University Hospital Rikshospitalet, Oslo, Norway,

[6]Centre of excellence for Health, Immunity and Infections, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark,

[7]Department of Infectious Diseases, Copenhagen University Hospital – Amager and Hvidovre, Hvidovre, Denmark,

[8]Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, ANA Futura, Campus Flemingsberg, S-14152 Stockholm, Sweden.

[9]Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, Sweden;

[10]Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, United Kingdom

[11]Institute of Clinical Medicine, Oslo, Norway

†Equal contribution

*Corresponding authors. Email: flora.mikaeloff@ki.se, ujjwal.neogi@ki.se

Abstract:

Applying omics technologies improve biological understanding of pathology and systemic health status in people living with HIV on antiretroviral therapy ($PLWH_{ART}$). Still, a systematic and in-depth characterization of metabolic risk profile during successful long-term treatment is lacking. In this study, we used multi-omics (plasma lipidomic and metabolomic, and fecal 16s microbiome) data-driven stratification and characterization to identify the metabolic at-risk profile within $PLWH_{ART}$. Through network analysis and similarity network fusion (SNF), we identified three groups of $PLWH_{ART}$ (SNF-1 to 3). The SNF-2 (n=44; 45%) had obesity-related clinical features (increased visceral adipose tissue, BMI), higher incidence of metabolic syndrome (MetS), and increased di- and triglycerides despite having higher $CD4^+$ T-cell counts than the other two clusters. The influence of sexual orientation in the microbiome was observed in the SNF-2 and SNF-3, with increased α-diversity compared to SNF-1. Our multi-omics integrative analysis reveals a complex microbial interplay by microbiome-derived metabolites in $PLWH_{ART}$. It identifies a risk cluster of PLWH who may benefit from personalized medicine and lifestyle intervention to improve their metabolic profile and enhance healthy aging.

Introduction:

Antiretroviral therapy (ART) has improved the immune profile by suppressing viral replication and reducing the morbidity and mortality of people living with HIV (PLWH). Yet living with HIV under ART induces a strong metabolic perturbation in the body due to virus persistence, immune activation, chronic low-grade inflammation, and treatment toxicity, mostly with older antiretrovirals (Yoshimura, 2017). The biological shifts due to a mixed effect of drugs and viruses are also highly personalized depending on the patient genetic background, age, gender, immunological, and lifestyle factors (Pelchen-Matthews *et al*, 2018). The long-term HIV infection, even with successful ART, is associated with an accentuated onset of non-AIDS-related comorbidities (Deeks, 2011). Consequently, diseases of the aged population appear in relatively young HIV patients, including cardiovascular disease, liver-kidney disease, and neurocognitive and metabolic disorders (Nasi *et al*, 2017).

Systems biological analyses are valuable methodologies for systematically understanding pathology and identifying potential novel treatment strategies (Karahalil, 2016). Microbiome studies provided enormous knowledge about the microbial association with the HIV status, sexual practice, and gender (Gelpi *et al*, 2020; Noguera-Julian *et al*, 2016; Zhou *et al*, 2020) and the possible interplay between HIV-related gut microbiota, immune dysfunction, and comorbidities like metabolic syndrome (MetS) and visceral adipose tissue (VAT) accumulation (Gelpi *et al.*, 2020). Our extensive metabolomics studies from three different cohorts from India (Babu *et al*, 2019), Cameroon (Mikaeloff *et al*, 2022), and Denmark (Gelpi *et al*, 2021) with more than 500 PLWH indicated that disrupted amino acid (AA) metabolism in PLWH with ART ($PLWH_{ART}$) following prolonged ART that plays the central role in the comorbidities such as MetS (Gelpi *et al.*, 2021).

Multi-omic characterizations may offer insights into understanding the mechanisms underlying biological processes in a specific disease condition. The application of integrative omics to understand the disease pathogenesis in PLWH under suppressive ART is lacking. To the best of our knowledge, no integrative omics studies have been performed to understand complex biological phenotypes in PLWH during prolonged suppressive ART ($PLWH_{ART}$). A recent longitudinal study integrating metabolomics, plasma protein biomarkers, and transcriptomics in patients' samples identified potential lipid and amino acid metabolism perturbations in PLWH with immune reconstitution inflammatory syndrome (IRIS) (Pei *et al*, 2021). Our recent network-based integrative plasma lipidomics, metabolic biomarker, and clinical data indicated a coordinated role of clinical parameters like accumulation of visceral adipose tissue (VAT) and exposure to earlier generations of antiretrovirals with glycerolipids and glutamate metabolism in the pathogenesis of PLWH with MetS (Olund Villumsen *et al*, 2021).

The present study aimed to identify the molecular data-driven phenotypic patient stratification using network-based integration of plasma metabolomics/lipidomics and fecal microbiota in a cohort of $PLWH_{ART}$ with prolonged suppressive therapy to identify the at-risk metabolic profile following long-term successful therapy. We further investigated the underlying factors differing from these profiles and the link to their clinical phenotype to clarify risk factors for metabolic disease further.

Results

Comprehensive multi-omics characterization of PLWH on successful cART: In this study, we used untargeted plasma metabolomics (877 metabolites) (Gelpi *et al.*, 2021), lipidomics (977 lipids) (Olund Villumsen *et al.*, 2021), and fecal 16s rRNA microbiome [241 operational taxonomic unit (OTU)] data (Gelpi *et al.*, 2020) from 97 $PLWH_{ART}$ from the Copenhagen Comorbidity (COCOMO) cohort (Gelpi *et al*, 2018) where we have three levels of omics data available. Additionally, we included 48 clinical and demographical features comprising lifestyle habits (food, medicine, alcohol, smoking), comorbidities linked to obesity, and HIV-related measurements (viral load, treatment history, CD4 T-cell count, CD8 T-cell counts) (Table S1). The PLWH were mainly male (86%; 84/97), of Caucasian ethnic origin (81%, 79/97) with a median (IQR) age of 54 (48-63) years. The median (IQR) duration of the treatment was 15 (9-18) years. At the time of sample collection, the viral load was below detection level with successful immune reconstitution [median (IQR) CD4 T-cell count 713 (570-900) cells/µL]. Additionally, 20 HIV-negative controls (HC) with similar sex proportions (90%, 18/20) and median age (IQR) 56 (50-67) years were used to reference multi-omics.

Integrative omics-based similarity network fusion (SNF) identifies three clusters in $PLWH_{ART}$: To stratify the $PLWH_{ART}$ based on their molecular signature, we used Similarity Network Fusion (SNF) that constructs similarity matrices and networks of $PLWH_{ART}$ for each of the omics and fuses into one network that represents the full spectrum of the underlying data and disease status in $PLWH_{ART}$ (Wang *et al*, 2014). We identified three clusters of patients, defined as SNF-1 (N=19), SNF-2 (N=44), and SNF-3 (N=34) (Fig 1A). The concordance matrix based on Normalized Mutual Information (NMI) score (0=no mutual information, 1=perfect correlation) showed that lipids had the most influence in the final network (NMI=0.6), followed by metabolites (NMI=0.4) and finally, microbiome (NMI=0.3) (Fig 1B). A clear segregation of the SNF clusters (Fig 1C) was observed in the PCA based on the fused network values (Fig 1D) and PCA of single omics for lipidomics and metabolomics but not microbiome (Fig S1). The addition of HC with the clusters showed that the SNF-3 had an HC-like profile in metabolomics and lipidomics PCA plots (Fig S1).

Lipids and metabolites highlight clinical differences between patient clusters: To characterize the molecular data-driven clusters of $PLWH_{ART}$, we used clinical and single omics analysis. Cluster-specific clinical characteristics of $PLWH_{ART}$ are presented in Table 1. Clusters were not statistically different for age, gender, duration of ART, and type of ART($p>0.05$). On the other hand, SNF-1 had the healthiest profile, SNF-3 an intermediate, and SNF-2 the most severe metabolic perturbations indicating at-risk metabolic profile. SNF-2 represented patients with high BMI, central obesity, higher VAT, and incidence of MetS (all $p<0.05$). Regardless of disease severity, SNF-2 had a higher CD4+ T-cell count at the time of sample collection and more men who have sex with men (MSM) as transmission mode compared to the other clusters (all $p<0.05$). The SNF-2 and SNF-3 had a significantly higher subcutaneous adipose tissue (SAT) and incidence of Hypertension compared to SNF-1 (all $p<0.05$). SNF-1 had the lowest BMI, SAT, VAT, and incidence of Hypertension (all $p<0.05$). A similar lipid class profile was observed between SNF-1, SNF-3, and HC (Table S2). Patients from SNF-2 showed a drastic increase in diglycerides (DAG) (Fig 1E) and triglycerides (TAG) (Fig 1F) compared to SNF-1, SNF-3, and HC (all FDR<0.1) as well as other lipids classes which coordinate with their clinical metabolic profile (Fig S2). In this analysis, the relation between cluster and ART class was not significant (X2, FDR = 0,45). Still, we can mention that the three groups had an important proportion of missing data for this variable (16%, 29% and 29% respectively).

However, comparing the metabolites (Table S3), SNF-3 was the closest to HC, followed by SNF-2, and finally, SNF-1, the centroid of which was close to the SNF-2 centroid (Fig 1G). The differential metabolite abundance (DMA) was presented in Fig 1H. Among these highly differing metabolites, most perturbations were observed between HC and SNF-1 (124/159) and HC and SNF-2 (62/159). These clusters showed an up-regulation of the metabolites in the xenobiotics, nucleotides, and AA pathways compared to HC (HC vs. SNF-1, 97/124, HC vs. SNF-2, 45/62). SNF-3 and HC had only nine metabolites differing, in line with the high clustering of both groups shown with PCA. In turn, SNF-1 and SNF-2 showed similar metabolic profiles. The lipids (SNF-2 vs. SNF-3, 10/19, SNF-1 vs. SNF-3, 12/21) and amino acids (SNF-2 vs. SNF-3, 7/19) differed significantly between the clusters. Among these metabolites, 50 had a low or moderate association with age and BMI (Spearman correlation, absolute $R<0.4$, $p<0.1$) and 51 with gender (chi-squared test, $p<0.1$), showing the modest influence of individual characteristics on metabolomics profile. Combining the in-depth metabolomics and lipidomic data indicated more personalized risk factors for $PLWH_{ART}$ that cannot be explained by the clinical features and a complex interplay between the multi-omics layers define overall health status.

Sexual preferences influence the clusters' differences driven by the microbiome. As the metabolic aberrations were closely linked with the microbiome profile, we investigated the microbiome's impact in PLWH clusters. The α-diversity indices indicated a loss of diversity according to Observed, ACE, se.ACE, Chao1, and Fisher indices in SNF-1 compared to SNF-2 (Mann Whitney, FDR<0.05) (Fig 2A, Fig S3, Table S4). A non-metric multidimensional scaling (NMDS) ordination of the dissimilarity-based index (Bray-Curtis) of diversity at the OTU level was performed to measure the inter-individual differences between clusters (β-diversity) (Fig 2B). Based on NMDS plot axis coordinate 1, SNF-1 was segregated separately from SNF-2 and SNF-3 (Mann Whitney, FDR<0.05, Fig 2C). The relative abundance of fecal microbiota was more influenced by the transmission mode than the cluster itself (Fig S4a). No other comorbidities on the microbiome profile were observed (Fig S4b-d). The SNF-2 had a significantly higher number of MSM compared to the other clusters (Table 1). While grouping SNF-2 and SNF-3, we count 54/78 (69%) MSM for SNF-2/SNF-3 and 9/19 (47%) MSM for SNF-1. This indicated that sexual preferences and the HIV-1 transmission mode relate to compositional differences in fecal microbiota between SNF clusters. Permutational multivariate analysis of variance (PERMANOVA) at the family level showed that the centroids of the SNF-1 samples were different from SNF-2 (FDR<0.001) and SNF-3 samples (FDR=0.0054) (Table S5), indicating that there is only a location effect as permutation test for homogeneity of multivariate dispersions was not significant between the clusters (FDR>0.05). No statistical difference was observed between SNF-2 and SNF-3 in both tests (FDR=0.38). SNF-1 was enriched in Bacteroides and Lachnospira, while SNF-2 and SNF-3 were enriched in Prevotella, Veillonella, and Succinivibrio (Fig 2D-2E). These families were also among 54 significantly discriminative features between SNF-1 and SNF-2/SNF-3 as shown with linear discriminant analysis effect size (LefSe) (Fig 2F). Mann Whitney U test between clusters at the family level also found Prevotella and Bacteroides to be statistically distinct between these clusters (FDR<0.05, Table S6). Our data thus support the potential role of the Prevotella and Bacteroides in the cluster separation that could be mediated by the sexual preferences in $PLWH_{ART}$.

Factor and network analysis indicated the importance of microbiome-derived metabolites:

To identify the molecular and clinical factors driving SNF cluster separation at the multi-omic level, we employed the Multi-Omic Factor Analysis (MOFA) tool for multi-omics integration (Argelaguet *et al*, 2018). After low variance filtering, the MOFA model was built using three views: microbiome with 173 OTUS, metabolome with 676 metabolites, and lipidome with 709 lipids. The model found 15 uncorrelated latent factors (Fig S5), i.e., combinations of features at the multi-omic level. The total variance was explained at 80% by the lipidome, 22% by the

metabolome, and 2% by the microbiome, agreeing with the SNF analysis (Fig 3A). No factor explained most of the variance in the three views (Fig 3B). After selecting features with the largest weight in each cluster-associated factor (Fig 3C). Features with the most importance based on the top 10% of absolute weight were selected in each view, resulting in 396 features (263 lipids, 111 metabolites, and 22 OTUs). A good cluster separation based on hierarchical clustering of Spearman correlation confirmed the relevance of this subset of features (Fig 3D). We also extracted the top 20 features for each view based on this subset (Fig 3E). Bacteroides and Firmicutes were found in the phylum with the highest weight confirming our results from microbiome analysis and the importance of these microbial communities for cluster separation. Nevertheless, the microbiome had a lower weight than metabolites and lipids in MOFA factors. Among the top 20 metabolite features, three metabolites derived or partially derived from microbiota (MDM) (3,4−dihydroxybutyrate, 2−oxindole−3−acetate, and indoleacetylglutamine) were found (Fig 3E). To investigate the coordinated role of MDM, we performed the consensus association analysis (Fig S6). To balance the different number of features in each of the three omics, we randomly selected 241 metabolites, 241 lipids, and 241 OTUs 1000 times. Significant pairwise correlations (FDR<10^{-6}) found in >90% of comparisons were used to build a positive co-expression network, and community detection was performed, resulting in a network with 1324 nodes, (694 lipids, 536 metabolites, 94 microbial communities), 131863 edges and eight multi-omic communities (N > 30) (Fig S11). To refine this network, we selected the 396 features based on MOFA differing the most clusters (Fig 3D) in the co-expression network (Fig 3F). The most central communities (Average degree C1=444, Average degree C2=364) were lipid specific (SNF-1, lipids=122/124, SNF-2, lipids=127/128), while metabolites enriched communities were sparser with a lower average degree (C3=26, C4=22, C6=10, C7=6) but still connected to lipids with 86 edges between lipids and metabolites. Microbiome enriched community (c8) did not correlate with metabolites or lipids. However, eight MDMs were found in the network, mostly in c6 (5/21), showing that MDMs were highly intercorrelated and can have a potential role in shaping the systemic metabolic and lipid profile.

MDM is highly associated with clinical features driven by bile acid metabolism and indole derivatives

We observed a high correlation among the MDMs (Fig 3F). Therefore, to further investigate their the role in PLWH, we retrieved 69 metabolites defined as, i) produced by intestinal bacterial mainly part of secondary bile acid metabolism (n=22) and ii) produced by host modified by bacteria (n=47, polyamines, propionate, acetate, butyrate, and indole derivatives) as reported (Table S7) (Postler & Ghosh, 2017). Differential abundance analysis 19 MDMs differed between HC and PLWH irrespective of the SNF clusters (Fig 4A). The

propionate and indole derivates were significantly (FDR<*0.05*) increased in PLWH compared to HC. As observed in the whole metabolomics profile, SNF-3 had a more similar profile to HC than SNF-1 and SNF-2, while SNF-1 and SNF-2 had identical profiles. We performed univariate linear regression to investigate the link between microbiome-derived metabolites and clinical parameters (Table S8). Lithocholate sulfate was associated with obesity-related comorbidities (MetS, SAT, VAT, Hypertension, central obesity) and 2-aminobutyrate and deoxycholic acid 12-sulfate. Several lifestyle parameters impacted MDM, such as poultry and vegetable intake, smoking, and alcohol. The use of medication as antihypertensives was also associated with three MDMs. Glycolithocholate and glycoursodeoxycholic acid sulfate were linked to HIV-related parameters (CD4 nadir, CD4 at study entry) and patients' demography and lifestyle parameters. The SNF cluster was linked to lithocholate sulfate, 3-ureidopropionate, and imidazole propionate (Fig 4B). Finally, to measure the influence of MDM on plasma metabolomics profile, we performed association analysis and community detection on metabolomics data only (Fig 4C). We obtained a co-expression network with 843 nodes and 15490 edges (FDR<0.02) and observed seven communities (c1-c7) (Fig 4C). The c4 contained all the secondary bile acid metabolites. Though the differential abundance analysis did not show all MDM differences between the SNF clusters and HC, they were highly correlated in PLWH, with significant MDMs differing between the groups (Fig 4D). Combining all the data, we showed an essential role of MDMs in the system-level metabolic profile of PLWH on successful therapy.

DISCUSSION:

In this study, we used network and factorization-based integrative analysis of plasma metabolomics, lipidomics, and microbiome profile to characterize clinical phenotypes in the $PLWH_{ART.}$ We identified three different diseases' state -omics phenotypes (SNF-1 to SNF-3) within $PLWH_{ART}$ driven by metabolomics, lipidomics, and microbiome that a single omics or clinical features could not explain. The integrative omics highlighted the importance of highly intercorrelated microbiome-derived metabolites and their association with the clinical parameters in $PLWH_{ART}$ clusters separation shaping their systemic health profile. The cluster SNF-2 has the at-risk metabolic profile characterized by an increase in TAG and DAG, highest median BMI, MetS incidence, VAT, and SAT, but had a higher CD4 T-cell count at sample collection compared to SNF-1 and SNF-3, which displayed an HC like lipidomic profile. However, SNF-1 and SNF-2 had a similar metabolic profile differing from HC, with dysregulation of AA metabolism. At the microbiome profile, SNF-1 had a lower α-diversity, a lower proportion of MSM, and was enriched in Bacteroides. In contrast, in SNF-2 and SNF-3, there was an increase in Prevotella, with a high proportion of MSM confirming the influence

of sexual orientation on the microbiome profile (Noguera-Julian *et al.*, 2016). Our study thus identified a risk group of PLWH with successful treatment with a dysregulated metabolic profile potentiate metabolic diseases that could be barriers to healthy aging.

Similarity network analysis reduces the high-dimensional nature and different variance of multi-omics data to group patients based on the most similar profile (Wang *et al.*, 2014). One of the main advantages of this method is the possibility to compare the networks' similarities to find out which layer has the most similarity with the final network. The similarity network fusion-based patient stratification has been used primarily in non-communicable diseases like cancer [to identify cancer subtypes (Chierici *et al*, 2020; Wang *et al.*, 2014) and prognosis(Wang *et al*, 2021)], respiratory diseases (Narayana *et al*, 2021) and to study the influence of diet on human health (Burton-Pimentel *et al*, 2021). Recently we developed SNF-based patient stratification by integrating transcriptomics and metabolomics to define disease severity in COVID-19 that are predictive of the most robust biological features (Ambikan *et al*, 2022). We also reported the influence of gut microbiota on the systemic metabolic profile associated with disease severity (Albrich *et al*, 2022). However, no data were presented to stratify the $PLWH_{ART}$ to fingerprint their disease status. The SNF has shown that the most crucial omics layer in cluster separation was lipids (NMI=0.6), supported by the MOFA analysis. A study reported that ART and HIV reservoirs are responsible for changes in adipose tissue and lipids metabolism in PLWH (Lagathu *et al*, 2019). Dyslipidemia represents the increase in triglycerides, low-density lipoprotein cholesterol (LDL-C), total cholesterol (TC), and decrease of high-density lipoprotein cholesterol (HDL-C) cholesterol in the blood is a well-recognized complication observed in PLWH both naïve (Wang *et al*, 2016) and after ART initiation leading to cardiovascular diseases and mortality (Bowman & Funderburg, 2019; Fiseha *et al*, 2021). We found that the SNF-2 cluster (44/97) had most lipids classes upregulated, especially TAG, DAG, and CER, compared to the other groups, while SNF-3 and SNF-1 had no difference with HC. SNF-2 also has more patients with high BMI, VAT, SAT, and incidence of MetS. DAG and TAG high levels have been linked to cardiovascular events (Bowman & Funderburg, 2019; Stegemann *et al*, 2014). The TAG levels have been linked to insulin resistance and increased diabetes risk (Bowman & Funderburg, 2019), confirming this cluster group's qualification as patients with dysregulated lipid profiles and metabolic disease risk. The association of lipid profile with CD4 counts is still debated. It is positively associated (Fiseha *et al.*, 2021; Ji *et al*, 2019), and negatively (Ombeni & Kamuhabwa, 2016) associated with the high abundant lipid profile. Interestingly, we found the SNF-2 cluster to have the highest CD4 count and suppressed viremia but have dysregulated lipid profile that could be reasoned for unhealthy aging and adverse cardio-metabolic health. Therefore, we propose using a holistic view to

define the clinical and immunological treatment success of $PLWH_{ART}$ beyond viral suppression and immune reconstitution.

The second omics defining clusters were metabolites (NMI=0.4). Interestingly, the metabolic profile was not completely overlapping with the lipid profile showing the complexity associated with the disease. $PLWH_{ART}$ in the SNF-1 cluster were the most different from the HC regarding their healthy-like clinical parameter with the lowest BMI, VAT, and SAT. Nevertheless, 32% of $PLWH_{ART}$ SNF-1 had MetS, which was half of SNF-2 (70%) but double SNF-3 (17%), indicating a possible lipid independent metabolic dysregulation. Still, the SNF-3 had the profile of the most HC-like, similarly to the lipids, despite having a significantly higher number of patients with Hypertension than SNF-1. SNF-1 and SNF-2 showed an up-regulation of the metabolites in the xenobiotics, nucleotides, and AA metabolism, indicating a potential role of diet. We previously showed that the glutamate metabolism was highly disrupted in $PLWH_{ART}$ with MetS in the same COCOMO cohort (Gelpi *et al.*, 2021), which can be responsible for late immune recovery in short-term ART patients (Rosado-Sánchez *et al*, 2019). Also, short-chain dicarboxylacylcarnitines (SCDA) and glutamine/valine were higher in PLWH with coronary artery disease than in controls (Okeke *et al*, 2018). In our cohort, we observed glutamate, N-acetyl-glutamate, phenyl-acetyl-glutamate, gamma-glutamylglutamate, and 4-hydroxyglutamate to be upregulated between SNF-2 and SNF-1 than HC and SNF-3. 4-hydroxyglutamate was increased in SNF-2 compared to SNF-1.

The microbiome network had a modest similarity with the final SNF network (NMI=0.3), and clustering was not observed on the PCA plot. Metabolism and immunity of the host have been shown to be affected by bacteria and disrupted microbiome linked to illness (Sun *et al*, 2016). More importantly, there is a high variability of microbiota among individuals based on lifestyle, diet, medication, and physiology (Knight *et al*, 2018). Increased α-diversity is associated with good health and decreased diversity in several diseases, including HIV (Zhou *et al.*, 2020). A meta-analysis reported that HIV status was not associated with decreased α-diversity in MSM, perhaps due to sexual behaviors, but was decreased in PLWH with heterosexual transmission(Tuddenham *et al*, 2020). Despite having healthy clinical and metabolic profiles, we observed an increase in α-diversity in SNF-2 compared to SNF-1 driven by MSM and no differences between SNF-1 and SNF-3. In terms of bacterial composition, studies reported that PLWH has a higher abundance of Prevotella and a lower abundance of Bacteroides(Neff *et al*, 2018). The crucial point is that the ratio is associated more with MSM than HIV status and has been shown by several studies (Gelpi *et al.*, 2020; Noguera-Julian *et al.*, 2016; Zhou *et al.*, 2020). Our study observed that SNF-2 was enriched in Prevotella and depleted in Bacteriodes compared to SNF-1. Interestingly, the decrease of

Bacteroides in obese patients was inversely correlated with serum glutamate (Wu *et al*, 2021), which was also observed in SNF-2 patients. On the other hand, some Prevotella species have proinflammatory effects, leading to intestinal inflammation, bacterial translocation, and microbiome dysbiosis (Iljazovic *et al*, 2021). In general, the cohort is mainly composed of MSM (63/97). As described above, it confirmed that the difference in the microbiome is driven by MSM status in SNF-2, as SNF-2 patients have 81% MSM. SNF-3, even if there is no difference with SNF-2 according to PERMANOVA, has the same proportion of MSM as SNF-1. It has been proposed that early regulation of the MSM-related microbiome could help prevent HIV infection (Zhou *et al.*, 2020). However, the question remains whether the MSM-related microbiome is a potential driving force of metabolic comorbidities or whether MSM is a confounding factor disturbing a potentially clinical signal from a disturbed microbiome.

Microbial compositions have implications for metabolism and metabolic diseases, notably through the production of MDMs (Agus *et al*, 2021). Secondary bile acids transformed from primary bile acids by bacteria have a role in lipid digestion. It regulates host metabolism through signaling and can inhibit the production of proinflammatory cytokines by immune cells (Postler & Ghosh, 2017). Lipid metabolism, including triglyceride trafficking, is influenced by bile acids through the interaction with the Farnesoid X receptor (FXR) receptor and has been implicated in mice's metabolic disorder (Schoeler & Caesar, 2019). A bile acid, glycolithocholate was found upregulated in SNF-1 and SNF-3 compared to controls shown previously associated with insulin resistance (Diboun *et al*, 2021). It was highly negatively associated with food elements such as vegetable intake and choice of fat for cooking, alcohol, and HIV-related parameters such as CD4 levels (nadir and at ART initiation) and HIV duration. High glycodeoxycholate was observed in SNF-1 especially compared to SNF-3, while the glycodeoxycholic acid has been shown to be negatively associated with insulin resistance (Wu *et al.*, 2021). Glycocholenate sulfate was down-regulated in the three clusters compared to controls. All secondary bile acids were shown to be highly intercorrelated in co-expression analysis. Three other bile acids, lithocholate sulfate, glycousodesoxycholic acid sulfate, and deoxycholic acid 12-sulfate, were negatively associated with metabolic perturbations including MetS, VAT, and central obesity. Acetate, propionates, and butyrate are part of short-chain fatty acids (SCFAs) and are obtained from the fiber bacterial fermentation in the colon that the host's enzymes cannot digest (Alwin & Karst, 2021). Proprionate derivates were upregulated in SNF-1 and SNF-2. Acetate and butyrate derivates had a more variable profile. Imidazole propionate (IMP) and 3-ureidopropionate were linked to the SNF clusters. In our study, the IMP was also linked to vegetable intake, which was reported to be involved in insulin resistance(Agus *et al.*, 2021).

The Bacteroides metabolize most of the acetate and propionate from polysaccharides, and Firmicutes produce butyrate (Postler & Ghosh, 2017), which does not explain the relationship within the SNF clusters indicating a more complex interplay between the MDMs and bacterial community in a diseased condition. Tryptophan is converted by bacterial tryptophanase into indole, and indole derivates are involved in host-microbiota homeostasis (Krautkramer *et al*, 2021). Indoles derivates mainly were upregulated in SNF-1 and SNF-2. Our data thus suggested the role of MDMs in shaping the clinical phenotype and systemic health profile in $PLWH_{ART,}$ which could be a therapeutic target for improving health.

Though our study is the first to demonstrate an integrative multi-omics approach to the role of MDMs in systemic alterations in $PLWH_{ART}$, our study has limitations that merit comments. First, the study is cross-sectional and therefore restricted to predicting dynamic interactions of different omics layers. Second, the microbiome data analysis is through 16S methodologies with a higher level of the missing data at the genus and species level than metagenomics. Third, although the network-based analysis and the observational data suggest a potential causal association of altered metabolic profile with clinical features, other factors may drive observed effects. Fouth, though this is the largest study to date to perform integrative omics in PLWH, the number of samples were relatively low. Finally, both microbiome and metabolomics are highly dependent upon the genetics, environment, and diet of an individual. The interaction noted may characterize the epiphenomena of a personalized immune system that can be an avenue for future studies to develop a more personalized model for integrative omics to phenotype the disease states that we recently reported (Ambikan *et al.*, 2022).

In conclusion, we performed a multi-omics analysis of $PLWH_{ART}$ with different clinical features. We identified the diversity of $PLWH_{ART}$ in HIV-related biological alterations regardless of immunological recovery and virological suppression. A proportion of $PLWH_{ART}$ (SNF-2; around 45% in the present cohort) showed highly dysregulated lipidomics (increased TAG and DAG) and clinical profile (increased BMI and obesity-related features) with increased Prevotella and decreased Bacteroides, the latter being related to MSM transmission. However, alterations in the metabolomics profile and higher CD4 T-cell count at the time of sample collection indicate a complex systemic interplay between host immunity and metabolic health that might affect healthy aging in this population. Integrative analytical approaches that reflect the overall systemic health profile of $PLWH_{ART}$ may improve patient stratification and individual therapeutic and preventive strategies. Developing a more personalized model or targeting the interaction networks rather than individual clinical or

omics features may provide novel treatment strategies in countering dysregulated metabolic traits, aiming to achieve healthier aging.

Methods:

Patient Cohort and Multiomics data: The cohort comprises 97 $PLWH_{ART}$ from the Copenhagen Comorbidity (COCOMO) Cohort, a prospective cohort of $PLWH_{ART}$. This study used untargeted metabolomics (Gelpi *et al.*, 2021), complex lipid profile (Olund Villumsen *et al.*, 2021), and 16s microbiome data (Gelpi *et al.*, 2020) reported earlier for the larger cohorts. We also extracted clinical and demographical data from the COCOMO database. The HIV-negative controls (HC) (n=20) were used to understand the basal level of omics.

Similarity network fusion (SNF): Lipids and metabolites with low variance (<0.3) were removed from the data. The three layers of omics (microbiome, lipidome, metabolome) were standard normalized before analysis. Analysis was processed using the package SNFtool as described (Wang *et al.*, 2014). Pairwise samples distances were calculated with the function dist2 followed by the construction of similarity graphs (number of neighbors, K=13, hyperparameter, alpha=0.8) for each layer. The similarity network fusion (SNF) was used to fuse all the networks (K=13, number of iterations, T=10). Spectral clustering was applied to the fused network (Number of clusters C=3). The parameters (K, alpha, T, C) were chosen to maximize Eigen gap and minimize rotation cost. A concordance matrix was calculated based on the similarity in cluster assignments in each network, and the fused network was calculated in normalized mutual information (NMI).

Clinical data statistics: Clinical characteristics were compared between clusters using pairwise tests. Welch's T-test and Mann-Whitney U tests were used to compare normally distributed and non-normally distributed continuous variables were compared using R (Team, 2010). Chi-Square Test was used to compare discrete variables if the expected values of the contingency table were five or more. Otherwise, Fisher's Exact Test was used. Univariate linear regression was performed with the function lm, and correlations were calculated using the cor.test and cor functions from the R stat package (Team, 2010).

Lipidomics and metabolomics analysis: Untargeted metabolomics and lipidomics were log2 transformed before analysis was performed in Metabolon™, USA. Lipid data were grouped by lipid classes as in the following.

$$[Class_j] = \sum_{i=1}^{n} [species_i]$$

$$[Class_j] = \text{Concentration of the lipid class j}$$

$$[species_j] = \text{Concentration of the molecular species i}$$
$$\text{n} = \text{number of molecular species of a class j}$$

Differential abundance analysis was performed pairwise with the R package limma between groups (HC, SNF-1, SNF-2, SNF-3) for each data set (lipidomics classes, lipidomics, metabolomics). Benjamini-Hochberg (BH) adjustment was applied, and FDR was set up to 0.05 for statistical significance. Deviations were mentioned in the respective analysis. The default p-value cutoff was set to 0.05. Other p-values cutoffs are adapted for a specific analysis depending upon the number of significance and to minimize the false positivity.

Microbiome analysis: Microbiome data analysis was performed using the R package phyloseq (McMurdie & Holmes, 2013). Alpha diversity estimates were calculated using the estimate_richness function and the following measures: Observed, ACE, se.ACE, Chao1, Shannon, Simpson, InvSimpson, and Fisher. NMDS ordinations based on Bray-Curtis distances between all samples were calculated using the ordinate function. Otu table was converted to relative abundances for further analysis. The Vegan package (Jari Oksanen *et al*, 2020) was used to perform PERMANOVA. Equal multivariate dispersion was verified using the betadisper function applying Marti Anderson's PERMDISP2 procedure. Pairwise PERMANOVA test was done between groups using the adonis function, Bray distance, and bonferroni correction. The cutoff for the adjusted p value was set up to 0.05. Galaxy module LDA Effect Size (LEfSe) was used to find microbial communities (at genus, family or higher level) specific to one specific cluster (Segata *et al*, 2011). The multiclass analysis approach was one against all. First, non-parametric factorial Kruskal-Wallis (KW) sum-rank test was performed with clusters (cutoff alpha=0.05), followed by pairwise Wilcoxon rank-sum tests between clusters (cutoff alpha=0.05), and then effect size calculation for each significant feature was done using discriminant analysis (absolute LDA score>2). Results are represented using a cladogram produced by the module.

Microbiome-derived metabolites: Microbiome-derived metabolites, groups, and subgroups were retrieved from the previous literature (Postler & Ghosh, 2017).

Multi-omics Factor analysis (MOFA): Filtered data for SNF was also used for MOFA analysis (Argelaguet *et al.*, 2018). Microbiome data were rarefied by filtering based on variance (>0.2). In addition, the microbiome data were center log-ratio (CLR) transformed to follow a normal distribution. The MOFA model was trained using default parameters, and sample metadata was added to the model. Total variance explained per view was used to

see the weight of each omics layer. A correlation plot was used to verify the low correlation between factors. A variance decomposition plot was used to determine the percentage of variance explained by each factor and omics layer. Association analysis of the factors with clinical features was done using MOFA function correlate_factors_with_covariates and factors associated with the SNF cluster selected. 5 and 95 % quantile weights for each view were selected for each factor. Pathway analysis was performed on factors using the MOFA function run_enrichment for each view, with the parametric statistical test, FDR-adjusted p-values, and separated positive and negative values. Annotation libraries were made from Metabolon™ super pathways for metabolomics and lipidomics and Division level for the microbiome.

Co-expression analysis: Pairwise Spearman correlations between features were calculated, and the cutoff for FDR of significant correlations was selected to minimize the number of false positives. The positive and negative networks were built using python igraph (Nepusz, 2005) and compared to random networks of the same size. Leiden community detection was applied to find groups of interconnected features, and the mean degree was calculated to represent community centrality (Blondel *et al*, 2008). Communities of less than 30 features were excluded. Consensus association analysis was performed to integrate the three layers of omics using 1000 iterations. At each iteration, pairwise correlations between OTUs (N=241), 241 metabolites and 241 lipids selected randomly were run, and significant positive correlations (Spearman, FDR<0.001) were kept as an association. Associations found in 90% of the comparisons over all iterations were kept building the final network as described above.

Visualization: Scatter plots, PCA plots, box plots, NMDS plots, circular heatmap, and bar plots were generated using ggplot2 (Wickham, 2016). Heatmaps were generated using ComplexHeatmap (Gu *et al*, 2016). Sankey plot was made using the R package ggalluvial (Brunson, 2020). Networks were plotted using Cytoscape v3.6.1 (Shannon *et al*, 2003).

ETHICAL CLEARANCES:

Ethical approval was obtained by the Regional Ethics Committee of Copenhagen (COCOMO: H-15017350) and Etikprövningsmyndigheten, Sweden (Dnr: 2022-01353-01). Informed consent was obtained from all participants and delinked before analysis.

DATA AVAILABILITY

The metabolomics and lipidomics data are available from 10.6084/m9.figshare.14356754 and 10.6084/m9.figshare.14509452. All the codes are available at github: https://github.com/neogilab/HIV_multiomics

ACKNOWLEDGMENTS

The study is funded by the Swedish Research Council grants 2017-01330, 2018-06156, and 2021-01756 to UN. Novo Nordic Foundation, Lundbeck Foundation, Augustinus Foundation, Region Hovedstaden, and Rigshospitalet Research Council to SDN.

AUTHOR CONTRIBUTIONS

Conceptualization: UN, FM, RB; Clinical Study Design: SDN, TB, MG, ADK; Methodology: FM, MG, RB, ADK, AM, MT; Investigation, clinical: SDN, TB, MG, ADK; Investigation, computational and laboratory: FM, MG, RB, ADK, BV, JH, JRH; Visualization: FM, UN; Funding acquisition: UN, SDN; Project administration: SDN, RB, UN; Supervision: SDN, RB, UN; Writing – original draft: FM, UN; Writing – review & editing: MG, RB, ADK, BV, JH, JRH, TB, DM, CGG, MT, SDN

DECLARATION OF INTERESTS

The authors declare that they have no competing interests.

FIGURE LEGENDS

Figure 1: Similarity network fusion based on lipidomics, metabolomics, and microbiome integration. (a) Scatter plot showing the maximization of Eigen gap and the minimization of rotation cost for the optimization of the number of clusters. **(b)** Concordance matrix between the combined network (SNF) and each omics network based on NMI calculation (0=no mutual information, 1=perfect correlation). **(c)** SNF-combined similarity network colored by clusters (SNF-1=blue, SNF-2=yellow, SNF-3=grey) obtained after spectral clustering. Edges' color indicates the strength of the similarity (black=strong, grey=weak). **(d)** PCA plot of samples based on fused network. Samples are colored by condition (SNF-1=blue, SNF-2=yellow, SNF-3 = grey). **(e)** Boxplots of DAG from untargeted lipid classes separated by groups. Significant stars are displayed for each comparison with *FDR <0.05, **FDR<0.01, ***FDR<0.001 (LIMMA). **(f)** Boxplots of TAG from untargeted lipid classes separated by groups. **(g)** PCA plot of samples after prior standardization based on significant metabolites between at least one pairwise comparison (LIMMA, FDR<0,05). Variance proportions are written on each component axis. Samples are colored by condition (Ctrl=green, SNF-1=blue, SNF-2=yellow, SNF-3=grey). **(h)** Circular heatmap of the top 159 metabolites (FDR<0.005). Metabolites are represented as slices and labeled around the plot. LogFC (LIMMA, FDR<0.05) from significant metabolites between groups are displayed in the first six outer layers. The 7th to 9th layers represents the coefficient of correlation between metabolites and BMI, metabolites respectively, and age (Spearman, p-value<0.1, R>0.15)

and the p-value from significant associations between metabolites and gender (Chi-squared, p value<0.1). The inner layer represents the pathway of each metabolite.

Figure 2: Transmission mode drove cluster differences in microbiome data (a) Boxplots of alpha diversity indices (Observed, ACE, Chao1, Fisher) separated by HIV cluster. Significant stars are shown for each comparison (Mann Whitney U test). **(b)** Non-metric multidimensional scaling (NMDS) plot of Bray-Curtis distances. Samples are colored by clusters. Boxplots based on NMDS1 and NMDS2 are represented. **(c)** Barplot representing the relative abundance of bacteria at the family level for each patient. Patient information are displayed above the barplot, including cluster, metabolic syndrome (MetS: yes/no), Hypertension (yes/no), transmission mode, and gender. **(d)** Barplot showing the top microbial families by representing their coefficient from PERMANOVA between SNF-1 and SNF-2. **(e)** Barplot showing the top microbial families between SNF-1 and SNF-3. **(f)** LEfSe cladogram representing cluster-specific microbial communities to SNF-1 and to SNF-2/SNF-3. Top families from PERMANOVA are labeled. **(g)** Boxplot of relative abundance at family level for Bacteroides (top) and Prevotella (bottom). Significant stars are shown for significant comparisons (Mann Whitney U test).

Figure 3: Factor analysis highlights the essential features for clusters separation and potential microbiome-derived metabolites importance (a) Barplot of total variance explained by MOFA model per view. **(b)** Variance decomposition plot. The percentage of variance is explained by each factor for each view. **(c)** External covariate association with factors plot. Association is represented with log10 adjusted p-values from Pearson correlation. **(d)** Heatmap representing levels of microbial communities, metabolites, and lipids with the higher absolute weight in MOFA factors associated with cluster (F1, F2, F3, F5, F8). Samples are labeled according to the study group (SNF-1, SNF-2, SNF-3). Data were Z-score transformed. Type of data (lipid, metabolite, microbe) is displayed on the right. **(e)** Top 20 features with higher absolute weight in MOFA factors associated with cluster (F1, F2, F3, F5, F8) from lipidome, metabolome, and microbiome. Microbiome-derived metabolites and bacterial phylum of interest are colored in blue and red, respectively. **(f)** MOFA features differing clusters and interactions extracted from the 3-layers consensus co-expression network. Microbiome-derived metabolites are labeled.

Figure 4: Microbiome-derived metabolites are affected in HIV clusters (a) Heatmap representing abundances of microbiome-derived metabolites differing at least one comparison. Data were Z-score transformed. Significant logFC (LIMMA, FDR < 0.05) of pairwise comparisons between conditions, groups, and under groups of microbiome-derived

metabolites are displayed on the right. **(b)** Cytoscape network showing significant positive and negative associations between clinical parameters and microbiome-derived metabolites (univariate linear regression, FDR<0.05). Clinical parameters are colored based on categories. **(c)** Co-expression network of metabolomics data in PLWH. Metabolites are grouped by communities, and microbiome-derived metabolites are labeled, and colored based on the subgroup. **(d)** The subset of microbiome-derived metabolites from the co-expression network. Non-significant metabolites in all comparisons are displayed with transparency. Significant microbiome-derived metabolites between at least two conditions are labeled.

References:

Agus A, Clement K, Sokol H (2021) Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut* 70: 1174-1182

Albrich WC, Ghosh TS, Ahearn-Ford S, Mikaeloff F, Lunjani N, Forde B, Suh N, Kleger GR, Pietsch U, Frischknecht M *et al* (2022) A high-risk gut microbiota configuration associates with fatal hyperinflammatory immune and metabolic responses to SARS-CoV-2. *Gut microbes* 14: 2073131

Alwin A, Karst SM (2021) The influence of microbiota-derived metabolites on viral infections. *Curr Opin Virol* 49: 151-156

Ambikan AT, Yang H, Krishnan S, Svensson-Akusjärvi S, Gupta S, Lourda M, Sperk M, Arif M, Zhang C, Nordqvist H *et al* (2022) Multiomics Personalized Network Analyses Highlight Progressive Disruption of Central Metabolism Associated with COVID-19 Severity. *Cell Systems* (in press)

Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, Buettner F, Huber W, Stegle O (2018) Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol* 14: e8124

Babu H, Sperk M, Ambikan AT, Rachel G, Viswanathan VK, Tripathy SP, Nowak P, Hanna LE, Neogi U (2019) Plasma Metabolic Signature and Abnormalities in HIV-Infected Individuals on Long-Term Successful Antiretroviral Therapy. *Metabolites* 9

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008

Bowman E, Funderburg NT (2019) Lipidome Abnormalities and Cardiovascular Disease Risk in HIV Infection. *Curr HIV/AIDS Rep* 16: 214-223

Brunson J (2020) ggalluvial: Layered Grammar for Alluvial Plots. *Journal of Open Source Software* 5

Burton-Pimentel KJ, Pimentel G, Hughes M, Michielsen CC, Fatima A, Vionnet N, Afman LA, Roche HM, Brennan L, Ibberson M *et al* (2021) Discriminating Dietary Responses by Combining Transcriptomics and Metabolomics Data in Nutrition Intervention Studies. *Mol Nutr Food Res* 65: e2000647

Chierici M, Bussola N, Marcolini A, Francescatto M, Zandona A, Trastulla L, Agostinelli C, Jurman G, Furlanello C (2020) Integrative Network Fusion: A Multi-Omics Approach in Molecular Profiling. *Front Oncol* 10: 1065

Deeks SG (2011) HIV infection, inflammation, immunosenescence, and aging. *Annual review of medicine* 62: 141-155
Diboun I, Al-Mansoori L, Al-Jaber H, Albagha O, Elrayess MA (2021) Metabolomics of Lean/Overweight Insulin-Resistant Females Reveals Alterations in Steroids and Fatty Acids. *J Clin Endocrinol Metab* 106: e638-e649
Fiseha T, Alemu W, Dereje H, Tamir Z, Gebreweld A (2021) Prevalence of dyslipidaemia among HIV-infected patients receiving combination antiretroviral therapy in North Shewa, Ethiopia. *PLoS One* 16: e0250328
Gelpi M, Afzal S, Lundgren J, Ronit A, Roen A, Mocroft A, Gerstoft J, Lebech AM, Lindegaard B, Kofoed KF *et al* (2018) Higher Risk of Abdominal Obesity, Elevated Low-Density Lipoprotein Cholesterol, and Hypertriglyceridemia, but not of Hypertension, in People Living With Human Immunodeficiency Virus (HIV): Results From the Copenhagen Comorbidity in HIV Infection Study. *Clin Infect Dis* 67: 579-586
Gelpi M, Mikaeloff F, Knudsen AD, Benfeitas R, Krishnan S, Svenssson Akusjärvi S, Høgh J, Murray DD, Ullum H, Neogi U *et al* (2021) The central role of the glutamate metabolism in long-term antiretroviral treated HIV-infected individuals with metabolic syndrome. *Aging* 13: 22732-22751
Gelpi M, Vestad B, Hansen SH, Holm K, Drivsholm N, Goetz A, Kirkby NS, Lindegaard B, Lebech AM, Hoel H *et al* (2020) Impact of Human Immunodeficiency Virus-Related Gut Microbiota Alterations on Metabolic Comorbid Conditions. *Clin Infect Dis* 71: e359-e367
Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32: 2847-2849
Iljazovic A, Roy U, Galvez EJC, Lesker TR, Zhao B, Gronow A, Amend L, Will SE, Hofmann JD, Pils MC *et al* (2021) Perturbation of the gut microbiome by Prevotella spp. enhances host susceptibility to mucosal inflammation. *Mucosal Immunol* 14: 113-124
Jari Oksanen FGB, Michael Friendly, Roeland Kindt,, Pierre Legendre DM, Peter R. Minchin, R. B. O'Hara, Gavin, L. Simpson PS, M. Henry H. Stevens, Eduard Szoecs and, Wagner H, 2020. vegan: Community Ecology Package. R package, version 2.5-7 ed.
Ji S, Xu Y, Han D, Peng X, Lu X, Brockmeyer NH, Wu N (2019) Changes in Lipid Indices in HIV+ Cases on HAART. *Biomed Res Int* 2019: 2870647
Karahalil B (2016) Overview of Systems Biology and Omics Technologies. *Curr Med Chem* 23: 4221-4230
Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall LI, McDonald D *et al* (2018) Best practices for analysing microbiomes. *Nat Rev Microbiol* 16: 410-422
Krautkramer KA, Fan J, Backhed F (2021) Gut microbial metabolites as multi-kingdom intermediates. *Nat Rev Microbiol* 19: 77-94
Lagathu C, Bereziat V, Gorwood J, Fellahi S, Bastard JP, Vigouroux C, Boccara F, Capeau J (2019) Metabolic complications affecting adipose tissue, lipid and glucose metabolism associated with HIV antiretroviral treatment. *Expert Opin Drug Saf* 18: 829-840
McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8: e61217
Mikaeloff F, Svensson Akusjarvi S, Ikomey GM, Krishnan S, Sperk M, Gupta S, Magdaleno GDV, Escos A, Lyonga E, Okomo MC *et al* (2022) Trans cohort metabolic reprogramming towards glutaminolysis in long-term successfully treated HIV-infection. *Commun Biol* 5: 27
Narayana JK, Mac Aogain M, Ali N, Tsaneva-Atanasova K, Chotirmall SH (2021) Similarity network fusion for the integration of multi-omics and microbiomes in respiratory disease. *Eur Respir J* 58

Nasi M, De Biasi S, Gibellini L, Bianchini E, Pecorini S, Bacca V, Guaraldi G, Mussini C, Pinti M, Cossarizza A (2017) Ageing and inflammation in patients with HIV infection. *Clin Exp Immunol* 187: 44-52
Neff CP, Krueger O, Xiong K, Arif S, Nusbacher N, Schneider JM, Cunningham AW, Armstrong A, Li S, McCarter MD *et al* (2018) Fecal Microbiota Composition Drives Immune Activation in HIV-infected Individuals. *EBioMedicine* 30: 192-202
Nepusz GCaT, 2005. The Igraph Software Package for Complex Network Research. Budapest, Hungary.
Noguera-Julian M, Rocafort M, Guillen Y, Rivera J, Casadella M, Nowak P, Hildebrand F, Zeller G, Parera M, Bellido R *et al* (2016) Gut Microbiota Linked to Sexual Preference and HIV Infection. *EBioMedicine* 5: 135-146
Okeke NL, Craig DM, Muehlbauer MJ, Ilkayeva O, Clement ME, Naggie S, Shah SH (2018) Metabolites predict cardiovascular disease events in persons living with HIV: a pilot case-control study. *Metabolomics* 14
Olund Villumsen S, Benfeitas R, Knudsen AD, Gelpi M, Hogh J, Thomsen MT, Murray D, Ullum H, Neogi U, Nielsen SD (2021) Integrative Lipidomics and Metabolomics for System-Level Understanding of the Metabolic Syndrome in Long-Term Treated HIV-Infected Individuals. *Front Immunol* 12: 742736
Ombeni W, Kamuhabwa AR (2016) Lipid Profile in HIV-Infected Patients Using First-Line Antiretroviral Drugs. *J Int Assoc Provid AIDS Care* 15: 164-171
Pei L, Fukutani KF, Tiburcio R, Rupert A, Dahlstrom EW, Galindo F, Laidlaw E, Lisco A, Manion M, Andrade BB *et al* (2021) Plasma Metabolomics Reveals Dysregulated Metabolic Signatures in HIV-Associated Immune Reconstitution Inflammatory Syndrome. *Front Immunol* 12: 693074
Pelchen-Matthews A, Ryom L, Borges AH, Edwards S, Duvivier C, Stephan C, Sambatakou H, Maciejewska K, Portu JJ, Weber J *et al* (2018) Aging and the evolution of comorbidities among HIV-positive individuals in a European cohort. *AIDS* 32: 2405-2416
Postler TS, Ghosh S (2017) Understanding the Holobiont: How Microbial Metabolites Affect Human Health and Shape the Immune System. *Cell Metab* 26: 110-130
Rosado-Sánchez I, Rodríguez-Gallego E, Peraire J, Viladés C, Herrero P, Fanjul F, Gutiérrez F, Bernal E, Pelazas R, Leal M *et al* (2019) Glutaminolysis and lipoproteins are key factors in late immune recovery in successfully treated HIV-infected patients. *Clin Sci (Lond)* 133: 997-1010
Schoeler M, Caesar R (2019) Dietary lipids, gut microbiota and lipid metabolism. *Rev Endocr Metab Disord* 20: 461-472
Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* 12: R60
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504
Stegemann C, Pechlaner R, Willeit P, Langley SR, Mangino M, Mayr U, Menni C, Moayyeri A, Santer P, Rungger G *et al* (2014) Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. *Circulation* 129: 1821-1831
Sun Y, Ma Y, Lin P, Tang YW, Yang L, Shen Y, Zhang R, Liu L, Cheng J, Shao J *et al* (2016) Fecal bacterial microbiome diversity in chronic HIV-infected patients in China. *Emerg Microbes Infect* 5: e31
Team TRDC, 2010. R: A Language and Environment for Statistical Computing, 4.1.2 ed. R Foundation for Statistical Computing, Vienna, Austria.

Tuddenham SA, Koay WLA, Zhao N, White JR, Ghanem KG, Sears CL, Consortium HIVMR-a (2020) The Impact of Human Immunodeficiency Virus Infection on Gut Microbiota alpha-Diversity: An Individual-level Meta-analysis. *Clin Infect Dis* 70: 615-627
Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A (2014) Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 11: 333-337
Wang Q, Ding H, Xu J, Geng W, Liu J, Guo X, Kang J, Li X, Jiang Y, Shang H (2016) 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
Wang TH, Lee CY, Lee TY, Huang HD, Hsu JB, Chang TH (2021) Biomarker Identification through Multiomics Data Analysis of Prostate Cancer Prognostication Using a Deep Learning Model and Similarity Network Fusion. *Cancers (Basel)* 13
Wickham H, 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
Wu J, Wang K, Wang X, Pang Y, Jiang C (2021) The role of the gut microbiome and its metabolites in metabolic diseases. *Protein Cell* 12: 360-373
Yoshimura K (2017) Current status of HIV/AIDS in the ART era. *J Infect Chemother* 23: 12-16
Zhou J, Zhang Y, Cui P, Luo L, Chen H, Liang B, Jiang J, Ning C, Tian L, Zhong X *et al* (2020) Gut Microbiome Changes Associated With HIV Infection and Sexual Orientation. *Front Cell Infect Microbiol* 10: 434

Supplementary Tables

Table S1. List of parameters used in the study.
Table S2a. Table of differential lipid abundance analysis on individual lipids abundances.
Table S2b. Table of differential lipid abundance analysis grouped by lipids classes.
Table S3. Table of differential metabolite abundance analysis.
Table S4. Alpha diversity indices statistics.
Table S5. Permutational multivariate analysis of variance at the family level.
Table S6. Microbiome statistics at a family level.
Table S7. List of microbiome-derived metabolites
Table S8. Univariate linear regression between clinical parameters and microbiome-derived metabolites differing groups.

Supplementary Figures

Fig S1. PCA plot of samples after prior standardization based on a) Lipidomics b) Metabolomics c) Microbiome. Variance proportions are written on each component axis. Samples are colored by condition (Ctrl = green, SNF-1 = blue, SNF-2 = yellow, SNF-3 = grey).
Fig S2. Boxplots of untargeted lipid classes separated by groups. Color is based on groups (Ctrl = green, SNF-1 = blue, SNF-2 = yellow, SNF-3 = grey). P values are displayed for each comparison (Mann Withney U Test).

Fig S3. Boxplots of alpha diversity indices (se.chao1,Simpson, Shannon, se.ACE, InvSimpson) separated by HIV-cluster. Color is based on groups (Ctrl = green, SNF-1 = blue, SNF-2 = yellow, SNF-3 = grey).

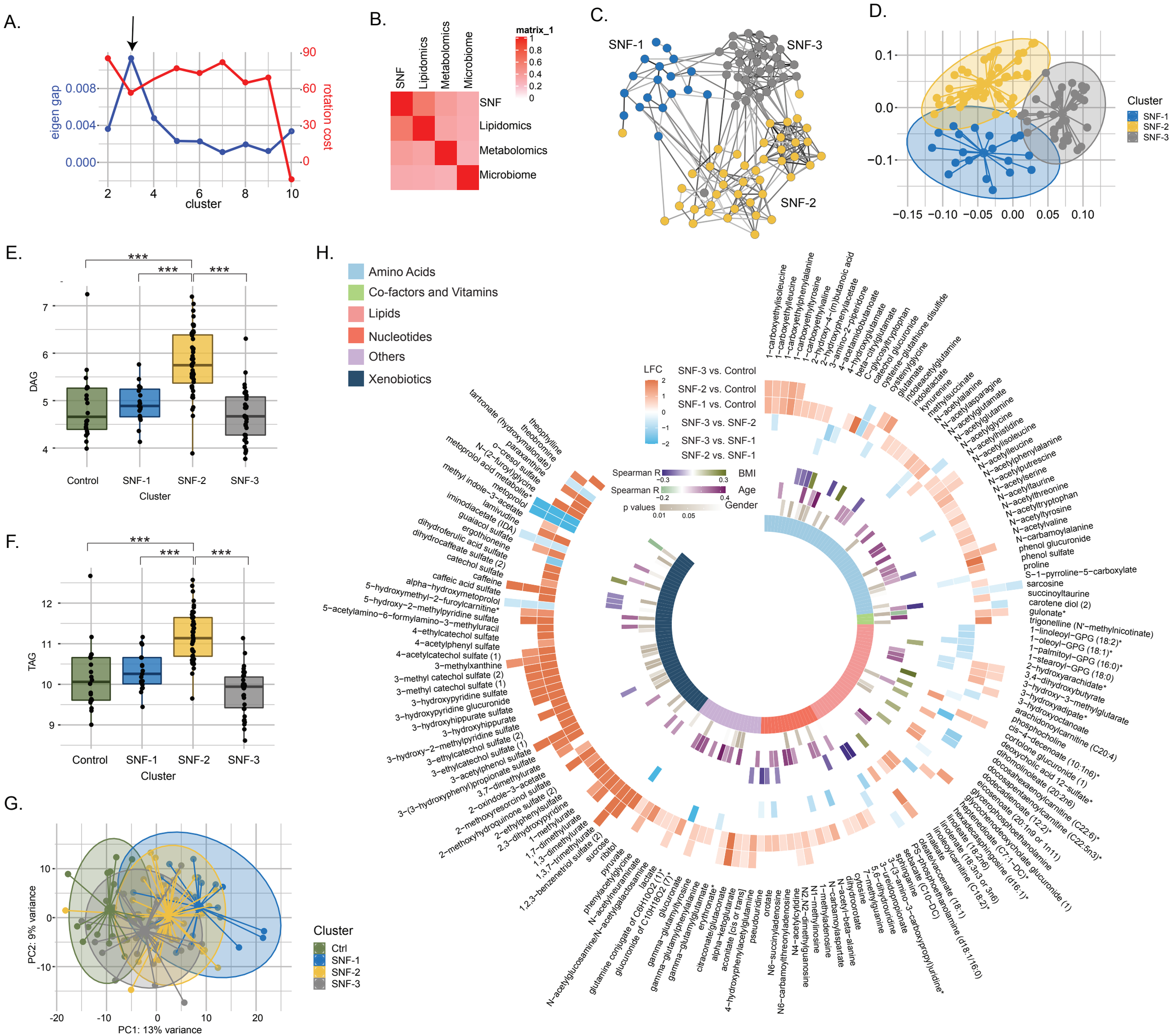
Fig S4. Non-metric multidimensional scaling (NMDS) plot of Bray-Curtis distances. Samples are colored by A) Transmission mode B) Central obesity C) Metabolic Syndrome D) Hypertension.

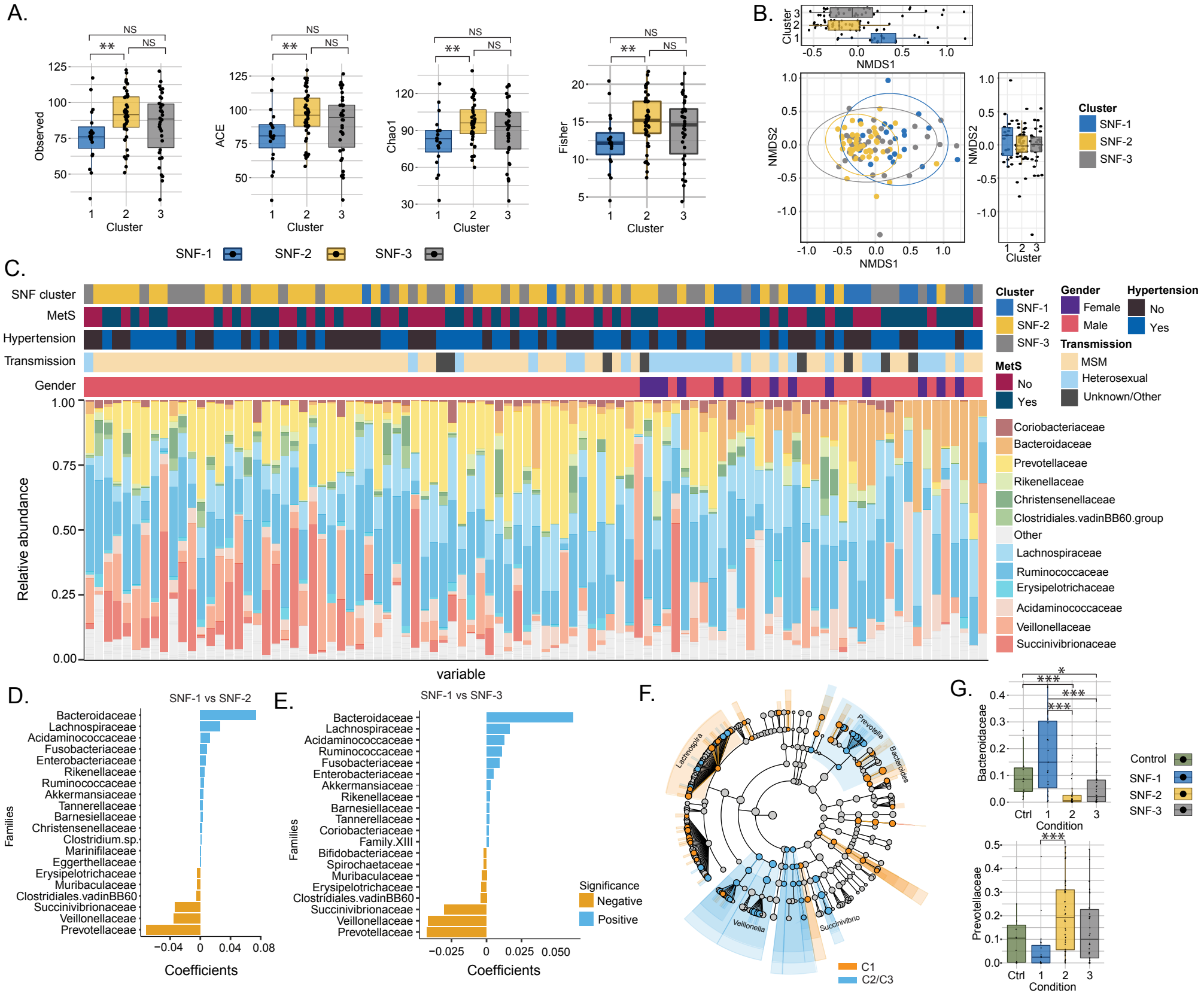
Fig S5. Correlation matrix of MOFA factors. Size and transparency are proportional to absolute coefficient of correlation. Color is displayed as a gradient-based coefficient of correlation from -1 (red) to 1 (blue).

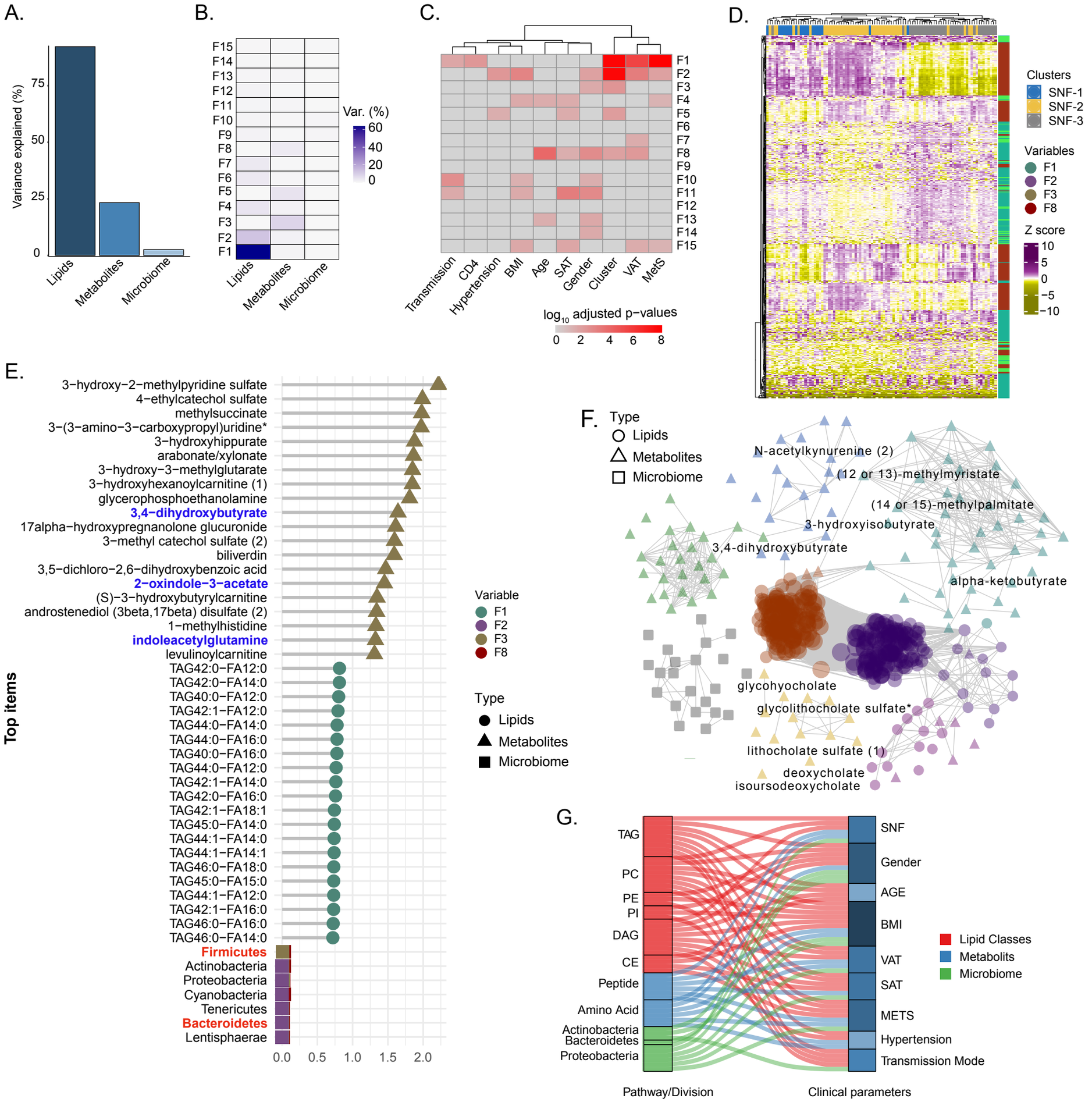
Fig S6. Cytoscape consensus co-expression network. Color and label are based on communities.

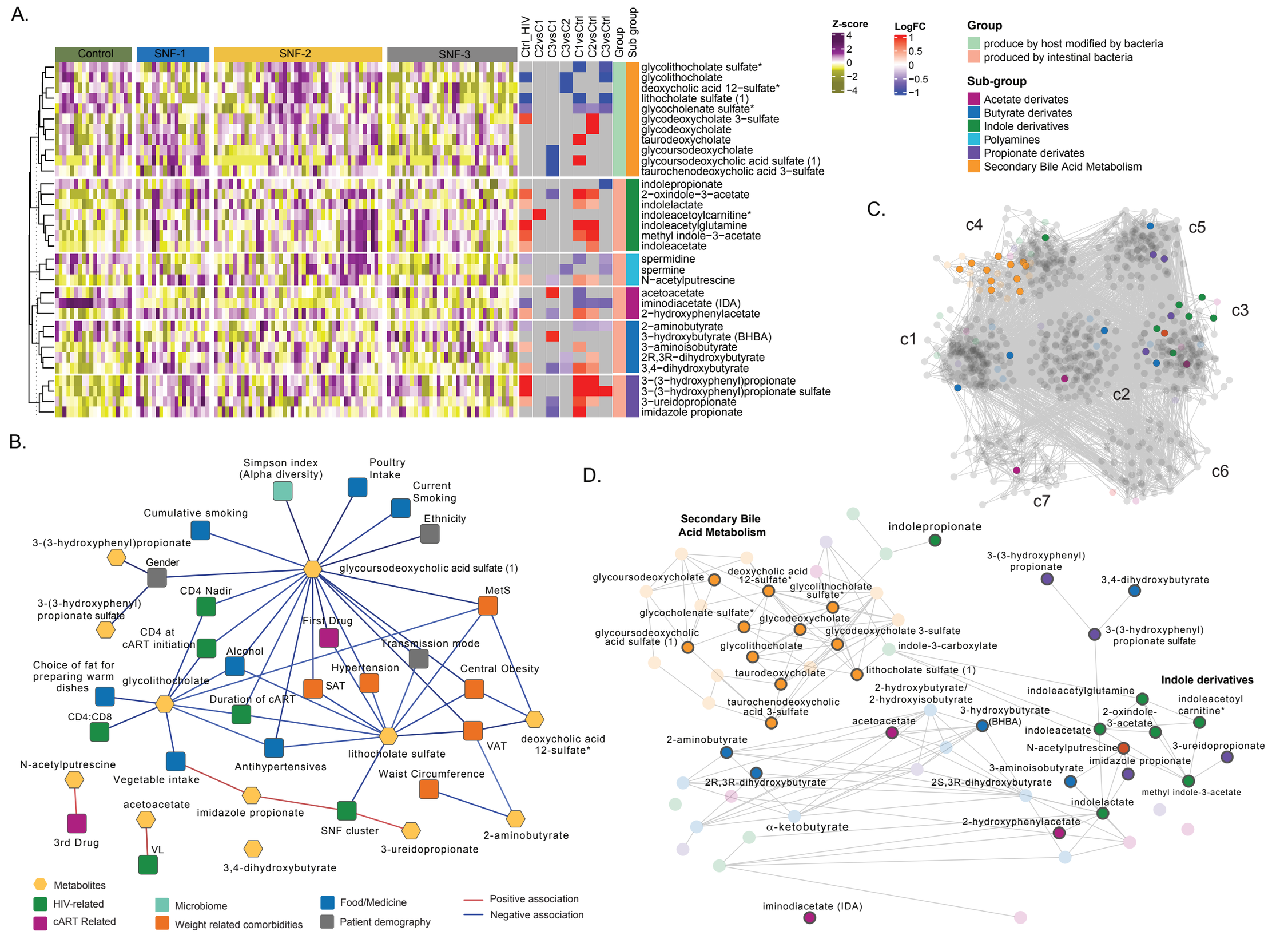
Table 1. Patient characteristics

	Complete Cohort	SNF-1	SNF- 2	SNF-3	P values
N	97	19	44	34	
Age in years, Median (IQR)	54 (48-63)	60 (48-68)	54 (48-62)	54 (51-60)	0.75
Gender, Male, N (%)	84 (87)	15 (79)	40 (91)	29 (85)	0.36
Ethnicity Caucasian, N (%)	79 (81)	15 (79)	38 (87)	26 (77)	0.49
Mode of transmission, N (%)					0.017
Homosexual/bisexual	63 (65)	9 (47)	36 (81)	18 (53)	
Heterosexual	26 (27)	7 (37)	6 (14)	13 (38)	
Other/unknown	8 (8)	3 (16)	2 (5)	3 (9)	
CD4 Nadir, cells/μL, Median (IQR)	235 (123-320)	240 (127-330)	240 (145-365)	223 (42-290)	0.49
CD4 at ART Initiation, cells/μL, Median (IQR)	287 (155-410)	270 (120-360)	318 (192-463)	240 (108-320)	0.11
Viral Load at ART initiation, log copies/mL, Median (IQR)	5.02 (4.34-5.61)	4.87 (4.32-5.5)	5.11 (4.74-5.61)	4.94 (4.2-5.55)	0.35
CD4 at sampling, cells/μL, Median(IQR)	713 (570-900)	680 (540-958)	762 (689-923)	610 (475-819)	0.015
CD8 at sampling, cells/μL, Median (IQR)	775 (600-1100)	780 (630-879)	894 (638-1300)	700 (530-870)	0.054
Viral load (<50 copies/mL), N (%)	97 (100)	19 (100)	44 (100)	34 (100)	1
Duration of treatment in years, median (IQR)	15 (9-18)	15 (13-18)	15 (8 -18)	14 (7-17)	0.73
Current Treatment, 1st drug, N (%)					0.84
ABC	31 (32)	8 (42)	13 (30)	10 (29)	
TDF/TAF	42 (43)	8 (42)	19 (43)	15 (44)	
Other	24 (25)	3 (16)	12 (27)	9 (27)	
Current Treatment, 3rd drug, N (%)					0.45
NNRTI	38 (39)	8 (42)	14 (32)	16 (47)	
PI/r	18 (19)	4 (21)	11 (25)	3 (9)	
INSTI	15 (15)	4 (21)	6 (14)	5 (15)	
Other	26 (27)	3 (16)	13 (29)	10 (29)	
BMI, Mean (SD)	24 (22-27)	22 (19-25)	26 (23-28)	24 (22-27)	0.003
VAT, Median (IQR)	89 (36-142)	41 (19-106)	127 (79-196)	69 (26-100)	0.0001
SAT, Median (IQR)	111 (70-167)	69 (33-115)	117 (82 -174)	119 (83-190)	0.02
MetS, N (%)	43 (44)	6 (32)	31 (70)	6 (17)	0.000009
Hypertension, N (%)	49 (51)	5 (26)	23 (52)	21 (62)	0.04
Central obesity, N(%)	57 (59)	8 (42)	32 (73)	17 (50)	0.033
Waist circumference (cm)	94 (87-101)	90 (84-95)	100 (91-105)	90 (87-97)	0.0007


A.
eigen gap
rotation cost
cluster
B.
matrix_1
SNF
Lipidomics
Metabolomics
Microbiome
C.
SNF-1
SNF-2
SNF-3
D.
Cluster
SNF-1
SNF-2
SNF-3
E.
DAG
Control
SNF-1
SNF-2
SNF-3
Cluster
F.
TAG
Control
SNF-1
SNF-2
SNF-3
Cluster
G.
PC2: 9% variance
PC1: 13% variance
Cluster
Ctrl
SNF-1
SNF-2
SNF-3
H.
Amino Acids
Co-factors and Vitamins
Lipids
Nucleotides
Others
Xenobiotics
LFC
SNF-3 vs. Control
SNF-2 vs. Control
SNF-1 vs. Control
SNF-3 vs. SNF-2
SNF-3 vs. SNF-1
SNF-2 vs. SNF-1
Spearman R
BMI
Age
Gender
p values


A.
Observed
ACE
Chao1
Fisher
Cluster
NS
**
SNF-1
SNF-2
SNF-3
B.
NMDS1
NMDS2
Cluster
SNF-1
SNF-2
SNF-3
C.
SNF cluster
MetS
Hypertension
Transmission
Gender
Relative abundance
variable
Cluster
SNF-1
SNF-2
SNF-3
Gender
Female
Male
Hypertension
No
Yes
Transmission
MSM
Heterosexual
Unknown/Other
MetS
No
Yes
Coriobacteriaceae
Bacteroidaceae
Prevotellaceae
Rikenellaceae
Christensenellaceae
Clostridiales.vadinBB60.group
Other
Lachnospiraceae
Ruminococcaceae
Erysipelotrichaceae
Acidaminococcaceae
Veillonellaceae
Succinivibrionaceae
D.
SNF-1 vs SNF-2
Families
Coefficients
Bacteroidaceae
Lachnospiraceae
Acidaminococcaceae
Fusobacteriaceae
Enterobacteriaceae
Rikenellaceae
Ruminococcaceae
Akkermansiaceae
Tannerellaceae
Barnesiellaceae
Christensenellaceae
Clostridium.sp.
Marinifilaceae
Eggerthellaceae
Erysipelotrichaceae
Muribaculaceae
Clostridiales.vadinBB60
Succinivibrionaceae
Veillonellaceae
Prevotellaceae
E.
SNF-1 vs SNF-3
Families
Coefficients
Significance
Negative
Positive
F.
Lachnospira
Prevotella
Bacteroides
Veillonella
Succinivibrio
C1
C2/C3
G.
Bacteroidaceae
Prevotellaceae
Condition
Ctrl
Control
SNF-1
SNF-2
SNF-3


A.
Variance explained (%)
75
50
25
0
Lipids
Metabolites
Microbiome
B.
F15
F14
F13
F12
F11
F10
F9
F8
F7
F6
F5
F4
F3
F2
F1
Lipids
Metabolites
Microbiome
Var. (%)
60
40
20
0
C.
F1
F2
F3
F4
F5
F6
F7
F8
F9
F10
F11
F12
F13
F14
F15
Transmission
CD4
Hypertension
BMI
Age
SAT
Gender
Cluster
VAT
MetS
log10 adjusted p−values
0 2 4 6 8
D.
Clusters
SNF-1
SNF-2
SNF-3
Variables
F1
F2
F3
F8
Z score
10
5
0
−5
−10
E.
3−hydroxy−2−methylpyridine sulfate
4−ethylcatechol sulfate
methylsuccinate
3−(3−amino−3−carboxypropyl)uridine*
3−hydroxyhippurate
arabonate/xylonate
3−hydroxy−3−methylglutarate
3−hydroxyhexanoylcarnitine (1)
glycerophosphoethanolamine
3,4−dihydroxybutyrate
17alpha−hydroxypregnanolone glucuronide
3−methyl catechol sulfate (2)
biliverdin
3,5−dichloro−2,6−dihydroxybenzoic acid
2−oxindole−3−acetate
(S)−3−hydroxybutyrylcarnitine
androstenediol (3beta,17beta) disulfate (2)
1−methylhistidine
indoleacetylglutamine
levulinoylcarnitine
TAG42:0−FA12:0
TAG42:0−FA14:0
TAG40:0−FA12:0
TAG42:1−FA12:0
TAG44:0−FA14:0
TAG44:0−FA16:0
TAG40:0−FA16:0
TAG44:0−FA12:0
TAG42:1−FA14:0
TAG42:0−FA16:0
TAG42:1−FA18:1
TAG45:0−FA14:0
TAG44:1−FA14:0
TAG44:1−FA14:1
TAG46:0−FA18:0
TAG45:0−FA15:0
TAG44:1−FA12:0
TAG42:1−FA16:0
TAG46:0−FA16:0
TAG46:0−FA14:0
Firmicutes
Actinobacteria
Proteobacteria
Cyanobacteria
Tenericutes
Bacteroidetes
Lentisphaerae
Top items
0.0 0.5 1.0 1.5 2.0
Variable
F1
F2
F3
F8
Type
Lipids
Metabolites
Microbiome
F.
Type
Lipids
Metabolites
Microbiome
N-acetylkynurenine (2)
(12 or 13)-methylmyristate
(14 or 15)-methylpalmitate
3-hydroxyisobutyrate
3,4-dihydroxybutyrate
alpha-ketobutyrate
glycohyocholate
glycolithocholate sulfate*
lithocholate sulfate (1)
deoxycholate
isoursodeoxycholate
G.
TAG
PC
PE
PI
DAG
CE
Peptide
Amino Acid
Actinobacteria
Bacteroidetes
Proteobacteria
SNF
Gender
AGE
BMI
VAT
SAT
METS
Hypertension
Transmission Mode
Lipid Classes
Metabolits
Microbiome
Pathway/Division
Clinical parameters


A.
Control
SNF-1
SNF-2
SNF-3
Ctrl_HIV
C2vsC1
C3vsC1
C3vsC2
C1vsCtrl
C2vsCtrl
C3vsCtrl
Group
Sub group
Z-score
4
2
0
−2
−4
LogFC
1
0.5
0
−0.5
−1
Group
produce by host modified by bacteria
produced by intestinal bacteria
Sub-group
Acetate derivates
Butyrate derivates
Indole derivatives
Polyamines
Propionate derivates
Secondary Bile Acid Metabolism
glycolithocholate sulfate*
glycolithocholate
deoxycholic acid 12−sulfate*
lithocholate sulfate (1)
glycocholenate sulfate*
glycodeoxycholate 3−sulfate
glycodeoxycholate
taurodeoxycholate
glycoursodeoxycholate
glycoursodeoxycholic acid sulfate (1)
taurochenodeoxycholic acid 3−sulfate
indolepropionate
2−oxindole−3−acetate
indolelactate
indoleacetoylcarnitine*
indoleacetylglutamine
methyl indole−3−acetate
indoleacetate
spermidine
spermine
N−acetylputrescine
acetoacetate
iminodiacetate (IDA)
2−hydroxyphenylacetate
2−aminobutyrate
3−hydroxybutyrate (BHBA)
3−aminoisobutyrate
2R,3R−dihydroxybutyrate
3,4−dihydroxybutyrate
3−(3−hydroxyphenyl)propionate
3−(3−hydroxyphenyl)propionate sulfate
3−ureidopropionate
imidazole propionate
C.
c1
c2
c3
c4
c5
c6
c7
B.
Simpson index (Alpha diversity)
Poultry Intake
Current Smoking
Ethnicity
Cumulative smoking
3-(3-hydroxyphenyl)propionate
Gender
glycoursodeoxycholic acid sulfate (1)
CD4 Nadir
MetS
3-(3-hydroxyphenyl) propionate sulfate
First Drug
CD4 at cART initiation
Alcohol
Transmission mode
Choice of fat for preparing warm dishes
glycolithocholate
Hypertension
Central Obesity
SAT
Duration of cART
CD4:CD8
deoxycholic acid 12-sulfate*
lithocholate sulfate
VAT
N-acetylputrescine
Vegetable intake
Antihypertensives
Waist Circumference
acetoacetate
imidazole propionate
SNF cluster
2-aminobutyrate
3rd Drug
VL
3-ureidopropionate
3,4-dihydroxybutyrate
Metabolites
HIV-related
cART Related
Microbiome
Weight related comorbidities
Food/Medicine
Patient demography
Positive association
Negative association
D.
Secondary Bile Acid Metabolism
indolepropionate
3-(3-hydroxyphenyl) propionate
3,4-dihydroxybutyrate
glycoursodeoxycholate
deoxycholic acid 12-sulfate*
glycolithocholate sulfate*
glycocholenate sulfate*
glycodeoxycholate
glycoursodeoxycholic acid sulfate (1)
glycodeoxycholate 3-sulfate
3-(3-hydroxyphenyl) propionate sulfate
glycolithocholate
indole-3-carboxylate
taurodeoxycholate
lithocholate sulfate (1)
Indole derivatives
2-hydroxybutyrate/ 2-hydroxyisobutyrate
indoleacetylglutamine
indoleacetoyl carnitine*
taurochenodeoxycholic acid 3-sulfate
3-hydroxybutyrate (BHBA)
2-oxindole-3-acetate
acetoacetate
indoleacetate
2-aminobutyrate
N-acetylputrescine
3-ureidopropionate
imidazole propionate
3-aminoisobutyrate
2R,3R-dihydroxybutyrate
2S,3R-dihydroxybutyrate
methyl indole-3-acetate
indolelactate
α-ketobutyrate
2-hydroxyphenylacetate
iminodiacetate (IDA)