

Multi-omics Clustering Reveals Distinct HIV Reservoir Profiles in the 2000HIV Cohort

Victoria Rios-Vazquez¹, Mareva Delporte², Vasiliki Matzaraki¹, Wilhelm Vos^{1,3}, Marc Blaauw¹, Louise van Eekeren¹, Albert Groenendijk⁴, Jéssica C. dos Santos¹, Mihai G. Netea¹, Andre van der Ven¹, Linos Vanderkerckhove²
¹Radboudumc (Department of Internal Medicine and Radboud Center of Infectious Diseases), Nijmegen, The Netherlands, ²Ghent University Hospital (HIV Cure Research Center), Ghent, Belgium, ³OLVG (Department of Internal Medicine), Amsterdam, The Netherlands, ⁴ErasmusMC (Department of Internal Medicine, Medical Microbiology, and Infectious Diseases), Rotterdam, The Netherlands

BACKGROUND

Multi-omics data integration offers the potential to uncover hidden patterns and heterogeneity among people living with HIV (PLWH). Understanding these molecular differences in host multi-omics as well as HIV viral reservoir (DNA) profiles may inform personalized therapeutic strategies. This study aimed to identify clusters of individuals with distinct host multi-omics and HIV viral reservoir profiles in the 2000HIV cohort.

METHODS

Study design

Multi-omics data from 1,236 PLWH in the 2000HIV cohort (NCT03994835) were analyzed, incorporating:

- **Gene Expression:** 58,347 genes (bulk RNA-seq from PBMCs)
- **DNA Methylation:** 793,775 CpG sites (Illumina Infinium MethylationEPIC BeadChip array from whole blood)
- **Plasma Proteins:** 2,367 proteins (Olink Explore from plasma)
- **Immune Cell Markers:** 355 markers (Flowcytometric immunophenotyping (CytoFLEX) from whole blood)
- **Cytokine Production Capacity:** 90 markers (ex vivo stimulation assays from PBMCs)
- **HIV Reservoir Measures:** Total HIV DNA and intact HIV DNA (copies/million CD4 T cells)

Clustering Approach & Driver Marker Identification

Data preprocessing included normalization, filtering, and feature selection across all layers. Multi-omics clustering and the identification of biological markers distinguishing the clusters were performed using **MoCluster**.

RESULTS

• Cluster selection confirmed by Consensus Partition Index (CPI), Gap Statistics, and area under the CDF curve identified three distinct clusters in the 2000HIV cohort ($k=3$) as shown in **Figure 1**:

- **High Total Low Intact (HTLI, n=348)** – High total reservoir DNA, low intactness
- **Low Total Low Intact (LTLI, n=421)** – Lowest total reservoir DNA and intactness
- **High Total High Intact (HTHI, n=467)** – High total reservoir DNA, highest intactness

• **Driver Markers:** The top 10 driver markers from each omics layer were identified, with the HIV (DNA) viral reservoir and bulk transcriptomics layers showing the strongest contributions (**Figure 2**).

- **In HIV (DNA) viral reservoir**, LTLI exhibited the lowest total reservoir and intactness, while HTLI and HTHI had high total reservoir levels with differing intactness (**Figure 3**). These differences were statistically significant ($P<0.05$).
- **In bulk transcriptomics**, key driver genes included *ACVR1C*, *AK5*, *LRP6*, *MMP28*, *SLC22A17*, *SLC16A10*, *OBSCN*, *MAN1C1*, *EDAR*, and *LEF1* ($P<0.05$), which were highly expressed ($P<0.05$) in LTLI compared to HTLI and HTHI (**Figure 4**).
- **These genes are involved** in T-cell development, immune response, macrophage function, apoptosis regulation, mitochondrial biogenesis, and HIV-1 transcriptional regulation.

2. Multi-Omics Clustering Heatmap Reveals Three Distinct Groups of PLWH

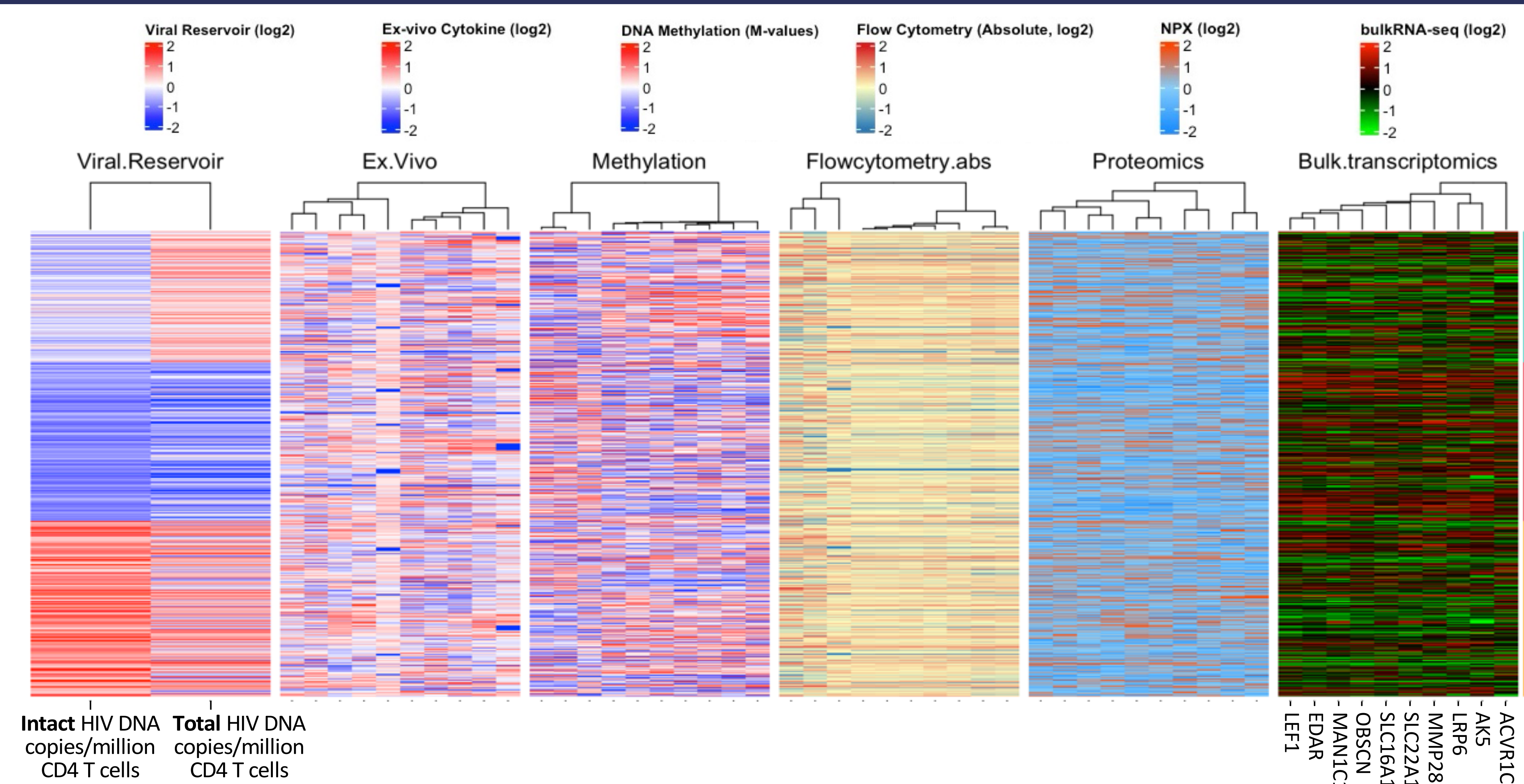
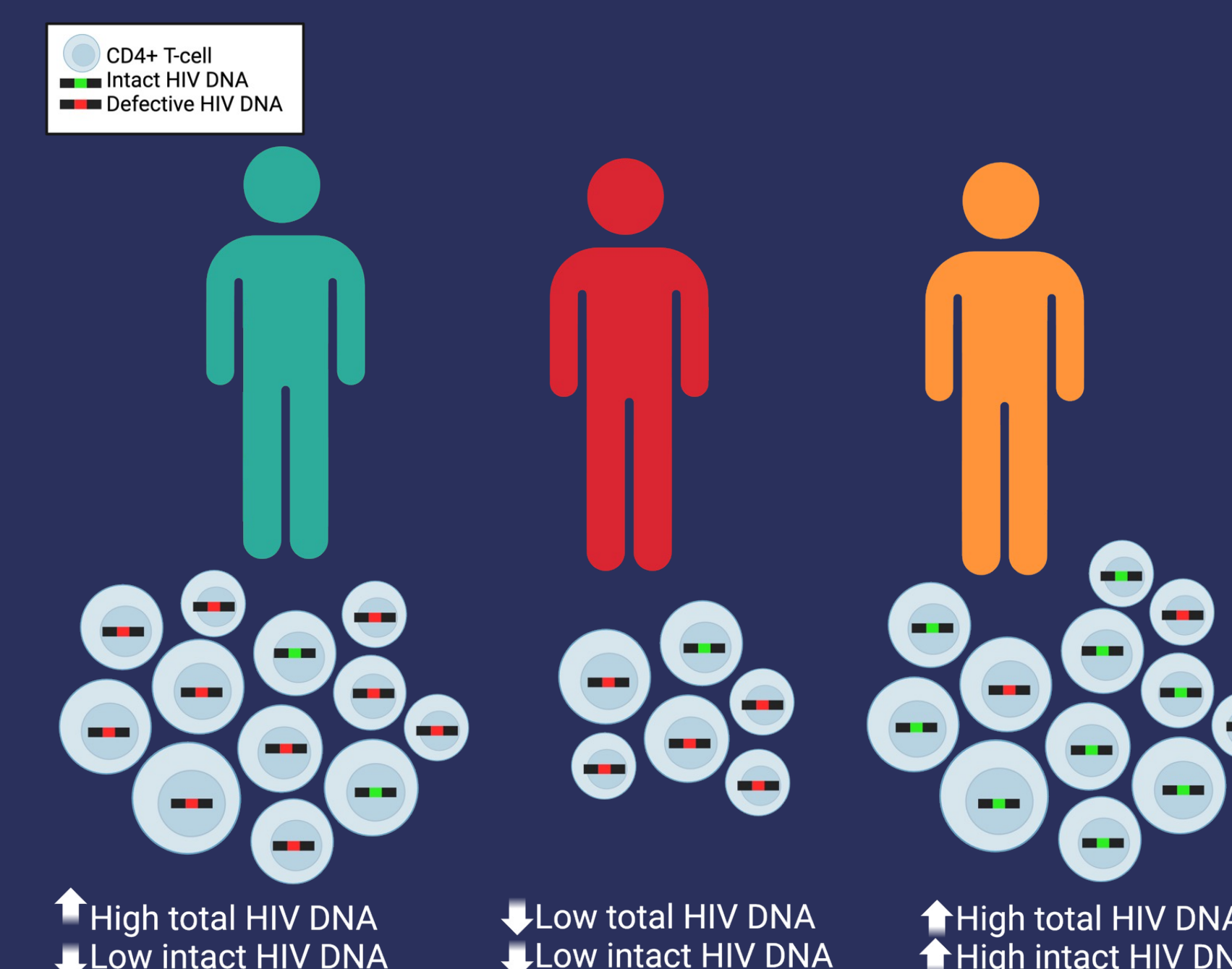
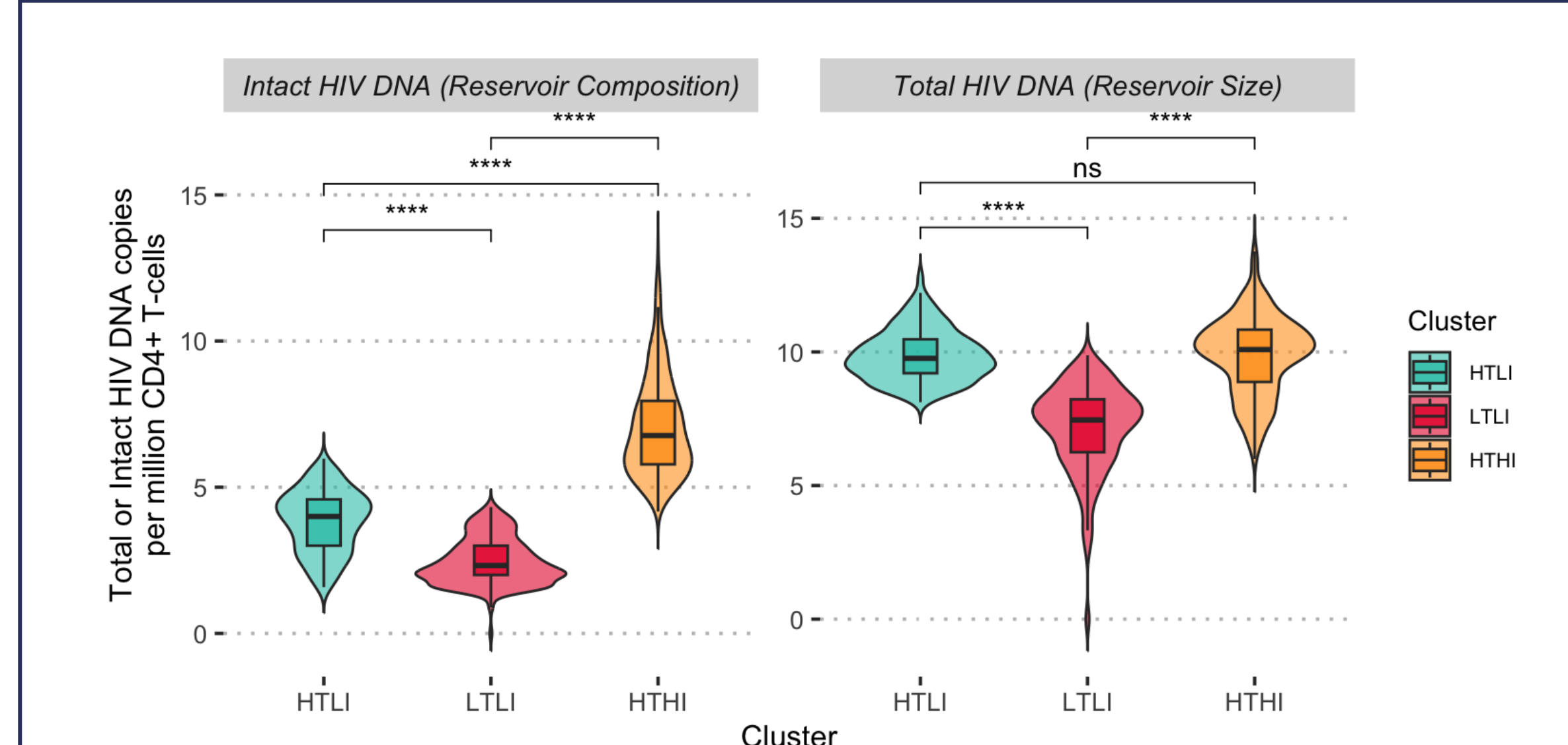


Figure 2. This heatmap illustrates the clustering results of the complete cohort using the MoCluster algorithm, which identified three distinct multi-omics clusters based on host multi-omics and viral reservoir layers. Columns represent individual omics markers, while rows correspond to PLWH, highlighting variations in omics features across clusters.

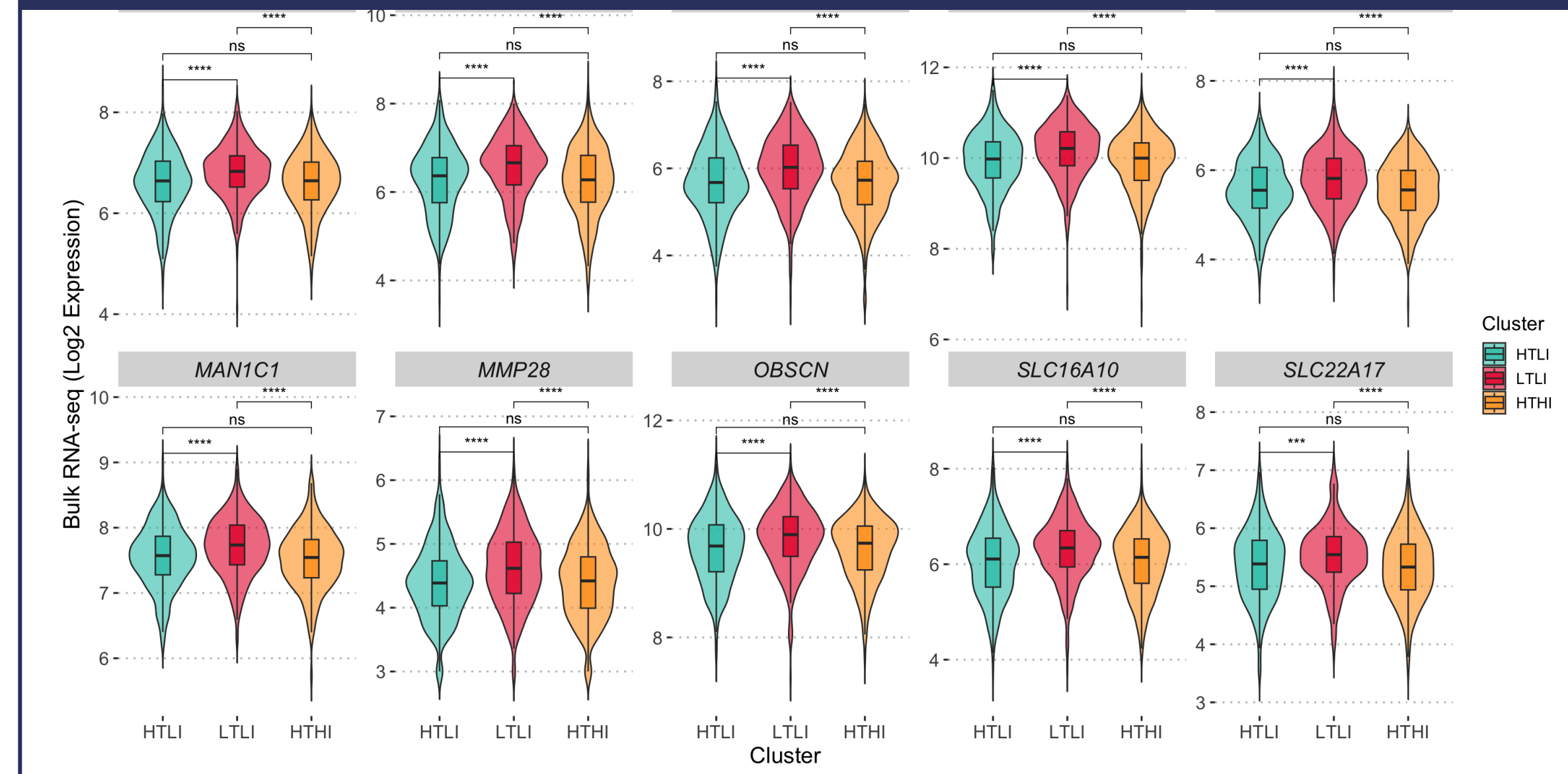
Multi-omics clustering revealed three distinct groups of PLWH, each with unique host molecular profiles and HIV reservoir size and composition



3. Variation in HIV Viral Reservoir and Intactness Across Clusters



4. Upregulation of Bulk Transcriptomics Driver Markers in the LTLI Cluster



CONCLUSIONS

Multi-omics clustering revealed significant heterogeneity in the 2000HIV cohort of PLWH under long-term treatment, identifying three distinct clusters with unique host molecular and HIV reservoir signatures. Notably, differences in the HIV reservoir and bulk transcriptomics layers were important drivers of cluster separation, highlighting variations in immune response markers and reservoir composition. Importantly, beyond the expected **high total/high intact** and **low total/low intact** groups, a distinct **high total/low intact** cluster was identified. These findings underscore the potential for personalized therapeutic strategies informed by multi-omics biomarkers.

ADDITIONAL KEY INFORMATION

Correspondence: Victoria.RiosVazquez@Radboudumc.nl

1. Three Optimal Clusters Identified via CPI and Gap Statistics

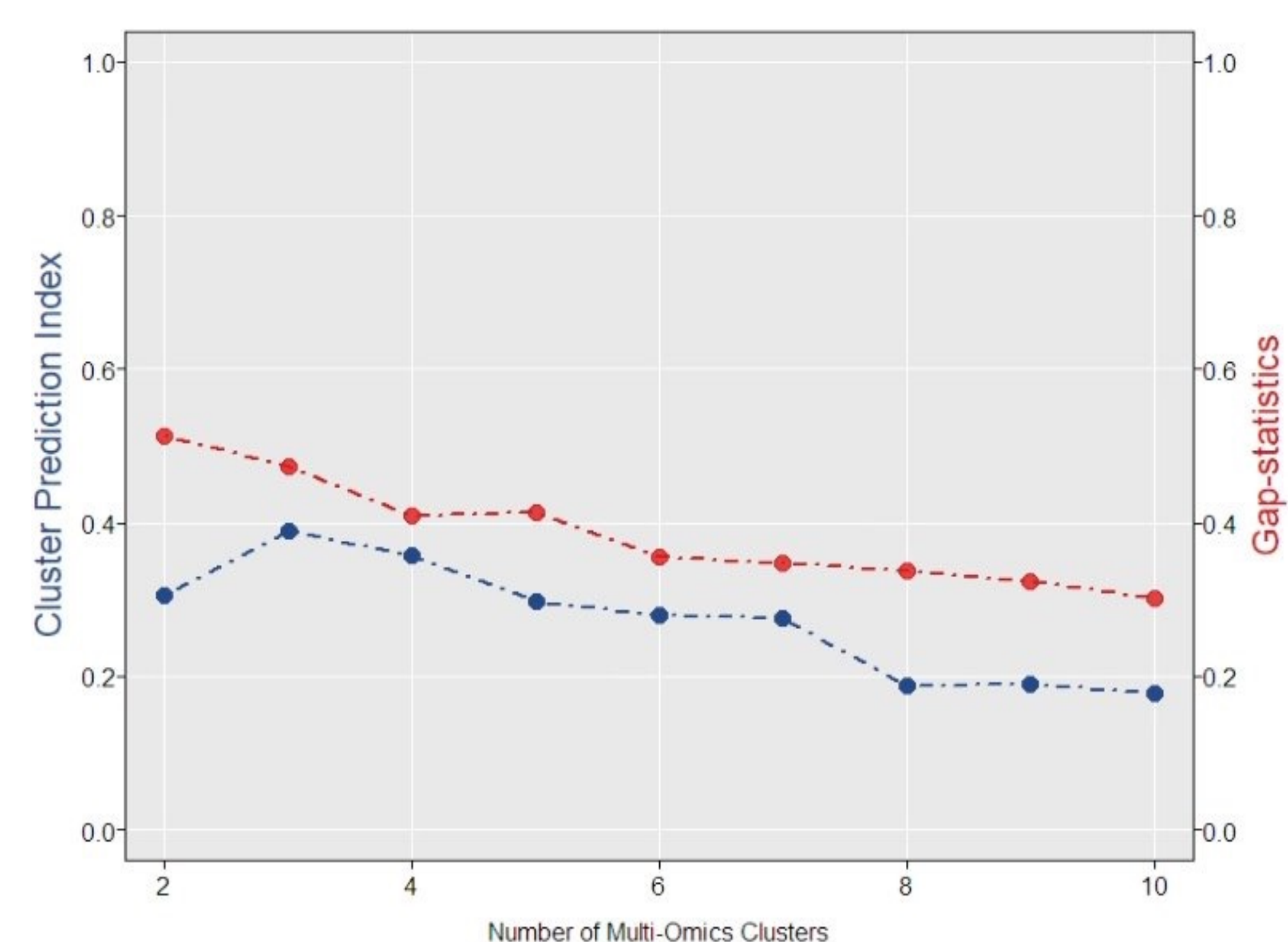


Figure 1. The Consensus Partition Index (CPI) and Gap Statistics were used to determine the optimal number of clusters (k). Visual inspection of their peaks identified $k=3$ as the optimal cluster number for the complete cohort.

