discovery are not effective enough. We have developed a bioinformatics pipeline to identify and classify all known viruses present in a metagenomic sample. Viral NGS reads are identified using a protein-based alignment method, DIAMOND, which is substantially faster than the standard BLAST method, and more reliable for viruses. These reads are automatically assembled into contigs using SPAdes, a *de novo* assembler. The contigs are then used to classify the virus at species level using a pan-viral typing tool based on all available taxonomic reference sequences from the International Committee on Taxonomy of Viruses (ICTV) database. This bioinformatics pipeline is Java-encoded and will include an easy-to-use web interface that is fit-for-purpose for researchers or clinicians. This tool can assemble viral contigs from paired-end reads generated by an Illumina MiSeq sequencer. So far 1865 viruses can be identified at species level resolution and 10 viruses (chikungunya virus, dengue virus, HBV, HCV, HHV8, HIV-1, HPV, HTLV-1, YFV, and Zika virus) at the genotype level. A web version of the pan-viral typing tool is already available and a web version with extended NGS functionality is currently being evaluated. Eliminating the need for virus-specific laboratory techniques, or targeted sequence capture, means a virome can be profiled in the context of its non-viral microbiome. Preliminary findings suggest our tool offers greater functionality than existing alternatives, with greater sensitivity to known viruses (including bacteriophages), automatic assembly and good quality phylogenetic analyses. A systematic comparison is underway.

### A35 Viral evolution and innate immune responses during acute HIV-1 infection and their association with disease pathogenesis

A.S. Hassan,[1,*] J. Hare,[2] G. Kamini,[3] L.M. Yindom,[4] A. Kamali,[5] E. Karita,[6] W. Kilemba,[6] M.A. Price,[7,8] P. Borrow,[4] P. Bjorkman,[9] J. Albert,[10] P. Kaleebu,[5] S. Allan,[6,11] P. Fast,[7] E. Hunter,[6,11] J. Gilmour,[2] T. Ndung'u,[3] S. Rowland-Jones,[4] E.J. Sanders,[1,4] J. Esbjornsson,[4,9,10]

[1]KEMRI/Wellcome Trust Research Programme, Kilifi, Kenya, [2]IAVI Human Immunology Laboratory, Lomndon, UK, [3]Kwazulu-Natal Research Institute for Tuberculosis and HIV, Durban, South Africa, [4]Nuffield Department of Medicine, University of Oxford, UK, [5]Medical Research Council/Uganda Virus Research Institute (MRC/UVRI), Uganda, [6] Rwanda and Lusaka, Rwanda/Zambia HIV Research Group (RZHRG) Kigali, Zambia, [7]IAVI, New York, NY, USA, [8]UCSF Department of Epidemiology and Biostatistics, San Francisco, CA, USA, [9]Department of laboratory medicine, Lund University, Sweden, [10]Department of Microbiology Tumor and Cell Biology, Karolinska Institute, Sweden and [11]Emory Vaccine Center Emory University, Atlanta, USA

The rate of HIV-1 disease progression varies widely between individuals. This has been attributed to a combination of virological and immunological events during acute HIV infection (AHI). However, the exact mechanisms explaining the relationship between HIV-1 diversity, evolutionary dynamics and host immune responses, and their effect on disease pathogenesis remain unclear. We aim to dissect HIV-1 viral diversity, evolutionary dynamics and select parts of the innate immune responses observed during AHI, and elucidate virus-host mechanisms involved in the regulation of HIV-1 disease pathogenesis during the acute and chronic stages of infection. A retrospective longitudinal study design from well-characterized AHI cohorts will be used. Archived samples from about 122 patients with AHI (defined as HIV-1 antibody negative and RNA or p24 antigen positive) from Europe (Sweden [$n=32$]) and Africa (Kenya [$n=32$], Rwanda [$n=14$], Uganda [$n=13$], Zambia [$n=15$], and South Africa [$n=16$]) will be included. Each patient will contribute plasma samples from four serial time points (<14, 30 [+/− 15], 90 [+/− 30] and 360 [+/− 180] days post estimated date of infection, EDI) collected prior to treatment initiation. HIV-1 env sequences determined by single genome sequencing (SGS), with 20 SGS clones from each time point, will be generated. In addition, a selected panel of innate immune markers will be profiled using the Meso Scale Discovery (MSD) electro-chemiluminescence-based platform and/or ELISAs. A multi-dimensional Bayesian framework of hierarchical phylogenetic models (HPM) will be applied, allowing for both fixed and random effects prior specifications to test for differences associated between and within patient group parameters, and where all measured virus-host parameters will be considered simultaneously. In addition, evolutionary parameters in different stages of the disease i.e. acute and chronic phases, will also be measured and accounted for in the HPM by addition of the epoch modeling approach to quantify the relationships between viral parameters, innate responses and their effect on disease pathogenesis. The proposed study is likely to constitute one of the largest virus-host dataset of longitudinally collected data of both virus sequences (covering a wide range of HIV-1 subtypes) and innate immune markers to date. The results of the proposed analyses will increase our understanding of HIV-1 pathogenesis and may have implications for therapeutic and prophylactic vaccine design.

### A36 Prevalence of HIV-1 subtypes in Slovenia with an emphasis on molecular and phylogenetic investigation of subtype A

J. Mlakar,[1,*] Maja M. Lunar,[1] A.B. Abecasis,[2] A.-M. Vandamme,[2,3] J. Tomažić,[4] T.D. Vovko,[4] B. Pećavar,[4] G. Volčanšek,[4] M. Poljak,[1]

[1]Faculty of Medicine, Institute of Microbiology and Immunology, University of Ljubljana, Ljubljana, Slovenia, [2]Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical Universidade Nova de Lisboa, Lisbon, Portugal, [3]Clinical and Epidemiological Virology, Rega Institute for Medical Research, K. U. Leuven, Leuven, Belgium and [4]Department of Infectious Diseases, University Medical Center Ljubljana, Ljubljana, Slovenia

In Slovenia, a small country in Central Europe, less than 1 per 1,000 inhabitants are estimated to be infected with HIV-1. As in most of the Central and Western European countries, the majority of patients diagnosed with HIV-1 are infected with subtype B. However, due to migration, other subtypes can become more prevalent in the country. The aim of this study was to determine HIV-1 subtypes circulating in Slovenia and to further examine the molecular epidemiology of subtype A. A total of 367 Slovenian HIV-1 sequences were included in the study, representing 58% of all patients diagnosed in Slovenia until the end of the year 2013. Subtype was assigned by employing different HIV subtyping tools coupled with Maximum likelihood phylogenetic analysis. The latter was performed to examine the molecular epidemiology of subtype A as well. Identified clusters of Slovenian subtype A sequences were further analyzed for the determination of the time of the most recent common ancestor (tMRCA) by using Monte Carlo Markov chain (MCMC) method available in BEAST 2.1.3 software. We determined the prevalence of subtype B to be 85.3%, while subtype A was the most prevalent non-B subtype found in 18 patients (4.9%), followed by CRF02_AG (2.4%), subtype C (1.1%), subtypes D, G and CRF01_AE (0.8% each) and subtypes F1 and CRF22_01A1 (0.3% each). Subtypes could not be assigned to 12 sequences (3.3%). The phylogenetic tree obtained by ML analysis of the subtype A and subtype A related recombinants revealed that Slovenian sequences were part of 6 major international clusters. Two clusters consisting only of Slovenian sequences were identified and thus additional MCMC analysis was employed. Results of a Slovenian cluster of 4 subtype A sequences showed a posterior probability value of 1 and a tMRCA between the years 1985 and 2008, with a mean in the year 2001. In conclusion, in a Central

European country, where subtype B predominates, the second most common subtype was found to be subtype A. Non-B subtypes were observed in one out of seven patients in Slovenia, a fraction that is not negligible, thus proving importance of surveillance of HIV subtype diversity and corresponding molecular epidemiology of non-B subtypes.

### A37 HIV drug resistance monitoring in children receiving first line antiretroviral therapy at two pediatric hospitals in Ho Chi Minh City

Anh Q. Luong,[1] Ton Tran,[1] Khanh Thu H. Hoang,[1] Ngoc Thao T. Do,[1] Thinh X. Vu,[1] Khanh H. Truong,[2] Quy T. Du,[2] Kim Thoa P. Le,[2] Viet C. Do,[3] An T. Vu,[3] Thanh Thuy T. Le,[3] Kim Chi T. Nguyen,[4] Binh K. Nguyen,[4] Hien. T Bui,[4] Xuan Lien T. Truong,[1]

[1]Pasteur Institute, Ho Chi Minh City, [2]Nhi Dong 1 hospital, Ho Chi Minh City, [3]Nhi Dong 2 hospital, Ho Chi Minh City and [4]HHS/CDC, Vietnam

Drug resistance is the main reason for antiretroviral treatment (ART) failure. Information on the prevalence of pediatric HIV drug resistance (HIVDR) in Vietnam is important to assist in the determination of the optimal ART regimen. We enrolled a prospective cohort of children newly initiating ART at one of two main pediatric hospitals in Ho Chi Minh City from December 2011 to March 2014. Demographic and clinical data were collected at baseline and supplemented by genotyping and VL at start and after 12 months or when first-line ART ends if it comes first. Of 136 patients enrolled, the mean age was 4.7 years; 17 (12%) exposed to ARV to prevent maternal to child HIV transmission; seven (5.15%) carried at least one strain of HIV with mutations related to ARV resistance, two (1.47%) against Nucleotide Reverse Transcriptase Inhibitors (NRTIs) (AZT, D4T), one (0.74%) against Non-Nucleotide Reverse Transcriptase Inhibitors (NNRTIs) (NVP), and four (2.94%) against Nelfinavir of Protease Inhibitors (PIs). At 12 months, 121 (89%) children were still receiving the first line ART, 7 (5%) died and 8 (6%) were lost to follow-up. Among 121 children on ART, 107(88%) achieved VL suppression (<1,000 copies/mL); 9 (7%) had acquired HIVDR mutations, three against NRTIs only, and six (4.96%) against both NRTIs and NNRTIs. The most prevalent mutation was the M184V (4.96%, $n = 6$) causing high-level resistance to 3TC, FTC and low-level resistance to ddI and ABC. Some TAMs were also found (D67N, K70R, T215F, K219Q). No major resistance mutations to PIs were detected. Viral load at initiation is associated with HIVDR at 12 months. Low levels of virologic failure and HIVDR were observed in pediatric patients. However, since some multidrug-resistant or cross-resistant mutations were recorded, continued monitoring of HIV drug-resistance in pediatric patients is needed.

### A38 Diversity analyses of HIV-1 envelope glycoproteins in HIV-infected individuals with and without broadly neutralizing antibodies

B. Mabvakure,[1,2] C. Scheepers,[1,2] M. Nonyane,[1] B. Lambson,[1] S. Madzorera,[1] D. Kitchin,[1,2] J. Bhiman,[1,2] K. Wibmer,[1,2] S. Abdool Karim,[3] C. Williamson,[3,4] L. Morris,[1,2,3] P.L. Moore,[1,2,3]

[1]National Health Laboratory Service, Center for HIV and STIs, National Institute for Communicable Diseases, Johannesburg, [2]School of Pathology Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, [3]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Kwa-Zulu Natal and [4]Division of Medical Virology, Institute of Infectious Disease and Molecular Medicine University of Cape Town

High levels of HIV envelope glycoprotein (Env) diversity have been associated with the development of broadly neutralizing antibodies (bNAbs). Here, we compare chronically HIV-1 infected subjects who develop bNAbs with those who did not, to assess whether lack of breadth can be attributed to low levels of viral diversity. Env nucleotide sequences were generated using Single Genome Amplification from four CAPRISA 002 cohort participants. Two participants developed neutralization breadth (CAP256 and CAP257) whereas the other two did not (CAP88 and CAP228) despite equivalently long duration of infection. Longitudinal diversity analyses were performed using Sequence Demarcation Tool (SDT). Phylogenetic analyses were performed using Bayesian Evolutionary Analysis by Sampling Trees (BEAST) software. Overall diversity increased with time in all subjects, as expected. Highest diversity was observed in CAP256 and CAP228, followed by CAP257 and least diversity in CAP88. The highest nucleotide substitution rates were observed in CAP257 (2.63 substitutions/100 nucleotides/year), CAP256 (2.28 subs/100n/yr.) and CAP228 (2.07 subs/100n/yr.), and the lowest in CAP088 (0.99 subs/100n/yr.). The time to the most recent common ancestor (tMRCA) inferred from BEAST was longer than the actual time of infection for CAP256 and CAP228, suggesting the possibility of super-infection or multivariant infection. We conclude that the absence of viral diversity may limit bNAb development, as in CAP88. However, increased diversity through high mutation rates and/or recombination, while likely necessary, is not sufficient for driving the development of bNAbs.

### A39 Human exome sequencing to evaluate the impact of rare coding variation on HIV-1 control

P.J. McLaren,[1,*] P.R. Shea,[2] I. Bartha,[3] J. Fellay,[3]

[1]National HIV and Retrovirology Laboratory, Public Health Agency of Canada, Winnipeg, Canada, [2]Institute for Genomic Medicine Columbia University, New York, NY, USA and [3]School of Life Sciences, Swiss Federal Institute of Technology, Lausanne, Switzerland

Common variants (>5% frequency) in the MHC and CCR5 regions are known to influence set point HIV-1 viral load (spVL) yet explain only a portion of the total trait variance. The impact of rare coding variation on HIV-1 disease progression has not been as thoroughly investigated. Here we utilize exome sequencing in 392 HIV-1 infected individuals with stable spVL to look for rare and functional variants that mediate control of HIV-1 infection. Set point HIV-1 viral load was calculated as the average of at least 3 measurements obtained during the chronic phase of infection. We captured and sequenced all coding exons in 392 HIV-1 infected individuals of the Swiss HIV Cohort Study using the Illumina Truseq 65Mb enrichment kit and the Illumina HiSeq2000. Quality control and variant calling were performed using the GATK and variant functional annotation was performed using snpEff version 3.3. Individual variants were tested for association using linear regression. Testing of the combined effects of multiple low frequency variants across each of >18,000 genes was performed using SCORE-Seq and SKAT. Consistent with previous results, single marker variant tests showed a strong signal of association in the MHC. The top association was observed between spVL and rs1131446 ($P = 2.3 \times 10^{-11}$) in exon 3 of HLA-B. Conditioning on this SNP, residual association was observed at rs2308622 (conditional $P = 2.2 \times 10^{-6}$) in HLA-C. Accounting for these two SNPs, no other variants showed evidence for association. Analyses aimed at detecting the combined effect of multiple low-frequency variants within a gene showed no significant associations. Restricting this analysis to only those variants that result in a change in protein sequence did not reveal further signals. Outside of the MHC, no significant impact of rare variation on spVL was detected by exome sequencing in 392 individuals.