Investigation of Immunological and Virological Markers in HIV Perinatally Infected Children

Triantafylia Maria Gkouleli

Thesis for the Degree of Doctor of Philosophy

Infection, Immunity, and Inflammation

Great Ormond Street Institute of Child Health

University College London

DECLARATION

I, Triantafylia Maria Gkouleli confirm that all the experimental and analytical work presented in this thesis is my own. Where others contributed to the study design or subsequent work presented, it has been appropriately indicated in the thesis.

This PhD project was funded by ViiV Healthcare and the UCL Institute of Child Health (ICH) as part of the EPIICAL consortium. The overall study was designed by my supervisors Professor Eleni Nastouli, Professor Nigel Klein and Dr Athina Soragia Gkazi.

In Chapter 3, The following people contributed to the work:
The RT-qPCR assay was designed and developed with the guidance of Dr Sarah Watters and Dr Kathleen Gartner. Initial analysis of the raw data was performed with the assistance of Dr Kathleen Gartner. The HIV-1 RNA fragment used to generate the standard curve was provided by Dr Eloise Busby, a research analyst at the LGC Group (formerly known as Laboratory of the Government Chemist).

In Chapters 4-6, the following people contributed to the work presented:
Sorting of the different cell populations used was performed at the Flow Cytometry Core Facility (UCL ICH) by Stephanie Canning and Dr Ayad Eddaoudi. The LD sub-study samples, used in Chapter 5, were prepared for sorting by Dr Katrine Schou Sandgaard. The PENTA11 sub-study samples, used in Chapter 6, were prepared for sorting, and extracted by Dr Athina Soragia Gkazi and Dr Katrine Schou Sandgaard. T-cell receptor excision circles (TRECs) were kindly measured by Dr Stuart Adams and Susanne Kricke at the SIHMDS Molecular Department of Haematology (Camelia Botnar Laboratories, Great Ormond Street Hospital). R and python scripts used for analysis were written and provided by my supervisor Dr Athina Soragia Gkazi and former PhD students of our group, Dr Ben Margetts, and Dr Katrine Schou Sandgaard.

ABSTRACT

Recent years have seen tremendous progress in HIV prevention methods and an impressive 70% decline in newly acquired paediatric infections. However, even at approximately 30%, the rate of new infections remains unacceptably high. Additionally, an increasing number of perinatally HIV-1 infected children are reaching adolescence and adulthood. Despite the significant patient care challenges that will unavoidably arise due to viral persistence, in cases suboptimal ART adherence, and the lack of a cure, this important cohort remains understudied. There is still important work to be done to ensure optimal treatment management and best quality of life in the years to come, by better understanding their viral and immunological dynamics.

This thesis aims to investigate immunological and virological markers by using three distinct paediatric HIV-1 cohorts. A cohort of early-treated, well-suppressed, perinatally infected children and adolescents, part of the CARMA study, is used to study virological and immunological characteristics. The size and composition of their HIV-1 reservoir is evaluated by developing a RT-qPCR assay to examine the presence of cell-associated HIV-1 RNA. Immunological characteristics, such as the TCR repertoire diversity, clonotype sharing, antigen "specificity", and the presence of invariant MAIT cells are studied by using a next generation sequencing method. Late-treated children, part of the LD study, and children that underwent planned treatment interruption, part of the PENTA11 study, are selected to study the immunological effects of deferred or well-monitored interrupted treatment.

Despite limitations and restrictions arising from studying small numbers of paediatric patients, the work here highlights the benefits and importance of early ART initiation and patient monitoring. Younger age at treatment initiation appears to have a beneficial effect in reducing the size of the viral reservoir, however HIV-1 CA-RNA is still detectable in 65% of the CARMA study participants. The greater thymic function capability of children, younger age at ART initiation and uninterrupted treatment and the presence of invariant immune cells, can provide an immunological advantage, with a higher repertoire diversity, great capacity for antigen recognition and immune response.

This work demonstrates that future longitudinal studies, that include larger international cohorts of HIV-1 perinatally infected children and adolescents, as well as age-matched healthy controls, will be necessary to determine the factors that affect the immunological and virological parameters in well-suppressed individuals, such as chronic immune activation and dysregulation, and HIV-1 reservoir replenishment. By continuously studying and elucidating their virological and immunological profiles, and by having an array of markers to follow, we will be able to predict therapy responses and inform novel treatment strategies.

IMPACT STATEMENT

This PhD project explores the impact of ART initiation timing as well as the impact of treatment strategies on specific immunological and virological markers. More specifically, this thesis explores the effects of ART initiation timing and potential treatment interruption on the immune system by studying characteristics such as TCR repertoire diversity, clonotype sharing, potential for antigen recognition and presence of an invariant immune cell population. Additionally, it investigates the effects of early ART initiation on the presence of HIV-1 CA-RNA, and subsequently the size of the viral reservoir and the presence of ongoing viral transcription. With this, I aim to contribute to the ever-expanding research field of paediatric HIV-1 infection.

Academically, this thesis will hopefully give rise to additional research projects that will further optimise the methods outlined here, both experimental and analytical. In the upcoming months, some of the research findings presented here will be submitted to peer-reviewed publications, ultimately contributing to the advancement of scientific knowledge in the HIV-1 research field.

Clinically, following the optimisation and streamlining required for introduction to clinical settings, these methods will be used to closely monitor HIV-1 paediatric patients when testing novel therapies to predict responses and avoid/prevent any adverse effects.

Finally, on a personal level, this PhD project enabled my training as a highly skilled researcher and majorly contributed to my career progression. It has enabled me to build strong collaborations and use skills and expertise acquired during this period to assist and inform patient care at a world-renowned paediatric hospital, Great Ormond Street Hospital, as a healthcare scientist (SIHMDS- Molecular Haematology).

ACKNOWLEDGEMENTS

This thesis would not have been possible without the funding from ViiV Healthcare and the EPIICAL consortium. But most importantly the patients and their families that allow us to conduct research to promote our scientific knowledge and contribute to the improvement of their lives. Without them nothing would be feasible.

To my supervisors, Nigel, Athina and Eleni, thank you for conceiving this project and for trusting me to complete it. It was a challenging journey and I appreciate your support throughout. Eleni, Kathleen, and Sarah thank you for providing me with your virological knowledge and guiding me with your HIV expertise. I always found our discussions challenging and motivating. Nigel, thank you for always being there for me even at difficult times and supporting me throughout this journey. I appreciate everything you did for me, all the help and support you provided all these years, all the stories you shared. Your optimism and enthusiasm, your immense immunological knowledge, and your encouragement even at points when I felt lost will not be forgotten. Athina, words cannot describe how lucky I feel that our paths crossed. I would not be here today if it wasn't for you. Thank you for not giving up on me even when things got tough, for training me and supporting me, for making me a better scientist, but most importantly a better person. Thank you for persisting and checking up on me, for being a great motivator, a great teacher and most of all a great friend. Thank you for all the phone calls and discussions, for letting me work through my fears and worries and for showing me there is always a way to achieve everything you want in life. I hope one day I will be able to repay all this.

To Dr Ben Margetts, thank you for introducing me to the world of TCR sequencing analysis. For your patience, advice, and guidance and for always being here to answer my R related questions, no matter how silly they were. To Dr Katrine Schou Sandgaard, thank you for all your help and feedback throughout all these years. It was great sharing the lab and working with you on all these projects. Your help with the sequencing analysis was invaluable. Thank you for all the conversations, your guidance and support.

To my friends and former colleagues at UCL Genomics, Paola, Tony, and Mark thank you for giving me the space I needed to complete my sequencing experiments. I am thankful for your support, advise and knowledge and I am happy your trusted me to be a part of the UCLG/PGU team last year.

To my friends, Erin, Claire, Jenny, Paola, and Maria thank you for being there for me. Through the ups and downs of our PhDs and lab work. Through all the walks and the trips, the coffee and theatre dates, the endless conversations, you were all the best support system one could have asked for. Can't wait to see what the future has in store for us.

To Jack, thank you for being the best partner in crime. For being so patient with me, for supporting and taking care of me these past few months when it wasn't always easy. Thank for all long walks and wonderful distractions, at a time when I needed it the most.

To family, especially my parents, Vaggelis and Emily and my sister Louiza. I wouldn't be here without you. You have been the biggest source of support and inspiration throughout my life. You let me spread my wings when I know it wasn't always an easy decision for you. You believed in me when no one else did and helped me become the person I am today. I hope I continue making you proud in the years to come and always be a positive light in your lives. To all of you I am eternally grateful.

TABLE OF CONTENTS

DECLARATION *2*

ABSTRACT *3*

IMPACT STATEMENT *5*

ACKNOWLEDGEMENTS *6*

TABLE OF CONTENTS *8*

ABBREVIATIONS *11*

LIST OF FIGURES *13*

LIST OF TABLES *18*

CHAPTER 1: INTRODUCTION *21*

1.1 Thesis Background **21**

1.2 Thesis Aim **22**

1.3 Human Immunodeficiency Virus (HIV) **23**

1.3.1 AIDS and HIV 23

1.3.2 Origins and Diversity of HIV 25

1.3.3 Transmission and Disease Progression 27

1.3.4 HIV-1 Structure and Genome 28

1.3.5 HIV-1 Life Cycle 30

1.3.6 HIV-1 Latency and Reservoir 31

1.4 Antiretroviral Therapy and Novel Treatments **35**

1.5 Cure Strategies **36**

1.6 Cellular Immune Response to HIV-1 ..... **38**

1.6.1 Innate Immune Response 38

1.6.2 Adaptive Immune Response 39

1.6.3 Differences Between Paediatric and Adult Immune Responses to HIV-1 42

1.7 Importance of Paediatric HIV-1 Studies **43**

CHAPTER 2: MATERIALS AND METHODS *44*

2.1 Sample Collection and Ethics **44**

2.1.1 Child and Adolescent Reservoir Measurements on early suppressive ART (CARMA) Study 44

2.1.2 Impact of ART Initiation Timing on the TCR Repertoire Sub-Study 45

2.1.3 PENTA11 Sub-Study 46

2.2 Blood Sample Processing **47**

2.2.1 Plasma Isolation 48

2.2.2 Peripheral Blood Mononuclear Cell Isolation 48

2.3 Fluorescence-activated Cell Sorting (FACS) **50**

2.3.1 Thawing and counting cells 52

2.3.2 Surface staining 52

2.3.3 PBMC fixation 53

2.4 Nucleic acid extraction 54
2.4.1 RNA extraction from fixed/sorted T-cells 55
2.4.2 DNA extraction from fixed/sorted T-cells 55
2.5 Quality control methods for nucleic-acid extractions and library generation 55
2.5.1 NanoDrop 56
2.5.2 Qubit 56
2.5.3 TapeStation 56
2.6 Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) 56
2.6.1 Quantitation of HIV-1 Cell-Associated RNA (HIV-1 CA-RNA) 58
2.7 Next Generation Sequencing (NGS) 62
2.7.1 Library preparation and sequencing 64
2.7.2 Data analysis 65
CHAPTER 3: VIROLOGICAL CHARACTERISTICS OF THE HIV-1 RESERVOIR 68
3.1 Background and Objectives 68
3.2 Introduction 69
3.3 Material and Methods 71
3.4 Results 74
3.5 Discussion 81
3.5.1 Conclusions 83
CHAPTER 4: IMMUNOLOGICAL CHARACTERISTICS OF HIV-1 PERINATALLY INFECTED INDIVIDUALS – IMPACT OF EARLY ART INITIATION ON T-CELL RECEPTOR (TCR) REPERTOIRE DIVERSITY 84
4.1 Background and Objectives 84
4.2 Introduction 84
4.3 Materials and Methods 86
4.4 Results 90
4.5 Discussion 108
4.5.1 Conclusions 114
CHAPTER 5: IMMUNOLOGICAL CHARACTERISTICS OF HIV PERINATALLY INFECTED INDIVIDUALS- IMPACT OF ART INITIATION TIMING ON THE TCR REPERTOIRE 115
5.1 Background and Objectives 115
5.2 Introduction 115
5.3 Materials and Methods 117
5.4 Results 120
5.5 Discussion 138
5.5.1 Conclusions 141
CHAPTER 6: IMMUNOLOGICAL CHARACTERISTICS OF HIV-1 PERINATALLY INFECTED INDIVIDUALS- IMPACT OF PLANNED TREATMENT INTERRUPTION ON THE TCR REPERTOIRE OF THE CD4+ MEMORY T-CELL POPULATION 142
6.1 Background and Objectives 142
6.2 Introduction 143

6.3 Material and Methods 146

6.4 Results 148

6.5 Discussion 167

6.5.1 Conclusions 171

CHAPTER 7: THESIS DISCUSSION 172

7.1 Part I: Virological Characteristics of HIV-1 Perinatally Infected Children 173

7.2 Part II: Immunological Characteristics of HIV-1 Perinatally Infected Children 174

7.3 Final Conclusions 179

APPENDICES 180

APPENDIX 1 180

APPENDIX 2 182

APPENDIX 3 184

APPENDIX 4 185

APPENDIX 5 187

APPENDIX 6 188

APPENDIX 7 190

APPENDIX 8 192

REFERENCES 194

ABBREVIATIONS

AIDS	Acquired Immune Deficiency Syndrome
ART	Antiretroviral Therapy
B-cell	B Lymphocyte
CA-DNA	Cell-associated Deoxyribonucleic Acid
CA-RNA	Cell-associated Ribonucleic Acid
CAR	Chimeric Antigen Receptor
CDR3	Complementarity Determining Region 3
CL3	Containment Level 3
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CT	Continuous Treatment
Ct	Cycle Threshold
DC	Dendritic Cells
DMSO	Dimethyl Sulfoxide
DS	Double Stranded
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence-activated Cell Sorting
FCS	Foetal Calf Serum
FDA	Food and Drug Administration
HIV	Human Immunodeficiency Virus
HTLV	Human T-cell lymphotropic Virus
ICH	Institute of Child Health
INSTI	Integrase Strand Transfer Inhibitor
IS	Incompletely Spliced
KS	Kaposi's Sarcoma
LAV	Lymphadenopathy Associated Virus
LGC	Laboratory of the Government Chemist
LRA	Latency Reversing Agents
LTR	Long Terminal Repeat
MAIT	Mucosal-associated Invariant T lymphocytes

MHC	Major Histocompatibility Complex
MS	Multiply Spliced
NK	Natural Killer
NTC	No Template Controls
NTRI	Nucleoside Reverse-Transcriptase Inhibitor
PBMC	Peripheral Blood Mononuclear Cell
PBS	Phosphate-buffered Saline
PCR	Polymerase Chain Reaction
PJP	*Pneumocystis jirovecii* Pneumonia
PI	Protease Inhibitor
PIC	Pre-integration Complex
PTI	Planned Treatment Interruption
qRT- PCR	Quantitative Reverse Transcription Polymerase Chain Reaction
QVOA	Quantitative Viral Outgrowth Assay
RT	Reverse Transcriptase
SIV	Simian Immunodeficiency Virus
SOP	Standard Operating Protocol
SS	Single Stranded
TALEN	Transcription Activator-like Nuclease
TBP1	TATA-box-binding Protein 1
T-cell	T Lymphocyte
TCR	T-cell Receptor
TILDA	*tat/rev* Induced Limiting Dilution Assay
TRECs	T-cell receptor excision circles
US	Unspliced
V(D)J	Variable, Diversity, Joining
WHO	World Health Organisation
ZNF	Zinc Finger Nuclease

LIST OF FIGURES

Figure 1 Thesis overview 22
Figure 2 Summary of the 2021 HIV epidemic data provided by the World Health Organisation (WHO). To date, approximately 38.4 million people are living with HIV. Despite all the progress made, a significant number of new HIV infections and related deaths are seen yearly [23]. 25
Figure 3 Phylogenetic tree of simian (SIV) and human (HIV) immunodeficiency viruses. For SIV, viruses originating from either chimpanzees or gorillas are presented. For HIV, the four groups N, O, P and M (with its different subtypes) are presented[33]. 26
Figure 4 Global distribution of HIV-1 group M subtypes. The different colours represent the different HIV-1 Group M subtypes and recombinants found predominantly found in each geographical area [35]. 26
Figure 5 The stages of HIV-1 infection. Increasing viral load and declining CD4+ T-cell counts mark disease progression [47]. 28
Figure 6 HIV-1 genome organisation and mature virion structure and protein expression[50] 29
Figure 7 Steps of the HIV-1 life cycle. Initially, the virus attaches to the host cell surface receptors. The viral genome enters the cytoplasm and is reverse transcribed into DNA. The viral DNA molecule then enters the nucleus as part of the pre-integration complex. Linear forms of viral DNA get integrated into the host cell genome where it serves as template for viral transcription. Transcription of the integrated provirus together with alternative RNA splicing create viral RNA species responsible for encoding both structural and accessory viral proteins. The generated transcripts are then exported into the cytoplasm where they assist in assembling new viral particles [62]. 30
Figure 8 Total HIV-1 CA-DNA consists of several different forms, both integrated and unintegrated, that can contribute to HIV pathogenesis via cell proliferation and persistent immune activation[71]. 32
Figure 9 Multiple HIV-1 RNA species, including unspliced, multiply (completely) and incompletely (singly) spliced, are generated through alternative splicing during the HIV-1 replication cycle[74]. 33
Figure 10 Antiretroviral drugs act by targeting different points of the HIV-1 life cycle. Entry inhibitors obstruct various proteins necessary for viral attachment, receptor binding and fusion. NRTIs and NNRTIs inhibit and block the viral reverse transcription process. Integrase inhibitors block the insertion of the proviral DNA into the host cell genome. Protease inhibitors interfere with the maturation of virions after budding from the host cell by inhibiting the necessary cleavage of viral polypeptides. Finally, maturation inhibitors similarly to protease inhibitors block the viral polypeptide cleavage [101]. 36
Figure 11 HIV-1 cure strategies. A) Latency reversing agents are being used to eradicate the latent HIV-1 reservoir in a “shock and kill” approach B) Gene therapy and gene editing approaches target the HIV-1 reservoir C) Latency inducing agents are being used to epigenetically silence the integrated viral genome in a “block and lock” approach D) Immune based therapeutic approaches such as vaccines, CAR-T cells and broadly neutralising antibodies are being used to enhance and support the host’s immune system and can be used in conjunction with other therapeutic approaches[108]. 38
Figure 12 Characteristics of the T-cell receptor: A) Random rearrangement of the variable (V), Diversity (D) and Joining (J) gene segments creates a diverse TCR repertoire. B) The TCR

is a heterodimeric polypeptide, consisting of either an alpha and beta chain or a gamma and delta chain[149]. 41
Figure 13 Representation of the haematocytometer surface. Sixteen small squares from the central square were used to count PBMCs. 50
Figure 14 Principles and main components of a flow cytometer: Cells suspended in liquid are transported in a single file towards the light source (laser beam) through the fluidics system. The optical system components collect the emitted scattered light and fluorescence. Two types of scattered light are collected and quantified, forward scatter (cell size) and side scatter (complexity and optical density of cells). The emitted fluorescence derives from the fluorescent dyes/antibodies used in the experiment. Photodetectors convert the generated light and fluorescence emissions into electrical signals and then in turn to digital signals by the electronics network. Specialised computer software presents the data in the form of dot plots and histograms (adapted from Cell Signalling Technologies)[190–192]. 51
Figure 15 Example of gating strategy followed at the Flow Cytometry Core Facility for the CARMA samples. Gating is a method used to select desired cell populations and exclude debris, thus allowing for the simultaneous analysis of an enormous number of parameters. Flow cytometry data are collected, and boxes are drawn around specific subpopulations, in this case CD4+ and CD8+ T-cells, for further analysis[194]. 54
Figure 16 Basic steps of PCR: Template DNA is denatured into single strands and primers are annealed onto target regions of each strand. DNA polymerase helps extend each primer along the template strand starting at the 3' end. These three basic steps are repeated to produce numerous copies of the selected DNA target (adapted from ThermoFisher Scientific). 57
Figure 17 Overview of the TaqMan probe-based qRT-PCR. This method takes advantage of the ability of Taq polymerase to cleave an oligonucleotide probe, generating a detectable fluorescent signal [196]. 58
Figure 18 Basic components of the one-step RT-qPCR reaction set up: The master mix and RNA template are combined under conditions that support the reverse transcription and subsequent PCR step (adapted from ThermoFisher Scientific). 59
Figure 19 Design of HXB2 plasmid insert for in vitro synthesis of RNA transcript (kindly provided by LGC Group research analyst E. Busby, 2018) 59
Figure 20 Illumina uses sequencing by synthesis technology. DNA molecules with attached adaptors are amplified on a flow cell. Following each round of synthesis fluorescence is recorded to determine which nucleotide is incorporated (https://frontlinegenomics.com/dna-sequencing-how-to-choose-the-right-technology/). .. 63
Figure 21 Outline of the TCR library preparation protocol (adapted by[202]) 64
Figure 22 Outline of the Demultiplexor computational pipeline for TCR analysis: This computational pipeline takes raw FASTQ files and with the help of the UMI sequences demultiplexes the data into files containing reads deriving from only one indexed sample. Decombinator takes the Decombinator output and identifies the relevant V and J gene information to assign them a five-part identifier. Collapsinator corrects the data for frequency and sequencing errors, so that CDR3 translator can extract the CDR3 amino acid sequences using the five-part identifier [202]. 66
Figure 23 Aspects of the persistent HIV reservoir. The involvement of numerous tissue and body compartments as well as various cell types, contributes to the complexity of the reservoir study (adapted from [216]). 69

Figure 24 Levels of A) total and B) unspliced HIV-1 CA-RNA detected by the one-step RT-qPCR in the 40 CARMA patients. HIV-1 CA-RNA was detectable in 65% of the CARMA study participants. 75
Figure 25 Quantified levels of different HIV-1 CA-RNA species, total and unspliced, in the CARMA patients. 77
Figure 26 Correlations between different variables measured at time of analysis and ART start. The heatmap indicates the strength of positive (blue) and negative (red) correlations present. The calculated Spearman's correlation coefficient (ρ) is also depicted in each corresponding square. 78
Figure 27 Example of the gating strategy used for the CARMA study samples. Sorting was performed using FACSAria III (BD Biosciences) to obtain CD4+ and CD8+ T-cells. 88
Figure 28 Correlations between different variables measured at time of analysis and ART start. The heatmap indicates the strength of positive (blue) and negative (red) correlations present. The calculated Spearman's correlation coefficient (ρ) is also depicted in each corresponding square. 91
Figure 29 TRECs number change as we age, with their levels constantly declining as life progresses. Here TRECs levels per million A) CD4+ and B) CD8+ T-cells are presented for the 19 UK CARMA patients. C) TRECs reference levels kindly provided by Dr Stuart Adams and the GOSH clinical lab team are shown. 92
Figure 30 Histogram presenting read number per sample for CD4+ A) TCRα and B) TCRβ sequences and CD8+ C) TCRα and D) TCRβ sequences. 95
Figure 31 Sequencing data were subsampled to allow for comparisons. Subsampled data were used to calculate the Gini coefficient presented here in the form of histogram plot, demonstrating similar diversity distribution between patients, for CD4+ A) TCRα, B) TCRβ sequences and CD8+ C) TCRα, D) TCRβ. 97
Figure 32 Plots presenting examples of CD4+ A) TCRα, B) TCRβ and CD8+ C) TCRα, D) TCRβ CDR3 sequence sharing and their frequencies between at least two UK CARMA patients. Only a limited number of the detected CDR3 sequences are shared between patients. 98
Figure 33 Frequency distribution of Hamming distances between CD4+ A) TCRα, B) TCRβ and CD8+ C) TCRα, D) TCRβ CDR3 sequences. A Hamming distance of 1 represents a single amino acid change between two neighbouring CDR3 sequences with the same length. 101
Figure 34 Clusters of closely related CD4+ TCRα CDR3 sequences. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. 102
Figure 35 Clusters of closely related CD8+ TCRα CDR3 sequences. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. Due to the lower subsampling level of some samples plots for UK01 and UK05 could not be generated. 103
Figure 36 TRECs levels per million CD4+ T-cells as measured for all seven sub-study participants. Age is one of the main factors affecting TRECs levels, with progressive loss being apparent as we age. 120
Figure 37 Subsampled sequencing data were used to calculate the Gini coefficient as seen here on this histogram plot. Similar clonotype abundance distribution was observed between all patients for CD4+ A) TCRα and B) TCRβ as well as CD8+ C) TCRα and D) TCRβ sequences. 122
Figure 38 Plots presenting the TCRα and TCRβ CD4+ CDR3 sequences shared between at least two patients in A) Group A, early treated CARMA study participants, B) Group B, late treated LD study participants and C) Group C, healthy controls. 123

Figure 39 Plots presenting the TCRα and TCRβ CD8+ CDR3 sequences shared between at least two patients in A) Group A, early treated CARMA study participants, B) Group B, late treated LD study participants and C) Group C, healthy controls. 124
Figure 40 Frequency distribution of the calculated Hamming distances of CD4+ α and β CDR3 sequences between the participants within A) Group A, B) Group B and C) Group C. Hamming distances represent the number of amino acid differences between two sequences of the same length. 128
Figure 41 Frequency distribution of the calculated Hamming distances of CD8+ α and β CDR3 sequences between the participants within A) Group A, B) Group B and C) Group C. Hamming distances represent the number of amino acid differences between two sequences of the same length. 130
Figure 42 Clusters of closely related CD4+ TCRα CDR3 sequences identified in each participant in A) Group A, B) Group B and C) Group C. 131
Figure 43 Clusters of closely related CD8+ TCRα CDR3 sequences identified in each participant in A) Group A, B) Group B and C) Group C. 132
Figure 44 Changes in the calculated Gini coefficient for α (blue) and β (pink) chains in the CD4+ naïve T-cell population during A) planned treatment interruption and B) continuous treatment. Increase in the number of clonal expansions was observed during weeks 24-48 in the PTI arm, returning though to pre PTI levels 3 years after ART re-initiation. No changes in the TCR repertoire diversity were observed in the CT arm of the study[204]. 144
Figure 45 Subsampled PENTA11 sequencing data (apart from data presented for patient G at the W0 timepoint) were used to calculate the Gini coefficient and investigate the effect of planned treatment interruption on the diversity of the TCR repertoire in the CD4+ memory T-cell population. As seen in these histograms, the Gini coefficient for both A) α and B) β chains remains at similar levels in either study arm. 151
Figure 46 Plots presenting the TCRα and TCRβ sequences shared between the different PENTA11 study timepoints (W0, W12/W24, W48 and W150) in the three patients receiving planned treatment interruption. 154
Figure 47 Plots presenting the TCRα and TCRβ sequences shared between the different PENTA11 study timepoints (W0/W2 and W150) in the three patients receiving continuous treatment. 157
Figure 48 Frequency distribution of Hamming distances between the different timepoints for each PENTA11 participants undergoing planned treatment interruption. Hamming distances represent the number of amino acid changes between two neighbouring CDR3 sequences of the same length. 161
Figure 49 Frequency distribution of Hamming distances between the different timepoints for each PENTA11 participants on continuous treatment. Hamming distances represent the number of amino acid changes between two neighbouring CDR3 sequences of the same length. 163
Figure 50 Clusters of closely related TCRα CDR3 sequences present at each PENTA11 study timepoint (W0, W12/W24, W48 and W150) in the planned treatment interruption arm. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. 164
Figure 51 Clusters of closely related α CDR3 sequences present at each PENTA11 study timepoint (W0/W2 and W150) in the continuous treatment arm. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. Lower

subsampling levels affect the number of clusters that can be detected leading to obvious differences in the presented plots. 165
Figure 52 Thesis overview 172
Figure 53 Due to similarities observed between participants in the Gini coefficient when calculated using subsampled sequencing data, it was also calculated using raw sequencing data, as seen here in these histogram plots. Similar clonotype abundance distribution patterns were observed between all patients for CD4+ A) TCRα and B) TCRβ as well as CD8+ C) TCRα and D) TCRβ sequences. The higher Gini coefficient present here when compared to the subsampled data could be attributed to the presence of higher frequencies of certain clonotypes. 187
Figure 54 Clusters of closely related CD4+ TCRβ CDR3 sequences. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. 188
Figure 55 Clusters of closely related CD8+ TCRβ CDR3 sequences. Due to the subsampling levels of some samples (UK01, UK04 and UK06) it was not possible to generate plots. 189
Figure 56 Clusters of closely related CD4+ TCRβ CDR3 sequences identified in participants in A) Group A, B) Group B and C) Group C. Participants in Group A, part of the CARMA study, had a higher number of clusters of closely related CDR3 sequences when compared to participants of the other two groups. Lower subsampling levels can affect the detection of clusters. 190
Figure 57 Clusters of closely related CD8+ TCRβ CDR3 sequences identified in participants of A) Group A, B) Group B and C) Group C. Participants of Group B (apart from V30 that was subsampled to a lower level) presented with higher numbers of closely related CDR3 clusters when compared to participants from Groups A and C. No plot could be generated for Group A participant UK13. 191
Figure 58 Clusters of closely related TCRβ CDR3 sequences present at each PENTA11 study timepoint (W0, W12/W24, W48 and W150) in the planned treatment interruption arm. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. No plot could be generated for timepoint W48 for participant C. 192
Figure 59 Clusters of closely related β CDR3 sequences present at each PENTA11 study timepoint (W0/W2 and W150) in the continuous treatment arm. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. Lower subsampling levels affect the number of clusters that can be detected leading to obvious differences in the presented plots. 193

LIST OF TABLES

Table 1 Characteristics of the CARMA cohort (adapted from [4]). Continuous data are presented as median (IQR), categorical data are presented as number of patients (%). 44
Table 2 Characteristics of individuals selected to investigate the impact of ART initiation timing (adapted from[4,178]). Continuous data are presented as median (IQR), categorical data are presented as number of patients (%). 46
Table 3 Characteristics of individuals selected to investigate the impact of treatment interruption on CD4+ memory T-cells (adapted from[181]). Continuous data are presented as median (IQR), categorical data are presented as number of patients (%) 47
Table 4 Age-dependent volume permitted per each blood draw for CARMA participants. .. 48
Table 5 Recommended volumes and tube sizes for PBMC isolation using Lymphoprep (adapted from Stem Cell Technologies) 49
Table 6 Reagents and volumes (µl) required for DNase treatment..... 60
Table 7 Reagents and volumes (µl) required for the total HIV CA-RNA assay. 61
Table 8 Reagents and volumes (µl) required for the detection of unspliced HIV-1 CA-RNA. . 61
Table 9 RT-qPCR temperature regimen on the Bio-Rad CFX96 instrument. 62
Table 10 Primer and probe sequences for the HIV CA-RNA assays[89]. 62
Table 11 Virological characteristics of the CARMA cohort (adapted from [4]). Unless otherwise stated, continuous data are presented as median (IQR). 71
Table 12 Calculated LOD and LOQ values for the HIV-1 CA-RNA RT qPCR assays. 74
Table 13 Summary of HIV-1 CA-RNA results, between the two assays detecting total or unspliced HIV CA-RNA. Copies/10^6 PBMCs are presented as median (IQR). 76
Table 14 Summary of virological and immunological parameters compared by total HIV-1 CA-RNA levels. Unless otherwise stated, all continuous data are presented as median (IQR). 79
Table 15 Summary of virological and immunological parameters compared by unspliced HIV-1 CA-RNA levels. Unless otherwise stated, all continuous data are presented as median (IQR)..... 80
Table 16 Characteristics of the 19 UK individuals, part of the CARMA cohort (adapted from [4]). Unless otherwise stated, continuous data are presented as median (IQR). 86
Table 17 Absolute CD4+ and CD8+ counts, relative subset sizes and CD4:CD8 rations, depending on age group, as measured in healthy paediatric populations (adapted from [236]). Unless otherwise stated, continuous data are presented as median, 10^{th} to 90^{th} percentiles. 86
Table 18 Age at analysis (in years) and TRECs levels (per million CD4+ and CD8+ T-cells) are presented for all 19 UK CARMA patients..... 94
Table 19 Antigen specificity found in the combined TCRα and TCRβ shared CDR3 sequencing data by using the VDJdb database. Multiple antigen matches can be seen for CD4+ CDR3 sequences, implying cross-reactivity. An example of the matches can be seen below. 99
Table 20 Antigen specificity in TCRα and TCRβ CD8+ CDR3 sequences assigned by the VDJdb database. Multiple antigen matches can be seen for CD8+ CDR3 sequences. An example can be seen below. 100
Table 21 Sequence similarity between CD4+ CDR3s and MAIT cells. Presented in this table are only the 50 first sequences, sorted by their similarity score..... 103

Table 22 Sequence similarity between CD8+ CDR3s and MAIT cells. Presented at this table are only the first 50 sequences, sorted by their similarity score. 105
Table 23 Characteristics of the 12 individuals selected to investigate the impact of ART initiation timing on the TCR repertoire (adapted from[4,178]). The participants were separated into three groups, Group A included individuals from the CARMA study that initiated ART early, Group B included individuals from the LD study that initiated ART late, and Group C included healthy controls. The participants were selected with their age as the main criterion, as there was an effort to obtain relatively age-matched participants for each group. Participants in Control Group C are of a slightly older age, as it wasn't possible to obtain samples from healthy volunteers below the 18th year of age. Continuous data are presented as median (IQR), categorical data are presented as number of patients (%). 117
Table 24 TRECs per million CD4+ cell reference data provided by the SIHMDS Molecular Department of Haematology lab, presented here as median and 10th centile. 119
Table 25 Example of antigen "specificity" as identified by the VDJdb database. Cross-reactivity with multiple pathogens was observed for the majority of the detected CDR3 sequences. 125
Table 26 Example of antigen "specificity" as identified by the VDJdb database. Cross-reactivity with various antigens was observed for numerous of the identified CDR3 sequences. 126
Table 27 Number of MAIT cell matched sequences per sub-study participant, as identified by the MAIT MATCH server (DTU). Sequences presenting with a similarity score of one were considered a perfect match. 132
Table 28 MAIT cell matched sequences identified within the CD4+ samples of all 12 sub-study participants, using the MAIT MATCH server (DTU). 133
Table 29 MAIT cell matched sequences identified within the CD8+ samples of all 12 sub-study participants, using the MAIT MATCH server (DTU). 136
Table 30 Characteristics of the 9 PENTA11 participants selected for this sub study (adapted from [181]). Unless otherwise stated, continuous data are presented as median (IQR). 146
Table 31 Read depth achieved per patient (for each timepoint and each TCR chain) using the Illumina MiSeq system. Three patients had to be excluded from the study due to obtaining insufficient or no number of reads. Two samples from patient G (highlighted in red) could not be subsampled due to their sequencing depth. 148
Table 32 Top five expanded CDR3 sequences in PTI participants, per chain (α and β), and their frequencies at each study timepoint (W0, W12/W24, W48 and W150). 152
Table 33 Top five expanded CDR3 sequences in CT participants, per chain, and their frequencies at each study timepoint (W0/W2 and W150). 155
Table 34 Example of antigen specificity as identified by the VDJdb database. Cross-reactivity was observed for multiple CDR3 sequences found within the PENTA11 data. 158
Table 35 HIV-1 specific CDR3 sequences present in PENTA11 PTI patients and their frequencies at the four different timepoints. Antigen specificity was determined using the VDJdb database. No HIV-1 related sequences were identified in the three CT patients. 158
Table 36 Sequence similarity between CDR3 sequences identified in the PENTA11 patients and known MAIT cell sequences. A similarity score of 1 represents an exact match to a known MAIT cell sequence. No MAIT cell sequences with a similarity score of one were found in patients F_CT and G_CT. 166
Table 37 The CARMA study included 40 samples from sites across Europe (UK, Spain, and Italy). After sample collection, PBMC isolation and DNA extraction, different assays were

performed across the sites. Results from the total HIV-1 CA-DNA assay are shown below[4]. 180
Table 38 UK CARMA samples selected for TCR sequencing for the research presented in Chapter 4. All samples (n=19) were sorted into two T-cell populations, CD4+ and CD8+ and RNA was extracted before library preparation and TCR sequencing. 182
Table 39 Participants of the LD study (V=vertical transmission, C=healthy control) selected for comparison to the early treated CARMA study participants, in Chapter 5[4,14]. All samples were sorted into two T-cell populations, CD4+ and CD8+ and RNA was extracted before library preparation and TCR sequencing. 184
Table 40 RNA samples selected for the PENTA11 sub-study presented in Chapter 6, investigating TCR differences in the CD4+ memory T-cell subpopulation. Four timepoints for patients under PTI and two timepoints for patients under CT were chosen. Cell numbers after FACS sorting and RNA concentration after extraction are shown below. 185

CHAPTER 1: INTRODUCTION

1.1 Thesis Background

Increased survival rates of HIV-1 perinatally infected children and adolescents and their progression into adulthood, has set up a significant challenge for patient care in the coming decades[1]. The paediatric HIV-1 epidemic is currently joining a new phase, since an increasing number of children on antiretroviral therapy are moving into adolescence and adulthood and living with HIV-1 presents new challenges. These individuals present a unique and rather uniform cohort since they had a developing immune system at the time of infection and to a point had suboptimal therapy options and adherence[1]. There is still much work to be done to ensure their optimal management and survival in the years to come.

Despite tremendous progress in the HIV-1 research field, there are limited studies addressing this cohort of patients. Studies on ART initiation in infancy have shown that it can result in the development of reduced viral reservoirs, leading to the hypothesis that these patients represent a unique cohort that could undergo ART interruption and achieve post-treatment control or even remission[2–5]. Paediatric cases, such as the "Mississippi baby", the South African child enrolled in the "Children with HIV-1 Early antiretroviral therapy (CHER) trial" and the French patient, that have demonstrated that long-term virological control and delayed viral rebound post treatment interruption are achievable and positively correlated with the time of ART initiation[6–11]. Additionally, further elucidating the effect of HIV-1 on the T-cell thymic output and on the developing immune system could help predict patient responses to novel immunotherapeutic interventions. Thus far, it has been suggested that children initiating ART early can still have a relatively active thymus leading to better immune reconstitution[12–14]. It is apparent that paediatric HIV-1 studies are necessary and will continue to play a vital role in informing patient management and new treatment strategies. These studies will be able to help patients reach adulthood with minimal detrimental clinical outcomes, intact immune systems, minimal drug resistance and toxicity[15].

1.2 Thesis Aim

Due to the legal, ethical, and physiological limitations surrounding paediatric studies, the impact of early ART on HIV-1 persistence and evolution, as well as the effects on the developing immune system of perinatally infected children/adolescents has not been studied extensively. This thesis aimed to contribute to the expanding HIV-1 research field and offer further insights on the viral and immunological dynamics of perinatally HIV-1 infected individuals.

The overall aims of this project were to investigate:

a. The size and composition of the HIV-1 reservoir, by examining the presence of cell-associated HIV-1 RNA in long supressed early treated perinatally infected children/adolescents with HIV-1.
b. The effects of different antiretroviral therapy strategies on the developing immune system of perinatally infected children, by examining their T-cell receptor (TCR) repertoire characteristics.

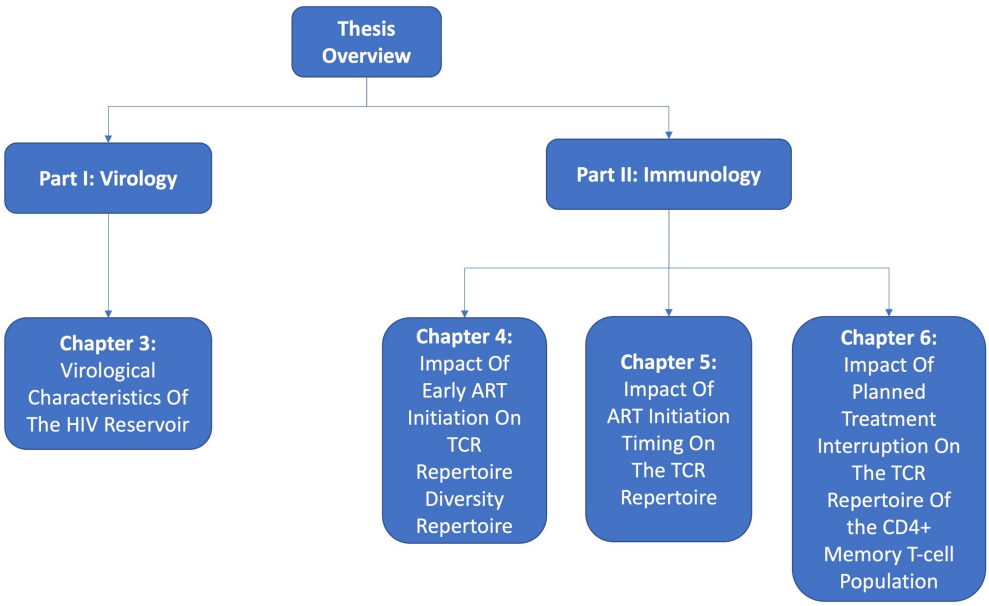


Figure 1 Thesis overview

Research presented in Chapter 3, aimed to help characterise the viral reservoir of perinatally HIV-1-infected children, that initiated therapy early in life and had been virally suppressed and provide key data on the established viral reservoir, its size and composition. This was achieved by optimising the tools that investigate determinants of the HIV-1 reservoir with biological significance, more specifically with the quantification of cell-associated HIV-1 RNA (HIV-1 CA-RNA). During this project, possible evidence of ongoing viral replication in early treated individuals that have undetectable plasma viremia were investigated. The generated data showed that HIV-1 CA-RNA, both in spliced and unspliced forms, is detectable in patient PBMCs suggesting ongoing virion production contributing to replenishment of the reservoirs and chronic immune activation in these children.

Research presented in Chapters 4-6, aimed to elucidate the effects of HIV-1 infection, timing of ART initiation and treatment interruption on the developing immune system of perinatally infected children and adolescents, by investigating the characteristics of their TCR repertoire. Thus far, most TCR studies have been conducted using samples from adult patients. Since the immune system of children and adults presents differences, this project aimed to address a gap in our existing knowledge and complement our groups ongoing research and findings. This was achieved through investigating the TCR repertoire by using samples from three distinct cohorts and a next generation sequencing method.

1.3 Human Immunodeficiency Virus (HIV)

1.3.1 AIDS and HIV

Acquired Immune Deficiency Syndrome (AIDS) was first recognised in 1981, when several homosexual men in the USA presented with unusual opportunistic infections and malignancies, such as *Pneumocystis jirovecii* Pneumonia (PJP) (formerly known as *Pneumocystis carinii* Pneumonia) and Kaposi's Sarcoma (KS)[16,17]. In 1983, Francoise Barré-Sinoussi et al. isolated what they believed to be the causative agent of AIDS, a T-lymphotropic

retrovirus[18]. They concluded that it belonged to the family of the previously described human T-cell lymphotropic virus (HTLV) and named it Lymphadenopathy Associated Virus (LAV)[18]. A year later, in 1984, Robert Gallo et al. reported the isolation of a new retrovirus termed HTLV-III[19]. It was later confirmed that these newly discovered retroviruses were the same viral species and in 1986 the International Committee on the Taxonomy of Viruses introduced the name Human Immunodeficiency Virus (HIV)[20].

According to the latest data released by UNAIDS (United Nations Programme on HIV/AIDS), since the start of the epidemic 84.2 million people have become infected with HIV and approximately 40.1 million have died of AIDS-related illnesses. In 2021, 1.5 million new infections and 650,000 AIDS related deaths were reported[21]. Despite all the progress that has been made 38.4 million people, including 1.7 million children (<15years)[22], are currently living with HIV (Figure 2).

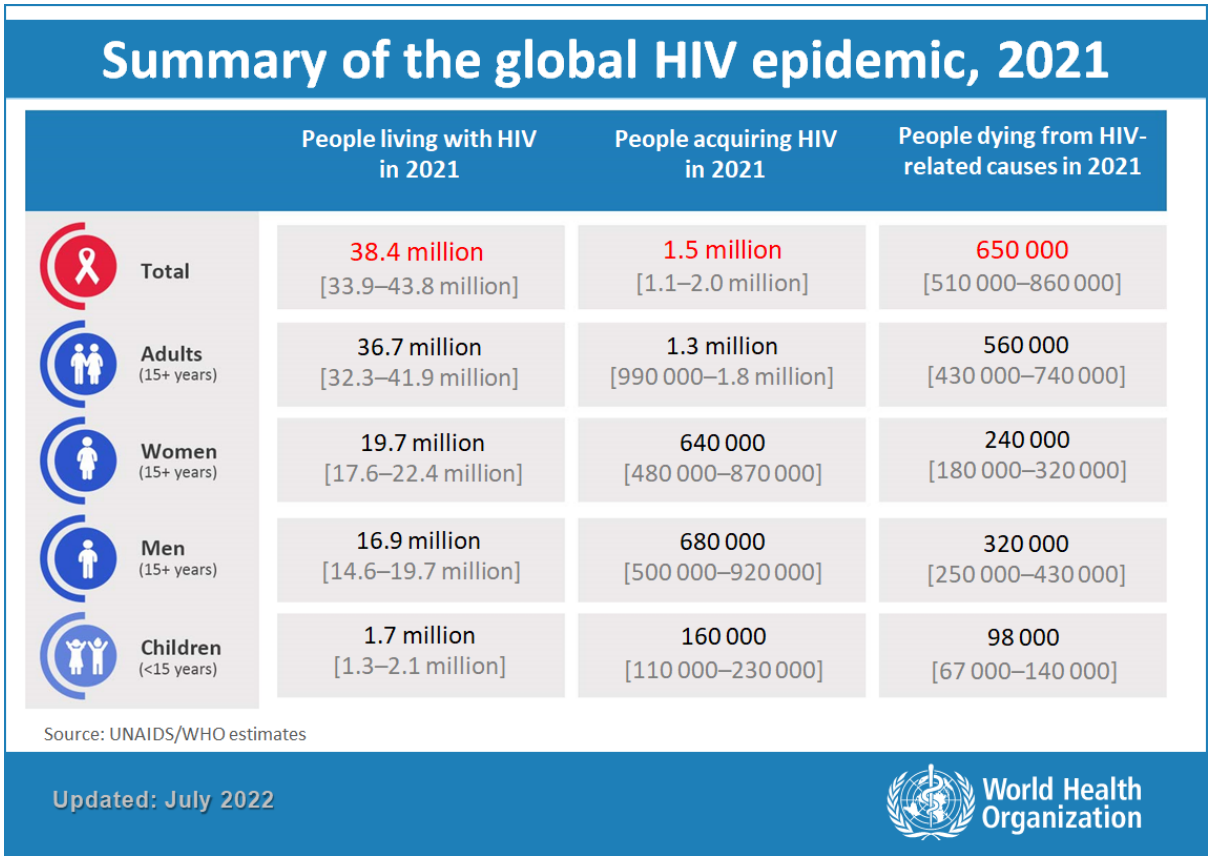


Figure 2 Summary of the 2021 HIV epidemic data provided by the World Health Organisation (WHO). To date, approximately 38.4 million people are living with HIV. Despite all the progress made, a significant number of new HIV infections and related deaths are seen yearly [23].

1.3.2 Origins and Diversity of HIV

In the past decades, there has been great progress in our understanding of the origin, genetic variability, and diversity of HIV. There are two main strains of HIV, HIV-1 and HIV-2[24,25]. Despite their unique characteristics, both types present many similarities including gene organisation, routes of transmission and viral replication pathways[26,27].

HIV-1 is mainly responsible for the global pandemic and accounts for most HIV infections and AIDS-related deaths worldwide[25]. Due to its high mutation, recombination and replication rates, several HIV-1 groups and subtypes have heterogeneously emerged world-wide[28]. It can be divided in four groups: M (main/major), O (outlier), N (non-major and non-outlier) and P[29,30]. Groups M and N are believed to have originated from the simian immunodeficiency virus (SIV) found in chimpanzees (SIVcpz), whereas groups O and P originated from gorillas (SIVgor)[31–33](Figure 3).

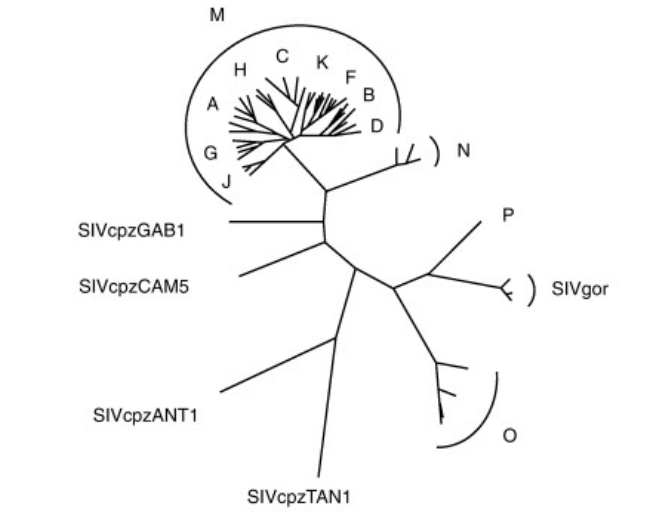


Figure 3 Phylogenetic tree of simian (SIV) and human (HIV) immunodeficiency viruses. For SIV, viruses originating from either chimpanzees or gorillas are presented. For HIV, the four groups N, O, P and M (with its different subtypes) are presented[33].

As the name suggests, group M accounts for most HIV-1 infections and AIDS-related deaths worldwide. Group M is further divided into subtypes or clades (Figure 3), designated as A, B, C, D, F, G, H, J, K and recombinant forms[28,29,34].

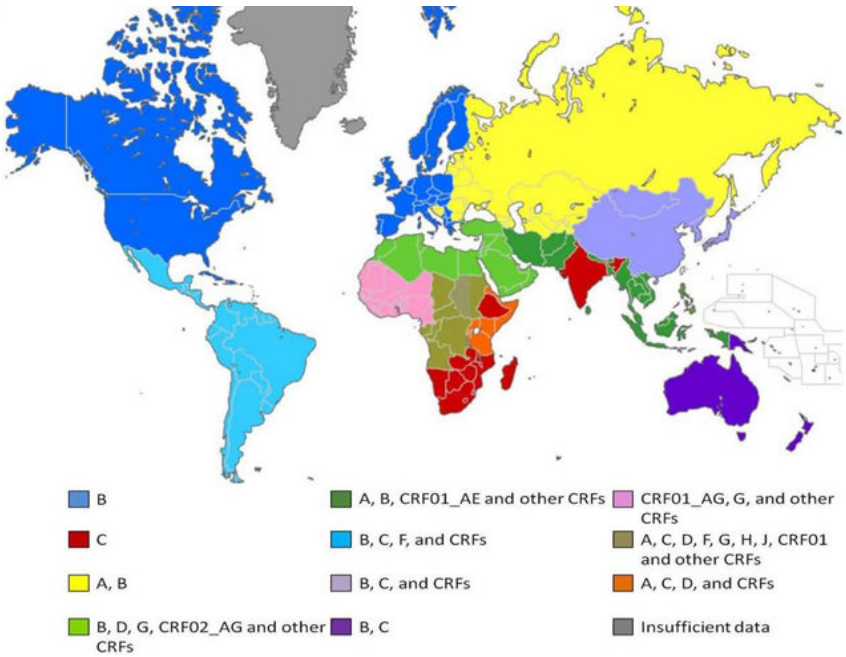


Figure 4 Global distribution of HIV-1 group M subtypes. The different colours represent the different HIV-1 Group M subtypes and recombinants found predominantly found in each geographical area [35].

Recent data suggest an increase in the prevalence of recombinant forms, especially in regions where multiple subtypes co-circulate (Figure 4)[28]. Groups O, N and P remain much less prevalent, with group P being documented in only two patients from Cameroon[25,32].

To date, HIV-2 has remained mainly concentrated to West African countries and is considered less common and infectious than HIV-1[36]. It is closely related to SIV from sooty mangabeys (SIVsmm) and can be divided into eight groups (A-H), with groups A and B being the most predominant[37,38].

1.3.3 Transmission and Disease Progression

There are two modes of HIV-1 transmission, horizontal and vertical[39]. In adults, the main routes of horizontal transmission include unprotected sexual contact, transfer of infected blood and blood products and sharing of contaminated needles and sharp objects[39,40]. On the other hand, in children, the transmission is usually vertical and happens from mother-to-child either during pregnancy, childbirth or through breastfeeding[41,42].

HIV-1 infects and kills vital cells of the immune system, such as CD4+ T lymphocytes (T-cells), macrophages and dendritic cells[43–45]. One of the main characteristics of HIV-1 infection is the declining numbers of CD4+ T cells[46]. If left untreated, it can lead to AIDS and susceptibility to opportunistic infections[47].

In general, there are three major stages of HIV-1 infection: acute primary HIV-1 infection, clinical latency, and AIDS[47]. The acute stage develops during the first few weeks after infection and is characterised by plummeting CD4+ T-cell counts and increasing copy numbers of HIV-1 RNA in plasma[47]. If left untreated, during the clinical latency, HIV-1 continues to actively replicate at lower levels and the CD4+ T-cell count slightly recovers[48]. Due to severe immunological damage over approximately a decade, the viral load starts to increase significantly and the CD4+ T-cell count to decrease, leading to AIDS[49].

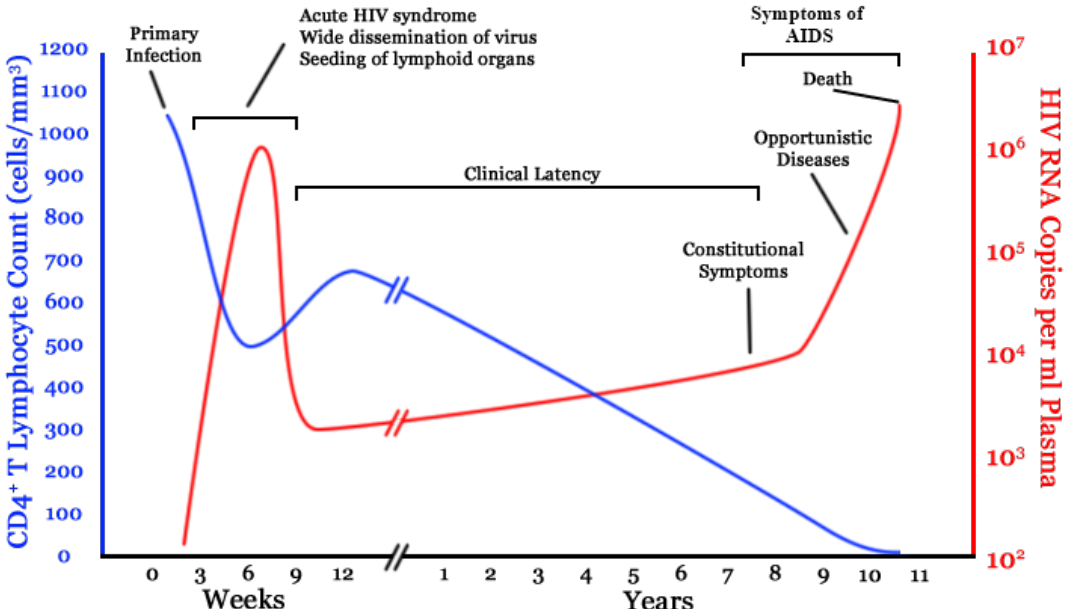


Figure 5 The stages of HIV-1 infection. Increasing viral load and declining CD4+ T-cell counts mark disease progression [47].

1.3.4 HIV-1 Structure and Genome

HIV-1 is a single stranded, positive-sense, enveloped RNA virus that belongs to the *Lentivirus* genus *of the Retroviridae* family[50]. It is spherical in shape and has a diameter of approximately 100-130nm[51]. After reverse transcription into DNA and integration into the host cell genome the ~9.7 kb provirus comprises of two long terminal repeats (LTRs) flanking nine genes: gag, pol, env, tat, rev, nef, vif, vpu and vpr (Figure 6)[50,52].

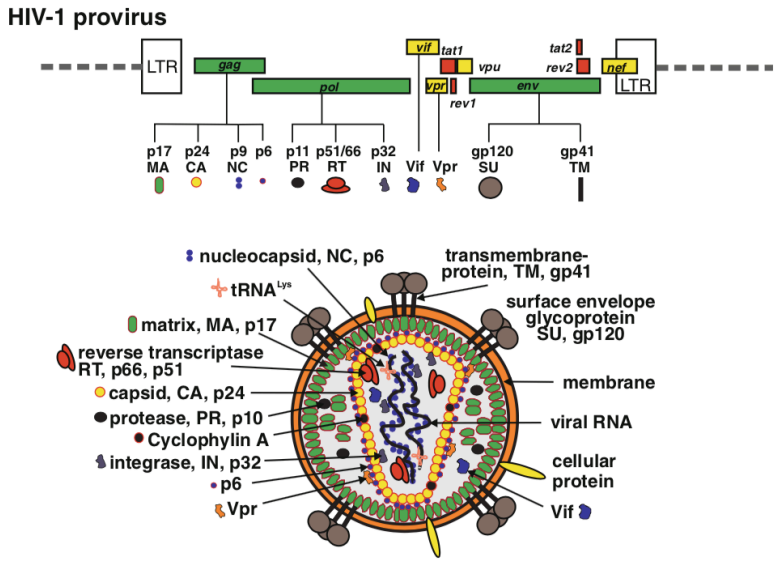


Figure 6 HIV-1 genome organisation and mature virion structure and protein expression[50]

Gag, pol and env are common to all retroviruses and encode structural and enzymatic proteins[50,53]. The gag gene gives rise to a precursor protein called p55 that is cleaved into four smaller proteins during viral maturation, matrix (MA-p17), capsid (CA-p24), nucleocapsid (NC-p9) and p6[53]. The pol gene encodes for two enzymes necessary for virus replication and insertion into the genome of an infected cell, RT (reverse transcriptase) and IN (integrase), and a third enzyme necessary for virus maturation, PR (protease)[54]. Env encodes for the external glycoprotein 120 (gp120) and the transmembrane glycoprotein 41 (gp41) that facilitate viral attachment and fusion of the viral envelop with the cellular membrane, respectively[55]. This leads to entry of HIV-1 into its target cells. From the regulatory genes, tat (transactivator) enhances transcription of proviral mRNAs, rev plays an important role in the export of viral mRNAs into the cytoplasm[56]. The remaining four genes, nef, vif, vpu and vpr, were called “accessory”, since initially they were deemed unnecessary for viral replication[57]. However, it is now known that they are responsible for several other critical functions, such as the facilitation of viral immune invasion and the antagonization of host antiretroviral factors[58,59].

1.3.5 HIV-1 Life Cycle

The HIV-1 life cycle comprises of the entry, reverse transcription of viral RNA into DNA, uncoating, nuclear import of viral DNA, integration, transcription of viral RNAs, RNA export, translation, assembly, budding and maturation stages (Figure 7)[50]. It can be divided into an early and a late stage. The early phase commences with the attachment of the virion to the cell surface and ends with proviral DNA integration[60]. The late phase includes all the steps between the proviral transcription to the release of new infectious virions[50,61].

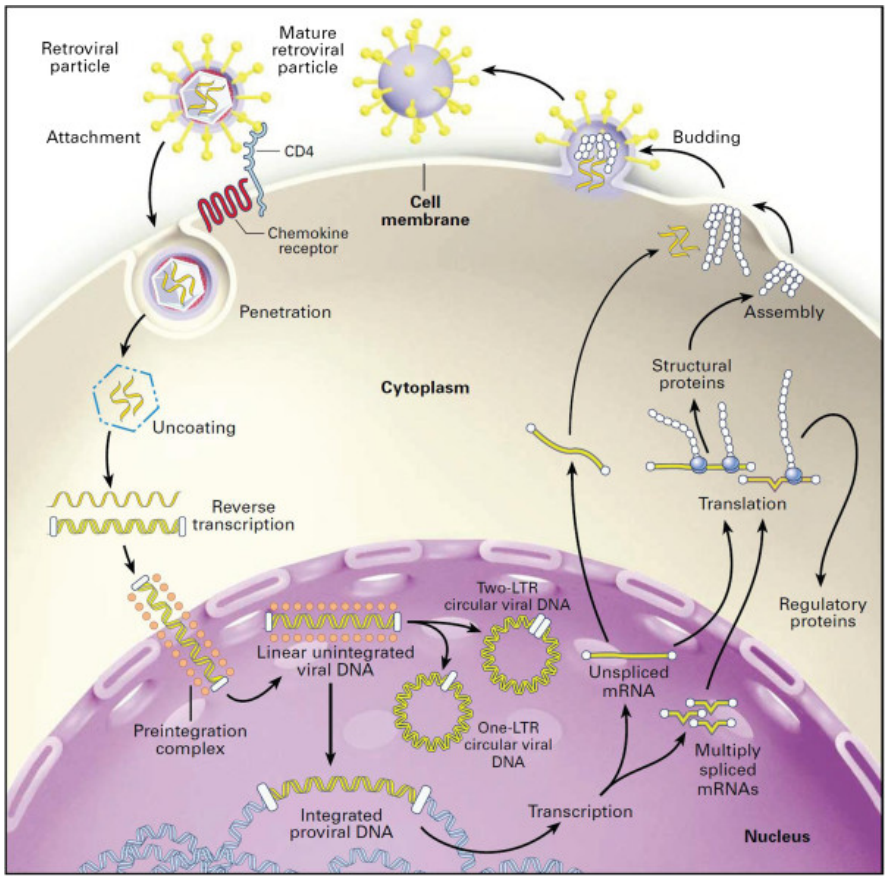


Figure 7 Steps of the HIV-1 life cycle. Initially, the virus attaches to the host cell surface receptors. The viral genome enters the cytoplasm and is reverse transcribed into DNA. The viral DNA molecule then enters the nucleus as part of the pre-integration complex. Linear forms of viral DNA get integrated into the host cell genome where it serves as template for viral transcription. Transcription of the integrated provirus together with alternative RNA splicing create viral RNA species responsible for encoding both structural and accessory viral proteins. The generated transcripts are then exported into the cytoplasm where they assist in assembling new viral particles [62].

At first, the env protein in the viral membrane binds to a receptor (CD4) and co-receptor (CCR5 for R5-tropic or CXCR4 for X4-tropic viruses) on the host cell surface, causing conformational changes in env necessary to allow fusion of the two membranes[47]. Once fusion is completed, the contents of the HIV-1 particle are released into the cytoplasm of the host-cell and through reverse transcription the single-stranded RNA (ss-RNA) is converted into double-stranded DNA (ds-DNA)[63]. After reverse transcription is completed, viral and cellular proteins associate with the viral ds-DNA and form the pre-integration complex (PIC), which can enter the nucleus and eventually lead to integration into the host cell genome (HIV-1 provirus)[64]. In the case of T-cell activation, transcription of the integrated HIV-1 DNA begins and leads to the generation of different RNA transcripts[65]. Initially, short (~2kb) multiply spliced (ms) RNA is produced that encodes the tat, rev, and nef genes. As the infection progresses there is a shift towards the production of unspliced (us) and incompletely spliced (is) transcripts (~9kb and ~4kb respectively)[50]. The export of the viral mRNAs from the nucleus to the cytoplasm and the translation of viral proteins lead to the subsequent assembly of a progeny virus particle[61]. The newly formed HIV-1 progeny virus exits the host cell at the cellular membrane through a process called budding[61]. Finally, activation of the viral protease leads to the maturation of the viral particle, leading to major structural changes within the particle, which can then infect a new cell[50,52].

1.3.6 HIV-1 Latency and Reservoir

Despite considerable advances in our understanding of HIV-1 infection, a major barrier for HIV-1 eradication remains. That barrier is the result of proviral integration into the host cell genome early on during the acute phase of infection[66]. There, the virus can remain in a latent state (not actively replicating) and establish a stable reservoir, even in the presence of suppressive antiretroviral therapy[67]. In the case of treatment interruption or cell re-activation, this reservoir can lead to the production of new infectious virions, thus making its study and detection vital for patient management and wellfare[67]. It is generally accepted that the main pool of HIV-1 latency is CD4+ T-cells, although other cell populations such as macrophages and dendritic cells have also been implicated[68–70]. Since suppressive ART can

make viral load undetectable, different markers have been investigated and used to gauge the size of the HIV-1 reservoir. Two of the most widely used markers are total cell-associated HIV-1 DNA (CA-DNA) and cell-associated HIV-1 RNA (CA-RNA)[62].

Cell-associated HIV-1 DNA has been proven to be one of the most clinically relevant reservoir markers and has been used extensively in studies, since its levels present in blood and PBMC samples can predict disease progression and response to treatment[71,72]. In most settings, total HIV-1 CA-DNA is quantified and used as a guide of the reservoir size. Total HIV-1 CA-DNA consists of integrated and unintegrated (linear, episomal 1-LTR and 2-LTR) forms, deriving from both replication competent and defective genomes[71,73]. It has been hypothesised that total HIV-1 CA-DNA can contribute to pathogenesis and reservoir maintenance via cell proliferation and immune activation (Figure 8).

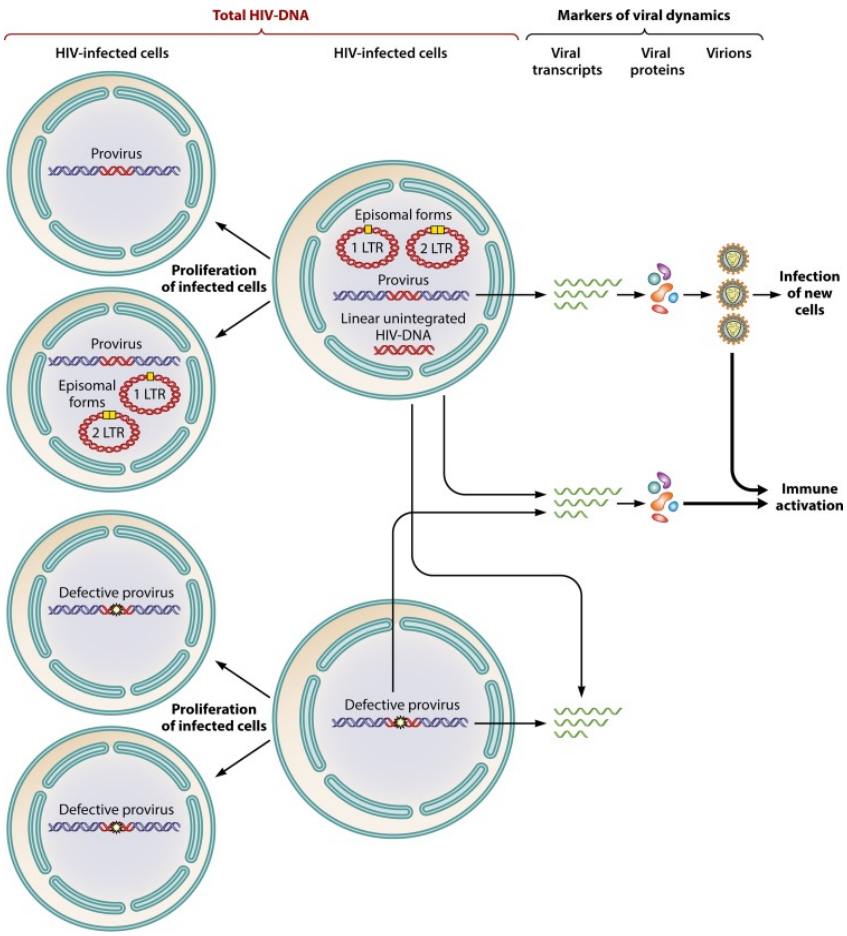


Figure 8 Total HIV-1 CA-DNA consists of several different forms, both integrated and unintegrated, that can contribute to HIV pathogenesis via cell proliferation and persistent immune activation[71].

During the viral replication cycle, it has been reported that more than 40 different viral RNAs are generated in HIV-infected cells by alternative splicing (examples can be seen in Figure 9)[62,74]. The species of viral RNAs produced depend on the stage of the HIV-1 replication cycle and certain protein thresholds that need to be achieved. As mentioned in the section above, short multiply (completely) spliced (ms) transcripts are produced at the initial stages, encoding the Tat, Rev and Nef proteins[62,75]. As the replication cycle progresses and a certain threshold level of the Rev protein is achieved that prevents further transcript splicing and degradation, there is a shift towards production of unspliced (us) and incompletely (or singly) spliced (is) transcripts. These longer transcripts encode the remaining viral proteins (Pol, Gag, Env, Vif, Vpu and Vpr)[75]. This shift in transcript production facilitates the export of the usRNA and isRNA molecules from the nucleus to the cytoplasm, by Rev binding to the Rev-responsive element (RRE)[62,75]. In the final stages of the HIV-1 replication cycle, usRNA is packaged into progeny virions as genomic RNA.

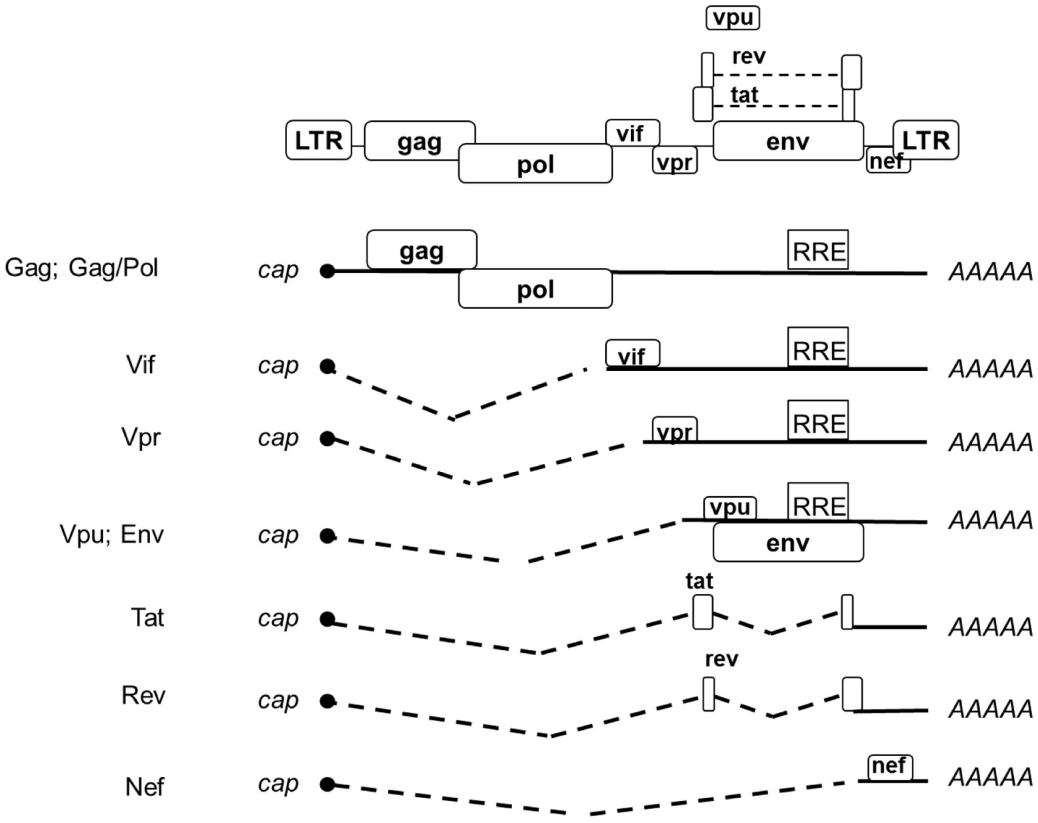


Figure 9 Multiple HIV-1 RNA species, including unspliced, multiply (completely) and incompletely (singly) spliced, are generated through alternative splicing during the HIV-1 replication cycle[74].

It is plausible that in ART-treated suppressed children residual transcription of HIV-1 CA-RNA and protein expression, cause persistent immune activation, inflammation and immunosenescence[76]. All these can in turn lead to continuous strain on the immune system and additional morbidity. Still, little is known about the cells that express proviral RNA in vivo or about the expression levels within them[75,77].

Currently, there are several methods in use for total HIV-1 CA-DNA and CA-RNA quantification, each with their own advantages and limitations. To inform novel treatment strategies and provide better management of patients, our ability to study and measure the viral reservoir using sensitive assays is imperative. The Quantitative Viral Outgrowth Assay (QVOA) is considered the gold standard for the quantification of the replication-competent (non-defective) HIV-1 reservoir[78–80]. However, it is expensive, time-consuming, requires large sample volumes from patients and ultimately underestimates the size of the replication-competent reservoir[81]. An alternative to the QVOA that measures CA-RNA production is the *tat/rev* Induced Limiting Dilution Assay (TILDA)[82]. It is much faster and sensitive than QVOA and requires a lower volume of sample, but it can overestimate the size of the replication-competent reservoir[83] . The most used assays for HIV-1 CA-DNA and CA-RNA detection, especially in clinical settings and paediatric studies, are quantitative Polymerase Chain Reaction-based (qPCR) assays[83–87]. Generally, they are reproducible, quick, simple, cheap and do not require large volumes of patient samples or viral cultures, that could be a major obstacle in large scale clinical studies[88]. One example is the total HIV-1 CA-DNA qPCR assay that can measure viral DNA but cannot distinguish between replication-competent and defective viruses, leading to an overestimation of the reservoir[72,85,89–91]. The Alu-gag PCR assay can quantify the levels of integrated HIV-1 DNA in patients[86,92]. The CA-RNA reverse transcription (RT) qPCR assays can be used to account for ongoing expression of HIV-1 and target different types of the viral transcripts, such as the poly(A) tail of all HIV mRNAs, or specific types of HIV-1 RNA (e.g., usRNA and msRNA) depending on primer location[75,84,93,94].

1.4 Antiretroviral Therapy and Novel Treatments

Several different classes of antiretroviral drugs have been developed to prevent and treat HIV-1/AIDS. The first U.S. Food and Drug Administration (FDA) approved drug was a nucleoside reverse-transcriptase inhibitor (NTRI) called azidothymidine (AZT)[95,96]. Due to the high mutation rate of HIV-1 and resulting development of drug resistance as well as the development of drug-related toxicity side-effects, it became apparent very early on that single-drug treatment regiments were limiting, thus introducing the notion of double or triple drug therapy, also called highly active antiretroviral therapy (HAART)[97,98]. Currently, there are more than 30 antiretroviral drugs available that work through different mechanisms and target various steps of the HIV-1 replication cycle (Figure 10)[99]. These include entry inhibitors, NRTIs, non-NRTIs (NNRTIs), protease inhibitors (PIs), CCR5 antagonists, integrase strand transfer inhibitors (INSTIs), fusion inhibitors, or post-attachment inhibitors[97,99]. As their names suggest, drugs belonging to the NRTI, NNRTI, PI and INSTI categories act to interfere and block the viral reverse transcription and integration into the host genome processes[100]. Entry and fusion inhibitors act to prevent binding with the CCR5 co-receptors and stop conformational changes in viral envelope proteins necessary for viral fusion, respectively[100].

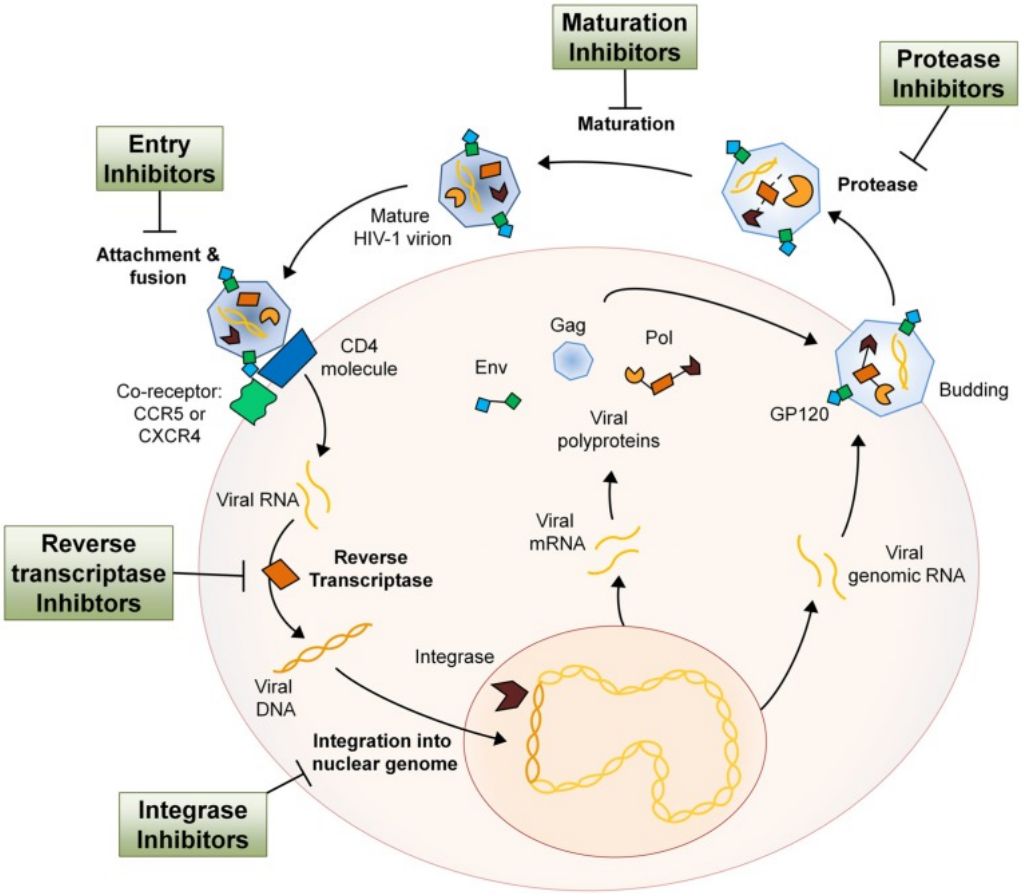


Figure 10 Antiretroviral drugs act by targeting different points of the HIV-1 life cycle. Entry inhibitors obstruct various proteins necessary for viral attachment, receptor binding and fusion. NRTIs and NNRTIs inhibit and block the viral reverse transcription process. Integrase inhibitors block the insertion of the proviral DNA into the host cell genome. Protease inhibitors interfere with the maturation of virions after budding from the host cell by inhibiting the necessary cleavage of viral polypeptides. Finally, maturation inhibitors similarly to protease inhibitors block the viral polypeptide cleavage [101].

1.5 Cure Strategies

To this day, most treatment strategies do not offer a sterilising cure (complete elimination of replication-competent virus) but instead aim at minimising the latent HIV-1 reservoir and controlling HIV-1 replication (functional cure)[102,103]. Thus far there have been several treatment strategies for an HIV-1 cure, functional or sterilising, with moderate to no success[103]. For this reason, better understanding of the viral dynamics of HIV-1 replication is essential in evaluating novel treatment strategies and setting up large scale clinical trial studies in the future.

HIV-1 treatment strategies tested include gene therapy and gene editing approaches (using ZFN, TALEN, CRISPR based technologies), immune-based approaches (vaccines, broadly neutralising antibodies and chimeric antigen receptor T-cells), the "shock and kill" approach (using latency reversing agents), the "block and lock" approach (using latency inducing agents) and haematopoietic stem cell transplants (Figure 11)[104–106]. Gene therapy/editing approaches have seen tremendous progress in the recent years and attempt to interrupt the normal functions of the integrated proviruses by using techniques such as zinc finger nucleases (ZFN), transcription activator-like nucleases (TALENS) or clustered regularly interspaced short palindromic repeats- associated protein 9 (CRISPR/Cas9)[107,108]. The "shock and kill" therapeutic approach uses latency reversing agents (LRAs) to reactivate the latent reservoir and expose it for immune clearance[109,110]. The "block and lock" strategy, aims to permanently silence the integrated proviruses by targeting it at a transcriptional level[111]. In the scope of immune-based cure strategies, broadly neutralising antibodies have been used to target the HIV-1 envelope, help neutralise the virus and block infection[112–114]. Due to their successful use as cancer therapies, chimeric antigen receptor (CAR) T-cells have been seen as a promising approach for an HIV-1 functional cure, but to this day they have not achieved any beneficial effects in HIV-1 clinical trials[112,115]. Similarly, HIV-1 vaccine trials through the years have provided less than efficacious results, potentially due to the viral genetic diversity and high mutation rates[116]. Haematopoietic stem cell transplants showed some promise in unique cases like those of the Berlin and London patients, however they are still deemed high risk and non-practical, since they haven't always been successful and were only used as an option in cases of leukemic patients[108,117]. Nonetheless, none of these HIV-1 cure strategies have led to prolonged remission in clinical studies, highlighting the need for more research and candidate functional cures.

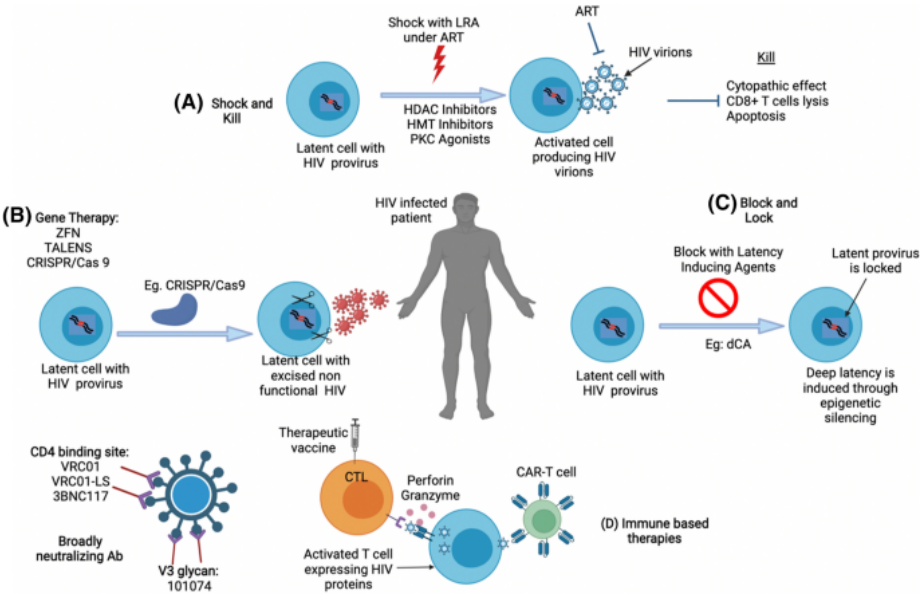


Figure 11 HIV-1 cure strategies. A) Latency reversing agents are being used to eradicate the latent HIV-1 reservoir in a "shock and kill" approach B) Gene therapy and gene editing approaches target the HIV-1 reservoir C) Latency inducing agents are being used to epigenetically silence the integrated viral genome in a "block and lock" approach D) Immune based therapeutic approaches such as vaccines, CAR-T cells and broadly neutralising antibodies are being used to enhance and support the host's immune system and can be used in conjunction with other therapeutic approaches[108].

1.6 Cellular Immune Response to HIV-1

HIV-1 infections give rise to both innate and adaptive immune responses. The generated immune responses are vital for infection control and disease progression delay[46].However, they are undermined by the fact that this virus can target and infect key cells of the immune system, thus compromising the generated immune response and severely affecting the regular function of certain cell populations, like CD4+ and CD8+ T-cells[118].

1.6.1 Innate Immune Response

The first cells responding to an HIV-1 infection are cells of the innate immune system, like macrophages, natural killer (NK) cells and dendric cells (DC)[119–121]. Generally, macrophages have a role in both the innate and adaptive immune system by phagocytosing pathogens and

cellular debris as well as acting as antigen presenting cells and aiding antibody responses[122]. NK cells possess functions that promote both cytotoxicity and cytokine production[123].DCs capture, process and present antigens to lymphocytes thus initiating the adaptive immune response[124]. Thus, these cell types are necessary for early pathogen containment and have a key role in ensuring the quality of the adaptive immune response that will follow[125]. In the context of an acute HIV-1 infection, NK cells, especially NK subsets with cytolytic abilities rapidly expand[126]. However, the virus has developed mechanisms to evade detection by NK cells or to downregulate their functions[127]. Macrophages trigger T-cell responses by presenting HIV-1 derived peptides to CD4+ T-cells and activating CD8+ T-cells by presenting HIV-1 antigens[128,129]. Due to their localisation in different tissues and their ability to infiltrate various organs, it has been reported that macrophages might have a critical role in the spread of HIV-1 within a host and they have also been implicated in mother-to-child transmission cases through breastfeeding[130]. Similarly, DCs also appear to have a dual role in the disease progression. As one of the first cell types that come across this pathogen, they act as antigen presenting cells assisting in subsequent T-cell activation[131,132]. However, as seen before with macrophages, the virus can take advantage of their normal innate immunity roles to gain faster access to CD4+ T-cells[131,132].

1.6.2 Adaptive Immune Response

B lymphocytes (B-cells) derive from haematopoietic stem cells and are responsible for humoral activity in humans[133]. Their development is a lifelong process that starts in primary lymphoid tissues, in the liver prenatally and in the bone marrow throughout life, and continues with maturation in secondary lymphoid tissues, like the spleen and lymph nodes[134]. Generally, B-cells have multiple key roles to support the adaptive immune response, including acting as antigen-presenting cells, producing antibodies, assisting the initiation of T-cell immune responses, and directly contributing to inflammatory pathways[133,134]. In the context of an HIV-1 infection, B-cell abnormalities arise leasing to a defective antibody response against the virus[135,136]. Initial B-cell responses are mostly directed towards non-neutralising HIV-1 epitopes, while later responses fail to act against the rapidly diversifying virus[135,136]. Additionally, the enabling of cell-to-cell transmission, increase of circulating immature B-cells,

loss of the memory B-cell population and the functional exhaustion as the infection progresses have been reported[135,136].

T-cells play an essential role in the cell-mediated adaptive immune response and can provide long-term protection[137]. They are involved in responses specific to pathogens, allergens, and tumours, as well as playing a role in autoimmune and inflammatory diseases[138–140]. T-cells derive from progenitor cells that migrate from the bone marrow to the thymus for maturation, selection events and export to the periphery[141]. Initially, these progenitor cells lack CD4+ and CD8+ receptor expression and must undergo T-cell receptor (TCR) rearrangement to generate double positive (CD4+ CD8+) thymocytes. The double positive cells undergo selection to generate single positive (CD4+ or CD8+) cells that get into the periphery, where they can be separated into different subsets, including naïve and memory T-cells[138]. Naïve T-cells have the capacity to respond to new antigens whereas memory T-cells derive from previous antigen exposure and activation events. In early life, most circulating T-cells are naïve. Memory T-cells increase and accumulate with antigen exposures and vaccinations during childhood but reach a plateau during adulthood, that is maintained over the next decades[138]. This memory T-cell population ensures the generation of a rapid and effective immune response upon antigen re-exposure, through rapid clonal expansion, functional activation, and cytokine secretion[142–144]. The memory T-cell response during secondary pathogen exposure is usually accompanied by epigenetic changes, like DNA methylation, histone modification and chromatin re-structuring, as well as the upregulation of various transcription factors[144,145]. Unfortunately, in cases of multiple pathogen infections memory T-cells might fail to persist and survive[142].

Each T-cell expresses a unique TCR, which can recognise a specific antigen when presented by the major histocompatibility complex (MHC) and MHC-like molecules on antigen-presenting cells and can trigger a cytotoxic response against infected cells or the release of cytokines to regulate inflammation[146,147]. TCRs consist of a variable and a constant domain and a combination of either alpha (α) and beta (β) chains (approximately 95% of TCRs), or gamma (γ) and delta (δ) chains (Figure 12)[148].

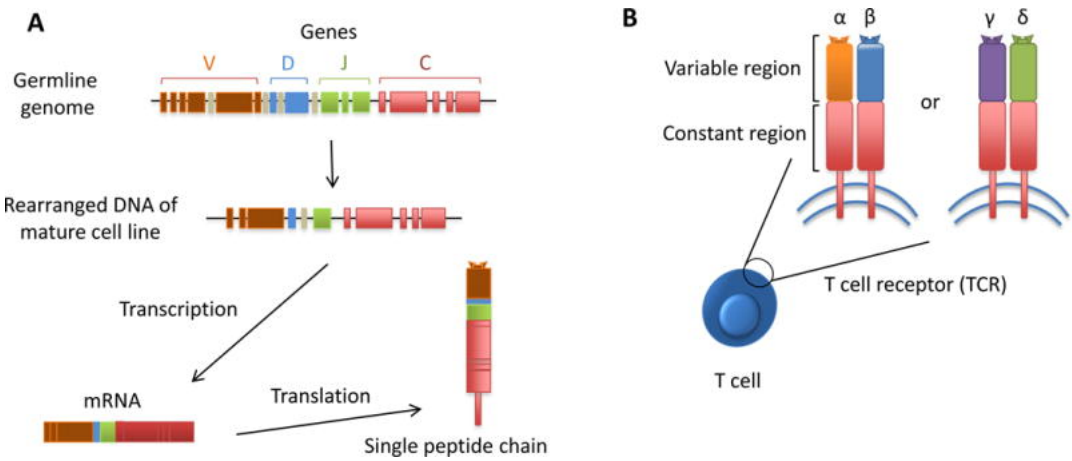


Figure 12 Characteristics of the T-cell receptor: A) Random rearrangement of the variable (V), Diversity (D) and Joining (J) gene segments creates a diverse TCR repertoire. B) The TCR is a heterodimeric polypeptide, consisting of either an alpha and beta chain or a gamma and delta chain[149].

The highly diverse TCR repertoire is created when T-cells randomly re-arrange different variable (V), diversity (D) and joining (J) gene segments within the hypervariable complementarity determining region 3 (CDR3), through a process called V(D)J recombination[149] (Figure 12). In the thymus β chains are rearranged first, with α chains following[150,151]. It has been estimated that a single β-chain can dimerise with as many as 25 different α-chains, contributing this way to the high TCR repertoire diversity[151–153]. Consequently, each TCR sequence is so unique that it enables the recognition and tracking of each specific clonotype, making it a unique biomarker that can be used in the future to inform novel treatment strategies[154]. The human body is estimated to generate approximately 10^{13} different TCR clonotypes[155]. HIV-1 infection can severely affect the immune system, with adult studies showing that HIV-1+ individuals present with less diverse TCR repertoires, depletion of public TCR sequences and small numbers of highly expanded sequences[156].

As previously mentioned, HIV-1 primarily infects CD4+ T-cells and leads to substantial reduction and ultimately depletion of their numbers if the infection remains untreated[46]. During the initial stages of an acute HIV-1 infection, CD4+ T-cells have a contradicting role, by assisting the production of additional virions but also stimulating cellular immune responses, like assisting in the activation of cytotoxic CD8+ T-cells[157]. CD8+ cytotoxic T-cells play a very crucial role in the control of viral replication and suppression of HIV-1 binding and

transcription. Early on following the onset of the acute phase of infection, they can recognise HIV-1 infected cells and lead to their destruction through an MHC-dependent manner[158–160]. It is worth noting that timing of ART initiation is very important as it can positively affect the quality of HIV-1 specific CD4+ T-cell responses and ultimately prevent their depletion[161].

Recently another T-cell subset that is abundant in the mucosal tissues, the liver and peripheral blood, called mucosal-associated invariant T-cells (MAIT), has gained interest in HIV-1 infection studies. MAIT cells develop in a thymus-dependent manner and express a CD161 receptor and a semi-invariant TCR receptor, with an invariant alpha and a restricted beta chain repertoire[162,163]. They recognise bacterial and fungal derived metabolites, thus playing a crucial yet not completely understood role in the protection against microbial infections[164–166]. In healthy individuals, MAIT cells can represent up to 10% of the circulating T-cells[167]. From those, CD8+ cells constitute the majority of the MAIT cell pool (approximately 80%) and CD4+ and CD4-CD8- cells constitute the minority (5% and 15% respectively)[164,165]. It has been reported that MAIT cell levels are diminished during chronic HIV-1 infection[168]. Their cytotoxic (e.g., Perforin, Granzyme B etc.) and cytokine (e.g., IFN-γ, TNF, IL-17 etc.) enzyme production is also impaired because of immune exhaustion, with ART only partially restoring their function[168–170]. To this day, the mechanisms behind MAIT cell depletion and dysfunction remain unclear, but it has been suggested that these impair the antimicrobial immune defences of HIV-1+ patients, making them susceptible to opportunistic infections[156,164,169,171].

1.6.3 Differences Between Paediatric and Adult Immune Responses to HIV-1

Throughout the years, studies have shown that paediatric immune responses to HIV-1 can differ from those of adults. That can potentially be attributed to the fact that infancy is critical period for the development of the immune system. The immune system goes through a lot of changes in the early stages of life due to the abrupt transition from a sterile environment to an environment with repeated immune stimuli[172]. In contrast to the developing and adapting immune system of children, immunologically matured adults are acclimatised to antigen exposure and have a different immune response. Moreover, substantial phenotypic and functional differences in immune cell populations can be seen[172]. It has been previously

reported that children have a greater percentage of naïve T-cells and a decreased percentage of memory T-cells, in contrast to adults where the reverse proportion of T-cell populations are observed[173,174]. In terms of HIV-1 infections, adults present better virological and worse immunological responses. In contrast, children have better immunological responses, but their poorer virological outcomes may lead to ART resistance[175].

1.7 Importance of Paediatric HIV-1 Studies

Despite the tremendous progress made towards preventing vertical HIV-1 transmission, paediatric infections continue to be a reality. Management of this cohort remains challenging due to limitations in the existing treatment options, compared to adult patients. Though, recent advances in formulations and ART drug options have been promising[176,177].

To date, most studies have been conducted in adult cohorts. However, adults and children differ in terms of both immune and treatment responses. For that reason, the guidelines for paediatric care and treatment should be adapted depending on the evidence presented in paediatric studies. Due to the existing gaps in our knowledge, it is apparent that paediatric studies are vital for patient management.

CHAPTER 2: MATERIALS AND METHODS

This chapter describes the general materials and methods used throughout the thesis. Any methods modified for specific chapters will be discussed in detail in the relevant sections.

2.1 Sample Collection and Ethics

All research described in Chapters 3-6, was conducted in accordance with the recommendations of Great Ormond Street Hospital for Children NHS Trust and the Institute of Child Health (ICH) Research Ethics committee with written informed consent from all participants. Blood samples from healthy adult volunteers were collected to be used as controls.

2.1.1 Child and Adolescent Reservoir Measurements on early suppressive ART (CARMA) Study

For the CARMA research study (UCL R&D: 17IR17), peripheral blood mononuclear cells (PBMCs) were isolated from 19 UK patients (APPENDIX 2) and subsequently used for T-cell receptor repertoire sequencing and thymic output quantification (Chapter 4). Additionally, nucleic acid samples isolated from all 40 patients (APPENDIX 1), collected from 7 sites across Europe involved in this study (UK, Italy, and Spain), were used for determining the presence of HIV-1 CA-RNA (Chapter 3)[4]. To our knowledge, all CARMA participants were unrelated. An example of the cohort characteristics is presented below (

Table 1).

Table 1 Characteristics of the CARMA cohort (adapted from [4]). Continuous data are presented as median (IQR), categorical data are presented as number of patients (%).

	All (n=40)

Region, No. (%)	
United Kingdom	19 (47.5)
Italy	16 (40.0)
Spain	5 (12.5)
Sex, No. (%)	
Female	27 (67.5)
Male	13 (32.5)
Race/ethnicity, No. (%)	
White	13 (32.5)
Black/Black African	21 (52.5)
Other	6 (15.0)
Age at HIV diagnosis, months	4.17 (2.19-6.32)
AIDS diagnosis, No. (%)	
No	25 (62.5)
Yes	15 (37.5)
Age at ART start, months	4.08 (0.25-6.23)
Follow-up, years	13.0 (8.19-16.1)

*Abbreviations: HIV, Human Immunodeficiency Virus; AIDS, Acquired Immune Deficiency Syndrome; ART, Antiretroviral Therapy

2.1.2 Impact of ART Initiation Timing on the TCR Repertoire Sub-Study

Chapter 5 is a sub-study investigating the TCR repertoire by comparing samples sequenced from 5 UK CARMA (Group A) patients to 7 samples sequenced during a collaborative project with Dr Katrine Schou Sandgaard (Lymphocyte development (LD) in paediatric HIV infection study, 17/SC/0218, UCL R&D: 17IR08). From the LD study, samples from 5 HIV-1+ patients (Group B), that initiated ART late, and 2 healthy controls (Group C) were selected (APPENDIX 3). The LD samples were initially collected as part of a vaccine study (10/H0713/1, R&D: 08ID21), investigating the impact of vertically acquired HIV-1+ on humoral immunity development[178].

Table 2 Characteristics of individuals selected to investigate the impact of ART initiation timing (adapted from[4,178]). Continuous data are presented as median (IQR), categorical data are presented as number of patients (%).

	HIV+ Group A (n=5)	HIV+ Group B (n=5)	Control Group C (n=2)
Sex, No. (%)			
Female	3 (60.0)	3 (60.0)	0 (0.0)
Male	2 (40.0)	2 (40.0)	2 (100.0)
Race/ethnicity, No. (%)			
White	0 (0.0)	0 (0.0)	1 (50.0)
Black/Black African	3 (60.0)	5 (100.0)	1 (50.0)
Other	2 (40.0)	0 (0.0)	0 (0.0)
Age at HIV diagnosis, years	0.18 (0.14-0.29)	1.3 (0.42-4.75)	NA
Age at ART start, years	0.22 (0.01-0.38)	10.65 (10.30-10.93)	NA
Age at study sample collection point, years	17.22 (16.06-17.66)	18.54 (15.18-18.61)	22.38 (22.32-22.44)
CD4+ count at sample collection point, cells/µl	1043 (652.0-1130.0)	320.0 (220.0-740.0)	924.0 (924.0-924.0)
CD4 %	41.0 (37.0-52.0)	28.7 (22.0-45.3)	68.4 (67.6-69.2)
CD8+ count at sample collection point, cells/µl	636.0 (450.0-665.0)	889.0 (852.44-969.83)	209.46 (187.72-231.20)
CD8 %	28.0 (27.0-34.0)	71.2 (43.3-73.0)	21.5 (21.3-21.6)
CD4:CD8	1.5 (1.1-2.1)	0.36 (0.19-0.92)	4.61 (4.13-5.09)

*Abbreviations: HIV, Human Immunodeficiency Virus; ART, Antiretroviral Therapy

2.1.3 PENTA11 Sub-Study

For the PENTA11 (main trial: ISRCTN 36694210) sub-study presented in Chapter 6, samples from previously sorted and extracted, by Dr Athina Soragia Gkazi and Dr Katrine Schou Sandgaard, CD4+ memory T-cells (stored at -80˚C) collected from nine patients, were selected to investigate the longitudinal effects of treatment interruption on the T-cell receptor repertoire (APPENDIX 4)[179–181]. The selected samples included 4 different timepoints (weeks

0/2, 12, 48, 150) for 4 patients under planned treatment interruption (PTI) and 2 timepoints (Weeks 0/2 and 150) for 5 patients on continuous treatment (CT).

Table 3 Characteristics of individuals selected to investigate the impact of treatment interruption on CD4+ memory T-cells (adapted from[181]). Continuous data are presented as median (IQR), categorical data are presented as number of patients (%)

	PTI Group (n=4)	**CT Group (n=5)**
Age at first timepoint, years	9.92 (8.94-10.83)	9.58 (5.33-10.08)
Age at final timepoint, years	15.55 (14.82-15.92)	14.08 (11.83-15.33)
CD4+ count at first timepoint, cells/ µl	969.5 (870.0-1091.75)	1360.0 (930.0-1370.0)
CD4+ count at final timepoint, cells/ µl	935.5 (843.25-960.0)	967.0 (646.0-1030.0)
≤50 copies/ml Viral load at first timepoint, (%)	4 (100.0)	5 (100.0)
≤50 copies/ml Viral load at last timepoint, (%)	4 (100.0)	5 (100.0)

*Abbreviations: PTI, Planned Treatment Interruption; CT, Continuous Treatment

2.2 Blood Sample Processing

For the CARMA study, a protocol for blood sample processing was established and followed across all sites. A total of 36-80ml of whole blood per patient (Table 4) were collected in Ethylenediaminetetraacetic acid (EDTA) vacutainer tubes. Sample volumes were determined depending on the weight of the participants and were conforming with the guidelines present in paediatric research[182]. All samples were processed in the ICH Containment Level 3 (CL3) laboratory within 12 hours of the blood draw in accordance with the standard operating

protocol (SOP) followed across the participating sites and local CL3 laboratory regulations. It has previously been reported that the time between venepuncture to peripheral blood mononuclear cell (PBMC) isolation and cryopreservation can be crucial to PBMC viability and can affect their performance in different assays[183].

Table 4 Age-dependent volume permitted per each blood draw for CARMA participants.

Weight (kg)	First extraction (ml)	Second extraction (ml)	Third extraction (ml)
20.0-29.9	16-24	16-24	16-24
30.0-60.0	25-48	25-30	-
>60.0	50-70	-	-

2.2.1 Plasma Isolation

EDTA vacutainer tubes were centrifuged at 800xg for 10 minutes with brake (at 18-25°C). Following centrifugation, within a class II biosafety cabinet, the plasma was carefully aspirated, so not to disturb the buffy coat layer. Processing was completed by a second centrifugation of the plasma at 800xg for 10 minutes. Aliquots of 1ml plasma were transferred into the appropriate number of microtubes and immediately stored at -80°C for future use.

2.2.2 Peripheral Blood Mononuclear Cell Isolation

PBMCs were isolated by density gradient centrifugation using Lymphoprep (STEMCELL Technologies). This method is based on the principle that different cell types present with differences in density, thus making it possible to create separate layers during centrifugation. Generally, granulocytes and erythrocytes have higher density than mononuclear cells, which remain at the Lymphoprep interface[184]. Following plasma removal, an equivalent volume of sterile Gibco Phosphate-buffered saline (PBS), pH 7.4 (ThermoFisher Scientific) was added to the remaining sample to bring it back to its original whole blood volume. The PBS/blood mix was transferred to a T75cm^2 tissue culture flask and depending on the estimated blood

volume, was diluted at a 1:1 ratio with sterile Gibco RPMI 1640 Medium (ThermoFisher Scientific). According to the total volume of blood/RPMI mixture and the manufacturer's guidelines, 50ml centrifuge tubes were prepared containing the appropriate volume of Lymphoprep (Table 5).

Table 5 Recommended volumes and tube sizes for PBMC isolation using Lymphoprep (adapted from Stem Cell Technologies)

PBS/Blood Mix (ml)	RPMI (ml)	Lymphoprep (ml)	Falcon tube size (ml)
1	1	1.5	5
2	2	3	15
5	5	10	50
10	10	15	50
15	15	15	50
12.5	12.5	25	50

The RPMI/blood mix was then slowly (dropwise) and carefully added on top of the Lymphoprep in a sharp angle, so not to mix the two layers. The samples were centrifuged immediately at 800xg for 30 minutes, without brake. Using a 2ml sterile Pasteur Pipette (pastette) the white buffy coat layer was aspirated and transferred to a 50ml falcon tube. The PBMCs were then washed with RPMI by a centrifugation at 800xg for 10 minutes with the maximum brake. After the cells were pelleted, the supernatant was aspirated carefully using a 10ml or 25ml (depending on sample volume) Serological Pipette (Stripette). The cells were washed again with RPMI by centrifugation at 800xg for 10 minutes. The supernatant was removed, the pellet was resuspended in 10ml of RPMI and centrifuged at 800xg for 5 minutes. After centrifugation the pellet was once again resuspended in 10ml RPMI and the cells were counted using the following steps: in an 1.5ml microcentrifuge tube (Eppendorf), 10µl of resuspended cells were mixed with 10µl Trypan Blue dye. 10µl of the cell/Trypan Blue mix was transferred onto a haemocytometer (Figure 13). Lymphocytes were counted in total of sixteen small squares (contained within the red marked area in the figure below), using the following equation: Number of counted cells x dilution factor in grid x 10^4. For example: 50

cells x 2 (10µl of cell suspension and 10µl of Trypan Blue) x 10 (10ml of resuspended cells) x 10^4.

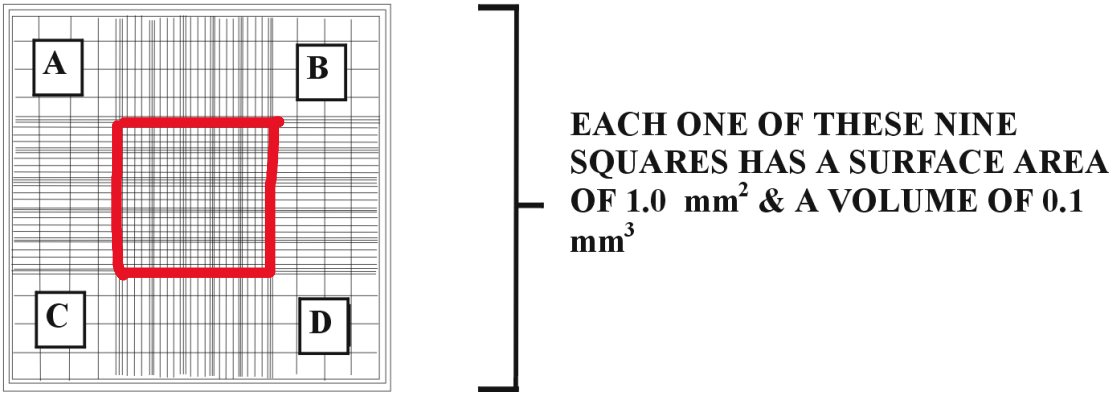


Figure 13 Representation of the haematocytometer surface. Sixteen small squares from the central square were used to count PBMCs.

After counting, the resuspended PBMCs were centrifuged one last time at 800xg for 10 minutes, with the brake on. 1ml PBMC aliquots were produced at concentrations of 0.5 to 5 x10^6 cells per ml in 1.5ml cryovial tubes, after resuspension in the appropriate volume of pre-chilled Foetal Calf Serum (FCS) and pre-chilled freeze medium (80% FCS +20% Dimethyl Sulfoxide (DMSO)), at a a 1:1 ratio. Thus, bringing the final mix concentration to FCS +10% DMSO. The freeze medium prevents the intracellular formation of ice crystals, that would otherwise compromise the cell structure during the thawing process thus decreasing cell viability[185–187]. To ensure stepwise temperature decrease and avoid intracellular freezing, all aliquots were immediately placed in a "Mr. Frosty" freezing container (ThermoFisher Scientific) and moved to a -80˚C freezer[188]. After 72 hours the PBMC aliquots were moved into a liquid nitrogen (LN_2) storage tank.

2.3 Fluorescence-activated Cell Sorting (FACS)

Fluorescence-activated cell sorting (FACS) is an important tool used in biological research. It is a type of flow cytometry with the added benefit of enabling users to separate a heterogenous cell mixture into different populations. Like a flow cytometer, a FACS cytometer

comprises of a fluidics system, an optical system, an electronic network, and a computer (Figure 14)[189]. In summary, cells suspended in liquid pass through a laser beam and their light scatter and fluorescence is detected. Based on the parameters set by the users, an electrical charge is added to the cells which helps the cells to be sorted into the appropriate collection tubes[189,190].

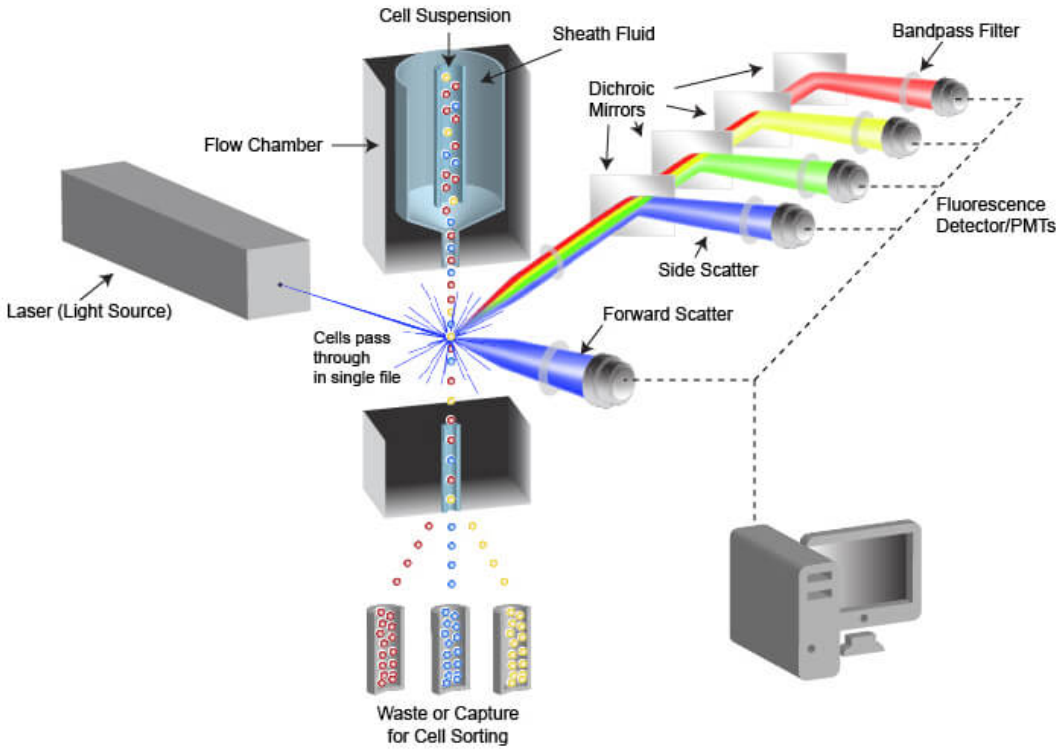


Figure 14 Principles and main components of a flow cytometer: Cells suspended in liquid are transported in a single file towards the light source (laser beam) through the fluidics system. The optical system components collect the emitted scattered light and fluorescence. Two types of scattered light are collected and quantified, forward scatter (cell size) and side scatter (complexity and optical density of cells). The emitted fluorescence derives from the fluorescent dyes/antibodies used in the experiment. Photodetectors convert the generated light and fluorescence emissions into electrical signals and then in turn to digital signals by the electronics network. Specialised computer software presents the data in the form of dot plots and histograms (adapted from Cell Signalling Technologies)[190–192].

For the purposes of the experiments described in Chapters 4 and 5, CD4+ and CD8+ T-cell populations were sorted by FACS using the following protocol.

2.3.1 Thawing and counting cells

Complete RPMI (cRPMI) media was prepared by adding 10% FCS (StemCell Technologies), 1% penicillin-streptomycin (ThermoFisher Scientific) to sterile Gibco RPMI 1640 Medium with L-Glutamine (ThermoFisher Scientific). Due to the lack of access to a heat block or water bath (37^{o}C) in the CL3 laboratory and the requirement to handle cryovial tubes within a Class I microbiology safety cabinet, PBMCs were thawed by hand-warming for approximately 60 seconds, after which a small volume of preheated cRPMI media was added to the vial and mixed thoroughly by pipetting up and down. The resuspended cells were then transferred in a 15ml centrifuge tube with pre-heated cRPMI. 500µl cRPMI were added in the cryovial tubes to retrieve any remaining cells. The 15ml tubes were centrifuged at 1,500rpm for 10 minutes to pellet the cells. The supernatant was discarded, and the cells were resuspended in 1ml FACS buffer (PBS and 2% FCS). The PBMCs were then counted (as described in section 2.2 of this chapter) and divided in five aliquots (containing ~0.5x10^{6} PBMCs each) for subsequent FACS.

2.3.2 Surface staining

PBMC aliquots were spun down for 10 minutes at 1,500rpm in FACS buffer. The supernatant was discarded, and the cells were resuspended in 100µl FACS buffer. After turning off the light in the safety cabinet to avoid bleaching of the fluorescent stains, 2µl of the corresponding surface antibodies were added to the following tubes:

i. APC anti-human CD4 (APC CD4) antibody (BioLegend)
ii. APC/Cyanine7 anti-human CD8 (APC/Cy7) antibody (BioLegend)
iii. APC CD4 and APC/Cy7 antibodies

For the correct interpretation of the data that would be acquired, proper controls were prepared and used. The first two tubes, containing each antibody separately were used for compensation, to determine any spectral overlap[193]. A viability stain was also included to

enable the exclusion of dead cells from the analysis. After an incubation for 10 minutes in the dark (at room temperature), 2µl of LIVE/DEAD Fixable Aqua-Dead Cell Stain (ThermoFisher Scientific) were added to the tubes above and the tubes below:

iv. LIVE/DEAD Fixable Aqua-Dead Cell stain
v. Unstained

All samples were thoroughly mixed and incubated for another 10 minutes in the dark. 1ml PBS was then added and the samples were centrifuged at 1,500rpm for 10 minutes. The supernatant was discarded.

2.3.3 PBMC fixation

BD CellFix (BD Biosciences) was diluted 1:10 in sterile water at room temperature and the appropriate volume (200µl per 0.5 - $1x10^5$ PBMCs) was added in each tube. Samples were placed at 4˚C for 30 minutes. After the incubation, all samples were centrifuged at 1,500rpm for 10 minutes, the supernatant was discarded, and they were re-suspended in 200µl BD CellFix (BD Biosciences). Finally, they were transferred to the Flow Cytometry Core Facility (UCL GOS Institute of Child Health). Using the appropriate gating strategy (Figure 15) all samples were sorted, by Stephanie Canning and Dr Ayad Eddaoudi, into the required CD4+ and CD8+ populations, using the FACSAria III cell sorter (BD Biosciences).

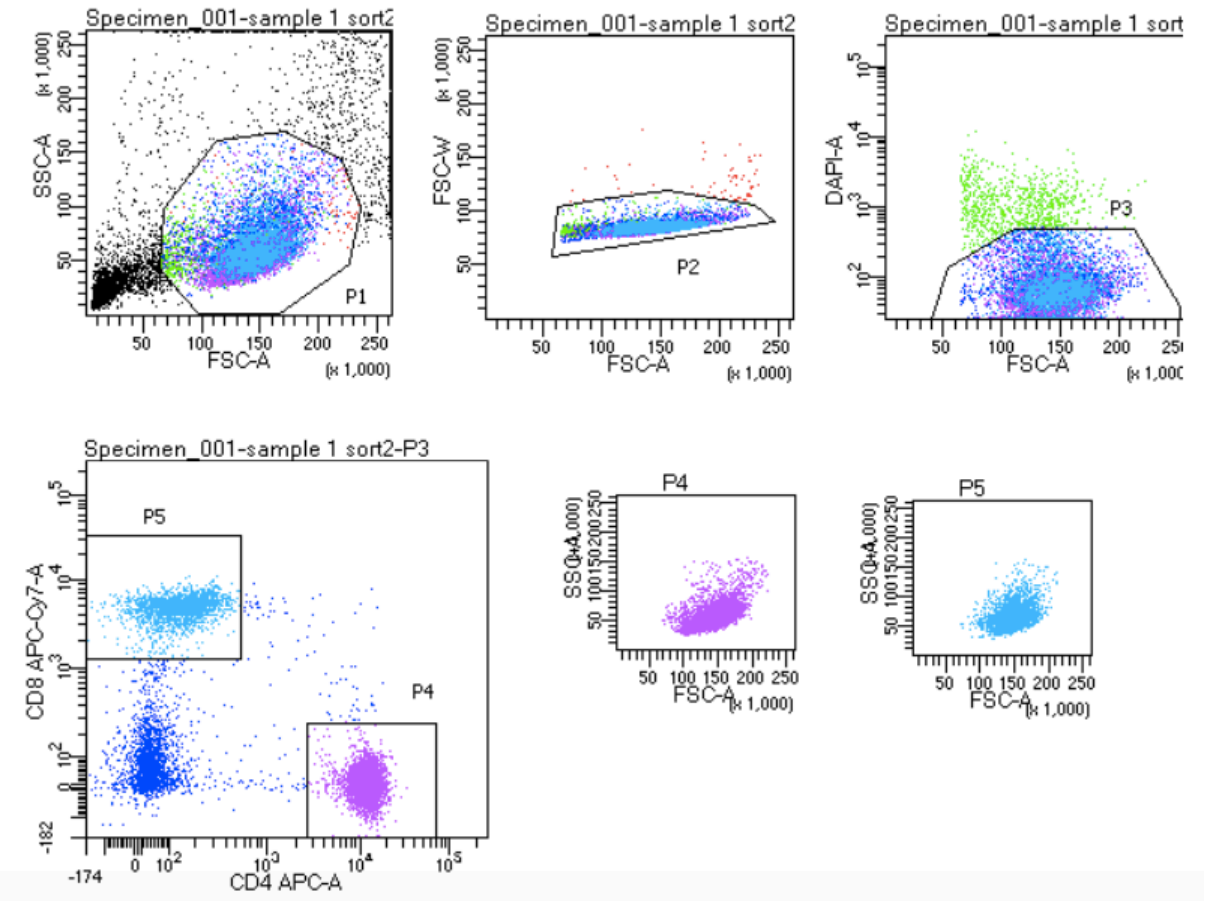


Figure 15 Example of gating strategy followed at the Flow Cytometry Core Facility for the CARMA samples. Gating is a method used to select desired cell populations and exclude debris, thus allowing for the simultaneous analysis of an enormous number of parameters. Flow cytometry data are collected, and boxes are drawn around specific subpopulations, in this case CD4+ and CD8+ T-cells, for further analysis[194].

2.4 Nucleic acid extraction

Following sorting of the PBMC samples into CD4+ and CD8+ T-cell populations on the FACSAria III sorter (BD Biosciences), all samples were separated and used for RNA (2/3 of each sorted sample) and DNA extraction (1/3 of each sorted sample). Since all samples consisted of relatively low T-cell numbers, the nucleic acid extractions were done following previously optimised, by lab group members, protocols.

2.4.1 RNA extraction from fixed/sorted T-cells

RNA was extracted with a combination of the FFPE RNA kit (QIAGEN) and the RNA Micro kit (QIAGEN) manufacturer's protocols. T-cells were pelleted by centrifugation at 6,000rpm for 10 minutes. The supernatant was removed and 150µl of Buffer PKD were added and mixed by vortexing, before the addition of 10µl proteinase K. The samples were incubated at 56˚C for 15 minutes, followed by incubations at 80˚C for 15 minutes and on ice for 3 minutes. 320µl of Buffer RBC were added to the lysate and mixed thoroughly. To continue, 720µl of 100% Ethanol were mixed in by pipetting. 700µl of the sample were transferred to a RNeasy MinElute spin column placed in a 2ml collection tube and centrifuged for 15 seconds at 8,000xg. The flow-through was discarded and the step was repeated until the entire sample had passed through the column. To finish the extraction the RNA Micro kit (QIAGEN) manufacturer's protocol was followed, starting with the addition of 350µl of Buffer RW1 onto the column. The RNA was eluted in 16µl RNase-free water.

2.4.2 DNA extraction from fixed/sorted T-cells

DNA was extracted by following an optimised version of the DNA Micro kit (QIAGEN) manufacturer's protocol. To start, T-cells were pelleted by centrifugation at 6,000rpm for 10 minutes. The supernatant was then removed, the pellet was resuspended in 100µl of Buffer ALT and 10µl of Proteinase K were added. Following 15 seconds of vortexing, all samples were incubated at 56˚C for 1 hour, followed by a second incubation at 90˚C for another hour. 100µl of Buffer AL were added and the samples were mixed by vortexing for 15 seconds. Then, 100µl of 100% Ethanol were added, followed by vortexing and a 3-minute incubation at room temperature. To continue, the samples were carefully transferred onto the QIAamp MinElute columns and the manufacturer's protocol was followed. The DNA was eluted in 35µl Buffer AE.

2.5 Quality control methods for nucleic-acid extractions and library generation

The isolated nucleic acid concentration was established by the following methods:

2.5.1 NanoDrop

DNA and RNA samples were quantified using the NanoDrop (ThermoFisher Scientific) spectrophotometer. NanoDrop allowed for quick quantification of nucleic acids and detection of sample purity, using the 260/280 absorbance ratio (~1.8 for DNA purity and ~2.0 for RNA purity ratios are generally acceptable). Due to limitations caused by NanoDrop's capacity to overestimate concentrations by measuring nucleic acids without clearly discriminating between DNA and RNA, Qubit was used in parallel.

2.5.2 Qubit

Quantitation of DNA and RNA was performed using the Qubit 4 Fluorometer (Invitrogen) and following the recommended protocols for the Qubit dsDNA HS and RNA HS assay kits (ThermoFisher Scientific) according to the manufacturers protocol. Qubit can offer a more accurate quantification method, compared to NanoDrop, since the assay kits that are being used are specific to the target of interest and the dyes fluoresce only when bound to the selected molecule.

2.5.3 TapeStation

The integrity, size, and quality of the NGS amplicons was analysed using the High Sensitivity D1000 ScreenTape (Agilent) alongside the 2200 TapeStation instrument (Agilent). This was carried out as per manufacturer's protocol and analysed using the accompanying software.

2.6 Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR)

Polymerase Chain Reaction (PCR) is one of the fundamental methods used in research. It was first developed by Kary Mullis in 1983, and utilises an enzyme called DNA polymerase, along

with target specific primers to exponentially amplify a single copy of DNA. This amplification is achieved by three repeating steps that include denaturation, annealing and extension (Figure 16).

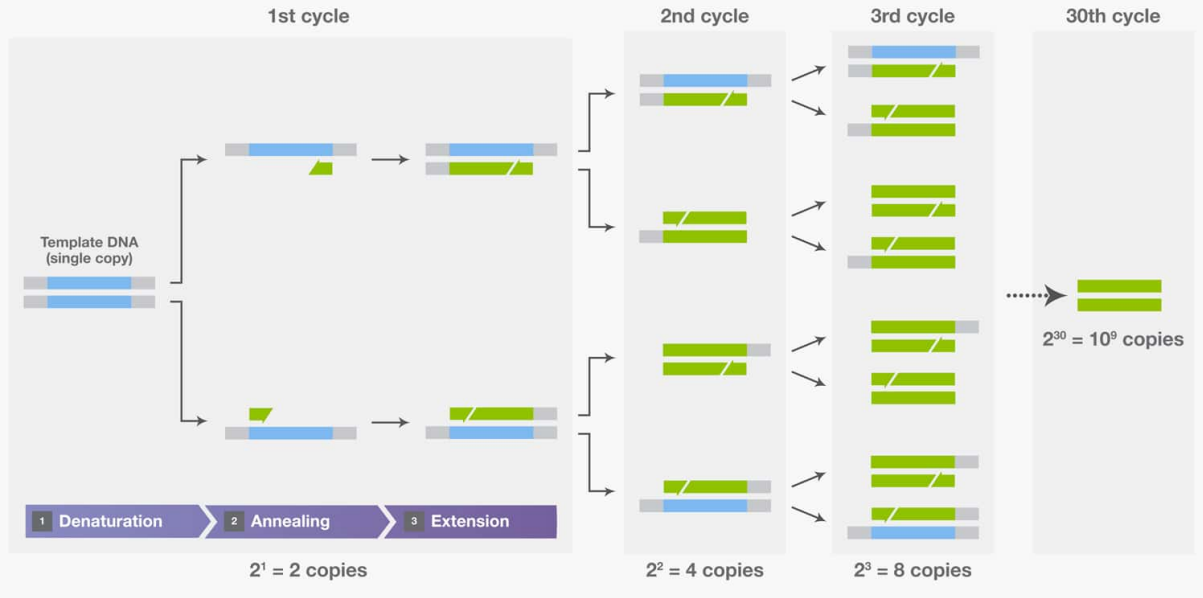


Figure 16 Basic steps of PCR: Template DNA is denatured into single strands and primers are annealed onto target regions of each strand. DNA polymerase helps extend each primer along the template strand starting at the 3' end. These three basic steps are repeated to produce numerous copies of the selected DNA target (adapted from ThermoFisher Scientific).

For the purposes of the experiments presented in Chapter 3, a one-step Quantitative Reverse Transcription PCR (qRT-PCR) method was developed utilising RNA as its template. This method, based on the basic principles of the PCR, can be used to measure the products generated during each PCR cycle by utilising fluorescent-labelled probes, containing a reporter and a quencher fluorophore[195]. The fluorescence is measured during each cycle since the signal should increase proportionally to the amount of replicated DNA. The fluorescent signal is generated during the extension and probe cleavage step (Figure 17). During this step, the probe is degraded leading to the emission of fluorescence by the reporter that is no longer quenched[195].

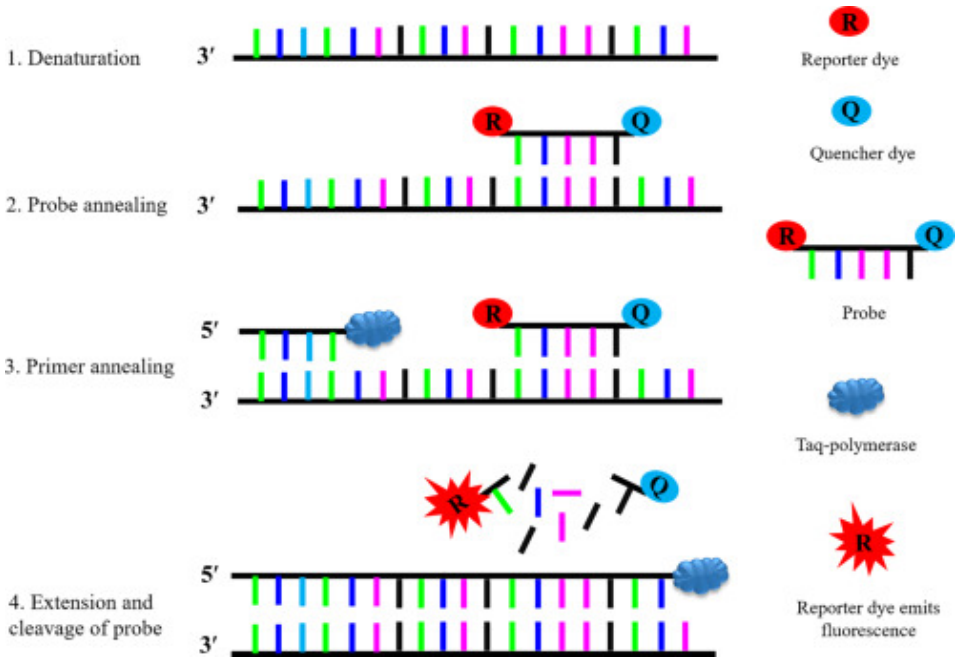


Figure 17 Overview of the TaqMan probe-based qRT-PCR. This method takes advantage of the ability of Taq polymerase to cleave an oligonucleotide probe, generating a detectable fluorescent signal [196].

2.6.1 Quantitation of HIV-1 Cell-Associated RNA (HIV-1 CA-RNA)

To quantify the amount of HIV-1 CA-RNA present in the CARMA samples a one-step qRT-PCR assay was developed in collaboration with Dr Sarah Watters and Dr Kathleen Gartner. For the one-step qRT-PCR set up, RNA samples were added into a reaction along with sequence specific primers, probes labelled with fluorescent dyes and a master mix, containing DNA polymerase, reverse transcriptase, buffer, dNTPs, and RNase inhibitor (Figure 18).

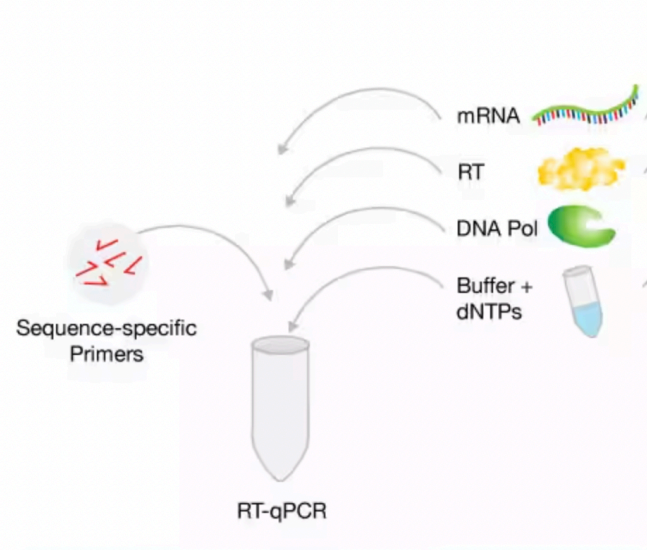


Figure 18 Basic components of the one-step RT-qPCR reaction set up: The master mix and RNA template are combined under conditions that support the reverse transcription and subsequent PCR step (adapted from ThermoFisher Scientific).

A standard curve was generated for this assay by utilising an *in vitro* transcribed positive strand HIV-1 RNA was diluted in an RNA extract derived from HL-60 cells. This was done to stabilise the RNA and offer a background for the reference genes. The HIV-1 RNA fragment used was derived from the HXB2 HIV-1 reference genome sequence (accession K03455.1; positions 1 to 5,619 presented in Figure 19) and was generated using MEGAscript T7 RNA polymerase (Life Technologies) by Dr Eloise Busby (research analyst, LGC Group).

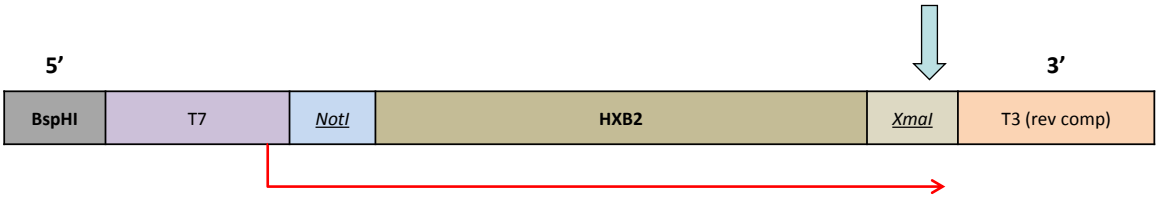


Figure 19 Design of HXB2 plasmid insert for in vitro synthesis of RNA transcript (kindly provided by LGC Group research analyst E. Busby, 2018)

The RNA transcript was diluted to approximately ~2x10^{5}copies/µl in RNA storage solution (ThermoFisher Scientific) and stored at -80°C. The transcript includes sequences for HIV-1 5'-

LTR, pol, gag and vif genes. The final stock concentration of the assay standard was 6.525ng/µl or $1x10^4$copies/µl of HIV-1 RNA.

Before the qRT-PCR, all samples were DNase treated (for volumes refer to Table 6) followed by heat inactivation of the enzyme using a temperature regimen of 37˚C and 75˚C degrees for 10 minutes respectively.

Table 6 Reagents and volumes (µl) required for DNase treatment.

Reagent	Volume (1x)
DNase Buffer (10x)	1
DNase 1 (2U/µl)	0.5
Template RNA	10
Total	**11.5**

Once the standard curve was generated, the RNA samples were amplified using an qRT-PCR for LTR and pol to detect total and unspliced HIV-1 CA-RNA respectively. TATA-box-binding protein 1 (TBP1) and Importin-8 (IPO8) were used as reference genes[197,198]. The volumes required for the set-up of each qRT-PCR, the primer sequences, and the selected settings on the Bio-Rad CFX96 instrument can be found in Tables 7-10. The standard curve was prepared by performing five 10-fold serial dilutions on the DNase treated assay standard. As a positive control, a known dilution of 8E5 cells was used (ATCC, CRL-8993). The 8E5 cells theoretically contain a single defective HIV-1 DNA copy/cell. No template controls (NTC) were included in each run to detect any possible contamination. Initial raw data analysis was performed with the help of Dr Kathleen Gartner, using the Bio-Rad CFX Maestro software. Excel (Microsoft, version 16.65) packages and the Prism software (GraphPad, version 9.0.0) were used for statistical analysis of the data and plot generation.

Table 7 Reagents and volumes (µl) required for the total HIV CA-RNA assay.

Reagent	Volume (1x)
HIV LTR F (10µM)	0.2
HIV LTR R (10µM)	0.2
HIV LTR probe (10µM)	0.2
TBP1 primer/probe	0.5
IPO8 primer/probe	1
TaqMan Fast Virus 1-Step Master Mix (4x) (ThermoFisher Scientific)	5
H_2O	2.9
RNA template	10
Total	20

Table 8 Reagents and volumes (µl) required for the detection of unspliced HIV-1 CA-RNA.

Reagent	Volume (1x)
HIV Pol F (10µM)	0.8
HIV Pol R (10µM)	0.8
HIV Pol PG probe (10µM)	0.4
TBP1 primer/probe	0.5
IPO8 primer/probe	1
TaqMan Fast Virus 1-Step Master Mix (4x) (ThermoFisher Scientific, Cat#: 4444427)	5
H_2O	1.5
RNA template	10
Total	20

Table 9 RT-qPCR temperature regimen on the Bio-Rad CFX96 instrument.

Step	Temperature (˚C)	Time	Cycles
RT	50	5 minutes	1
RT Inactivation/Denaturation	95	20 seconds	1
Amplification	94	15 seconds	45
	60	90 seconds	

Table 10 Primer and probe sequences for the HIV CA-RNA assays[89].

Oligonucleotide Primer/Probe Name	Sequence (5' → 3')	HXB2 nucleotides
HIV1LTRTaq1 F	GCCTCAATAAAGCTTGCCTTGA	525-543
HIV1LTRTaq2 R	GGCGCCACTGCTAGAGATTTT	622-639
HIV1LTRTaqPR	FAM-TGTGACTCTGGTAACTAGAGATCCCTCAGAC-TAMRA	577-604
HIVPolF	TGTACCAGTAAAATTAAAGCCAGGAA	2570-2595
HIVPolR	TATGGATTTTCAGGCCCAATT	2596-2626
HIVPol Probe	FAM-TGGATGGBCCAARRGTYAAACARTGGCCATT-TAMRA	2695-2716
TBP1 primer/probe	Bio-Rad	
IPO8 primer/probe	Bio-Rad	

2.7 Next Generation Sequencing (NGS)

Next Generation Sequencing (NGS) is a term used to describe a sequencing technology, that in contrast to previous technologies like Sanger sequencing, allows for parallel and high depth sequencing[199]. This technique can be used to sequence entire genomes or like in this project

for specific areas of interest. There are several different NGS platforms and two main sequencing approaches, short-read, and long-read sequencing, available. In this instance, short-read sequencing using the Illumina MiSeq platform was utilised. The Illumina platform uses a sequencing by synthesis technology, where a DNA polymerase or ligase is used to parallelly extend numerous strands of DNA and fluorescently labelled reversible terminators are used to detect and image each dNTP as it is added (Figure 20)[200,201].

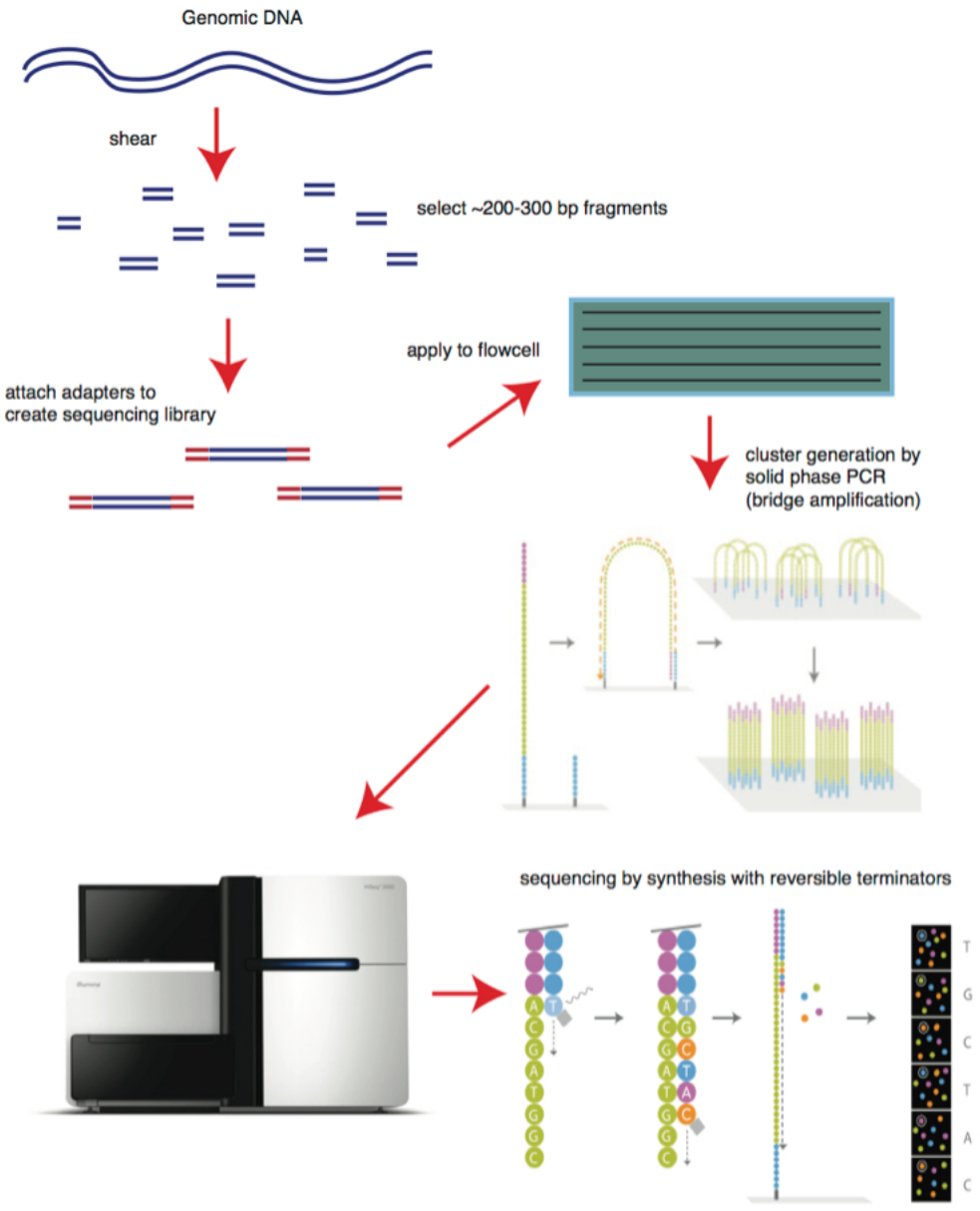


Figure 20 Illumina uses sequencing by synthesis technology. DNA molecules with attached adaptors are amplified on a flow cell. Following each round of synthesis fluorescence is recorded to determine which nucleotide is incorporated (https://frontlinegenomics.com/dna-sequencing-how-to-choose-the-right-technology/).

For the experiments described in Chapters 4-6, a previously published NGS method was used to assess the TCR repertoire [202].

2.7.1 Library preparation and sequencing

RNA templates from sorted T-cell subpopulations were used to generate libraries as previously described with minor amendments, previously instigated by Dr Athina Soragia Gkazi (Figure 21)[202]. More specifically, ProNex purification beads (Promega) with the corresponding ratios for library purification were used throughout all the steps, instead the recommended by the protocol AmPure purification beads (Beckman Coulter).

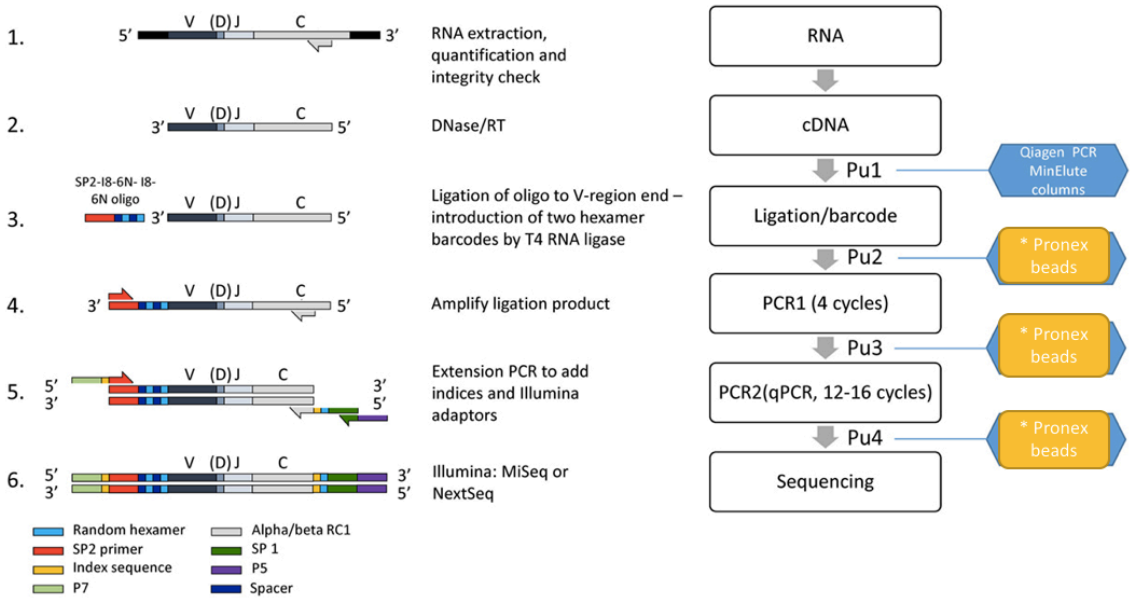


Figure 21 Outline of the TCR library preparation protocol (adapted by[202])

In summary, RNA samples were first DNase treated, to remove any residual DNA, and then reverse transcribed into cDNA using primers that are located close to the 5' end of the TCR constant region. An oligonucleotide containing the Illumina SP2 primer, and a unique molecular identifier (UMI) was ligated to the 3' end of the cDNA. The ligated products were

amplified through a PCR, using nested primers in the alpha and beta constant region with the Illumina SP2 primer. The products were amplified further to incorporate the Illumina adaptor and SP1 primer, as well as indices for multiplexing[202,203]. The final, purified products were used to generate pooled libraries that were sequenced using the Illumina MiSeq system and the MiSeq Reagent Kit V2 (500-cycles).

2.7.2 Data analysis

The generated FASTQ files were demultiplexed and error-corrected (through unique molecular identifiers -UMIs) by using Decombinator, a freely available suite of Python scripts (https://github.com/innate2adaptive/Decombinator, v4.0.3). This pipeline consists of four scripts that are applied consecutively. Initially Demultiplexor, using the random barcode sequences introduced through the library preparation protocol, separated the raw FASTQ files into individual sample sequencing files. Then Decombinator searched the sample specific files rearranged TCR sequences and assigned them a five-part index (called DCR), representing the rearrangement. The DCR output files were error-corrected through Collapsinator using the barcode sequences as reference for amplification and sequencing mistakes. The CDR3translator script translated the error-corrected data into CDR3 amino acid and nucleotide sequences, V and J genes, rearrangement abundance, productivity etc (Figure 22).

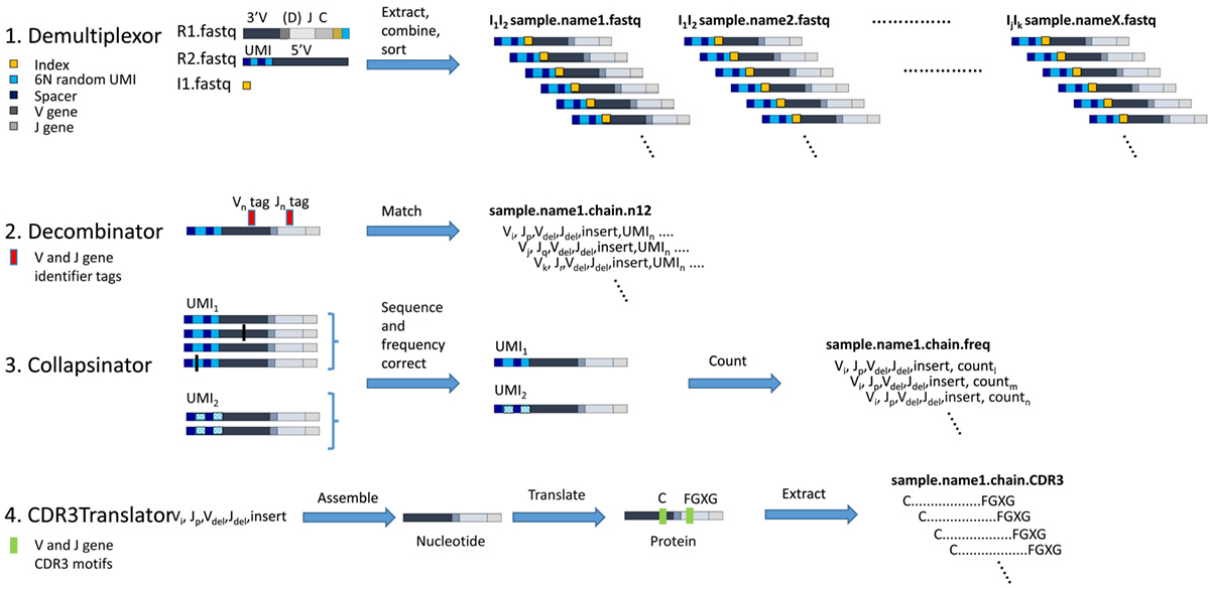


Figure 22 Outline of the Demultiplexor computational pipeline for TCR analysis: This computational pipeline takes raw FASTQ files and with the help of the UMI sequences demultiplexes the data into files containing reads deriving from only one indexed sample. Decombinator takes the Decombinator output and identifies the relevant V and J gene information to assign them a five-part identifier. Collapsinator corrects the data for frequency and sequencing errors, so that CDR3 translator can extract the CDR3 amino acid sequences using the five-part identifier [202].

The produced CDR3 sequences were subsequently used for TCR repertoire analysis. An in-house subsampling R script (R, version 3.6.3) written by Dr Ben Margetts was used to compute the minimum representative subsampling depth per sample to make the subsequent sample analysis more comparable[203]. The Gini coefficient, an economic tool originally used for measuring wealth distribution, was utilised in this thesis to measure intrasample inequality. The Gini coefficient scale ranges between 0 and 1, where 0 accounts for a completely equal sample with identical CDR3 clone frequency and 1 accounts for a completely unequal sample showing oligoclonality[203]. Excel (Microsoft, version 16.65) was utilised to determine any potential clonotype sharing by using the "MATCH" function. As previously described, the VDJdb database (version "SearchTable-2021-05-20 23_12_40.221") and an in-house R script were used to investigate CDR3 "specificity" towards pathogens[14,203,204]. CDR3 sequence similarity was investigated by calculating Hamming distances and studying the presence of closely related TCR clusters, using scripts provided by Dr Katrine Schou Sandgaard[14,204]. Hamming distance is a metric that helps compute the number of amino acid (AA) changes

between two CDR3 sequences of the same length, where a Hamming distance of 1 reflects a single AA change[14,204,205]. Closely related CDR3 sequences were visualised by computing similarity using a shared triplet metric. Clusters of related CDR3 sequences are represented as nodes in a network[206]. The open access MAIT Match Server database (DTU Bioinformatics) was used to compare the identified sequences with known MAIT cells sequences. Statistical analysis was performed using Prism (GraphPad, version 9.0.0). Finally, plots were generated using an R package (ggplot2, version 3.3.5) and the Prism software (GraphPad, version 9.0.0).

CHAPTER 3: VIROLOGICAL CHARACTERISTICS OF THE HIV-1 RESERVOIR

3.1 Background and Objectives

Thus far, most paediatric studies have investigated the effects of ART on the levels of HIV-1 CA-DNA, showing that it can in some cases be detectable even after years on therapy[4,10,207–209]. It has been reported that the timing of ART initiation can have a significant impact on the reservoir size and HIV-1 CA-DNA, with reduced levels being associated with younger age of treatment onset[4,90]. Lately, another complementary biomarker, CA-RNA, has gained ground in the study of the HIV-1 reservoir, with similarly reduced levels being reported in HIV-1 infected children on early ART[210,211]. One of the biggest challenges we face is the ability to detect and quantify these minute amounts of HIV-1 nucleic acids (DNA and RNA) in limited biological sample volumes from paediatric patients on suppressive ART. So, the development and use of highly sensitive, economical, and reproducible assays is necessary.

The aim of this chapter was to develop and optimise a method to investigate the levels of HIV-1 CA-RNA and subsequently determine the presence of ongoing viral transcription in 40 HIV-1 perinatally infected well-supressed children (viral load <50 copies/ml) that initiated ART early on in life (<2 years of age), taking part in the CARMA study. The CARMA study included children from 7 European centres and is part of the EPIICAL (Early Treated Perinatally HIV Infected Individuals: Improving Children's Actual Life) consortium that aims to select new and promising therapeutic strategies by identifying ideal candidates by considering their virological and immunological profiles[4,212]. By combining the information provided by different complementary markers, such as total HIV-1 CA-DNA, we can obtain a more complete picture of the HIV reservoir and the effects of early ART[4]. Better understanding of the viral and immunological dynamics present within these patients could assist in predicting their response not only to their current treatment regiments but also to novel strategies in future clinical trials.

Due to the limited sample volume available per patient two qRT-PCR assay versions were developed, targeting total and unspliced HIV-1 CA-RNA respectively. Unspliced HIV-1 CA-RNA

was selected since it more abundant and easily detectable in chronically suppressed individuals[62]. Studies have shown that in ART-treated patients with undetectable viral load, unspliced HIV-1 CA-RNA can act as a predictive marker for treatment outcomes, with higher levels being associated with potential treatment failure[213]. Other species of HIV-1 CA-RNA, like multiply-spliced, decay much faster upon ART initiation thus making it difficult to measure unless there is a stimulation step ex vivo and excluding it as target for our study at this point[75]. For that reason, levels of total HIV-1 CA-RNA were investigated to determine the potential presence of other HIV-1 CA-RNA species additionally to unspliced.

3.2 Introduction

To this day, HIV proviral integration in immune system cells and the establishment of a reservoir during the early stages of infection remains one of the major obstacles in achieving an HIV cure[66,214,215]. This complex reservoir, in terms of tissue compartment and cell type involvement (some of which are presented in Figure 23), remains dormant (does not actively produce new HIV virions) in the presence of ART, but can potentially lead to a productive infection due to viral reactivation in the case of treatment failure or interruption[67,216].

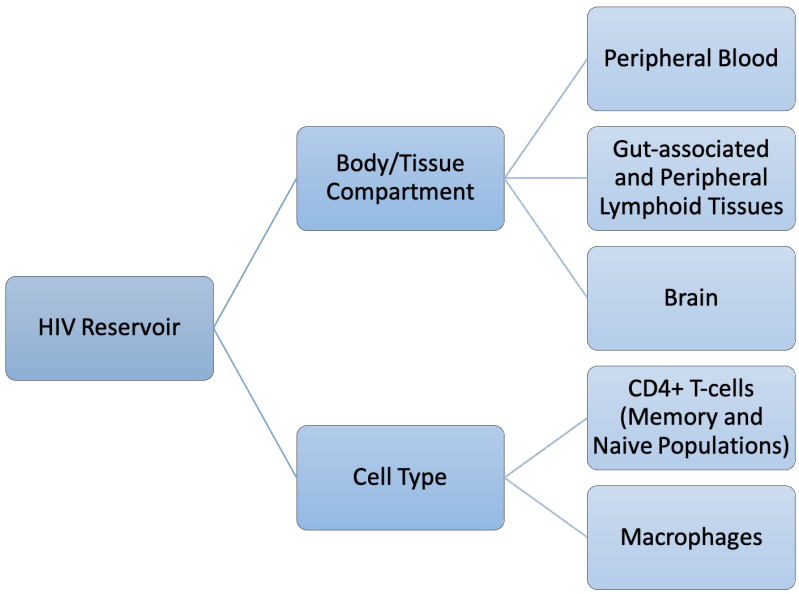


Figure 23 Aspects of the persistent HIV reservoir. The involvement of numerous tissue and body compartments as well as various cell types, contributes to the complexity of the reservoir study (adapted from [216]).

Understanding the dynamics of the viral reservoir, especially in children on early ART with undetectable plasma viremia, is vital and can lead to better informed treatment strategies and patient management in the future. Since in most individuals on continuous ART plasma viremia remains undetectable, additional virological markers, such as cell-associated HIV-1 DNA (CA-DNA) and cell associated HIV-1 RNA (CA-RNA), are required to monitor, and predict therapy responses[62]. The development of sensitive and efficient methods to measure these HIV-1 markers is required.

3.3 Material and Methods

The methods used in this chapter are described in detail in Chapter 2. As part of the CARMA study, samples from 40 HIV-1 perinatally infected children, that started therapy before two years of age and had been virally suppressed (viral load (VL) <50 copies/ml) for at least five years, were tested[4]. For this study single viral load blips and/or annual viral spikes (plasma VL rise from 50 to 399 copies/ml and 400-999 copies/ml respectively, returning to <50 copies/ml on subsequent follow-up sampling) were permitted [4]. The virological characteristics of all participants are presented bellow (Table 11).

Table 11 Virological characteristics of the CARMA cohort (adapted from [4]). Unless otherwise stated, continuous data are presented as median (IQR).

Virological Characteristics	**All (n=40)**
VL at diagnosis, log copies/ml	5.60 (4.98-5.93)
VL at ART start, log copies/ml	5.28 (4.07-5.70)
Time to viral suppression, months	4.69 (2.52-6.26)
VL blip, No. (%)	11 (27.5)
Time to blip, years	3.7 (2.35-7.26)
Viral spike, No. (%)	5 (12.5)
Time to spike, years	1.53 (0.86-3.53)

*Abbreviations: IQR, Interquartile Range; VL, Viral Load; ART, Antiretroviral Therapy

In summary, total nucleic acids from 10µl of extract were treated by adding 0.5µl DNase I (NEB) and 1µl DNase I buffer (NEB) by an initial 10-minute incubation step at 37°C, followed by a 10-minute inactivation step at 72°C. Total or unspliced HIV-1 CA-RNA was detected using 10µl of DNase I treated nucleic acids and a one-step RT qPCR using a TaqMan Fast Virus master mix (ThermoFisher), in a final reaction volume of 20µl. Primers and probes were designed for the HIV LTR and HIV pol gene (as previously described in [89]), for the total and unspliced HIV-1 CA-RNA assays respectively. IPO8 (Bio-Rad) and TBP1 (Bio-Rad) were used as reference genes, to support confidence in the qPCR results, since they should be stably expressed in PBMCs[197,198]. In both versions of the assay, the reactions were run in duplicate

and with the following conditions: 50°C for 5 minutes, 95°C for 20 seconds followed by 45 cycles of 94°C for 15 seconds and 60°C for 90 seconds, for amplification. The RT qPCRs were performed on a CFX96 Touch Deep Well Real-Time PCR System (Bio-Rad). Serial dilutions of an RNA transcript with known copy numbers were used to generate a standard curve. Assay results were analysed using the CFX Maestro software (Bio-Rad). Generated mean cycle threshold (Ct) values were plotted against the standard curve to obtain copy numbers per reaction. HIV-1 CA-RNA copy numbers were then normalised against the two reference genes and copy numbers per 10^6 PBMCs were calculated.

The limit of detection (LOD) and limit of quantitation (LOQ), for both CA-RNA qPCR methods, were determined by amplifying serial dilutions (containing 0.3-500 copies) of the assay standard in replicates. The LOD refers to the lowest concentration of the analyte that can be detected in our sample, but not necessarily quantified[217]. In contrast, the LOQ refers to the lowest concentration of the analyte that can be precisely and accurately quantified in our sample[217]. The efficiency was evaluated through linear regression analysis, by considering the analytical noise (values of 3.3 and 10, for detection and quantitation respectively), the slope of the linear logarithmic phase of the reaction (values between -3.0 and -3.6 are expected) and the y-intercept (theoretical LOD of the reaction). All calculations were performed using previously described methods and results were reported as copies/reaction[217]. The equations used are shown below (the STEYX function returns the standard deviation of the predicted y-intercept).

(A) $LOD = (3.3 \times STEYX) \div slope$

(B) $LOQ = (10 \times STEYX) \div slope$

Statistical analysis was performed using Excel (Microsoft, version 16.65) packages and the Prism (GraphPad, version 9.0.0) software. HIV-1 CA-RNA data were separated in two groups depending on the presence or absence of CA-RNA. Participants belonging to the detectable HIV-1 CA-RNA group presented with >0 copies/10^6 PBMCs (with cycle threshold [Ct] values <40) and participants belonging to the undetectable HIV-1 CA-RNA group presented with 0 copies/10^6 PBMCs (Ct values ≥40). For continuous variables median and interquartile range

(IQR) were calculated and reported. For the statistical analysis, using the Prism software (GraphPad, version 9.0.0), the data were presumed non-parametric, and the Mann-Whitney test was performed. For categorical variables counts and percentages were assessed and reported, and the Fisher's exact test was performed. P-values of ≤ 0.05 were considered as statistically significant. To assess any correlation between continuous variables the Spearman test was performed.

3.4 Results

During the optimisation stage, the LOD was calculated at 11 copies per million PBMCs for both assays. The LOQ was reported at 34 and 33 copies per million PBMCs, for the total and unspliced HIV-1 CA-RNA RT qPCR assays respectively (Table 12).

Table 12 Calculated LOD and LOQ values for the HIV-1 CA-RNA RT qPCR assays.

Assay	Slope	STEYX	LOD (copies/10^6 PBMCs)	LOQ (copies/10^6 PBMCs)
Total HIV-1 CA-RNA RT qPCR	-0.03605	0.1224	11	34
Unspliced-1 HIV CA-RNA RT qPCR	-0.03532	0.1162	11	33

Forty HIV-1 perinatally infected children that initiated ART early on in life and were part of the CARMA cohort, were selected and investigated for the presence of residual HIV-1 CA-RNA levels[4]. HIV-1 CA-RNA, total and/or unspliced was detectable in 65% (26/40) of the patients. More specifically, total HIV-1 CA-RNA was detectable in 57.5% (23/40) of the patients, with levels ranging from 1.2-5789.5 copies/10^6 PBMCs. Unspliced HIV-1 CA-RNA was detectable in 47.5% (19/40) of the CARMA patients, with levels ranging from 0.04-274.5 copies/10^6 PBMCs (Figure 24).

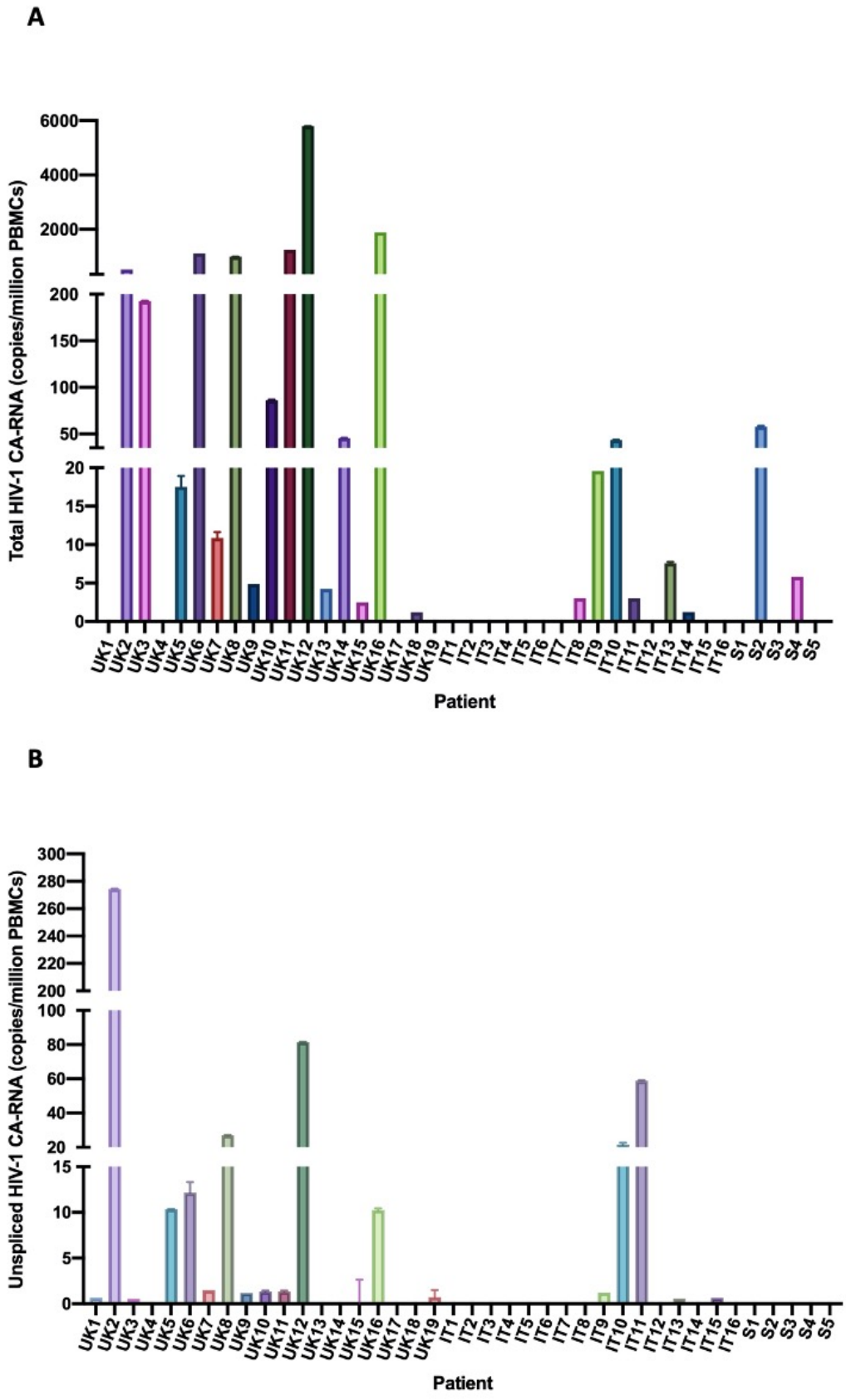


Figure 24 Levels of A) total and B) unspliced HIV-1 CA-RNA detected by the one-step RT-qPCR in the 40 CARMA patients. HIV-1 CA-RNA was detectable in 65% of the CARMA study participants.

A summary of the results for both qRT-PCR assays is presented in Table 13, with the second column reporting the numbers and percentages of patients for each category and the third column reporting the median and IQR detectable levels of HIV-1 CA-RNA copies/10^6 PBMCs assessed for both assays. In the 23 children with detectable total HIV-1 CA-RNA, the median was 19.57 (4.23-522.5) copies/10^6 PBMCs. In the 19 children with detectable unspliced HIV-1 CA-RNA, the median was 1.48 (0.67-21.62) copies/10^6 PBMCs. More analytically as seen in the figures above, out of the 23 children with detectable total HIV-1 CA-RNA only 16 had detectable unspliced HIV-1 CA-RNA simultaneously. Three participants (UK1, UK19 and IT15) that presented with low level (<0.5 copies/10^6 PBMCs) unspliced HIV-1 CA-RNA did not have detectable levels of total HIV-1 RNA.

Table 13 Summary of HIV-1 CA-RNA results, between the two assays detecting total or unspliced HIV CA-RNA. Copies/10^6 PBMCs are presented as median (IQR).

	All (n=40)	**Copies/10^6 PBMCs (IQR)**
Total HIV-1 CA-RNA		
Detectable, No. (%)	23 (57.5)	19.57 (4.23-522.50)
Undetectable, No. (%)	17 (42.5)	-
Unspliced HIV-1 CA-RNA		
Detectable, No. (%)	19 (47.5)	1.48 (0.67-21.62)
Undetectable, No. (%)	21 (52.5)	-

As expected, the quantified levels of total HIV CA-RNA were higher than the levels of unspliced HIV-1 CA-RNA for most patients, indicating the presence of more CA-RNA species, such as multiply spliced HIV-1 CA-RNA (Figure 25). Only four participants, UK1, UK19, IT11 and IT15 presented slightly higher levels of unspliced HIV-1 CA-RNA (Figure 24).

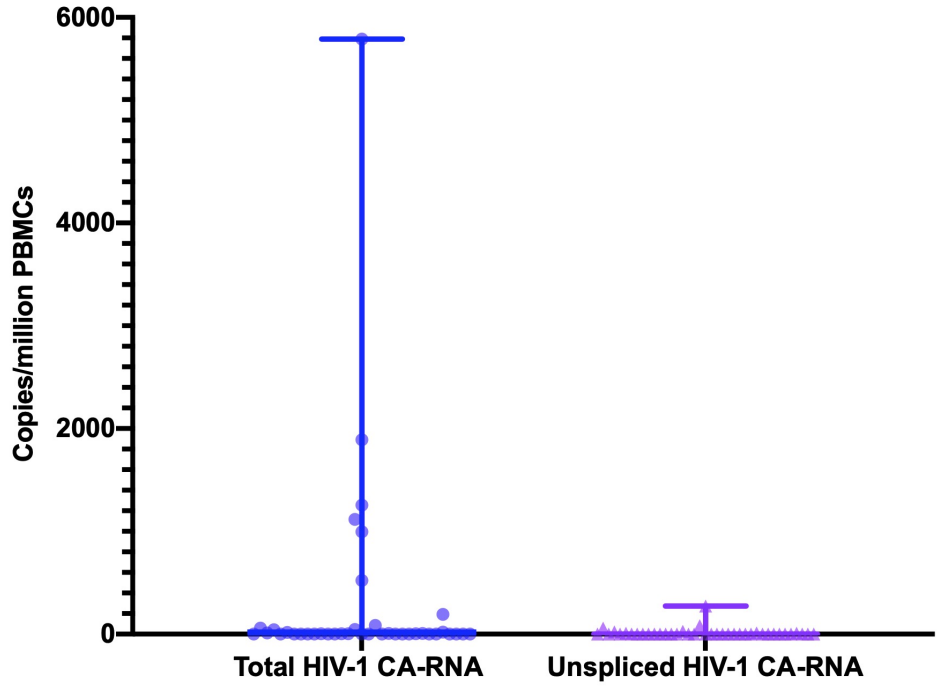


Figure 25 Quantified levels of different HIV-1 CA-RNA species, total and unspliced, in the CARMA patients.

The strongest positive correlation was shown when comparing the levels of total and unspliced HIV-1 CA-RNA (ρ=0.685, p<.001). Total HIV-1 CA-DNA levels, previously reported in 2020 by Foster et al., positively correlated with both total HIV-1 CA-RNA (ρ=0.481, p=0.002) and unspliced HIV-1 CA-RNA (ρ=0.456, p=0.003) levels. Additionally, higher levels of unspliced HIV-1 CA-RNA correlated with older age (ρ=0.462, p=0.002) and higher CD8% (ρ=0.390, p=0.012) at ART initiation (Figure 26).

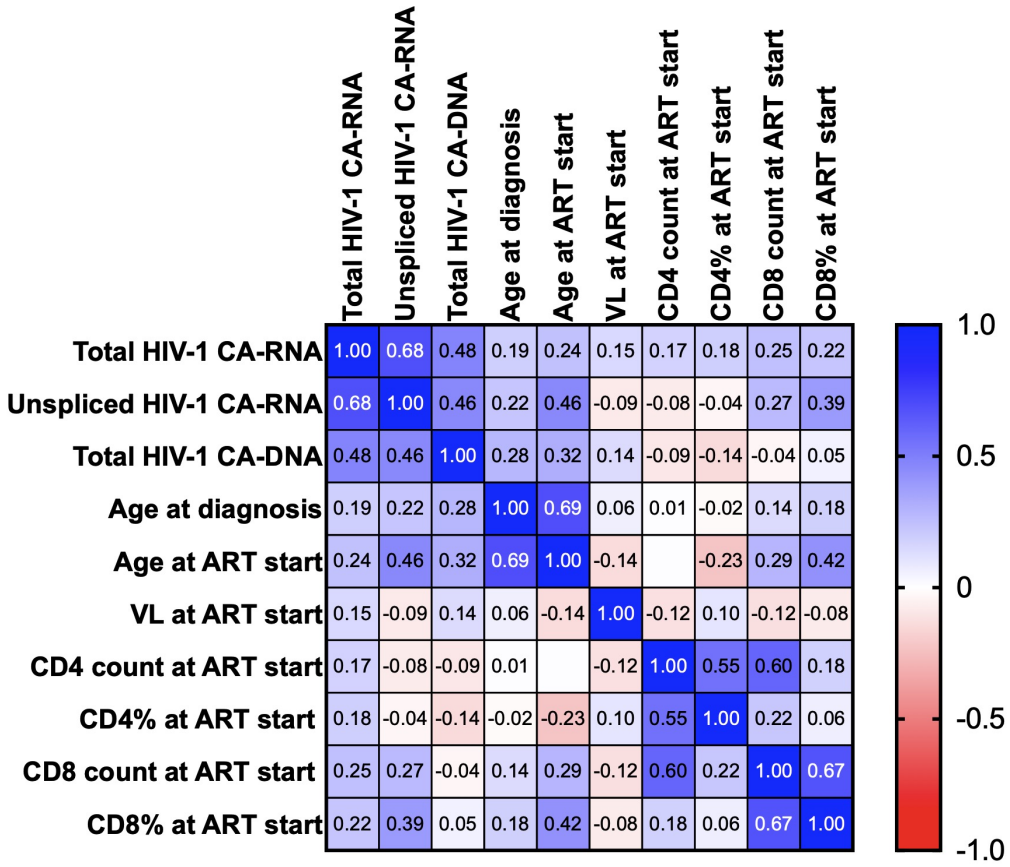


Figure 26 Correlations between different variables measured at time of analysis and ART start. The heatmap indicates the strength of positive (blue) and negative (red) correlations present. The calculated Spearman's correlation coefficient (ρ) is also depicted in each corresponding square.

Following the separation of the CARMA patients into two groups depending on the quantified levels of HIV-1 CA-RNA (undetectable and detectable), analysis was performed to demine any possible associations with parameters such as sex, age at HIV diagnosis, age, viral load, CD4+ and CD8+ T-cell levels (total count and %) at the ART initiation point. All results and the predicted p-values are summarised below (Table 14 & Table 15).

As mentioned before, total HIV-1 CA-RNA levels only presented a strong association with total HIV-1 CA-DNA (p=0.0249, Table 14). Generally, individuals that presented with undetectable levels of total HIV-1 CA-RNA had lower levels of total HIV-1 CA-DNA, with a median of 41.1 (0.05-61.35) copies/10^6 PBMCs.

Table 14 Summary of virological and immunological parameters compared by total HIV-1 CA-RNA levels. Unless otherwise stated, all continuous data are presented as median (IQR).

	All (n=40)	Undetectable, 0 copies/10^6 PBMCs (n=17)	Detectable, <0 copies/10^6 PBMCs (n=23)	p-value
Total HIV-1 CA-DNA, copies/10^6 PBMCs	48.35 (5.95-112.7)	41.10 (0.05-61.35)	67.70 (17.90-197.0)	0.0249
Sex, No. (%)				0.3326
Female	27.0 (67.5)	13.0 (76.5)	14.0 (60.9)	
Male	13.0 (32.5)	4.0 (23.5)	9.0 (39.1)	
Age at HIV diagnosis, months	4.17 (2.17-6.43)	4.03 (1.05-6.62)	4.2 (3.08-6.49)	0.5559
Age at ART start, months	4.28 (0.24-6.90)	3.61 (0.1-6.29)	4.43 (2.66-8.56)	0.3804
VL at ART start, copies/ml	145535 (6048-500000)	46000 (728-360000)	204000 (43756-500000)	0.1582
CD4+ count at ART start, cells/µl	853.5 (23.75-2977)	637.0 (0.0-3080)	1000 (380-2943)	0.3631
CD4 % at ART start	25.5 (15.25-40.50)	18.0 (0.0-36.5)	29.0 (19.0-43.0)	0.1537
CD8+ count at ART start, cells/µl	623.5 (0.0-1332)	0.0 (0.0-1499)	670 (0.0-1380)	0.4036
CD8 % at ART start	27.0 (0-37.75)	21.0 (0.0-34.0)	27.0 (17.0-40.0)	0.2745

*Abbreviations: IQR, Interquartile Range; ART, Antiretroviral Therapy; VL, Viral Load

As with total HIV-1 CA-RNA, unspliced CA-RNA levels presented an association with total HIV-1 CA-DNA (p=0.0077). Undetectable unspliced HIV-1 CA-RNA levels associated with reduced total HIV-1 CA-DNA, with a median of 17.9 (0.55-55.1) copies/10^6 PBMCs. Furthermore, an association was shown with younger age (p=0.0166, median of 2.66 (0.0-6.29) months) and lower CD8 % (p=0.0289, median of 21.0% (0.0-30.0)) at the ART initiation point. A summary of all the results is presented below (Table 15).

Table 15 Summary of virological and immunological parameters compared by unspliced HIV-1 CA-RNA levels. Unless otherwise stated, all continuous data are presented as median (IQR).

	All (n=40)	Undetectable, 0 copies/10^6 PBMCs (n=21)	Detectable, <0 copies/10^6 PBMCs (n=19)	p-value
Total HIV-1 CA-DNA, copies/10^6 PBMCs	48.35 (5.95-112.7)	17.9 (0.55-55.1)	87.9 (33.1-225.8)	0.0077
Sex, No. (%)				0.9999
Female	27.0 (67.5)	14.0 (66.7)	13.0 (68.4)	
Male	13.0 (32.5)	7.0 (33.3)	6.0 (31.6)	
Age at HIV diagnosis, months	4.17 (2.17-6.43)	4.13 (1.36-6.62)	4.2 (3.15-6.49)	0.5688
Age at ART start, months	4.28 (0.24-6.90)	2.66 (0-6.29)	4.59 (3.87-12.92)	0.0166
VL at ART start, copies/ml	145535 (6048-500000)	204000 (6930-916500)	100210 (5600-490000)	0.3573
CD4+ count at ART start, cells/ µl	853.5 (23.75-2977)	1621.0 (47.5-4188)	810.0 (0.0-1390)	0.2935
CD4 % at ART start	25.5 (15.25-40.50)	22.0 (6.0-44.5)	27.0 (17.0-35.0)	0.8563
CD8+ count at ART start, cells/ µl	623.5 (0.0-1332)	251.0 (0.0-946.5)	720.0 (0.0-1970)	0.2285
CD8 % at ART start	27.0 (0.0-37.75)	21.0 (0.0-30.0)	32.0 (18.0-40.0)	0.0289

*Abbreviations: IQR, Interquartile Range; ART, Antiretroviral Therapy; VL, Viral Load

3.5 Discussion

In 2010, WHO published guidelines that recommended starting ART in all infected children younger than 24 months of age, immediately regardless of their virological and immunological status. Cases like those of the "Mississippi baby", the "French teenager" and the "South African child" have demonstrated that early ART can lead to long-term post-treatment virological control[7,8,10,11,218]. Even though, some of these patients might eventually relapse, there is no denying that early ART initiation has created a unique cohort of children that will reach adolescence and adulthood with reportedly reduced HIV reservoirs[4,9,207,219,220]. Since these children must be on ART for the entirety of their lives, they could be ideal candidates for new treatment approaches that could potentially ease the burden created by drug toxicity and reduce the dangers associate with developing drug resistance mutations.

Here, the presence of HIV-1 CA-RNA and the potential of ongoing viral transcription in well-suppressed, perinatally infected individuals that started ART before their second year of age, were investigated in the context of the CARMA study[4]. It was hypothesised that early ART initiation leads to low or undetectable levels of HIV-1 CA-RNA. CA-RNA along with other virological and immunological markers, could be used to identify individuals that might sustain better viral control in future clinical trials. To this day, the use of HIV-1 CA-RNA as a biomarker in clinical settings is surrounded by controversy. Its presence might not always signify productive HIV-1 infection but also transcription from a stable reservoir[75,210,214,221]. The detected transcripts might be defective at the sequence level or get destroyed by post-transcriptional modifications, thus not leading to the formation of new virions. To analyse whether viral transcription is deriving from replication competent proviruses, further sequencing and viral outgrowth studies are necessary[62,75]. Nonetheless, HIV-1 CA-RNA has been previously shown to predict treatment failure and correlates with ART adherence[213,222].

Two of the main obstacles in paediatric HIV-1 studies, like the one presented here, are the small sample volumes (due to regulations and ethical considerations) available for the investigation of all the necessary virological and immunological markers, and the reduced number of participants. These two factors can hinder our understanding of the reservoir dynamics since they make it difficult to see potential trends in patient characteristics. Due to

the advantages of being sensitive, quicker, reproducible, cheaper, and not requiring large sample volumes qPCR assays are widely used in HIV-1 studies. Thus far, several versions have been published, targeting different types of HIV-1 CA-DNA and CA-RNA[4,89,93,223]. Even though both biomarkers can overestimate the size of the reservoir and can't offer a distinction between defective and replication-competent viruses on their own, their presence is a potential indicator of ongoing viral replication and protein production, even during long-term suppressive ART. For the purposes of the CARMA study, two one-step qRT-PCR assays were developed to investigate the presence of total and unspliced HIV-1 CA-RNA and their association with age at ART start.

Of the 40 perinatally infected and well-supressed children, 35% had no detectable levels of total and/or unspliced HIV-1 CA-RNA. More analytically, total HIV-1 CA-RNA was undetectable in 42.5% and unspliced HIV-1 CA-RNA was undetectable in 52.5% of the CARMA patients. As expected, lower levels of total and unspliced HIV-1 CA-RNA correlated with lower levels of HIV-1 CA-DNA, supporting the evidence of a reduced viral reservoir size in this paediatric cohort. Surprisingly, age at the point of ART initiation only showed an association with unspliced HIV-1 CA-RNA, with younger age (median of 4.13 months) correlating with undetectable levels (p=0.0166). Undetectable levels of unspliced HIV-1 CA-RNA also showed an association with lower CD8% (median of 21.0%) at ART start (p=0.0289). No other patient characteristics at ART start (e.g., VL, CD4+ T-cell count) seem to affect the levels of HIV-1 CA-RNA. Sequencing and viral outgrowth studies are currently ongoing to determine whether the detected viral transcription derives from defective or replication competent viruses.

One of the main limitations of our study was that the number of enrolled patients was relatively small and that the reservoir dynamics were only investigated at one timepoint, thus making it impossible to determine the actual effects of detectable HIV-1 RNA levels in the scope of early initiated and continuous ART. Notably, during the course of the CARMA study, it was reported that participant UK12 presented with suboptimal ART adherence for a brief time period and had a viral load blip (>50 copies/ml). Despite being requested, it was impossible at this point to obtain additional clinical data for this participant to determine factors that might have led and contributed to the viral blip, as well as additional clinical features during this period (like CD4+ T-cell counts, time to suppression etc). However, at the

point of analysis, interestingly this participant (UK12) presented with the highest total and second highest HIV-1 CA-RNA levels respectively. This fact could potentially strengthen the case that paediatric HIV-1 patients should be monitored longitudinally as changes in their CA-RNA levels could be a predictive marker of treatment failure. In the future, providing participants already enrolled to the CARMA study with the opportunity to give longitudinal samples to track potential changes in biomarker levels, could be both beneficial and informative. Following all the publications that are being generated from the different study sites the CARMA study cohort could become one of the best virologically and immunologically characterised HIV-1+ paediatric populations. Finally, continuous efforts for additional and bigger paediatric cohorts to join studies like CARMA would be highly advantageous for HIV-1 reservoir investigations.

To date, HIV-1 CA-RNA has not been developed for clinical use. Even though it has received a lot of attention in recent years as a viral reservoir marker and several quantification methods have been published, there is still ongoing debate on how to accurately measure its levels and what how it is associated with existing clinical endpoints[62,75]. Recently, a study investigating HIV-1 CA-RNA levels demonstrated its predictive value over plasma viral load for virological and immunological responses in early treated patients, association with the severity and time to viral rebound following ART discontinuation, as well as the disease progression[224]. However, limitations arising from the method set up, such as processing blood samples, isolating RNA and needing an internal control to normalise assay input, might make its adoption in clinical routine in the near future challenging.

3.5.1 Conclusions

Despite all the limitations present in this study, concordance was observed with previously published reports of HIV-1 CA-RNA presence in ART well-suppressed children[210,225]. This highlights the value of this biomarker for potential use in clinical settings for patient monitoring. In the future, large-scale global paediatric and adult research studies will be necessary to better understand and characterise the dynamics of viral reservoir of this unique cohort early treated HIV-1 perinatally infected individuals, giving them access to novel treatment strategies and ultimately improving their quality of life.

CHAPTER 4: IMMUNOLOGICAL CHARACTERISTICS OF HIV-1 PERINATALLY INFECTED INDIVIDUALS – IMPACT OF EARLY ART INITIATION ON T-CELL RECEPTOR (TCR) REPERTOIRE DIVERSITY

4.1 Background and Objectives

Early ART initiation, within days or months of the HIV-1 infection, has been associated with reduced T-cell activation and preservation of T-cell numbers, with levels like those observed in healthy individuals[226,227]. However, the impact of early ART on the TCR repertoire in children, has yet to be extensively explored and reported.

Previous work in our group has demonstrated that HIV-1 infected individuals that present good viral control show higher thymic output, reduced numbers of clonal expansions in CD4+ and CD8+ T-cells and increased TCR repertoire diversity[14]. In this chapter, it is hypothesised that early ART initiation can have a positive impact on the developing immune system of HIV-1 perinatally infected children providing them with greater TCR diversity and protection against infections.

4.2 Introduction

To date there have been various studies investigating the effect of ART, short-term or long-term, on the immune system and TCR repertoire of adults living with HIV-1. Even though ART leads to decreased HIV specific immune responses and changes in the abundance of certain TCR sequences, their TCR repertoire remains depleted compared to healthy individuals potentially due to their low thymic output[156,228,229]. However, not many reports on the effect of continuous long-term ART in the TCR repertoire of early treated paediatric patients can be found in the literature. Can ART initiation at a very young age, during immune system development, have more of a beneficial effect on the TCR repertoire of HIV+ children?

This chapter aims to explore the impact of early ART initiation on the thymic output as well as use next generation sequencing techniques to examine the CD4+ and CD8+ TCR repertoire in the patients enrolled in the UK arm of the CARMA study. Both alpha and beta TCRs were investigated for their diversity, sharing between patients, antigen "specificity" and similarity to invariant MAIT cell sequences.

Overall, the CARMA study aimed to investigate the impact of early ART initiation on virological and immunological factors in a paediatric HIV cohort of 40 well-suppressed patients and was designed as part of the EPIICAL consortium collaboration[4]. As mentioned previously in this thesis, the EPIICAL consortium was founded with the belief that perinatally infected individuals, treated early with ART comprise an ideal population that will best inform new therapeutic strategies and make ART-free remission possible in the future[230]. This will be made possible by conducting studies that will allow for the development of predictive immunological and virological models. Thus far, studies such as VISCONTI, CHER, and the case of the "Mississippi baby" have yielded promising results, by demonstrating that early ART at a time when thymic function is at its highest, can lead to lower mortality rates, greater immune reconstitution, and even long-term viral control in some individuals in cases of treatment interruption[8,9,13,204,218,231,232]. Most of these studies mentioned above have focused on parameters that determine the HIV-1 reservoir size like cell-associated DNA and RNA, plasma HIV-1 RNA levels and HIV-1 immune responses. The CARMA study was designed to support and compliment previous reports as well as provide new insights into the dynamics of paediatric HIV-1+ populations.

4.3 Materials and Methods

The methods used in this chapter are described in detail in Chapter 2. Samples from the 19 UK based individuals (13 females and 6 males) enrolled in the CARMA study were selected and further investigated for their TCR repertoires[4]. The age and immunological characteristics of the participants at the point of ART initiation and analysis (when study blood samples were obtained) are presented below (Table 16). Normal CD4+ and CD8+ count ranges, subset size and CD4:CD8 ratios, depending on age, are presented below (Table 17). In HIV+ paediatric patients CD4+ and CD8+ counts and ranges depend on not only their age but also the immunosuppression severity[233,234]. CD4:CD8 ratios between 1.5-2.5 are considered normal, however HIV+ patients with altered immune functions, present with low or inverted ratios[235].

Table 16 Characteristics of the 19 UK individuals, part of the CARMA cohort (adapted from [4]). Unless otherwise stated, continuous data are presented as median (IQR).

All (n=19)	At ART initiation	At Analysis
Age	4.56 months (2.66-18.72)	13.96 years (10.61-16.06)
CD4+ count, cells/µl	810 (450-1390)	975 (734-1216)
CD4%	21.0 (16.0-35.0)	42.0 (37.0-50.0)
CD8+ count, cells/µl	960 (290-1970)	639 (450-731)
CD8%	30.0 (16.0-40.0)	26.0 (22.0-30.0)
CD4:CD8	0.61 (0.27-1.17)	1.67 (1.34-2.05)

*Abbreviations: IQR, Interquartile Range; ART, Antiretroviral Therapy

Table 17 Absolute CD4+ and CD8+ counts, relative subset sizes and CD4:CD8 rations, depending on age group, as measured in healthy paediatric populations (adapted from [236]). Unless otherwise stated, continuous data are presented as median, 10th to 90th percentiles.

T-cell Subset	Age Group					
	0-3 months	3-12 months	1-2 years	2-6 years	6-12 years	12-18 years

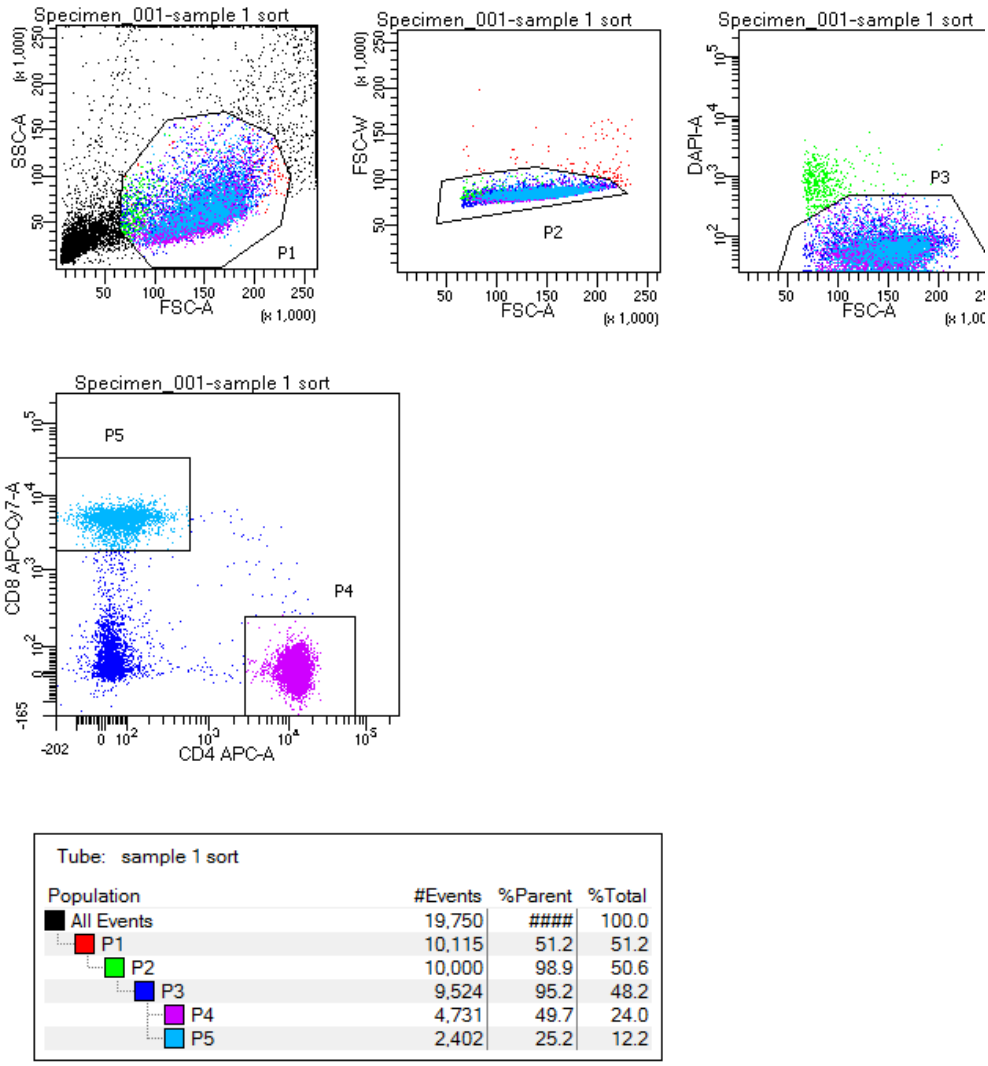
CD4+ count, cells/µl	3079 (2330-3617)	2492 (1523-3472)	1866 (1573-2949)	1448 (870-2144)	1030 (646-1515)	887 (610-1446)
CD4%	53.2 (42.8-65.7)	43.6 (34.9-53.1)	41.2 (35.0-51.9)	38.0 (31.1-47.4)	39.9 (31.7-47.0)	44.0 (32.6-51.5)
CD8+ count, cells/µl	1048 (712-1361)	976 (524-1583)	884 (656-1432)	804 (472-1107)	595 (365-945)	518 (282-749)
CD8%	18.4 (15.0-23.0)	16.2 (12.8-27.1)	19.3 (16.1-29.4)	21.0 (16.0-26.9)	24.0 (17.1-30.0)	23.0 (19.0-29.0)
CD4:CD8	2.90 (1.93-4.19)	2.64 (1.48-3.77)	2.00 (1.34-3.04)	1.77 (1.26-2.90)	1.67 (1.18-2.65)	2.05 (1.21-2.64)

In summary, PBMCs were isolated by density gradient centrifugation (section 2.2) and resuspended in the appropriate volume of pre-chilled FCS and pre-chilled freeze medium (FCS +10% DMSO) to produce 1ml aliquots containing 4-5 x10^6 PBMCs/ml. All aliquots were stored into a liquid nitrogen (LN_2) storage tank.

FACS was used to obtain total CD4+ and CD8+ T-cell populations from the stored PBMCs (as described in section 2.3). PBMCs were thawed using cRPMI, counted, divided in aliquots, and stained using the following surface antibodies and stains:

vi. APC anti-human CD4 (APC CD4) antibody (BioLegend)
vii. APC/Cyanine7 anti-human CD8 (APC/Cy7) antibody (BioLegend)
viii. LIVE/DEAD Fixable Aqua Dead Cell Stain (ThermoFisher Scientific)

All aliquots were then fixed using BD CellFix (BD Biosciences) and transferred to the Flow Cytometry Core Facility (UCL GOS Institute of Child Health). Using the appropriate gating strategy (Figure 27) all samples were sorted, by Dr Ayad Eddaoudi, into the required CD4+ and CD8+ populations, using the FACSAria III sorter (BD Biosciences).



Tube: sample 1 sort			
Population	#Events	%Parent	%Total
All Events	19,750	####	100.0
P1	10,115	51.2	51.2
P2	10,000	98.9	50.6
P3	9,524	95.2	48.2
P4	4,731	49.7	24.0
P5	2,402	25.2	12.2

Figure 27 Example of the gating strategy used for the CARMA study samples. Sorting was performed using FACSAria III (BD Biosciences) to obtain CD4+ and CD8+ T-cells.

Following CD4+ and CD8+ T-cell sorting all samples were divided into two aliquots and used for RNA and DNA extractions, using protocols optimised for fixed samples (section 2.4). The isolated nucleic acid concentration was determined using NanoDrop.

DNA isolates from all 19 individuals were transferred to the SIHMDS Molecular Department of Haematology (Camelia Botnar Laboratories, Great Ormond Street Hospital) and T-cell receptor excision circles (TRECs) were kindly measured by Dr Stuart Adams and Susanne Kricke, using a quantitative real-time PCR assay as previously described[237]. TRECs are circular fragments of DNA formed and removed during the V(D)J recombination process[237,238]. Briefly,

they are formed during the ligation of the recombination signal sequences leading to the deletion of the TCRD locus from the TCRA locus on the α-chain[237,238]Because TRECs are stable, do not duplicate during mitosis and do not degrade easily, they can be used to measure recent thymic emigrants and evaluate thymic function[238,239]. TRECs levels are significantly higher in healthy children and decrease with age[238]. The measurement of TRECs levels has been used in clinical settings to assist with primary and acquired immunodeficiency diagnosis, as well as monitor immune reconstitution following haematopoietic stem cell and thymus transplantations and gene therapy[238–240].

An NGS method targeting the CDR3 region was utilised to assess the TCR repertoire. More specifically, RNA isolates were used to generate libraries as previously described (section 2.7)[202]. Following DNase treatment cDNA was synthesised using primers located close to the 5' end of the TCR constant region. An oligonucleotide containing the Illumina SP2 primer, and a unique molecular identifier (UMI) were then ligated to the 3' end of the cDNA. The ligated products were amplified further through PCRs using primers in the alpha and beta constant region as well as Illumina adaptors, the SP1 primer and indexes for multiplexing. The final libraries were purified, pooled, and sequenced the Illumina MiSeq system and the MiSeq Reagent Kit V2 (500-cycles).

The generated FASTQ files were demultiplexed using the Decombinator suite of python scripts as previously described[202,203,241,242]. In-house R scripts (R, version 3.6.3) provided by Dr Ben Margetts and Dr Katrine Schou Sandgaard, were adapted, and used for subsampling and calculating the Gini coefficient[203,204]. The Excel (Microsoft, version 16.65) "MATCH" function was used to determine shared clonotypes between patients. The VDJdb database (records version "SearchTable-2021-05-20 23_12_40.221") and an in-house R script were used to determine CDR3 specificity[203,204,243]. CDR3 sequence similarity was investigated as previously described by calculating Hamming distances and studying the presence of closely related CDR3 sequence clusters[204,206]. The MAIT Match Server database (DTU Bioinformatics) was used to determine similarity between our generated CDR3 and MAIT sequences. R packages (ggplot2, version 3.3.5) and the Prism 9.0.0 (GraphPad) software were utilised to generate plots (section 2.7.2). Statistical analysis of the chapter results was completed using the Prism software (GraphPad, version 9.0.0).

4.4 Results

For the UK patients enrolled in the CARMA study, CD4+ counts showed an increase in most individuals at the point of analysis, compared to the baseline levels at the point of ART initiation. To determine any relationship between factors like age at ART and time of analysis (i.e., the time the CARMA study samples were taken), CD4+ and CD8+ counts, percentages, and ratios, as well as TRECs, the Spearman's rank correlation coefficient (ρ) was calculated (Figure 28). As shown below, there is an apparent strong negative correlation between age at the point of analysis and both CD4+ and CD8+ TRECs levels (ρ=-0.73, p=0.0004 and ρ=-0.68, p=0.0013 respectively). CARMA study participants that presented with lower CD4:CD8 ratios at the point of ART initiation, presented with higher levels of total HIV-1 CA-DNA at the point of analysis for the cohort (ρ=-0.59, p=0.008), indicating that delays in treatment can have potential negative effects for the patient's immune system due to the higher HIV-1 reservoir levels. Generally, as observed in the Figure below, younger age at ART initiation correlated with a better virological and immunological profile at the point of analysis, with lower levels of HIV-1 CA-DNA and CA-RNA as well as improved CD4:CD8 cell ratios present.

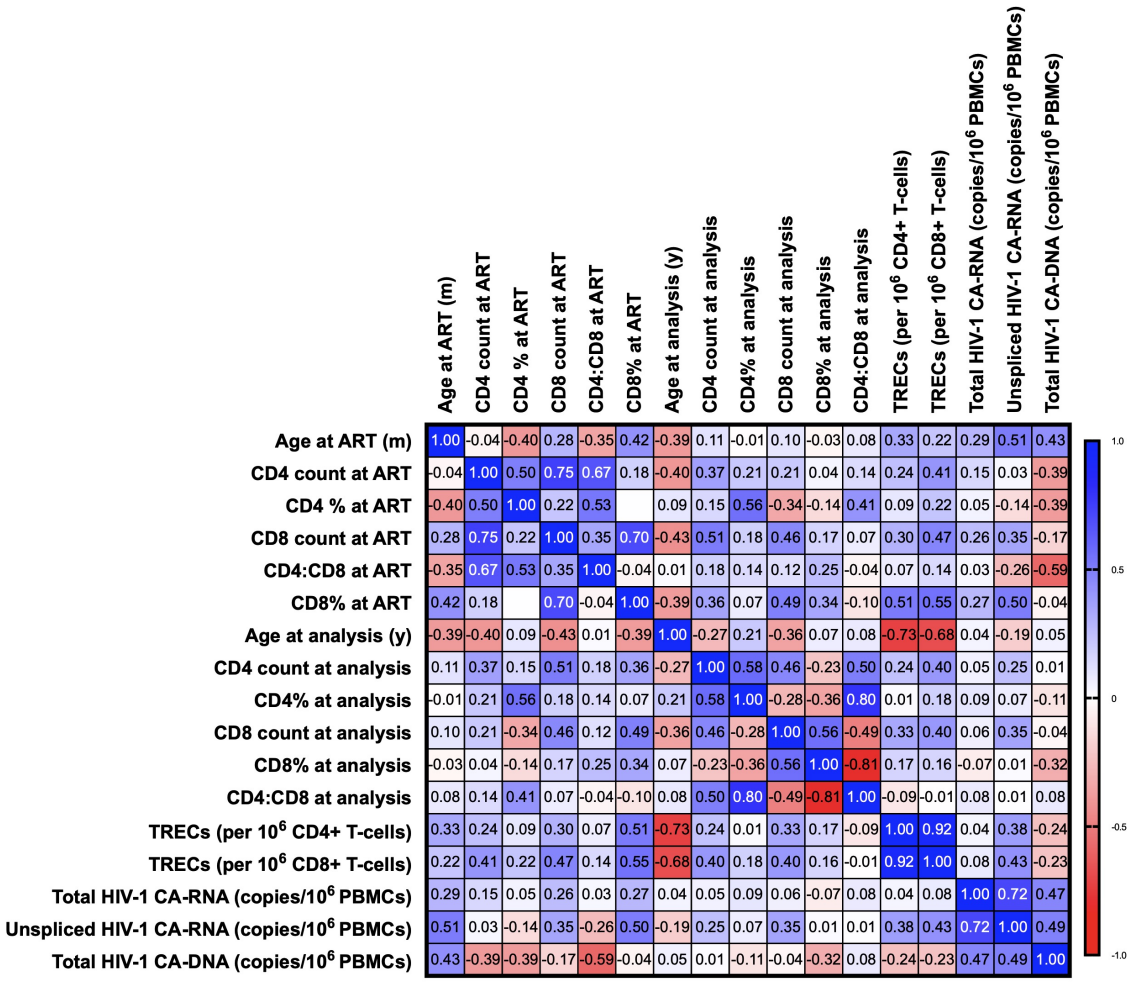
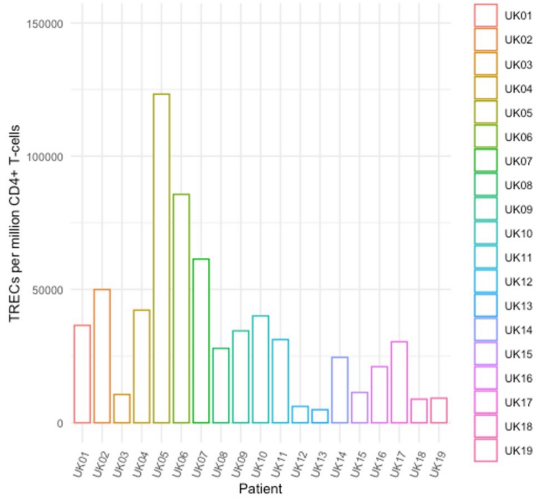


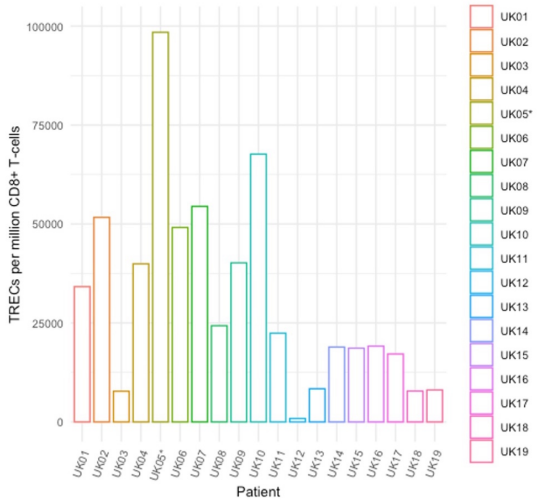
Figure 28 Correlations between different variables measured at time of analysis and ART start. The heatmap indicates the strength of positive (blue) and negative (red) correlations present. The calculated Spearman's correlation coefficient (ρ) is also depicted in each corresponding square.

In this chapter, assessment of thymic output was performed by quantifying TRECs levels in sorted CD4+ and CD8+ T-cells (Figure 29). For CD4+ T-cells TRECs levels were ranging between 4905-123290 per million cells. For CD8+ T-cells TRECs levels were ranging between 860-98432 per million cells. TRECs reference levels were kindly provided by the GOSH clinical lab and Dr Stuart Adams.

A



B



C

Age Group (Years)	TRECs per 10^6 CD4+ T-cell		TRECs per 10^6 CD8+ T-cell	
	Median	10th Centile	Median	10th Centile
0-2	39039	23755	42251	28635
2-5	20813	11476	26731	11462
6-12	28183	11196	24193	9319
12-18	17262	N/A	23716	N/A
Adult	3859	595	13100	3237

Figure 29 TRECs number change as we age, with their levels constantly declining as life progresses. Here TRECs levels per million A) CD4+ and B) CD8+ T-cells are presented for the 19 UK CARMA patients. C) TRECs reference levels kindly provided by Dr Stuart Adams and the GOSH clinical lab team are shown.

As seen on Table 18, most of the study participants present TRECs levels that correspond with their age ranges. As seen in the figure above, no reference data were available to account for the 10th centile in the 12-18 age group. Thus, no definite conclusions can be drawn for CARMA participants in that age group that present TRECS levels below the reference median (e.g., UK03). A notable difference can be seen with participant UK12 where TRECs per million CD8+ T-cells are much lower compared to other study participants, even ones of similar age. This could be associated with possible suboptimal ART adherence discussed with the clinical team at the time of the study.

Table 18 Age at analysis (in years) and TRECs levels (per million CD4+ and CD8+ T-cells) are presented for all 19 UK CARMA patients.

Patient ID	Age at Analysis (y)	TRECs per 10^6 CD4+ T-cells	TRECs per 10^6 CD8+ T-cells
UK01	10.6	36535	34160
UK02	9.4	49951	51678
UK03	17.7	10615	7747
UK04	9.0	42245	39931
UK05	9.1	123290	98432
UK06	12.4	85653	49123
UK07	14.0	61367	54451
UK08	15.7	27875	24302
UK09	14.0	34458	40185
UK10	13.8	40063	67661
UK11	13.8	31213	22416
UK12	19.2	6142	860
UK13	16.1	4905	8370
UK14	14.3	24512	18907
UK15	17.2	11387	18628
UK16	6.9	21006	19142
UK17	10.8	30379	17155
UK18	18.4	8861	7788
UK19	14.8	9234	8059

Through RNA sequencing and primary analysis of the generated FASTQ files, CDR3 sequences for all 19 samples were extracted. As seen in Figure 30, there was a range between the read number obtained per sample, possibly affected by the sample quality and the number of cells available for sorting and RNA extraction.

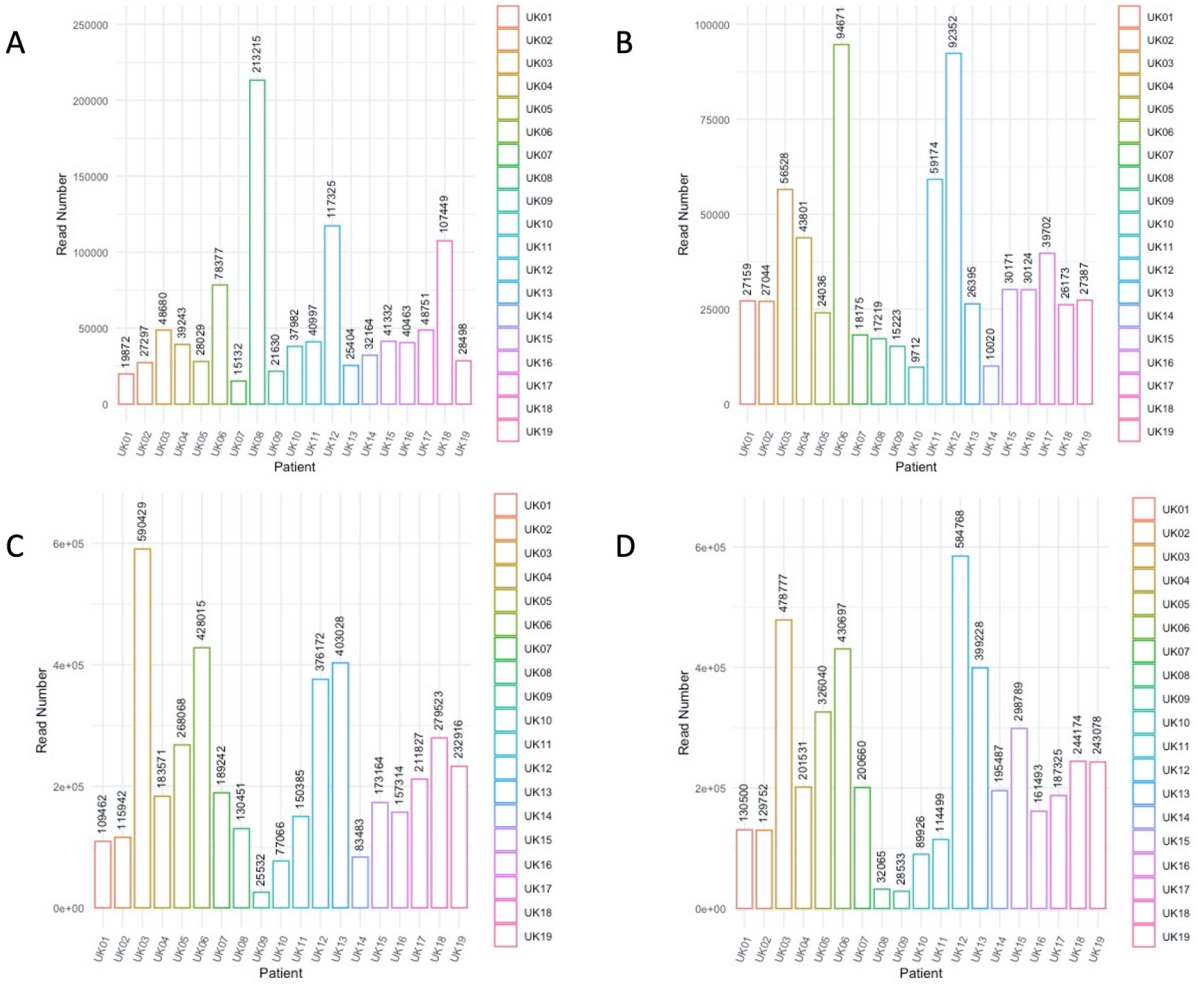


Figure 30 Histogram presenting read number per sample for CD4+ A) TCRα and B) TCRβ sequences and CD8+ C) TCRα and D) TCRβ sequences.

For comparison purposes and to determine the TCR diversity more accurately, data deriving from TCRα and TCRβ sequences were subsampled to the same read depth level, using an in-house algorithm developed to calculate the minimum representative subsampling depth per sample[203]. CD4+ TCRα and TCRβ data were subsampled together at a read depth of 5983 reads. Due to the differences in sequencing read numbers between samples, a small number of them was excluded from the initial subsampling analysis. This is due to the limitations of the algorithm used for subsampling, since the maximum number of samples to be included is optimised based on the total number of Decombinator-identified reads and their minimum subsampling read depth[203]. For that reason, samples that present much higher or much lower sequencing read depth numbers are automatically rejected from the pipeline. These excluded

samples had to be subsampled either to a higher read depth of 49584 (UK08 CD4+ TCRα, UK13 CD4+ TCRα and TCRβ) or a lower read depth of 1905 (UK06 CD4+ TCRα and TCRβ and UK11 CD4+ TCRα), so their diversity metrics should be interpreted with caution when comparing them with the remaining CARMA samples.

The majority of CD8+ TCRα and TCRβ data were subsampled at a read depth of 1595 reads. Samples with lower numbers of raw reads (UK01, UK05, UK06, UK14 and UK19) were excluded from the subsampling algorithm used and had to be further subsampled to 500 reads. Changing the subsampling parameters to the lowest number of reads (500 in this case) would have led to the loss of valuable information and would have affected the subsequent analysis and would have biased the conclusions presented in this chapter. Samples that had higher numbers of raw reads were also excluded and subsequently subsampled at 29968 reads.

The TCR repertoire diversity was calculated using the subsampled data and the Gini coefficient measure. As seen in Figure 31 below, both in CD4+ and CD8+ TCRs the Gini coefficient presented similar levels, demonstrating similarities in the degrees of clonal expansions and a high TCR repertoire diversity between early treated patients.

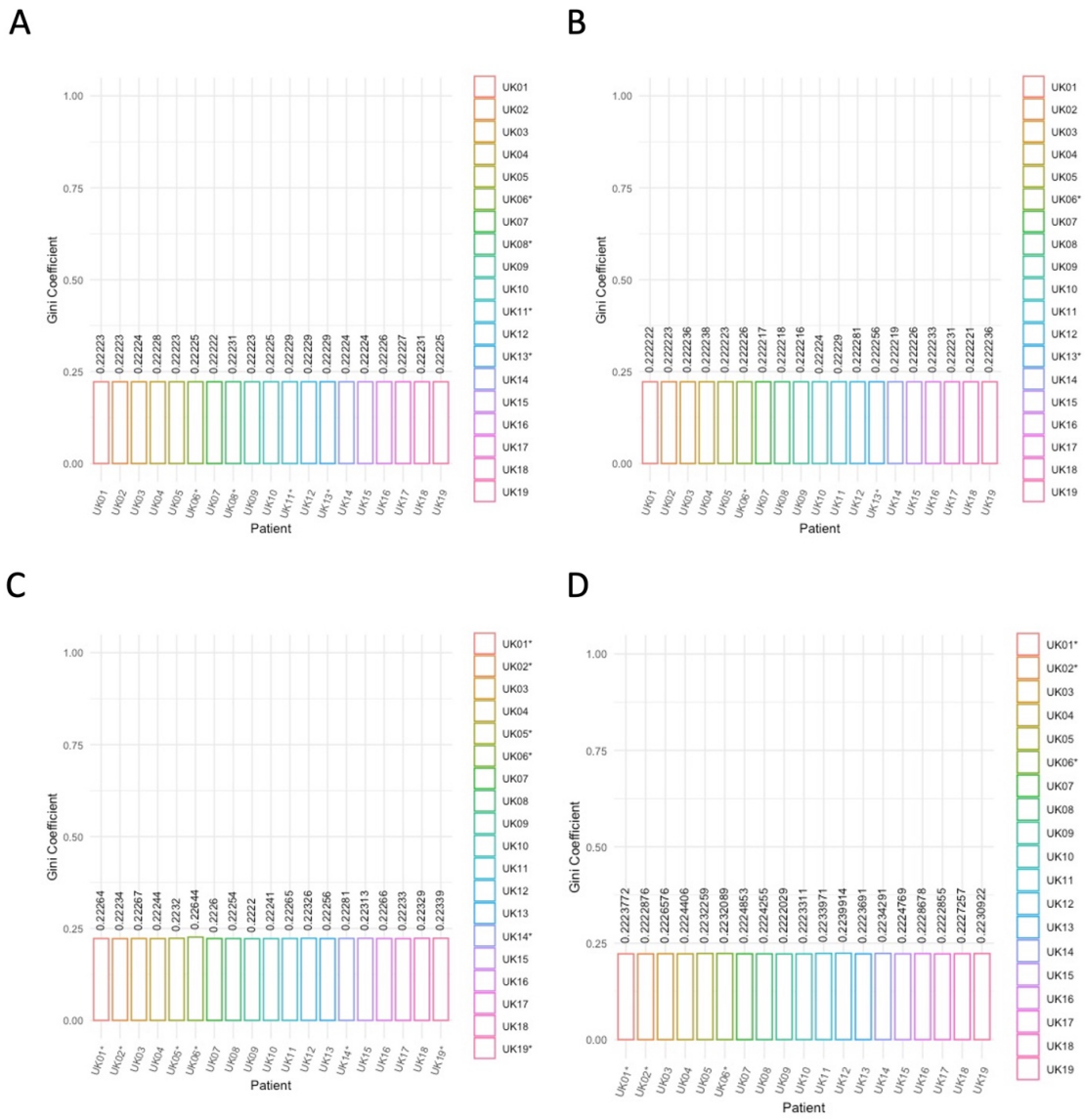


Figure 31 Sequencing data were subsampled to allow for comparisons. Subsampled data were used to calculate the Gini coefficient presented here in the form of histogram plot, demonstrating similar diversity distribution between patients, for CD4+ A) TCRα, B) TCRβ sequences and CD8+ C) TCRα, D) TCRβ.

To assess the degree clonotype sharing between patients, Excel (Microsoft, version 16.65) was utilised. An example of shared CDR3 sequences and the frequencies in which they are seen in a patient can be seen in the figure below (Figure 32). Out of the subsampled CD4+ TCRα sequencing reads, 118741 different CDR3 sequences were detected. Out of those CDR3s only 14829 (12.5%) clonotypes were shared between at least two patients. Similarly, in CD4+

TCRβ samples 128462 different CDR3 sequences were found, most being unique between patients. Only 2836 (2.2%) sequences shared between two or more patients.

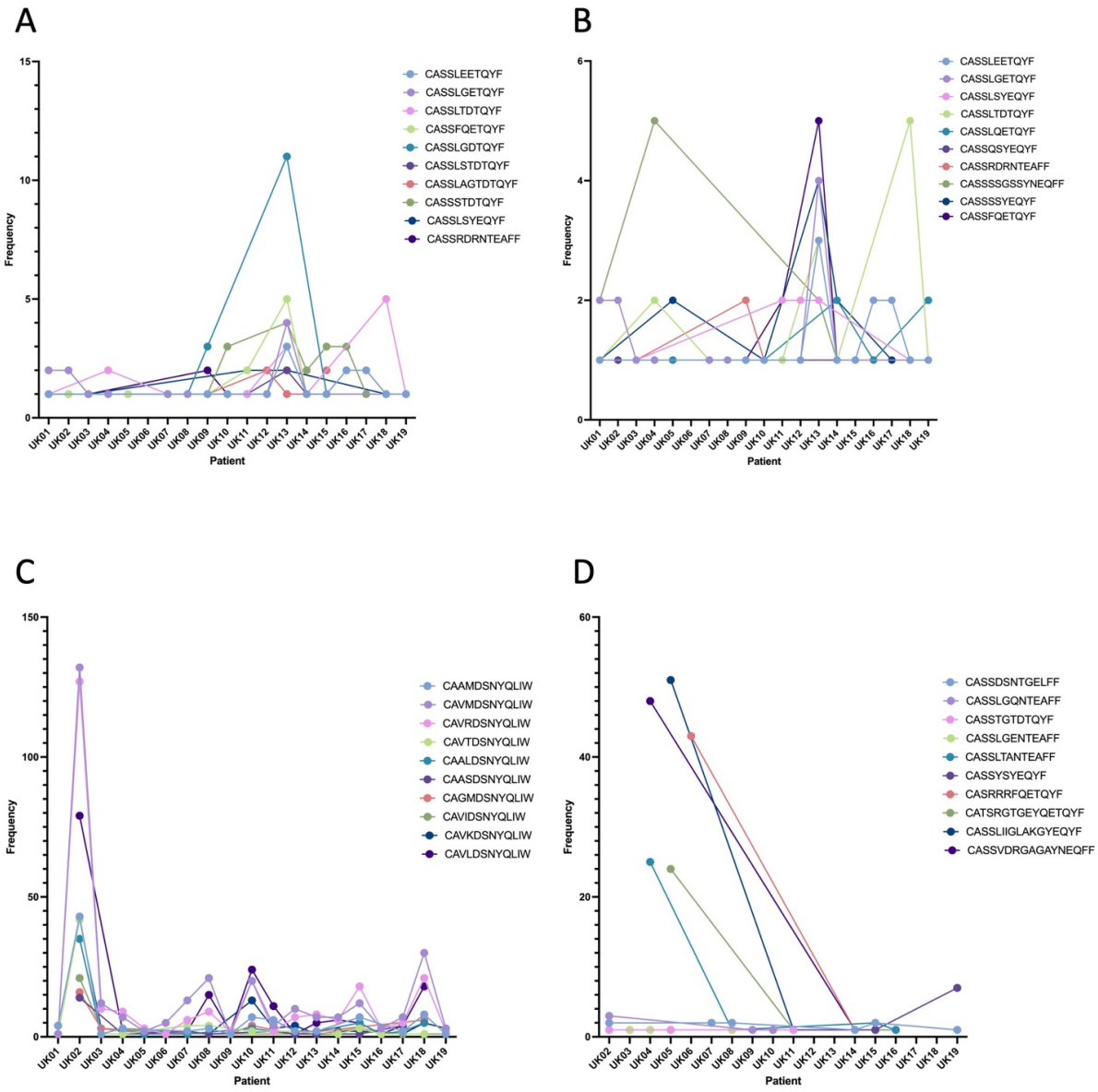


Figure 32 Plots presenting examples of CD4+ A) TCRα, B) TCRβ and CD8+ C) TCRα, D) TCRβ CDR3 sequence sharing and their frequencies between at least two UK CARMA patients. Only a limited number of the detected CDR3 sequences are shared between patients.

In CD8+ TCRα samples, 30336 different CDR3 sequences were identified, 2456 (8.09%) were shared between at least two patients. A snapshot of the shared CDR3 sequences can be seen

in Figure 32 above. Two sequences, CAVMDSNYQLIW and CAVRDSNYQLIW, were present in all 19 CARMA study participants. These two sequences are highly similar, having only one amino acid different, suggesting possible similarity in their function. They have previously been associated with invariant T-cell sequences[244]. In CD8+ TCRβ samples, 38130 different clonotypes were found, 558 (1.46%) of which were shared between at least two patients.

Using VDJdb, an open-access TCR database (data from version "SearchTable-2021-05-20 23_12_40.221") and combined TCRα and TCRβ shared CDR3 sequences between patients in the UK CARMA cohort, potential antigen specificity/epitope recognition was investigated. Several shared CDR3 sequences appear to be specific to more than one antigen. An example of CDR3 sequences and their reported antigen specificity/epitope recognition according to the VDJdb database can be seen in the table below (Table 19). Out of the 17664 shared CD4+ CDR3 sequences, 67 showed potential specificity to HIV-1, accounting for 0.38%.

Table 19 Antigen specificity found in the combined TCRα and TCRβ shared CDR3 sequencing data by using the VDJdb database. Multiple antigen matches can be seen for CD4+ CDR3 sequences, implying cross-reactivity. An example of the matches can be seen below.

CDR3 Sequence	Gene	Antigen Specificity/Match
CAALTGGGNKLTF	TRA	CMV, EBV, HIV-1
CAAMDSNYQLIW	TRA	CMV, EBV, Influenza A, HIV-1
CAERTGGGNKLTF	TRA	HIV-1
CAFTSGTYKYIF	TRA	CMV, EBV, HIV-1
CAGADDKIIF	TRA	YFV, HIV-1
CALGGSYIPTF	TRA	HIV-1
CASMDSNYQLIW	TRA	CMV, EBV, HIV-1
CASSLDRETQYF	TRA	HIV-1
CASRDSSYEQYF	TRA	HIV-1

*Abbreviations: CMV, Cytomegalovirus; EBV, Epstein-Barr virus; YFV, Yellow Fever Virus

Similarly, shared CD8+ CDR3 (both TCRα and TCRβ) sequences matched with specificity against multiple antigens including CMV, EBV and Influenza A (Table 20). Out of the 3014 shared CD8+ CDR3 sequences, only 16 have previously been associated with HIV-1 specificity, accounting for 0.53%.

Table 20 Antigen specificity in TCRα and TCRβ CD8+ CDR3 sequences assigned by the VDJdb database. Multiple antigen matches can be seen for CD8+ CDR3 sequences. An example can be seen below.

CDR3 Sequence	Gene	Antigen Specificity/Match
CAALDSNYQLIW	TRA	CMV, EBV
CAAMDSNYQLIW	TRA	CMV, Influenza A
CASMDSNYQLIW	TRA	CMV, EBV, HIV-1
CASSLDSYEQYF	TRB	CMV, Influenza A, HIV-1
CASSQDRREQYF	TRB	Influenza A, HIV-1
CAVDSNYQLIW	TRA	CMV, Influenza A, HIV-1, YFV
CAVDTNAGKSTF	TRA	CMV, HIV-1
CAVNNARLMF	TRA	CMV, HIV-1
CAVRDSNYQLIW	TRA	CMV, EBV, HIV-1

*Abbreviations: CMV, Cytomegalovirus; EBV, Epstein-Barr virus; YFV, Yellow Fever Virus

Hamming distances were calculated for both CD4+ and CD8+ (TCRα and TCRβ) populations to investigate sequence similarities by determining the number of amino-acid (AA) changes present between CDR3 sequences of the same overall length (Figure 33). The calculated hamming distances between the study participants presented with similar distribution patterns, with mostly dissimilar sequences in the CD4+ TCRα and TCRβ, as well as the CD8+ TCRα populations. In these populations, very low frequencies of low Hamming distances (e.g., distances equal to 1, 2, 3, 4) were observed indicating the presence of a diverse TCR repertoire. A notable difference in the frequency distributions was observed for the CD8+ TCRβ population with some patients, like UK12, presenting with slightly higher frequencies of low hamming distances indicating the presence of more similar CDR3 sequences.

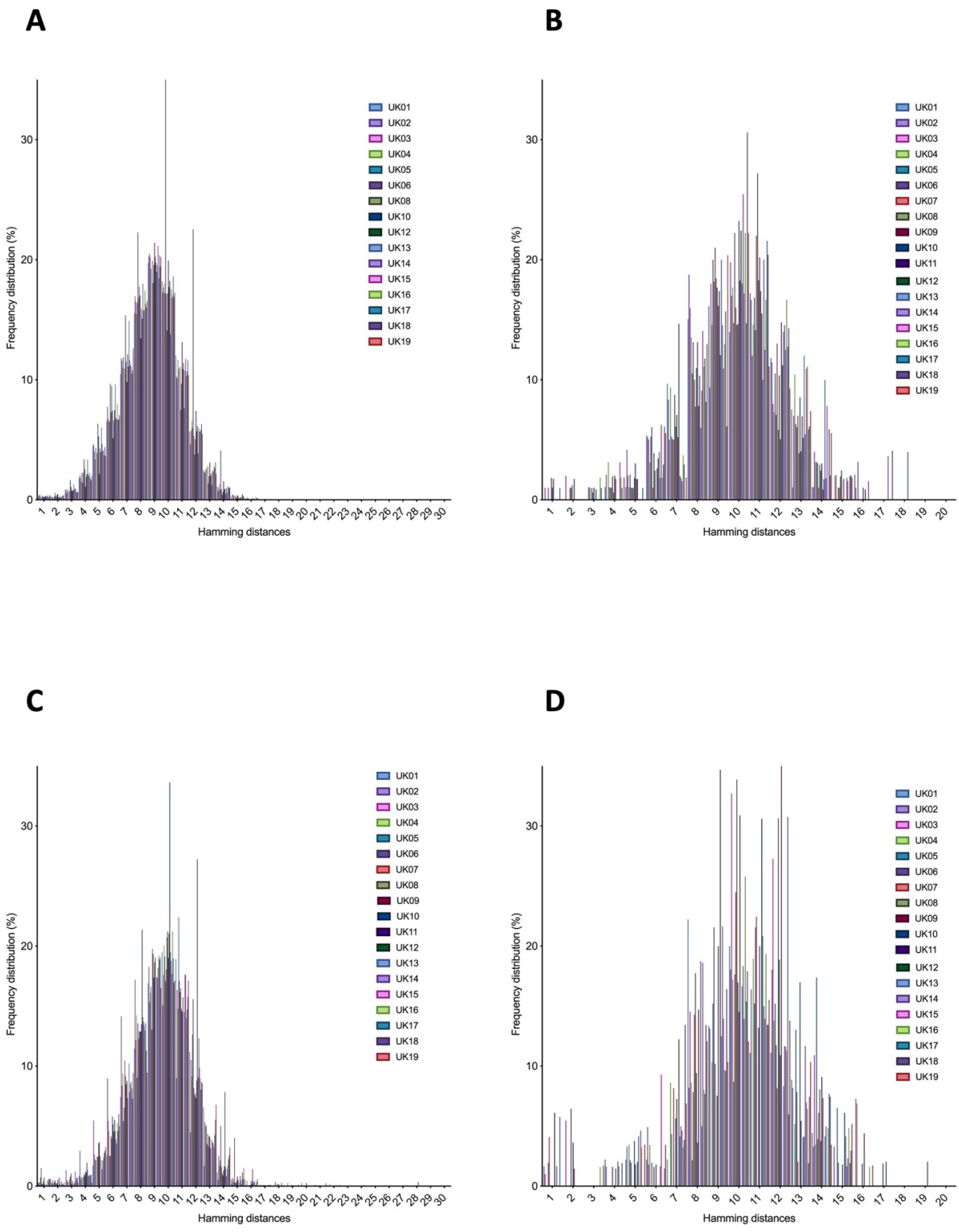


Figure 33 Frequency distribution of Hamming distances between CD4+ A) TCRα, B) TCRβ and CD8+ C) TCRα, D) TCRβ CDR3 sequences. A Hamming distance of 1 represents a single amino acid change between two neighbouring CDR3 sequences with the same length.

Finally, the presence of clusters of closely related CDR3 sequences was identified, as previously described by the detection of amino acid triplet sharing within the subsampled sequencing data per CARMA patient (*Figure* 34 and *Figure* 35)[206]. CD4+ and CD8+ TCRα CDR3 sequences are selected for data visualisation. CD4+ and CD8+ TCRβ can be found in APPENDIX 6. Since some of the CD8+ TCRα data were subsampled at a lower level (500 reads) they comprise of less data to determine the presence of CDR3 clusters. For that reason, it was not possible to obtain cluster plots for UK01 and UK05.

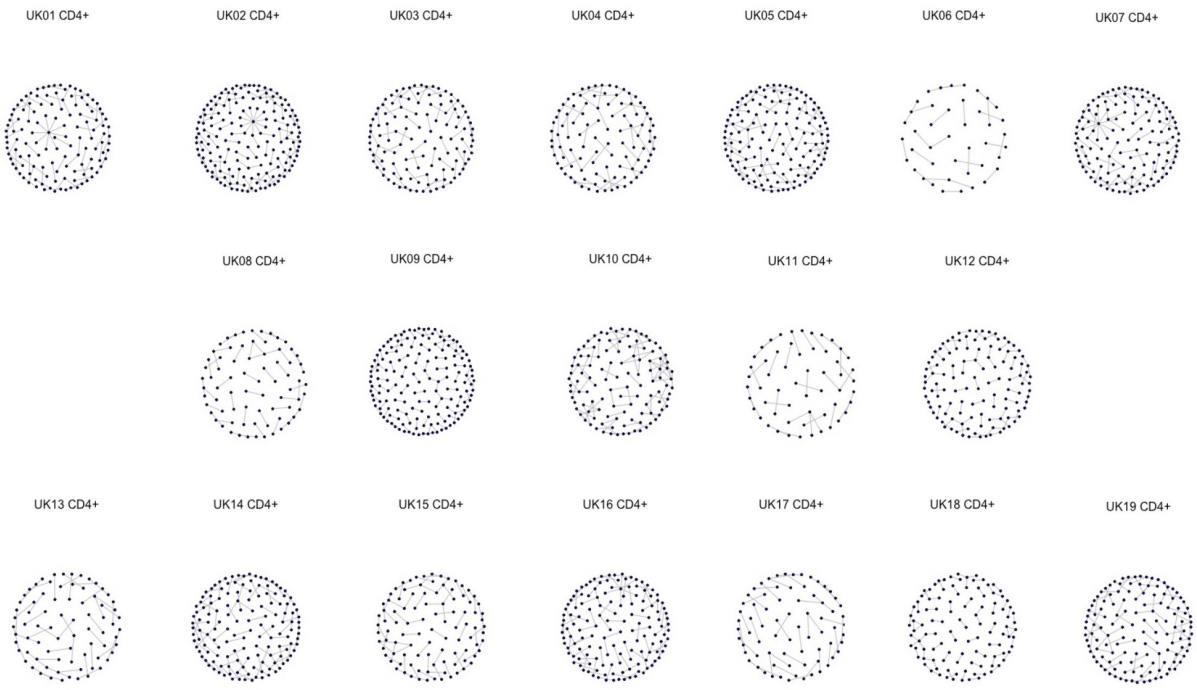


Figure 34 Clusters of closely related CD4+ TCRα CDR3 sequences. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes.

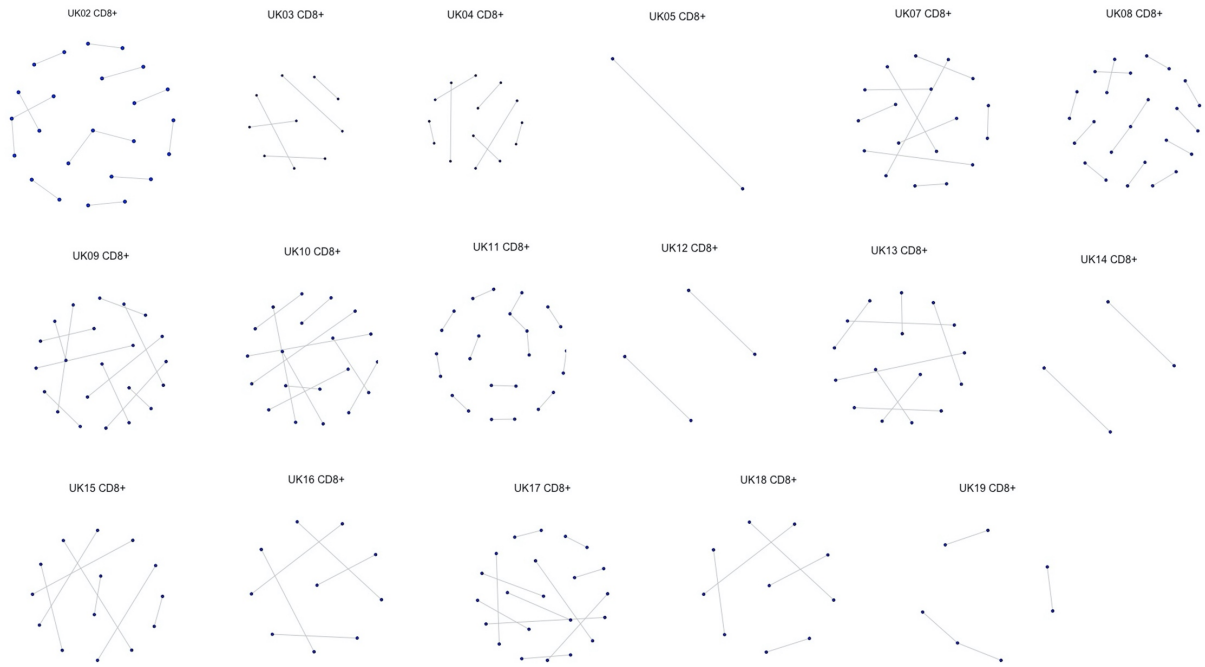


Figure 35 Clusters of closely related CD8+ TCRα CDR3 sequences. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. Due to the lower subsampling level of some samples plots for UK01 and UK05 could not be generated.

Finally, the shared CDR3 sequences were compared to known MAIT cell sequences using the MAIT Match Server. This open-access database annotates a score from 0 to 1 to each sequence, depending on its similarity to previously published MAIT sequences (with 1 being a perfect match). Out of the 17664 CD4+ CDR3 sequences 55 (0.31%), appear to be a perfect match to known MAIT cell sequences (Table 21).

Table 21 Sequence similarity between CD4+ CDR3s and MAIT cells. Presented in this table are only the 50 first sequences, sorted by their similarity score.

CDR3 Sequence	MAIT Sequence	Similarity Score
CAASKAAGNKLTF	CAASKAAGNKLTF	1
CAGMDSNYQLIW	CAGMDSNYQLIW	1
CAVGDNAGNMLTF	CAVGDNAGNMLTF	1
CAVNNNNDMRF	CAVNNNNDMRF	1

CAVPSGGSYIPTF	CAVPSGGSYIPTF	1
CAVRDSNYQLIW	CAVRDSNYQLIW	1
CAALDSNYQLIW	CAALDSNYQLIW	1
CARMDSNYQLIW	CARMDSNYQLIW	1
CASMDSNYQLIW	CASMDSNYQLIW	1
CAVLNTGGFKTIF	CAVLNTGGFKTIF	1
CAVMDSSYKLIF	CAVMDSSYKLIF	1
CAVSDSNYQLIW	CAVSDSNYQLIW	1
CAVTDSNYQLIW	CAVTDSNYQLIW	1
CAASDSNYQLIW	CAASDSNYQLIW	1
CAGGTGTASKLTF	CAGGTGTASKLTF	1
CAPMDSNYQLIW	CAPMDSNYQLIW	1
CAVGDSNYQLIW	CAVGDSNYQLIW	1
CAVKDSNYQLIW	CAVKDSNYQLIW	1
CAVMDSNYQLIW	CAVMDSNYQLIW	1
CAVRDSSYKLIF	CAVRDSSYKLIF	1
CAVRDTGFQKLVF	CAVRDTGFQKLVF	1
CAVRGNFNKFYF	CAVRGNFNKFYF	1
CAVSDSSYKLIF	CAVSDSSYKLIF	1
CVPMDSNYQLIW	CVPMDSNYQLIW	1
CAAMDSNYQLIW	CAAMDSNYQLIW	1
CALNDMRF	CALNDMRF	1
CAPSGSARQLTF	CAPSGSARQLTF	1
CAVGYSSASKIIF	CAVGYSSASKIIF	1
CAVRDNDYKLSF	CAVRDNDYKLSF	1
CAVVDSNYQLIW	CAVVDSNYQLIW	1
CAVRDSDYKLSF	CAVRDSDYKLSF	1
CAAVDSNYQLIW	CAAVDSNYQLIW	1
CAFMDSNYQLIW	CAFMDSNYQLIW	1
CAGLDSNYQLIW	CAGLDSNYQLIW	1
CAIMDSNYQLIW	CAIMDSNYQLIW	1

CATMDSNYQLIW	CATMDSNYQLIW	1
CAVDDNTDKLIF	CAVDDNTDKLIF	1
CAVFSDGQKLLF	CAVFSDGQKLLF	1
CAVGADDYKLSF	CAVGADDYKLSF	1
CAVIDSNYQLIW	CAVIDSNYQLIW	1
CAVLDSNYQLIW	CAVLDSNYQLIW	1
CAVLDSSYKLIF	CAVLDSSYKLIF	1
CAVPNSGGSNYKLTF	CAVPNSGGSNYKLTF	1
CAVQDSNYQLIW	CAVQDSNYQLIW	1
CAVRDGDYKLSF	CAVRDGDYKLSF	1
CAVRDGNYQLIW	CAVRDGNYQLIW	1
CAVRDNFNKFYF	CAVRDNFNKFYF	1
CAVRDRDYQLIW	CAVRDRDYQLIW	1
CAVRGGDYKLSF	CAVRGGDYKLSF	1
CAASKAAGNKLTF	CAASKAAGNKLTF	1

Out of the 3014 CD8+ CDR3 sequences 39 (1.3%) appear to be a perfect match to known MAIT cell sequences (Table 22).

Table 22 Sequence similarity between CD8+ CDR3s and MAIT cells. Presented at this table are only the first 50 sequences, sorted by their similarity score.

CDR3 Sequence	MAIT Sequence	Similarity Score
CAAMDSNYQLIW	CAAMDSNYQLIW	1
CASMDSNYQLIW	CASMDSNYQLIW	1
CAVLDSSYKLIF	CAVLDSSYKLIF	1
CAVMDSNYQLIW	CAVMDSNYQLIW	1
CAVRDSNYQLIW	CAVRDSNYQLIW	1
CAVRRRDDKIIF	CAVRRRDDKIIF	1
CAVTDSNYQLIW	CAVTDSNYQLIW	1
CAAFDSNYQLIW	CAAFDSNYQLIW	1
CAALDSNYQLIW	CAALDSNYQLIW	1

CAASDSNYQLIW	CAASDSNYQLIW	1
CAAVDSNYQLIW	CAAVDSNYQLIW	1
CAFMDSNYQLIW	CAFMDSNYQLIW	1
CAGLDSNYQLIW	CAGLDSNYQLIW	1
CAGMDSNYQLIW	CAGMDSNYQLIW	1
CAIMDSNYQLIW	CAIMDSNYQLIW	1
CAPLDSNYQLIW	CAPLDSNYQLIW	1
CAPMDSNYQLIW	CAPMDSNYQLIW	1
CARMDSNYQLIW	CARMDSNYQLIW	1
CATMDSNYQLIW	CATMDSNYQLIW	1
CAVGDSNYQLIW	CAVGDSNYQLIW	1
CAVIDSNYQLIW	CAVIDSNYQLIW	1
CAVKDSNYQLIW	CAVKDSNYQLIW	1
CAVLDSNYQLIW	CAVLDSNYQLIW	1
CAVLSNDYKLSF	CAVLSNDYKLSF	1
CAVMDSSYKLIF	CAVMDSSYKLIF	1
CAVPSGGSYIPTF	CAVPSGGSYIPTF	1
CAVQDSNYQLIW	CAVQDSNYQLIW	1
CAVRDGDYKLSF	CAVRDGDYKLSF	1
CAVRDNDYKLSF	CAVRDNDYKLSF	1
CAVRDNFNKFYF	CAVRDNFNKFYF	1
CAVRDRDYKLSF	CAVRDRDYKLSF	1
CAVRDRDYQLIW	CAVRDRDYQLIW	1
CAVRDSSYKLIF	CAVRDSSYKLIF	1
CAVRRDDKIIF	CAVRRDDKIIF	1
CAVSDSNYQLIW	CAVSDSNYQLIW	1
CAVVDSNYQLIW	CAVVDSNYQLIW	1
CAVYDSNYQLIW	CAVYDSNYQLIW	1
CVPMDSNYQLIW	CVPMDSNYQLIW	1
CAVRDGNYQLIW	CAVRDGNYQLIW	1
CAVINYGQNFVF	CAVVNYGQNFVF	0.9887

CAALDSSYKLIF	CASLDSSYKLIF	0.9833
CALMDSNYQLIW	CAIMDSNYQLIW	0.9831
CAAMDSSYKLIF	CASMDSSYKLIF	0.9827
CALSDSSYKLIF	CAVSDSSYKLIF	0.9822
CAVISNDYKLSF	CAVLSNDYKLSF	0.9819
CALESNSGYALNF	CAVESNSGYALNF	0.9816
CASLDSNYQLIW	CAALDSNYQLIW	0.9815
CVVMDSSYKLIF	CAVMDSSYKLIF	0.9813
CALMDSSYKLIF	CAVMDSSYKLIF	0.9811
CVVSDSNYQLIW	CAVSDSNYQLIW	0.9806

4.5 Discussion

Generally, studies assessing the impact of ART on the TCR repertoire have been very limited and have primarily involved adult cohorts[156,241]. The impact of early ART on the immunological markers in paediatric cohorts has been severely understudied. Thus far, there have been reports that early treatment initiation leads to improved CD4+ T-cell levels and CD4:CD8 ratios[4,245]. These reports come into agreement with the findings of this study, where apart from an increase in CD4+ T-cell counts after years of continuous ART, the CD4:CD8 T-cell ratio presented an almost three-fold increase.

As we know the TCR repertoire is generated in the thymus, but as individuals reach puberty and adulthood the thymic output is progressively reduced[246]. This declining thymic activity leads to naïve T-cell depletion and memory T-cell accumulation[247]. In this study, TRECS levels were used as a measure of thymic output, as it can be a useful clinical marker of the ability of patients to generate T-cells to replenish the ones lost due to the HIV-1 infection[237,248]. It is worth noting that even though TRECS are a reliable measure of thymic function, they do not account for ongoing cell proliferation in the periphery[249,250]. However, since the patients included in the CARMA study were treated early and have been well suppressed for years, division levels of peripheral T-cell are expected to be low[251,252]. As shown by the Spearman's correlation coefficient there is a strong negative correlation between age and TRECS levels, with younger age corresponding to higher TRECs levels for both CD4+ and CD8+ T-cell populations. These results are in accordance with previously published TRECs studies reporting decreased TRECs levels as life progresses[248,253,254]. In the past years, the units used to measure TRECs levels have been quite wide-ranging, from TRECs/ml, TRECs/µl to TRECS/10^6 T-cells etc. This makes it harder to obtain comparable data to determine differences in TRECs levels between different cohorts and different age groups. For this reason, TRECs range levels were kindly provided to us by the clinical laboratory at GOSH that also tested the samples from the 19 UK CARMA participants. For most of CARMA study participants thymic output appears to be maintained at substantial levels, indicating that their thymus is still contributing to the reconstitution of their immune system when on sustained long-term ART. This could be because ART initiation happened very early on at an age where thymic output is high, limiting the potential damaging effects that an uncontrolled HIV-1

infection can have[251,252]. However, one patient, UK12, presented with almost diminished CD8+ TRECs levels, potentially due to immune activation and an increase to the CD8+ memory T-cell population caused by suboptimal ART adherence and a viral blip during the CARMA study, as reported by the clinical team. From the study baseline clinical data, participant UK12 appeared to be effectively suppressed previously, with viral load levels <50 cells/ml. During the viral blip episode, the CD4+ T-cell count fell to 230 cells/µl. Once ART was re-initiated CD4+ T-cell count levels recovered (1200 cells/µl). More clinical data for this patient was requested to help better elucidate their immunological profile and factors that might have affected and compromised it. However, despite our efforts it was impossible at this instance. Follow-up samples from this participant would have also been useful to determine the effects of treatment re-initiation and viral suppression post the blip episode. General differences in TRECs levels between the CARMA study participants may be an indicator of the natural variability caused by differences in thymic size[248]. It would be of interest to follow this cohort longitudinally and monitor any possible changes in their thymic output as they reach adulthood.

To this day, the dynamics of the TCR repertoire during childhood remain understudied. It is believed that major events and numerous pathogen encounters during childhood can shape the composition of the immune system and T-cell repertoire[138]. Thus far, there are no published paediatric studies highlighting the effect of early ART on the TCR repertoire, constituting this project as quite unique. Our group has previously shown that different ART strategies can affect the composition of the TCR repertoire in paediatric patients[14,204]. Treatment interruption drives clonal expansions in the CD4+ T-cell memory population, changes the repertoire landscape by decreasing diversity and promotes the formation of closely related TCR clusters possibly due to the generation of an antigen specific response[204]. Moreover, we have previously reported that adolescents that present with poor viral control present with lower TCR repertoire diversity, clonal expansions and decreased thymic output[14].

For the purposes of this thesis, the TCR clonal distribution was explored using the Gini coefficient, a measure of inequality originally used in economic studies[203,255,256]. The Gini coefficient has a scale ranging from 0 to 1, with samples tending towards 0 representing

equality i.e., several clonotypes with almost identical frequencies (e.g., a lot of different clonotypes appearing only once), and samples tending towards 1 representing inequality/oligoclonality i.e., few highly expanded clonotypes present (e.g., a few clonotypes appearing numerous times in a study participant)[203]. Generally, in published studies, the diversity of the TCR repertoire is determined by the number of different sequences that are present in an individual as well as the abundance of certain clonotypes. Increased diversity, with a great number of different clonotypes that are not highly expanded, likely shows the ability of an individual to recognise and respond to different antigens[257]. Previous HIV-1 studies have demonstrated an increased Gini coefficient in HIV-1+ adult individuals compared to healthy controls (for controls Gini coefficient values <0.4 were observed), reporting reduced diversity and the presence of a small number of highly expanded TCR sequences[156]. For the CARMA study, it was hypothesised that early ART initiation and continuous treatment can restore the TCR repertoire diversity to almost normal levels. As seen in section 4.4 of this chapter, the reported Gini coefficient amongst this cohort demonstrates homogeneity and a low Gini coefficient. For all CARMA participants, the Gini coefficient presented similarly low levels showing a relatively diverse repertoire with very few clonal expansions and a lot of unique TCR sequences (appearing with the same frequencies). This could imply that early and timely ART initiation reduces the potential damage to the TCR repertoire and the developing immune system, leading to better responses to pathogen exposure.

Next, CDR3 sequence sharing between the study participants was investigated. It has been previously reported that the number of shared sequences between HIV-1+ individuals is substantially reduced[156]. Following subsampling, 14829 (12.5%) out of the 118741 identified TCRα and 2836 (2.2%) out of the 128462 TCRβ CDR3 sequences were shared between at least two CARMA study participants. 2456 (8.1%) out of the 30336 TCRα and 558 (1.5%) TCRβ CDR3 sequences were shared between at least two participants. It is known that sequences found to be shared between many individuals could be characterised as "public"[156,258]. However, since the shared "public" sequences reported in this thesis have been obtained though single (α or β) chain comparisons only, it is most likely that the actual percentage of CDR3 sequence sharing between patients would be lower if both chains were paired[259].

Hamming distances and CDR3 clusters were investigated to further look at the TCR repertoire diversity. In concordance to previous published reports optimal ART adherence and viral control leads to more diversity in the TCR repertoire[14]. The study participants showed similar patterns in their Hamming distances, favouring sequences with more amino acid changes between them, meaning a greater diversity in their repertoire. However, a difference in the hamming distance frequency patterns was noticed when looking at the CD8+ TCRβ population, showing a slight increase in very low and very high hamming distances (towards the two ends of the bell-shaped curves observed on the plot). This could be an indicator of an immune response to a certain pathogen. The presence of closely related TCR clusters was reported for the CD4+ TCRα population shown here. However, despite an even greater number of clusters expected to be present within the CD8+ TCRα population, only a lower number was found. This can most likely be attributed to the lower subsampling levels and the lack of data available for the comparisons. As with TCRα, TCRβ CD4+ and CD8+ populations showed similar patterns, with CD8+ TCRβ samples presenting reduced numbers of clusters (APPENDIX 6).

Next, VDJdb, a public TCR database containing CDR3 sequences with their known antigen specificities, was used as a reference in the analysis of our generated CDR3 data. Previous studies have shown that contrary to initial theories, the TCR repertoire can be biased. The recombination process has been shown to preferentially generate TCR sequences specific to common pathogens, like CMV and EBV[259,260]. This can also be seen from the data presented in this thesis, with specificity/reactivity towards pathogens like CMV, EBV and Influenza A being most common amongst our shared TCRs. An absence of high frequencies of TCR sequences specific towards HIV-1 is apparent in this cohort, potentially attributed to early and sustained ART[156]. Nonetheless, data generated using the VDJdb database should be treated with caution since previous reports have demonstrated cross-reactivity, with even what have been previously reported as HIV-1 specific TCR sequences found in HIV-1 negative individuals[203]. This leads us to believe that a lot of TCR sequences might not be antigen specific, but instead be cross-reactive against multiple antigens[203]. Another reason why the data presented here should be interpreted with caution, is the inability to pair the generated alpha and beta chain data, making it impossible to determine and report the exact specificity

of each clonotype[261,262]. In the future, this issue could be resolved by using single-cell sequencing protocols that allow for α-β chain pairing[263].

Finally, the generated data were for sequences attributed to MAIT cells. Generally, MAIT cells constitute up to 10% of the circulating T-cells in humans[167]. Adult studies on HIV-1+ patients have demonstrated a sharp decrease in their levels suggesting that they are vulnerable to the immune exhaustion caused by the chronic infection[156,264]. Moreover, ART doesn't appear to fully restore their functionality, potentially making these patients more susceptible to opportunistic infections[156,264,265]. Despite the increased scientific interest in this cell population in HIV-1 infection, the mechanisms leading to their depletion and reduced functionality, as well as the effect that these could have in the health of patients remain poorly studied[264]. In this chapter, the intension was to determine if any of the CD4+ and CD8+ CDR3 sequences found in the CARMA cohort could be attributed to known invariant cell sequences, and more specifically those that have been associated with MAIT cells. Despite chronic HIV-1 infection and continuous ART, both CD4+ and CD8+ MAIT cell sequences were identified. Two CDR3 sequences, CAVMDSNYQLIW and CAVRDSNYQLIW, that have been found to be present in all 19 study participants have been previously identified as invariant MAIT cell sequences[244]. This could be pointing to potential similar immunological functions. As expected, there was a higher percentage of CD8+ MAIT cells present, supporting the reports that they constitute the largest pool amongst MAIT cell subpopulations[164]. In this instance the described analysis approach had severe limitations and it was impossible to report on total numbers present individually in these patients or their functionality. These issues arise from the small participant numbers, restricted sample volumes and the reliance on sequences that have already been identified and published as MAIT and exist on the public database. However, this data could still be a promising indicator of the immune system capacity of paediatric patients. If MAIT cells are still present in paediatric cohorts, it could potentially indicate an improved response against opportunistic infections. Thus, warranting further studies both in paediatric and adult HIV-1+ cohorts, to determine the impact of HIV-1 infection on MAIT cell levels and their antimicrobial function, as well as studies in healthy cohorts.

For this chapter, available samples from 19 HIV-1+ individuals enrolled in the UK arm of the CARMA study were processed. The small number of study participants, as well as differences in ethnicity, sex, and age, can be a great limitation in terms of the applicability of our conclusions, since we must err on the side of caution when interpreting all generated data. Further studies on bigger cohorts are necessary to validate our findings. Follow-up studies, that would monitor this cohort longitudinally, would be highly beneficial and enable us to track any changes in their TCR repertoire as these paediatric patients reach adulthood.

A major obstacle of this TCR experimental pipeline and study design was the inability to pair α-β TCR chain data by using single cell sequencing due to funding limitations, even though this option would have provided much lower sequencing throughput[262]. Differences presented here could potentially be due to the availability of unpaired α-β data and the abundance of α chains when compared to β chains, since as previously mentioned one β chain can pair up with as many as 25 different α chains[151]. Moreover, current regulations present at UCL ICH leading to working with "fixed" sorted cells. Specifically, according to the local GOSH ICH CL3 SOP and Risk Assessment (title: 17IR08) in place, the risk of exposure to biological samples containing DNA from pathogens, including blood-borne viruses, such as in this case HIV-1 is high. For that reason, all work including preparation of cells for sorting was carried out within the CL3 class II biosafety cabinet as appropriate to decrease the risk of exposure. Since no samples containing the live virus were allowed to leave the CL3 laboratory, CellFix (BD Biosciences) was used for the fixation of the samples prior to transport to the Flow Cytometry Core Facility. However, cell fixation can potentially affect the quality of the RNA extracted from the sorted CD4+ and CD8+ samples, subsequently affecting the quality and size of sequencing data that are obtained.

Finally, as previously mentioned due to the small number of PBMCs available for each patient, it was not possible to sort and sequence more than the CD4+ and CD8+ total T-cell populations. In future studies, it would be interesting to sort other subpopulations like MAIT T-cells and investigate the TCR sequences present. Further experiments are required to address the abundance and functionality of MAIT T-cells, in HIV-1+ early treated, well-suppressed paediatric patients. Longitudinal studies could also assist in exploring any progressive depletion of this cell population.

4.5.1 Conclusions

This thesis highlights the importance of TCR studies when trying to investigate the effects of pathogens, like HIV-1, on the immune system of individuals, through the course of infection and following continuous treatment. NGS is a powerful tool that can be used to provide important data on repertoire diversity, clonotype presence and "specificity" in individuals[203].These data can help optimise future therapeutic approaches as they can be used to monitor patients throughout different stages of their care.

In conclusion, we believe that early ART initiation can help children form a relatively normal immune system. The presence of relatively normal thymic output levels for their age, a diverse T-cell repertoire, and the presence of MAIT cells, could assist in generating T-cell responses against new antigens and aid patients in cohorts like this one to better respond against opportunistic infections. Studies like this are vital to gage the effect of early ART on the developing immune system, as more and more paediatric HIV-1+ patients are reach adolescence and adulthood.

CHAPTER 5: IMMUNOLOGICAL CHARACTERISTICS OF HIV PERINATALLY INFECTED INDIVIDUALS- IMPACT OF ART INITIATION TIMING ON THE TCR REPERTOIRE

5.1 Background and Objectives

Despite all the progress that has been made towards the prevention of vertical HIV-1 transmission and the reduction in reported cases, children continue to get infected[266]. According to the data published by UNAIDS, in 2021, approximately 1.7 million children (ages 0-14 yeas) are living with HIV-1 but unfortunately only 52% is receiving life-saving ART[267]. To this day, HIV-1 infections present as one of the highest health burdens among children and adolescents globally and their management has proven challenging at times[266,268]. Since 2013, progress in fighting paediatric HIV infections has been made by amending the ART initiation timing requirements[269]. Immediate and early ART initiation at the point of diagnosis is now recommended, regardless of immunological criterial, such as CD4+ counts and their percentages[269].

In this chapter, paediatric patients that received ART early on in life are compared to ones that received ART later in life, as well as healthy controls. It is hypothesised that early ART initiation has a positive impact on the developing immune system of children, by helping develop a high TCR repertoire diversity.

5.2 Introduction

Research has shown that early ART initiation has multiple clinical, virological, and immunological benefits for paediatric patients. For example, it reduces mortality and morbidity levels and positively effects the growth of these children, by improving their weight and height taking them to levels like the ones observed for their age-matched HIV-1 uninfected peers[231,270,271]. Virologically, as discussed earlier in this thesis in Chapter 3, low levels of both total HIV-1 DNA and cell-associated HIV-1 RNA indicate the presence of a

decreased HIV-1 reservoir[2,4,211,219,272]. Immunologically, early ART leads to higher thymic output and appears to prevent the loss of CD4+ T-cells during the initial stages of the infection[5,270,273]. Moreover, as also demonstrated in Chapter 4, timely ART initiation can lead to the development of a relatively normal immune system, with good levels of thymic output and high TCR repertoire diversity.

In this chapter, sequencing data from CD4+ and CD8+ T-cell populations, deriving from children that started ART early or late, as well as healthy controls, are compared to gain a better insight into the effects of treatment initiation timing to the diversity of the TCR repertoire.

5.3 Materials and Methods

The sub-study presented in Chapter 5 was designed as a continuation of our group's research on the effects of ART initiation timing. For this purpose, five participants from the UK arm of the CARMA study that started ART early on in life (UCL R&D: 17IR17) were selected and compared with five similarly aged children and two healthy controls from the "Lymphocyte development (LD) in paediatric HIV-1 infection study (UCL R&D: 17IR08)[4,14,178]. As seen in the table below (Table 23), the participants were separated in three main groups, Group A included early treated CARMA participants, Group B included late treated LD participants and Group C included healthy controls (APPENDIX 2 and APPENDIX 3).

Table 23 Characteristics of the 12 individuals selected to investigate the impact of ART initiation timing on the TCR repertoire (adapted from[4,178]). The participants were separated into three groups, Group A included individuals from the CARMA study that initiated ART early, Group B included individuals from the LD study that initiated ART late, and Group C included healthy controls. The participants were selected with their age as the main criterion, as there was an effort to obtain relatively age-matched participants for each group. Participants in Control Group C are of a slightly older age, as it wasn't possible to obtain samples from healthy volunteers below the 18th year of age. Continuous data are presented as median (IQR), categorical data are presented as number of patients (%).

	Group A (n=5)	**Group B (n=5)**	**Control Group C (n=2)**
Sex, No. (%)			
Female	3 (60.0)	3 (60.0)	0 (0.0)
Male	2 (40.0)	2 (40.0)	2 (100.0)
Race/ethnicity, No. (%)			
White	0 (0.0)	0 (0.0)	1 (50.0)
Black/Black African	3 (60.0)	5 (100.0)	1 (50.0)
Other	2 (40.0)	0 (0.0)	0 (0.0)
Age at HIV diagnosis, years	0.18 (0.14-0.29)	1.3 (0.42-4.75)	NA
Age at ART start, years	0.22 (0.01-0.38)	10.65 (10.30-10.93)	NA
Age at study follow-up sample collection point, years	17.22 (16.06-17.66)	18.54 (15.18-18.61)	22.38 (22.32-22.44)

CD4+ count at follow-up sample collection point, cells/µl	1043 (652.0-1130.0)	320.0 (220.0-740.0)	924.0 (924.0-924.0)
CD4 %	41.0 (37.0-52.0)	28.7 (22.0-45.3)	68.4 (67.6-69.2)
CD8+ count at follow-up sample collection point, cells/µl	636.0 (450.0-665.0)	889.0 (852.4-969.8)	209.5 (187.7-231.2)
CD8 %	28.0 (27.0-34.0)	71.2 (43.3-73.0)	21.5 (21.3-21.6)
CD4:CD8	1.5 (1.1-2.1)	0.4 (0.2-0.9)	4.6 (4.1-5.1)

*Abbreviations: HIV, Human Immunodeficiency Virus; ART, Antiretroviral Therapy

RNA isolates from sorted CD4+ and CD8+ T-cell populations were used to generate libraries that were in turn sequenced using the Illumina MiSeq system and a V2 reagent kit (500 cycles)[202]. As analytically described in Chapter 2, the generated FASTQ files were demultiplexed using the Decombinator suite of python scripts[202,241]. In-house R-scripts (R version 3.6.3) were used for subsampling and for the Gini coefficient calculations[203]. Excel (version 16.65, Microsoft) and the "MATCH" function were utilised to detect any shared CDR3 sequences between the sub-study participants. CDR3 "specificity" was determined by using the VDJdb database (version "SearchTable-2021-05-20 23_12_40.221") and an in-house R-script (R version 3.6.3)[203,204,243,274]. Hamming distances and CDR3 sequence clusters were used to investigate sequence similarity[14,204]. Detection of MAIT cell sequences amongst the generated sequencing data was done using the MAIT Match Server database (DTU bioinformatics). Finally, Prism (GraphPad, version 9.0.0) and the ggplot2 R package (version 3.3.5) were utilised for plot generation.

TRECs (per million CD4+ cells) data for the selected 12 participants were kindly measured by Dr Stuart Adams and Susanne Kirke at the SIHMDS Molecular Department of Haematology (Camelia Botnar Laboratories, Great Ormond Street Hospital). Reference data on expected TRECs levels for specific age groups were also provided (Table 24).

Table 24 TRECs per million CD4+ cell reference data provided by the SIHMDS Molecular Department of Haematology lab, presented here as median and 10th centile.

Age Group (years)	Median	10th Centile
0-2	39039	23755
2-5	20813	11476
6-12	28183	11196
12-18	17262	N/A
Adult	3859	595

5.4 Results

For this Chapter, the thymic output per participant was evaluated by quantifying the TRECs levels in sorted CD4+ T-cells. As it can be seen in the figure below (Figure 36), most CARMA participants presented with higher TRECs per million CD4+ cells levels when compared to LD participants and healthy controls. As per the reference levels provided above (Table 24) all participants that could be classed in the adult category (over 18 years of age) present with "normal" TRECs levels for their age. For LD participants UK03, UK13, UK15, V68 and V78, the levels present were lower than the reference median TRECs per million CD4+ levels, however no conclusions can be drawn since no data are available for the 10th centile of the 12-18 years age group.

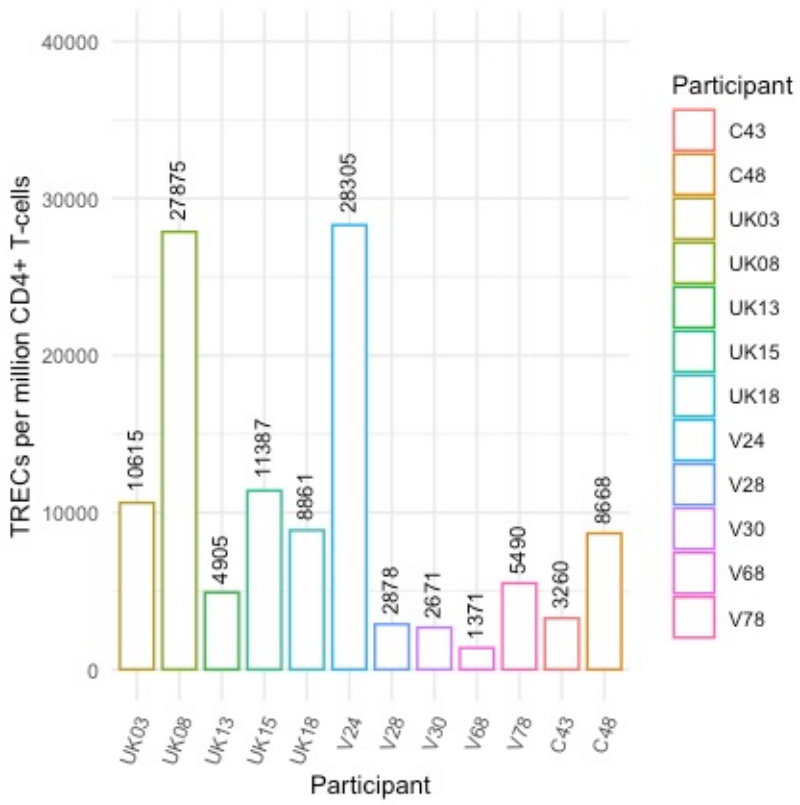


Figure 36 TRECs levels per million CD4+ T-cells as measured for all seven sub-study participants. Age is one of the main factors affecting TRECs levels, with progressive loss being apparent as we age.

For comparison purposes, all raw sequencing data deriving from both CD4+ and CD8+ T-cell populations were subsampled down to the same read depth. As previously mentioned in this thesis, the subsampling algorithm used here computes the minimum representative read depth for each sample and optimises the maximum number of samples that can be included in the study depending also on the total number of sequences present in each of them[203]. For that reason, samples that have either too high or too low total sequences included in their raw sequencing files get excluded from the initial subsampling analysis. Here data from both CD4+ and CD8+ populations were subsampled in two groups respectively to account for the difference in total read numbers identified in the raw samples. Specifically, CD4+ T-cell sequences were subsampled down to either 11252 (UK03, UK08, UK13, UK15, UK18b, V24a, V68 and V78a) or 2021 (UK18a, C43, C48, V24b, V28, V30 and V78b) reads. CD8+ T-cell sequences were subsampled down to either 12606 (V24, V28, V68 and V78) or 1995 (C43, C48, V30, UK03, UK08, UK13, UK15 and UK18) reads.

TCR repertoire diversity was calculated by using the Gini coefficient, a measure of inequality originally used in the field of economics to determine wealth distribution. For the Gini coefficient the scale ranges from 0 to 1, with 0 indicating complete equality (i.e., x number of clones with identical frequencies) and 1 indicating complete inequality (i.e., oligoclonal samples present). In this case, the Gini coefficient allows for a summary of the clonotype abundance distribution amongst our samples. As seen below (Figure 37), all 7 selected samples present with similar Gini coefficient values (between 0.2-0.3) indicating a relatively normal TCR repertoire present in all three groups.

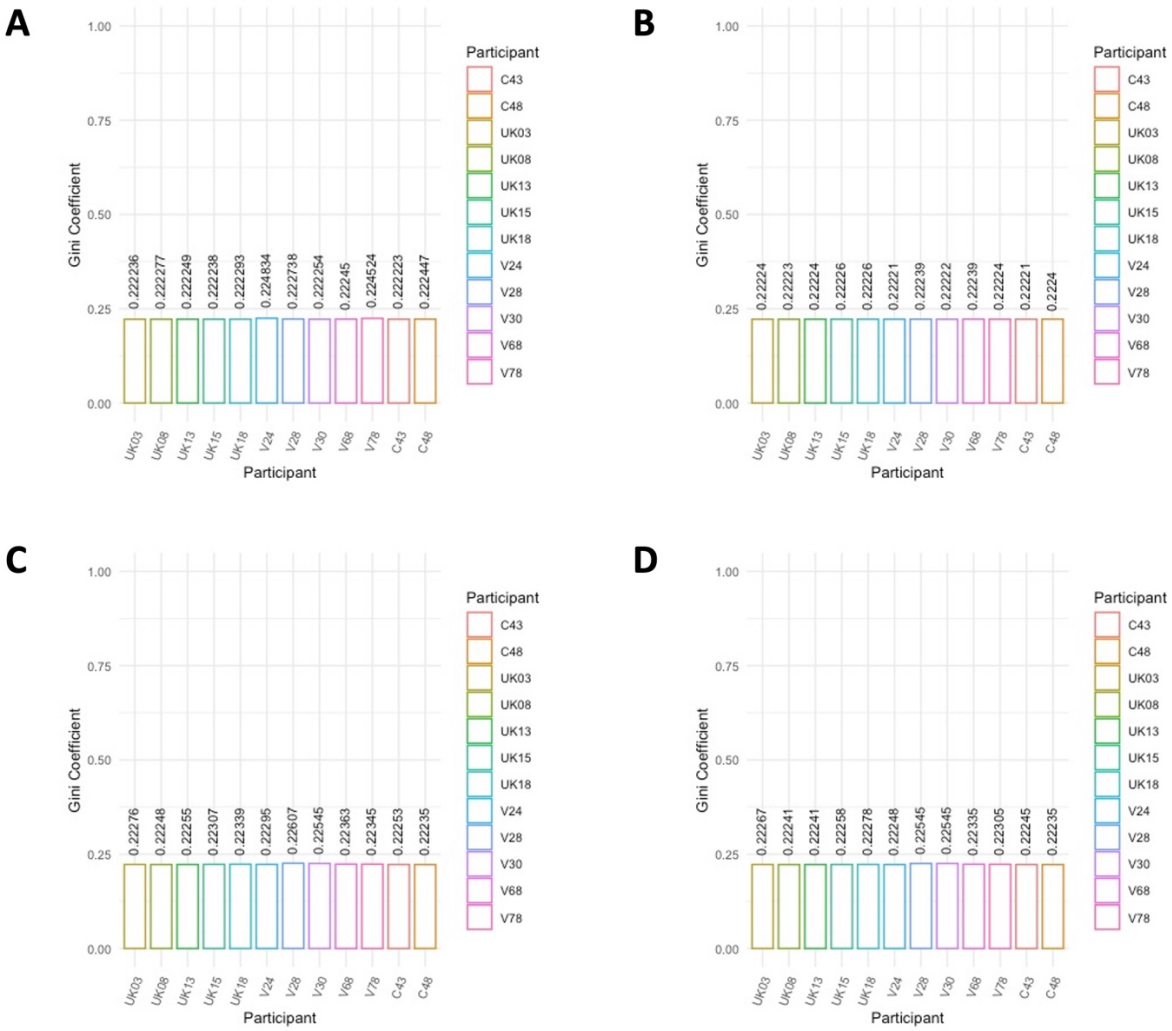


Figure 37 Subsampled sequencing data were used to calculate the Gini coefficient as seen here on this histogram plot. Similar clonotype abundance distribution was observed between all patients for CD4+ A) TCRα and B) TCRβ as well as CD8+ C) TCRα and D) TCRβ sequences.

Using Excel (Microsoft, version 16.65) the degree of clonotype sharing between the different groups included in this sub-study was assessed. For the CD4+ TCRα samples, 6.2%, 2.5% and 0.7% of the identified CDR3 sequences were shared between at least two participants within Group A (CARMA), Group B (LD) and Group C (healthy controls) respectively. For the CD4+ TCRβ samples, 1%, 0.4% and 0.1% of the detected CDR3 sequences are shared between Group A, Group B and Group C respectively (Figure 38).

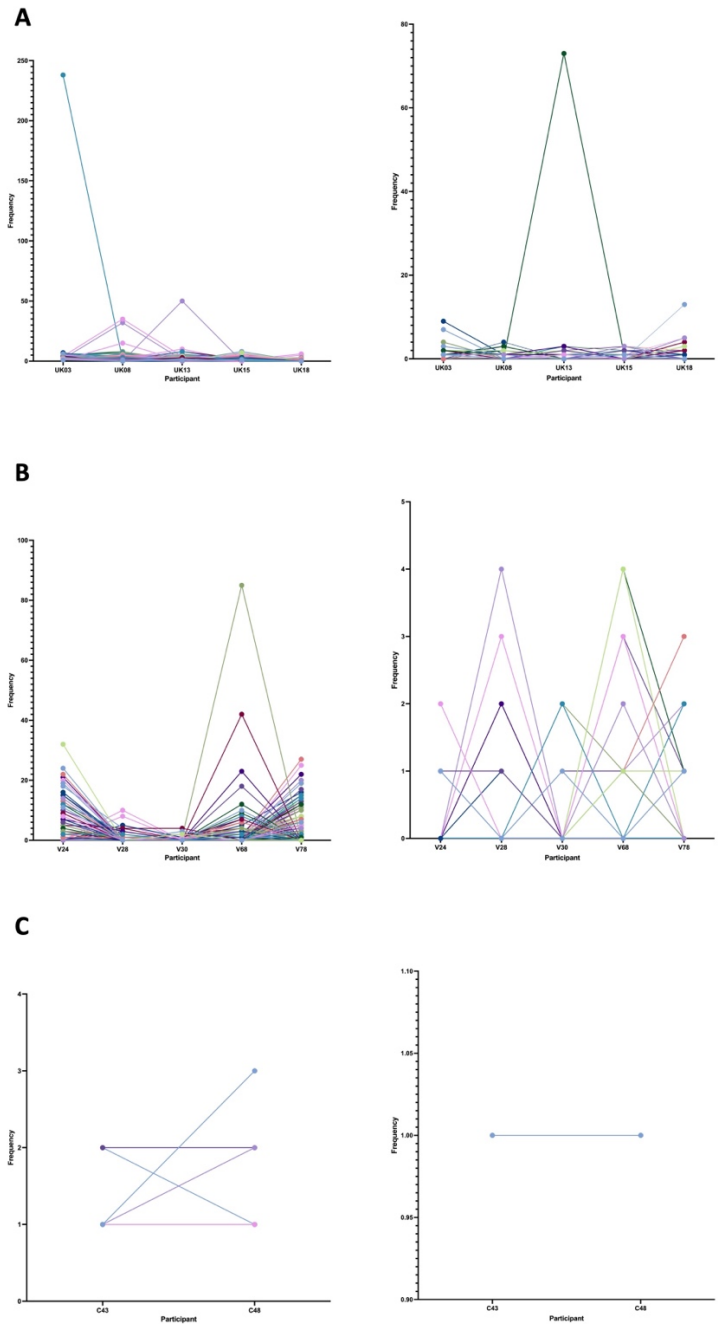


Figure 38 Plots presenting the TCRα and TCRβ CD4+ CDR3 sequences shared between at least two patients in A) Group A, early treated CARMA study participants, B) Group B, late treated LD study participants and C) Group C, healthy controls.

For the CD8+ TCRα samples, 2.1%, 2.4% and 1.1% of the CDR3 sequences were shared between at least two participants within Group A, Group B and Group C respectively. For the CD8+ TCRβ samples, 0.5%, 0.6% and 0.1% of the identified CDR3 sequences were shared

between at least two of the participants within each of the three groups respectively (Figure 39).

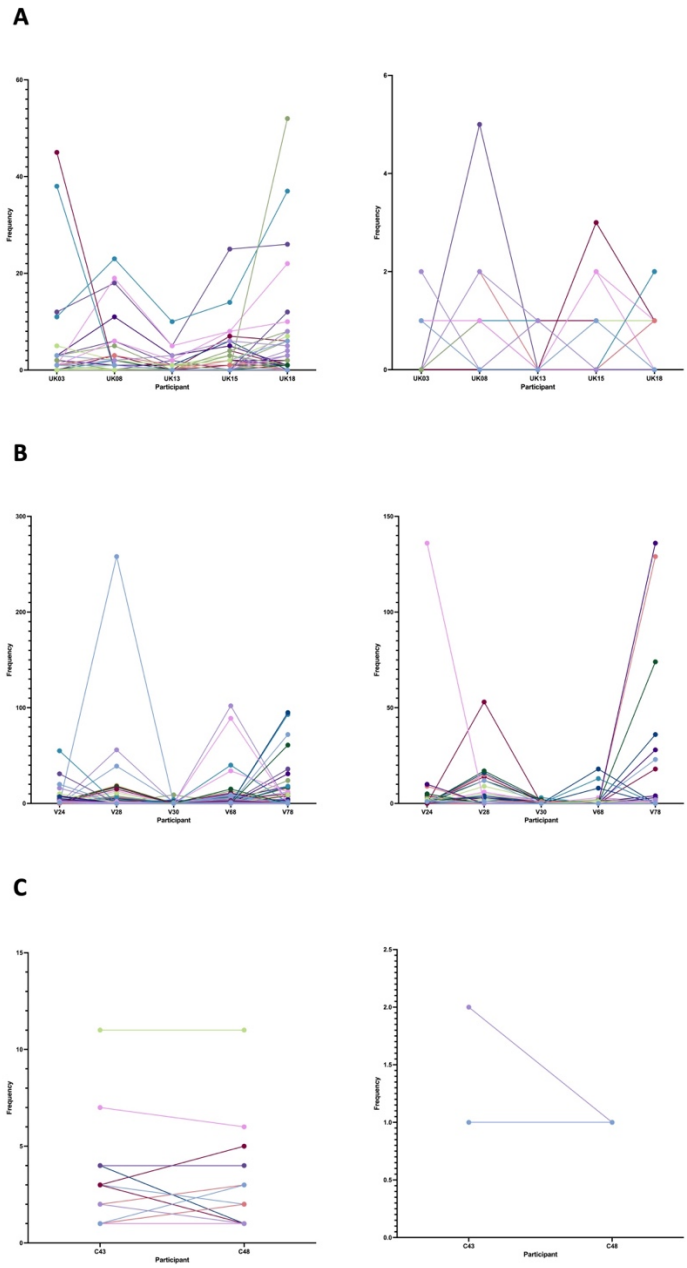


Figure 39 Plots presenting the TCRα and TCRβ CD8+ CDR3 sequences shared between at least two patients in A) Group A, early treated CARMA study participants, B) Group B, late treated LD study participants and C) Group C, healthy controls.

As described in the previous chapter, antigen "specificity" was explored using the VDJdb database[274]. Sequences shared between at least two patients within each group were selected for the investigation. As expected, several of the identified CDR3 sequences show "specificity" towards multiple epitopes. The majority appear to be specific towards pathogens like EBV, CMV and InfluenzaA. An example of "specificity" found in CD4+ CDR3 sequences in the three different groups can be seen below (Table 25). Eleven CDR3 sequences appeared to show potential "specificity" to HIV-1 in Group A, where only two sequences were matching in Group B. Sequences investigated in group C presented with "specificity" to InfluenzaA, EBV and CMV only.

Table 25 Example of antigen "specificity" as identified by the VDJdb database. Cross-reactivity with multiple pathogens was observed for the majority of the detected CDR3 sequences.

Group	CDR3 Sequence	Gene	Antigen Specificity/Match
A, B	CAAGGSQGNLIF	TRA	InfluenzaA, EBV, CMV
A	CAADSNYQLIW	TRA	EBV, CMV
A	CAASGGSYIPTF	TRA	EBV, CMV, YFV
A	CASSFSTDTQYF	TRB	InfluenzaA, CMV, EBV
A	CAAMDSNYQLIW	TRA	HIV-1, InfluenzaA, CMV, EBV
A, B	CASMDSNYQLIW	TRA	HIV-1, EBV, CMV
A	CASSLGYGYTF	TRB	HIV-1, CMV, EBV
A	CASSLRGTGELFF	TRB	HIV-1
A	CAVDSNYQLIW	TRA	HIV-1, InfluenzaA, CMV
A	CAVNGNKLVF	TRA	HIV-1, EBV, CMV
A	CAVNNARLMF	TRA	HIV-1, EBV, CMV
A	CAVRDSNYQLIW	TRA	HIV-1, EBV, CMV
A	CAVRGATNKLIF	TRA	HIV-1, CMV
A	CAVRMDSSYKLIF	TRA	HIV-1, EBV, CMV
A	CAVTRDDKIIF	TRA	HIV-1, EBV, CMV
B	CASSLTGNTEAFF	TRB	HIV-1, CMV
A, B, C	CAVEAAGNKLTF	TRA	EBV, CMV
A, C	CAVKAAGNKLTF	TRA	EBV, CMV

A, C	CALGGSQGNLIF	TRA	InfluenzaA

In a similar manner, shared CD8+ CDR3 TCRα and TCRβ sequences displayed specificity towards multiple antigens, predominantly CMV, EBV and InfluenzaA. Four HIV-1 "specific" sequences were found in Group A and five sequences in Group B. As with CD4+ CDR3 sequences, in Group C CMV, EBV and InfluenzaA matches were seen. Despite all Group C participants being confirmed as HIV-1 negative, one of the detected sequences (CAAMDSNYQLIW) has been shown to also recognise HIV-1 epitopes.

Table 26 Example of antigen "specificity" as identified by the VDJdb database. Cross-reactivity with various antigens was observed for numerous of the identified CDR3 sequences.

Group	CDR3 Sequence	Gene	Antigen Specificity/Match
A, B, C	CAAMDSNYQLIW	TRA	HIV-1, EBV, CMV, InfluenzaA
A	CAASANDYKLSF	TRA	EBV, CMV
A	CASSLGGSNQPQHF	TRB	EBV, CMV
A, B	CAFTSGTYKYIF	TRA	HIV-1, EBV, CMV
A, B	CASMDSNYQLIW	TRA	HIV-1, EBV, CMV
A, B	CAVRDSNYQLIW	TRA	HIV-1, EBV, CMV
A, B	CAALDSNYQLIW	TRA	EBV, CMV
B	CAASGGGSQGNLIF	TRA	EBV, InfluenzaA
B	CASSLGGYEQYF	TRB	EBV, InfluenzaA
B	CASSLPGELFF	TRB	HIV-1
A, B, C	CAVKDSNYQLIW	TRA	EBV, CMV
A, B, C	CAVLDSNYQLIW	TRA	EBV, CMV, InfluenzaA

Hamming distances were calculated for each of the three groups to investigate CDR3 sequence similarity. This measure of similarity looks at CDR3 sequences of similar length and reports the number of amino acid changes between them. Higher hamming distances imply the presence of more dissimilar CDR3 sequences. Even though, for CD4+ sequences, hamming

distances are detected at relatively similar levels within the different groups, some differences are apparent. As seen below in Figure 40, participant UK18 presents with increased frequencies of higher hamming distances for α chain CDR3 sequences compared to the other participants included in Group A. In Group B, the frequency distribution presents with a shift towards the left, with higher frequencies of lower hamming distances. A more apparent shift was noted for participant V68, both in α and β CDR3s. The observed frequency distribution for participant V30 presented with a slight shift towards the right (higher frequencies of higher hamming distances) in CDR3 β sequences. For the two participants included in the control group, Group C, hamming distances appear at similar frequencies for α CDR3 sequences, however, for β CDR3 sequences two opposite distribution patterns are observed. C43 presents with slightly increased frequencies of lower hamming distances whereas C48 presents with increased frequencies of higher hamming distances.

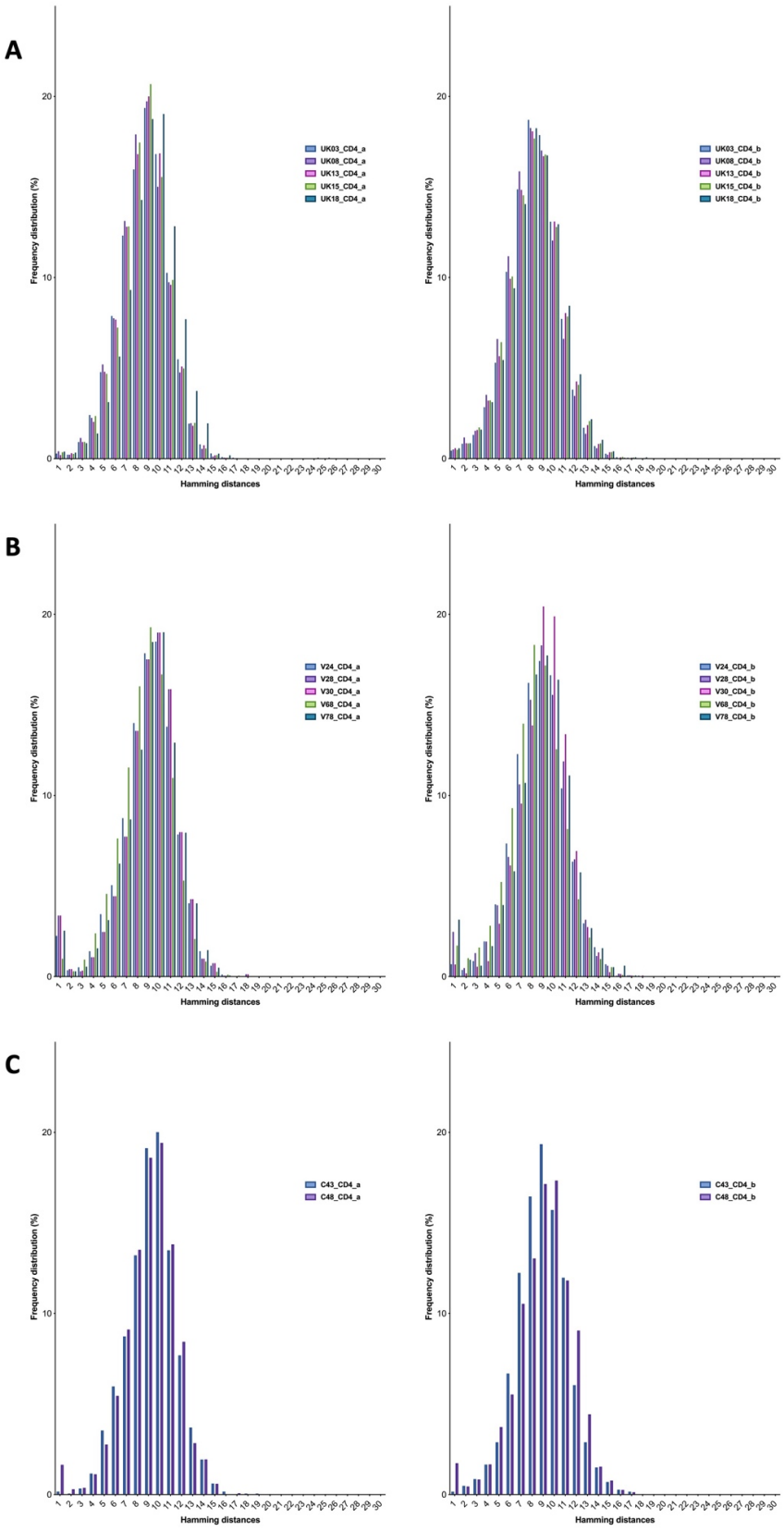


Figure 40 Frequency distribution of the calculated Hamming distances of CD4+ α and β CDR3 sequences between the participants within A) Group A, B) Group B and C) Group C. Hamming distances represent the number of amino acid differences between two sequences of the same length.

Similarly, when examining CD8+ CDR3 α and β sequences between the three groups, a few differences were observed. As with CD4+ CDR3 sequences, increased frequencies of lower hamming distances are observed in Group B, compared to Groups A and C. When looking closely at specific participants within the groups, UK13 (Group A) presented with higher frequencies of hamming distances equal to 6, 8, 10, 12 and 14 for α CDR3 sequences when compared to the others in the group. In Group B, increased frequencies of hamming distances equal to 8, 10 and 12 for α CDR3 sequences were noted for participants V24 and V28. For Group C, similar patterns in the frequency distribution were found, with a few exceptions as seen in Figure 41.

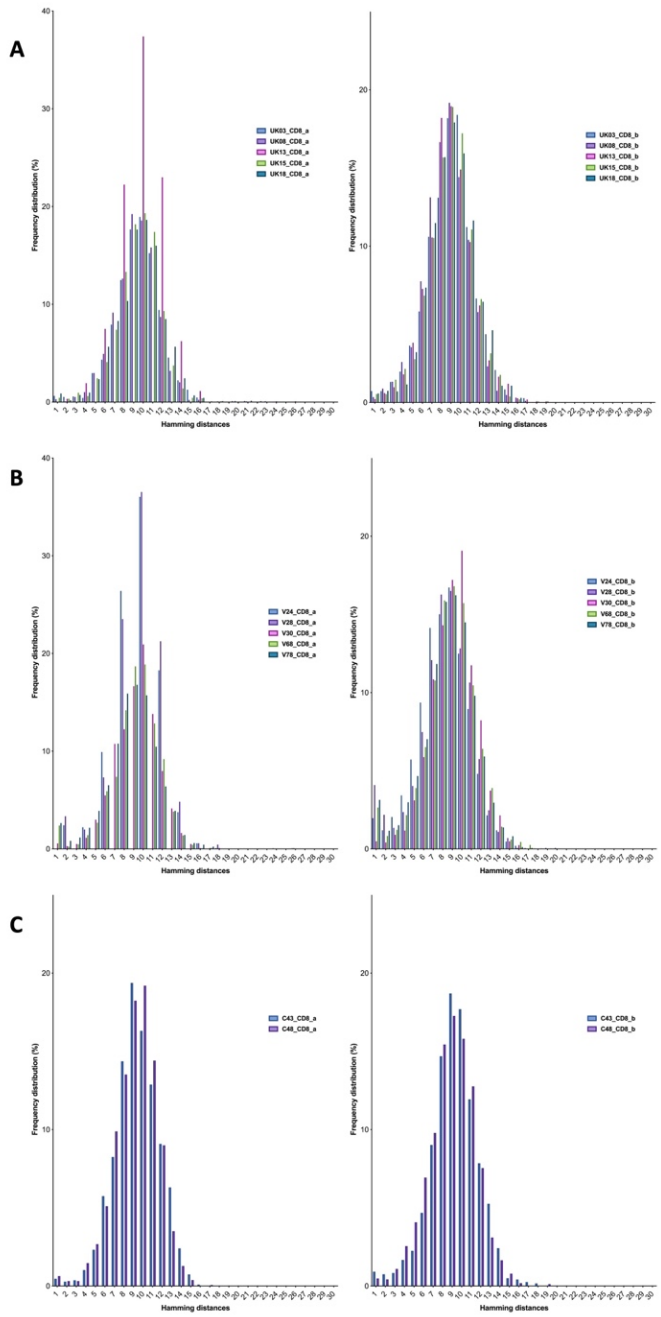


Figure 41 Frequency distribution of the calculated Hamming distances of CD8+ α and β CDR3 sequences between the participants within A) Group A, B) Group B and C) Group C. Hamming distances represent the number of amino acid differences between two sequences of the same length.

Similarity was also investigated by the generation of cluster plots for α chain CD4+ and CD8+ CDR3 sequences[206]. These plots display the presence of clusters of closely related CDR3 sequences, sharing amino acid triplets. CDR3 α chain sequencing data are used here to show

the presence of similar sequence clusters (Figure 42 and Figure 43). An increased number of CD4+ TCRα CDR3 clusters can be seen for participants UK03, UK08 and UK13.

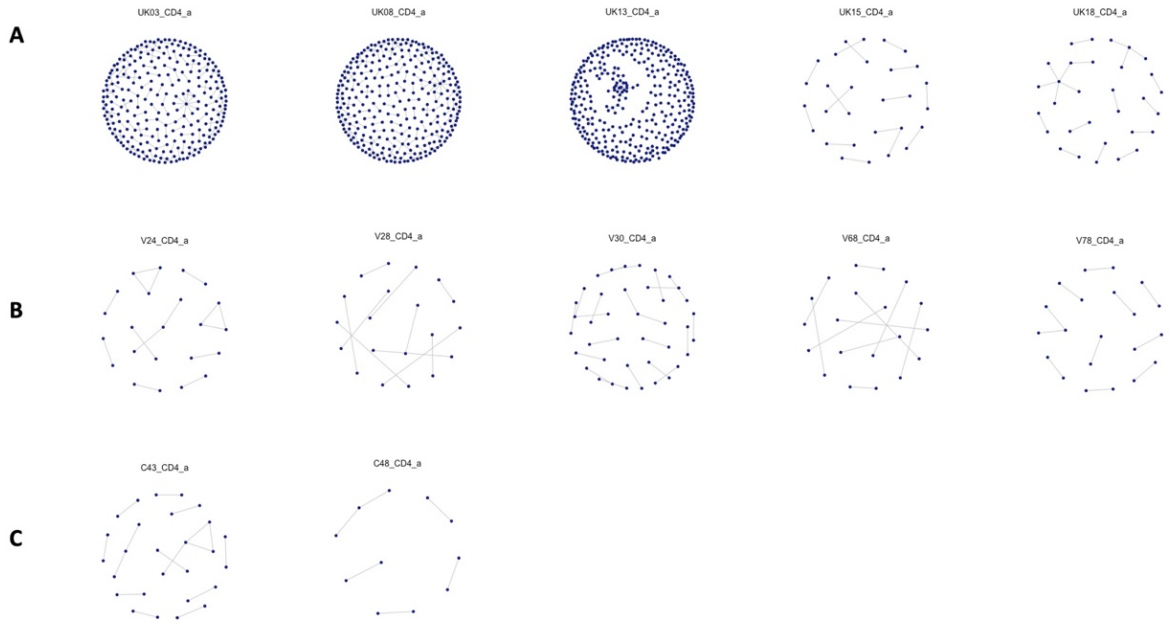


Figure 42 Clusters of closely related CD4+ TCRα CDR3 sequences identified in each participant in A) Group A, B) Group B and C) Group C.

Participants V68 and V78 appear to have more clusters of closely related CD8+ TCRα CDR3 sequences compared to all the other participants of the sub-study. Custers identified for CD4+ and CD8+ TCRβ sequences can be found in APPENDIX 7.

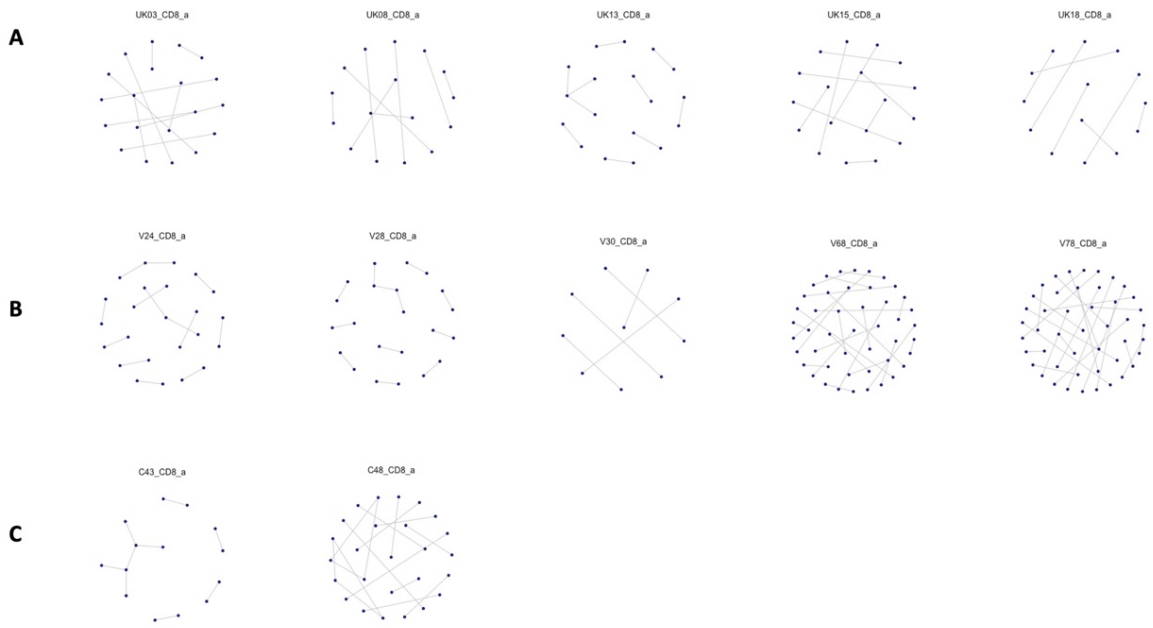


Figure 43 Clusters of closely related CD8+ TCRα CDR3 sequences identified in each participant in A) Group A, B) Group B and C) Group C.

Finally, the subsampled CDR3 sequences per participant were uploaded to the MAIT MATCH server (DTU) to detect the presence of any MAIT cell sequences. A similarity score of 1 implied a perfect match to a known MAIT cells sequence. A summary of the number and percentage of MAIT cell sequences found per participant can be seen below (Table 27).

Table 27 Number of MAIT cell matched sequences per sub-study participant, as identified by the MAIT MATCH server (DTU). Sequences presenting with a similarity score of one were considered a perfect match.

Participant	Cell Population	MAIT Cell Matched Sequences (n)	Percentage (%)
UK03	CD4	17	0.1
	CD8	20	1.0
UK08	CD4	36	0.2
	CD8	20	0.7
UK13	CD4	11	0.1

	CD8	18	0.7
UK15	CD4	19	0.1
	CD8	25	1.0
UK18	CD4	13	0.1
	CD8	22	1.0
V24	CD4	3	0.1
	CD8	8	0.1
V28	CD4	6	0.2
	CD8	13	0.4
V30	CD4	0	0.0
	CD8	9	0.3
V68	CD4	12	0.1
	CD8	19	0.3
V78	CD4	3	0.1
	CD8	16	0.3
C43	CD4	6	0.2
	CD8	15	0.7
C48	CD4	3	0.1
	CD8	17	0.5

The detected CD4+ and CD8+ MAIT cell sequences were in turn cross-checked to explore any sharing between the participants of each of the three sub-study groups (Table 28 and Table 29).

Table 28 MAIT cell matched sequences identified within the CD4+ samples of all 12 sub-study participants, using the MAIT MATCH server (DTU).

MAIT Cell Sequence	Participant
CAASDSNYQLIW	UK03, UK08, UK15, C48
CAFMDSNYQLIW	UK03, UK08, UK13

CAPMDSNYQLIW	UK03, UK08
CAVGADDYKLSF	UK03
CAVGDSNYQLIW	UK03, UK08, UK15, V68
CAVKDSNYQLIW	UK03, UK08, UK18, V68
CAVLDSNYQLIW	UK03, UK08, UK15, UK18, C43
CAVLNTGGFKTIF	UK03, UK15, V68, C43
CAVMDSNYQLIW	UK03, UK08, UK13, UK18, V28, C48
CAVMDSSYKLIF	UK03, UK08, UK13, UK15, UK18, V68
CAVPNSGGSNYKLTF	UK03, UK15
CAVRDSSYKLIF	UK03
CAVRDTGFQKLVF	UK03
CAVRGNFNKFYF	UK03, UK13, UK15
CAVRNSNYQLIW	UK03, UK15
CAVSDSNYQLIW	UK03, UK08, UK15, UK18
CAVSDSSYKLIF	UK03, UK15
CAALDSNYQLIW	UK08, V28
CAAMDSNYQLIW	UK08, UK18
CAASKAAGNKLTF	UK08, UK13, UK15, V68
CAAVDSNYQLIW	UK08
CAGLDSNYQLIW	UK08, UK18
CAGMDSNYQLIW	UK08, UK15, UK18, V28, V78
CAGWDSNYQLIW	UK08
CAIMDSNYQLIW	UK08
CALLDSNYQLIW	UK08
CALNDMRF	UK08
CASMDSNYQLIW	UK08, UK18, V24, V78, C43
CASMDSSYKLIF	UK08
CASRSYNTDKLIF	UK08
CATMDSNYQLIW	UK08
CAVDDNTDKLIF	UK08, UK13
CAVFSDGQKLLF	UK08

CAVIDSNYQLIW	UK08, C43
CAVLDSSYKLIF	UK08
CAVLSNDYKLSF	UK08, V78
CAVQDSNYQLIW	UK08
CAVRDGDYKLSF	UK08, UK18, V28, C43
CAVRDNDYKLSF	UK08
CAVRDRDYKLSF	UK08
CAVRDSDYKLSF	UK08
CAVRDSNYQLIW	UK08, UK15, UK18
CAVTDSNYQLIW	UK08, UK15, UK18, V28, C43
CAVVDSNYQLIW	UK08, UK13, UK15, UK18
CAVGYSSASKIIF	UK13
CAVPNYNQGGKLIF	UK13
CAVRDNFNKFYF	UK13, V68
CAVRRDDKIIF	UK13, V68
CAAIDSNYQLIW	UK15
CAPSGSARQLTF	UK15
CAVNNNNDMRF	UK15
CAVPSGGSYIPTF	UK15, V24, V68
CVVKDSGYALNF	V24
CAVRESNYQLIW	V28, C48
CAALSNDYKLSF	V68
CAAPTGRRALTF	V68
CAVRVDRGSTLGRLYF	V68
CAVYDSNYQLIW	V68

Table 29 MAIT cell matched sequences identified within the CD8+ samples of all 12 sub-study participants, using the MAIT MATCH server (DTU).

MAIT Cell Sequence	Participant
CAALDSNYQLIW	UK03, UK08, UK13, UK15, UK18, V28, V30, V68, C48
CAAMDSNYQLIW	UK03, UK08, UK13, UK15, UK18, V28, V30, V68, V78, C43, C48
CAASDSNYQLIW	UK03, UK08, UK13, UK15
CAGMDSNYQLIW	UK03, UK08, UK13, UK15, UK18, V24, C43, C48
CAPMDSNYQLIW	UK03, UK15, V28, V68, C48
CASMDSNYQLIW	UK03, UK13, UK18, V28, V68, V78, C43
CATMDSNYQLIW	UK03
CAVGDSNYQLIW	UK03, UK18, V28, V30, V68
CAVKDSNYQLIW	UK03, UK08, UK13, UK15, UK18, V30, V68, V78, C43, C48
CAVLDSNYQLIW	UK03, UK08, UK13, UK15, UK18, V24, V30, V68, V78, C43, C48
CAVLSNDYKLSF	UK03, V68
CAVMDSNYQLIW	UK03, UK08, UK13, UK15, UK18, V24, V28, V30, V68, V78, C43, C48
CAVMDSSYKLIF	UK03, UK08, UK13, UK18
CAVPSGGSYIPTF	UK03
CAVRDGDYKLSF	UK03, UK08, UK13, UK18, C43, C48
CAVRDNDYKLSF	UK03
CAVRDSDYKLSF	UK03, UK08, V28, C48
CAVRDSNYQLIW	UK03, UK08, UK13, UK15, UK18, V28, V68, C43
CAVSDSNYQLIW	UK03, UK08, UK15, UK18, V30, V68, V78, C43, C48
CAVTDSNYQLIW	UK03, UK08, UK13, UK15, UK18, V28, V68, V78, C43, C48
CAGLDSNYQLIW	UK08, UK15, UK18, V24, C43
CAPLDSNYQLIW	UK08, UK13, UK15, UK18, V78
CAVGYSSASKIIF	UK08, V68
CAVIDSNYQLIW	UK08, UK15, UK18, V24, V68, V78, C43, C48
CAVLDSSYKLIF	UK08, UK15
CAVSDSSYKLIF	UK08, UK15
CAVVDSNYQLIW	UK08, UK15, UK18, V28, V30, V68, C48
CAAVDSNYQLIW	UK13, UK15, C43

CALLDSNYQLIW	UK13, V78, C43
CARMDSNYQLIW	UK13
CAVQDSNYQLIW	UK13, UK15, UK18, V78
CAVRDRQAGTALIF	UK13, V78
CAFMDSNYQLIW	UK15, C48
CAVRDGNYQLIW	UK15
CAVRDRDYKLSF	UK15, V68
CAVRDRDYQLIW	UK15
CAVRRRDDKIIF	UK15, V78
CVPMDSNYQLIW	UK15
CAAEDSNYQLIW	UK18
CAIMDSNYQLIW	UK18, C43, C48
CAVFSDGQKLLF	UK18
CAVRGGDYKLSF	UK18
CAALSDYKLSF	V24
CAVGADDYKLSF	V24
CAVSGDYKLSF	V24
CALNDMRF	V28, V78
CAVRDNFNKFYF	V28
CAVRDSSYKLIF	V28
CAALSNDYKLSF	V30
CAAFDSNYQLIW	V68
CAVIQGDYKLSF	V68
CAVLNTGGFKTIF	V68
CAAIDSNYQLIW	V78
CAVEEDNYGQNFVF	V78
CAPSGSARQLTF	C48
CAVRESNYQLIW	C48

5.5 Discussion

To this day, studies investigating the impact of ART initiation timing to the TCR repertoire, especially in paediatric settings, have been very limited. Significant gaps in our knowledge of how early or late ART initiation affects the TCR repertoire diversity remain.

In this chapter, a cohort, comprising of 5 early-treated and 5-late treated paediatric patients as well as 2 healthy individuals, was studied to investigate the effects of ART timing on the TCR repertoire. For comparison purposes, an effort was made to select relatively age-matched participants for each group.

As it has been previously reported, age plays an important role on thymic output. As we age, the levels of thymic output progressively decrease, leading to diminished levels during adulthood[275]. Here, thymic output was evaluated by calculating TRECs per million CD4+ T-cells and comparing the results to reference levels kindly provided by the clinical lab at GOSH. Depending on their age, all 12 participants were sorted into either the 12-18 years or the adult (over 18 years of age) category. Even though most participants presented with normal TRECs per million CD4+ levels for their age, no definitive conclusions can be drawn since no reference data were available for the 10th centile of the 12-18 years age group. It is worth noting that all participants in Group B, apart from participant V78, presented with lower TRECs levels when compared to participants in Group B. This could be an indication that the timing of ART initiation can affect thymic function and output, however the number of participants in each group is too low to be able to draw any informative conclusions.

TCR repertoire diversity was investigated by using the Gini coefficient, a measure of inequality[14,156,203]. A Gini coefficient that tends towards zero dictates the presence of a uniform population where all TCR sequences are found to be equally present. A Gini coefficient that tends towards one, represents a highly inequal population with increased oligoclonality. However, this measure can be severely affected by the number of CDR3 sequences available per patient. To make the samples more “comparable”, all sequencing data were subsampled down to the same levels depending on the amount of raw data

available (11252 or 2021 reads for CD4+ sequencing data, and 12606 or 1995 reads for CD8+ sequencing data). A relative uniformity with minimal differences in the levels of the Gini coefficient (ranging between 0.2-0.3) was observed between all three participant groups, early treated (Group A), late-treated (Group B) and healthy controls (Group C). This could be indicative of a potentially biased subsampling method, that tends to remove/reduce the frequencies of most of the clonal expansions present in the samples, leading to uniform TCR populations with all CDR3 sequences present at equal frequencies, tending towards a relatively diverse repertoire.

The subsampled sequencing data were also used to explore sequence sharing between the individuals in the three sub-study groups. Research has previously shown that CDR3 sequence sharing decreases between HIV+ individuals[156]. For the CD4+ TCRα and TCRβ populations, 6.2% and 1% of the sequences were shared between at least two individuals in Group A, 2.5% and 0.4% shared in Group B and 0.7% and 0.1% shared in Group C, respectively. For the CD8+ TCRα and TCRβ populations, 2.1% and 0.5% were shared between at least two individuals in Group A, 2.4% and 0.6% shared in Group B, 1.1% and 0.1% shared in Group C. The unexpected lack of public sequence sharing found in the healthy controls group (Group C) could potentially be attributed to the lack of available data, due to the lower sequencing read numbers and the fact that only two individuals were included.

Hamming distances were calculated to investigate the presence and frequency of closely related CDR3 sequences. A Hamming distance of 1 represents highly similar sequences, with only one amino acid change required for two same-length sequences to become identical. Higher Hamming distance measurements dictate the presence of a more diverse TCR repertoire[14,205]. Each of the three sub-study groups were investigated for sequence similarity, within both CD4+ and CD8+ TCR α and TCR β populations. No major differences were found in the patterns of hamming distances observed in CD4+ and CD8+ TCR populations, between Groups A and C. However, Group B presented with slightly higher frequencies of lower hamming distances, potentially indicating that delayed treatment initiation could lead to a less diverse TCR repertoire. This finding, compliments work previously published by our group, indicating that suboptimal HIV-1 infection control can lead to reduced TCR repertoire diversity[14].

The presence of closely related CDR3 clusters was used as a second measure of sequence similarity. For the CD4+ TCRα population most participants in Group A (60%) presented with a much higher number of CDR3 clusters when compared with participants in Group B and Group C. For the CD8+ TCRα population, 40% of the Group B participants had higher numbers of CDR3 clusters when compared to the participants in Groups A and C. However, the differences in related CDR3 sequence clusters might be skewed due to the available subsampled data per participant. Generally, according to previously published studies, healthy individuals as well as early treated-well suppressed HIV-1+ patients (like in Group A) are expected to have a higher number of closely related CDR3 clusters present. However, due to the low number of participants in each group, and the limitations of the sequencing pipeline and data analysis, the true effect of ART initiation timing on the CDR3 clusters and TCR diversity might be obscured.

As before, the public VDJdb database was used to explore possible antigen "specificities"[274]. For this purpose, shared sequences between the participants in the three different groups were used. Most of the generated CDR3 sequences appeared to be reactive towards common pathogens such as EBV, CMV and InfluenzaA. For the CD4+ population, only a small fraction of frequencies appeared to be specific towards HIV-1 (11 in Group A and 2 in Group B). Again, for the CD8+ population, a limited number of sequences presented with HIV-1 specificity (4 in Group A and 5 in Group B). However, one of the reported as HIV-1 specific sequences (CAAMDSNYQLIW) was also detected in the healthy control group (Group C). This finding supports previous reports of cross-reactivity amongst the TCR repertoire, with TCRs potentially binding to more than one antigen[203,276]. For that reason, the term "specificity" should be used with caution when interpreting TCR data. Another reason why the term "specificity" should be carefully used is the inability to pair α and β chain sequences here, since no definitive conclusions can be drawn by using data derived from only one chain.

Finally, the generated subsampled data were searched to detect any sequences associated with the MAIT cell population. In a healthy population, MAIT cells constitute approximately 10% of the CD8+ pool[277]. In adult studies, this population appears to be irreversibly depleted, potentially increasing the susceptibility to certain opportunistic infections[156,277]. Within the

data presented in this chapter, a small percentage could be attributed to MAIT cell sequences. As expected, out of the identified MAIT cell sequences the majority were CD8+ and only a very small fraction was CD4+, being in accordance with previously published data reporting that they predominantly express the CD8 co-receptor. Groups A and C, presented with slightly higher percentages of CD8+ MAIT cells when compared to Group B. This could be an indicator of the potential negative effect and number depletion that late ART initiation can have in the MAIT cell population of HIV-1+ infected individuals, when compared to early ART initiation.

5.5.1 Conclusions

This chapter explored the effects of ART initiation timing on the TCR repertoire of children/adolescents. This was achieved by using and comparing CD4+ and CD8+ T-cell population sequencing data, deriving from 5 early-treated, 5-late treated HIV-1+ individuals and 2 healthy controls.

However, this sub-study presents with a major limitation due to the small number of subjects included in its design, especially when it comes to healthy controls. This limitation, combined with the differences in sequencing read depth seen per participant, prevents any statistically significant conclusions from being drawn.

Nonetheless, a few differences were observed in the late-treated group (Group B), such as decreased CD4+ sequence sharing, increased frequencies of low Hamming distances, decreased numbers of CD4+ CDR3 clusters and lower percentages of CD8+ MAIT cell sequences. All these could be indicators of how delayed treatment initiation could lead to a less diverse TCR repertoire. In the future, studies including bigger cohorts, that would also include higher numbers of age-matched healthy controls could help expand our knowledge by further elucidating the TCR repertoire landscape.

CHAPTER 6: IMMUNOLOGICAL CHARACTERISTICS OF HIV-1 PERINATALLY INFECTED INDIVIDUALS- IMPACT OF PLANNED TREATMENT INTERRUPTION ON THE TCR REPERTOIRE OF THE CD4+ MEMORY T-CELL POPULATION

6.1 Background and Objectives

The use of ART and the advances made in the available treatment options have led to a decrease in morbidity and mortality rates in HIV patients[278,279]. However, we are still unable to achieve a sterilising cure and continuous ART can, in some cases, be ineffective and/or cause serious side-effects[280–282]. For this reason, planned treatment interruption (PTI) strategies have been studied in both adult and paediatric cohorts[231,283–288]. These studies mainly aimed to explore the presence of immune responses against HIV following viremia suppression, as well as the life quality improvement and minimisation of viral resistance development[289]. So far data, especially in adult studies, have shown that continuous treatment remains far superior to treatment interruption strategies[283,290,291]. For most of the adult studies, no benefits were observed after treatment interruption. On the contrary, individuals enrolled in treatment interruption studies, such as the Strategies for Management of Antiretroviral Therapy (SMART) study, presented with an increase in opportunistic infections and deaths[283,291,292]. For this reason, interruption studies are not recommended for virologically controlled adult patients.

Nonetheless, paediatric studies have shown different and cautiously promising results. As we know, children have a greater potential of immune recovery since they have an immature immune system and greater thymic output[12,13,293,294]. The first paediatric PTI study, the Paediatric European Network for Treatment of AIDS (PENTA) 11 trial was completed in 2008[295]. This trial demonstrated that, if well-planned and well-monitored, PTIs might still be a viable option for children. Although PTI cannot be recommended as a standalone therapeutic option, it could potentially be combined with other novel therapeutic strategies, such as the use of broadly neutralising antibodies. In this chapter, the potential impact of PTI compared to continuous treatment (CT) on the TCR repertoire diversity of CD4+ memory T-cells was explored.

6.2 Introduction

The PENTA11 trial aimed to determine whether PTI could be a safe strategy for children[296]. In contrast to previously described adult studies, it was hypothesized that the response to PTI might differ in perinatally infected children, due to their developing immune system and early ART initiation[181]. It included 109 participants between the ages of 2 and 15 years that had been on ART for more than 24 weeks, had undetectable viral load (below 50 copies/ml) and a good CD4+ count (depending on their age)[296]. This randomised, CD4-guided phase II trial compared two groups of patients, one group where ART was taken consistently (CT) as per current standard of care and one group where PTI was introduced. After trial completion patients re-initiated ART and were monitored for a period of two years to determine any possible consequences of PTI. At follow-up very few or no adverse clinical outcomes, virological or immunological, were observed[180,297]. Recently, in a PENTA-11 sub-study published by our group, investigated changes in the thymic output and TCR repertoire of CD4+ naïve and CD8+ memory T-cells[204]. Despite the limited number of patients that were included in this sub-study, increased thymic output of CD8+ T-cells, increased TCR clonal expansions in the naïve CD4+ T-cell population and a sharp increase in TCR sharing following the PTI protocol initiation were some of the common denominators observed in all selected patients of the PTI arm[204].

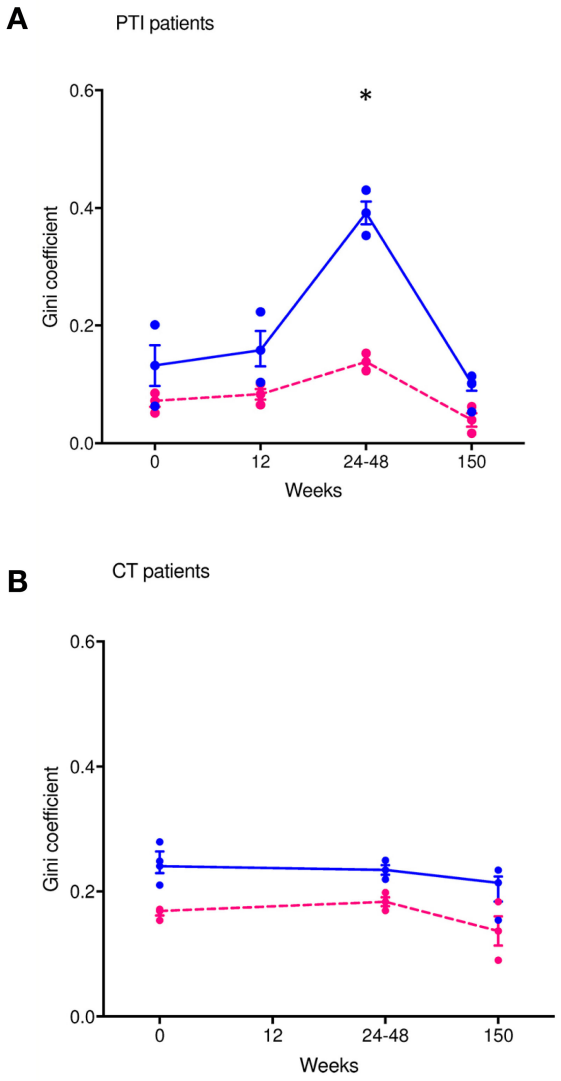


Figure 44 Changes in the calculated Gini coefficient for α (blue) and β (pink) chains in the CD4+ naïve T-cell population during A) planned treatment interruption and B) continuous treatment. Increase in the number of clonal expansions was observed during weeks 24-48 in the PTI arm, returning though to pre PTI levels 3 years after ART re-initiation. No changes in the TCR repertoire diversity were observed in the CT arm of the study[204].

This chapter comes as a continuation of our group's research on the immunological characteristics of PENTA11 patients. Its aim was to explore the impact of planned treatment interruption in another T-cell population, the memory CD4+ T-cells, especially since this it is one of the main cell targets of HIV-1 infections and it is severely affected by disease

progression and ART treatment[298,299]. For that purpose, as in previous chapters of this thesis, next generation sequencing techniques were used to obtain data on both α and β chains and to investigate TCR repertoire diversity, CDR3 sequence sharing between timepoints for the two different treatment arms of this study and determine potential antigen specificity.

6.3 Material and Methods

The sub-study presented in this chapter was designed as a continuation to our research groups PENTA11 studies. The methods used to complete the experiments described below are described in detail in Chapter 2. To summarise, a small number of patients, both in the CT and PTI PENTA11 trial arms, were selected to investigate the TCR repertoire in their CD4+ memory T-cells and explore the potential differences that may arise from the different treatment strategies.

As part PENTA11 sub-study (main trial: ISRCTN 36694210), extracted RNA samples from sorted memory subpopulation of CD4+ T-cells (stored at -80˚C) collected from nine patients, were selected to investigate the T-cell receptor repertoire longitudinally. The median age and CD4+ counts at the start and end of the study are presented below (Table 30). The selected samples included 4 different timepoints, weeks 0/2, 12/24, 48, 150, for 4 patients under PTI and 2 timepoints, weeks 0/2 and 150, for 5 patients on CT. Samples from week 0/2 signify the study baseline when ART was either stopped or continued. Weeks 12/24 indicate a timepoint during which treatment was ceased for the PTI arm. Week 48 is the timepoint right before ART re-initiation and week 150 is the long-term follow-up timepoint, 3 years following the study end. In summary, these RNA templates were used to generate libraries[202] that were sequenced using the Illumina MiSeq system and the MiSeq Reagent Kit V2 (500-cycles).

Table 30 Characteristics of the 9 PENTA11 participants selected for this sub study (adapted from [181]). Unless otherwise stated, continuous data are presented as median (IQR).

All (n=9)	PTI Group (n=4)	CT Group (n=5)
Age at first timepoint, years	9.9 (8.9-10.8)	9.6 (5.3-10.1)
Age at final timepoint, years	15.6 (14.8-15.9)	14.1 (11.8-15.3)
CD4+ count at first timepoint, cells/µl	969.5 (870.0-1091.8)	1360.0 (930.0-1370.0)
CD4+ count at final timepoint, cells/µl	935.5 (843.6-960.0)	967.0 (646.0-1030.0)

The generated FASTQ files were demultiplexed and error-corrected (through unique molecular identifiers -UMIs) by using an automated analysis pipeline that incorporated Decombinator Python script (https://github.com/innate2adaptive/Decombinator)[202]. Analysis including subsampling and calculating the Gini coefficient was produced using in-house R scripts (R, version 3.6.3) and the output from the Decombinator pipeline[203]. The excel (Microsoft, version 16.65) "MATCH" function was used to determine shared clonotypes between patients. VDJdb (version "SearchTable-2021-05-20 23_12_40.221") and an in-house R script (R, version 3.6.3) were used to determine CDR3 specificity[203,204,243]. CDR3 sequence similarity was explored by calculating Hamming distances and by searching for clusters of related TCR sequences[14,204,206]. MAIT sequence similarity was determined by using the MAIT Match Server database (DTU Bioinformatics). Finally, plots were generated using R packages (ggplot2, version 3.3.5) and the Prism software (GraphPad, version 9.0.0 (as described in section 2.7.2).

6.4 Results

From the 9 PENTA11 patients included in this sub-study, sequencing was only successful for 6, possibly due to poor RNA sample quality and concentration. Samples from three participants had to be excluded due to failure to pass my sequencing and analysis requirements (Table 31). More specifically, one patient included in the PTI arm, D, was excluded since insufficient numbers of sequencing reads for all timepoints were obtained. Two patients from the CTI arm, H, and I, were also excluded due to low or no available number of sequencing reads.

Table 31 Read depth achieved per patient (for each timepoint and each TCR chain) using the Illumina MiSeq system. Three patients had to be excluded from the study due to obtaining insufficient or no number of reads. Two samples from patient G (highlighted in red) could not be subsampled due to their sequencing depth.

Sample ID	Treatment Arm	Timepoint (Week)	TCR Chain	MiSeq Number of Reads
A	PTI	0	alpha	3744
			beta	5017
		12	alpha	561
			beta	1124
		48	alpha	1402
			beta	2546
		150	alpha	3430
			beta	6702
B	PTI	0	alpha	8693
			beta	33193
		24	alpha	12531
			beta	41200
		48	alpha	10420
			beta	35218
		150	alpha	26092

			beta	60770
C	PTI	0	alpha	2778
			beta	7963
		12	alpha	5294
			beta	7357
		48	alpha	941
			beta	8750
		150	alpha	11262
			beta	19999
D	PTI	0	alpha	Low reads
			beta	Low reads
		12	alpha	Low reads
			beta	Low reads
		48	alpha	Low reads
			beta	Low reads
		150	alpha	Low reads
			beta	Low reads
E	CT	2	alpha	1259
			beta	15887
		150	alpha	16212
			beta	19383
F	CT	0	alpha	4887
			beta	28299
		150	alpha	1390
			beta	12892
G	CT	0	alpha	367
			beta	283
		150	alpha	3227
			beta	3654
H	CT	2	alpha	Low reads
			beta	Low reads

		150	alpha	Low reads
			beta	Low reads
I	CT	0	alpha	No reads
			beta	No reads
		150	alpha	No reads
			beta	No reads

As in chapters 4 and 5, sequencing data deriving from this study's subjects were subsampled to allow for comparisons using an in-house algorithm developed to calculate the minimum representative subsampling depth per sample[203]. As discussed previously in this thesis (section 4.4, Chapter 4), this algorithm presents some limitations as it excludes samples that present much higher or much lower initial read numbers leading to a separate subsampling analysis. For that reason, even though most of the samples were subsampled down to 2544 reads, samples that had a high number of reads were subsampled down to 19999 reads and samples that had a low read number available (below 1500 raw reads) were subsampled down to 557 reads. For timepoint W0 for patient G both α and β chains a low amount of raw sequencing data, 367 and 283 reads respectively, was obtained, making it impossible to include to any of the subsampling categories mentioned above. Consequently, results presented for this timepoint for patient G derive from the raw data. As previously discussed, due to the differences in subsampling levels, results should be interpreted with caution.

TCR repertoire diversity was calculated using a measure of inequality, the Gini coefficient[203]. As seen in Figure 45 below, for both CD4+ memory α and β chain samples the Gini coefficient presented similar levels, meaning the TCR repertoire diversity of the selected patients is relatively normal with minimal clonal expansions present. When looking at the Gini coefficient, PTI does not appear to cause any adverse effects in the repertoire diversity of the CD4+ memory T-cell compartment.

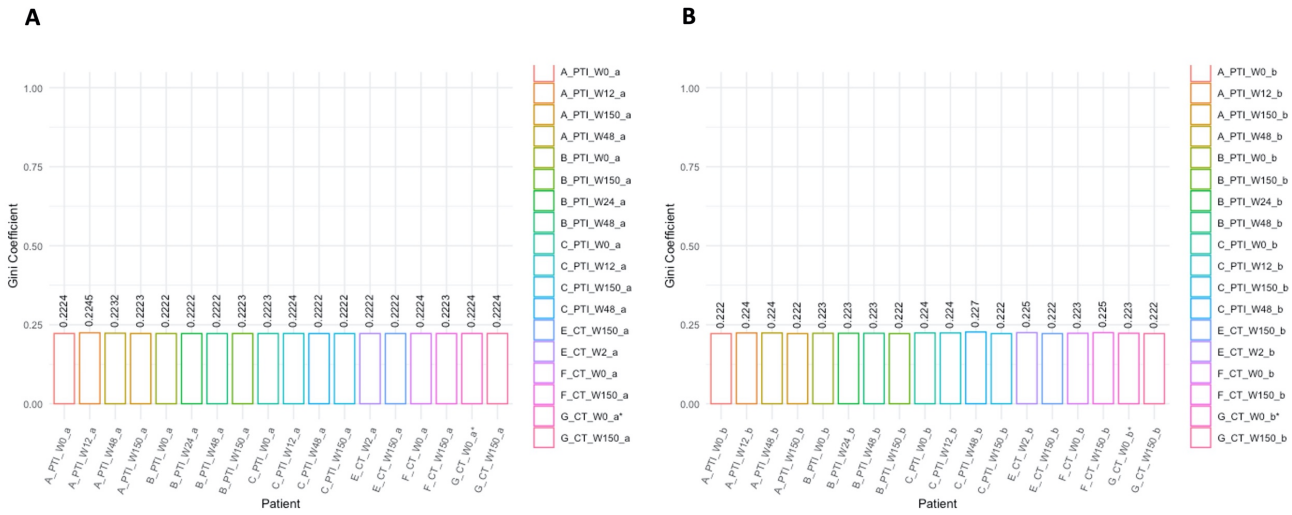


Figure 45 Subsampled PENTA11 sequencing data (apart from data presented for patient G at the W0 timepoint) were used to calculate the Gini coefficient and investigate the effect of planned treatment interruption on the diversity of the TCR repertoire in the CD4+ memory T-cell population. As seen in these histograms, the Gini coefficient for both A) α and B) β chains remains at similar levels in either study arm.

Using Excel (Microsoft, version 16.65) the sequence sharing in the PENTA11 cohort between different timepoints was investigated. The aim was to determine whether the two different treatment strategies that were followed during the PENTA11 study, PTI or CT, have any effects on the clonotypes and their frequencies that are present in the patients. Both α and β chain CDR3 sequences were investigated for sharing through the different timepoints.

For the three patients included in the PTI group, most of the identified CDR3 sequences appeared to be unique to each different timepoint. Only a small percentage of sequences (<10%) appear to be shared between at least two of the timepoints for each patient. More specifically, for patient A, a 3.72% of CDR3 α and 3.42% of CDR3 β shared sequences was identified. For patient B, only 5.89% and 9% of the CDR3 α and β sequences were shared. For patient C, 4.25% CDR3 α and 6.89% CDR3 β sequences were shared between timepoints. As seen in Figure 46, when looking at α chain CDR3 sequences, a drop in frequencies and sequence sharing was noticeable between W0 and W12 of the study. Several sequences appeared to expand again at W48 and/or at the 3-year follow up point (W150). Noticeably,

for participant A, frequencies of several β chain CDR3 sequencies expanded at W48 and dropped again at W150. For PTI participant B, more CDR3 β chain sequences were available, since high numbers of raw sequencing reads that were subsampled at a higher level (19999 reads) was obtained. A number of these CDR3 sequences seemed to obtain similar frequency levels through the different timepoints under investigation. An example of the top five expanded CDR3 sequences found per PTI patient present at W0 and their changing frequencies throughout timepoints are presented below (Table 32).

Table 32 Top five expanded CDR3 sequences in PTI participants, per chain (α and β), and their frequencies at each study timepoint (W0, W12/W24, W48 and W150).

Sample ID	Chain	CDR3 Sequence	Frequency (Read Numbers)			
			W0	W12/W24	W48	W150
A	α	CAVSEGTGNQFYF	38	20	11	19
		CVVNPTGRRALTF	27	8	9	4
		CVVIRTGANNLFF	24	9	14	18
		CAARDVGNFNKFYF	16	10	3	14
		CVVNPNSGYSTLTF	15	14	4	0
	β	CSVLDRSDSPLHF	49	21	79	13
		CASSLGQAGVADTQYF	23	7	19	0
		CASSLASGETAEQYF	19	4	43	10
		CATSDMGAGGYNEQFF	18	13	55	12
		CSAKTPGRYYEQYF	16	7	78	9
B	α	CALGEHSGGGADGLTF	16	5	3	0
		CAVSAGGGADGLTF	8	2	3	7
		CADMDSNYQLIW	7	3	1	15
		CAVNMKSASKIIF	7	1	1	22
		CAARFSGGGADGLTF	5	1	0	1
	β	CASSLTTASSVSYEQYF	101	89	57	0
		CSVGVTRPDNEQFF	92	62	70	52
		CASSLRQAGGPDTQYF	78	56	58	55

C	α	CASSVTGTSGQETQYF	66	54	64	38
		CSASDPGAFF	63	62	60	47
		CAASSTGRRALTF	170	205	73	39
		CALSGGYNKLIF	47	52	16	23
		CAVSDWGAIGFGNVLHC	38	60	14	23
		CVVRLGGGGADGLTF	23	14	1	1
		CIVRVNTGGFKTIF	13	9	3	3
	β	CAISDRQGQDTQYF	46	93	19	194
		CSVEVVHGDTQYF	20	8	8	59
		CSGKRTNTEAFF	19	10	18	45
		CASSSSTGYGYTF	15	8	6	34
		CASSTGNTDTQYF	15	8	7	41

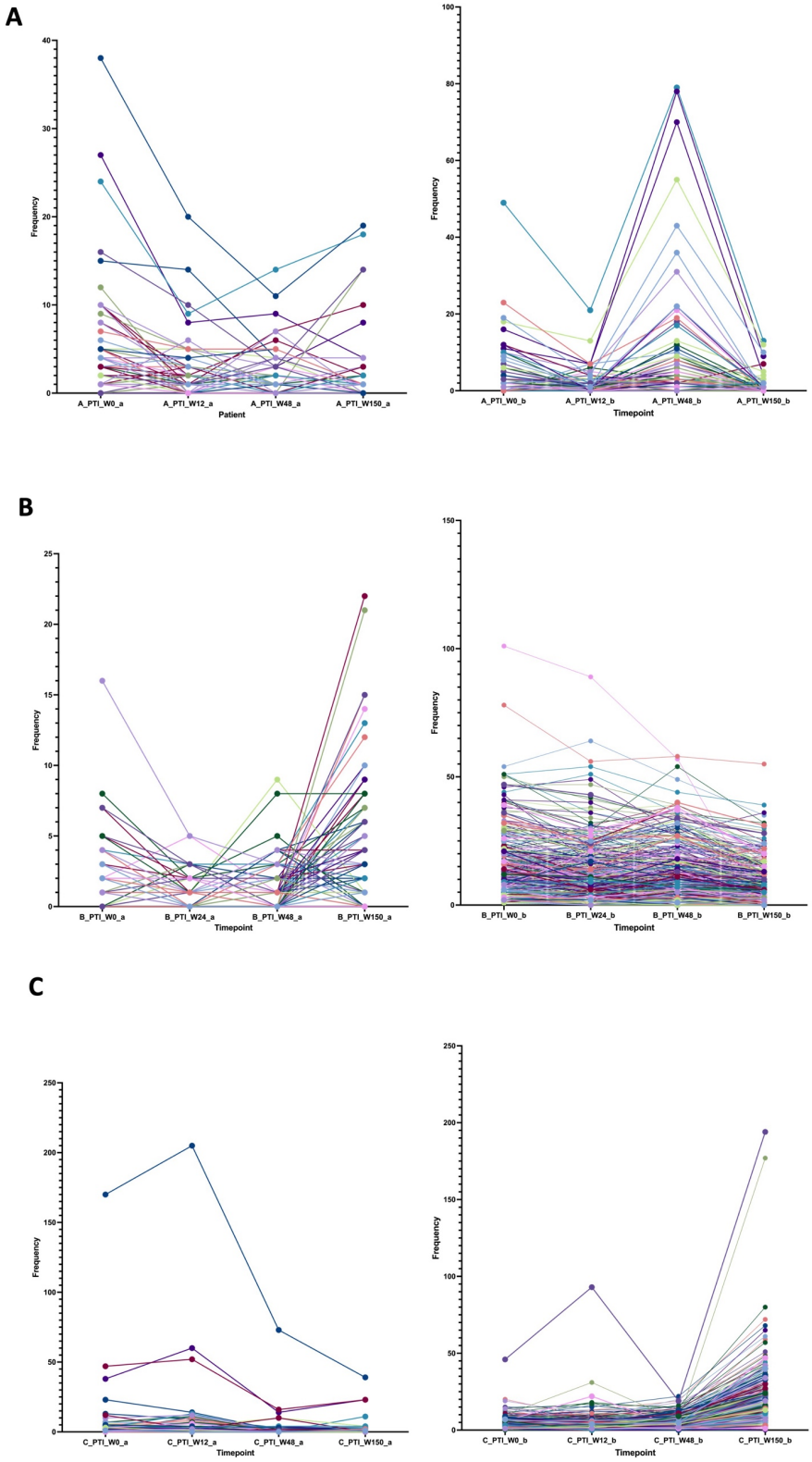


Figure 46 Plots presenting the TCRα and TCRβ sequences shared between the different PENTA11 study timepoints (W0, W12/W24, W48 and W150) in the three patients receiving planned treatment interruption.

For patients receiving continuous treatment, the majority of the shared CDR3 sequences present with similar abundances between the two study timepoints (Figure 47). A slight decrease in frequencies in the majority of CDR3 sequences at W150 can be seen in patients E, F and G. For participant E, 1.47% CDR3 α sequences were shared between the two timepoints (W0 and W150). Most of these sequences, appeared only once at each timepoint. For participant F, 0.59% α chain CDR3 sequences were shared between the timepoints most of which were detected only once at either. Since, more CDR3 β chain sequences were available for W0 and raw data were subsampled at a higher level (19999 reads), a higher number of expanded CDR3s can be seen. For patient G, samples for both α and β chain W0 sequences, were excluded from subsampling since it only had a small number of raw reads available (367 and 283 respectively). Out of the identified CDR3 α and β sequences, only 1.2% and 3.23% respectively were shared between timepoints. An example of expanded CDR3 sequences that were identified in the CT group and shared between timepoints as well as their observed frequencies can be seen at the table below (Table 33).

Table 33 Top five expanded CDR3 sequences in CT participants, per chain, and their frequencies at each study timepoint (W0/W2 and W150).

Sample ID	Chain	CDR3 Sequence	Frequency (Read Numbers)	
			W0/W2	W150
E	α	CAVNMKSASKIIF	5	1
		CADMDSNYQLIW	4	2
		CAVSEGTGNQFYF	3	1
		CAAYNNNDMRF	2	1
		CATDDRGSTLGRLYF	2	1
	β	CASSKGLAAYEQYF	16	4
		CASSLMGASNQPQHF	16	3
		CSATDGINTIYF	16	6
		CSVDPTGLNYGYTF	16	5
		CSVEINYGYTF	15	10
F	α	CAVNMKSASKIIF	11	3

G		CADMDSNYQLIW	5	4
		CAASGGYQKVTF	4	2
		CAASIIQGGGADGLTF	2	1
		CALMNTGFQKLVF	2	1
		CSVGVTRPDNEQFF	100	17
		CASSVTGTSGQETQYF	82	15
	β	CASSLRQAGGPDTQYF	81	10
		CSASDPGAFF	78	20
		CASSQQKKTQYF	74	19
		CAVGNNAGNMLTF	11	1
		CAMREASFSGGYNKLIF	7	3
	α	CAASNTGNQFYF	5	4
		CAVTEGGKLIF	5	3
		CAASSSNTGKLIF	4	11
		CASSFRTGTDTQYF	6	2
		CASSFRTGTDTQYF	6	2
	β	CASRVGAEGYTF	5	3
		CASRVGAEGYTF	5	3
		CASSPSGASTDTQYF	4	1

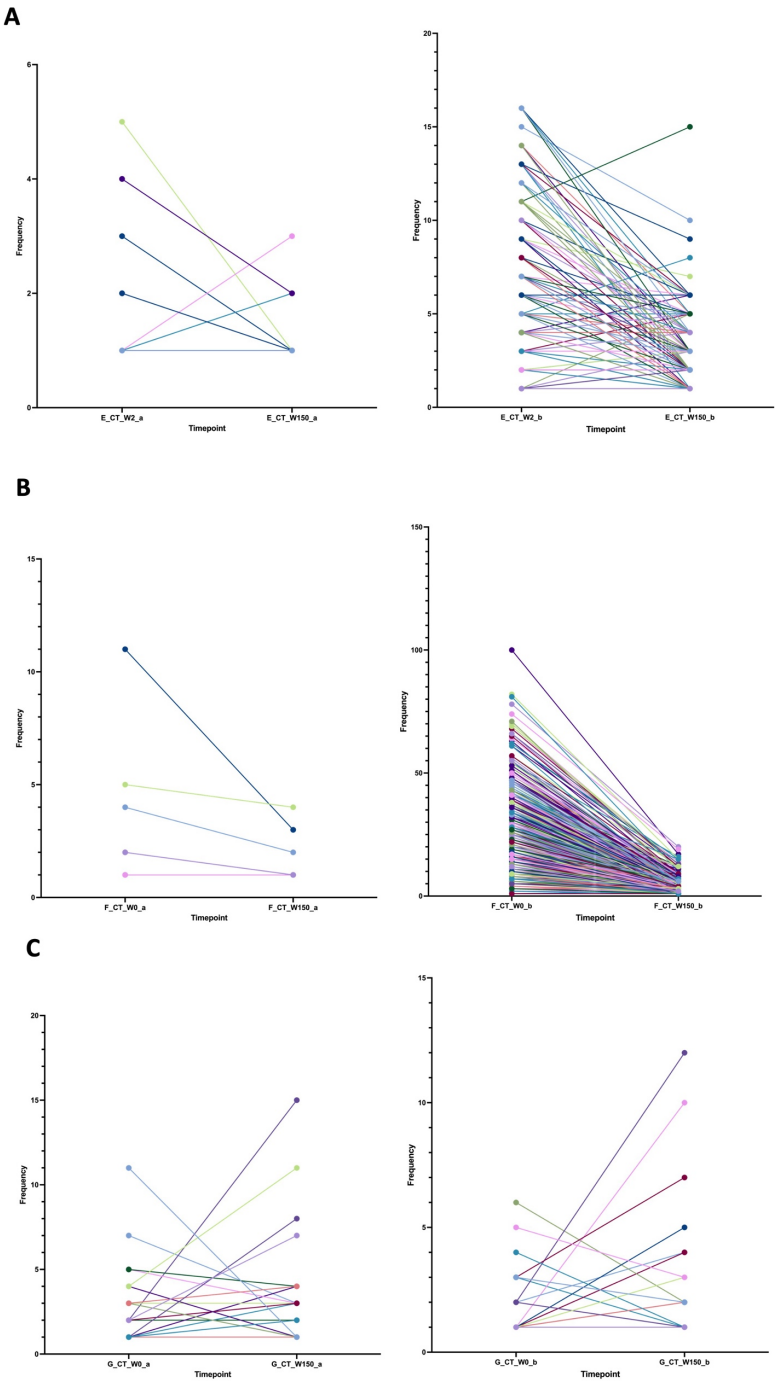


Figure 47 Plots presenting the TCRα and TCRβ sequences shared between the different PENTA11 study timepoints (W0/W2 and W150) in the three patients receiving continuous treatment.

Antigen specificity/epitope cross-reactivity was investigated using the VDJdb database. Similarly, to the analysis described in the previous chapters, several of the identified CDR3 sequences in the PENTA11 cohort are specific to more than one epitope (Table 34). The majority of the CDR3 sequences present in both the PTI and CT patients are EBV or CVM specific, with only a small number being associated with HIV-1 specificity (Table 35). Interestingly, no HIV-1 specific sequences were identified in any of the three CT patients.

Table 34 Example of antigen specificity as identified by the VDJdb database. Cross-reactivity was observed for multiple CDR3 sequences found within the PENTA11 data.

CDR3 Sequence	Gene	Antigen Specificity/Match
CAAPTGNQFYF	TRA	EBV
CALSETSYDKVIF	TRA	EBV, CMV
CASSSRGGQETQYF	TRB	EBV, CMV
CAVDSNYQLIW	TRA	CMV, InfluenzaA, HIV-1
CAALTGGGNKLTF	TRA	EBV, CMV, HIV-1
CAAREETSGSRLTF	TRA	EBV, CMV, HIV-1
CASSLTGNTEAFF	TRB	CMV, HIV-1
CAAMDSNYQLIW	TRA	EBV, CMV, InfluenzaA, HIV-1
CAVRDSNYQLIW	TRA	EBV, CMV, HIV-1
CAVSGYSTLTF	TRA	EBV, CMV, InfluenzaA, HFV

Table 35 HIV-1 specific CDR3 sequences present in PENTA11 PTI patients and their frequencies at the four different timepoints. Antigen specificity was determined using the VDJdb database. No HIV-1 related sequences were identified in the three CT patients.

CDR3 Sequence	Patient	Timepoint	Frequency
CAVDSNYQLIW	A_PTI	W0	4
		W12	2
		W48	1
		W150	0
CAALTGGGNKLTF	B_PTI	W0	0

		W24	0
		W48	1
		W150	2
CAAREETSGSRLTF	B_PTI	W0	1
		W24	0
		W48	0
		W150	3
CASSEAGGTEAFF	B_PTI	W0	2
		W24	1
		W48	0
		W150	0
CASSLGTGGNQPQHF	B_PTI	W0	0
		W24	1
		W48	1
		W150	0
CASSLTGNTEAFF	B_PTI	W0	4
		W24	0
		W48	2
		W150	4
CAVDSNYQLIW	B_PTI	W0	0
		W24	1
		W48	0
		W150	2
CAVYTGNQFYF	B_PTI	W0	0
		W24	0
		W48	1
		W150	3
CAAMDSNYQLIW	C_PTI	W0	1
		W12	2
		W48	0
		W150	0

CAVRDSNYQLIW	C_PTI	W0	1
		W12	0
		W48	0
		W150	1

Next, hamming distances were calculated for each one of the PENTA11 to investigate the possible effects of treatment interruption. As previously discussed, hamming distances look at CDR3 sequence similarity by determining the number of amino-acid (AA) changes present between sequences of similar length. The presence of higher hamming distances is an indicator of sequence dissimilarity and higher TCR repertoire diversity. As seen in Figure 48, all three patients included in the PTI group, present with high hamming distances in average. Higher differences in hamming distances between the four timepoints are apparent for patient C, with greater similarity between the detected CDR3 α sequences appearing at timepoints W0 and W150 but higher hamming distances (sequence dissimilarity) present at W48. Between the CDR3 β sequences found for this patient, W48 presents a shift in distribution towards higher hamming distances, while W150 demonstrates a shift towards lower hamming distances and higher sequence similarity.

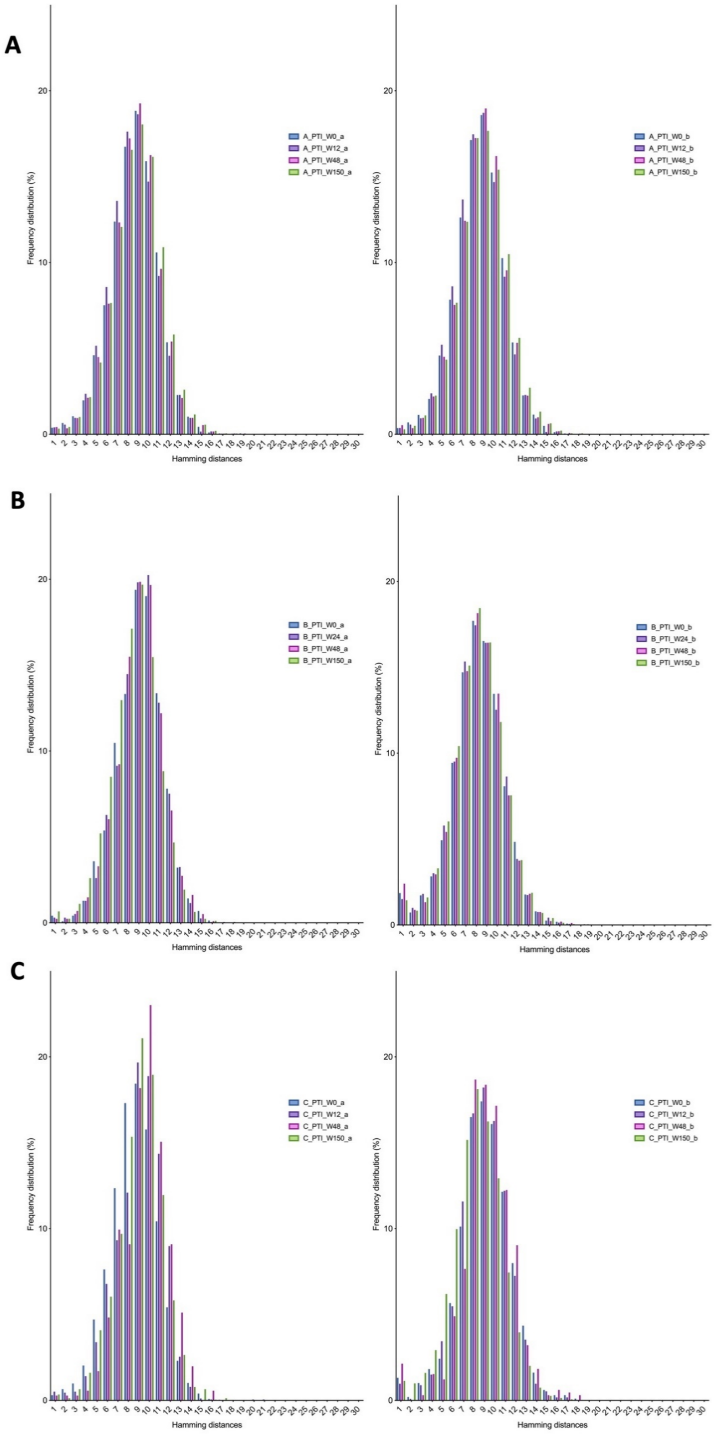


Figure 48 Frequency distribution of Hamming distances between the different timepoints for each PENTA11 participants undergoing planned treatment interruption. Hamming distances represent the number of amino acid changes between two neighbouring CDR3 sequences of the same length.

As seen in the figure below (Figure 49), on average, there is an increased frequency of higher hamming distances present in the CT group. A shift in the hamming distance distribution between W0 and W150 is more apparent in this group, with patient G demonstrating the greatest differences, potentially due to the W0 data availability.

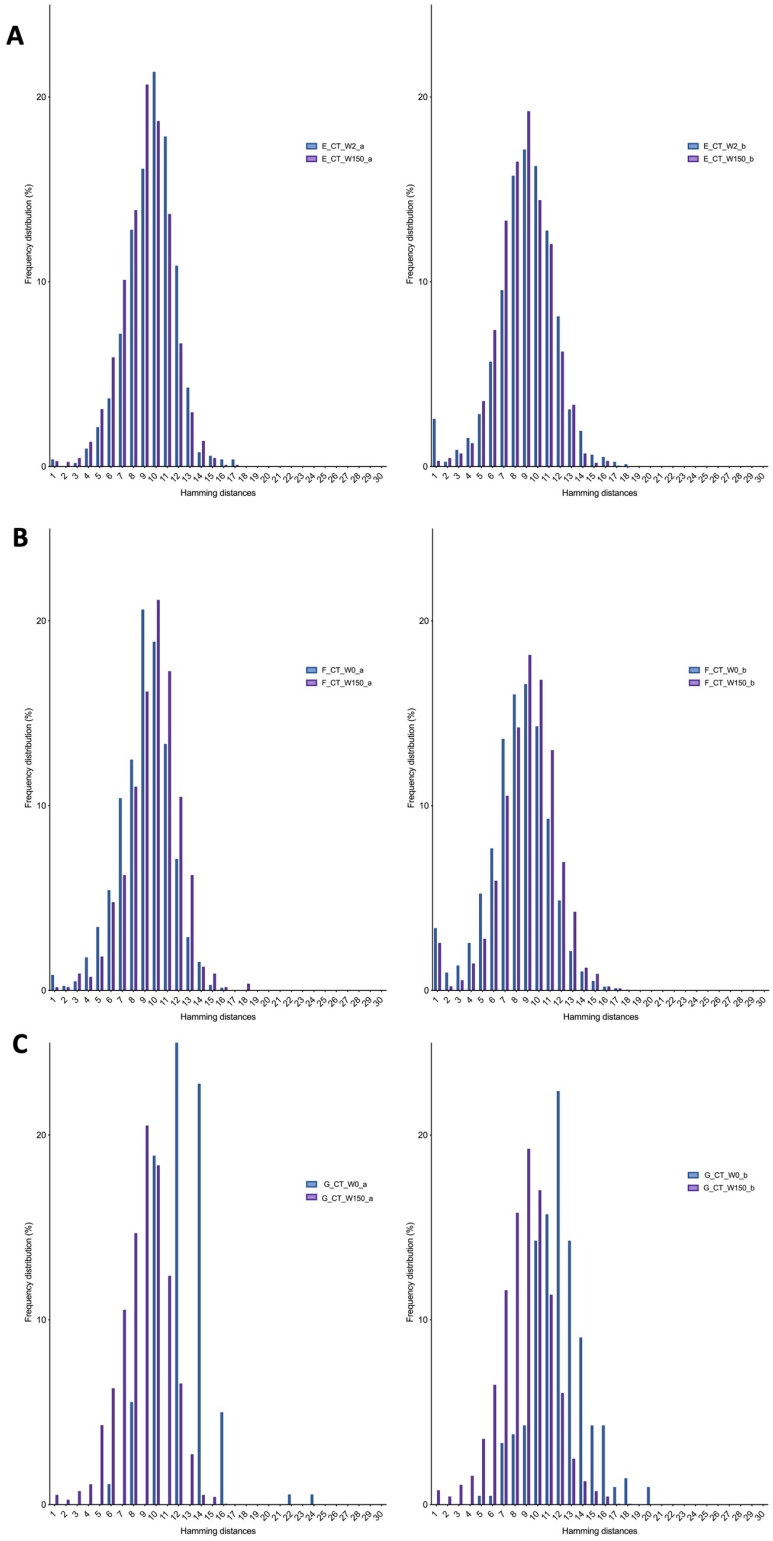


Figure 49 Frequency distribution of Hamming distances between the different timepoints for each PENTA11 participants on continuous treatment. Hamming distances represent the number of amino acid changes between two neighbouring CDR3 sequences of the same length.

The final measure of sequence similarity used for this chapter was performed by investigating the presence of closely related CDR3 sequence clusters, as previously described[14,204]. Clusters generated by subsampled (apart from patient G's W0) CDR3 α chain sequence data were selected to demonstrate the differences in sequence similarity between timepoints. B chain cluster data can be found in APPENDIX 8. As stated in earlier chapters of this thesis, lower subsampling levels and sequence availability affects the number of clusters that can be generated using this R script. In Figure 50, clusters of related CDR3 sequences in the three PTI patients and the effects of treatment secession is depicted. Loss of clusters can be seen in patient A for W12 and W48, and in patient C for W48. For all these three timepoints, the raw sequencing data were subsampled down to 557 reads (the lowest of the three categories), potentially explaining the presented differences.

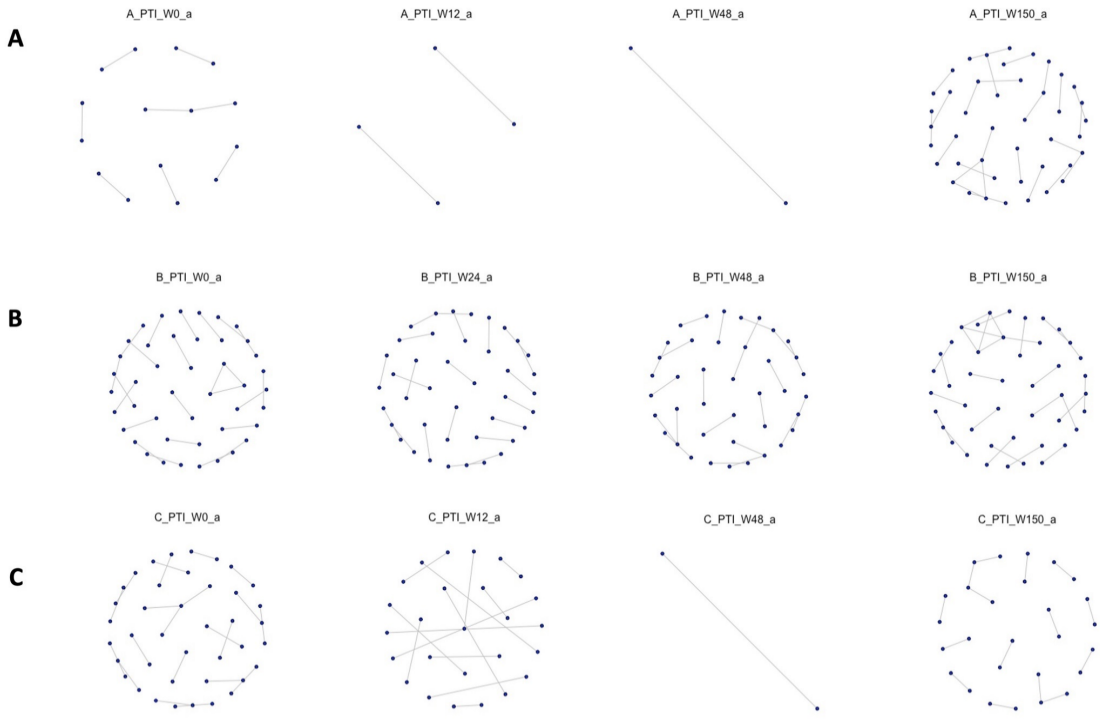


Figure 50 Clusters of closely related TCRα CDR3 sequences present at each PENTA11 study timepoint (W0, W12/W24, W48 and W150) in the planned treatment interruption arm. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes.

Similarly for the three patients included in the CT group, lack of closely related CDR3 sequences for patients E (W2), F (W150) and G (W0) could be attributed for the low numbers of sequences available for investigation.

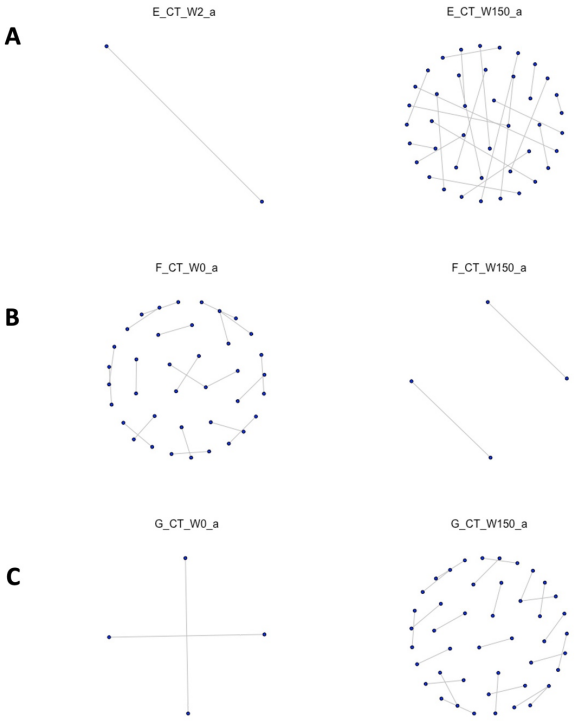


Figure 51 Clusters of closely related α CDR3 sequences present at each PENTA11 study timepoint (W0/W2 and W150) in the continuous treatment arm. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. Lower subsampling levels affect the number of clusters that can be detected leading to obvious differences in the presented plots.

The shared CDR3 sequences per patient were uploaded to the MAIT Match Server and compared to known MAIT cell sequences. Sequences with a similarity score of 1 were considered a perfect match. Out of all the available sequences for all the 6 patients, only 8 appeared to be a perfect match to MAIT cells (Table 36). No MAIT cell sequences were identified in patients F and G of the CT arm.

Table 36 Sequence similarity between CDR3 sequences identified in the PENTA11 patients and known MAIT cell sequences. A similarity score of 1 represents an exact match to a known MAIT cell sequence. No MAIT cell sequences with a similarity score of one were found in patients F_CT and G_CT.

CDR3 Sequence	MAIT Sequence	Similarity Score	Patient
CAVSGDYKLSF	CAVSGDYKLSF	1	A_PTI
CAVSDSSYKLIF	CAVSDSSYKLIF	1	A_PTI
CAASKAAGNKLTF	CAASKAAGNKLTF	1	B_PTI
CAVLDSNYQLIW	CAVLDSNYQLIW	1	B_PTI
CAAMDSNYQLIW	CAAMDSNYQLIW	1	C_PTI
CAVRDSNYQLIW	CAVRDSNYQLIW	1	C_PTI
CAVLNTGGFKTIF	CAVLNTGGFKTIF	1	C_PTI
CAPSGSARQLTF	CAPSGSARQLTF	1	E_CT

6.5 Discussion

To date, the lack of an HIV-1 cure has led to the use of lifelong continuous ART being the only option to successfully manage disease levels in both adult and paediatric patients. However, even though continuous treatment is the gold standard, it can lead to issues such as side-effects caused by drug toxicity, suboptimal adherence, and the development of viral resistance [300,301]. These are some of the reasons that highlight the importance of investigating new treatment strategies and why planned treatment interruption has been explored in both adult and paediatric cohorts. Unfortunately, this strategy has presented challenges and has mostly been viewed in a negative light. Adult PTI studies, such as the SMART trial, showed an increase in opportunistic infections and even deaths within their cohort[283].The Staccato study, reported low CD4+ counts, increased incidence of opportunistic infections (such as candidiasis) and an increase of HIV-1 viral load amongst the participants included in the PTI arm of the trial[302]. The Stop Antiretroviral Therapy (STOPAR) trial showed an increase in cardiovascular risk in the participants included in the treatment interruption group[303].

In contrast, paediatric patients taking part on treatment interruption studies, appeared to fair better than adults. The PENTA11 paediatric study, was a 72-week randomised phase II trial. Children (n=109) with HIV-1 RNA levels below 50copies/ml and CD4+ percentages of at least 25-30%, depending on age, were selected, and split into two groups (CT and PTI)[304]. In the PTI group, ART was re-initiated if the percentage of CD4+ T-cells dropped below 20% or once the 48-week mark was reached[304]. This trial demonstrated that when well monitored, PTI can be a viable option for children, since no or limited adverse virological or immunological outcomes were reported[180,181,305,306]. The Optimising Paediatric HIV-1 Treatment study (OPH-03) investigated children that received treatment before 13 months of age and had adhered to it for 2 years prior to the study[307]. It showed that short treatment interruptions did not cause any adverse side effects, such as loss of viral control, drop in CD4+ levels or morbidity, when compared to continuous treatment strategies in infants[307]. Similarly, the BREATHER trial included 199 participants and demonstrated no significant adverse effects of short treatment interruptions when compared to continuous treatment, suggesting it could be a potentially viable therapeutic strategy in HIV-infected children[308].

In this chapter, the focus was on the PENTA11 trial, by investigating the effects of PTI and CT to the CD4+ memory TCR repertoire. The results presented and discussed here should be viewed carefully, under the scope of the study limitations. Initially, only a small number of patient samples was available. Out of the stored RNA samples from 9 patients that were chosen, only 6 (3 for the PTI group and 3 for the CT group) were successfully sequenced and selected for further analysis. The inability to obtain sequencing data from three patients and the amount of usable sequencing reads generated for the remaining 6 patients could stem from the low PBMC numbers and subsequent reduced RNA quality of PENTA11 samples. This could potentially be attributed to inappropriate cell storage conditions and poor recovery following cell thawing as previously reported[181,297]. Additionally, another limiting factor that affected the sample quality and potential RNA yield stemmed from the necessity to fix the stained PBMCs prior to sorting due to local safety regulations, as previously discussed in this thesis.

Previous work from our group, on the TCR repertoire of CD4+ naïve and CD8+ memory T-cells in a small subset of PENTA11 patients, revealed that PTI decreased the TCR repertoire diversity of the CD4+ naïve T-cell population and increased the number of clonal expansions[204]. However, the TCR repertoire diversity returned to levels like the ones observed at the start of the trial. No changes in diversity were observed in the CD8+ memory population[204]. In a similar fashion, the Gini co-efficient was used to gage at the diversity of the CD4+ memory TCR repertoire[204]. As with the CD8+ memory population, the TCR diversity of the patients remained at similar levels (0.2-0.3) throughout the different study timepoints, from the start of the study (W0/W2) until the long-term follow-up point (W150). A low Gini co-efficient demonstrates a relatively evenly distributed cell population, where sequences appear at similar frequencies. Generally, exposure to antigens, like in this case exposure to HIV-1 during the PTI period, is expected to trigger a massive expansion of antigen specific T-cells leading to TCR repertoire bias and oligoclonality[202,204,309]. Here, the observation of an evenly distributed CD4+ T-cell memory population should be interpreted with caution. No definitive conclusions can be drawn on the effects of PTI and CT on the CD4+ memory populations as the analysed data could be skewed due to limitations present, including low sample quality and sequencing analysis restrictions.

Next, the focus was turned on the TCR repertoire composition by investigating CDR3 sequence sharing within each patient throughout the different timepoints. For the three patients included in the PTI group (A, B and C) only a small percentage (<10%) of sequences were shared between any two timepoints. Most of the identified sequences appeared to be unique for each one of the timepoints (W0, W12/W24, W48 and W150). Moreover, a sequence sharing trend was noticeable in all three patients, a drop in sequence sharing and their frequencies was observed following treatment interruption. Sequence frequencies started to recover once treatment was re-instated and clonal expansions were seen at W150. For the three patients in the CT group (E, F and G), most of the shared sequences showed similar frequencies between the two timepoints (W0 and W150). As with the PTI group, only a small percentage (<5%) of sequences were shared between the two timepoints in all three patients. The lack of sequence sharing, and the low read number frequencies observed in both groups could be an artifact of subsampling due to the restricted amounts of raw data available for my analysis.

Using the identified shared sequences per patient, antigen "specificity" was determined by using the VDJdb database. Even though the term "specificity" is used throughout this thesis, it is not necessarily the correct one when talking about antigenic epitope recognition. It is known that TCR sequences exhibit cross-reactivity to different epitopes, even for antigens that have not been encountered previously[203,310,311]. As expected, several of the identified CDR3 sequences recognised more than one epitope and showed specificity towards different pathogens, such as EBV, CMV, InfluenzaA, HFV and HIV-1. EBV and CMV appear to be the two predominant specificities that appear in this cohort, potentially showing greater levels of CMV exposure as have previously been reported[204]. Interestingly, only a very small number of the identified sequences showed "specificity" to HIV-1 within the PTI group. No HIV-1 "specific" sequences were found in any of the three CT patients, possibly due to the presence of suppressive ART. As previously stated, the antigen "specificity" data presented here derived by investigating α and β chains separately, hence the conclusions should be treated with caution.

The calculation of hamming distances and the generation of TCR sequence clusters were used to investigate CDR3 sequence similarity amongst the 6 patients. Within the PTI group, high

hamming distances were observed, implying the presence of mostly dissimilar CDR3 sequences throughout the study timeframe. Slight shifts in the hamming distance frequency distribution, or even more apparent changes between timepoints as seen in patient C, could be an artifact of different subsampling depth and sequencing data availability. A greater frequency of increased hamming distances was observed within the CT group and a noticeable distribution shift between the two timepoints. Patient G presents with the greatest differences in sequence similarity between W0 and W150. This result could be potentially attributed to the very low number of sequencing reads available for the W0 timepoint. The presence of clusters of closely related α and β CDR3 sequences were investigated as an additional measure of sequence similarity. Here, CDR3α sequences were selected to visualise any differences cause by the different treatment strategy. TCRβ CDR3 clusters can be found in APPENDIX 8. Unfortunately, the script used for cluster generation is highly dependent on the number of sequences available for investigation. In this case lower subsampling levels led to the generation of reduced TCR clusters, without necessarily representing the true effect of PTI, thus making it impossible to draw accurate conclusions on the actual diversity of sequences present at any given timepoint of the study. In the two figures presented (Figure 50 and Figure 51) all samples that present with an obvious loss of clusters belong to the lower subsampling depth category, potentially accounting for the differences.

Finally, the shared CDR3 sequences found in each patient were investigated to determine similarity to known MAIT cell sequences. Sequences presenting a similarity score of 1 were considered a perfect MAIT cell match. As expected, out of the total number of sequences available for this enquiry, only a very small number (8) matched to a known MAIT cell sequence. Most of the classed MAIT cell sequences were found in the PTI group, while sequences from patients F and G from the CT group returned no matches. MAIT cells comprise 0.1-10% of the peripheral blood population[312]. Identifying a very small number of MAIT cell sequenced within the CD4+ T-cell memory cohort is to be expected, since it has been previously reported that the majority of MAIT cells are CD8+ (80% approximately) and only a small percentage is either CD4+ or CD4-CD8-[313]. Moreover, the limited number of the available TCR sequences for this sub-study, restricts this type of analysis.

6.5.1 Conclusions

This chapter highlights the potential of using planned treatment interruption as a strategy in paediatric patients. By using CD4+ memory T-cell populations of a small number of individuals that took part in the PENTA11 study, no major effects in the TCR repertoire between the CT and the PTI groups was seen. This sub-study comes to support previous reports on this cohort indicating that no adverse immunological effects are caused in children going through a controlled treatment interruption and that TCR diversity can recover once treatment is re-initiated.

Nonetheless, the low number of patients included in this study, 3 in the PTI and 3 in the CT group, and the inability to pair α and β chains present great limitation and prevent us from drawing statistically significant conclusions. Low sequencing read numbers could have also skewed the results presented in this study to a degree and might have masked differences between the two groups.

CHAPTER 7: THESIS DISCUSSION

This thesis aimed to expand our knowledge and contribute to the ever-expanding field of paediatric HIV-1 infections and can be separated into two distinct parts (Figure 52). Part I, aimed to explore the virological characteristics of early treated, HIV-1 perinatally infected individuals by developing and optimising an assay detecting cell-associated HIV-1 RNA. Part II, aimed to explore the immunological characteristics of HIV-1 perinatally infected children by studying three paediatric cohorts and investigating the effects of HIV-1 infection and ART on the T-cell receptor repertoire.

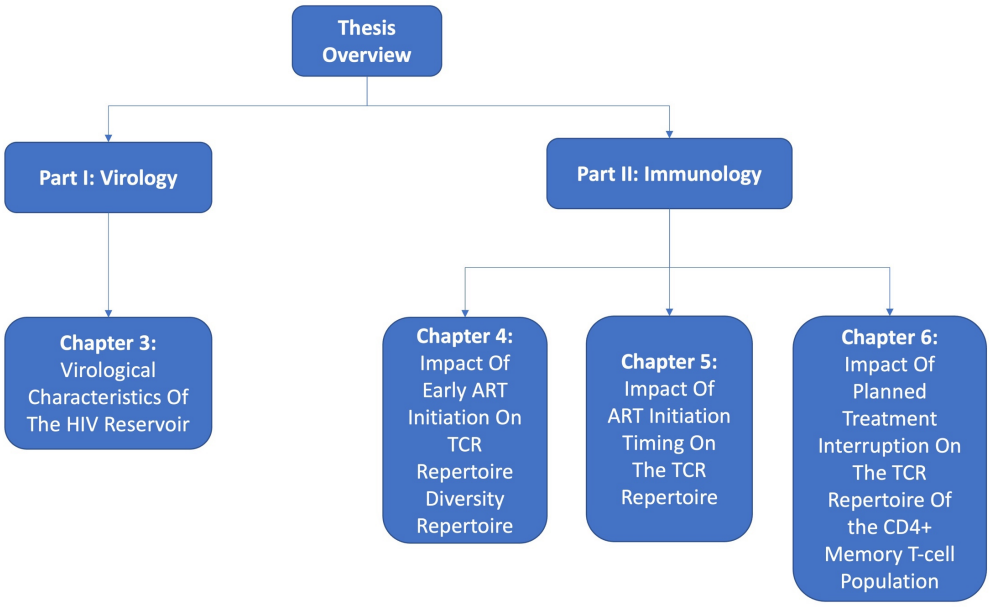


Figure 52 Thesis overview

7.1 Part I: Virological Characteristics of HIV-1 Perinatally Infected Children

Chapter 3 detailed the development and optimisation of an assay detecting the presence of HIV-1 CA-RNA, as a marker of the HIV-1 reservoir. To this day, the establishment of the viral reservoir early on during infection remains the main barrier in achieving a cure[314]. Several paediatric studies have reported that early ART initiation leads to the establishment of a reduced viral reservoir[7,8,211,314,315]. However, even reduced HIV-1 reservoirs can eventually lead to viral rebound once ART is stopped, making their monitoring via sensitive assays vital for the patients' health and treatment management[9].

In well-suppressed individuals on continuous ART, plasma viremia remains undetectable. In recent years alternative virological markers, such as cell-associated HIV-1 DNA and RNA, are being used to monitor the established viral reservoir as well as patient therapy response[71,211,316,317]. Here a cohort of 40 HIV-1 perinatally infected, well-suppressed, children was investigated for the presence of HIV-1 CA-RNA levels potential ongoing viral transcription. These individuals were part of the CARMA study, that previously published a report correlating ART initiation at younger age with the presence of lower HIV-1 DNA levels[4]. In this thesis, it was hypothesised that early ART initiation also leads to the presence of reduced HIV-1 CA-RNA levels. As with previous reports on HIV-1 DNA levels, younger age at ART initiation correlated with lower levels of HIV-1 CA-RNA. Additionally, HIV-1 CA-RNA was undetectable in 35% of the CARMA study participants. These finding are complimenting current literature supporting that early ART initiation favours the establishment of a reduced HIV-1 reservoir[4,318,319]. However, the finding that HIV-1 CA-RNA, even at low levels in some cases, can be detected in most well-suppressed individuals under continuous ART warrants further investigations to better elucidate the landscape of the HIV-1 viral reservoir and evaluate the significance of its clinical value.

Several limitations can be observed when investigating the findings presented here that could affect the way data are interpreted. These include the small number of patients in the study and the selection of only one study timepoint. Low participant numbers in study can cause issues when trying to draw statistically significant conclusions as the true effects of early ART initiation on the presence of HIV-1 CA-RNA might be slightly masked. The choice to have only

one timepoint in a study, hinders the ability to study viral reservoir dynamics throughout the years to determine any factors that could potentially have a positive or detrimental effect as these children begin to reach adulthood. It also prevents the proper evaluation the predictive strength of this virological marker in terms of treatment failure or the time and extent of a viral rebound. Moreover, in this instance it was impossible to determine if the detected HIV-1 CA-RNA derives from defective or replication competent HIV-1 integrated genomes. Future work, that would include longitudinal patient samples and viral sequencing studies could help in further advancing our knowledge on the effects early ART initiation on the HIV-1 viral dynamics.

It is true that the idea of using HIV-1 CA-RNA as a virological marker in clinical settings to monitor patients and treatment response is still in its infancy. A lot of questions remain on its significance, association with other virological and immunological markers, accuracy and time-effectiveness of published methods and ease of data interpretation. However, similarly to this study, recent publications have highlighted its potential and called for further research as better understanding and monitoring of the HIV-1 reservoir is important for the HIV-1+ paediatric population facing a lifetime of ART.

7.2 Part II: Immunological Characteristics of HIV-1 Perinatally Infected Children

Chapters 4-6 detailed the investigation of the effects of ART on the T-cell receptor repertoire by using a next generation sequencing method and three distinct cohorts. Thus far, studies assessing the impact of ART on the TCR repertoire have been limited and mostly focused on adult cohorts[156,229,320]. These studies report that even though ART helps restore the TCR repertoire diversity to a certain degree, it remains significantly perturbed, potentially due to the low thymic output levels and chronic immune activation seen in adults[229]. On the contrary, paediatric TCR repertoire studies, including ones conducted by our research group, have demonstrated promising data, showing that children and adolescents present with higher TCR repertoire diversity, possibly due to their developing immune systems and higher thymic output levels[14,204,321].

In Chapter 4, 19 CARMA study participants (from the UK) were selected to study the effects of early ART initiation on the developing immune system and the TCR repertoire diversity of HIV-1 perinatally infected children. Thymic output was investigated by studying TRECs levels. The results were in accordance with reports showing positive effects in the thymic output caused by ART, as well as reports showing decreased thymic output levels with increased age. [12,237,248,275,322]. As for the TCR repertoire diversity, homogeneity and a relatively normal TCR diversity was observed between the CARMA study participants when calculating the Gini coefficient. Out of the identified TCR α and β sequences only a small fraction was shared between at least two patients in this study. Moreover, using VDJdb, the identified sequences were matched with "specificity" towards common pathogens like EBV, CMV and InfluenzaA and only a small number matched with HIV-1. Higher frequencies of increased hamming distances and the presence of CDR3 clusters were another indicator of good TCR diversity in this cohort. Finally, known MAIT cell sequences were identified in our pool of samples, potentially indicating a good ability to respond against opportunistic infections. All these findings combined could be a testament of the positive effects that early and timely ART initiation has on the developing immune system and the TCR repertoire.

In Chapter 5, 5 early-treated CARMA study participants (from the UK) were selected and compared to 5 relatively age-matched late-treated adolescents and two healthy young-adult controls to investigate any potential differences. Like mentioned above, TRECs levels were used as a measure of thymic output. Depending on their age group, most participants presented with normal TRECs levels, however no definitive conclusions could be drawn since no reference data were available for the 10th centile of the 12-18 years age group. The Gini coefficient presented with relative uniformity within the three distinct groups, being a potential indicator that in paediatric settings TCR repertoire diversity can recover and reach relatively normal levels, following ART initiation. For CD4+ TCRα and TCRβ sequences a higher degree of sharing was observed within the early-treated group when compared to the late-treated group, but similar degrees of sharing were observed for CD8+ TCRα and TCRβ sequences. Hamming distances presented with similar patterns in the early-treated and healthy control groups, however higher frequencies of lower hamming distances were observed within the late-treated group, possibly pointing towards a less diverse repertoire. As before, the presence of closely related TCR clusters was used as another indicator of

sequence similarity and TCR diversity. For the CD4+ population most of the early-treated individuals presented with the highest number of TCR clusters, in the contrary to the CD8+ population where late-treated individuals presented with higher numbers. Cross-reactivity amongst the identified TCR sequences was apparent once more with only a small fraction being "specific" for HIV-1 and the majority being "specific" towards pathogens like EBV and CMV. Finally, the data were searched for any known MAIT cell sequences, showing higher numbers within the early-treated and healthy control groups when compared to the late-treated group, potentially suggesting that early ART could provide their immune system with better potential to fight certain infections.

In Chapter 6, participants from the PENTA11 study, 3 within the planned treatment interruption group and 3 within the continuous treatment group, were selected to investigate the effects of ART treatment strategies on the CD4+ memory T-cell population. As the first measure of TCR diversity the Gini coefficient was calculated and showed no differences between the two different groups at any timepoint. The relatively low Gini coefficient found indicated an evenly distributed TCR population. Even though only a small portion of sequences were shared between the timepoints investigated for each group, a noticeable drop in sharing was observed in the PTI group following treatment interruption. However, the degree of sequence sharing appeared to recover once ART was re-introduced. As previously seen in Chapters 4 and 5, most of the identified shared sequences deemed to be cross-reactive and "specific" towards common pathogens, predominantly CMV and EBV. Unexpectedly, no HIV-1 "specific" sequences were found to be shared in any of the three participants of the CT group. Hamming distances and clusters of closely related CDR3 sequences were investigated to further study any differences that can be caused to the TCR repertoire diversity due to treatment interruption. However, no definitive conclusions can be drawn as the small differences observed between patients and treatment groups could be attributed to low subsampling levels and data availability. Finally, as expected from previous publications, not many of the identified sequences within the groups matched with known MAIT cell sequences as most of them are CD8+ and only a small percentage of those found in the peripheral blood population are CD4+ or CD8-CD4-[163,168].

Even though, all three of the studies presented in Chapters 4-6 provide suggestions on the potential effects of ART on the TCR repertoire, they also present with great limitations that urge us to err in the side of caution when interpreting the data. Firstly, in all cases, only a small number of participants were included, thus restricting the strength of the conclusions. Studies with bigger cohorts, that would also account for differences caused by ethnicity, sex and age would be necessary to further elucidate and validate the findings presented in this thesis. Longitudinal follow-up studies would also enable the tracking of differences caused by age as these patients reach adolescence and later adulthood. Generally, TCR diversity levels are expected to fall during adulthood, due to the decrease in the naïve T-cell population and the increase of the memory T-cell population. Moreover, as studies on adults leaving with HIV-1 have shown that long-term ART only leads to partial recovery and a perturbed TCR repertoire, it would be of interest to explore the TCR dynamics of early-treated children once they reach adulthood[323,324]. The markers used here to investigate TCR diversity, like the Gini coefficient, Hamming distances, and presence of closely related TCR clusters can be influenced by data availability and computational analysis pipelines. The fact that presence of uniform and relatively similarly abundant TCR populations (e.g., highly similar low Gini coefficient values) throughout the chapters of this thesis can be caused or influenced by low sample numbers, sequencing data availability and quality, should not be disregarded. In future studies a different design and analysis approach could assist in obtaining a more accurate picture of the TCR repertoire landscape of HIV-1 paediatric patients and avoid analysis shortcomings. For example, additionally to recruiting higher numbers of participants as mentioned above, protocol changes on how PBMC cells are isolated and stored, removal of the cell "fixation step" (if possible) and improved RNA extraction methods would lead to an improved sequencing input and in turn data output. Furthermore, future studies including bigger sample sizes, would allow for the sorting and proper investigation of additional cell populations like the MAIT cells, and assist in accurate identification of associated sequences, even novel ones. The approach presented in this thesis was very limited as it was very depended on sequencing data deriving from general CD4+ and CD8+ populations and a database that includes previously sequences. Its main aim was to explore the possibility of identifying MAIT cell sequences in limited paediatric samples, and only provided a snapshot of their presence. Important questions of their total numbers and functionality in ART treated children remain.

In terms of analysis, a better subsampling method will be necessary to decrease the number of groupings and include all participant sequencing samples under one read depth, thus making them more comparable. As TCR analysis scripts and platforms are improving and new ones are being released, a critical evaluation of their analytical strength, suitability and accuracy should be performed to select the ones with the ability to provide the most accurate and informative data. Future studies using a single cell sequencing approach could assist in more accurate interpretation of the generated data, as α and β chain data would be paired thus providing a better understanding of TCR "specificity" and diversity.

It is worth noting that additional clinical information was requested for all study participants, including infective history, nutrition status, viral blips post study, possibility of relation, ethnicity, and type of ART. This additional information would have been useful for the thesis data interpretation and could potentially strengthen the observations. For example, it would be of interest to explore whether individuals with detectable HIV-1 CA-RNA levels had any viral blips or spikes following the end of the CARMA study, and how this virological marker could have better informed treatment management. Additionally, the possibility that the TCR repertoire homogeneity observed amongst different groups, is by any chance affected by different factors like ART regiments, gender, or ethnicity could have been explored. Information on previous infective status, if a CMV, EBV etc infection diagnosis exists, would have strengthen the observations of the detected TCR "specificity". In contrast, if the study participants had never been diagnosed with any of these infections, it would strengthen the hypothesis that the TCR repertoire is not necessarily always specific towards one pathogen, but instead is able to recognise a multitude of different antigenic epitopes. Unfortunately, despite the best efforts, including asking the supervisors and group's research associates for assistance, it was impossible to obtain the requested clinical information at this point.

7.3 Final Conclusions

To conclude, the research presented in this thesis highlights the importance of paediatric HIV-1 studies. Currently, an increasing number of HIV-1 perinatally infected children are reaching adolescence and adulthood, facing a lifetime on ART. Studies like this one, investigating the dynamics of the latent HIV-1 reservoir and the TCR repertoire diversity, are necessary to better understand the different virological and immunological characteristics of these unique cohorts and optimal patient management in the future. In this thesis, the potential effects of early ART initiation at a time when the immune system is still developing were presented. Future studies, including higher number of participants will be necessary to further elucidate the virological and immunological landscape of children born and living with HIV-1.

APPENDICES

APPENDIX 1

Table 37 The CARMA study included 40 samples from sites across Europe (UK, Spain, and Italy). After sample collection, PBMC isolation and DNA extraction, different assays were performed across the sites. Results from the total HIV-1 CA-DNA assay are shown below[4].

Sample ID	Total HIV-1 CA-DNA (HIV copies/10^6 PBMCs)
UK01	0.0
UK02	121.8
UK03	81.9
UK04	59.3
UK05	108.3
UK06	49.1
UK07	5.6
UK08	260.5
UK09	87.9
UK10	32.8
UK11	197.0
UK12	260.7
UK13	12.3
UK14	12.3
UK15	171.6
UK16	265.1
UK17	50.9
UK18	48.4
UK19	225.8
IT01	7.0
IT02	112.4
IT03	0.1

IT04	0.0
IT05	0.0
IT06	1.6
IT07	112.8
IT08	0.3
IT09	256.3
IT10	33.1
IT11	67.7
IT12	48.3
IT13	46.1
IT14	17.9
IT15	4.4
IT16	48.1
S01	0.0
S02	323.1
S03	41.1
S04	0.8
S05	63.4

APPENDIX 2

Table 38 UK CARMA samples selected for TCR sequencing for the research presented in Chapter 4. All samples (n=19) were sorted into two T-cell populations, CD4+ and CD8+ and RNA was extracted before library preparation and TCR sequencing.

Sample ID	Cell Population	Starting Sorted T-cell Numbers	Extracted RNA Concentration (ng/µl)
UK01	CD4+	1,728,624	136.3
	CD8+	800,040	51.4
UK02	CD4+	1,229,822	66.0
	CD8+	1,038,263	44.3
UK03	CD4+	600,621	42.7
	CD8+	598,031	39.1
UK04	CD4+	1,184,362	133.8
	CD8+	814,280	73.4
UK05	CD4+	649,839	33.9
	CD8+	272,078	22.6
UK06	CD4+	454,762	25.4
	CD8+	329,890	22.7
UK07	CD4+	1,003,895	65.8
	CD8+	424,394	30.8
UK08	CD4+	881,475	88.9
	CD8+	474,402	41.1
UK09	CD4+	823,260	102.9
	CD8+	545,324	45.7
UK10	CD4+	467,463	31.0
	CD8+	266,744	19.8
UK11	CD4+	452,006	36.1
	CD8+	233,081	21.4

UK12	CD4+	451,122	34.8
	CD8+	708,635	51.1
UK13	CD4+	1,046,905	86.8
	CD8+	594,873	40.1
UK14	CD4+	989,704	81.2
	CD8+	408,436	34.4
UK15	CD4+	1,357,929	129.2
	CD8+	367,540	19.3
UK16	CD4+	757,023	65.2
	CD8+	384,478	37.7
UK17	CD4+	1,087,998	78.9
	CD8+	612,573	52.7
UK18	CD4+	1,241,451	139.2
	CD8+	807,022	90.6
UK19	CD4+	736,921	65.8
	CD8+	416,750	45.8

APPENDIX 3

Table 39 Participants of the LD study (V=vertical transmission, C=healthy control) selected for comparison to the early treated CARMA study participants, in Chapter 5[4,14]. All samples were sorted into two T-cell populations, CD4+ and CD8+ and RNA was extracted before library preparation and TCR sequencing.

Sample ID	Cell Population	Starting Sorted T-cell Numbers	Extracted RNA Concentration (ng/µl)
V24	CD4+	1,475,223	138.2
	CD8+	1,359,164	69.9
V28	CD4+	66,179	70.7
	CD8+	1,874,537	86.2
V30	CD4+	301,239	25.0
	CD8+	571,044	42.8
V68	CD4+	513,595	52.4
	CD8+	2,408,997	80.0
V78	CD4+	281,731	18.4
	CD8+	190,689	92.8
C43	CD4+	1,064,312	183.0
	CD8+	1,371,323	73.0
C48	CD4+	2,185,262	219.6
	CD8+	1,923,733	50.7

APPENDIX 4

Table 40 RNA samples selected for the PENTA11 sub-study presented in Chapter 6, investigating TCR differences in the CD4+ memory T-cell subpopulation. Four timepoints for patients under PTI and two timepoints for patients under CT were chosen. Cell numbers after FACS sorting and RNA concentration after extraction are shown below.

Sample ID	Treatment	Week	Starting Sorted T-cell Numbers	Extracted RNA Concentration (ng/µl)
Patient A	PTI	0	300,822	71.2
		12	45,698	31.7
		48	168,624	12.4
		150	710,242	9.77
Patient B	PTI	Screen (before W0)	667,861	56.0
		24	412,683	34.6
		48	318,374	49.9
		150	675,280	66.4
Patient C	PTI	0	313,977	9.2
		12	339,759	26.8
		48	63,129	12.3
		150	720,138	80.8
Patient D	PTI	0	747,526	60.8
		12	395,495	52.4
		48	202,340	35.6
		150	552,390	64.1
Patient E	CT	2	286,738	28.8
		150	1,378,465	102.8
Patient F	CT	0	149,395	17.1
		150	1,109,380	58.5
Patient G	CT	0	236,598	2.7
		150	707,342	5.0
Patient H	CT	2	47,608	5.8

Patient I	CT	150	1,064,077	123.1
		0	170,891	15.0
		150	760,155	123.7

APPENDIX 5

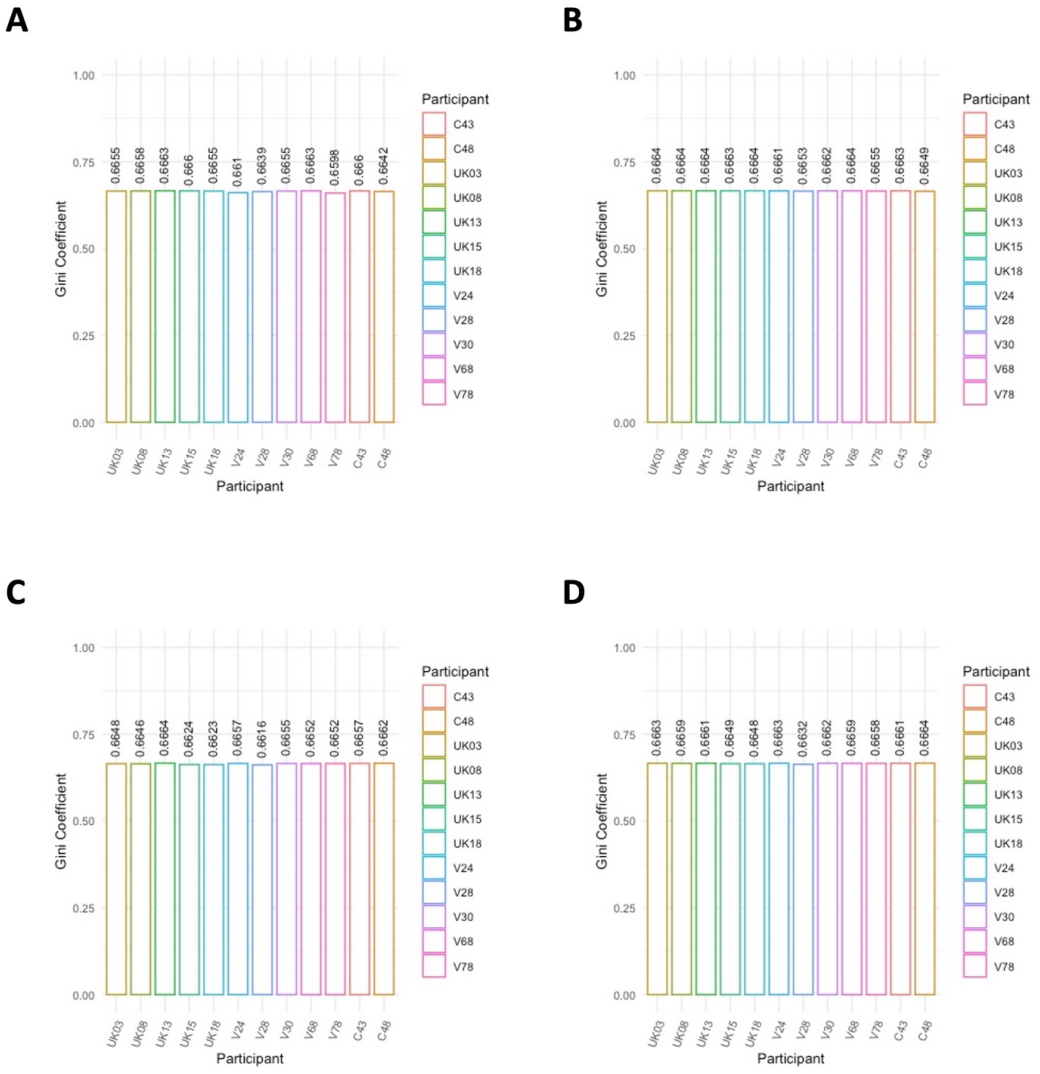


Figure 53 Due to similarities observed between participants in the Gini coefficient when calculated using subsampled sequencing data, it was also calculated using raw sequencing data, as seen here in these histogram plots. Similar clonotype abundance distribution patterns were observed between all patients for CD4+ A) TCRα and B) TCRβ as well as CD8+ C) TCRα and D) TCRβ sequences. The higher Gini coefficient present here when compared to the subsampled data could be attributed to the presence of higher frequencies of certain clonotypes.

APPENDIX 6

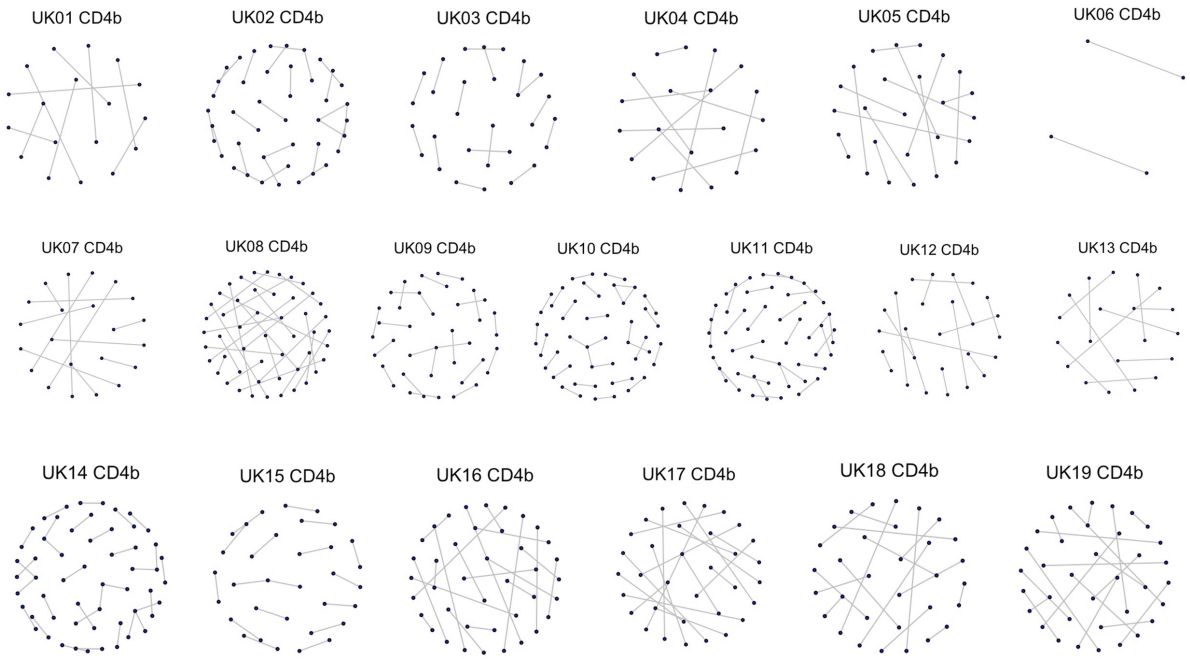


Figure 54 Clusters of closely related CD4+ TCRβ CDR3 sequences. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes.

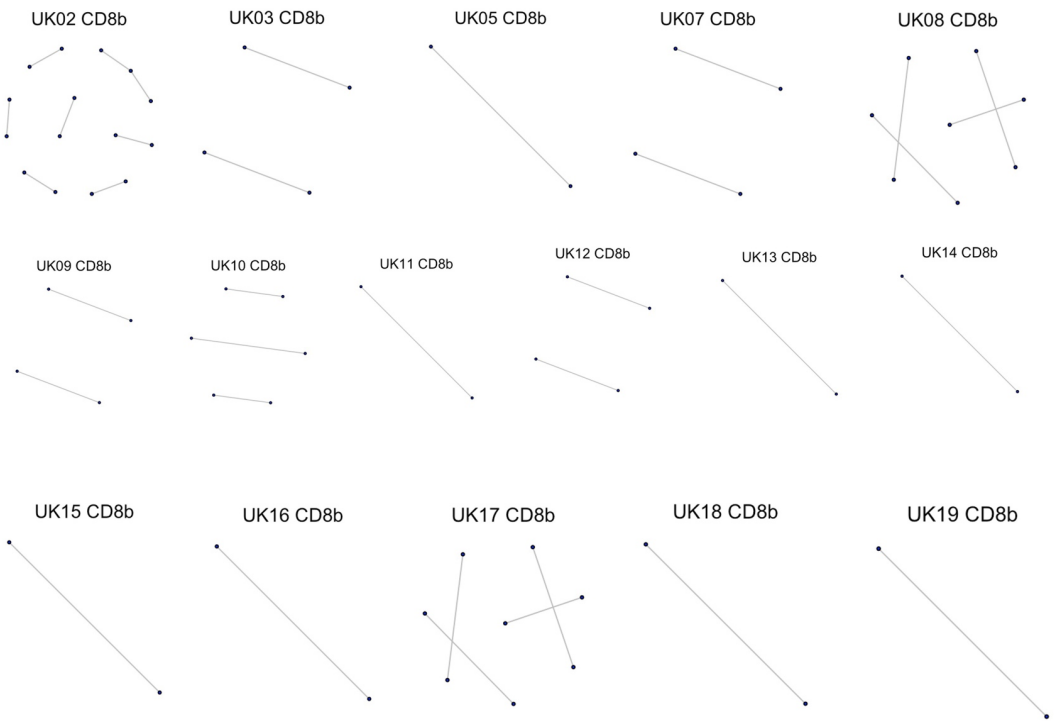


Figure 55 Clusters of closely related CD8+ TCRβ CDR3 sequences. Due to the subsampling levels of some samples (UK01, UK04 and UK06) it was not possible to generate plots.

APPENDIX 7

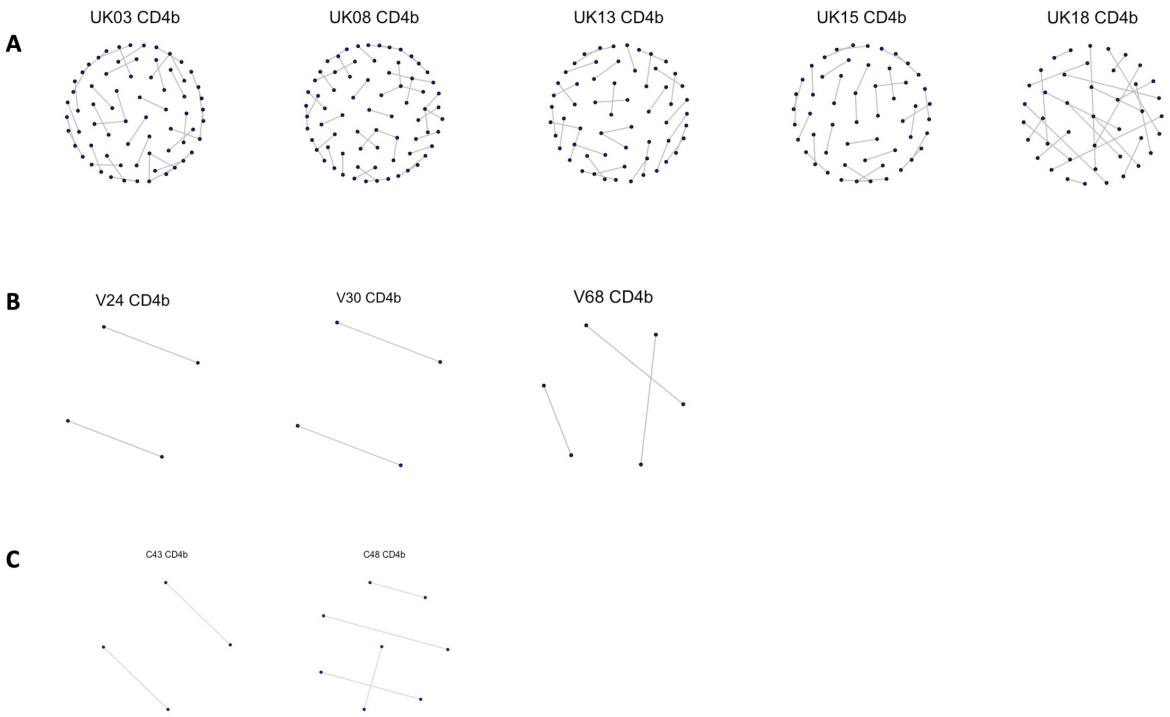


Figure 56 Clusters of closely related CD4+ TCRβ CDR3 sequences identified in participants in A) Group A, B) Group B and C) Group C. Participants in Group A, part of the CARMA study, had a higher number of clusters of closely related CDR3 sequences when compared to participants of the other two groups. Lower subsampling levels can affect the detection of clusters.

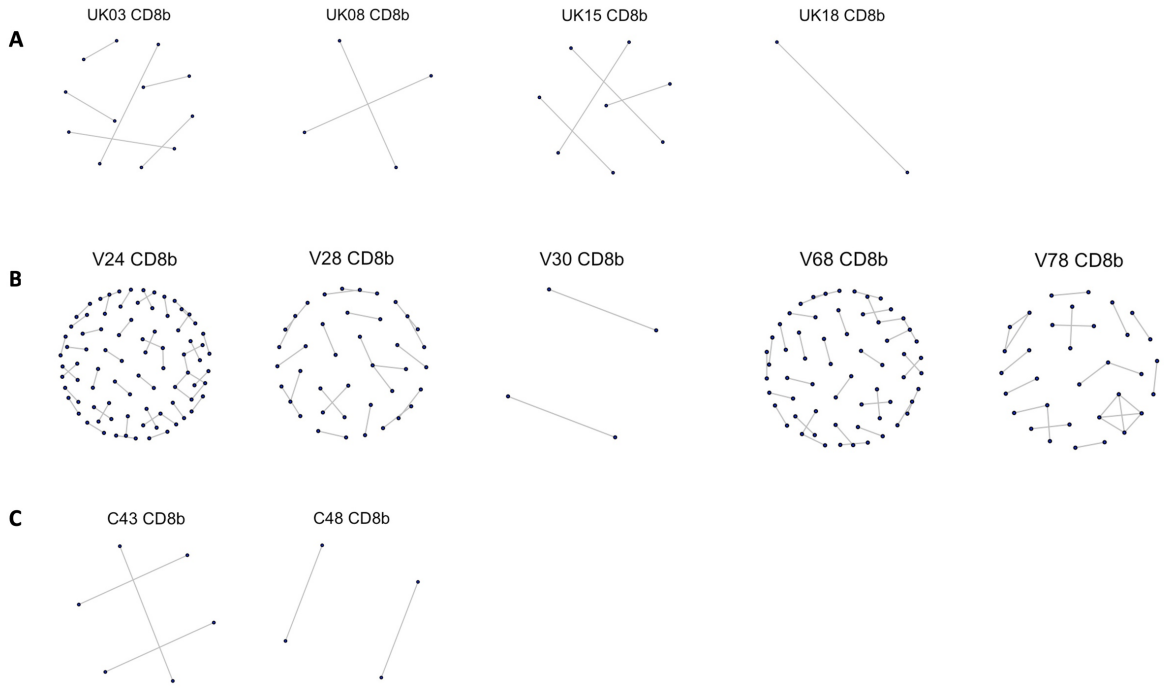


Figure 57 Clusters of closely related CD8+ TCRβ CDR3 sequences identified in participants of A) Group A, B) Group B and C) Group C. Participants of Group B (apart from V30 that was subsampled to a lower level) presented with higher numbers of closely related CDR3 clusters when compared to participants from Groups A and C. No plot could be generated for Group A participant UK13.

APPENDIX 8

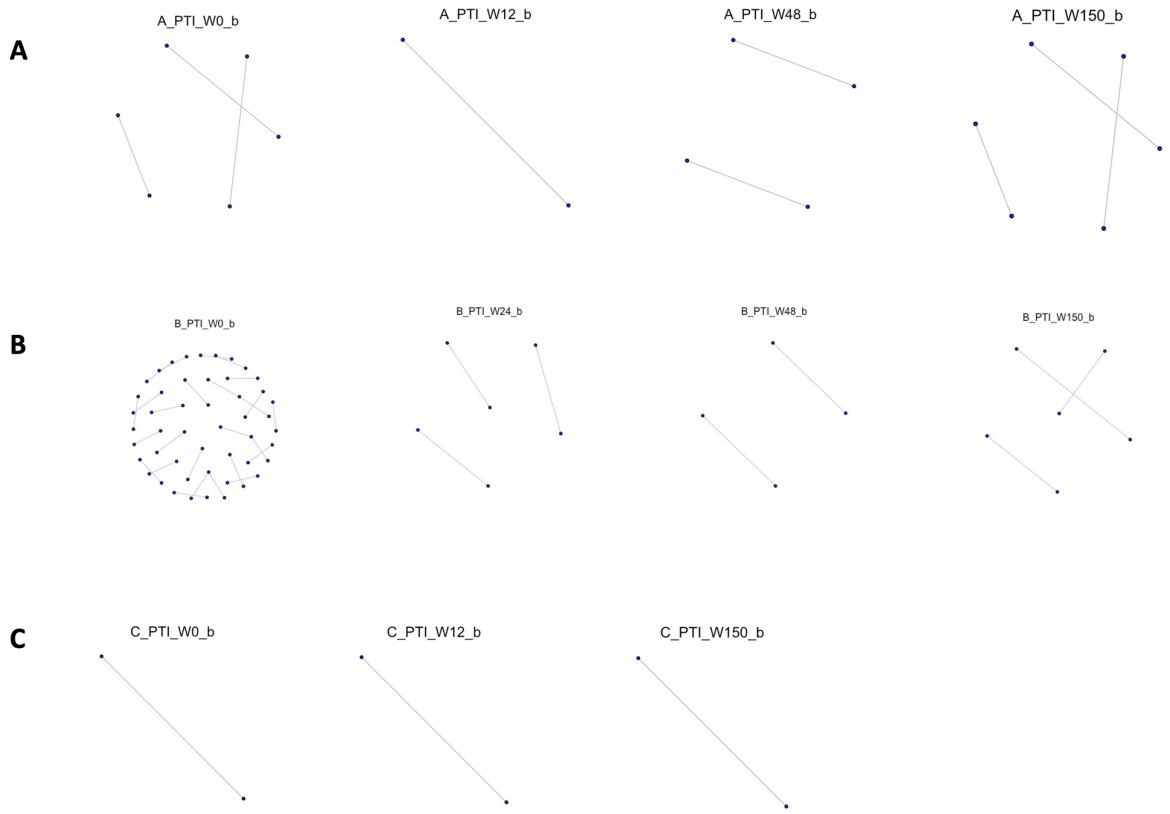


Figure 58 Clusters of closely related TCRβ CDR3 sequences present at each PENTA11 study timepoint (W0, W12/W24, W48 and W150) in the planned treatment interruption arm. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. No plot could be generated for timepoint W48 for participant C.

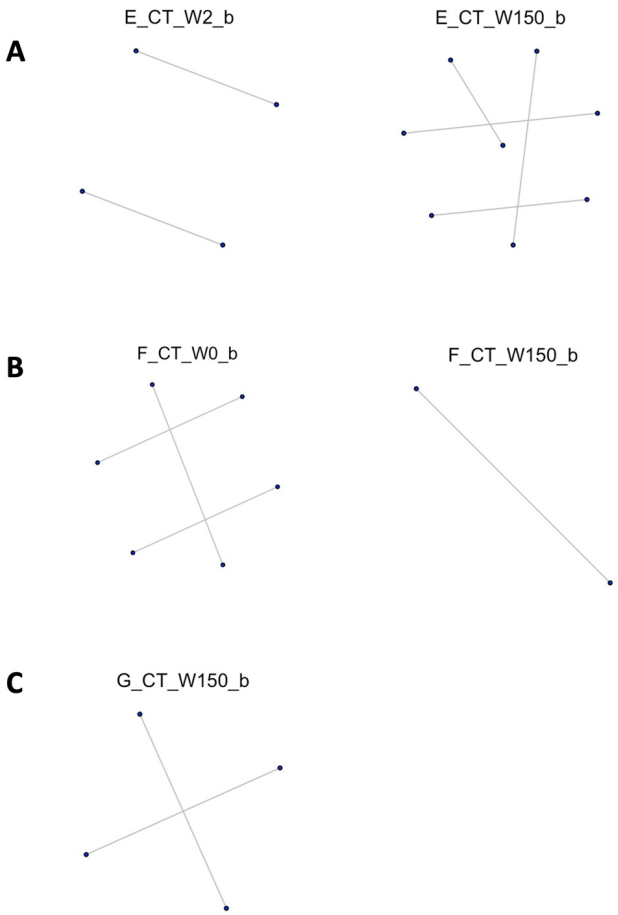


Figure 59 Clusters of closely related β CDR3 sequences present at each PENTA11 study timepoint (W0/W2 and W150) in the continuous treatment arm. CDR3s with the same motifs in their sequences form a network of clusters represented here as nodes. Lower subsampling levels affect the number of clusters that can be detected leading to obvious differences in the presented plots.

REFERENCES

1. Sohn, A. H. & Hazra, R. The changing epidemiology of the global paediatric HIV epidemic: keeping track of perinatally HIV-infected adolescents. *J Int AIDS Soc* **16**, 18555 (2013).
2. Tagarro, A. *et al.* Early and highly suppressive antiretroviral therapy are main factors associated with low viral reservoir in European perinatally HIV-infected children. in *Journal of Acquired Immune Deficiency Syndromes* vol. 79 269–276 (Lippincott Williams and Wilkins, 2018).
3. Rocca, S. *et al.* Human Immunodeficiency Virus (HIV)-Antibody Repertoire Estimates Reservoir Size and Time of Antiretroviral Therapy Initiation in Virally Suppressed Perinatally HIV-Infected Children. *J Pediatric Infect Dis Soc* (2018) doi:10.1093/jpids/piy080.
4. Foster, C. *et al.* The CARMA Study: Early Infant Antiretroviral Therapy—Timing Impacts on Total HIV-1 DNA Quantitation 12 Years Later. *J Pediatric Infect Dis Soc* 1–7 (2020) doi:10.1093/jpids/piaa071.
5. Domínguez-Rodríguez, S. *et al.* Clinical, Virological and Immunological Subphenotypes in a Cohort of Early Treated HIV-Infected Children. *Front Immunol* **13**, (2022).
6. Martínez-Bonet, M. *et al.* Establishment and Replenishment of the Viral Reservoir in Perinatally HIV-1-infected Children Initiating Very Early Antiretroviral Therapy. *Clin Infect Dis* **61**, 1169–78 (2015).
7. Rainwater-Lovett, K., Luzuriaga, K. & Persaud, D. Very early combination antiretroviral therapy in infants: Prospects for cure. *Current Opinion in HIV and AIDS* vol. 10 4–11 Preprint at https://doi.org/10.1097/COH.0000000000000127 (2015).
8. Persaud, D. *et al.* Absence of Detectable HIV-1 Viremia after Treatment Cessation in an Infant. *New England Journal of Medicine* **369**, 1828–1835 (2013).
9. Luzuriaga, K. *et al.* Viremic Relapse after HIV-1 Remission in a Perinatally Infected Child. *New England Journal of Medicine* **372**, 786–788 (2015).
10. Violari, A. *et al.* A child with perinatal HIV infection and long-term sustained virological control following antiretroviral treatment cessation. *Nat Commun* **10**, 412 (2019).
11. Frange, P. *et al.* HIV-1 virological remission lasting more than 12 years after interruption of early antiretroviral therapy in a perinatally infected teenager enrolled in the French ANRS EPF-CO10 paediatric cohort: a case report. *Lancet HIV* **3**, e49–e54 (2016).
12. Sandgaard, K. S., Lewis, J., Adams, S., Klein, N. & Callard, R. Antiretroviral therapy increases thymic output in children with HIV. *AIDS* **28**, 209–214 (2014).
13. Lewis, J. *et al.* Thymic Output and CD4 T-Cell Reconstitution in HIV-Infected Children on Early and Interrupted Antiretroviral Treatment: Evidence from the Children with HIV Early Antiretroviral Therapy Trial. *Front Immunol* **8**, 1162 (2017).
14. Sandgaard, K. S. *et al.* The importance of taking ART appropriately in children and adolescents with HIV-1 to reach the highest capacity of immune function later in life. *Front Immunol* **13**, 3885 (2022).
15. Bamford, A. *et al.* Paediatric European Network for Treatment of AIDS (PENTA) guidelines for treatment of paediatric HIV-1 infection 2015: optimizing health in preparation for adult life. *HIV Med* **19**, e1–e42 (2018).

16. Hymes, KennethB. *et al.* KAPOSI'S SARCOMA IN HOMOSEXUAL MEN—A REPORT OF EIGHT CASES. *The Lancet* **318**, 598–600 (1981).
17. Gottlieb, M. S. *et al. Pneumocystis carinii* Pneumonia and Mucosal Candidiasis in Previously Healthy Homosexual Men. *New England Journal of Medicine* **305**, 1425–1431 (1981).
18. Barré-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (1979)* **220**, 868–871 (1983).
19. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, Redfield R, Oleske J, Safai B, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science (1979)* **224**, 500–503 (1984).
20. Case, K. Nomenclature: Human Immunodeficiency Virus. *Ann Intern Med* **105**, 133 (1986).
21. Unaids. F A C T S H E E T 2 0 2 2.
22. Unaids. *Fact sheet - Latest global and regional statistics on the status of the AIDS epidemic*.
23. summary-of-the-global-hiv-epidemic-2021.png (1146×810). https://www.who.int/images/default-source/departments/hiv/summary-of-the-global-hiv-epidemic-2021.png?sfvrsn=73ac5b6a_9.
24. Clavel, F. *et al.* Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**, 343–6 (1986).
25. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med* **1**, a006841 (2011).
26. Nyamweya, S. *et al.* Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. *Rev Med Virol* **23**, 221–240 (2013).
27. Zagury, J. F. *et al.* Genetic variability between isolates of human immunodeficiency virus (HIV) type 2 is comparable to the variability among HIV type 1. *Proceedings of the National Academy of Sciences* **85**, 5941–5945 (1988).
28. Bbosa, N., Kaleebu, P. & Ssemwanga, D. HIV subtype diversity worldwide. *Curr Opin HIV AIDS* **14**, 153–160 (2019).
29. Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E. & Hammer, S. M. The challenge of HIV-1 subtype diversity. *N Engl J Med* **358**, 1590–602 (2008).
30. Vallari, A. *et al.* Confirmation of Putative HIV-1 Group P in Cameroon. *J Virol* **85**, 1403–1407 (2011).
31. Huet, T., Cheynier, R., Meyerhans, A., Roelants, G. & Wain-Hobson, S. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345**, 356–359 (1990).
32. Mourez, T., Simon, F. & Plantier, J.-C. Non-M variants of human immunodeficiency virus type 1. *Clin Microbiol Rev* **26**, 448–61 (2013).
33. Hemelaar, J. The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine* vol. 18 182–192 Preprint at https://doi.org/10.1016/j.molmed.2011.12.001 (2012).
34. Buonaguro, L., Tornesello, M. L. & Buonaguro, F. M. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J Virol* **81**, 10209–19 (2007).
35. Tyor, W., Fritz-French, C. & Nath, A. Effect of HIV clade differences on the onset and severity of HIV-associated neurocognitive disorders. *Journal of NeuroVirology* vol. 19 515–522 Preprint at https://doi.org/10.1007/s13365-013-0206-6 (2013).
36. Azevedo-Pereira, J. M. & Santos-Costa, Q. HIV Interaction With Human Host: HIV-2 As a Model of a Less Virulent Infection. *AIDS Rev* **18**, 44–53.

37. Hirsch, V. M., Olmsted, R. A., Murphey-Corb, M., Purcell, R. H. & Johnson, P. R. An African primate lentivirus (SIVsmclosely related to HIV-2. *Nature* **339**, 389–392 (1989).
38. Visseaux, B., Damond, F., Matheron, S., Descamps, D. & Charpentier, C. HIV-2 molecular epidemiology. *Infection, Genetics and Evolution* **46**, 233–240 (2016).
39. Shaw, G. M. & Hunter, E. HIV transmission. *Cold Spring Harb Perspect Med* **2**, (2012).
40. Kilmarx, P. H. Global epidemiology of HIV. *Curr Opin HIV AIDS* **4**, 240–246 (2009).
41. Scarlatti, G. Mother-to-child transmission of HIV-1: advances and controversies of the twentieth centuries. *AIDS Rev* **6**, 67–78.
42. Prendergast, A. J. *et al.* Transmission of CMV, HTLV-1, and HIV through breastmilk. *Lancet Child Adolesc Health* **3**, 264–273 (2019).
43. Douek, D. C. *et al.* HIV preferentially infects HIV-specific CD4+ T cells. *Nature* **417**, 95–98 (2002).
44. Sattentau, Q. J. & Stevenson, M. Macrophages and HIV-1: An Unhealthy Constellation. *Cell Host Microbe* **19**, 304–10 (2016).
45. Manches, O., Frleta, D. & Bhardwaj, N. Dendritic cells in progression and pathology of HIV infection. *Trends Immunol* **35**, 114–22 (2014).
46. Okoye, A. A. & Picker, L. J. CD4+ T cell depletion in HIV infection: mechanisms of immunological failure. *Immunol Rev* **254**, 54 (2013).
47. Naif, H. M. Pathogenesis of HIV Infection. *Infect Dis Rep* **5**, e6 (2013).
48. Swanstrom, R. & Coffin, J. HIV-1 pathogenesis: the virus. *Cold Spring Harb Perspect Med* **2**, a007443 (2012).
49. Coffin, J. & Swanstrom, R. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harb Perspect Med* **3**, a012526 (2013).
50. Kirchhoff, F. HIV Life Cycle : Overview HIV Life Cycle : Overview. (2016) doi:10.1007/978-1-4614-9610-6.
51. German Advisory Committee Blood (Arbeitskreis Blut), S. 'Assessment of P. T. by B. Human Immunodeficiency Virus (HIV). *Transfusion Medicine and Hemotherapy* **43**, 203 (2016).
52. Adamson, C. S. & Freed, E. O. Human Immunodeficiency Virus Type 1 Assembly, Release, and Maturation. *Adv Pharmacol* **55**, 347–387 (2007).
53. Bieniasz, P. D. The Cell Biology of HIV-1 Virion Genesis. *Cell Host Microbe* **5**, 550–558 (2009).
54. German Advisory Committee Blood (Arbeitskreis Blut), Subgroup 'Assessment of Pathogens Transmissible by Blood', G. A. C. B. (Arbeitskreis & Blood', S. 'Assessment of P. T. by. Human Immunodeficiency Virus (HIV). *Transfus Med Hemother* **43**, 203–22 (2016).
55. Arrildt, K. T., Joseph, S. B. & Swanstrom, R. The HIV-1 env protein: a coat of many colors. *Curr HIV/AIDS Rep* **9**, 52–63 (2012).
56. Karn, J. & Stoltzfus, C. M. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med* **2**, a006916 (2012).
57. Malim, M. H. & Emerman, M. HIV-1 Accessory Proteins—Ensuring Viral Survival in a Hostile Environment. *Cell Host Microbe* **3**, 388–398 (2008).
58. Faust, T. B., Binning, J. M., Gross, J. D. & Frankel, A. D. Making Sense of Multifunctional Proteins: Human Immunodeficiency Virus Type 1 Accessory and Regulatory Proteins and Connections to Transcription. *Annu Rev Virol* **4**, 241–260 (2017).
59. Richter, S., Frasson, I. & Palu, G. Strategies for inhibiting function of HIV-1 accessory proteins: a necessary route to AIDS therapy? *Curr Med Chem* **16**, 267–286 (2009).

60. Nisole, S. & Saïb, A. Early steps of retrovirus replicative cycle. *Retrovirology* **1**, 9 (2004).
61. Sundquist, W. I. & Krausslich, H.-G. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harb Perspect Med* **2**, a006924–a006924 (2012).
62. Pasternak, A. O., Lukashov, V. V & Berkhout, B. Cell-associated HIV RNA: a dynamic biomarker of viral persistence. *Retrovirology* **10**, 41 (2013).
63. Hu, W.-S. & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* **2**, (2012).
64. Craigie, R. & Bushman, F. D. HIV DNA integration. *Cold Spring Harb Perspect Med* **2**, a006890 (2012).
65. Purcell, D. F. & Martin, M. A. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J Virol* **67**, 6365–78 (1993).
66. Siliciano, R. F. & Greene, W. C. HIV Latency. 1–20 (2017).
67. Datta, P. K. *et al.* HIV-1 Latency and Eradication: Past, Present and Future. *Curr HIV Res* **14**, 431–441 (2016).
68. Agosto, L. M. & Henderson, A. J. CD4 + T Cell Subsets and Pathways to HIV Latency . *AIDS Res Hum Retroviruses* **34**, 780–789 (2018).
69. Kumar, A., Abbas, W. & Herbein, G. HIV-1 Latency in Monocytes/Macrophages. *Viruses* **6**, 1837 (2014).
70. Kumar, N. A. *et al.* Myeloid Dendritic Cells Induce HIV Latency in Proliferating CD4 $^{+}$ T Cells. *The Journal of Immunology* **201**, 1468–1477 (2018).
71. Avettand-Fènoël, V. *et al.* Total HIV-1 DNA, a Marker of Viral Reservoir Dynamics with Clinical Implications. *Clin Microbiol Rev* **29**, (2016).
72. Rouzioux, C. & Fenoël, V. A. Total HIV DNA : a global marker of HIV persistence. *Retrovirology* 1–7 (2018) doi:10.1186/s12977-018-0412-7.
73. Trémeaux, P. *et al.* Increasing contribution of integrated forms to total HIV DNA in blood during HIV disease progression from primary infection. *EBioMedicine* **41**, 455–464 (2019).
74. Kuzembayeva, M., Dilley, K., Sardo, L. & Hu, W. S. Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* vols 454–455 362–370 Preprint at https://doi.org/10.1016/j.virol.2014.01.019 (2014).
75. Pasternak, A. O. & Berkhout, B. What do we measure when we measure cell - associated HIV RNA. *Retrovirology* 1–16 (2018) doi:10.1186/s12977-018-0397-2.
76. Patterson, B. K. *et al.* Persistence of intracellular HIV-1 mRNA correlates with HIV-1-specific immune responses in infected subjects on stable HAART. *AIDS* **15**, 1635–1641 (2001).
77. Wiegand, A. *et al.* Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART. *Proceedings of the National Academy of Sciences* **114**, E3659–E3668 (2017).
78. Laird, G. M., Rosenbloom, D. I. S., Lai, J., Siliciano, R. F. & Siliciano, J. D. Measuring the Frequency of Latent HIV-1 in Resting CD4+ T Cells Using a Limiting Dilution Coculture Assay. in *Methods in molecular biology (Clifton, N.J.)* vol. 1354 239–253 (2016).
79. Chun, T. W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183–188 (1997).
80. Siliciano, J. D. & Siliciano, R. F. Enhanced culture assay for detection and quantitation of latently infected, resting CD4+ T-cells carrying replication-competent virus in HIV-1-infected individuals. *Methods Mol Biol* **304**, 3–15 (2005).

81. Pardons, M. *et al.* Single-cell characterization and quantification of translation-competent viral reservoirs in treated and untreated HIV infection. *PLoS One* **15**, 1–28 (2019).
82. Procopio, F. A. *et al.* A Novel Assay to Measure the Magnitude of the Inducible Viral Reservoir in HIV-infected Individuals. *EBioMedicine* **2**, 874–883 (2015).
83. Plantin, J., Massanella, M. & Chomont, N. Inducible HIV RNA transcription assays to measure HIV persistence: Pros and cons of a compromise. *Retrovirology* **15**, 1–11 (2018).
84. Hong, F. *et al.* Novel Assays for Measurement of Total Cell-Associated HIV-1 DNA and RNA. *J Clin Microbiol* **54**, 902–11 (2016).
85. Pasternak, A. O. *et al.* Highly Sensitive Methods Based on Seminested Real-Time Reverse Transcription-PCR for Quantitation of Human Immunodeficiency Virus Type 1 Unspliced and Multiply Spliced RNA and Proviral DNA □. *J Clin Microbiol* **46**, 2206–2211 (2008).
86. O'Doherty, U., Swiggard, W. J., Jeyakumar, D., McGain, D. & Malim, M. H. A Sensitive, Quantitative Assay for Human Immunodeficiency Virus Type 1 Integration. *J Virol* **76**, 10942–10950 (2002).
87. van der Sluis, R. M. *et al.* Quantitation of HIV-1 DNA with a sensitive TaqMan assay that has broad subtype specificity. *J Virol Methods* **187**, 94–102 (2013).
88. Eriksson, S. *et al.* Comparative Analysis of Measures of Viral Reservoirs in HIV-1 Eradication Studies. *PLoS Pathog* **9**, e1003174 (2013).
89. Busby, E. *et al.* Instability of 8E5 calibration standard revealed by digital PCR risks inaccurate quantification of HIV DNA in clinical samples by qPCR. *Sci Rep* **7**, 1209 (2017).
90. Tagarro, A. *et al.* Early and highly suppressive antiretroviral therapy are main factors associated with low viral reservoir in European perinatally HIV-infected children. in *Journal of Acquired Immune Deficiency Syndromes* vol. 79 269–276 (Lippincott Williams and Wilkins, 2018).
91. F, H. *et al.* Novel Assays for Measurement of Total Cell-Associated HIV-1 DNA and RNA. *J Clin Microbiol* **54**, (2016).
92. Yu, J. J. *et al.* A more precise HIV integration assay designed to detect small differences finds lower levels of integrated DNA in HAART treated patients. *Virology* **379**, 78–86 (2008).
93. Archin, N. M. *et al.* Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature* **487**, 482–485 (2012).
94. Shan, L. *et al.* A Novel PCR Assay for Quantification of HIV-1 RNA. *J Virol* **87**, 6521–6525 (2013).
95. Nakashima, H. *et al.* Inhibition of replication and cytopathic effect of human T cell lymphotropic virus type III/lymphadenopathy-associated virus by 3'-azido-3'-deoxythymidine in vitro. *Antimicrob Agents Chemother* **30**, 933–7 (1986).
96. Prokofjeva, M. M., Kochetkov, S. N. & Prassolov, V. S. Therapy of HIV Infection: Current Approaches and Prospects. *Acta Naturae* **8**, 23–32 (2016).
97. Arts, E. J. & Hazuda, D. J. HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med* **2**, a007161 (2012).
98. Kantor, R. *et al.* Evolution of resistance to drugs in HIV-1-infected patients failing antiretroviral therapy. *AIDS* **18**, 1503–11 (2004).
99. Zulfiqar, H. F. *et al.* HIV Diagnosis and Treatment through Advanced Technologies. *Front Public Health* **5**, 32 (2017).
100. Beck, B. J., Freudenreich, O. & Worth, J. L. Patients with Human Immunodeficiency Virus Infection and Acquired Immunodeficiency Syndrome. *Massachusetts General*

Hospital Handbook of General Hospital Psychiatry 353–370 (2010) doi:10.1016/B978-1-4377-1927-7.00026-1.
101. Smith, R. L., de Boer, R., Brul, S., Budovskaya, Y. & van Spek, H. Premature and accelerated aging: HIV or HAART? *Front Genet* **3**, 328 (2012).
102. Liu, C., Ma, X., Liu, B., Chen, C. & Zhang, H. HIV-1 functional cure: will the dream come true? *BMC Med* **13**, 284 (2015).
103. Martin, A. R. & Siliciano, R. F. Progress Toward HIV Eradication: Case Reports, Current Efforts, and the Challenges Associated with Cure. *Annu Rev Med* **67**, 215–228 (2016).
104. Pham, H. T. & Mesplède, T. The latest evidence for possible HIV-1 curative strategies. *Drugs Context* **7**, 212522 (2018).
105. Liu, B. *et al.* Chimeric Antigen Receptor T Cells Guided by the Single-Chain Fv of a Broadly Neutralizing Antibody Specifically and Effectively Eradicate Virus Reactivated from Latency in CD4+ T Lymphocytes Isolated from HIV-1-Infected Individuals Receiving Suppressive Combined Antiretroviral Therapy. *J Virol* **90**, 9712–9724 (2016).
106. Deeks, S. G. Shock and kill. *Nature* **487**, 439–440 (2012).
107. Pauza, C. D., Huang, K. & Bordon, J. Advances in cell and gene therapy for HIV disease: it is good to be specific. *Curr Opin HIV AIDS* **16**, 83–87 (2021).
108. Abana, C. Z. Y., Lamptey, H., Bonney, E. Y. & Kyei, G. B. HIV cure strategies: which ones are appropriate for Africa? *Cellular and Molecular Life Sciences 2022 79:8* **79**, 1–16 (2022).
109. Jimmy Yeh, Y. H. & Ho, Y. C. Shock-and-kill versus block-and-lock: Targeting the fluctuating and heterogeneous HIV-1 gene expression. *Proc Natl Acad Sci U S A* **118**, e2103692118 (2021).
110. Lichterfeld, M. Reactivation of latent HIV moves shock-and-kill treatments forward. *Nature 2021 578:7793* **578**, 42–43 (2020).
111. Vansant, G., Bruggemans, A., Janssens, J. & Debyser, Z. Block-And-Lock Strategies to Cure HIV Infection. *Viruses* **12**, (2020).
112. Ward, A. R., Mota, T. M. & Jones, R. B. Immunological approaches to HIV cure. *Semin Immunol* **51**, 101412 (2021).
113. Hsu, D. C., Mellors, J. W. & Vasan, S. Can Broadly Neutralizing HIV-1 Antibodies Help Achieve an ART-Free Remission? *Front Immunol* **12**, 2771 (2021).
114. Spencer, D. A., Shapiro, M. B., Haigwood, N. L. & Hessell, A. J. Advancing HIV Broadly Neutralizing Antibodies: From Discovery to the Clinic. *Front Public Health* **9**, 610 (2021).
115. Miliotou, A. N. & Papadopoulou, L. C. CAR T-cell Therapy: A New Era in Cancer Immunotherapy. *Curr Pharm Biotechnol* **19**, 5–18 (2018).
116. Ng'uni, T., Chasara, C. & Ndhlovu, Z. M. Major Scientific Hurdles in HIV Vaccine Development: Historical Perspective and Future Directions. *Front Immunol* **11**, 2761 (2020).
117. Ding, J., Liu, Y. & Lai, Y. Knowledge From London and Berlin: Finding Threads to a Functional HIV Cure. *Front Immunol* **12**, 1852 (2021).
118. McMichael, A. J. & Rowland-Jones, S. L. Cellular immune responses to HIV. *Nature 2001 410:6831* **410**, 980–987 (2001).
119. McMichael, A. J., Borrow, P., Tomaras, G. D., Goonetilleke, N. & Haynes, B. F. The immune response during acute HIV-1 infection: Clues for vaccine development. *Nature Reviews Immunology* vol. 10 11–23 Preprint at https://doi.org/10.1038/nri2674 (2010).

120. Borrow, P. Innate immunity in acute HIV-1 infection. *Curr Opin HIV AIDS* **6**, 353–63 (2011).
121. Chakrabarti, L. A. & Simon, V. Immune mechanisms of HIV control. *Current Opinion in Immunology* vol. 22 488–496 Preprint at https://doi.org/10.1016/j.coi.2010.06.006 (2010).
122. Murray, P. J. & Wynn, T. A. Protective and pathogenic functions of macrophage subsets. *Nat Rev Immunol* **11**, 723 (2011).
123. Vivier, E., Tomasello, E., Baratin, M., Walzer, T. & Ugolini, S. Functions of natural killer cells. *Nature Immunology 2008 9:5* **9**, 503–510 (2008).
124. Song, L., Dong, G., Guo, L. & Graves, D. T. The function of dendritic cells in modulating the host response. *Mol Oral Microbiol* **33**, 13–21 (2018).
125. Carrington, M. & Alter, G. Innate Immune Control of HIV. *Cold Spring Harb Perspect Med* **2**, (2012).
126. Alter, G. *et al.* HLA Class I Subtype-Dependent Expansion of KIR3DS1+ and KIR3DL1+ NK Cells during Acute Human Immunodeficiency Virus Type 1 Infection. *J Virol* **83**, 6798 (2009).
127. Lodoen, M. B. & Lanier, L. L. Viral modulation of NK cell immunity. *Nat Rev Microbiol* **3**, 59–69 (2005).
128. Ackerman, A. L. & Cresswell, P. Cellular mechanisms governing cross-presentation of exogenous antigens. *Nat Immunol* **5**, 678–684 (2004).
129. Murray, P. J. & Wynn, T. A. Protective and pathogenic functions of macrophage subsets. *Nat Rev Immunol* **11**, 723 (2011).
130. Koppensteiner, H., Brack-Werner, R. & Schindler, M. Macrophages and their relevance in Human Immunodeficiency Virus Type I infection. *Retrovirology* **9**, 1–11 (2012).
131. Martín-Moreno, A. & Muñoz-Fernández, M. A. Dendritic cells, the double agent in the war against hiv-1. *Front Immunol* **10**, 2485 (2019).
132. Wu, L. & KewalRamani, V. N. Dendritic-cell interactions with HIV: infection and viral dissemination. *Nature Reviews Immunology 2006 6:11* **6**, 859–868 (2006).
133. Althuwaiqeb, S. A. & Bordoni, B. Histology, B Cell Lymphocyte. *StatPearls* (2023).
134. Lebien, T. W. & Tedder, T. F. B lymphocytes: how they develop and function. *Blood* **112**, 1570 (2008).
135. Moir, S. *et al.* B cells in early and chronic HIV infection: evidence for preservation of immune function associated with early initiation of antiretroviral therapy. *Blood* **116**, 5571–5579 (2010).
136. Moir, S. & Fauci, A. S. B cells in HIV infection and disease. *Nat Rev Immunol* **9**, 235 (2009).
137. Actor, J. K. *Introductory Immunology*. *Introductory Immunology* (Elsevier, 2014). doi:10.1016/c2013-0-09949-5.
138. Kumar, B. V., Connors, T. J. & Farber, D. L. Human T Cell Development, Localization, and Function throughout Life. *Immunity* vol. 48 202–213 Preprint at https://doi.org/10.1016/j.immuni.2018.01.007 (2018).
139. Pennock, N. D. *et al.* T cell responses: naive to memory and everything in between. *Adv Physiol Educ* **37**, 273–83 (2013).
140. Farber, D. L. Form and function for T cells in health and disease. *Nat Rev Immunol* **20**, 83–84 (2020).
141. Chaplin, D. D. Overview of the immune response. *Journal of Allergy and Clinical Immunology* **125**, S3 (2010).
142. Tanel, A. *et al.* Cellular and molecular mechanisms of memory T-cell survival. *Expert Rev Vaccines* **8**, 299 (2009).

143. Pennock, N. D. *et al.* T cell responses: naïve to memory and everything in between. *Adv Physiol Educ* **37**, 273 (2013).
144. Gray, J. I., Westerhof, L. M. & MacLeod, M. K. L. The roles of resident, central and effector memory CD4 T-cells in protective immunity following infection or vaccination. *Immunology* **154**, 574–581 (2018).
145. Swain, S. L., McKinstry, K. K. & Strutt, T. M. Expanding roles for CD4+ T cells in immunity to viruses. *Nature Reviews Immunology 2012 12:2* **12**, 136–148 (2012).
146. Bartolo, V. Di. Cell Biology of T Cell Receptor Expression and Regulation. (2018).
147. Wucherpfennig, K. W., Gagnon, E., Call, M. J., Huseby, E. S. & Call, M. E. Structural Biology of the T-cell Receptor : Insights into Receptor Assembly , Ligand Recognition , and Initiation of Signaling. 1–14 (2010).
148. Clambey, E. T., Davenport, B., Kappler, J. W., Marrack, P. & Homann, D. Molecules in medicine mini review: the αβ T cell receptor. *J Mol Med (Berl)* **92**, 735–41 (2014).
149. Matos, T. R., de Rie, M. A. & Teunissen, M. B. M. Research Techniques Made Simple: High-Throughput Sequencing of the T-Cell Receptor. *Journal of Investigative Dermatology* **137**, e131–e138 (2017).
150. Pardoll, D. M., Fowlkes, B. J., Lechler, R. I., Germain, R. N. & Schwartz, R. H. Early genetic events in T cell development analyzed by in situ hybridization. *J Exp Med* **165**, 1624 (1987).
151. Tanno, H. *et al.* Determinants governing T cell receptor α/β-chain pairing in repertoire formation of identical twins. *Proc Natl Acad Sci U S A* **117**, 532–540 (2020).
152. Keşmir, C., Borghans, J. A. M. & Boer, R. J. de. Diversity of Human αβ T Cell Receptors. *Science (1979)* **288**, 1135–1135 (2000).
153. Charles A Janeway, J., Travers, P., Walport, M. & Shlomchik, M. J. T-cell receptor gene rearrangement. (2001).
154. Heather, J. M., Ismail, M., Oakes, T. & Chain, B. High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. *Brief Bioinform* **19**, 554–565 (2018).
155. Bianconi, E. *et al.* An estimation of the number of cells in the human body. *Ann Hum Biol* **40**, 463–471 (2013).
156. Heather, J. M. *et al.* Dynamic Perturbations of the T-Cell Receptor Repertoire in Chronic HIV Infection and following Antiretroviral. **6**, 1–15 (2016).
157. Aavani, P. & Allen, L. J. S. The role of CD4 T cells in immune system activation and viral reproduction in a simple model for HIV infection. *Appl Math Model* **75**, 210–222 (2019).
158. Buggert, M., Japp, A. S. & Betts, M. R. Everything in its right place: resident memory CD8+ T cell immunosurveillance of HIV infection. *Curr Opin HIV AIDS* **14**, 93–99 (2019).
159. Huang, S. *et al.* Latent HIV reservoirs exhibit inherent resistance to elimination by CD8 + T cells. **128**, (2018).
160. Goepfert, P. Understanding the CD8 T-cell response in natural HIV control [version 1 ; referees : 3 approved] Sushma Boppana Referee Status : **7**, (2018).
161. Rinaldi, S. *et al.* Impact of Early Antiretroviral Therapy Initiation on HIV-Specific CD4 and CD8 T Cell Function in Perinatally Infected Children. *The Journal of Immunology* **204**, 540–549 (2020).
162. Martin, E. *et al.* Stepwise development of mait cells in mouse and human. *PLoS Biol* **7**, 0525–0536 (2009).
163. Hinks, T. S. C. & Zhang, X.-W. MAIT Cell Activation and Functions. *Front Immunol* **11**, 1014 (2020).

164. Lal, K. G. *et al.* Dynamic MAIT cell response with progressively enhanced innateness during acute HIV-1 infection. *Nat Commun* **11**, 1–13 (2020).
165. Godfrey, D. I., Koay, H. F., McCluskey, J. & Gherardin, N. A. The biology and functional importance of MAIT cells. *Nature Immunology* vol. 20 1110–1128 Preprint at https://doi.org/10.1038/s41590-019-0444-8 (2019).
166. Toubal, A., Nel, I., Lotersztajn, S. & Lehuen, A. Mucosal-associated invariant T cells and disease. *Nature Reviews Immunology 2019 19:10* **19**, 643–657 (2019).
167. Gherardin, N. A. *et al.* Human blood MAIT cell subsets defined using MR1 tetramers. *Immunol Cell Biol* **96**, 507–525 (2018).
168. Leeansyah, E. *et al.* Activation, exhaustion, and persistent decline of the antimicrobial MR1-restricted MAIT-cell population in chronic HIV-1 infection. *Blood* **121**, 1124–35 (2013).
169. Spaan, M. *et al.* Frequencies of circulating MAIT cells are diminished in chronic hCV, HIV and HCV/ HIV Co-Infection and do not recover during therapy. *PLoS One* **11**, e0159243 (2016).
170. Le Bourhis, L. *et al.* Antimicrobial activity of mucosal-associated invariant T cells. *Nat Immunol* **11**, 701–708 (2010).
171. Saeidi, A. *et al.* Functional role of mucosal-associated invariant T cells in HIV infection. *J Leukoc Biol* **100**, 305–314 (2016).
172. Martinez, D. R., Permar, S. R. & Fouda, G. G. Contrasting Adult and Infant Immune Responses to HIV Infection and Vaccination. *Clinical and Vaccine Immunology* vol. 23 84–94 Preprint at https://doi.org/10.1128/CVI.00565-15 (2016).
173. Rudolph, M. E. *et al.* Differences between pediatric and adult T Cell Responses to in vitro staphylococcal enterotoxin B stimulation. *Front Immunol* **9**, (2018).
174. Saule, P. *et al.* Accumulation of memory T cells from childhood to old age: central and effector memory cells in CD4(+) versus effector memory and terminally differentiated memory cells in CD8(+) compartment. *Mech Ageing Dev* **127**, 274–281 (2006).
175. Sabin, C. Response to combination antiretroviral therapy: Variation by age. *AIDS* **22**, 1463–1473 (2008).
176. Turkova, A. *et al.* Dolutegravir as First- or Second-Line Treatment for HIV-1 Infection in Children. *New England Journal of Medicine* **385**, 2531–2543 (2021).
177. Nalwanga, D. & Musiime, V. Children living with HIV: a narrative review of recent advances in pediatric HIV research and their implications for clinical practice. *Ther Adv Infect Dis* **9**, (2022).
178. Eisen, S. *et al.* B-cell development and pneumococcal immunity in vertically acquired HIV infection. *AIDS* **30**, 1867–1876 (2016).
179. Lewis, J. *et al.* Thymic Output and CD4 T-Cell Reconstitution in HIV-Infected Children on Early and Interrupted Antiretroviral Treatment: Evidence from the Children with HIV Early Antiretroviral Therapy Trial. *Front Immunol* **8**, (2017).
180. Bunupuradah, T. *et al.* Outcomes after reinitiating antiretroviral therapy in children randomized to planned treatment interruptions. *AIDS* **27**, 579–589 (2013).
181. Klein, N. *et al.* The immunological and virological consequences of planned treatment interruptions in children with HIV infection. *PLoS One* **8**, e76582 (2013).
182. Howie, S. R. Blood sample volumes in child health research: review of safe limits. *Bull World Health Organ* **89**, 46–53 (2011).
183. Bull, M. *et al.* Defining blood processing parameters for optimal detection of cryopreserved antigen-specific responses for HIV vaccine trials. *J Immunol Methods* **322**, 57–69 (2007).

184. Böyum, A. Isolation of leucocytes from human blood. Further observations. Methylcellulose, dextran, and ficoll as erythrocyteaggregating agents. *Scandinavian Journal of Clinical and Laboratory Investigation, Supplement* **97**, 31–50 (1968).
185. Wohnhaas, C. T. *et al.* DMSO cryopreservation is the method of choice to preserve cells for droplet-based single-cell RNA sequencing. *Scientific Reports 2019 9:1* **9**, 1–14 (2019).
186. Chaytor, J. L. *et al.* Inhibiting ice recrystallization and optimization of cell viability after cryopreservation. *Glycobiology* **22**, 123–133 (2012).
187. Kartberg, A.-J. *et al.* Vitrification with DMSO protects embryo membrane integrity better than solutions without DMSO. **17**, 378–384 (2008).
188. Mallone, R. *et al.* Isolation And Preservation Of Peripheral Blood Mononuclear cells for analysis of islet antigen-reactive T cell responses: Position statement of the T-Cell Workshop Committee of the Immunology of Diabetes Society. *Clinical and Experimental Immunology* vol. 163 33–49 Preprint at https://doi.org/10.1111/j.1365-2249.2010.04272.x (2011).
189. Adan, A., Alizada, G., Kiraz, Y., Baran, Y. & Nalbant, A. *Flow cytometry: basic principles and applications*. *Critical Reviews in Biotechnology* vol. 37 163–176 (Taylor and Francis Ltd, 2017).
190. Manohar, S. M., Shah, P. & Nair, A. Flow cytometry: Principles, applications and recent advances. *Bioanalysis* **13**, 185–202 (2021).
191. Robinson, J. P. Flow cytometry: past and future. *Biotechniques* **72**, 159–169 (2022).
192. Adan, A., Alizada, G., Kiraz, Y., Baran, Y. & Nalbant, A. Flow cytometry: basic principles and applications. *https://doi.org/10.3109/07388551.2015.1128876* **37**, 163–176 (2016).
193. Schmit, T., Klomp, M. & Khan, M. N. The Application of Flow Cytometry for Simultaneous and Multi-parametric Analysis of Heterogenous Cell Populations in Basic and Clinical Research. *Methods in Molecular Biology* **2223**, 183–200 (2021).
194. Cossarizza, A. *et al.* Guidelines for the use of flow cytometry and cell sorting in immunological studies*. *Eur J Immunol* **47**, 1584–1797 (2017).
195. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res* **6**, 986–994 (1996).
196. Roy, J., Jain, N., Singh, G., Das, B. & Mallick, B. Small RNA proteome as disease biomarker: An incognito treasure of clinical utility. in *AGO-Driven Non-Coding RNAs* 101–136 (Elsevier, 2019). doi:10.1016/b978-0-12-815669-8.00005-1.
197. Ledderose, C., Heyn, J., Limbeck, E. & Kreth, S. Selection of reliable reference genes for quantitative real-time PCR in human T cells and neutrophils. *BMC Res Notes* **4**, 427 (2011).
198. Dheda, K. *et al.* Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* **37**, 112–119 (2004).
199. Behjati, S. & Tarpey, P. S. What is next generation sequencing? *Arch Dis Child Educ Pract Ed* **98**, 236 (2013).
200. Fuller, C. W. *et al.* The challenges of sequencing by synthesis. *Nature Biotechnology 2009 27:11* **27**, 1013–1023 (2009).
201. Chen, C. Y. DNA polymerases drive DNA sequencing-by-synthesis technologies: Both past and present. *Front Microbiol* **5**, 305 (2014).
202. Oakes, T. *et al.* Quantitative characterization of the T cell receptor repertoire of naïve and memory subsets using an integrated experimental and computational pipeline which is robust, economical, and versatile. *Front Immunol* **8**, 1–17 (2017).

203. Gkazi, A. S. *et al.* Clinical T Cell Receptor Repertoire Deep Sequencing and Analysis: An Application to Monitor Immune Reconstitution Following Cord Blood Transplantation. *Front Immunol* **9**, 2547 (2018).
204. Sandgaard, K. S. *et al.* Plasticity of the Immune System in Children Following Treatment Interruption in HIV-1 Infection. *Front Immunol* **0**, 3098 (2021).
205. Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017).
206. Joshi, K. *et al.* Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. *Nat Med* **25**, 1549 (2019).
207. Kuhn, L. *et al.* Age at antiretroviral therapy initiation and cell-associated HIV-1 DNA levels in HIV-1-infected children. *PLoS One* **13**, e0195514 (2018).
208. Ometto, L. *et al.* Immune reconstitution in HIV-1-infected children on antiretroviral therapy: Role of thymic output and viral fitness. *AIDS* **16**, 839–849 (2002).
209. Saitoh, A. *et al.* Persistence of human immunodeficiency virus (HIV) type 1 DNA in peripheral blood despite prolonged suppression of plasma HIV-1 RNA in children. *Journal of Infectious Diseases* **185**, 1409–1416 (2002).
210. Zanchetta, M. *et al.* Long-term decay of the HIV-1 reservoir in HIV-1-infected children treated with highly active antiretroviral therapy. *Journal of Infectious Diseases* **193**, 1718–1727 (2006).
211. Uprety, P. *et al.* Cell-Associated HIV-1 DNA and RNA Decay Dynamics during Early Combination Antiretroviral Therapy in HIV-1-Infected Infants. *Clinical Infectious Diseases* **61**, 1862–1870 (2015).
212. Zangari, P. *et al.* Report from the First EPIICAL (Early-treated Perinatally HIV-infected Individuals: Improving Children's Actual Life with Novel Immunotherapeutic Strategies) General Assembly meeting, 9-11 November 2017, Rome, Italy. *J Virus Erad* **4**, 51–54 (2018).
213. Pasternak, A. O. *et al.* Cellular levels of HIV unspliced RNA from patients on combination antiretroviral therapy with undetectable plasma viremia predict the therapy outcome. *PLoS One* **4**, (2009).
214. Pitman, M. C., Lau, J. S. Y., McMahon, J. H. & Lewin, S. R. Barriers and strategies to achieve a cure for HIV. *The Lancet HIV* vol. 5 e317–e328 Preprint at https://doi.org/10.1016/S2352-3018(18)30039-0 (2018).
215. Chun, T. W. *et al.* Early establishment of a pool of latently infected, resting CD4+ T cells during primary HIV-1 infection. *Proc Natl Acad Sci U S A* **95**, 8869–8873 (1998).
216. Chun, T. W., Moir, S. & Fauci, A. S. HIV reservoirs as obstacles and opportunities for an HIV cure. *Nature Immunology* vol. 16 584–589 Preprint at https://doi.org/10.1038/ni.3152 (2015).
217. Shrivastava & Gupta, V. B. Methods for the determination of limit of detection and limit of quantitation of the analytical methods. *Chronicles of Young Scientists* **2**, 21 (2011).
218. Sáez-Cirión, A. *et al.* Post-Treatment HIV-1 Controllers with a Long-Term Virological Remission after the Interruption of Early Initiated Antiretroviral Therapy ANRS VISCONTI Study. *PLoS Pathog* **9**, (2013).
219. Luzuriaga, K. *et al.* HIV Type 1 (HIV-1) Proviral Reservoirs Decay Continuously Under Sustained Virologic Control in HIV-1–Infected Children Who Received Early Treatment. *Journal of Infectious Diseases* **210**, 1529–1538 (2014).
220. van Zyl, G. U. *et al.* Early Antiretroviral Therapy in South African Children Reduces HIV-1-Infected Cells and Cell-Associated HIV-1 RNA in Blood Mononuclear Cells. *Journal of Infectious Diseases* **212**, 39–43 (2015).

221. Sonza, S. *et al.* Selectively Reduced tat mRNA Heralds the Decline in Productive Human Immunodeficiency Virus Type 1 Infection in Monocyte-Derived Macrophages. *J Virol* **76**, 12611–12621 (2002).
222. Pasternak, A. O. *et al.* Cell-associated HIV-1 RNA predicts viral rebound and disease progression after discontinuation of temporary early ART. *JCI Insight* **5**, (2020).
223. Kiselinova, M., Pasternak, A. O., Spiegelaere, W. De, Vogelaers, D. & Berkhout, B. Comparison of Droplet Digital PCR and Seminested Real- Time PCR for Quantification of Cell-Associated HIV-1 RNA. **9**, 1–8 (2014).
224. Pasternak, A. O. *et al.* Cell-associated HIV-1 RNA predicts viral rebound and disease progression after discontinuation of temporary early ART. *JCI Insight* **5**, (2020).
225. Uprety, P. *et al.* Cell-Associated HIV-1 DNA and RNA Decay Dynamics during Early Combination Antiretroviral Therapy in HIV-1-Infected Infants. *Clinical Infectious Diseases* **61**, 1862–1870 (2015).
226. Rajasuriar, R., Wright, E. & Lewin, S. R. Impact of antiretroviral therapy (ART) timing on chronic immune activation/inflammation and end-organ damage. *Current Opinion in HIV and AIDS* vol. 10 35–42 Preprint at https://doi.org/10.1097/COH.0000000000000118 (2015).
227. Macatangay, B. J. C. & Rinaldo, C. R. Preserving HIV-specific T cell responses: Does timing of antiretroviral therapy help? *Current Opinion in HIV and AIDS* vol. 10 55–60 Preprint at https://doi.org/10.1097/COH.0000000000000124 (2015).
228. Conrad, J. A. *et al.* Antiretroviral Therapy Reduces the Magnitude and T Cell Receptor Repertoire Diversity of HIV-Specific T Cell Responses without Changing T Cell Clonotype Dominance. *J Virol* **86**, 4213–4221 (2012).
229. Turner, C. T. *et al.* Persistent T Cell Repertoire Perturbation and T Cell Activation in HIV After Long Term Treatment. *Front Immunol* **12**, (2021).
230. Palma, P. *et al.* The EPIICAL project: an emerging global collaboration to investigate immunotherapeutic strategies in HIV-infected children. *J Virus Erad* **1**, 134–139 (2015).
231. Cotton, M. F. *et al.* Early time-limited antiretroviral therapy versus deferred therapy in South African infants infected with HIV: Results from the children with HIV early antiretroviral (CHER) randomised trial. *The Lancet* **382**, 1555–1563 (2013).
232. Azzoni, L. *et al.* Early ART Results in Greater Immune Reconstitution Benefits in HIV-Infected Infants: Working with Data Missingness in a Longitudinal Dataset. *PLoS One* **10**, (2015).
233. Berhan, Y. Age and CD4 count of vertically HIV-infected children at the time of diagnosis: what are independent predictors for being symptomatic and CD4 counts drop? *J Trop Pediatr* **57**, 14–23 (2011).
234. Garcia, S. A. B. & Guzman, N. Acquired Immune Deficiency Syndrome CD4+ Count. *StatPearls* (2023).
235. McBride, J. A. & Striker, R. Imbalance in the game of T cells: What can the CD4/CD8 T-cell ratio tell us about HIV and health? *PLoS Pathog* **13**, (2017).
236. Tosato, F. *et al.* Lymphocytes subsets reference values in childhood. *Cytometry Part A* **87**, 81–85 (2015).
237. Adams, S. P., Kricke, S., Ralph, E., Gilmour, N. & Gilmour, K. C. A comparison of TRECs and flow cytometry for naive T cell quantification. *Clin Exp Immunol* **191**, 198–202 (2018).
238. Hazenberg, M. D., Verschuren, M. C., Hamann, D., Miedema, F. & Dongen, J. J. T cell receptor excision circles as markers for recent thymic emigrants: basic aspects, technical approach, and guidelines for interpretation. *J Mol Med (Berl)* **79**, 631–640 (2001).

239. Somech, R. T-cell receptor excision circles in primary immunodeficiencies and other T-cell immune disorders. *Curr Opin Allergy Clin Immunol* **11**, 517–524 (2011).
240. Levy, A. *et al.* T cell receptor excision circles as a tool for evaluating thymic function in young children. *Brazilian Journal of Medical and Biological Research* **52**, (2019).
241. Thomas, N., Heather, J., Ndifon, W., Shawe-Taylor, J. & Chain, B. Decombinator: A tool for fast, efficient gene assignment in T-cell receptor sequences using a finite state machine. *Bioinformatics* **29**, 542–550 (2013).
242. Duez, M. *et al.* Vidjil: A web platform for analysis of high-Throughput repertoire sequencing. *PLoS One* **11**, (2016).
243. Bagaev, D. v. *et al.* VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res* **48**, D1057–D1062 (2020).
244. Kitaura, K., Shini, T., Matsutani, T. & Suzuki, R. A new high-throughput sequencing method for determining diversity and similarity of T cell receptor (TCR) α and β repertoires and identifying potential new invariant TCR α chains. *BMC Immunol* **17**, 1–16 (2016).
245. Davy-Mendez, T. *et al.* Acute HIV infection and CD4/CD8 ratio normalization after antiretroviral therapy initiation. *J Acquir Immune Defic Syndr (1988)* **79**, 510–518 (2018).
246. Attaf, M. *et al.* Cytomegalovirus-Mediated T Cell Receptor Repertoire Perturbation Is Present in Early Life. *Front Immunol* **11**, 1587 (2020).
247. Salam, N. *et al.* T cell ageing: Effects of age on development, survival & function. *Indian J Med Res* **138**, 595 (2013).
248. Douek, D. C. *et al.* Changes in thymic function with age and during the treatment of HIV infection. *Nature 1998 396:6712* **396**, 690–695 (1998).
249. Levy, A. *et al.* T cell receptor excision circles as a tool for evaluating thymic function in young children. *Brazilian Journal of Medical and Biological Research* **52**, (2019).
250. Bains, I., Thiébaut, R., Yates, A. J. & Callard, R. Quantifying Thymic Export: Combining Models of Naive T Cell Proliferation and TCR Excision Circle Dynamics Gives an Explicit Measure of Thymic Output. *The Journal of Immunology* **183**, 4329–4336 (2009).
251. de Rossi, A. *et al.* Increased Thymic Output after Initiation of Antiretroviral Therapy in Human Immunodeficiency Virus Type 1-Infected Children in the Paediatric European Network for Treatment of AIDS (PENTA) 5 Trial.
252. Hazenberg, M. D. *et al.* Increased cell division but not thymic dysfunction rapidly affects the T-cell receptor excision circle content of the naive T cell population in HIV-1 infection. *Nat Med* **6**, 1036–1042 (2000).
253. Levy, A. *et al.* T cell receptor excision circles as a tool for evaluating thymic function in young children. *Brazilian Journal of Medical and Biological Research* **52**, (2019).
254. Kwok, J. S. Y. *et al.* Establishing Simultaneous T Cell Receptor Excision Circles (TREC) and K-Deleting Recombination Excision Circles (KREC) Quantification Assays and Laboratory Reference Intervals in Healthy Individuals of Different Age Groups in Hong Kong. *Front Immunol* **11**, 1411 (2020).
255. Farris, F. A. The gini index and measures of inequality. *American Mathematical Monthly* **117**, 851–864 (2010).
256. Rosati, E. *et al.* Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol* **17**, (2017).
257. Mora, T. & Walczak, A. M. How many different clonotypes do immune repertoires contain? *Curr Opin Syst Biol* **18**, 104–110 (2019).

258. Foth, S., Völkel, S., Bauersachs, D., Zemlin, M. & Skevaki, C. T Cell Repertoire During Ontogeny and Characteristics in Inflammatory Disorders in Adults and Childhood. *Front Immunol* **0**, 3826 (2021).
259. Thomas, P. G. & Crawford, J. C. Selected before selection: A case for inherent antigen bias in the T-cell receptor repertoire. *Current Opinion in Systems Biology* vol. 18 36–43 Preprint at https://doi.org/10.1016/j.coisb.2019.10.007 (2019).
260. Elhanati, Y. *et al.* Predicting the spectrum of TCR repertoire sharing with a data-driven model of recombination. *Immunol Rev* **284**, 167 (2018).
261. Chronister, W. D. *et al.* TCRMatch: Predicting T-Cell Receptor Specificity Based on Sequence Similarity to Previously Characterized Receptors. *Front Immunol* **12**, 673 (2021).
262. Vujovic, M. *et al.* T cell receptor sequence clustering and antigen specificity. *Comput Struct Biotechnol J* **18**, 2166 (2020).
263. Redmond, D., Poran, A. & Elemento, O. Single-cell TCRseq: Paired recovery of entire T-cell alpha and beta chain transcripts in T-cell receptors from single-cell RNAseq. *Genome Med* **8**, 1 (2016).
264. Saeidi, A. *et al.* Functional role of mucosal-associated invariant T cells in HIV infection. *J Leukoc Biol* **100**, 305–314 (2016).
265. Cosgrove, C. *et al.* Early and nonreversible decrease of CD161++/MAIT cells in HIV infection. *Blood* **121**, 951 (2013).
266. Nalwanga, D. & Musiime, V. Children living with HIV: a narrative review of recent advances in pediatric HIV research and their implications for clinical practice. *https://doi.org/10.1177/20499361221077544* **9**, (2022).
267. Geneva: Joint United Nations Programme on HIV/ & AIDS. IN DANGER: UNAIDS Global AIDS Update 2022. https://www.unaids.org/sites/default/files/media_asset/2022-global-aids-update_en.pdf (2022).
268. Foster, C., Ayers, S. & Fidler, S. Antiretroviral adherence for adolescents growing up with HIV: understanding real life, drug delivery and forgiveness. *Ther Adv Infect Dis* **7**, (2020).
269. Schomaker, M. *et al.* When to Start Antiretroviral Therapy in Children Aged 2–5 Years: A Collaborative Causal Modelling Analysis of Cohort Studies from Southern Africa. *PLoS Med* **10**, e1001555 (2013).
270. Frange, P. *et al.* Impact of Early Versus Late Antiretroviral Treatment Initiation on Naive T Lymphocytes in HIV-1-Infected Children and Adolescents – The-ANRS-EP59-CLEAC Study. *Front Immunol* **12**, 1202 (2021).
271. Simms, V. *et al.* Growth improvement following antiretroviral therapy initiation in children with perinatally-acquired HIV diagnosed in older childhood in Zimbabwe: a prospective cohort study. *BMC Pediatr* **22**, 1–13 (2022).
272. Zanchetta, M. *et al.* Early therapy in HIV-1-infected children: Effect on HIV-1 dynamics and HIV-1-specific immune response. *Antivir Ther* **13**, 47–55 (2008).
273. Picat, M. Q. *et al.* Predicting patterns of long-term CD4 reconstitution in HIV-infected children starting antiretroviral therapy in sub-Saharan Africa: a cohort-based modelling study. *PLoS Med* **10**, (2013).
274. Matthews, K. K. *et al.* VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res* **46**, D419–D427 (2017).
275. Palmer, D. B. The effect of age on thymic function. *Front Immunol* **4**, 316 (2013).
276. Lee, C. H. *et al.* Predicting Cross-Reactivity and Antigen Specificity of T Cell Receptors. *Front Immunol* **11**, 2498 (2020).

277. Cosgrove, C. *et al.* Early and nonreversible decrease of CD161++ /MAIT cells in HIV infection. *Blood* **121**, 951–961 (2013).
278. Flepp, M., Schiffer, V., Weber, R. & Hirschel, B. Modern anti-HIV therapy. *Swiss Medical Weekly* vol. 131 207–213 Preprint at https://doi.org/10.4414/smw.2001.06132 (2001).
279. D'Arminio Monforte, A. *et al.* The changing incidence of AIDS events in patients receiving highly active antiretroviral therapy. *Arch Intern Med* **165**, 416–423 (2005).
280. Carr, A., Miller, J., Matthew, L. & Cooper, D. A. A syndrome of lipoatrophy, lactic acidaemia and liver dysfunction associated with HIV nucleoside analogue therapy: Contribution to protease inhibitor-related lipodystrophy syndrome. *AIDS* **14**, (2000).
281. Group, T. D. C. on A. E. of A.-H. D. (DAD) S. Combination Antiretroviral Therapy and the Risk of Myocardial Infarction. *New England Journal of Medicine* **349**, 1993–2003 (2003).
282. Clavel, F. & Hance, A. J. HIV Drug Resistance. *New England Journal of Medicine* **350**, 1023–1035 (2004).
283. Group, T. S. for M. of A. T. (SMART) S. CD4+ Count–Guided Interruption of Antiretroviral Treatment. *New England Journal of Medicine* **355**, 2283–2296 (2006).
284. Wamalwa, D. *et al.* Treatment interruption after 2-year antiretroviral treatment initiated during acute/early HIV in infancy. *AIDS* **30**, 2303–2313 (2016).
285. Butler, K. *et al.* BREATHER (PENTA 16) short-cycle therapy (SCT) (5 days on/2 days off) in young people with chronic human immunodeficiency virus infection: An open, randomised, parallel-group phase II/III trial. *Health Technol Assess (Rockv)* **20**, 1–107 (2016).
286. Imaz, A. *et al.* Short-term and long-term clinical and immunological consequences of stopping antiretroviral therapy in HIV-infected patients with preserved immune function. *Antivir Ther* **18**, 125–130 (2013).
287. Siberry, G. K. *et al.* CD4+ Lymphocyte-Based Immunologic Outcomes of Perinatally HIV-Infected Children During Antiretroviral Therapy Interruption. *JAIDS Journal of Acquired Immune Deficiency Syndromes* **57**, 223–229 (2011).
288. Goujard, C. *et al.* Continuous versus intermittent treatment strategies during primary HIV-1 infection: The randomized ANRS INTERPRIM Trial. *AIDS* **26**, 1895–1905 (2012).
289. Hirschel, B. Planned interruptions of anti-HIV treatment. *Lancet Infectious Diseases* vol. 1 53–59 Preprint at https://doi.org/10.1016/S1473-3099(01)00022-6 (2001).
290. Latif, A. *et al.* Fixed duration interruptions are inferior to continuous treatment in African adults starting therapy with CD4 cell counts < 200 cells/µl. *AIDS* **22**, 237–247 (2008).
291. Danel, C. *et al.* CD4-guided structured antiretroviral treatment interruption strategy in HIV-infected adults in west Africa (Trivacan ANRS 1269 trial): a randomised trial. *Lancet* **367**, 1981–1989 (2006).
292. Ananworanich, J. *et al.* CD4-guided scheduled treatment interruptions compared with continuous therapy for patients infected with HIV-1: results of the Staccato randomised trial. *Lancet* **368**, 459–465 (2006).
293. Gibb, D. M. *et al.* Immune repopulation after HAART in previously untreated HIV-1-infected children. *Lancet* **355**, 1331–1332 (2000).
294. De Rossi, A. *et al.* Increased thymic output after initiation of antiretroviral therapy in human immunodeficiency virus type 1-infected children in the Paediatric European Network for Treatment of AIDS (PENTA) 5 trial. *Journal of Infectious Diseases* **186**, 312–320 (2002).

295. Castro, H. Response to planned treatment interruptions in HIV infection varies across childhood. *AIDS* **24**, 231–241 (2010).
296. Paediatric European Network for Treatment of AIDS. Response to planned treatment interruptions in HIV infection varies across childhood. *AIDS* **24**, 231–241 (2010).
297. Freguja, R. *et al.* Long-term clinical, virological and immunological outcomes following planned treatment interruption in HIV-infected children. *HIV Med* hiv.12986 (2020) doi:10.1111/hiv.12986.
298. Okoye, A. A. & Picker, L. J. CD4(+) T-cell depletion in HIV infection: mechanisms of immunological failure. *Immunol Rev* **254**, 54–64 (2013).
299. Song, C. B. *et al.* CD4+CD38+ central memory T cells contribute to HIV persistence in HIV-infected individuals on long-term ART. *J Transl Med* **18**, 95 (2020).
300. Fellay, J. *et al.* Prevalence of adverse events associated with potent antiretroviral treatment: Swiss HIV Cohort Study. *The Lancet* **358**, 1322–1327 (2001).
301. Kiekens, A. *et al.* Exploring the mechanisms behind HIV drug resistance in sub-Saharan Africa: conceptual mapping of a complex adaptive system based on multi-disciplinary expert insights. *BMC Public Health* **22**, 1–15 (2022).
302. Ananworanich, J. *et al.* CD4-guided scheduled treatment interruptions compared with continuous therapy for patients infected with HIV-1: results of the Staccato randomised trial. *The Lancet* **368**, 459–465 (2006).
303. Olmo, M. *et al.* Impact of antiretroviral therapy interruption on plasma biomarkers of cardiovascular risk and lipids: 144-week final data from the STOPAR study. *HIV Med* **13**, 488–498 (2012).
304. Castro, H. Response to planned treatment interruptions in hiv infection varies across childhood. *AIDS* **24**, 231–241 (2010).
305. Freguja, R. *et al.* Long-term clinical, virological and immunological outcomes following planned treatment interruption in HIV-infected children. *HIV Med* **22**, 172 (2021).
306. Group, T. S. for M. of A. T. (SMART) S. CD4+ Count–Guided Interruption of Antiretroviral Treatment. *New England Journal of Medicine* **355**, 2283–2296 (2006).
307. Wamalwa, D. *et al.* Treatment interruption after 2-year antiretroviral treatment (ART) initiated during acute/early HIV in infancy: a randomized trial. *AIDS* **30**, 2303 (2016).
308. Breather Penta, T. & Group, T. Weekends-off efavirenz-based antiretroviral therapy in HIV-infected children, adolescents, and young adults (BREATHER): a randomised, open-label, non-inferiority, phase 2/3 trial. *Lancet HIV* **3**, e421–e430 (2016).
309. Gutierrez, L., Beckford, J. & Alachkar, H. Deciphering the TCR Repertoire to Solve the COVID-19 Mystery. *Trends Pharmacol Sci* **41**, 518 (2020).
310. Farber, D. L., Yudanin, N. A. & Restifo, N. P. Human memory T cells: generation, compartmentalization and homeostasis. *Nat Rev Immunol* **14**, 24 (2014).
311. Su, L. F., Kidd, B. A., Han, A., Kotzin, J. J. & Davis, M. M. Virus-specific CD4+ memory phenotype T cells are abundant in unexposed adults. *Immunity* **38**, 373 (2013).
312. Terpstra, M. L. *et al.* Tissue-resident mucosal-associated invariant T (MAIT) cells in the human kidney represent a functionally distinct subset. *Eur J Immunol* **50**, 1783 (2020).
313. Chen, P. *et al.* Circulating mucosal-associated invariant T Cells in a large cohort of chiniese individuals from newborn to elderly. *Front Immunol* **10**, 260 (2019).
314. Sengupta, S. & Siliciano, R. F. Targeting the latent reservoir for HIV-1. *Immunity* **48**, 872 (2018).
315. Luzuriaga, K. Early Combination Antiretroviral Therapy Limits HIV-1 Persistence in Children. *Annu Rev Med* **67**, 201–213 (2016).

316. Massanella, M., Gianella, S., Lada, S. M., Richman, D. D. & Strain, M. C. Quantification of Total and 2-LTR (Long terminal repeat) HIV DNA, HIV RNA and Herpesvirus DNA in PBMCs. *Bio Protoc* **5**, 1–20 (2015).
317. Kuhn, L. *et al.* Age at antiretroviral therapy initiation and cell-associated HIV-1 DNA levels in HIV-1-infected children. *PLoS One* **13**, e0195514 (2018).
318. Ananworanich, J., Dubé, K. & Chomont, N. How does the timing of antiretroviral therapy initiation in acute infection affect HIV reservoirs? *Curr Opin HIV AIDS* **10**, 18 (2015).
319. de Clercq, J. *et al.* Benefits of antiretroviral therapy initiation during acute HIV infection. *https://doi.org/10.1080/17843286.2020.1770413* **77**, 168–176 (2020).
320. Baum, P. D. *et al.* Blood T-cell receptor diversity decreases during the course of HIV infection, but the potential for a diverse repertoire persists. *Blood* **119**, 3469 (2012).
321. Yin, L. *et al.* Antiretroviral therapy restores diversity in the T-cell receptor Vβ repertoire of CD4 T-cell subpopulations among human immunodeficiency virus type 1-infected children and adolescents. *Clinical and Vaccine Immunology* **16**, 1293–1301 (2009).
322. Douek, D. Thymic Output and HIV Infection: On the Right TREC. *Immunity* **21**, 744–745 (2004).
323. Turner, C. T. *et al.* Persistent T Cell Repertoire Perturbation and T Cell Activation in HIV After Long Term Treatment. *Front Immunol* **12**, 634489 (2021).
324. Towlerton, A. M. H., Ravishankar, S., Coffey, D. G., Puronen, C. E. & Warren, E. H. Serial Analysis of the T-Cell Receptor β-Chain Repertoire in People Living With HIV Reveals Incomplete Recovery After Long-Term Antiretroviral Therapy. *Front Immunol* **13**, 879190 (2022).