Novel Approaches Towards Characterising the HIV Reservoir:

A combined microfluidic and target enrichment approach



Mathew Lee Jones

St Cross College

University of Oxford

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Clinical Medicine

Hilary term 2019

This thesis is dedicated to
Peggy, Rosie, Ken and Judy.

This thesis was supervised by
Professor John Frater and Dr Azim Ansari

Acknowledgements

I would firstly like to express my most sincere thanks to Professor Frater; not only for originally taking me on as a research assistant, but also for assisting in my transfer to DPhil status to complete this thesis. Your guidance, belief in the concept, and at times a tug on the reins have allowed me to pursue a project that has managed to incorporate both the lowest and highest of tech, whilst maintaining a clear sense of the potential benefit to the field. By allowing me the independence to guide the direction of the project, I emerge the other side of this process a more complete (if not slightly grayer) scientist. I would also like to thank Dr Ansari. Despite coming on board later in the process, your guidance regarding the bioinformatics in the project has been invaluable and gave me the confidence to trust the analysis.

I owe a debt of gratitude to all those who have consented to participate in the SPARTAC study, the HEATHER study and those who selflessly continue to do so globally on a day to day basis. It has been a privilege to work in a field with such an active and engaged participant advocacy group. Furthermore, the interactions between the UK CHERUB collaboration and those it seeks to help are to be commended and should serve as a model for other disease areas.

To Lucia, the one rogue physicist in a building full of biologists, I owe a huge thankyou. Your knowledge, help and patience throughout this project has been critical in bringing a number of incredibly technical components together into one flowing pipeline.

For starting us off down the road to droplets, a big thankyou to Professor Aarts and Raam, the launchpad you gave me towards building our microfluidics lab from the ground up has saved us immeasurable time and the collaboration with the Chemistry department is something I hope will continue.

Many faces have come and gone from the Medawar during my time here and most have contributed to my development either personally or professionally. In particular, I would like to thank Helen and Nic for all they do in keeping the lab running on a day to day basis, I honestly think I would only be halfway through this project without you guys around. To the other members of John's group, both past and present, thankyou all. The PMB has always been a great place to work,

with a neverending supply of great minds to bounce ideas off and share a drink with. An extra thanks to Matty P for the cakes!!

To my family and friends, although it has become increasingly obvious that not a single one amongst you has any idea what I do, I know I can always rely on you to stand resolutely behind me while I do it. I thank you.

My final aliquot of thanks is reserved of course for my 'life-associate' Debbie. Your (almost) unwavering support throughout this process has allowed me to pursue a passion that no sensible 36 year old man should really be pursuing. You have had to endure more than could reasonably have been expected of a person over the past four years, but as of today, hopefully, *'normal'* service will be resumed. Oh, and of course Levi. *And Panagiota Zacharopoulou.*

Contributions to this thesis

The majority of the work presented herein is mine. A number of experiments have been carried out by others in the group with my input as detailed below:

Results chapter 1

- The Sureselect bait panel was designed with the assistance of Dr Jacob Hurst.
- The Breakalign script was written by Dr Emanuele Marchi.

Results chapter 2

- The cross contamination and full-length reconstruction pipelines were custom scripts written by Dr Nicholas Grayson.

Results chapter 3

- Alice Russell performed a number of experiments and created a subset of the figures for 4.8, 4.9, 4.11 & 4.12.
- The majority of droplet merging experiments were performed by Dr Lucia Parolini
- Droplet sorting experiments were jointly designed and performed with Dr Lucia Parolini

Results chapter 4

- A subset of the lysis experiments were performed by Dr Lucia Parolini.

Results chapter 5

- The majority of the sorting experiments were performed by Dr Lucia Parolini..

Abstract

The latent reservoir in human immunodeficiency virus (HIV) infection is a small population of long-lived infected cells that remain quiescent during suppressive antiretroviral therapy (ART). This reservoir is widely considered to be the major barrier to HIV cure, as in the absence of effective therapy the cells can become reactivated and cause recrudescent viraemia. A number of assays have been developed in an attempt to both quantify and improve our understanding of the reservoir. However, despite early promise the key factors that contribute to its establishment and persistence remain elusive.

The work undertaken in this thesis was directed primarily at improving the suite of tools available to clinicians and researchers targeting the latent reservoir in HIV infection. To date, the assays that have been designed to quantify and characterise the reservoir have limitations leading to either imprecise measurements or potentially biased analysis of viral sequences. By developing a modified target enrichment protocol, it was shown that diverse full-length proviral sequences can be captured whilst avoiding the bias commonly introduced by PCR-based sequencing methods. In addition, this pipeline allows for the identification of integration sites from the same sample.

The major hurdle facing reservoir assays is the inability to distinguish a latently infected cell from an uninfected cell by cell surface markers. To bypass this requirement, this thesis presents a microfluidics pipeline aimed at using the resident provirus as a means of identifying a latently infected cell. By encapsulating single cells in water-in-oil droplets, it is possible to lyse the cells and perform a PCR reaction targeting the HIV genome. In the presence of HIV, droplets fluoresce and are labelled for dielectrophoretic sorting. Post sorting, nucleic acids can be reclaimed from the droplets for downstream analysis on a purified subset of latently infected cells.

The two assays proposed within this thesis have the potential to profoundly change our understanding of HIV latency. In addition to this, as clinicians enter a new era of treatment interruption studies, changes in reservoir size and composition are becoming a desirable metric.

Overall, the foundation laid down in this thesis provides a solid base from which to begin applying both assays to clinical cohorts.

Contents

List of Figures **xiii**

List of Tables **xvii**

List of abbreviations and acronyms **xix**

1 Introduction **1**

1.1 Virology and epidemiology 2
1.1.1 A current snapshot of the global HIV epidemic 2
1.1.2 HIV structure and lifecycle 3
1.2 The successes of ART and other curative strategies. 8
1.3 The latent HIV reservoir as the barrier to cure. 11
1.3.1 Mechanisms contributing to the maintenance of the reservoir 14
1.4 Reservoir studies and curative strategies. 15
1.5 Current measures of latency 18
1.5.1 Quantitative Viral Outgrowth Assay (qVOA) 18
1.5.2 quantitative PCR (qPCR) Assay 19
1.5.3 Tat/rev induced Limiting Dilution Assay (TILDA) Assay . . 22
1.5.4 Intact proviral DNA assay (IPDA) 23
1.5.5 No single assay is capable of accurately measuring the reservoir 23
1.6 The goals and requirements of HIV provirus sequencing 23
1.6.1 Replication competence 24
1.6.2 Integration site analysis 25
1.6.3 Age of individual proviruses 26
1.7 Droplet microfluidics 26
1.7.1 Microfluidics and emulsions 27
1.7.2 Why the HIV latency field needs droplets 30
1.7.3 Droplet microfluidics revolutionising single-cell analysis . . . 31
1.8 Aims of this thesis 32
1.9 Thesis structure 33

2 Target enrichment method development 35
2.1 Introduction . . . 37
2.1.1 Description of the standard Sureselect protocol . . . 39
2.1.2 Bait design . . . 39
2.2 The standard Sureselect protocol requires modifications to work in the context of HIV proviral sequencing. . . . 40
2.2.1 Initial testing with the standard protocol . . . 41
2.2.2 Increasing input amount and reducing PCR cycles increases sample complexity . . . 48
2.2.3 Unavoidable sample loss occurs during end-repair, a-tailing and adapter ligation . . . 50
2.2.4 Increased hybridisation and wash temperatures improve enrichment efficiency whilst retaining specificity . . . 52
2.2.5 Recommended bait concentration can be significantly reduced to save costs . . . 57
2.3 Bioinformatic pipeline . . . 59
2.3.1 Sequencing read alignment . . . 59
2.3.2 Optimising the integration site pipeline . . . 60
2.3.3 Integration site visualisation . . . 62
2.3.4 Integration site validation . . . 63
2.4 Summary . . . 63
2.4.1 Limitations and further improvements . . . 65

3 Target enrichment results 67
3.1 Context . . . 68
3.2 Sequence quality and cross-contamination assessment . . . 68
3.3 Sequencing depth and coverage . . . 71
3.3.1 New de-novo analysis techniques to maximise HIV read allocation and build full-length contigs. . . . 75
3.4 Integration site identification . . . 77
3.4.1 PCR methods for identifying integration sites and full-length sequencing have the potential for bias. . . . 78
3.5 Summary . . . 80

4 Droplet microfluidics: Droplet production and manipulation method development 81
4.1 Context . . . 81
4.1.1 Lessons from the cleanroom . . . 82
4.2 Method development . . . 82
4.2.1 Microfluidic chip manufacture . . . 82

4.2.2 WO production . . . 85
4.2.3 Water-in-oil-in-water (WOW) droplet formation . . . 90
4.2.4 Wetting in microfluidic devices . . . 96
4.2.5 Chip geometry, flow-rates and surfactant concentration determine obtainable droplet sizes . . . 98
4.2.6 Continuous phase and fluorosurfactant selection . . . 101
4.2.7 Single-cell encapsulation in droplets . . . 104
4.2.8 Diluting lysis reagents within droplets . . . 107
4.2.9 Droplet reinjection . . . 119
4.2.10 Droplet sorting . . . 121
4.3 Summary . . . 135

5 Droplet microfluidics: In-drop PCR **137**
5.1 Introduction . . . 137
5.2 Droplet thermostability . . . 138
5.2.1 Igepal CA-360 as a surprise surfactant . . . 143
5.3 PCR compatible in-droplet lysis . . . 147
5.3.1 PCR reagent optimisation . . . 165

6 Droplet microfluidics: Results **173**
6.1 Context . . . 174
6.2 DNA enrichment pipeline . . . 175
6.2.1 DNA encapsulation . . . 175
6.2.2 DNA droplet thermocycling . . . 176
6.2.3 DNA droplet sorting . . . 178
6.2.4 Nucleic acid quantitation. . . . 178
6.2.5 DNA shearing leads to proviruses being split between droplets. 182
6.2.6 DNA enrichment pipeline discussion . . . 184
6.3 Cell enrichment pipeline testing . . . 185
6.3.1 Testing the full cell enrichment pipeline using uninfected PBMCs. . . . 185
6.3.2 Cell enrichment pipeline discussion . . . 187
6.4 Conclusion . . . 187

7 General discussion **189**
7.1 General discussion . . . 190
7.1.1 Limitations of this work. . . . 192
7.1.2 Implications of this thesis for the HIV cure field and beyond. 194
7.2 Publications . . . 195
7.2.1 Papers in preparation. . . . 195
7.3 Papers to follow . . . 196
7.4 Papers published during the thesis period . . . 196

8 General methods **199**
8.1 Cell lines . . . 199
8.2 General techniques for ex-vivo sample handling . . . 200
8.2.1 Processing and cryopreservation of blood samples . . . 200
8.2.2 CD4+ T-cell enrichment . . . 201
8.3 Nucleic acid extraction . . . 201
8.4 HIV reservoir quantification . . . 201
8.5 Statistical analysis . . . 201
8.6 Target capture and next-gen sequencing . . . 203
8.6.1 Sequencing analysis . . . 203
8.7 Droplet microfluidics . . . 204
8.7.1 Photolithography masks . . . 204
8.7.2 Silicon wafer master production . . . 204
8.7.3 PDMS chip production . . . 204
8.7.4 Running microfluidic devices . . . 205
8.7.5 Droplet merging . . . 205
8.7.6 Droplet sorting . . . 205
8.7.7 Emulsion breaking . . . 205
8.8 Clinical trials . . . 205
8.8.1 SPARTAC . . . 205
8.8.2 HEATHER . . . 206

References **207**

9 Full length contig reconstruction **223**
9.1 Full-length contig reconstruction . . . 223

10 Copyright permission **237**
10.1 Springer Nature copyright permission . . . 237

List of Figures

1.1 The UNAIDS 90:90:90 targets . . . 2
1.2 The virion of HIV-1 . . . 4
1.3 The HIV genome . . . 6
1.4 HIV DNA decay is biphasic and occurs slowly on ART. . . . 13
1.5 Potential fates of HIV DNA following an infection event. . . . 19
1.6 qPCR quantitation assays are unable to detect lethal defects in the HIV provirus. . . . 20
1.7 Defining 'rain' in the ddPCR platform. . . . 22
1.8 Commercial microfluidics chips . . . 28

2.1 SureSelect workflow . . . 40
2.2 Bait Design . . . 41
2.3 Nucleotide frequencies . . . 42
2.4 Barplot showing reads mapping to either HIV or human for cell-line testing . . . 44
2.5 Bar plots showing sequencing depth across the genome from SPARTAC PLWH samples. . . . 47
2.6 SureSelect HIV-coverage correlation . . . 47
2.7 Sampling issues . . . 49
2.8 Ligation sample loss . . . 51
2.9 Barplots showing the effect of increasing hybridisation and wash temperature. . . . 55
2.10 Effect of hybridisation temperature, wash temperature and bait concentration on enrichment efficiency . . . 56
2.11 Level of evaporation observed during hybridisation steps . . . 58
2.12 Aligner testing . . . 60
2.13 RNA/DNA species present after Sureselect hybridisation . . . 61
2.14 Integration site identification and visualisation . . . 64

3.1 Qscores are high and remain stable throughout the entire Miseq run. 69
3.2 Indexing is largely proportional to expected input . . . 70

3.3 Contamination between samples is minimal . . . 71
3.4 Sequencing depth produced using the modified Sureselect protocol. 72
3.5 Total HIV DNA does not correlate with the number of HIV mapped reads for this dataset. . . . 74
3.6 The modified Sureselect protocol improves the sensitivity of sequencing depth and increases genome coverage . . . 75
3.7 Viral LTR profiling. . . . 78
3.8 Integration site and overlapping amplicon PCR primers may lead to biased results. . . . 79

4.1 PDMS device fabrication . . . 85
4.2 Stages of droplet production in a flow-focusing junction . . . 87
4.3 PDMS device fabrication and syringe pump setup. . . . 89
4.4 WOWs . . . 91
4.5 WOWs in flow cytometry . . . 93
4.6 FACS fluorophores survive thermocycling . . . 96
4.7 Chip geometry parameters tested in relation to droplet size . . . 99
4.8 Flow-rate and geometry optimisation . . . 100
4.9 Wo thermostability is influenced by surfactant concentration and droplet size. . . . 103
4.10 Cell encapsulation in a microfluidic device . . . 104
4.11 Bead distribution from encapsulation experiment . . . 106
4.12 Particle sedimentation during sorting . . . 108
4.13 Schematic drawing of controlling droplet spacing through channel geometry . . . 109
4.14 Schematic drawing of merging during droplet separation . . . 110
4.15 Soft lithography masks of merging devices . . . 111
4.16 Schematic showing droplet re-injection and droplet catch-up. . . . 112
4.17 Brightfield images of droplet re-injection and droplet catch-up. . . . 113
4.18 Still images showing droplet merging in serpentine channels . . . 116
4.19 On-demand picoinjection of reagents into stabilised droplets . . . 118
4.20 Example sorting chip . . . 122
4.21 Excitation source testing . . . 125
4.22 Flow of data processing for droplet sorting . . . 127
4.23 Example plot showing initial droplet threshold determination . . . 129
4.24 Schematic detailing the paired droplet flow regime . . . 130
4.25 Hydrodynamic resistance . . . 131
4.26 Dielectrophoretic sorting . . . 132
4.27 Lithography masks of sorting devices used in this project . . . 133
4.28 Schematic of sorting for second geometry . . . 135

5.1 PCR mix droplet thermostability . . . 139
5.2 Removal of the PCR enzyme recovers droplet thermostability . . . 140
5.3 Glycerol and enzyme droplet thermostability . . . 140
5.4 Droplet thermostability of different PCR components . . . 142
5.5 Glycerol-free enzyme does not cause coalescence . . . 143
5.6 Igepal mediated droplet thermostability . . . 144
5.7 Schematic of an Igepal stabilised droplet . . . 144
5.8 Droplet size analysis performed in MATLAB . . . 146
5.9 qPCR compatibility testing of Igepal . . . 147
5.10 Schematic of cell membranes . . . 149
5.11 Thermal lysis . . . 151
5.12 End-point albumin PCR on heat-lysed cells and extracted DNA . . . 152
5.13 Test for albumin PCR amplification in-drop . . . 153
5.14 Test of direct lysis kit . . . 155
5.15 Localised fluorescence is not evidence of target amplification . . . 157
5.16 In-drop lysis outcomes . . . 159
5.17 Qiagen lysis buffers completely inhibit PCR . . . 160
5.18 Lysis buffers able to lyse cells inhibit PCR . . . 161
5.19 SDS PCR inhibition can be partially rescued by the addition of Tween-20 . . . 163
5.20 Optimisation of in-drop qPCR reagents . . . 165
5.21 Fluorescence results from a positive population of albumin droplets as measured by a PMT. . . . 167
5.22 Comparison between single-quenched and double-quenched probes. . . . 169
5.23 The double-quenched probe provides a greater fluorescent signal whe compared with te single-quenched probe . . . 171

6.1 The droplet pipelines. . . . 174
6.2 High DNA concentrations lead to jetting during WO production. . . . 175
6.3 PCR can be performed within thermostable droplets. . . . 177
6.4 Still images of droplet sorting. . . . 179
6.5 Potential droplet types dependent upon provirus shearing. . . . 183
6.6 Colocalisation of HIV qPCR assay fluorescence for Qiagen column extracted DNA. . . . 183
6.7 Colocalisation of HIV qPCR assay fluorescence for Qiagen Puregene extracted DNA. . . . 184
6.8 Colocalisation of HIV qPCR assay fluorescence for Qiagen Puregene extracted DNA. . . . 186

9.1 Full-length contig - view 1 . . . 224

9.2 Full-length contig - view 2 225
9.3 Full-length contig - view 3 226
9.4 Full-length contig - view 4 227
9.5 Full-length contig - view 5 228
9.6 Full-length contig - view 6 229
9.7 Full-length contig - view 7 230
9.8 Full-length contig - view 8 231
9.9 Full-length contig - view 9 232
9.10 Full-length contig - view 10 233
9.11 Full-length contig - view 11 234
9.12 Full-length contig - view 12 235
9.13 Full-length contig - view 13 236

List of Tables

1.1 Genes and proteins coded for by HIV-1 5

2.1 Samples used for standard protocol testing 43
2.2 Coverage and sequencing depth results 43
2.3 Clinical samples used for standard protocol testing 46
2.4 Coverage and sequencing depth results 48
2.5 Increasing temperature leads to a decrease in library yield but an increase in on-target capture 53
2.6 Table detailing the experimental matrix to test hybridisation temperatures, wash temperatures and bait concetration. 54

3.1 Sequencing quality results 68
3.2 Clinical data from the HEATHER samples 72
3.3 Coverage and sequencing depth results. Table showing coverage and sequencing depth results when aligned to the HIV genome. 73
3.4 Integration site candidates. 77

4.1 Properties of microfluidic chip materials 83
4.2 Obtainable depths using different SU-8 types. 84
4.3 Population frequencies from the WOW fluorophore experiment . . . 95
4.4 PDMS surface modifications 98
4.5 Encapsulation solutions 105
4.6 Sorting parameters 130

5.1 PCR components coalescence 139
5.2 PCR mix 2 components coalescence 141
5.3 The effect of Igepal on droplet radius post-thermocycling 145
5.4 Thermal lysis initial testing results 150
5.5 Mismatched primer/probe experiment design 155
5.6 An overview of the reagents tested as part of the cell lysis experiments. 158
5.7 SDS and Tween-20 have an inhibitory effect on qPCR amplification. 164
5.8 Mismatched primer/probe experiment design 166

6.1 Droplet thermostability 176
6.2 Nucleic acid recovery 180
6.3 Nucleic acid recovery after concentration of the DNA 180
6.4 Potential droplet types dependent upon provirus shearing. 182

8.1 qPCR and in-drop PCR primer and probe sequences 202

List of abbreviations and acronyms

ACD tubes . . Acid citrate dextrose tubes

AIDS . . . . . Acquired immunodeficiency syndrome

ART . . . . . Antiretroviral therapy

BHQ . . . . . Black-hole quencher 1

BWA . . . . . Burrows-wheeler aligner

CAR-T . . . . Chimeric antigen recepter T-cell therapy

CCR5 . . . . . C-C chemokine receptor 5; CD195

CAS . . . . . CRISPR associated protein (Cas).

CRISPR . . . Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated protein (Cas).

CXCR4 . . . . C-X-C chemokine receptor 4; CD184

DC . . . . . Dendritic cells

DC-SIGN . . . Dendritic cell specific ICAM-3 grabbing nonintegrin

ddPCR . . . . Droplet digital PCR

DMSO . . . . . Dimethylsulfoxide

DNA . . . . . Deoxyribonucleic acid

EC . . . . . Elite controllers

ELISA . . . . . Enzyme-linked immunosorbent assay

FACS . . . . . Fluorescence Activated Cell Sorting

FAM . . . . . Fluoroscein amidite

FCS . . . . . Foetal calf serum

FFJ . . . . . Flow focusing junction

FSC . . . . . Forward scatter

HDACS . . . . Histone deacetylase inhibitors

HEATHER . . HIV Reservoir Targeting with Early Antiretroviral Therapy

HERV . . . . . Human endogenous retrovirus

HEX . . . . . Hexachlorofluorescein succinimidyl ester

HFE7500 . . . 2-(Trifluoromethyl)-3-ethoxydodecafluorohexane

HIV . . . . . Human immunodeficiency virus 1

HLA . . . . . Human leukocyte antigen

HMDS . . . . Hexamethyldisilazane

HXB2 . . . . . HXB2 HIV reference genome

IGV . . . . . Integrated genomics viewer

IPDA . . . . . Intact proviral DNA assay

IUPM . . . . . Infectious units per million

kHz . . . . . Kilohertz

LEDIF . . . . LED induced fluorescence

LIF . . . . . Laser induced fluorescence

LRA . . . . . Latency reversing agents

LTR . . . . . Long terminal repeat

mRNA . . . . Messenger RNA

msRNA . . . . Multiply spliced RNA

NaCl . . . . . Sodium chloride

OW . . . . . Oil-in-water-droplet

PBMC . . . . Peripheral blood mononuclear cell

PBS . . . . . Phosphate buffered saline

PCR . . . . . Polymerase chain reaction

PDMS . . . . . Polydimethylsiloxane

PEG . . . . . Polyethylene glycol

PFPE . . . . . Perfluoropolyether

PHA . . . . . Phytohemaglutinin

PHI . . . . . Primary HIV infection

PIC . . . . . Pre-integration complex

PLWH . . . . . Person/people living with HIV

PMA . . . . . Phorbol 12-myristate 13-acetate

PMT . . . . . Photomultiplier tube

PREP . . . . . Pre-exposure prophylaxis

PTFE . . . . . Polytetrafluoroethylene

PTC . . . . . Post treatment controllers

Q-Score . . . . Quality score

qPCR . . . . . Quantitative PCR

qVOA . . . . . Quantitative viral outgrowth assay

R5 Tropic . . . Strain of HIV which uses CCR5 for entry

Reservoir . . . Latent reservoir

rev . . . . . Regulator of expression of virion proteins

$\mathbf{R}_H$ . . . . . Hydrodynamic resistance

RNA . . . . . Ribonucleic acid

RNase-H . . . Ribonuclease H

RNA Seq . . . RNA Sequencing

RT . . . . . Reverse transcriptase

RT-PCR . . . Reverse Transcription PCR

SPARTAC . . Short Pulse Antiretroviral Therapy at HIV Seroconversion

SPRI . . . . . Solid phase reversible immobilisation beads

SSC . . . . . Side scatter

tat . . . . . Trans-activator of transcription

T-Cell . . . . . T-Lymphocyte

TI . . . . . Treatment interruption

TILDA . . . . Tat/rev induced Limiting Dilution Assay

UNAIDS . . . The Joint United Nations Programme on HIV/AIDS

WHO . . . . . World Health Organisation

WO . . . . . Water-in-Oil-droplet

$\mathbf{WO}_d$ . . . . . Droplet diameter

WOW . . . . . Water-in-Oil-in-Water-droplet

X4 Tropic . . . Strain of HIV which uses CXCR4 for entry

ZEN . . . . . Internal zen quencher

Chapter 1

Introduction

Contents

1.1 Virology and epidemiology 2
1.1.1 A current snapshot of the global HIV epidemic 2
1.1.2 HIV structure and lifecycle 3
1.2 The successes of ART and other curative strategies. . 8
1.3 The latent HIV reservoir as the barrier to cure. 11
1.3.1 Mechanisms contributing to the maintenance of the reservoir 14
1.4 Reservoir studies and curative strategies. 15
1.5 Current measures of latency 18
1.5.1 Quantitative Viral Outgrowth Assay (qVOA) 18
1.5.2 quantitative PCR (qPCR) Assay 19
1.5.3 Tat/rev induced Limiting Dilution Assay (TILDA) Assay 22
1.5.4 Intact proviral DNA assay (IPDA) 23
1.5.5 No single assay is capable of accurately measuring the reservoir 23
1.6 The goals and requirements of HIV provirus sequencing 23
1.6.1 Replication competence 24
1.6.2 Integration site analysis 25
1.6.3 Age of individual proviruses 26
1.7 Droplet microfluidics 26
1.7.1 Microfluidics and emulsions 27
1.7.2 Why the HIV latency field needs droplets 30
1.7.3 Droplet microfluidics revolutionising single-cell analysis 31
1.8 Aims of this thesis 32
1.9 Thesis structure 33

1.1 Virology and epidemiology

1.1.1 A current snapshot of the global HIV epidemic

Thirty seven years on from human immunodeficiency virus 1's (HIV) discovery [1, 2], it remains the cause of a major global burden of disease [3]. Advances in antiretroviral therapy (ART), have reduced HIV-1 associated morbidity and mortality. People living with HIV (PLWH) who adhere to the prescribed drug regimen quickly control viraemia to levels below the limit of detection of current assays effectively leaving them unable to transmit infection [4]. The most recent UNAIDS global report on HIV published in 2018 contains within it many achievements that warrant praise, yet is balanced with a cautionary tale of those same successes breeding complacency [5]. The report coincides with the halfway point of the Fast-Track phase of the Global Aids Response first set out in 2014 [6]. The major milestones of this international initiative are highlighted in figure 1.1. In brief, by 2020 the initiative firstly aimed to ensure that:

- 90% of all PLWH are aware of their HIV status
- 90% of all PLWH who are aware of their status are receiving treatment
- 90% of all PLWH who are receiving treatment have a suppressed viral load [5]

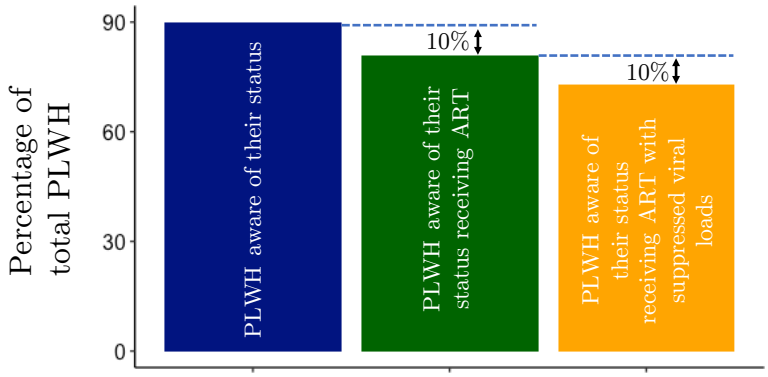


Figure 1.1: The UNAIDS 90:90:90 targets. Barplot depicting the milestones proposed by the Fast Track phase of the Global AIDS response

The 2018 document reports a similar story to previous annual reports where, despite strong successes the Global AIDS response remains behind schedule to hit the 2020 milestones [6]. The number of new HIV infections continues to fall year on year (~1.8 million as of 2017), however as the mortality rate also falls (940,000 as of 2017), the total number of PLWH globally rises. By approaching these data on a global population level there is a risk of oversimplifying the message; for example, a number of countries report a >50% reduction in new HIV infections whilst others still report increases in new infection rates [5]. Socio-economic obstacles to treatment and stigma surrounding HIV infection continue to play a dominant role in the successful implementation of the 90:90:90 targets. Should a vaccine or curative treatment be discovered in the laboratory, these factors will remain critical to the successful implementation of any potential global eradication schedule.

1.1.2 HIV structure and lifecycle

HIV is of the genus Lentivirus from within the Retroviridae family. The virion itself is an enveloped capsid containing within it two copies of its RNA genome and other essential proteins required for its continued propagation. Once released from a cell an individual virion has a lifespan of ~8 hours and the average complete lifecycle from viral budding, through infecting a new cell and then its progeny budding is ~2.6 days [7]. Each virion is ~100 nm.

1.1.2.1 Structure

HIV virions are comprised of:

- **Glycoprotein envelope** This envelope is comprised of a mixture of host cell membrane proteins and viral glycoproteins. The viral glycoproteins are comprised of gp120-gp41 heterodimers derived from the gp160 polypeptide precursor. gp120 constitutes the outer membrane of the virion and is involved in target-cell binding. gp41 is the transmembrane protein through which gp120 binds to the viral membrane and contains the fusion protein that is inserted into the target-cell membrane. The outward facing gp120 has

extensive glycosylation which may help prevent attack via the host humoral response. There are ~10 gp120-gp41 heterodimers embedded in each HIV virion [8].

- **Matrix protein** Attached to the inner surface of the viral envelope, p17 is a 132 amino acid polypeptide forming a protective shell around the nucleocapsid [9].

- **Nucleocapsid** This conical capsid core is comprised of an outer shell of p24 containing within it: two copies of the viral RNA genome, viral integrase, viral reverse transcriptase and the Vpr accessory protein [8].

A schematic diagram of a mature HIV virion is shown in figure 1.2.

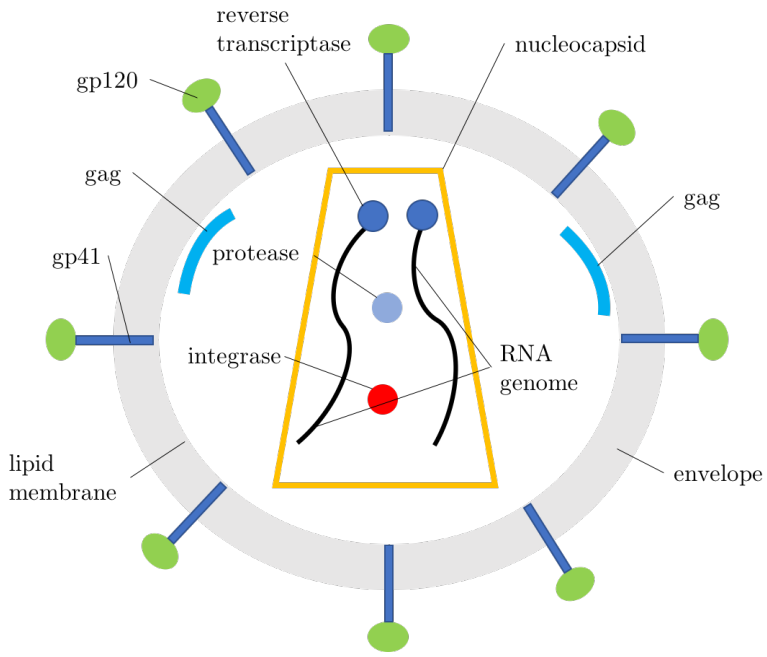


Figure 1.2: The virion of HIV-1. A schematic of a single HIV-1 virion. The numbers of each molecule are not representative.

1.1.2.2 Genome

The single-strand positive-sense RNA genome encodes three major proteins common to all retroviruses (gag, pol and env). These proteins are involved in viral structure and enzymatic activity. In addition, a number of smaller regulatory and

accessory proteins are encoded. In total, HIV encodes fifteen viral proteins from nine genes as shown in table 1.1. Two identical long-terminal repeats (LTRs) are situated at either end of the genome that play a role in integration, transcription promotion and viral RNA polyadenylation.

Table 1.1: Genes and proteins coded for by HIV-1

Gene	Protein	Function	Protein products
gag	p17	Structural	Matrix
gag	p24	Structural	Capsid
gag	p7	Structural	Nucleocapsid
gag	p6	Spacer	Spacer protein
pol	prot	Enzyme	Protease
pol	p51	Enzyme	Reverse transcriptase
pol	p31	Enzyme	Integrase
env	gp120	Structural	Membrane binding protein
env	gp41	Structural	Membrane fusion protein
tat	tat	Regulatory	tat
rev	rev	Regulatory	rev
nef	nef	Accessory	nef
vpr	vpr	Accessory	vpr
vif	vif	Accessory	vif
vpu	vpu	Accessory	vpu

The HIV genome has three open reading frames allowing all of its genes to be encoded in less than 10,000 nucleotides. Figure 1.3 shows the locus of each gene amongst the genome.

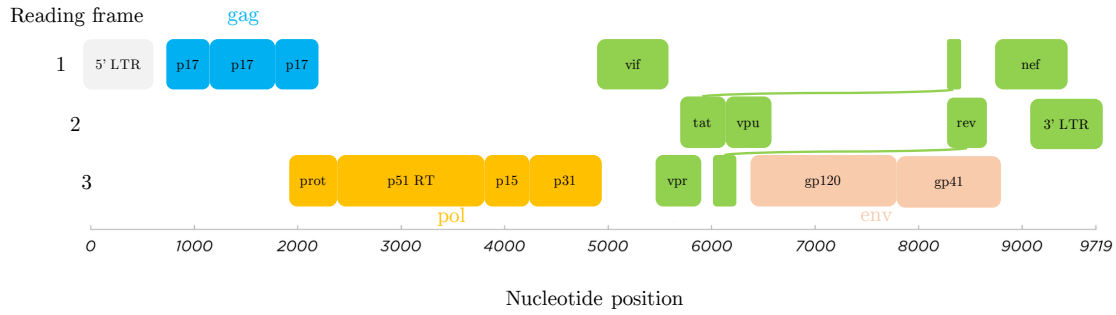


Figure 1.3: The HIV genome.

1.1.2.3 Lifecycle

HIV persists through repeated rounds of infection and replication within new cells. To gain entry to the cell, binding between virion and cell membrane must first be achieved, followed by fusion of the cell and viral membranes to allow delivery of the conical shaped viral core into the cytoplasm [10]. This process can be enhanced in the presence of attachment factors including $\alpha 4\beta 7$ or pattern recognition receptors such as dendritic cell specific ICAM-3 grabbing nonintegrin (DC-SIGN); however, such factors are not essential for viral entry [11, 12]. Beyond cell entry, attachment factors can play an additional role in HIV propagation; for example by adhering to dendritic cells (DCs) virions are able to utilise the DCs capacity to migrate from peripheral tissues to secondary lymphoid organs [13]. This mechanism allows the virion faster access to the CD4+ T-lymphocyte (CD4+ T-cell) compartments found in the lymphoid tissues. Research to date is inconclusive as to whether this DC mediated transport requires actual infection of the DC.

Following attachment, there are three steps that result in viral entry:

- Virion-CD4 binding
- Virion-Co-receptor binding
- Membrane fusion

HIV predominantly infects CD4+ T-cells via gp120 binding the CD4 receptor using either CCR5 or CXCR4 as a co-receptor. Despite gp120's preference for binding CD4, cells that do not express the marker are also capable of becoming

infected, albeit at low frequencies. The precise mechanisms governing these CD4 independent infection events is yet to be fully understood. Preliminary studies have identified C-type Lectin receptors such as DC-SIGN as likely candidates for binding due to the heavy glycosylation of gp120 with mannose and fucose saccharides [14]. Aside from the low frequency infection events described above, CD4 mediated gp120 binding is the preferential pathway for HIV propagation. The sequence of events that permit viral entry into the cytoplasm of the target cell are as follows:

1. Viral gp120 binds CD4 on the surface of the target cell inducing a conformational change that exposes the variable V3 loop of gp120.

2. The newly exposed V3 loop is recognised and bound by either the CCR5 chemokine receptor for R5 tropic viruses, or CXCR4 for X4 tropic viruses [15–17]

3. Binding of the co-receptor induces viral gp41-mediated membrane fusion and the viral capsid core is released into the cell cytoplasm.

Following entry into the cell, the capsid is uncoated releasing the two copies of the viral RNA genome for reverse transcription. For cells that go on to be successfully infected, the uncoating process typically begins ~30 minutes following entry into the cell. Uncoating can happen much faster or slower dependent upon the virion and host factors. However in a study by Mamede et al individual cells were live imaged to determine whether their associated virions led to successful viral DNA integration. In this study, the virions associated with cells deemed to be successfully infected began uncoating after ~30 minutes [18]. In addition, complete uncoating reduced the chances of successful nuclear import as a small amount of residual capsid protects the RNA from the host proteasome [19].

Following capsid uncoating the RNA genome must be reverse transcribed to DNA ready for nuclear translocation. The viral RT protein packaged within the viral capsid performs two essential functions during this stage of the lifecycle: duplication of the RNA template via a DNA polymerase and degradation of the

original RNA strand (still bound to the new DNA strand) via an RNase H [20]. The RNase H will only degrade DNA when it is part of an RNA-DNA hybrid. The double-stranded viral DNA genome is subsequently incorporated in to the pre-integration complex (PIC) where it translocates to the nucleus [21]. Formation of the PIC, nuclear import and integration of the viral DNA into the host genome are all dependent upon host cell proteins. Following nuclear import, the double stranded viral DNA genome is inserted into the host genome [21]. This process will be described in greater detail in section 1.6.2.

Once integrated into a host chromosome, the HIV genome is able to be transcribed through host machinery, usually upon cellular activation or division [22]. Initial transcription of the provirus leads to the production of the unspliced version of the RNA genome. This unspliced RNA has two roles: to act as the RNA genome in new virions and also to act as the messenger RNA (mRNA) for the viral proteins and enzymes that are packaged within them. The fate of each unspliced RNA is dictated by the start site used during transcription [23]. HIV transcription begins at one of three guanosine residues (either nucleotide position 454, 455 or 456). Transcripts beginning at 456 are more likely to dimerise and are most frequently fated to become packaged RNA genomes within progeny virions. Alternatively, those transcripts beginning at 454 or 455 most commonly remain monomeric and preferentially accumulate on polysomes and function as mRNAs [24].

Following the production of the RNA genome and the associated viral enzymes, the virus assembles at and buds from the plasma membrane. Upon budding the virion is considered immature and is non-infectious. To mature into an infectious virion, the gag polyproteins must be cleaved by the viral protease at five specific sites creating the functional proteins [8]. At this point the virion is considered mature and is able to infect new target cells as described above.

1.2 The successes of ART and other curative strategies.

The increasing effectiveness of ART over the past 15 years is among the most noteworthy medical successes of recent times. PLWH can now expect a lifespan in line with the general population and a vast reduction in the chance of HIV-related comorbidities [25–27]. A gradual decline in HIV related deaths has been reported year on year since its peak in 2004. By 2017, the mortality rate had fallen to 51% of the 2004 level (940,000 and 1.9 million deaths respectively) [28]. The WHO modified its guidelines in 2015 to recommend that any person newly infected with HIV should start ART immediately and remain on therapy [29]. This was a shift away from the previous guidelines which recommended the commencement of therapy once a CD4+ T-cell count $\leq$350 was reached [30–32]. This recommendation was made as a result of both clinical trial and research data. The Strategies for Management of Antiretroviral Therapy (SMART) trial in 2006 found that the episodic administration of ART guided by CD4 count significantly increased the risk of opportunistic disease and death when compared to those on continuous therapy [30]. In addition, episodic ART did not reduce the risk of ART associated adverse events. Further support for early ART came in 2011 from the HPTN 052 study where serodiscordant couples were followed to measure the incidence of linked HIV transmission in HIV negative partners. Of the 28 virologically linked transmissions, one occurred in the early ART group and the remaining 27 were in the delayed ART GROUP [31]. Finally, the Strategic Timing of Antiretroviral Therapy (START) trial was terminated early when an interim analysis showed that the risk of a serious AIDS related event or death was 57% less likely in the early therapy group when compared with those initiating therapy when their CD4 count was <350 cells/mm^3 [33]. Research data published at the time further strengthened this argument by suggesting that commencing ART immediately after diagnosis reduced both reservoir size and immune activation [34–36]. In one study, early ART reduced total HIV DNA by 20-fold after 2 weeks and 316-fold after 3 years [37].

For the majority of PLWH who are on therapy, HIV is suppressed in the body to a level that is below the limit of detection of the most sensitive assays available. The viraemic burden in PLWH on suppressive treatment is so small they are considered to be unable to transmit their virus [31, 38, 39]. In recent years ART has also been introduced as a pre-exposure prophylaxis (PREP) for HIV. In this setting people whose lifestyles place them in a high risk category of contracting the virus can take ART prior to high-risk activities in order to block the aquisition of the virus [40–42]. A recent study has identified a sex difference between the efficacy of PREP in which it was found to be more efficacious in men than women; the authors suggest this difference is likely due to the pharmacokinetic differences of tenofovir between vaginal and rectal tissues [43].

The new generation of combination drugs now available to PLWH has changed the face of the epidemic, especially in developed countries. This 'new normal' has also changed the expectations of the patient groups and the direction of HIV researchers. As HIV is now predominantly kept under pharmaceutical control in PLWH on therapy, efforts have shifted more towards a sterilising, or perhaps more probable, a functional cure. In this context a sterilising cure would be considered an intervention capable of removing any trace amounts of HIV from a PLWH including blood and any other anatomical reservoirs. This is believed to have been achieved for Timothy Brown, the first PLWH to have the virus completely eradicated from his body [44, 45]. Timothy Brown was diagnosed with Acute Myeloid Leukemia and received total body irradiation followed by an allogeneic stem-cell transplant from a donor who was homozygous for the 32-bp CCR5$^{\Delta}$32 deletion. This mutation leads to a non-functional CCR5 gene product and consequently confers resistance to R5 tropic HIV virions [46]. A number of attempts have been made to repeat this with limited success either due to the participants dying from their underlying malignancies or via rebound of an X4 tropic virus that does not require CCR5 as a co-receptor for cell entry. Two recent cases show further promise, one PLWH in London [47] and one in Düsseldorf [48]. Both patients received total body irradiation prior to an allogeneic stem-cell

transplant from a donor with the CCR5$^{\Delta}$32 deletion. To date, HIV has not been detected from any sites sampled in either of these patients. It is worth noting that although the London patient has been treatment and virus free for over 18 months, the Düsseldorf patient has only been so for 4 months. It is therefore possible that these patients could both rebound and that a sterilising cure has not been achieved.

The alternative therapeutic option to a sterilising cure is the concept of a functional cure. In this situation, it is hoped that a medical intervention will reduce the viral burden in a PLWH to a level that renders the virus incapable of seeding rebound viraemia in the absence of ART. It is currently unknown exactly what criteria will need to be met in order to achieve a functional cure, however it is entirely plausible that the activation of a single cell latently infected with a replication competent virus could be capable of seeding rebound viraemia [49]. As a result, it will likely require the development of effective host immunity to HIV, in order to provide lifelong suppression of the virus in the absence of ART.

1.3 The latent HIV reservoir as the barrier to cure.

The focal point of current cure research is what has been labelled the 'latent reservoir' (reservoir) and is widely accepted to be the major barrier to cure [50–63]. There is a level of ambiguity attached to this term, therefore in the context of this thesis it can be defined as the following:

> A long lived HIV infected cell capable of maintaining itself in a quiescent state, or through periodic cell division, without reactivating the latent virus. This cell must also be able to seed active viral replication in the event of the cell becoming activated. [64]

Despite prolonged ART, this population of cells harbour HIV DNA integrated into the host genome yet remain invisible to the host immune system [50, 65, 66]. The largest contributor and most extensively characterised of the cells that make up the reservoir are the resting memory CD4+ T-cells. HIV is capable of infecting resting or activated cells but shows a strong preference for infecting activated cells [67, 68]. It is hypothesised that these cells are generally:

- Infected whilst an activated cell is returning to a resting state
- Infected whilst activated and subsequently return to a resting state. In order to achieve this the cell would first have to survive the cytolytic effects of virion production and evade the immune system. [69]

One of the determining characteristics of the memory CD4+ T-cells is persistence [70]. By selectively integrating into these cells, a viral genome is able to remain hidden from host immune surveillance whilst serving as a long-term viral reservoir [69].

On ART, the plasma and intracellular HIV RNA burden quickly declines to undetectable levels whereas HIV proviral DNA, despite an early rapid decline, eventually decays at a much slower rate [71–74]. During this early phase viral RNA is not suppressed, hence active viral replication is taking place. HIV DNA decay is slow during this phase as labile pre-integration HIV DNA species are present within cells and are constantly replenished by ongoing infection [75]. As viral RNA is suppressed over time with ongoing ART, viral replication all but ceases resulting in a loss of the pre-integration form of latent reservoir cells. DNA decay is now centred around the more stable HIV proviruses resident in resting memory CD4+ T-cells. Consequently the half-life of the DNA reservoir increases leading to a distinct biphasic decline in HIV DNA for PLWH undergoing ART (fig 1.4). Viral species present in the slow decay portion of the reservoir have been shown to most closely relate to RNA viruses circulating in pre-therapy samples [76] suggesting that the long-lived portion of the DNA reservoir is largely stable once therapy has commenced.

The reservoir is established early in HIV infection and cannot be blocked by early ART. In one case-study, a baby started treatment within 30 hours of vertical infection, and received regular ART for 18 months. Treatment was stopped as the parents were lost to follow-up and the baby remained virus free off of therapy [77]. After 27 months of treatment interruption (TI), a decline in CD4+ T-cell count was observed followed shortly afterwards by rebound viraemia upon which the

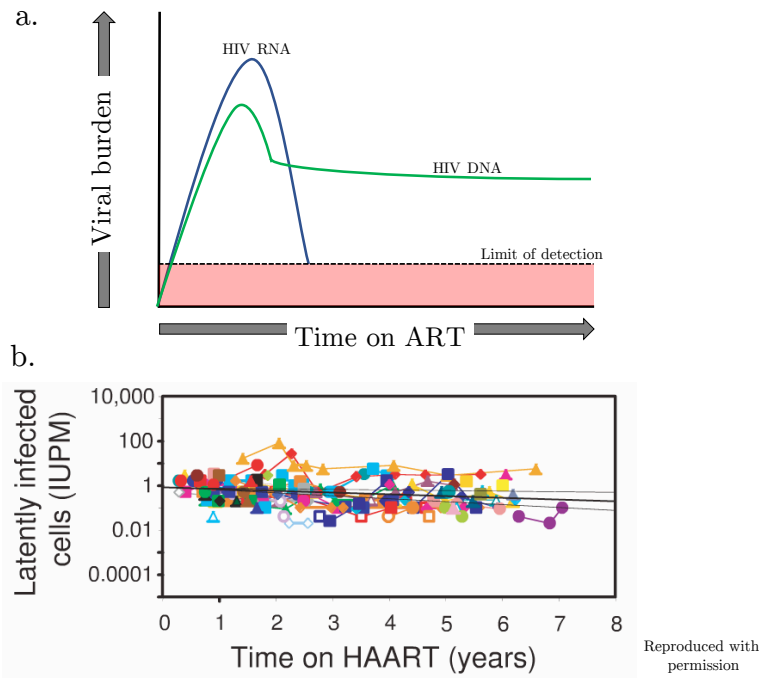


Figure 1.4: HIV DNA decay is biphasic and occurs slowly on ART.(a) Lineplot showing HIV RNA/DNA dynamics for PLWH on ART over time. (b) Figure 1.a from Siliciano et al [74] demonstrating the frequency of latently infected cells in 62 HIV-1-infected adults who maintained suppression of viremia on ART.

baby was placed back on ART [78]. This case-study showed that despite beginning ART the day after infection, the latent reservoir is already seeded and that early treatment is not a potential cure intervention.

The half-life of a latently infected cell has been estimated to be ~44 months [69]. This means that despite quickly suppressing HIV RNA, a PLWH with a reservoir of 10^6 cells would require ~73.4 years of ART to eradicate the reservoir [74]. Studies have been performed monitoring individuals on long term suppressive ART, who undergo treatment interruptions. Despite disease progression being delayed on ART [79], the majority (~ 85%) of patients will rebound within weeks if not days of ART cessation and will return to setpoint viral loads [80]. Recent evidence suggests there may be a low frequency of individuals (5 - 15%) who, if treated early in primary HIV infection (PHI) for a prolonged period, continue to control viraemia after TI [80–82]. These individuals, dubbed post treatment controllers (PTCs), differ from the well documented spontaneous elite controllers

(ECs). PTCs are not enriched for the common protective HLA alleles e.g HLA-B*52/27/57, exhibit higher viral loads in PHI and more frequently present with symptomatic PHI, none of which are expected in ECs [81, 83, 84]. PTC is not always observed however, with a success rate of between 1-16% comonly observed when patients are treated early. TI has been modelled and both EC and PTC have been predicted. EC is predicted in the presence of a strong, polyfunctional cytotoxic T-lymphocyte response (commonly seen in ECs) and PTC is predicted in the presence of a small reservoir (<1 cell/million CD4+ T-cells) and if the cell-mediated immune response is adequate [85]. PTC can also be difficult to classify as a number of PLWH who remain untreated are also able to exhibit long periods of transient viral remission following PHI [86].

1.3.1 Mechanisms contributing to the maintenance of the reservoir

It is not yet fully understood what factors contribute to HIV proviruses remaining latent and persisting during suppressive ART [87]. What is clear is that despite cell cycling, in which latently infected cells will eventually die, there must be at least one proliferative mechanism that allows the reservoir to persist beyond the half-life of a memory cell. A low level of viraemia is regularly detectable from plasma samples of PLWH on suppressive ART [88–90]. As current ART does not restrict viral transcription, translation, assembly or virion release, this data alone is unsurprising [91]. However, as virion fusion and entry are blocked by ART any released virions should be unable to reinfect new cells and contribute to reservoir maintenance.

A number of studies have suggested that the presence of viraemia coupled with a low level of viral evolution on ART is evidence enough to assume that ongoing replication occurs despite effective ART [57]. This remains a controversial view within the HIV research field and to date there is no consensus in literature as to what the implications of this are. ART penetration is not uniform across the body and ‘drug sanctuary sites’ have been discovered where ART penetration is limited in comparison to other areas such as the peripheral blood [92, 93]. However, it is noteworthy that a correlation is not always observed between drug concentrations

and residual viraemia in these sanctuary sites [92]. This data suggests that if drug concentration is a contributory factor to residual viraemia, low concentrations alone are not sufficient to permit viral replication and other pathways must be involved. Conversely, other studies propose that there is little evidence that ongoing viral replication on suppressive ART contributes to reservoir maintenance [93–96].

The dominant mechanism driving the maintenance of the reservoir is the proliferation of latently infected cells by either antigen-induced or homeostatic cytokine-mediated proliferation [97–101]. Crucial to HIV persistence, this proliferation of latently infected cells can occur without inducing viral transcription which would make the cell vulnerable to immune targeting [102, 103]. Any therapeutic approaches looking to target HIV persistence through this mechanism, for example by blocking T-cell proliferation, must do so whilst not having a detrimental effect on the uninfected T-cell population.

1.4 Reservoir studies and curative strategies.

Despite the resounding success of the latest generation of ART, it can not realistically be considered a curative strategy. In addition to this, the logistics and costs of trying to provide a lifetime of treatment to the 35 million PLWH dispersed across the globe (many in resource-poor countries with limited infrastructure) are immense. There exists therefore, a global desire for either a sterilising or functional cure for HIV. It is beyond the remit of this introduction to explain each of the potential avenues in detail but an overview of those most frequently presented in the literature are provided below:

- **Latency reversal or 'Kick and Kill'.** This branch of cure research seeks to eradicate or decrease the latent reservoir in PLWH by using latency reversing agents (LRAs) to reverse latency thereby inducing killing by viral cytolytic effects or immune-mediated targeting [104]. A class of drugs known as histone deacetylase inhibitors (HDACs) used in cancer treatment [105] are currently being tested in a number of kick and kill studies. HDACs trigger deacetylation of the genome, thereby releasing any transcriptional regulation imposed on

the virus by histone acetylation. The principle being that removing the histone-mediated transcriptional silencing will enhance viral protein expression and virion budding leading to increased immune clearance and virus induced cytolysis. HDACs such as Vorinostat [35, 106] and Romidepsin [107] have been shown to induce a modest increase of HIV transcription in-vivo. However, this increase is not reflected by an increase in plasma viral RNA [106]. In addition, the administration of Vorinostat in Wightman et al led to potentially harmful immunological changes including an increase in regulatory T-cells which are critical to regulating the activation and proliferation of effector T-cells [108]. In an extensive study of potential LRAs including three HDACs, a Protein kinase C agonist (bryostatin-1) and an LRA proposed to involve nuclear factor-κB (disulfiram), Bullen et al reported that only bryostatin-1 induced a significant increase in cell-associated HIV mRNA and this effect was only 4% of that seen for Phorbol 12-myristate 13-acetate (PMA) + ionomycin activation [109]. Combination therapy approaches with LRAs utilising different pathways have been shown to increase the level of latency reversal up to 50% of the activation observed with PMA + ionomycin. Research continues in this area to optimise drug concentrations and synergies. [110]

- **Vaccines.** In 1984 a speculative estimate for the time until an HIV vaccine would be developed was proposed to be two years [111]; 35 years later there is still no therapeutic or prophylactic vaccine available for HIV. Early attempts following traditional vaccine development strategies targeted 20+ recombinant env proteins in a bid to induce neutralising antibodies to the virus. These studies showed no decrease in new infections versus placebo controls and no delay in disease progression [112, 113]. Contemporary vaccine efforts can be compartmentalised into a number of approaches.

  - Env V2 specific antibody responses matched to local strains (RV144 trial [114])

  - Diverse polyfunctional antibody responses
  - Broadly neutralising antibody responses
  - Elicit effector memory cells via replicating vectors

  Despite good immunogenicity in-vitro, clinical trials thus far have failed to live up to the promise of the preliminary data. The RV144 trial is the only trial to have some success with 31.2% vaccine efficacy at 42 months. For those patients in whom HIV was later diagnosed, vaccination did not affect the viraemic burden or CD4 count [114]. Despite modest results, each new vaccine trial is providing researchers with a better understanding of the immune correlates of protection. Building on these foundations, there are currently a number of active trials underway that show promise.

- **Gene editing.** The success of the Timothy Brown case [44] has prompted research into alternative means of implanting an HIV-resistant immune system. Gene editing techniques including Crispr/Cas9 [115, 116] and Zinc finger nucleases [117] have been tested for their ability to disrupt CCR5 expresion in-vitro. These studies have shown by means of adenovirus and lentivirus delivery systems that there is a rapid selection of $CCR5^{-/-}$ cells in the context of HIV infection. Work continues in this exciting area of research with a newly ignited interest in CAR-T cells also becoming prominent in light of recent successes in treating haemotologic malignancies. This approach towards HIV eradication has recently been reviewed in Spragg et al [118].

One major limitation to all of these cure strategies is that for them to be successfully tested a TI is required. TI is the process by which a PLWH who has received a medical intervention designed to reduce the HIV reservoir stops taking therapy for a period of time to determine whether the virus will rebound. To perform a TI study as safely as possible it is necessary to monitor the virus [119]. This can be achieved through viral RNA levels, however as upon the recommencement of therapy this will likely return to undetectable, it is not a good

measure of the success of the treatment. Ideally, an accurate reservoir measurement would be recorded pre, during and post-intervention to determine the effect on the reservoir which is the primary goal for any of these studies. Further to assessing efficacy of the TI, eligibility criteria for PLWH entering these trials should consider the DNA reservoir level at TI as it has been shown that a high pre-TI DNA reservoir measurement decreases time to viral rebound [120]. Moving forward, DNA reservoir quantitation and sequencing stand to play a critical part in TI studies and at present there exist no accurate, scalable methods to quantify the latent reservoir. The following section will detail the current methods available and highlight the strengths and weaknesses of each.

1.5 Current measures of latency

Following the successful infection of a cell, newly transcribed HIV DNA has the potential to follow a number of different paths. The path the DNA follows dictates whether it has the potential to cause recrudescent viraemia or will play no further part in viral persistence. The potential fates of HIV DNA are described in figure 1.5.

In order to fully understand HIV latency, it is necessary to be able to detect and quantify each of the potential viral fates. To date, no assay is capable of doing so. The current state of the art in reservoir quantification and characterisation assays is reviewed below.

1.5.1 Quantitative Viral Outgrowth Assay (qVOA)

The qVOA has been seen as the 'gold standard' for measuring HIV reservoir DNA [123–125]. In the qVOA, resting CD4+ T-cells are activated using phytohemaglutinin (PHA) and subsequently co-cultured with allogeneic, irradiated, CD8+ depleted peripheral blood mononuclear cells (PBMCs) at an effector/target ratio of 1:5 [124, 126]. After 7 or 14 days, replication is measured by either, polyadenylated HIV-1 RT-PCR [62] or p24 ELISA [124]. Using the qVOA, only replication competent viruses will elicit a positive p24 response. The reservoir may be up to 60-fold higher than the qVOA assay result however, as a single round

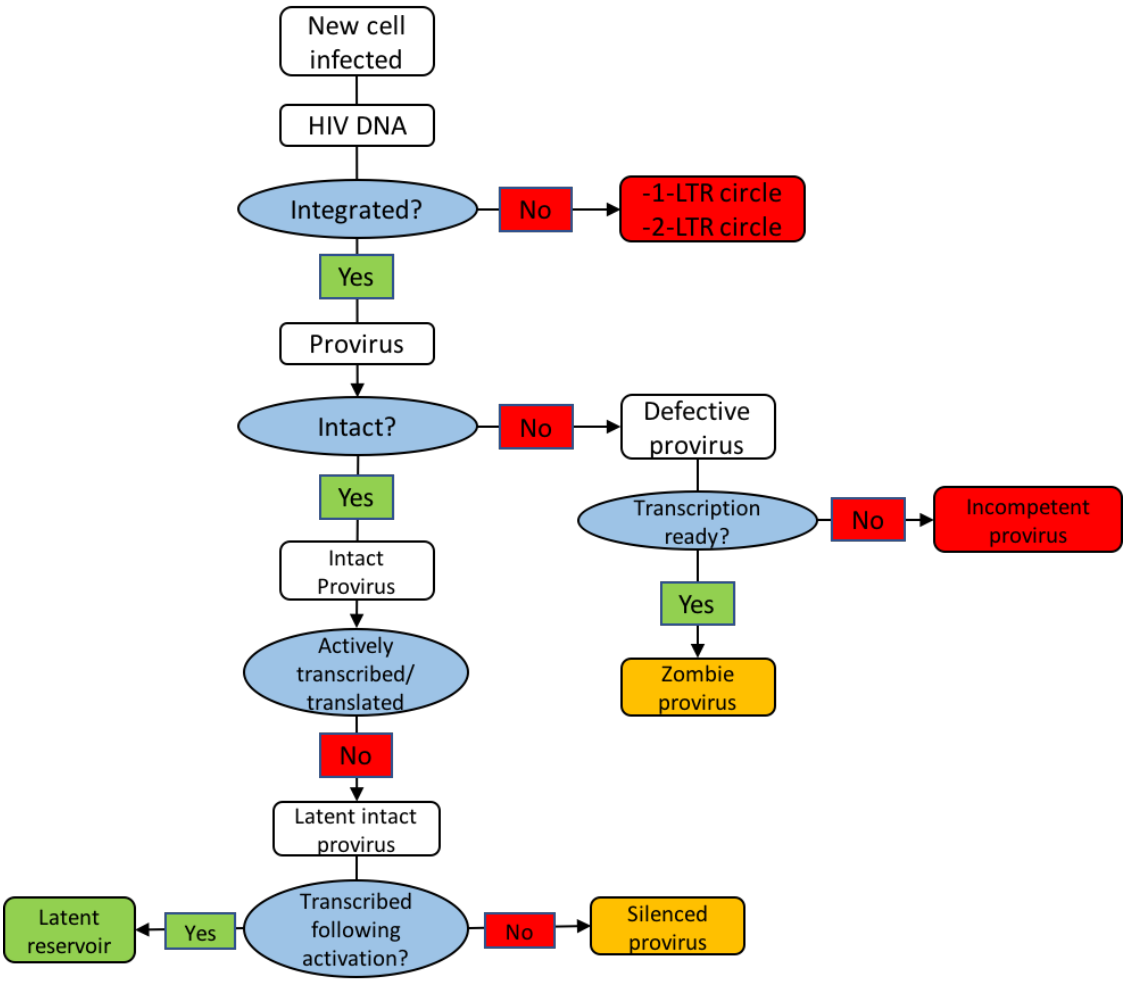


Figure 1.5: Potential fates of HIV DNA following an infection event. A decision tree detailing the potential fates of HIV DNA following transcription within a newly infected cell. Red termini indicate a genetic dead-end for the virus, orange termini indicate an infectious deadend for the virus and the green terminus indicates the potential for further infection. 1 and 2-LTR-circles are the result of unsuccessful integration events. Adapted from [121] under the Creative Commons licence [122].

of activation is not sufficient to induce all replication competent viruses able to establish recrudescent viraemia [127]. This was shown in Ho et al where non-induced proviruses were sequenced and reproduced by de-novo genome synthesis. 11.7% of the non-induced proviruses were found to have intact genomes and normal LTR function. In addition, 100% of the proviruses that were reconstructed showed equivalent replicative fitness to NL4-3 and other induced proviruses from the same patient [63].

1.5.2 quantitative PCR (qPCR) Assay

qPCR assays generally target a short (150-200bp) region of the HIV genome [128–130] and are time, sample and cost effective. The major limitation of qPCR assays is that both defective and replication-competent viral species will be quantified. As long as the qPCR target region is intact, even viruses with large structural deletions will be measured (fig 1.6). The qPCR method therefore overestimates the true size of the reservoir as it has been estimated that >90% of proviruses are replication incompetent [119, 131]. A study comparing qVOA results and qPCR results from the same set of patient samples to assess concordance between the two was performed by Eriksson et al in 2013. In this study the authors found that assays using PCR quantitation gave results at least 2 logs higher than the qVOA assay [132]. This relationship held even for patients who had started ART during acute infection where it had been hypothesised that the viral population would not have had enough time, or replication cycles to build up a large defective population. This data was reinforced by a later study that suggested defective proviruses rapidly accumulate during acute HIV infection [133].

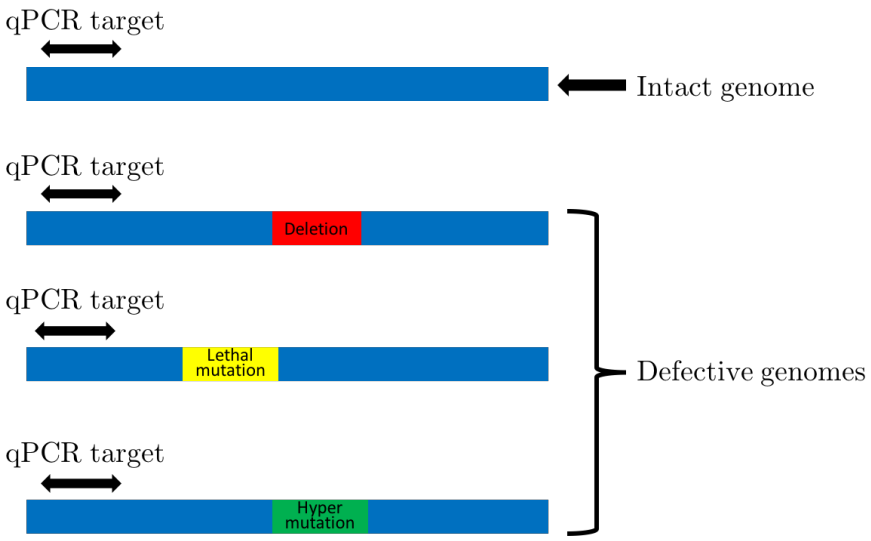


Figure 1.6: qPCR quantitation assays are unable to detect lethal defects in the HIV provirus.

1.5.2.1 Droplet digital PCR (ddPCR)

The development of HIV reservoir assays has occured in parallel with the rise of ddPCR. ddPCR centres around the principle of a limited dilution single-molecule PCR, a concept first described in the context of provirus quantitation over 15 years ago [134]. The assay requires diluting a sample to the extent that when divided among multiple reactions, the distribution of template molecules follows a Poisson distribution. Some reactions will not contain any copies of template and the rest will contain one or more. Using the frequency of positive and negative reactions it is then possible to deduce the concentration of template copies in the starting sample [135]. Recent technological advances in miniaturisation have allowed the implementation of greater reaction numbers by either multi-well plates or droplet microfluidics. The Biorad QX200 is one such instrument and has been widely used in HIV reservoir quantification over the past 5 years. As a result, any reference to ddPCR for the rest of this section will refer to Biorad's QX200 system. The QX200 partitions the bulk qPCR reaction by generating ~20,000 individual water-in-oil droplets. A major benefit of using ddPCR is the ability to quantify the sample without the requirement for a standard curve. For reservoir studies in particular, HIV DNA concentations are low, this can have a detrimental effect on reproducibility due to issues with sensitivity and accuracy when using low concentration PCR standard curves. As with qPCR, multiple species of HIV RNA and DNA can be measured using this approach. Only HIV DNA measurements are discussed here.

ddPCR has been shown to improve accuracy but not sensitivity for rare targets such as proviral HIV DNA [129, 135–139]. A major barrier limiting ddPCR's ability to accurately quantify low copy number samples is the phenomenon of 'rain' droplets in the analysis. It is impossible to determine whether rain droplets are positive or negative as their fluorescent intensity values fall close to the sorting threshold of the analysis (fig 1.7).

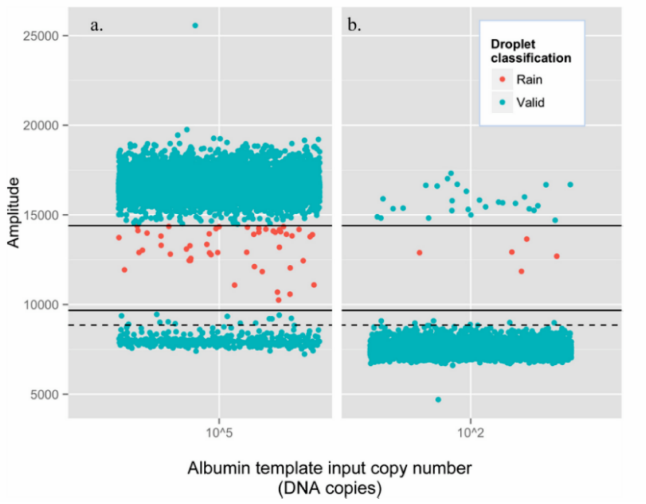


Figure 1.7: Defining rain in the ddPCR platform. Results of ddPCR from (a) high (10^5) and (b) low (10^2) copy number samples assayed in an albumin quantitation assay. Droplets that do not appear to be distinguishable as either positive or negative (rain) are coloured red. The dashed horizontal line indicates the threshold above which the instrument software determines a droplet to be positive. The full horizontal lines indicate limits that have been applied using definetherain on a positive control sample. Droplets that fall between these cut-offs are defined as rain and are not included in the calculation to determine copy number [139].

Numerous methods have been proposed to improve the way rain droplets are handled; these include the removal of ambiguous droplets [139] (fig 1.7), the implementation of a model based on extreme value theory that provides a single threshold [140] and cluster approximation based on a binormal distribution [129]. Before ddPCR can be used to reproducibly measure HIV DNA burden between different laboratories an different trials there is a need to standardise the droplet analysis method [137].

1.5.3 Tat/rev induced Limiting Dilution Assay (TILDA) Assay

The TILDA assay measures the frequency of cells that harbour viral genomes capable of producing tat/rev multiply spliced RNA (msRNA) upon stimulation [127]. This assay has the benefit of increased specificity over the qPCR assay as many viruses unable to produce rev and particularly tat transcripts have been shown to be non-replication competent [63]. Conversely, the ability to produce msRNA is not analogous to the virus being replication competent as they may

not be able to produce infective virions. When tested, TILDA correlated with the qPCR assay but not with qVOA. The DNA reservoir, as measured by TILDA in the Procopio study, was found to be a median of 24 cells/million CD4 cells. This was 48 times higher than the same samples as measured by qVOA and 6-27 times lower than the same samples as measured by total DNA qPCR [127].

1.5.4 Intact proviral DNA assay (IPDA)

A recent study published by Bruner et al proposed a new qPCR method that claims to be able to differentiate between intact and defective proviruses [119]. This IPDA assay uses two target amplicons, one at the 5' end of the virus and one at the 3' end that have previously been shown to be 90% predictive of defective genomes when one or both can not be amplified. An additional probe is also added for the 3' target that will preferentially bind when two APOBEC-3G induced mutations are present. This additional probe allows detection of defective proviruses induced by APOBEC-3G induced hypermutation. This assay is the first to become available that is scalable to medium-large clinical trials but to date has only been applied to a single cohort.

1.5.5 No single assay is capable of accurately measuring the reservoir

A significant amount of research and development has been employed over the past decade to try and produce a single assay capable of providing a true representation of the latent reservoir in PLWH. Each of the assays described above is able to quantify only a subset of the different provirus fates described in figure 1.5 and many are unable to distinguish between two or more of these separate fates. The different assays proposed all have different strengths and weaknesses however none are capable of producing a quantitative measure of a PLWH's functional DNA reservoir burden in a single scalable assay. As a direct consequence of these assays our understanding of the mechanisms of viral latency has improved. However,

moving forward into the new era of TI studies, the need persists to find a reproducible measure of what effect potential therapeutic interventions are having on the reservoir.

1.6 The goals and requirements of HIV provirus sequencing

Due to the high error rate of HIV's RT enzyme, high viral loads and the speed at which viral replication occurs, intrahost genetic diversity increases rapidly in acute infection [133, 141, 142]. These factors combined mean it is theoretically possible for a single-point mutation to occur at every position along the 10kb genome every day a patient is left untreated [143]. These defects are largely stochastic [144] and are often detrimental to the virus leading to either a decline in viral fitness or rendering them defunct. With the discovery that the majority of latent HIV proviruses residing in PLWH are defective [63], it has become clear that to effectively assess the replication competence of an individual's reservoir, some form of sequence analysis is required [64]. Proviral sequencing has the power to improve our understanding of a number of different aspects of HIV infection and latency including:

- Replication competence
- Integration site analysis
- Age of individual proviruses
- Drug resistance
- Strain pathogenesis
- HIV surveillance
- Transmission chains

Some of these potential applications most closely related to the aims of this thesis are discussed in more detail below.

1.6.1 Replication competence

As discussed previously, the replication competence of an HIV provirus is the key determinant of whether it has a potential role to play in viral rebound in the absence of therapy. It has been shown that the mere presence of a provirus as detected by qPCR, is not sufficient to cause recrudescent viraemia. Equally, the absence of successful virion production following host-cell activation is not evidence that the provirus is incapable of doing so stochastically in-vivo. Deep sequencing approaches utilised to date have either been performed on a bulk population of cells or have involved an upfront enrichment step based on the cell's response to activation such as in the qVOA. In an ideal world it would be possible to harness the power of full-length HIV deep sequencing on individual latently infected cells in order to be able to determine viral haplotype and its associated replication competence.

A further benefit of having deep sequence analysis for HIV proviruses is the ability to distinguish so-called "zombie" proviruses. These are non-replication-competent proviruses that are still able to act as template for select HIV transcripts. It is thought these zombie proviruses have a role to play in persistent immune exhaustion for PLWH on suppressive therapy [131, 145].

1.6.2 Integration site analysis

Integration of the provirus into the host genome is a stable and irreversible stage in the HIV lifecycle. Integration does not require a full-length or intact genome, in fact, fragments of HIV as small as one of the LTR regions of the virus have been found integrated into a host genome [63]. Integration can happen anywhere within the host genome however the process is not random. HIV has been found to preferentially integrate close to transcriptionally active genes [146, 147] and in areas just inside ($\leq 1\,\mu m$) the nuclear envelope [148]. This localisation of HIV DNA at the nuclear envelope is only seen for integrated DNA and when integratation is blocked, the unintegrated LTR-circles are found throughout the nucleus. The leading theory as to why HIV has evolved an integration preference for actively transcribed regions suggests it is linked to HIV's preference for infecting activated

CD4+ T-cells. These cells, once infected, have a short half life (~1.6 days) [7] and therefore HIV variants that propagate the most virions within that period of time will likely hold a selective advantage.

As discussed, proviral integration sites appear to have an effect on reservoir persistence. For this reason, being able to identify individual integration sites may aid in better understanding what enables some proviruses to persist for as long as they do. For example, integration within oncogenes or cell cycle regulatory genes appears to confer a selective advantage and allows these cells to persist during suppressive ART [101].

1.6.3 Age of individual proviruses

Until recently, attempts to determine the date of integration for a latently infected cell have been performed by measuring the genetic distance between the provirus in question and pre-therapy HIV RNA sequences [58]. One major limitation to previous implementations of this method is the failure to establish an intra-host evolutionary rate specific to each host. In a phylogenetic framework presented in 2018, Jones et al built upon the inter-host phylogenetic models that have been used to map the early stages of the HIV pandemic to its roots in Kinshasa [149]. Instead of predicting an inter-host molecular clock they calculated intra-host evolutionary rates from longitudinal RNA sequences with known sampling dates. A root sequence was selected from the training dataset that optimised the relationship between time of sampling and the evolutionary distance, a linear model was then fitted to explain the relationship between time of sampling and the root-tip evolutionary distance for each sample in the training dataset. For unknown query sequences, the evolutionary distance from the root can then be used to infer the date at which the virus became latent. When applied to patient samples where 50% of the samples were censored and 50% were used as the training dataset, the model predicted integration dates to within 16% of the total training data timespan [150].

This framework can be applied to better understand the dynamics of latency and reservoir persistence by tracking the rise and fall of different viral quasispecies on therapy and in the context of TI.

1.7 Droplet microfluidics

Modern molecular biology's demand for ever smaller scales, higher throughput and more data has forced technological advances in assay miniaturisation. Despite a steady stream of innovation, only small increments in throughput have been made. This is driven largely by the physical limitations of using plastic consumables. Wells have been reduced to a size that will only just accomodate the end of a pipette tip and have forced manufacturer's to solve fluid dynamics problems to be able to transport reagents. These barriers have been continually beaten down to the point we are at now where there is little scope for further improvement. In order to progress past the current cutting edge, a different approach is needed.

One area showing real promise is that of microfluidics. This chapter introduces the field and details ways in which a shift away from the macro-world has the potential to solve many barriers facing the continued miniaturisation of life science assays.

1.7.1 Microfluidics and emulsions

Microfluidics is "the science and technology of systems that process or manipulate small amounts of fluids, using channels with dimensions of tens to hundreds of micrometres" [151]. It has been almost two decades since it was first proposed that microfluidics harboured the potential to revolutionalise methods in the life sciences [152] and it is no coincidence that this technology has come to the fore at the same time as the field has reached a plateau in terms of increasing throughput with plastic consumables. Multi-well plates are now available with 1536 wells. However, as the 'big data' and single-cell revolutions continue to develop, there is an urgent need for a step-change in technology to allow higher throughput assays to be developed.

Microfluidics in its simplest guise can be seen in a number of commercial products currently available to a researcher in the life sciences:

- **Agilent's Bioanalyzer** has been available since 1999 and was among the first commercial instruments to use microfluidic technology for the analysis of biological samples (fig 1.8.a)[153]. The system utilises a series of small channels to perform gel electrophoresis on the micro-scale allowing greater sensitivity and accuracy when compared to standard electrophoresis and at a lower price point when compared with capillary electrophoresis [154].

- **Fluidigm's integrated fluidics circuits** can be used to perform multiple parallel PCR reactions on one chip. The highest capacity chip is able to run 9,216 parallel PCR reactions (fig 1.8.b)[155].

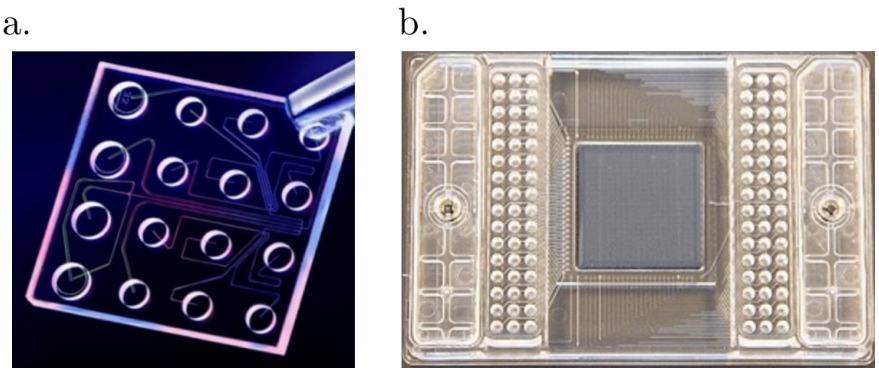


Figure 1.8: Commercial microfluidics chips (a) An example of the microfluidics chips used in Agilent's Bioanalyzer instrument. (b) An example of one of Fluidigm's integrated fluidics circuits.

One key feature of microfluidic devices that enables them to be used for such complicated systems is the predictability of flow that is not present in larger channels. This allows for device characteristics such as concentration gradients between two laminar flows and the temporal resolution of reactions governed by the controlled diffusion of reagents [156]. This feature is the result of a different set of forces becoming dominant on the microscale when compared to those determining fluid movement around us on a day to day basis. The effects that begin to dominate include:

- Laminar flow

- Fluidic resistance
- Surface tension
- Surface area - volume ratio

In order to determine whether a microfluidic device will operate on the microscale the Reynold's number of the fluid flow within the channels must be calculated. This number dictates whether the fluid flow is turbulent and consequently unpredictable, or laminar where two streams flowing in parallel will remain distinct except for mixing through diffusion [157]. The Reynold's number of a fluid flow can be calculated using:

$$Re = \frac{\rho \upsilon D_h}{\mu}$$

where ρ is the fluid density, υ is the average velocity and D_h is the hydraulic diameter [152]. Flow rates within microfluidics channels between 1 µm and 100 µm will commonly have flow rates between 1 µm/s and 1 cm/sec. These relatively low flow rates coupled with the small channel width typically lead to low Reynolds numbers [158]. A Reynolds number <2000 generally indicates a laminar flow with the flow becoming more turbulent as the number passes 2300 [152, 159].

Physical microfluidic chips have led to a large increase in capacity and throughput for a number of applications; for example, 384 PCR reactions using traditional consumables has been expanded to 9,216 by utilising microfluidic technology. Once again however, the need for plastic consumables in which to house these individual reactions is approaching a finite limit meaning these technologies may not be the step change required to future-proof biological assays.

Another series of platforms borne from microfluidics are those with the ability to generate dense emulsions. An emulsion is defined as a mixture of two immiscible liquids, where one liquid is dispersed inside the other, forming a droplet [160]. By dispersing an aqueous solution within an oil, it is possible to generate millions of individual droplets where each can be considered the equivalent to a single well on a multiwell plate. Coupled with droplet generation rates in the region of 1-20

kilohertz (kHz), these microfluidics systems have the potential to easily surpass current plastic based technologies in terms of throughput. However, one limitation of these emulsions is that they are unstable due to their large interfacial surface area and require additional agents to stabilise them [160].

Another consideration when assessing an emulsion's suitability as a microreactor is the ability to retain and utilise the droplet's contents in downstream applications. Despite not always being necessary (for example in Biorad's ddPCR system where droplets are stabilised to the extent that they are not easily broken) it is often beneficial to be able to rescue the sample from within the droplet. This is possible through a process known as emulsion breaking or resolution (named as the result is to separate the emulsion into its constituent parts). Emulsion breaking can be achieved by a variety of methods including: chemical, electrolytic and physical means [161].

The potential gains in throughput and cost from transferring biological assays to droplets are immense. In order to generate 1 million individual PCR reactions using a technology such as the Fluidigm platform would require ~108 individual chips at a cost of ~£6,500 and would take ~3 days to prep and analyse. The equivalent number of dropletised PCR reactions can be generated in <20 minutes for ~£2.50, all thermocycled simultaneously and analysed in ~42 minutes. In addition, any number of additional reactions can be produced with minimal additional cost and a linear increase in production and analysis time. In theory, there is no limit to the amount of individual reactions that can be produced; the main barrier however is analysis time as this is the bottleneck of the process. There are applications where platforms such as Fluidigm have made a notable contribution to the field, for example in the fields of DNA fingerprinting and genotyping [162]. In the context of HIV reservoir studies, target cells may be as infrequent as 1 latently infected cell/million, analysing ~800 cells for single-cell RNA sequencing (RNASeq) or ~9000 reactions for PCR as in Fluidigm is therefore of little value. In order for these systems to be suitable, it would first be required to perform an upfront enrichment for latently infected cells. As discussed previously, such a method does not exist to

isolate cells latently infected with HIV. For the reasons discussed above, this thesis has set out to develop an assay using the power of microfluidic emulsions that is able to isolate, quantify and sequence cells latently infected with HIV.

1.7.2 Why the HIV latency field needs droplets

The technical challenges currently facing researchers are borne largely from two interlinked characteristics of HIV latency:

1. Latently infected cells are extremely rare, therefore interpreting signal from bulk population sequencing is problematic due to the background signal from uninfected cells.

2. Cells have no cell surface phenotype or molecular markers that identify them as latently infected with HIV.

The ability to sort a cell based on whether or not it contains an HIV provirus would have profound implications on better understanding the mechanisms governing HIV latency and discovering potential therapeutic interventions. A number of methods are currently employed to try and characterise cells based on their infection status. However, all of these require either in-vivo activation or fixation, both of which modify the immunologic profile of the cell. Were an ideal assay to be developed it would perform the following:

- Identify whether the cell is latently infected
- Assess the cell for the presence of key cell surface markers relevant to latency e.g Tim-3 & PD-1
- Analyse the transcriptomic profile of the cell to identify any common pathways that may be dysregulated
- Analyse the integrated viral DNA
- Determine whether a latent cell can be activated using common activation cytokines

In order to overcome these issues I have turned to the field of droplet microfluidics. The concept being that it is possible to compartmentalise discrete volumes of a homogenous fluid alongside individual cells and progress them through various downstream processes. Optimising this single technology has the potential to enable multiple parallel downstream applications crucial to advancing the field.

1.7.3 Droplet microfluidics revolutionising single-cell analysis

As single-cell analysis methods have matured it has become clear that the amount of heterogeneity observed, even among the same cell type, is more than would have been predicted [163, 164]. This has led to a drive for higher throughput, lower cost 'omics' assays able to provide single-cell resolution. In recent years there has been an explosion of new methods being developed utilising droplet microfluidics as tiny individual reaction vessels for biological assays [165–170]. These assays cover a wide range of research applications including RNA sequencing, cell enrichment, directed enzyme evolution, rare pathogen detection and secretome analysis. The application with the largest uptake thus far is that of single-cell RNA sequencing.

The use of droplet microfluidics for single-cell RNA sequencing involves co-encapsulating a single cell with a bead inside a droplet. The bead is studded with thousands of unique poly-t tailed molecular barcodes that bind target RNA in the droplet. Droplet emulsions are subsequently broken and the bead-linked RNAs sequenced and demultiplexed using the molecular barcodes to determine the cell of origin. An extension of this assay is to implement a sorting step prior to sequencing where RTqPCR reagents are also co-encapsulated in the droplet and the droplets thermocycled. Post-thermocycling, the droplets containing the target of interest fluoresce and can be sorted from the 'dark' droplets using a microfluidic sorting chip. At present, only RNA targets can reliably be used to sort droplets due to the difficulty in lysing the cell nucleus to access the genomic DNA.

1.8 Aims of this thesis

The aims of this thesis are outlined below:

- To develop a target enrichment assay capable of enriching HIV proviral DNA from a background of bulk CD4+ T-cells extracted from PLWH.
- To develop a novel droplet microfluidics assay capable of identifying and isolating cells latently infected with HIV.
- To develop a novel droplet microfluidics assay capable of enriching for droplets containing DNA from latently infected cells.

In terms of contribution to the field, the key potential benefits are explained below:

- **Reservoir quantification** The pipeline proposed here combines the benefits of the HIV qPCR assay with a sequence based correction for replication incompetent proviruses. This correction is applied by interrogating next generation sequence data for common defects known to affect replication competence such as deletions, polymorphisms and hypermutation.
- **Reservoir phenotyping and isolation** To date, it has not been possible to profile latently infected cells as they lack a known immunologic signature that would identify them as being so. Possibilities exist here to stain for both immune markers and intracellular HIV DNA on the same cell. Those positive for HIV DNA can then be analysed for common markers that may be used to sort latently infected cells in future experiments.
- **Reservoir sequencing and transcription profiling** Full length high-throughput cell-associated DNA sequencing remains a challenge for the HIV cure field for reasons discussed earlier in this chapter. This platform has the potential to overcome these limitations as it will utilise single cell sequencing after enriching for a pure population of reservoir cells.

- **Integration site analysis** A major disadvantage of assays designed to assess the integration landscape in HIV latency is the inability to determine if the integration site is associated with a replication competent virus. The assays proposed here aim to capture full length genomes including the integration site, thereby improving integration characterisation in HIV latency.

1.9 Thesis structure

As this thesis will focus predominantly on method development work that has been undertaken to make these assays possible, the structure will be as outlined below:

- **Results chapter 1 - Target enrichment: Method development** This chapter will highlight the modifications made to the standard Sureselect enrichment method to make it applicable to the frequency of targets observed in HIV latency. The interpretation of the results in this chapter are discussed as they are presented.

- **Results chapter 2 - Target enrichment: Results** This chapter will show the results of the modified target enrichment assay as applied to clinical samples from PLWH.

- **Results chapter 3 - Droplet microfluidics: Droplet production and manipulation method development** This chapter will introduce the reader to the fundamental steps required to produce and manipulate droplets in microfluidic devices. This will be presented as a step by step walk through the pipeline introducing the key problems encountered and the solutions that were developed to overcome them. The interpretation of the results in this chapter are discussed as they are presented.

- **Results chapter 4 - Droplet microfluidics: In-drop PCR** This chapter introduces the problems associated with transferring a PCR pipeline from plastic consumables on the macroscale to droplet emulsions on the microscale. This will be presented as a step by step walk through the pipeline introducing

the key problems encountered and the solutions that were developed to overcome them. The interpretation of the results in this chapter are discussed as they are presented.

- **Results chapter 5 - Droplet microfluidics: Results** This chapter will provide a brief overview of the current stage of development for the droplet microfluidics assays. This will include an insight into the last remaining obstacles to be cleared before advancing to processing clinical samples.
- **General discussion** A summary of the work completed in this thesis, the remaining hurdles and the potential applications.
- **General materials and methods**

Chapter 2

Target enrichment method development

Contents

2.1 Introduction 37
2.1.1 Description of the standard Sureselect protocol 39
2.1.2 Bait design 39
2.2 The standard Sureselect protocol requires modifications to work in the context of HIV proviral sequencing. 40
2.2.1 Initial testing with the standard protocol 41
2.2.2 Increasing input amount and reducing PCR cycles increases sample complexity 48
2.2.3 Unavoidable sample loss occurs during end-repair, a-tailing and adapter ligation 50
2.2.4 Increased hybridisation and wash temperatures improve enrichment efficiency whilst retaining specificity 52
2.2.5 Recommended bait concentration can be significantly reduced to save costs 57
2.3 Bioinformatic pipeline 59
2.3.1 Sequencing read alignment 59
2.3.2 Optimising the integration site pipeline 60
2.3.3 Integration site visualisation 62
2.3.4 Integration site validation 63
2.4 Summary 63
2.4.1 Limitations and further improvements 65

This chapter is based on work presented at the following:

- Poster presentation at CROI, 2017
- Poster presentation at BHIVA, 2017
- Oral presentation at Oxford Viromics conference, 2019
- Invited speaker at the annual UK Agilent User Meeting 2018.

2.1 Introduction

In the previous chapter, it was shown how quantifying the HIV DNA reservoir remains an inexact science. This is largely due to the inability of conventional methods to determine the replication competence of the viruses they are quantifying. Full-length sequencing of the HIV provirus can allow researchers to determine replication competence. However, as discussed previously, current methods rely on PCR amplification that can introduce bias or CD4+ T-cell stimulation which is not sufficient to activate all latent viruses [62]. As cells latently infected with HIV can be present at a rate as low as one in one million, the sparsity of reservoir cells observed in most people living with HIV necessitates an upfront enrichment step in order to sequence bulk cells.

Target enrichment is most commonly performed for one of two reasons:

- To improve the cost per sample of sequencing for particular targets of interest
- To allow greater NGS sequencing depth for samples that when processed in bulk are too rare to provide any data

As NGS costs continue to fall, targeted enrichment of DNA samples has increased in popularity with a number of companies offering both standard and custom bait sets. Roche's SeqCap, Agilent's Sureselect and Illumina's Nextera Flex platforms are all examples of this. Experimental methods that were historically either prohibitively expensive or technically unfeasible, are now obtainable by coupling target enrichment with NGS. The basic principles behind these protocols are largely similar. A panel of biotinylated baits are designed to be complementary to the target of interest. Extracted DNA to be tested is then sheared to small fragments and incubated with the bait library to allow hybridisation between the DNA fragments and their complementary baits. Post hybridisation, the bait/DNA mixture is added to a suspension of paramagnetic beads coated in streptavidin. The biotin attached to the baits, binds the streptavidin on the beads and forms a new three-part complex of streptavidin bead, biotinylated bait and hybridised DNA

fragment. This mixture is placed on a magnetic rack where the bead+bait+DNA complex migrates to and pellets against the magnet. Un-bound off-target DNA contained within the supernatant is then aspirated from the sample tube and discarded. The now target-enriched DNA is eluted from the beads and processed for downstream sequencing.

One key benefit of this method over commonly used amplicon based techniques for sequencing is the greater level of sequencing complexity and uniformity achieved [171]. This is borne largely from the length of the baits used in target enrichment. PCR primers are ~20 nucleotides in length and are generally very specific. One nucleotide mismatch in these primers is often enough to prevent binding from occurring, this effect is particularly strong at the 3' end where absolute homology is required for binding to occur. Target enrichment baits are designed to be ~120 nucleotides in length. At this length, the baits have more binding flexibility than the smaller PCR primers. In a previous study by Bonsall et al [172], they estimated that <80% identity was required before a drop in enrichment was observed. Target enrichment baits therefore are particularly beneficial when working with diverse targets, this is in comparison with PCR based amplicon methods which can lead to a bias of viral species that share near exact homology at the PCR primer site.

This chapter will describe a novel method using biotinylated RNA baits to capture HIV proviral DNA for subsequent sequencing. In particular, the development of this method will be detailed and a working protocol proposed. This new methodology has the potential to allow HIV researchers to more accurately quantify the HIV latent reservoir based on a sequence corrected total HIV DNA qPCR quantification. The correction will use full length sequence data and provirus integration sites to determine what proportion of total proviruses may have the capacity to cause recrudescent viraemia.

This chapter will cover the following:

- Introduce the Agilent Sureselect protocol

- Describe the modifications required for Sureselect to be applicable to latent HIV provirus sequencing
- Present data from cell-line and clinical samples that show the level of enrichment achievable using this method

2.1.1 Description of the standard Sureselect protocol

The Agilent SureSelect XT enrichment protocol (fig 2.1) uses biotinylated 120mer nucleotide molecules (baits) homologous to the target sequence to enrich a bulk DNA sample. Extracted DNA is first sheared to a length compatible with the sequencing technology to be used, in this instance, 150bp on the Illumina MiSeq 2x150bp kit. Post-shearing, end repair is performed leaving blunt 5'-phosphorylated ends. An adenine tail is added to each fragment by means of a Klenow fragment upon which an Illumina sequencing adapter is ligated. 4 cycles of PCR are performed by priming on the Illumina adapter in order to amplify the adaptor-ligated library. 750ngs of amplified library are then incubated at 65℃ with the biotinylated baits for between 16-20 hours to allow bait-DNA hybrids to form. Post-hybridisation, baits are immobilised on streptavidin beads and progressed through three 65℃ washes. These washes remove DNA fragments that are not bound to a biotinylated bait, it is this step at which the sample is enriched as only the bait-bound on-target DNA fragments are retained on the beads. A final PCR is performed at this stage to produce enough material to sequence.

Sureselect is most commonly used for human exome enrichment or other custom targets, which occur at a greater frequency than integration events in HIV infection. Therefore, for this assay to be applicable to HIV reservoir sequencing, it has been necessary to perform a series of optimisation experiments.

2.1.2 Bait design

Single-stranded RNA baits are used in this protocol to bind the target DNA. 40,044 baits, 120 nucleotides in length were designed to be homologous to regions covering the full-length HIV genome. In order to design the baits, full-length sequences were

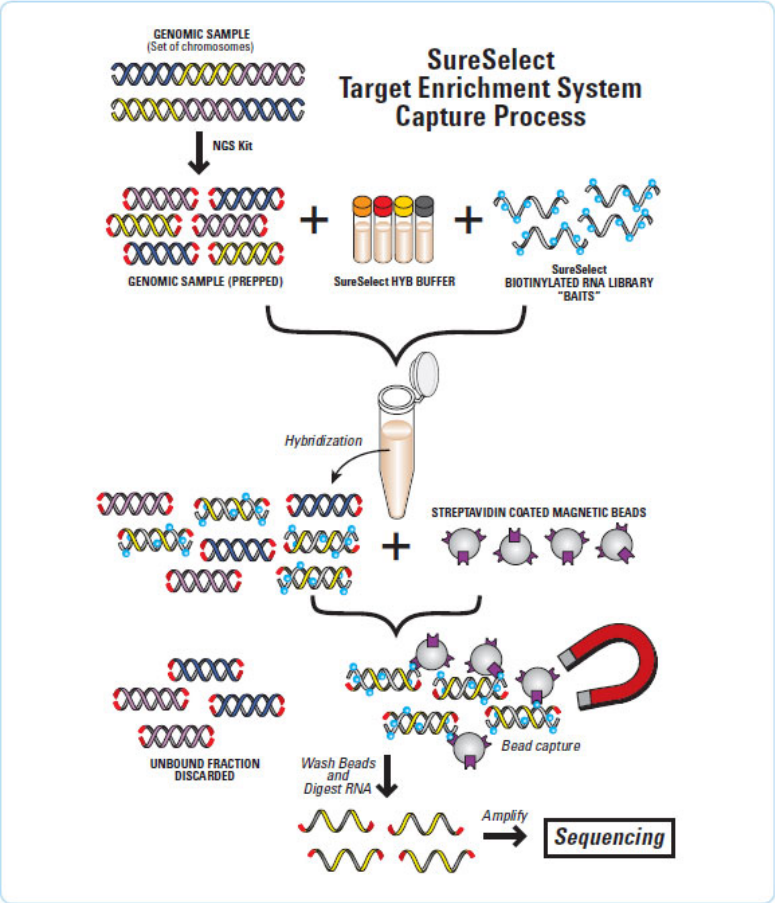


Figure 2.1: Schematic representation of the standard SureSelect workflow

downloaded from the Los Alamos National Laboratory database (www.hiv.lanl.gov) (n=675) and 53 of these were selected for the bait design process by randomly stepping through a phylogenetic tree of all downloaded sequences. The 53 selected full-length sequences were initially split into 120 nucleotide segments with each segment being used as a bait. Subsequently, starting from nucleotide position 12, the 53 sequences were again split into 120 nucleotide segments . This process was repeated until 40,044 baits had been generated (Fig 2.2).

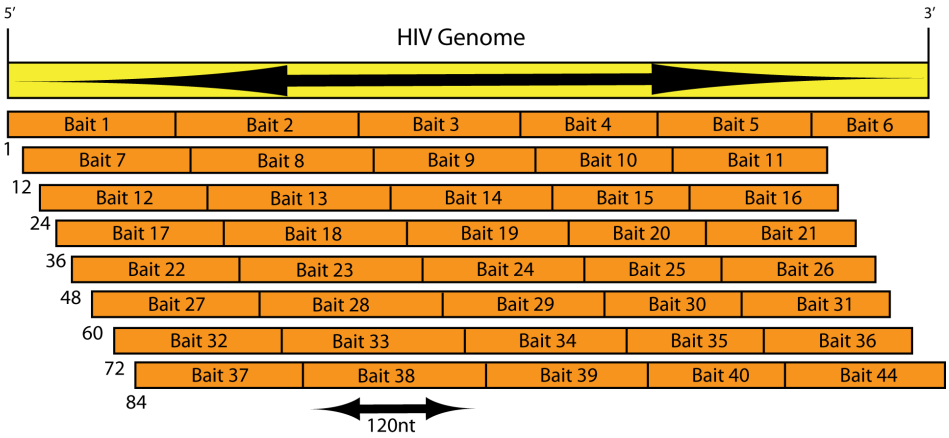


Figure 2.2: Sureselect bait tiling strategy. Schematic showing the bait tiling strategy for an example HIV genome. Orange rectangles represent individual baits spanning the entire genome. The number at the bottom left of each row of baits indicates the starting point on the genome for that line of baits.

2.2 The standard Sureselect protocol requires modifications to work in the context of HIV proviral sequencing.

The standard SureSelect protocol as depicted in figure 2.1 was originally designed for human exome enrichment and is predominantly used in situations where, if non-enriched bulk sample were to be sequenced, a satisfactory proportion of on target reads would be sequenced, albeit at lower frequencies than seen in amplicon sequencing. HIV reservoir enrichment is a different proposition. To compare, the human exome consists of ~180,000 exons totalling ~30 megabases of DNA, this equates to around 1% of the entire genome. In contrast, a single integrated copy of HIV is ~10 kilobases (0.0003% of the human genome, fig 2.3).

Using this example, were the standard Sureselect protocol to be used, a next generation sequencing run with 2 million 150bp reads would fail to sequence a single viral read due to the off-target host DNA monopolising the sequencing reagents.

In order to improve the enrichment efficiency, each step of the protocol was scrutinised for potential modifications to improve capture specificity.

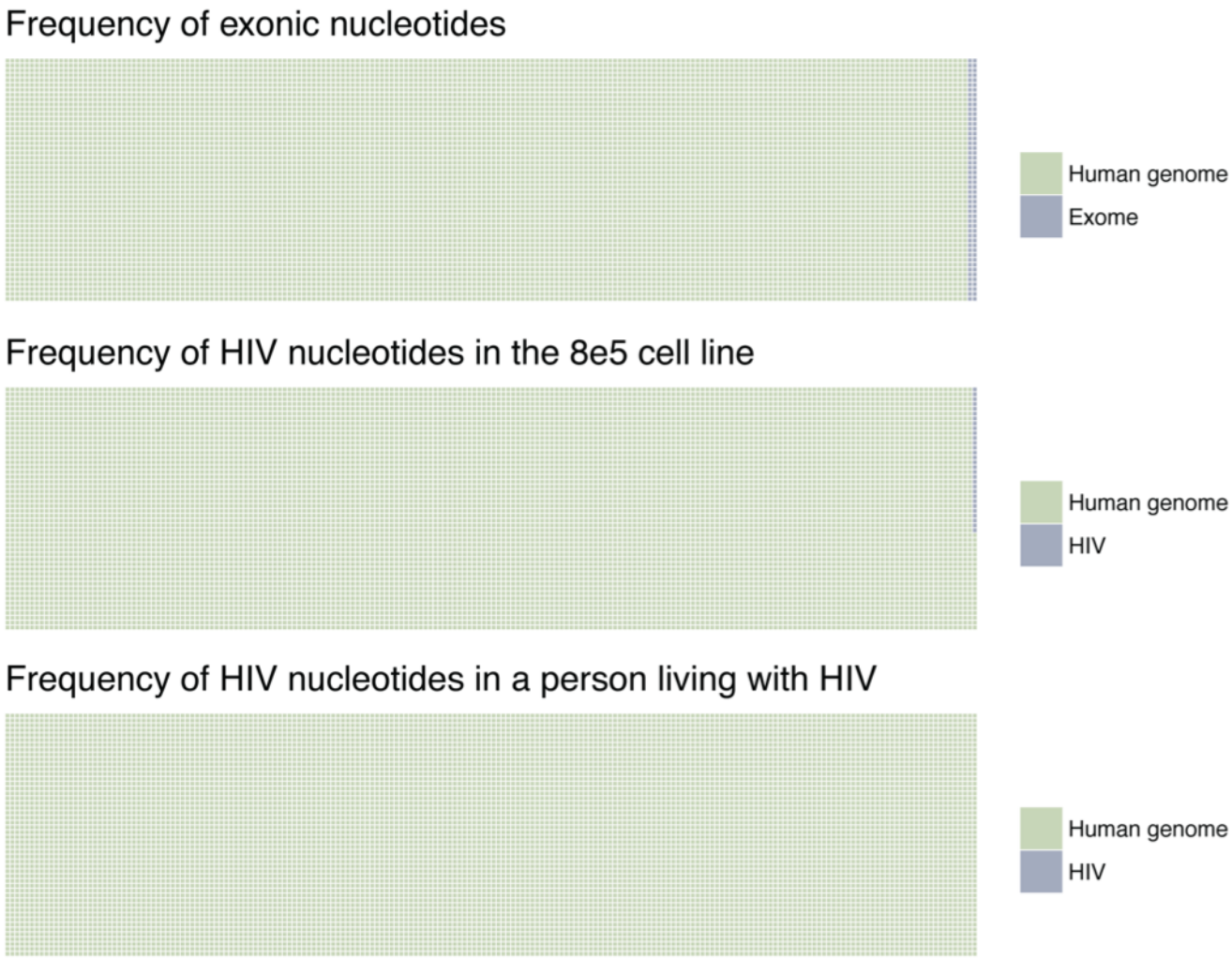


Figure 2.3: Waffle-plots showing relative nucleotide frequencies. Waffle plots indicating the frequency of either: exonic nucleotides present in the human genome, HIV nucleotides present in the 8e5 cell-line and HIV nucleotides present in a PLWH. Dark squares indicate on-target nucleotides in a total of 10^4 nucleotides.

2.2.1 Initial testing with the standard protocol

2.2.1.1 Testing on cell-line mixtures

In order to first establish the enrichment capacity of the standard protocol, it was applied, unmodified, to a series of cell-line mixtures with differing ratios of infected vs un-infected cells (table 2.1).

Table 2.1: Samples used for standard protocol testing

Sample ID	Total DNA (ng)	Percent of cells infected	Infected cell line used	Un-infected cell line/s used
U1:Jurk	3000	100	8e5	-
5:5:90	3000	10	U1/Jurkat	PBMC
8e5	3000	100	8e5	-
8e5200	200	100	8e5	-
NL43cDNA	200	100	U1	-

These mixtures ranged from a cDNA sample reverse transcribed from purified viral supernatant (NL43cDNA), to the 5:5:90 sample in which only 10% of the cells had an integrated copy. All samples used either 200 ng or 3 µgs of DNA as input as these are the two possible recommended starting concentrations for Sureselect. The protocol was followed as per the manufacturer's recommendations.

For all of the samples tested, the standard Agilent protocol was sufficient to provide near full-length coverage and an average read depth of $\geq$ 515x (table 2.2).

Samples tested using 3 µgs of DNA that contained one integrated copy in every cell (U1:Jurk & 8e5) performed well with average read depths of 18,605x and 6072x respectively. Despite promising results, these samples are not a true representation of a PLWH where the frequency of infected cells is generally between one in one thousand or one in one million. The lowest frequency of infected cells in this experiment was in the 5:5:90 sample where two infected cell lines (U1s and Jurkats) were mixed 50:50 and then spiked into PBMCs at a ratio of 1:10. The resulting

Table 2.2: Coverage and sequencing depth results

Sample ID	Number of reads	% mapped to HXB2	% mapped to human	HIV coverage (%)	Average read depth
U1:Jurk	$2.40x10^6$	15.4	60.1	100	18605
5:5:90	$0.35x10^6$	3.10	79.8	99.5	515
8e5	$0.63x10^6$	19.03	55.89	100	6072
8e5200	$2.51x10^6$	0.94	90.68	100	925
NL43cDNA	$1.08x10^6$	65.5	2.2	100	30632

sample had an infected:uninfected ratio of 1:10 with two different viral strains present in the infected population. Upon mapping to HXB2, this sample had an average read depth across the genome of 515. This value is consistent with the sample being a 1:10 dilution of the U1:Jurkat sample.

The 8e5 cell line was tested using both the 200 ng and 3 µg protocols. For the 3 µg protocol, 15x the amount of virus is used as input when compared to the 200 ng protocol. This difference is reflected in the average read depth where 8e5 provided a 20x higher average read depth compared to 8e5200. Supernatant from HIV virions was also tested using the 200 ng protocol and as would be expected, despite the low input the highest average read depth was observed (30632x). Figure 2.4 shows the percentage of total reads for each sample that mapped to either the HIV or human genome. Despite good HIV coverage being achieved for all samples tested, the percentage of reads mapping to the human genome far exceeds the percentage mapping to HIV. The exception to this was the NL43cDNA sample as it was purified from viral supernatant and should therefore not contain human DNA.

The conclusion drawn from this initial testing was that Sureselect showed promise in being able to enrich for HIV using the current bait design. Near full-length coverage with an adequate depth of sequencing was achieved in all of the samples tested. To further test Sureselect's suitability for HIV reservoir sequencing, it was necessary to test the protocol in clinical samples. This would determine

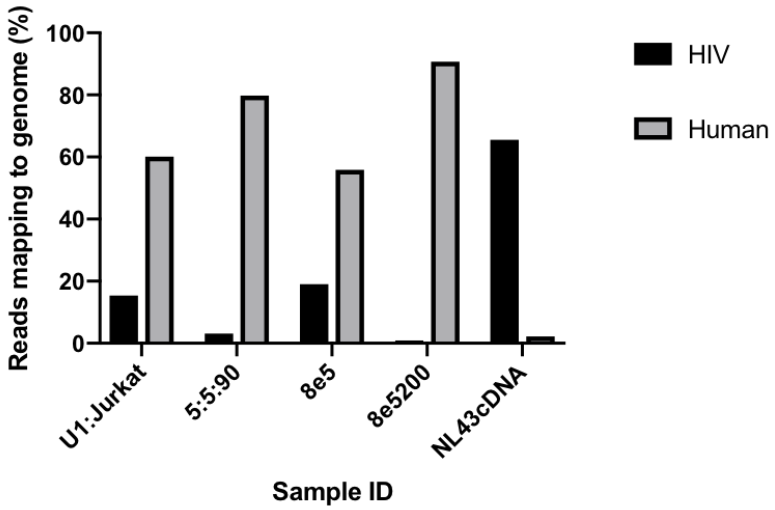


Figure 2.4: Barplot showing reads mapping to either HIV or human for cell-line testing. Barplot showing the percentage of total reads mapping to either the HIV or human genome for the cell-line testing experiment.

whether Sureselect had the required sensitivity and whether the additional sequence variation seen in clinical samples would impact upon genome coverage.

2.2.1.2 Clinical sample testing

Samples from 6 participants in the SPARTAC (See chapter 8.8.1) trial were used to test the standard Sureselect protocol. The size of the DNA reservoir varied between patients ranging from 1669.3 - 13186.7 copies/million PBMCs (Table 2.3). All participants were infected with clade B HIV and were taken from the week 108 study timepoint. All participants were on ART and consequently, all but one sample (SUQ036015) had an undetectable viral load; SUQ036015 had a viral load of 980.

Varying levels of coverage and sequencing depth were observed for the 6 samples. Two samples, SUP036078 and SUQ036015, achieved >93% genome coverage. SUM054007 and SUX036074 achieved between 75% and 85% whilst both SUA214014 and SUD036071 failed to provide >50% genome coverage (Table 2.4). The median sequencing depth followed a similar trend to genome coverage, with those samples that had higher coverage also having a higher median depth of sequencing. Similar to the cell-line tests discussed earlier, sequencing depth was

Table 2.3: Clinical samples used for standard protocol testing

Sample ID	Study time-point	DNA input (µg)	HIV copies/10^6 CD4+ cells	Viral load †	HIV clade
SUA214014	Week 108	3	4094.7	<40	B
SUD036071	Week 108	3	1669.3	<50	B
SUM054007	Week 108	3	5040	<50	B
SUP036078	Week 108	3	7723.2	<50	B
SUQ036015	Week 108	3	13186.7	980	B
SUX036074	Week 108	3	5973.3	<50	B

† Viral load measured by Roche COBAS 6800 assay

largely consistent across the genome with no major peaks observed at any particular locus (fig 2.5). A small number of peaks covering between 5-10 nucleotides were observed in all samples. These thin peaks, visible in figure 2.5, were observed at similar positions in all samples tested. This fact, coupled with the short length of the peaks suggests that they are likely an artefact of the bioinformatic alignment algorithm and may be indicative of homology between host and HIV at these positions.

HIV enrichment efficiency was positively correlated with the total DNA measurement for these samples with a Pearson correlation of 0.97 (p=0.0013, fig 2.6.b). In order to test the specificity of the pipeline, an uninfected control was run in parallel with a cell-line positive control in this experiment. The coverage plots for both can be seen in figure 2.6.a. Three reads from the negative control aligned to the HIV genome, however this frequency is consistent with the template switching phenomenon that is observed in next-generation sequencing platforms [173]. The

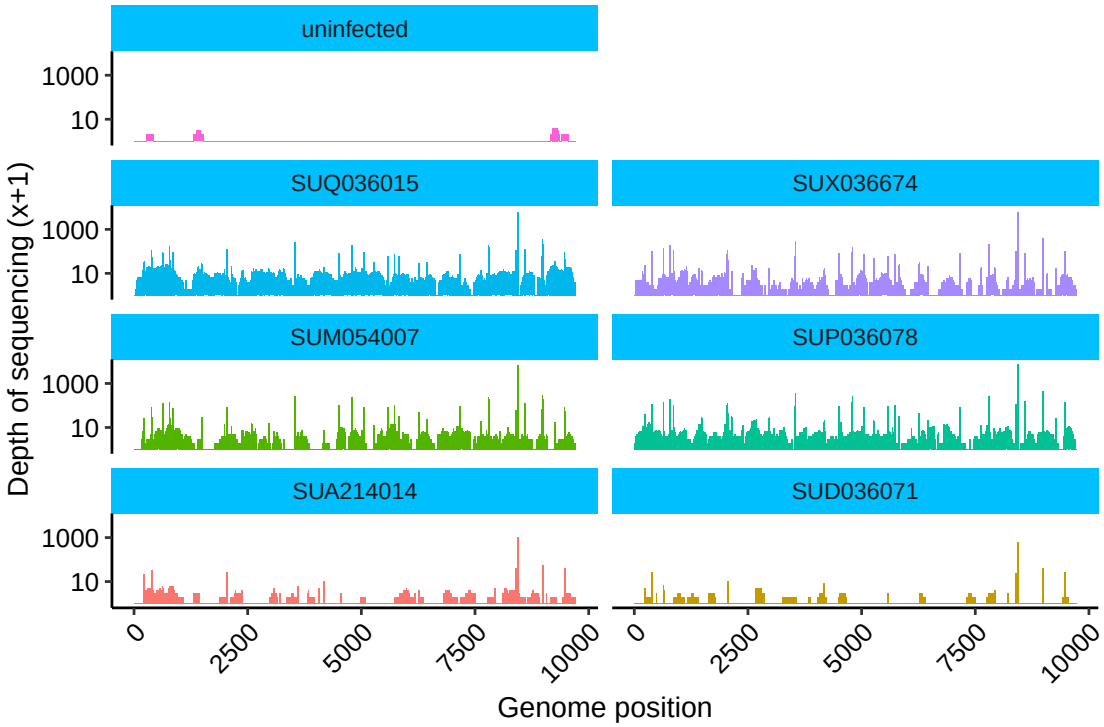


Figure 2.5: Bar plots showing sequencing depth across the genome from SPARTAC PLWH samples. Bar plots from individual PLWH samples showing the sequencing depth at each position along the HXB2 reference sequence.

positive control performed as expected.

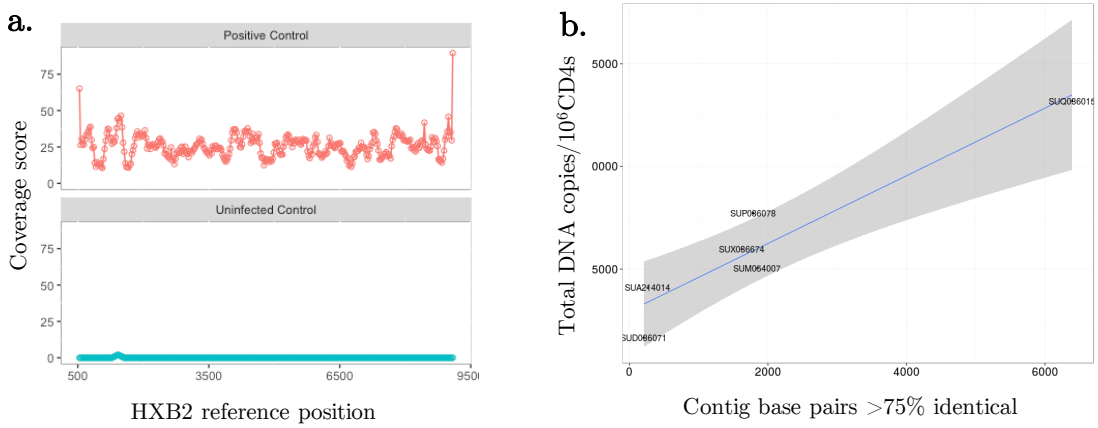


Figure 2.6: Plots showing the relationship between HIV reservoir size and observed coverage from SureSelect. (a) Dotplots showing HIV genome coverage for a positive control and an uninfected sample. (b)Scatterplot showing correlation between total DNA (as measured by qPCR) and the number of basepairs with >75% homology to HXB2 reference.

The clinical samples, all with lower total HIV DNA levels when compared to

Table 2.4: Coverage and sequencing depth results

Sample ID	Percent of genome covered	Median depth of coverage	Number of reads sequenced	Median depth per million reads
SUA214014	49.4	0	2559858	0
SUD036071	26.8	0	2220690	0
SUM054007	77.8	3	2090042	1.4
SUP036078	93.8	4	3017974	1.3
SUQ036015	96.0	7	2678486	2.6
SUX036074	82.3	3	2584421	1.2

the cell-lines produced inferior results in terms of both coverage and sequencing depth. SUM054007 was measured to have a total HIV DNA concentration of 5040 copies/10^6 cells. This value is a representative example of total HIV DNA levels seen in the clinic for PLWH on effective therapy. The inability to produce a full-length sequence for this sample therefore posits that the standard Sureselect protocol is not suitable for HIV reservoir sequencing of PLWH on suppressive ART. In order to determine whether this assay could be modified to be fit for purpose, a series of modifications were designed and tested.

2.2.2 Increasing input amount and reducing PCR cycles increases sample complexity

SureSelect attempts to accommodate for potential errors in the PCR cycle number selected by the user, by splitting the library in two at various stages in the protocol. This provides the user with two attempts to perform an appropriate number of PCR cycles so neither too little nor too much amplification occurs. These split steps however, are not suitable for extremely rare target enrichment. Splitting the sample, results in an increased risk of reducing the complexity of the sample as some quasispecies may not be represented in the half that is selected for hybridisation.

Figure 2.7.b shows a box containing the equivalent of one DNA sample. As this is a homogenous mix of on and off-target DNA, should only half of the sample be progressed, one can reasonably assume that not all of the on-target DNA will be assayed. Using the standard Sureselect method, only the viral quasispecies in the green shaded area would be taken forward into the hybridisation. In the optimised method described here, all viral quasispecies will be taken into the hybridisation. The number of upfront PCR cycles was also reduced from 6 cycles to 3 to account for the increased starting material. The reduced cycle number was crucial to ensure the minimum library size (750ngs) was achieved whilst not exceeding 2000ngs, the maximum concentration that has been successfully tested in the hybridisation step. Excessive upfront PCR cycles are detrimental as they increase both the probability of PCR error and the risk of producing over 2000ngs of DNA. In the latter case, as only 2000ngs can be loaded into the hybridisation, only a fraction of the library will be able to be loaded into the hybridisation and sample complexity will be lost (fig 2.7).

In addition, the initial sample input was changed from 3 µg to 4 µg of DNA. 3 µg is equivalent to ~4.5x10^5 cells. This relatively low cellular input, coupled with a low frequency of infected cells in a virally suppressed PLWH is a limiting factor for any methodologies of this type. By increasing the initial input amount to 4 µg the total number of cells being assayed increases to ~6x10^5 cells.

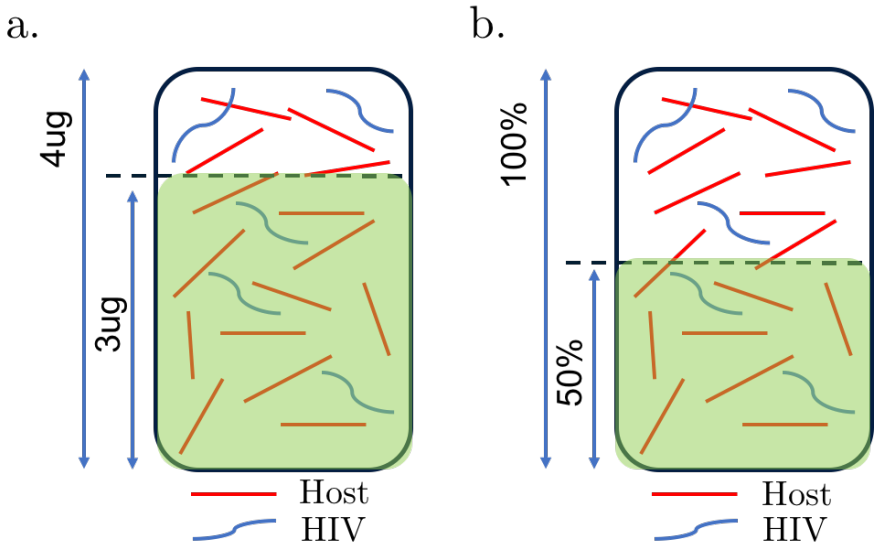


Figure 2.7: Sampling issues adressed as part of the Sureselect protocol. This figure details two modifications made to the Sureselect protocol to reduce sampling problems. The standard Sureselect protocol would only be able to capture fragments inside the green shaded areas whereas the modifications made are able to capture the entire sample. (a) A 4 µg starting sample increases complexity of the hybridisation mixture by including a greater number of viral fragments. (b) Progressing 100 % of the sample through the protocol increases the chance of capturing all species present in the mixture.

2.2.3 Unavoidable sample loss occurs during end-repair, a-tailing and adapter ligation

To be analysed using the Illumina Miseq, the following steps must be performed on the initial genomic DNA:

- Shear DNA to the required size dependent upon the sequencing chemistry to be used plus Ampure XP bead clean-up
- End-repair the sheared DNA fragment plus Ampure XP bead clean-up
- Ligate an overhanging adenine nucleotide to both the 5' and 3' end of the sheared fragment plus Ampure XP bead clean-up
- Use the overhanging adenine to ligate the generic Illumina sequencing adapter to the DNA fragment plus Ampure XP bead clean-up

It has been previously reported that substantial sample loss occurs throughout these steps due primarily to the inefficiency of the ligation reactions [174]. In order to fully understand where sample loss occurs in this process it was necessary to quantify the proportion of the inital library remaining after each of these early steps of the protocol. Upon completion of each of these steps, an aliquot of sample was taken to measure the concentration of DNA remaining. DNA concentration was measured by first staining double-stranded DNA using Picogreen, followed by measuring fluorescent intensity using a fluorometer (Fig 2.8).

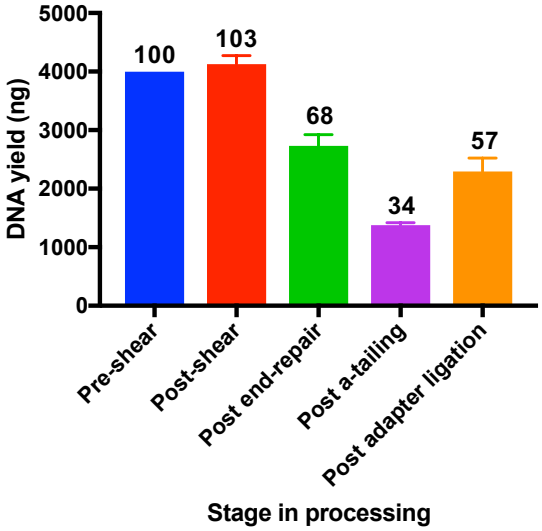


Figure 2.8: Loss of sample due to pre-hybridisation processing. This figure shows the loss of DNA observed at various stages of the pre-hybridisation process. Values displayed above bars indicate the percent of sample retained after each processing step.

In these experiments, DNA shearing, where genomic DNA is acoustically fragmented to a predetermined size had no significant effect on the DNA concentration. The next step, end-repair involves the following:

- 5' DNA overhangs are filled in by a Klenow DNA polymerase
- 3' overhangs are filled in by T4 DNA polymerase
- Blunt ends of the repaired DNA fragment are phosphorylated to enable subsequent ligation of Illumina adaptors

Following this end-repair step, DNA levels were significantly reduced with only 68% of the original 4 µg remaining. A-tailing utilises the Exo Klenow enzyme to add a non-templated adenine to the 3' end of the blunt-ended DNA fragment. The Exo enzyme is a mutated version of the Klenow fragment that retains its 5' - 3' polymerase activity but lacks any 5' or 3' exonuclease activity. Post a-tailing, another significant drop in DNA concentration was observed, with the DNA yield dropping to 34% of the original 4 µg DNA.

The final step prior to the pre-hybridisation PCR amplification is the ligation of the Illumina specific adaptors. This is achieved using a T4 DNA ligase capable of joining double-stranded DNA fragments with complementary, overhanging, single-stranded ends [175]. As shown in fig 2.8 it would appear there is a higher concentration of DNA present after adaptor ligation (57% in contrast with 34% after a-tailing). It is noteworthy however that this ligation step increases the fragment length considerably and thus will increase the DNA concentration of the sample as a whole.

These data indicate that despite ~4.5x10^5 cells being used as the starting input, only ~1.5x10^5 cells may actually make it to the hybridisation step.

2.2.4 Increased hybridisation and wash temperatures improve enrichment efficiency whilst retaining specificity

In any oligonucleotide duplex, there is a strong relationship between temperature and the strength of binding between the two strands. Lower temperatures will allow two strands with poor homology to bind whereas higher temperatures will eventually prevent binding, even of identical strands. It is this relationship that provides PCR amplification with its specificity. PCR primers are commonly designed to be ~20 nucleotides long with a melting temperature between 45℃ and 60℃. The melting temperature of an oligonucleotide duplex is influenced by a number of factors including [176]:

- Oligonucleotide sequence

- Nearest-neighbour interactions
- Oligonucleotide concentration
- Cation concentration
- dNTP concentration
- Salt concentration

As part of the Sureselect protocol, 65℃ incubations are used to provide the selectivity of the exome baits. Sureselect baits are 120 nucleotides in length and as a result have a melting temperature of ~71.2řC. 65℃ incubation and wash steps are adequate for exome enrichment, however, for very rare targets such as HIV provirus sequencing, 65℃ leads to an unacceptable amount of off-target capture. To assess the extent of the off-target capture, a series of pilot experiments (previously discussed in this chapter) were performed with the standard protocol at 65℃ to determine if it was sufficient for HIV proviral sequencing. The standard protocol does provide sufficient depth and coverage, assuming every cell contains an integrated copy of HIV. When applied to samples from PLWH with more representative reservoir levels however, full-length coverage is lost (fig 2.5).

It was hypothesised that the lack of sensitivity in samples with small reservoirs may be due in part to the permissive hybridisation temperatures used. In order to establish whether an increase in hybridisation and wash temperature would lead to an increase in sensitivity, a pilot study was performed testing a limited temperature range and assessing its effect on library yield and subsequent on-target sequencing. Initially, a test was performed on three temperatures (69, 71 & 73℃) to determine whether the increased temperature would lead to a terminal loss of library by washing the hybridised DNA from the beads. At 69, 71 and 73℃, observed library yields were 714.7, 108.1 and 64.4 nM respectively (fig 2.9). 64.4 nM is sufficient to form a sequencing library and therefore these samples were sequenced on a MiSeq to assess whether higher temperatures led to increased on-target sequencing as hypothesised.

Analysis of the sequencing output confirmed the initial hypothesis with a temperature dependent increase in HIV-mapped reads observed for every increase in temperature (fig 2.9 and table 2.5).

Table 2.5: Increasing temperature leads to a decrease in library yield but an increase in on-target capture

Hybridisation temperature (℃)	Wash temperature (℃)	Library yield (nM)	HIV reads/million reads
69	69	714.7	957.6
71	71	108.1	1882.7
73	73	64.4	2322.0

To further define this relationship, a series of experiments were performed to investigate the effect of hybridisation and wash temperature on enrichment efficiency for the 8e5 cell line. The experimental matrix tested is detailed in table 2.6.

Results from these experiments are shown in figures 2.9 and 2.10. Despite there being a PCR amplification step between the pre-hybridisation and post-hybridisation sampling points, there was a reduced number of DNA fragments containing the albumin target (used as a reference for host DNA) in the post-hybridisation sample for all temperatures tested (fig 2.10.a). No obvious trend was observed across the temperature range to suggest one condition had outperformed another in terms of enrichment. It is noteworthy that for every concentration tested, at least one of the two technical replicates saw a complete depletion of albumin.

The opposite effect was observed for the HIV LTR. All of the temperatures tested showed an increased number of HIV LTR fragments post-hybridisation when compared to the pre-hybridisation sample (fig 2.10.c). These data suggest that despite the pre-hybridisation sample having a markedly higher level of albumin (1501.3 copies/million PBMCs) than LTR (403.4 copies/million PBMCs), post-hybridisation, this situation had reversed (albumin = 5.7 copies/million PBMCs versus HIV LTR = 8.7×10^6 copies/million PBMCs). At the post-hybridisation

Table 2.6: Table detailing the experimental matrix to test hybridisation temperatures, wash temperatures and bait concetration.

ID	Hybridisation temp (℃)						Wash temp (℃)						Bait dilution			
	70	72	74	76	78	TD	70	72	74	76	78	TD	Neat	1:50	1:100	1:1000
1	X						X						X			
2	X						X						X			
3		X						X					X			
4		X						X					X			
5			X						X				X			
6			X						X				X			
7				X						X			X			
8				X						X			X			
9					X						X		X			
10					X						X		X			
11			X						X					X		
12			X						X					X		
13			X						X						X	
14			X						X						X	
15			X						X							X
16			X						X							X
17						X						X	X			
18						X						X	X			

'TD' indicates a touch down PCR from 80℃- 74℃. Each set of conditions were tested in duplicate.

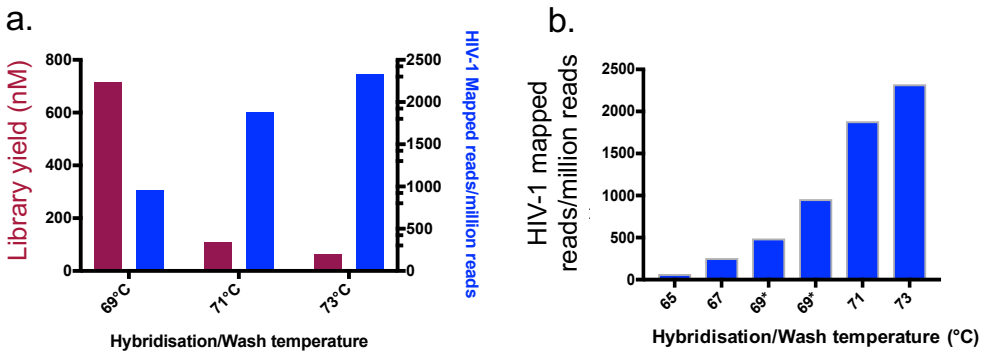


Figure 2.9: Barplots showing the effect of increasing hybridisation and wash temperature. (a) Barplot showing both the pre-sequencing library yields and the number of HIV mapped reads/million from three samples with three different temperatures. (b) Barplot showing the number of HIV mapped reads/million reads from an extended range of temperatures.

timepoint, PCR had been performed on the samples, however, as this was not HIV specific it would not bias the result in favour of the HIV qPCR target. Once again, at these higher temperatures, there did not appear to be any consistent trend towards high temperatures outperforming low temperatures when considering enrichment efficiencies (fig 2.10.a,c&e).

It is worth noting here that for the temperature range tested, a high level of evaporation was seen for a number of samples (fig 2.11.a). Due to the prolonged incubation at high temperatures that samples undergo as part of the Sureselect protocol, evaporation is a concern. In order to quantify potential losses, a series of experiments were perfomed using: different temperatures, different plastics, different positions on the thermocycler and different thermocyclers (ABI Veriti, MJ Research PTC100). Water was used as a surrogate for the hybridisation reaction mixture and 27µL was placed into either an ABI Veriti or MJ research PTC100 thermocyler for 20 hours. Figure 2.11.b shows that little evaporation was observed at the temperatures tested in either of the thermocyclers. In addition, both the type of tube and position in the thermocycler were tested to identify any potential link to increased evaporation. 27µL of water was placed into either Appleton Woods or Agilent tubes and incubated at 80°C for 20 hours in different areas of an MJ research PTC100 thermocyler (fig 2.11.c). Following incubation, levels remained

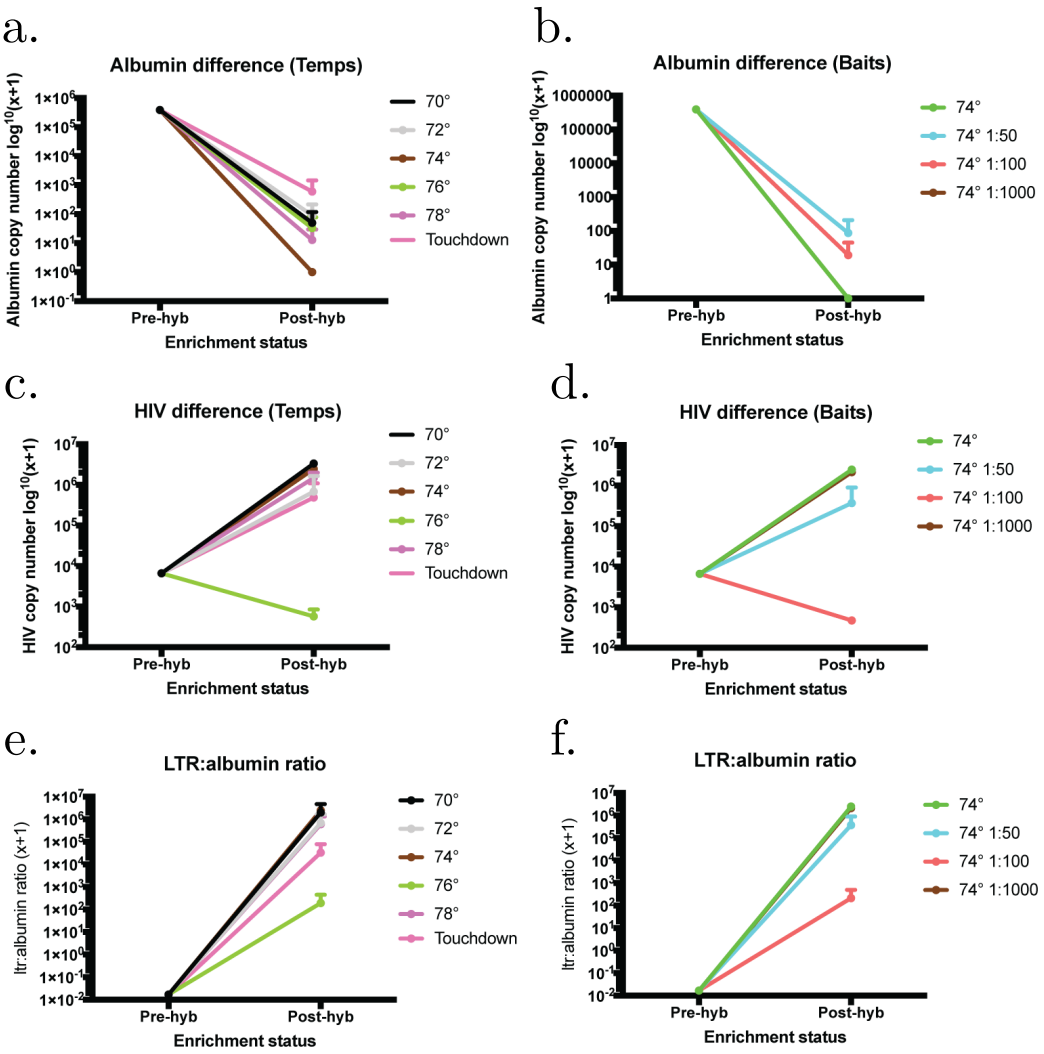


Figure 2.10: Quantification of the human albumin gene and HIV LTR to determine enrichment efficiency. (a,b) Plots showing the difference between pre-hybridisation and post-hybridisation levels of albumin for the temperature and bait optimisation experiments respectively. (c,d) Plots showing the difference between pre-hybridisation and post-hybridisation levels of HIV LTR for the temperature and bait optimisation experiments respectively. (e,f) Plots showing the LTR:albumin ratio between pre-hybridisation and post-hybridisation samples for the temperature and bait optimisation experiments respectively.

within acceptable limits (>24µL) with two notable exceptions; the bottom right and bottom left Appleton Woods tubes lost >5µL (fig 2.11.d). Upon closer inspection of the thermocycler post-incubation, a reason for these low values was observed. Due to the hinge position of the thermocycler, the two Appletwon Woods tubes in the bottom corners were not being compressed by the heated lid. As a result, this would have allowed for an increased level of evaporation from these two tubes and therefore did not give cause to modify the protocol. Based on these data, it is likely that the reagent transfer steps whilst the tubes remain on the thermocycler are responsible for the increased level of evaporation observed in sample tubes. This problem has no straight forward solution and as such must be taken into consideration when deciding upon temperatures to be used in this assay.

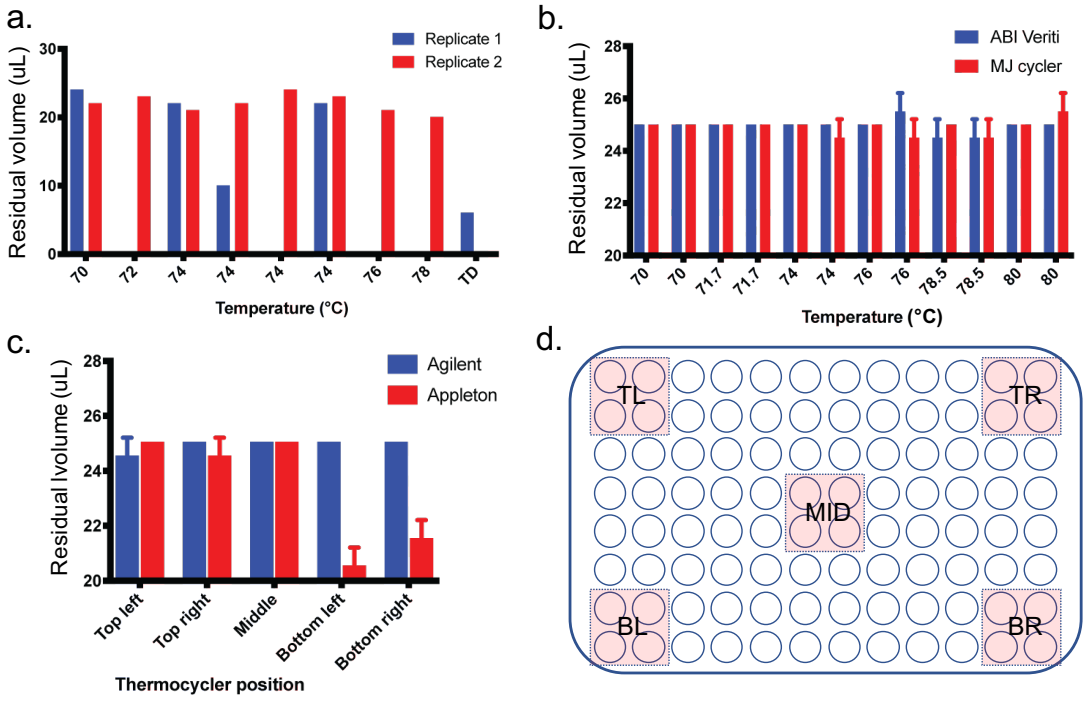


Figure 2.11: Plots showing residual volume following hybridisation at different temperatures and using different plastics. (a) Bar plot showing residual volumes post-hybridisation from the temperature optimisation experiment. (b) Bar plot showing residual volumes following water control experiment using two different thermocyclers. (c) Bar plot showing residual volumes following water control experiment using different plastics and thermocycler positions. (d) Plate setup for thermocycler position experiment (TL = Top left, TR = Top right, MID = Middle, BL = Bottom left & BR = Bottom right).

2.2.5 Recommended bait concentration can be significantly reduced to save costs

Sureselect bait libraries are proprietary and as such are not supplied with a known stock concentration. The recommendation for a capture target the size of HIV is to use 2µL of the bait stock per 27µL reaction. It was hypothesised that due to the redundancy built into the baits it would be possible to reduce the concentration used for each individual capture . This was tested by using three different bait concentrations to enrich the 8e5 cell-line (1:50, 1:100 & 1:1000 dilutions, (fig 2.10.b, d & f).

Reducing the bait concentration for the capture experiment did not negatively affect the enrichment efficiency of the assay (fig 2.10.f) A reduction in efficiency was observed for the 1:100 bait dilution however this sample exhibited a level of evaporation higher than those in the other conditions which would explain the reduced capture efficiency. No increase in either coverage or depth of sequencing was observed for this experiment however reducing the bait concentration allows for a considerable reduction in the cost per sample for this assay.

2.3 Bioinformatic pipeline

2.3.1 Sequencing read alignment

In addition to optimising the number of reads that are taken on to the sequencer, it is also critical to ensure that viral reads are not being lost during the subsequent bioinformatic processing. The diversity of HIV makes read mapping a particularly difficult prospect. Many organisms are simple to align using a reference sequence due to their well conserved genomes. The polymorphic nature of HIV can result in large heterogeneity within populations and consequently, viral reads can be missed during mapping. Most sequence aligners will fall under one of two umbrella terms that describe their mode of action: de-novo alignments and reference-based aligments. De-novo alignments are often favoured for highly diverse genomes such as HIV. In the context of proviral sequencing however, the read depth per sample is often insufficient to reliably perform de-novo alignment with a sequencing depth

of ~50x generally being required [177]. The performance of two reference-based aligners was tested: BWA-mem [178] (tested using both the single HXB2 reference sequence and a multi-fasta file containing 53 full length genomes sequenced from PLWH) and Novoalign [179]. A number of other reference-based aligners were excluded for various reasons that made them incompatible with HIV sequence analysis, for example unable to handle deletions.

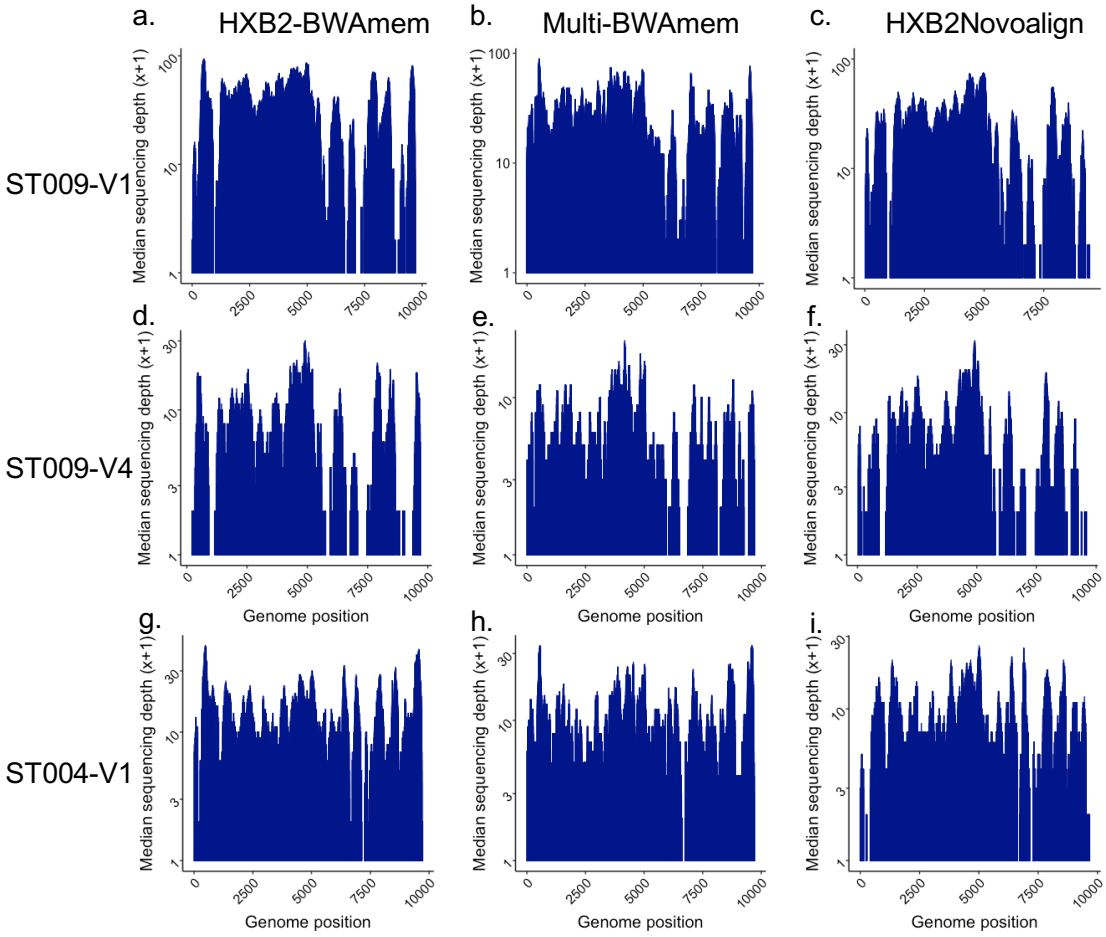


Figure 2.12: Aligner testing (a)

Alignment performance was tested by running the three alignment algorithms on three different PLWH samples. Coverage bar plots for each sample using each different aligner are shown in figure 2.12.

For well conserved regions of the genome, performance of the aligners was comparable. Poorly conserved areas such as ENV however, benefited from the extra

diversity provided by the multi-fasta approach. For all three samples, the BWA-multi alignment returns both a greater depth of sequencing and more complete coverage of the genome. Both HXB2-BWAmem and HXB2Novoalign for example have gaps in a highly diverse region around 7000bp. Multi-BWAmem however has improved coverage around this area however as the additional diversity represented in the reference sequence is able to align more of the sequencing reads to that area. Novoalign performance was better than BWA-HXB2 but was outperformed by BWA-Multi for this dataset. All future samples were therefore analysed using the BWA-Multi pipeline.

2.3.2 Optimising the integration site pipeline

The post-hybridisation reaction mixture consists of four potential species of DNA fragment (fig 2.13). One of these fragment types (chimeric fragments), enable the identification of HIV integration sites in the sample. Chimeric reads are composed of varying proportions of both HIV and host DNA and are found where an HIV genome has integrated into the host.

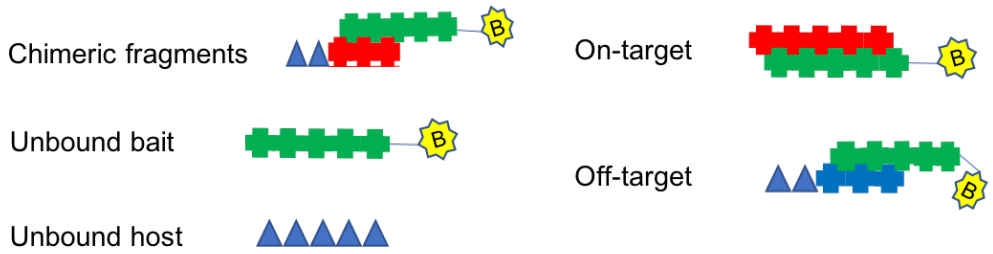
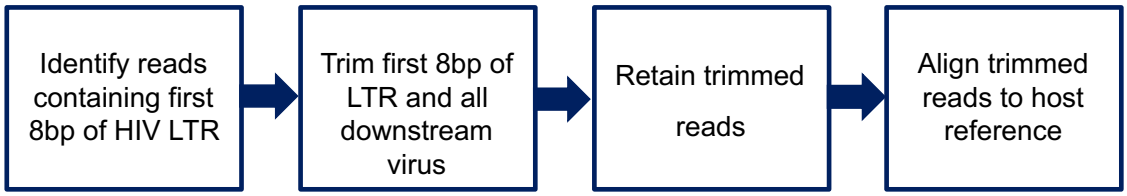


Figure 2.13: RNA/DNA species present after Sureselect hybridisation An overview of the four different DNA fragments and the unbound RNA baits possible following hybridisation. Green crosses represent baits, red crosses represent HIV nucleotide strings, blue triangles represent host nucleotide strings, blue crosses represent host nucleotide strings with partial homology to HIV and yellow stars represent biotin molecules.

Mapping of these fragments is problematic as the aligners will penalise the portion of the DNA fragment that does not match the reference sequence being used. Consequently, it is necessary to use a different bioinformatic pipeline whilst trying to identify these integration sites.

A simplified version of this pipeline is shown below:



This pipeline was designed to isolate, extract and map any HIV integration sites captured in the hybridisation reaction. This was achieved by first aligning all reads to the LTR region of the HXB2 reference sequence in order to look for any mutations that may result in lost integration sites in the following step. Post profiling, any read that contains the first 12bp of the LTR was identified and retained in a new BAM file. Subsequently, each read in the new BAM file had all viral nucleotides removed, this was achieved by identifying the first 12bp of the LTR and removing them. All nucleotides downstream of these 12bp were also removed leaving only the human insertion site remaining. This step was crucial in ensuring the maximum number of chimeric reads were aligned to the integration site (Fig 2.14.a). In a previous study by Sunshine et al [180], this step was excluded and the entire chimeric read was mapped to hg19 leading to a drop in the number of reads aligning to the integration site. The new trimmed BAM was then aligned to the human hg19 reference genome to establish the locus of the integration. Candidate integration sites produced using this method were assessed for potential false positives by ensuring it satisfied the following criteria:

- The candidate did not map on top of a human endogenous retrovirus (HERV).
- The candidate did not map in a repetitive region, these were indicated using the repeatmasker annotation built into the Integrative Genomics Viewer (IGV) [181].
- The candidate read had at least 20 nucleotides to map to hg19.
- The candidate read did not have a strong secondary mapping position in hg19 (>80% homology).

One limitation of this method for identifying integrations is that not all HIV proviruses will be intact and many may be missing one of the LTR regions. These truncated integrants will not be detected using this pipeline as it only looks for the first 12bp of the LTR. As this project is particularly focused on the replication competent reservoir these losses are acceptable. In order to include truncated integrants in the analysis, the pipeline could be modified to search for any HIV nucleotide string, not just the 12bp of the LTR, and subsequently trim that portion. It is noteworthy that this modification would require additional computing time and would likely involve a tool other than the 'grep' command which is currently used to identify the 12bp of LTR.

2.3.3 Integration site visualisation

There are currently no suitable tools available for clearly visualising integration sites. To address this, a perl script was written that allows a clear pictorial representation of the integration site and the associated sequencing reads called 'Breakalign'. Breakalign takes a small portion of a reference sequence and a fasta file of trimmed, and aligned reads. The script prints the nucleotide sequence of the reference portion across the top of the page and the aligned reads underneath so they are vertically aligned with the reference (fig 2.14). Visualising integration site data in this manner enables greater confidence in calling true integrations versus artefact. For example, in figure 2.14.b a cell line integration is shown. Due to the clonal nature of this integration it has numerous reads covering both ends of the virus and the host nucleotide string begins at the same location for all reads. This visualisation provides strong evidence that an integration is present at this locus.

2.3.4 Integration site validation

The 8e5 cell line is clonal and is reported to have a single integration site. In order to test this, sequence data from the earlier cell-line experiments was used to test if the pipeline could identify the 8e5 integration in the human genome. It was confirmed that the 8e5 cell line does contain one integration site at chr13:68060035.

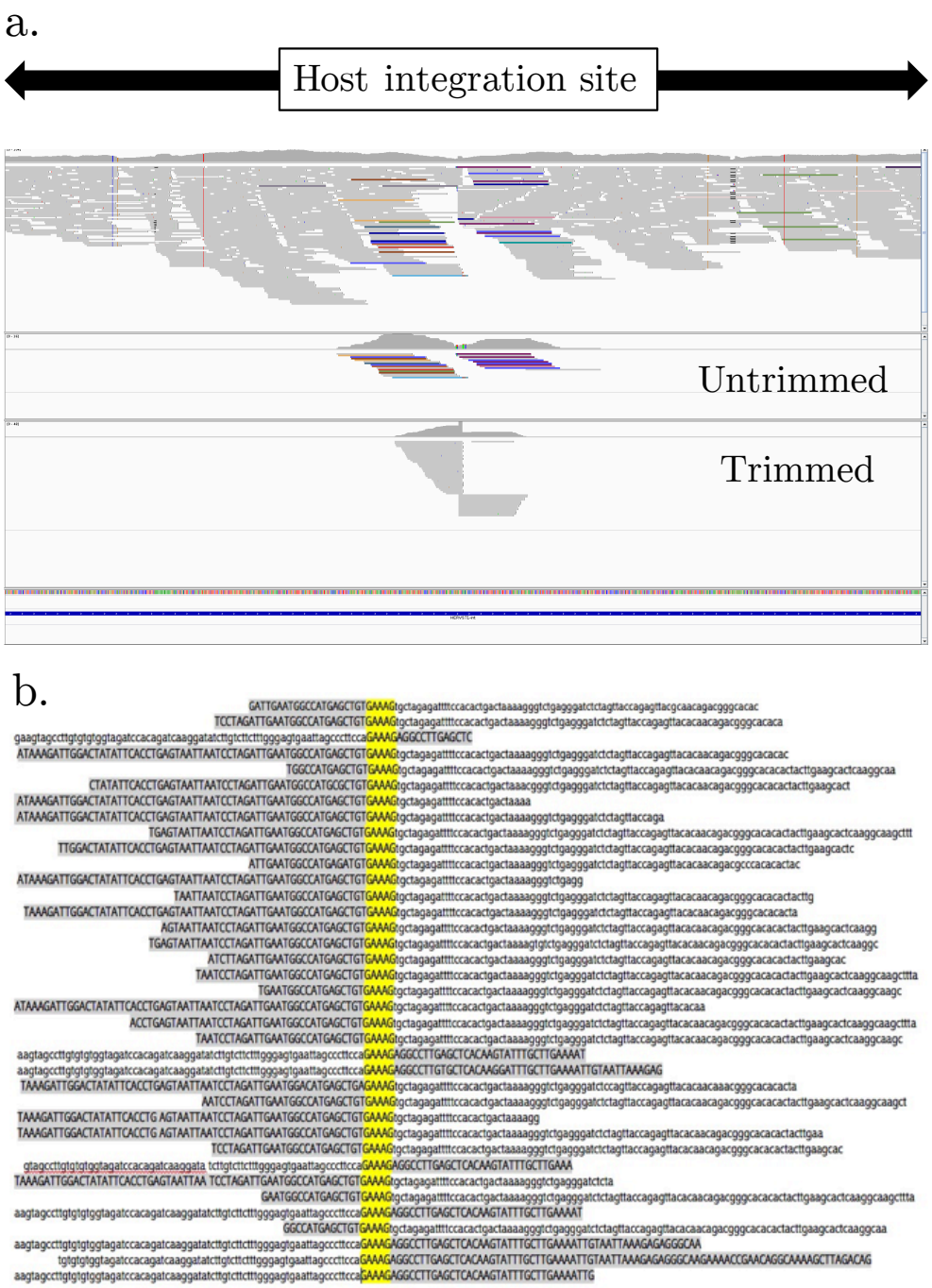


Figure 2.14: Integration site identification and visualisation (a) A represenative screenshot from the Integrative Genomics Viewer software (IGV) of an HIV integration event showing the read alignments before and after LTR trimming. (b) Integration site visualisation using Breakalign. Uppercase letters indicate host nucleotides, lowercase letters indicate HIV nucleotides and yellow highlighted letters indicate the target-site duplication common to retrovirus integrations.

Identifying integration sites in samples from PLWH on ART however, poses a different challenge, as there are likely only a small number of reads covering each integration site. The reason for this is that in clinical samples, clonal populations of HIV have a lower frequency when compared to cell-lines making them more difficult to detect. For example, should 100 cells be used as input for Sureselect, potentially 200x read depth would be obtained for the integration candidate for a cell-line. Should the same experiment be performed on a clinical sample, even assuming every cell contained a provirus, it is likely that each integration candidate would only have one or 1x or 2x read depth. This is due to a large number proviruses with unique integration sites being present in the reservoir. In these circumstances, it is necessary to examine the reads associated with the integration site candidates to determine whether they are likely to be real or an artefact of the pipeline.

2.4 Summary

The standard Sureselect protocol was unable to provide full-length genome coverage when assessing samples with a total HIV DNA reservoir <~5000 copies/10^6 PBMCs. For clinical samples isolated from ART treated PLWH, reservoir levels <~5000 copies/10^6 PBMCs are commonplace. As the Sureselect protocol showed promise to be able to provide the necessary data, a series of modifications were proposed and tested. Post-modification, the protocol is now able to provide full-length genome coverage at total DNA levels <500 copies/10^6 PBMCs. In reaching the final protocol, a number of modifications have been tested.

- Input template was modified from 3ug to 4ug. Due to the rarity of infected cells in the CD4+ T-cell population, using 4ug improves the chance of capturing viral sequences.

- Hybridisation input has been increased from 750ng to 1500-2000ng as loading only a fraction of the adaptor ligated library could result in the loss of viral variants.

- The hybridisation temperature has been modified to increase the specificity of the baits. As SureSelect was designed for moderate prevalence targets such as exome sequencing, binding stringency is less critical than in this application. Capturing reservoir cells requires high levels of enrichment, the higher hybridisation temperatures used here increase specificity by reducing the frequency of weakly bound off-target fragments.

- As in the point above, the post-hybridisation wash temperature was increased. As the baits are homologous to HIV, increasing wash temperatures cause weakly bound fragments to be washed from the beads.

- Due to excessive evaporation there is an upper limit on the temperatures that can be used. These should be optimised for individual labs and technicians.

2.4.1 Limitations and further improvements

The modified Sureselect protocol is now capable of providing full-length provirus sequence from ART suppressed PLWH. Despite this, sequencing depth remains too low to answer many of the questions still facing the reservoir field. Two key areas that show potential for improvement have been discussed previously, namely the sample input amount and the pre-hybridisation adaptor ligations. Here, I will briefly discuss two potential solutions to the problems facing the assay.

2.4.1.1 Droplet enriched samples to counter the template input issue

The low frequency of infected cells in PLWH coupled with the low maximum template input for Sureselect results in only a small number of infected cells being present in the hybridisation reaction. As it is not possible to add more cells to the reaction without overloading it, one possible solution is to increase the frequency of infected cells in the sample. Latently infected cells have no cell surface markers that identify them as being infected and therefore cannot be pre-sorted using a conventional cell sorter. Once DNA is extracted from these cells however it should be possible to pre-sort the DNA based on the presence of an HIV provirus. One method of achieving this is to use a modified version of the DNA sorting

methodology described in chapter 5. By pre-sorting DNA encapsulated in droplets it is possible to include more copies of HIV per 4 µg and therefore achieve a higher depth of sequencing per sample. As discussed in chapter 5, this additional work is nearing completion and when complete, will be used to pre-enrich clinical samples for HIV DNA prior to Sureselect enrichment.

2.4.1.2 Library prep methods

Since Sureselect's market introduction, a number of studies have tested its efficiency and looked for areas of improvement [164, 171, 182–186]. A large proportion of these studies have targeted the pre-hybridisation library preparation steps. In particular, combining end-repair and a-tailing steps leads to a marked increase in library yield [187]. Blunt ended ligations, as used in Sureselect, require three clean-up steps and two ligation steps to generate a library ready for hybridisation. This process can be reduced to just one reaction and one clean-up using dA extension and adaptor ligation. Using this modified step, it is possible to get a 50-fold increase in library yield from the same starting template [187]. By adopting this step into the modified Sureselect protocol it should ensure pre-hybridisation losses are minimised and lead to greater sample complexity.

Chapter 3

Target enrichment results

Contents

3.1 Context 68
3.2 Sequence quality and cross-contamination assessment 68
3.3 Sequencing depth and coverage 71
3.3.1 New de-novo analysis techniques to maximise HIV read allocation and build full-length contigs. 75
3.4 Integration site identification 77
3.4.1 PCR methods for identifying integration sites and full-length sequencing have the potential for bias. 78
3.5 Summary 80

3.1 Context

The transfer of reservoir assays from assaying cell-lines to samples from PLWH is not trivial due to the increased diversity and reduced frequency of viruses encountered in sequences from PLWH [180]. There is no guarantee that assays tested successfully on HIV cell lines will be sensitive enough to work on samples from PLWH. This chapter details the level of sequencing obtainable using the modified target enrichment protocol described in chapter 2 and preliminary analysis of the quality of the sequence data generated from samples taken from PLWH.

3.2 Sequence quality and cross-contamination assessment

As the protocol has been heavily modified it was necessary to ensure that the sequencing quality had not been negatively impacted. For all runs undertaken throughout this project the percentage of reads that returned a quality score (Qscore) of $\geq$ 30 was $\geq$ 79.8% (Table 3.1). A representative example of the distribution of Qscores from a typical run is shown in figure 3.1.a and similarly, a typical example of the longitudinal pattern in Qscores over the course of a run can be seen in figure 3.1.b. The two drops in Qscore observed in the first 10-20 cycles and the middle 10-20 cycles are a well described phenomena of Illumina chemistry and pose no cause for concern.

Table 3.1: Sequencing quality results

	Run 1	Run 2	Run 3	Run4	Run 5	Run 6	Run 7
Percentage of reads with Qscore $\geq$ 30	91.1	88.0	97.6	79.8	93.9	95.7	97.1

Overell, no discernable drop in sequence quality was observed from the protocol modifications.

Another potential problem for the modified protocol is to have poor index pooling. If this were to happen, sequencing reads would not be equally distributed

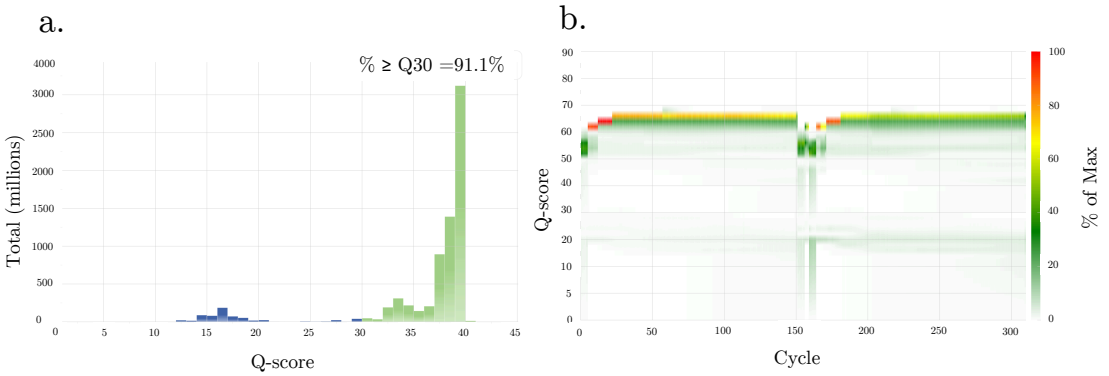


Figure 3.1: Qscores are high and remain stable throughout the entire Miseq run. (a) Barplot showing the distribution of Qscores for a representative MISeq run. (b) Dotplot showing the longitudinal Qscores across an entire 300 cycle Miseq run.

amongst the different samples loaded onto the Miseq flowcell resulting in a reduction in sequence complexity for the samples affected. Figure 3.2 shows the index distribution for 6 runs completed as part of this project. As all samples were pooled at equimolar concentrations (as per the protocol) prior to sequencing it is expected that an equal frequency of reads should be produced for each sample in the library pool. In reality, this is not the case. There are a number of instances where read distribution among the indexes is not equal. Run 5 in figure 3.2 for example has three samples that dominate the reads produced on the run. In this instance, these three samples were extracted from primary cells; the rest of the samples in this run were extracted from cell-lines. This may indicate there are processing differences between primary cells and cell-lines. For experiments that occurred after run 5, cell-lines were kept separate from patients and were not processed on the same Miseq run.

Cross-contamination is another key concern in this protocol. As it is necessary to perform a number of processing steps on a thermocycler, the chances of cross-contamination between samples increases. In addition, when processing samples with a mixture of total HIV DNA concentrations there is an increased chance of the high total HIV DNA samples contaminating those with low HIV DNA.

In order to test for contamination between samples, reads with exact matches to another sample on the same run were compared for three samples. Two of

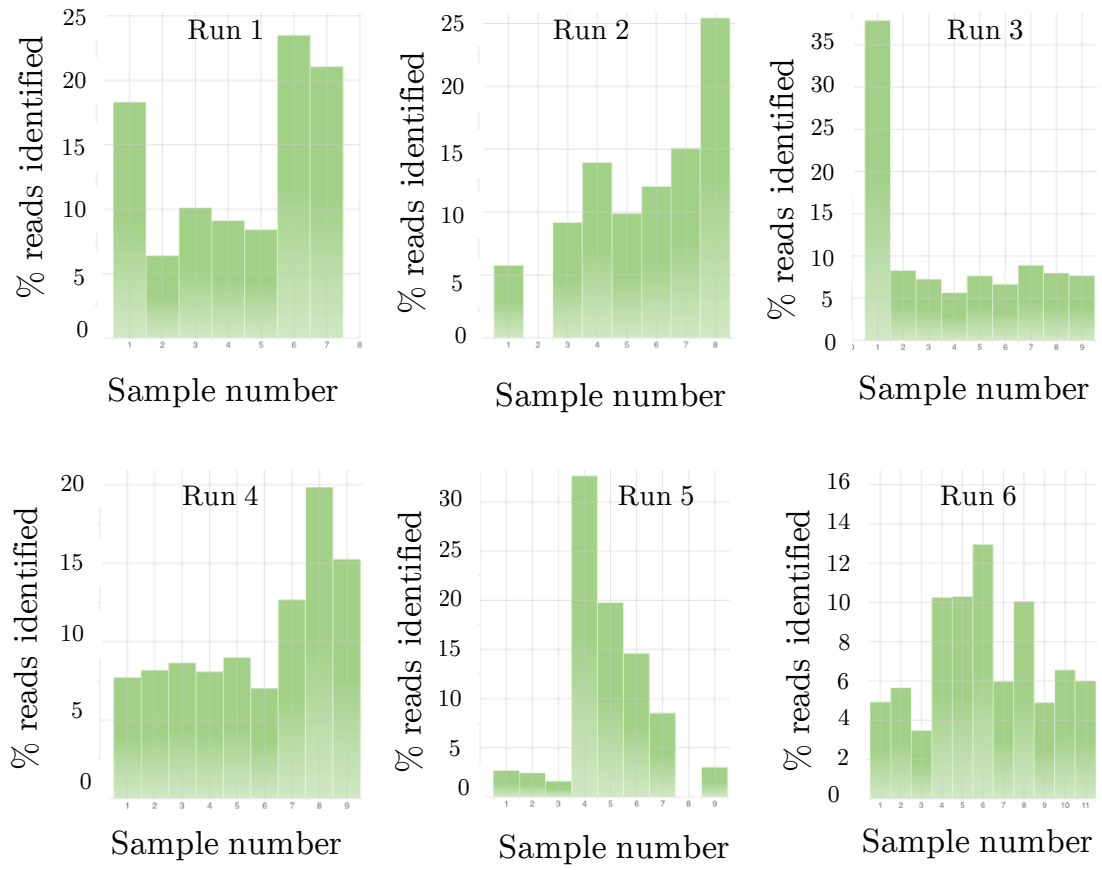


Figure 3.2: Indexing is predominantly proportional to expected input. Barplots showing the number of reads assigned to each individual barcode index for 6 Sureselect MiSeq runs on different samples. Samples in each plot are a mixture of cell-line and primary samples.

these samples were from the same PLWH taken at different timepoints and would therefore be expected to share sequence similarity. The third is an unrelated sample. As can be seen in figure 3.3, the two samples from the same PLWH (Samples 1 and 2) have extensive overlap when compared with sample 3. Although this has currently only been performed on a small subset of samples, the initial results are encouraging.

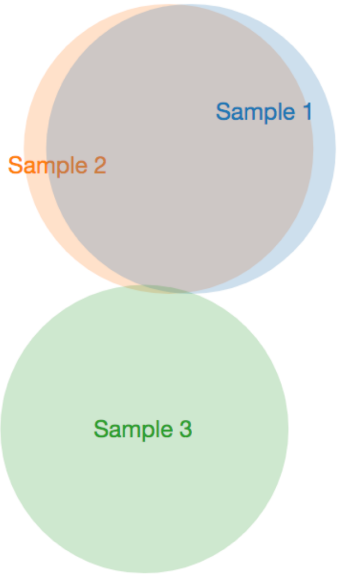


Figure 3.3: Contamination between samples is minimal. Venn diagram showing the overlap in matching reads between three samples as processed by the modified Sureselect protocol. Overlap between sample 1 and sample 2 is 348 reads, between sample 1 and sample 3 is 1 read and between sample 2 and sample 3 is 2 reads.

3.3 Sequencing depth and coverage

DNA was extracted from nine samples from the HEATHER cohort. Pertinent clinical data from the nine samples is detailed in table 3.2.

Each sample was progressed through the modified Sureselect protocol and an equimolar library pool was produced to be run on the MiSeq. Post sequencing, fastq files were filtered for quality and aligned using the Multi-BWAMEM pipeline described in chapter 3.1.

Near full-length genome coverage was observed for all nine patients assayed with median read depths of between 5x and 11x (fig 3.4 & table 3.3). Individual coverage and read depth results are shown in table 3.3. All participant samples used in this test had total HIV DNA measurements representative of PLWH on therapy in the HEATHER cohort.

Inconsistent with the cell-line testing in chapter 3.1, there was no correlation between total HIV DNA and the number of reads mapping to HIV/million reads for this set of samples ($p = 0.997$). Visual inspection of the data as plotted in

Table 3.2: Clinical data from the HEATHER samples

Participant ID	Visit	Clade	Viral load	Total HIV DNA (copies/million CD4s)
SM025	7	B	<20	1168.0
SM053	6	B	<20	1805.3
SM090	5	B	<20	922.7
ST391	6	B	<20	453.3
ST009	8	B	<20	1533.3
ST425	6	B	<20	3305.3
CW006	9	B	<20	2290.7
SM078	7	B	<20	2076.0
SM117	7	B	<20	537.3

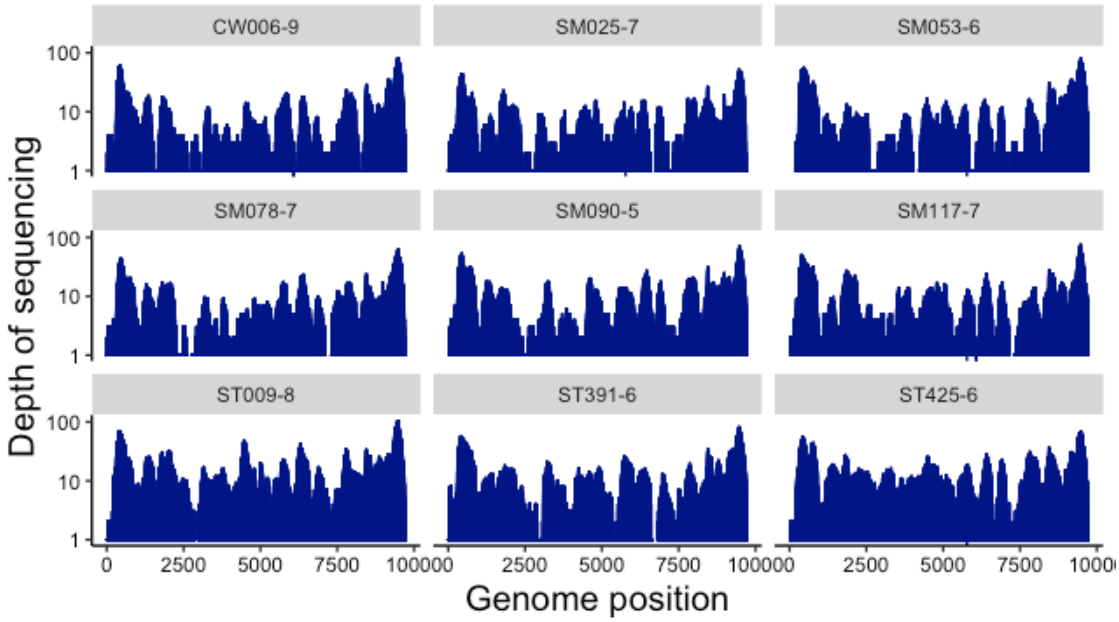


Figure 3.4: Sequencing depth produced using the modified Sureselect protocol. Barplots showing the depth of sequencing produced for 9 patients assayed using the modifed Sureselect protocol.

Table 3.3: Coverage and sequencing depth results. Table showing coverage and sequencing depth results when aligned to the HIV genome.

Sample ID	Percent of genome covered	Median depth of coverage	Total number of reads sequenced	Median depth per million reads
CW006-9	96.0	5	1019250	4.9
SM025-7	95.2	6	338880	17.7
SM053-6	92.5	6	310154	19.3
SM078-7	93.0	6	1222585	4.9
SM090-5	97.1	7	205354	34.1
SM117-7	95.7	7	379093	18.5
ST009-8	98.3	11	2472118	4.5
ST391-6	97.3	8	3970338	2.0
ST425-6	98.9	11	1813320	6.1

figure 3.5 shows two clear outliers in the dataset. It is likely that these potential outliers coupled with the small sample size mean the dataset is underpowered to draw any firm conclusions.

To demonstrate the additional sensitivity provided by the Sureselect modifications made as part of this thesis, a dataset processed using the standard protocol was compared to this dataset using the modified protocol. Two comparisons were performed:

- Total HIV DNA measurements as measured by qPCR were plotted against the median depth of sequencing
- Total HIV DNA measurements as measured by qPCR were plotted against the genome coverage achieved displayed as a percentage of HXB2.

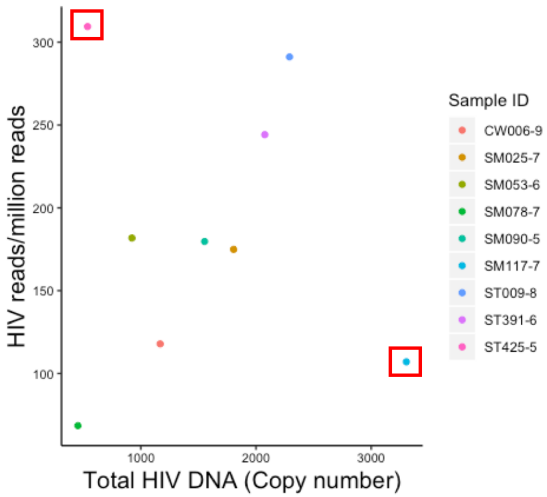


Figure 3.5: Total HIV DNA does not correlate with the number of HIV mapped reads for this dataset. Dotplot showing the relationship between Total HIV DNA as measured by qPCR and the number of reads mapping to HIV/million reads. Points in red squares indicate potential outliers.

A larger range of total HIV DNA measurements was tested for the standard protocol in order to ensure at least some level of depth and coverage were achieved for these samples. The modified protocol tested a narrow range of low total HIV DNA samples. For all samples tested with the modified protocol a median read depth of >5 was achieved. Samples tested using the standard protocol with a comparable total HIV DNA measurement exhibited a median read depth <1. To achieve a read depth of ≥5, samples processes using the standard protocol required total HIV DNA measurements of >7500 copies. This was compared to between 450 and 3300 copies for the modified protocol.

The percentage genome coverage data followed a similar trend where samples processed using the standard protocol with total HIV measurements <5000 copies produced a percent genome coverage <50%. The minimum genome coverage observed for the modified protocol, independent of the total HIV DNA measurement was 92.5%.

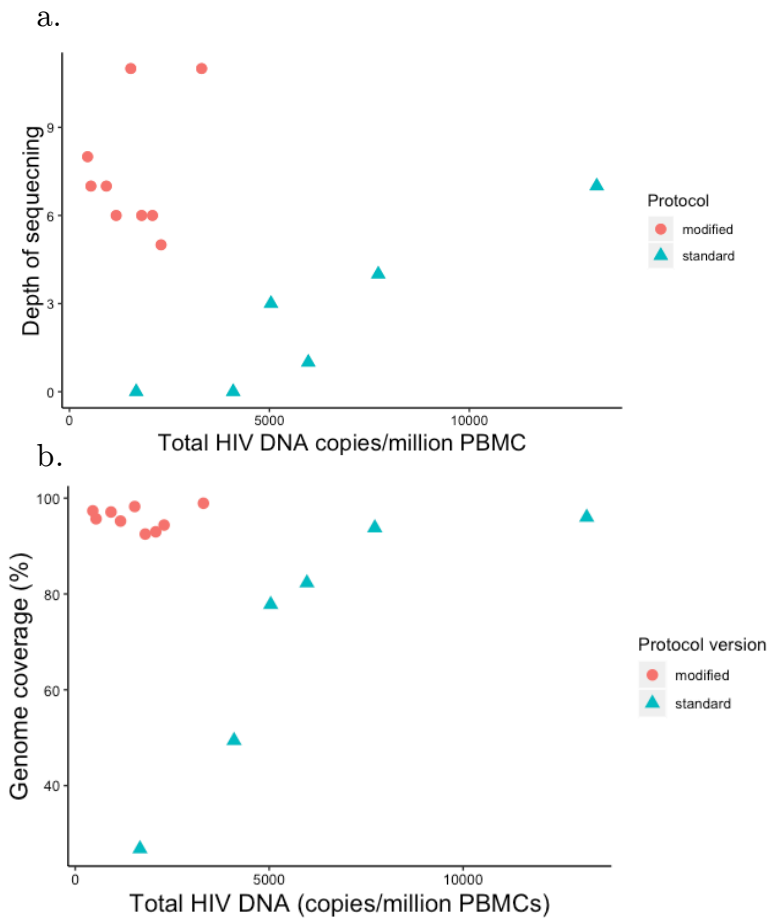


Figure 3.6: The modified Sureselect protocol improves the sensitivity of sequencing depth and increases genome coverage. (a) Dotplot showing the relationship between total HIV DNA in samples from PLWH as measured by qPCR and sequencing depth (median depth of reads across HXB2). (b) Dotplot showing the relationship between total HIV DNA in samples from PLWH as measured by qPCR and the percentage of the genome covered by sequencing. Red circles indicate samples assayed with the modified protocol and blue triangles indicate samples assayed using the standard protocol.

3.3.1 New de-novo analysis techniques to maximise HIV read allocation and build full-length contigs.

A new pipeline is in the early stages of development that has the potential to allocate more on-target reads from the raw fastq files. This is when compared to the current method that has been employed in this thesis using a reference based aligner.

Only preliminary data from a single PLWH sample is available at this point, however initial results are promising. From a PLWH sample with a total DNA measurement of 2950 copies/10^6 CD4+ T-cells, it was possible to reconstruct an

unbroken contig the entire length of the genome (Chapter 9). The output of this pipeline provides a string of nodes representing sliding windows across the genome. The algorithm builds a de-novo contig and based on other reads overlapping that region, provides alternative sequence paths or variants that are present in the sample. For example, should a clonal cell-line be analysed in this manner, the nodes would all be interconnected with no alternate pathways. Development of this pipeline has the potential to enable haplotype reconstruction from depths of sequencing lower than originally predicted.

3.4 Integration site identification

Integration site discovery was compared between one cell-line sample and one sample obtained from a PLWH. As described previously, integration sites were not counted if they mapped to a repetitive region of the human genome (as mapped by the repeatmasker annotation on IGV), if they had weak primary alignments (≤80%) or if they had strong secondary alignments (≥80%). The integration sites detected are detailed in table 3.4.

Table 3.4: Integration site candidates.

Sample ID	Integration candidates	Intragenic (%)	Number of clonal sequences
Patient sample	16	94	0
8e5 sample	2	N/A	0

16 integration site candidates were detected for the PLWH sample and 2 integration sites from the cell-line sample. 94% of the integration sites detected were identified as being intragenic. Of the integrations detected in the PLWH sample, none showed any evidence for favouring a particular gene and no evidence of clonal expansion was observed. One shared integration site was observed, the integration event known to be attributed to the 8e5 cell-line on chromosome 13. This integration was detected in both the cell-line and in the participant samples and may suggest that processing both high and low total HIV DNA concentration samples in parallel may increase the chance of cross-contamination.

LTR profiling was performed as part of the integration site identification pipeline in order to ensure there were no mismatches present in the first or last 12 nucleotides of the viral LTR sequences. An example LTR profiling pileup plot from the 8e5 cell-line is shown in figure 3.7. For all samples tested no mutations were found in the first or last 12 nucleotides of the viral LTRs.

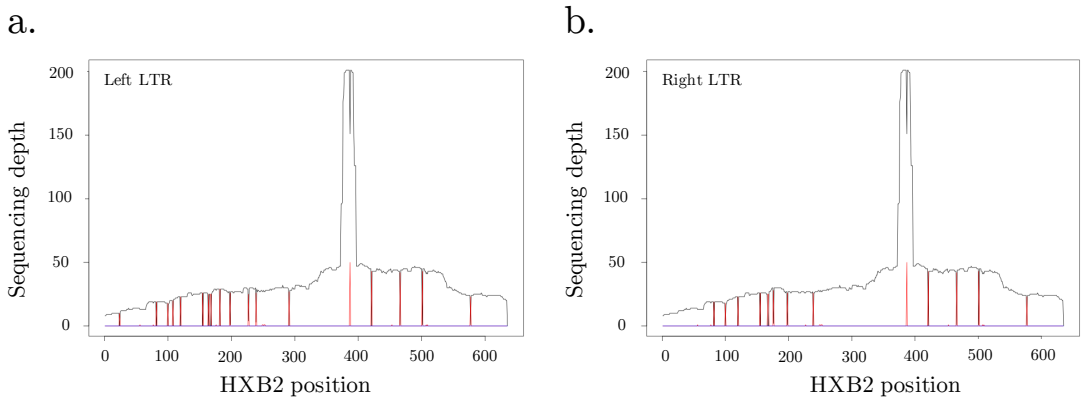


Figure 3.7: Viral LTR profiling. (a) Pileup plot showing the viral 5' LTR sequence from an example clinical sample in comparison to HXB2. (b) Pileup plot showing viral 3' LTR sequence from an example clinical sample in comparison to HXB2. Red lines indicate a deviation from the HXB2 reference sequence with the line height indicating the proportion of sequences containing the mutation.

3.4.1 PCR methods for identifying integration sites and full-length sequencing have the potential for bias.

Previous attempts to identify integration sites and to sequence full-length genomes associated with latent HIV infection have predominantly relied on PCR based methods. One potential fallibility of these methods is that they rely on the PCR primer binding site being intact. Due to the reduced stringency and increased redundancy observed in the Sureselect baits, this protocol is not dependent on primer sites being intact. Using this dataset, sequences were interrogated to determine whether or not commonly used primer sites for HIV integration site analysis and full-length genome sequencing were intact for these samples. Figure 3.8 shows 7 primer pairs commonly used in literature to sequence proviral DNA and each pair's homology to three of the sample sequences from PLWH obtained from the modified Sureselect protocol.

Of all of the primer pairs tested, assuming that a single mismatch is sufficient to hinder primer binding, not a single primer pair would have been able to amplify all of the sequences captured by the modified Sureselect protocol. Of all the combinations of primer pairs and PLWH samples analysed, only the A275F/A-3INOUT applied to Sample 3 had the potential to amplify all of the sequences

captured using the modified Sureselect method. In addition, 62% of all primer pair/sample combinations had more than one substitution present in the primer binding site and 67% had primer mismatches in over 50% of the sequences captured by the modified Sureselect protocol.

eople

Primer	Sample 1	Sample 2	Sample 3
BLOUTER-F			
BLOUTER-R	6	21	82
275F	85	83	83
280R	18	31	90
5GAGINF	91	81	21
5GAGINR	68	63	14
5ENVINF	19	35	78
5ENVINR	13	82	86
A275F			
A-3INOUT	86	86	
B-263F	91	85	83
B-3AccOut	5	5	
D-5ACCOUT			6
D-280R	18	31	90

No substitutions | 1 substitution | >1 substitution

Figure 3.8: Integration site and overlapping amplicon PCR primers may lead to biased results. Table detailing the frequency of primer mismatches observed for a number of commonly used full-length overlapping amplicon PCR primers and integration site identification primers for three PLWH samples. Green boxes indicate primer has no mismatches, orange boxes indicate one mismatch and red boxes indicate >1 mismatch. The numbers in the boxes indicate the percentage of reads at the locus with a mismatched nucleotide. Samples are from three participants who initiated therapy in the long arm of the SPARTAC trial.

3.5 Summary

The standard Sureselect protocol does not have the sensitivity required for HIV proviral sequencing. However, the modified protocol presented in this chapter has the ability to provide full-length coverage of HIV proviruses from PLWH on suppressive ART. In addition to the sequence data, it is possible to identify the sites in which the virus has integrated into the host genome. In its current state, despite providing full length coverage, the sequencing depth is low which presents problems for drawing bioinformatic conclusions from the data. It is hoped that with the application of the two potential modifications suggested in chapter 2.4.1, the sequencing depth can be increased. This would allow for a more powerful suite of bioinformatic tools to be applied to the data including:

- De-novo sequencing to further increase reads
- Haplotype reconstruction to determine replication competence
- Linked integration site and sequence analysis to provide an overview of clonal expansion during latency

Chapter 4

Droplet microfluidics: Droplet production and manipulation method development

Contents

4.1 Context 81
4.1.1 Lessons from the cleanroom 82
4.2 Method development 82
4.2.1 Microfluidic chip manufacture 82
4.2.2 WO production 85
4.2.3 Water-in-oil-in-water (WOW) droplet formation 90
4.2.4 Wetting in microfluidic devices 96
4.2.5 Chip geometry, flow-rates and surfactant concentration determine obtainable droplet sizes 98
4.2.6 Continuous phase and fluorosurfactant selection 101
4.2.7 Single-cell encapsulation in droplets 104
4.2.8 Diluting lysis reagents within droplets 107
4.2.9 Droplet reinjection 119
4.2.10 Droplet sorting 121
4.3 Summary 135

4.1 Context

This chapter details the different method development stages pursued from the initial generation of water-in-oil droplets (WOs) through to droplet sorting. The chapter is organised as a step by step walk through the droplet enrichment protocol.

4.1.1 Lessons from the cleanroom

The myriad obstacles to advancing this project from conception to reality have been technically challenging. Detailed below, are the key conceptual barriers that have had to be overcome to progress this project.

- Forming WOs in a microfluidic device
- Optimising device geometry
- Surfactant choice
- Stabilising WOs to prevent coalescence
- WO size control
- Droplet merging and pico-injection
- Creating double emulsion water-in-oil-in-water droplets (WOWs)
- WOW flow cytometry
- Emulsion breaking to release nucleic acids

This chapter describes the key observations from the method development process.

4.2 Method development

4.2.1 Microfluidic chip manufacture

The two common methods of manufacture for microfluidic chips are etching on to glass or by using soft lithography masks to generate disposable PDMS chips. As detailed below, both materials have their advantages and disadvantages:

As can be seen in table 4.1, both materials have their own desirable properties and drawbacks and often the choice as to which to use is application specific. Glass chips are often employed when a design is finalised and needs no further modification. In this circumstance, a chip can easily be washed and reused for

Property	Manufacture material	
	Glass	PDMS
Expensive	Yes	No
Washable	No	Yes
Deformable	Yes	No
Transparent	Yes	Limited
Autofluorescence	No	Low level
Gas permeable	No	Yes
Tubing-chip interface ports required	Yes	No

Table 4.1: Properties of microfluidic chip materials

different samples with no further manufacturing steps inbetween. PDMS chips are most suitable to situations where chips are under active development as they can be quickly modified and multiple different designs can be manufactured at once. Glass chips are non-deformable and the channels are etched which makes for very stable flow-rates. Most PDMS chips will have minor defects along their channels which will have subtle effects on flow-rate. In addition, a certain proportion will contain larger defects rendering them unusable. Dependent on the skill of the technician however, such large defects are rare. PDMS is deformable, which although can lead to less stable flow-rates, also means any fibres passing through the channels are less likely to create blockages. As it is non-deformable, a glass device that becomes blocked has to be washed with aggressive chemicals in order to dissolve the blockage. The relatively low cost of PDMS devices makes disposal due to a blockage a cost-effective option.

As the nature of this project was conceptual and designs would require multiple iterations with small changes, it was decided to proceed with manufacturing PDMS devices.

4.2.1.1 PDMS chip production

Manufacturing a PDMS device is a multi-stage process requiring specialist equipment and in an ideal situation, a cleanroom. In brief, a silicon wafer is baked to remove any moisture, cleaned using nitrogen gas and then one side is coated with a suitable negative photoresist (The negative photoresist, Su-8 epoxy resin, is a light sensitive material that when exposed to UV, and subsequently heated, becomes cross-linked and resistant to dissolution by a developer chemical). In order to control the depth and uniformity of the coating, the SU-8 is applied using a spin-coater. SU-8 selection is dependent upon the required thickness of the channels for the chip. Table 4.2 shows the three variations of SU-8 used in this project and the feature thicknesses they are each able to achieve.

SU-8 type	Viscosity (cSt)	Thickness (µm)
SU-8 2015	1250	21-25
SU-8 2025	4500	41-75
SU-8 2035	7000	35-110

Table 4.2: Obtainable depths using different SU-8 types.

The coated wafer is then baked, covered with an inverse patterned photolithographic mask of the chip design (fig 4.1.a) and exposed to UV light. This inverse mask is printed on acetate with UV absorbent ink covering most of the mask. Areas on the mask that do not have ink dictate where the SU-8 channel will form. A glass UV filter was secured on top of the mask to reduce the amount of channel sloping caused by Fresnel diffraction [188]. In areas where the UV is able to penetrate the mask, the SU-8 resin monomers become activated. The wafer is then heated, at which stage the epoxy groups in the activated SU-8 become cross-linked. Finally, the pattern is developed by immersing and agitating the wafer in a bath of solvent (Propylene glycol methyl ether acetate). The result of this process is a silicon wafer, known as a master, with a positive-relief network of SU-8 channels remaining on the surface (fig 4.1.b).

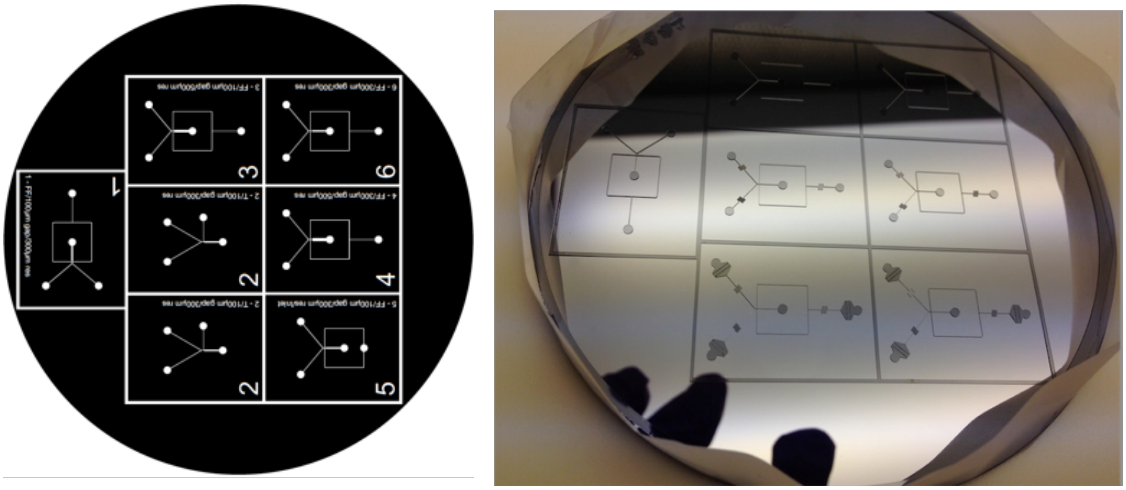


Figure 4.1: Examples of different stages in PDMS manufacture (a) Example lithography master with 7 chips printed. (b) Example silicon master with 7 positive SU-8 reliefs.

PDMS device fabrication

From a single master, multiple PDMS moulds can subsequently be created. To generate a PDMS mould, a silicone elastomer (Sylgard 184) is poured over the master and placed in a vacuum dessicator to remove any air bubbles trapped within. Once degassed, the mould is baked at 85℃ for at least 2 hours to cross-link the PDMS. Post baking, the mould is peeled from the master, cut to size and any necessary inlet/outlet holes are punched. Finally, the mould is inspected, washed and bonded to a glass slide using oxygen plasma treatment. Treating both the glass slide and the PDMS mould with oxygen plasma can improve adhesiveness by removing surface contaminants, crosslinking near-surface molecules and introducing reactive chemical groups [189, 190]

4.2.2 WO production

Generating WOs in microfluidic devices is achieved by introducing two immiscible phases at a junction in the device. At any droplet forming junction, the aqueous (dispersed) phase will pass the orifice of the junction where it meets the organic (continuous) phase and forms an interface. The interfacial tension is determined primarily by the composition of the two phases, and in particular the addition

of any surfactants to the mixture. Droplet formation can be broken down into three stages [191]:

- **Droplet expansion** - at this stage the two immiscible phases have met, the dispersed phase has proceeded beyond the junction orifice and begun to fill the main channel thereby blocking the continuous phase.
- **Necking stage** - the blocking of the continuous phase and consequent pressure buildup causes the deformation of the dispersed phase by the continuous phase and results in a thinning of the neck of the droplet.
- **Pinch-off stage** - the final separation of the droplet is determined by three forces: the interfacial tension, the shear stress acting on the droplet and the squeezing pressure across the droplet.

These three stages can be seen in figure 4.2.

4.2.2.1 Droplet forming regimes

Two different regimes of droplet formation are commonly seen in microfluidics devices: the dripping regime (fig 4.3a-c) and the jetting regime (fig 4.3d-f). Both result in stable droplets and have their own strengths dependent on the application. The two regimes can be described as follows:

- **Jetting** - The dispersed phase is squeezed by the continuous phase into a long thin jet that extends down the chip where it ultimately generates droplets with a diameter larger than the width of the jet. In this regime, the dispersed phase is driven mainly by the drag from the continuous phase and the length of jetting increases with a decreasing interfacial tension [191, 192].
- **Dripping** - After a droplet successfully pinches off, the dispersed phase first retracts and then begins to expand again forming a new droplet. A short neck is formed separating the dispersed phase and the new droplet, this neck continues to thin until the new droplet is detached [191, 193].

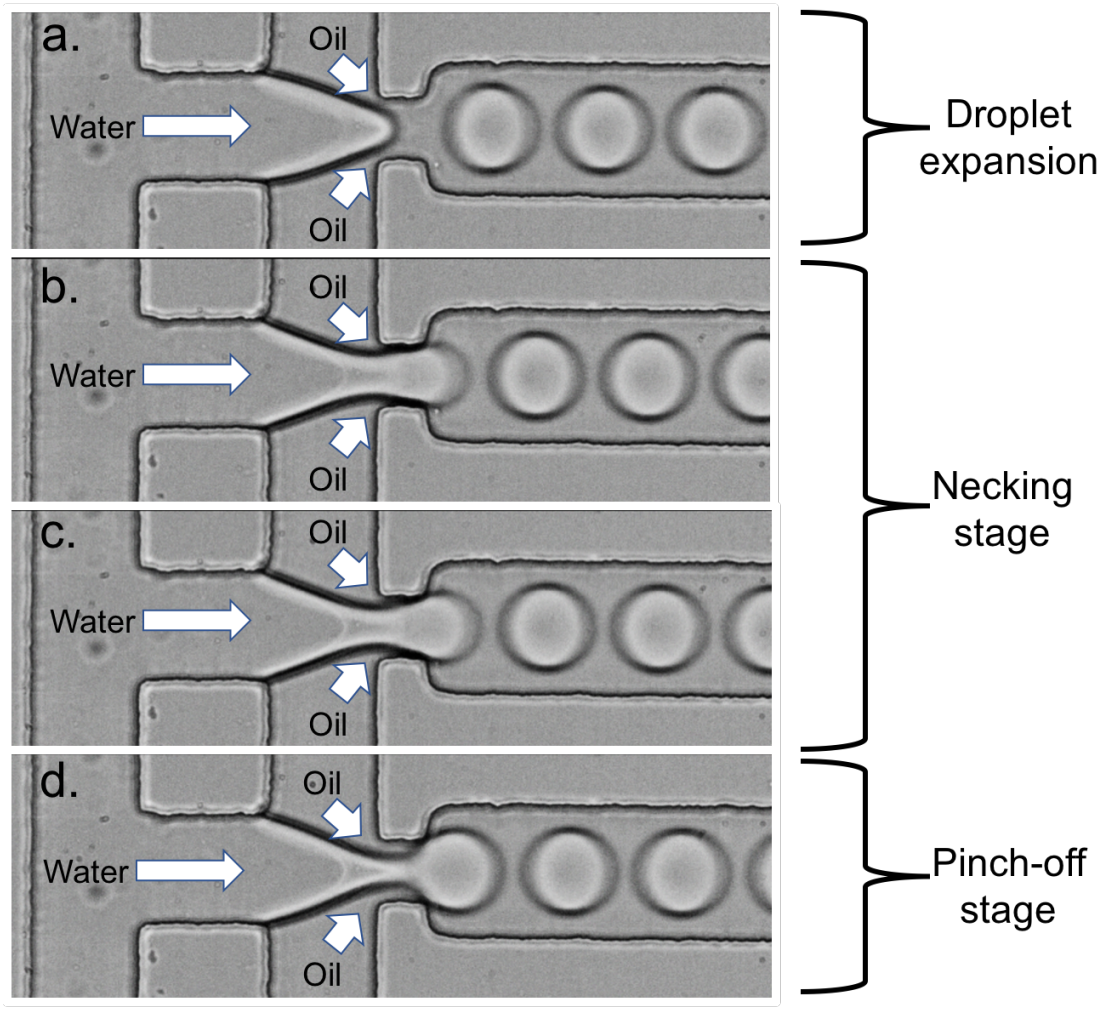


Figure 4.2: Stages of droplet production in a flow-focusing junction (a) The Droplet expansion stage as the dispersed phase is forced past the FFJ orifice. (b & c) The necking stage where the dispersed phase blocks the channel. The resulting pressure build-up causes the continuous phase to deform the neck of the dispersed phase. (c) The pinch-off stage where the droplet becomes separated from the bulk dispersed phase.

The jetting regime is well suited for producing small droplets at high frequencies as unlike dripping, it is not dependent on the width of the dispersed phase channel. One major limitation to jetting however is that it is a much less stable regime and can easily lead to greater droplet polydispersity and also transient episodes of stable co-flow. Stable co-flow is a situation in the device where the two phases remain distinct from each other and pass through the device as laminar flow (fig 4.3.g-i). Episodes of co-flow during a droplet production run lead to the formation of extremely large water droplets that have a series of negative consequences

downstream. The dripping regime, although limited in the size of the droplets it is able to produce, is capable of very stable monodisperse droplet production. For this reason, all droplets for the rest of the project were produced using a dripping regime.

4.2.2.2 Droplet production junctions in microfluidic devices

Droplet production can be achieved in a number of ways in a microfluidic device. The three most commonly used passive fluidic junctions are: flow-focusing junctions, T-shaped junctions and co-flow junctions.

These 3 junctions are illustrated in Figure 4.3 and described below:

- **Flow-focusing junctions** - One method to produce droplets is through the use of flow-focusing junctions. In this method, the dispersed phase is focused through a cross-road shaped junction. As the dispersed phase passes the junction opening, it joins with two oil channels (continuous phase), one from either side, and the stream is thinned. As the two phases are immiscible, they organise into a laminar flow towards the outlet junction of the crossroads. The expanding droplet phase (4.2.a) partially blocks the outlet junction and causes a build-up of pressure in the continuous phase channels. This pressure exerts a squeezing force on the dispersed phase forcing it to thin further leaving a thin bridge that keeps the bulk of the dispersed phase attached to the newly forming droplet (4.2.b,c). If flow-rates are constant, the building pressure will eventually cause the neck to separate and the droplet to continue down the outlet channel driven by the continuous phase.

- **T-shaped junctions** - First published in 2001 [194], T-shaped junctions were among the first geometries to be used for microfluidic droplet formation. Unlike the flow-focusing junctions where droplet generation depends on the inertia of the fluids [195, 196], T-shaped junction droplet production is determined by shear force and interfacial tension [197, 198] where a stream of the dispersed phase is broken by flowing a second immiscible fluid (continuous phase) across its path. This method of droplet production, although able

to generate higher frequencies of droplets, results in a more polydisperse population making this method less suitable for this project [199].

- **Co-flow junctions** - Co-flow junctions are distinct from the previous two geometries as they are not planar devices, meaning they have features on the chip of different depths. These chips flow the dispersed phase through a central inner channel that is nested within a larger outer channel containing the continuous phase. Both flows are parallel to each other and droplet generation is governed by the viscous shear of the continuous phase over the dispersed phase [200]. As the expanding droplet from the central channel grows, the viscous drag from the continuous phase increases until the drag matches the interfacial tension that keeps the droplet attached to the capillary tip. At this point, the newly formed droplet is sheared from the capillary tip and carried down the outer capillary in the continuous phase. Co-flow geometries are considerably more difficult to produce when compared to T-junctions and flow-focusing junctions and require an expensive mask aligner to achieve the best results.

Monodispersity of the droplets is perhaps the most important consideration for this project as droplets of different sizes will adversely affect the concentration of reagents in the dispersed phase. As T-shaped junctions have the poorest performance in terms of droplet monodispersity this geometry was excluded from the project. Both flow-focusing and co-flow geometries have similar performances in terms of monodispersity and frequency of droplets, however due to the ease of manufacture and less capital expenditure, flow-focusing junctions were deemed to be superior. For the rest of this project therefore, all droplets were produced using a dripping regime in a flow-focusing junction.

4.2.3 Water-in-oil-in-water (WOW) droplet formation

The ability to use a genomic target in the cell as a means of enriching for a certain population is an attractive one. Currently, most cell-based enrichment methods focus on FACS where a known panel of cell markers identifying a certain

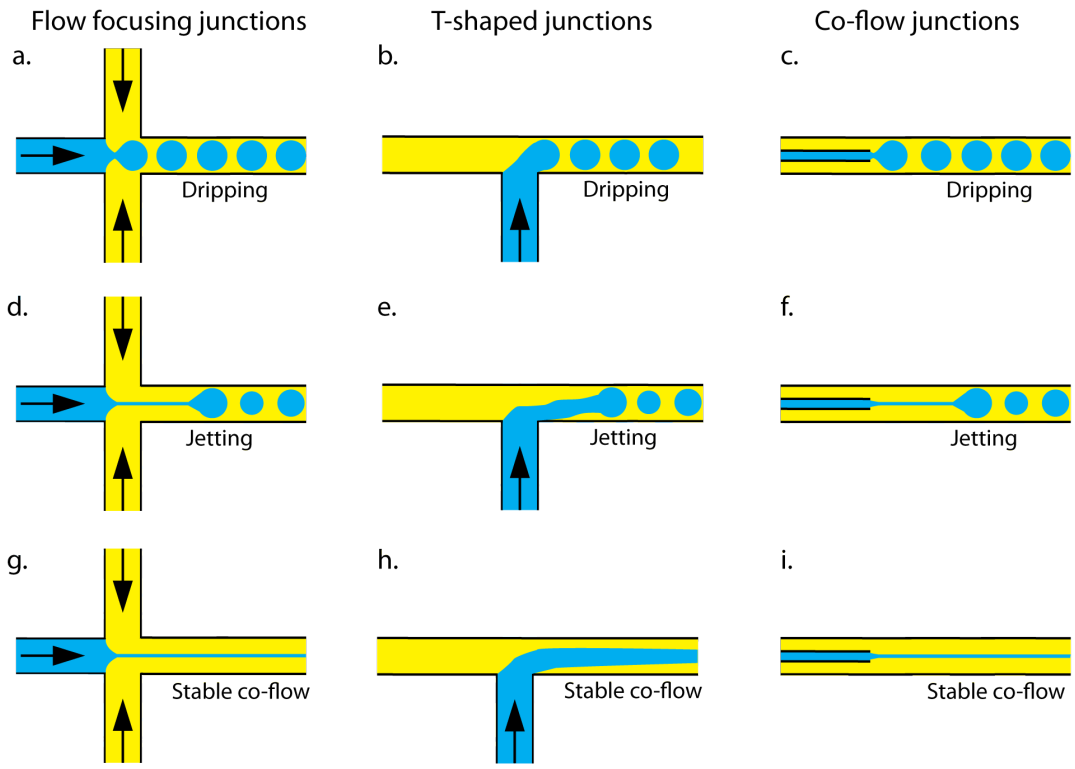


Figure 4.3: Three commonly used droplet generating geometries. (a,d,g) Flow-focusing junction. (b,e,h) T-junction (c,f,i) Co-flow device.

lineage or activation state can be used to isolate different populations of cells. One of the side-goals of this project is to explore the potential to combine the isolation of a cell based on an integrated HIV genome with the phenotypic readout of a technology like FACS. On-microscope droplet microfluidics chip setups are currently custom-built and require a scaffold framework attached to a microscope upon which peripherals are added. Until such a time that a commercial product is available with peripherals explicitly designed to be used with an integrated instrument, adding multiple channels for detection of fluorohpores is problematic. For each fluorophore under investigation, a separate detector (PMT) is required and dependent upon its speed, will likely require its own microcontroller. As these peripherals begin to stack around a standard bench microscope, the limitations of this approach become obvious. One elegant solution to this problem is to utilise an already existing technology for the analysis of the droplets, namely a flow cytometer.

Flow cytometers are in essence reusable microfluidics chips with a fixed channel geometry. Single cells are aspirated through the fluidics path where they pass one or more lasers and any fluorophores associated with that cell emit fluorescence that is captured by a series of detectors. As these fluidics paths are designed to carry cell suspensions, the inner surface of the channel is weakly hydrophilic. Should a WO droplet suspension be passed through this fluidics path, the aqueous core of the WOs would interact with the channel surfaces causing wetting, and as a consequence, coalescence of the emulsion. To overcome this barrier, a number of studies have created a double emulsion of water-in-oil-in-water droplets (WOWs) (Fig 4.4.a).

To produce WOWs, WOs are collected and then reinjected into a second droplet production chip with the phases reversed. In this chip, the reinjected droplets are the dispersed phase and a water/SDS (0.1%) mix is the continuous phase. The droplet reinjection inlet is constricted to ensure it is narrower than the droplet diameter; this constriction prevents the encapsulation of multiple WOs per WOW as only one WO can enter the flow-focusing junction at any one time. Converse to the WO production chips, WOW chips require a hydrophilic channel coating to prevent the oil shell surrounding the WO from interacting with the channel surface leading to wetting. Double emulsions remain stable due to their double interface. One interface between the aqueous core of the WO and the oil shell surrounding it, and another interface between the oil shell and the WOW continuous phase of water and SDS.

By producing double emulsions such as these it is possible to flow droplet populations within a conventional flow cytometer. As a proof of concept, a bulk albumin qPCR reaction was thermocycled (thereby increasing its fluorescent intensity), encapsulated in WOs and subsequently encapsulated within WOWs. To determine whether positive droplets could in fact be identified from negative droplets, a second population of WOs without DNA template were also encapsulated into WOWs. Both populations were then analysed on an LSRII flow cytometer (Becton Dickinson) for: forward scatter, side scatter and fluorescent intensity on the FITC

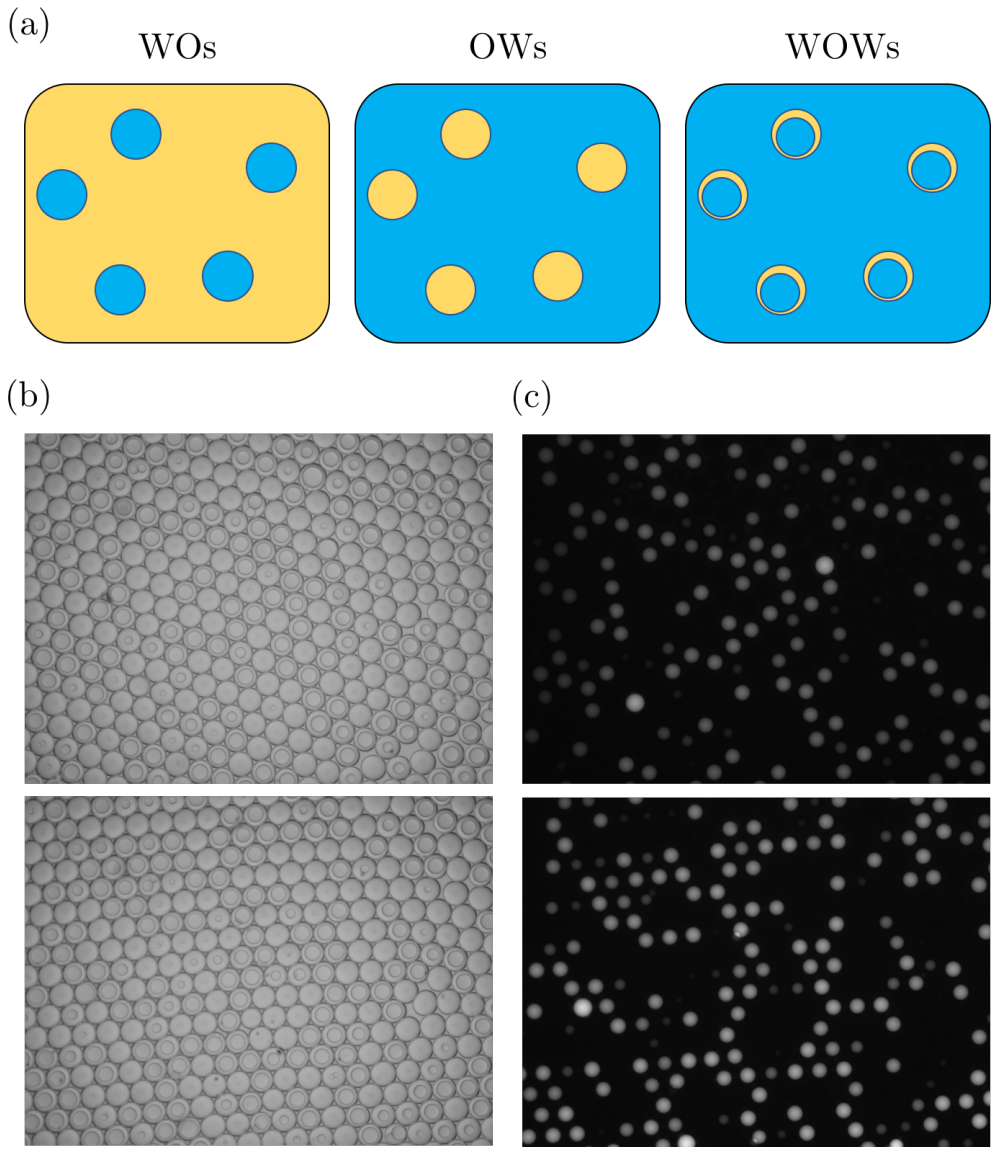


Figure 4.4: WOWs (a-c) Schematics of the three droplet types obtainable using oil and water: water-in-oil (WO), oil-in-water (OW) and water-in-oil-in-water (WOW). (b) Two brightfield images of WOWs. (c) Two fluorescent images of WOWs encapsulating fluorescein labelled WOs.

channel (qPCR probe channel). Any given population of WOWs that have been generated will likely be comprised of at least two distinct droplet types. Firstly, there will be oil-in-water droplets (OWs) that are formed when a volume of carrier oil passes the FFJ of the droplet production without carrying a WO. These droplets are considered single emulsions as they do not carry two interfaces within them. Secondly, there will be WOWs formed when the portion of carrier oil passing through the FFJ contains a WO. For this experiment, these WOWs can be further subdivided into two distinct populations: PCR^- WOWs and PCR^+ WOWs. PCR^- WOWs should appear identical to PCR^+ WOWs on the flow cytometer with the exception that they will lack fluorescence in the FITC channel. This is a result of the absence of DNA template in the starting PCR mixture. These three populations are described pictorially in fig 4.5(a-c).

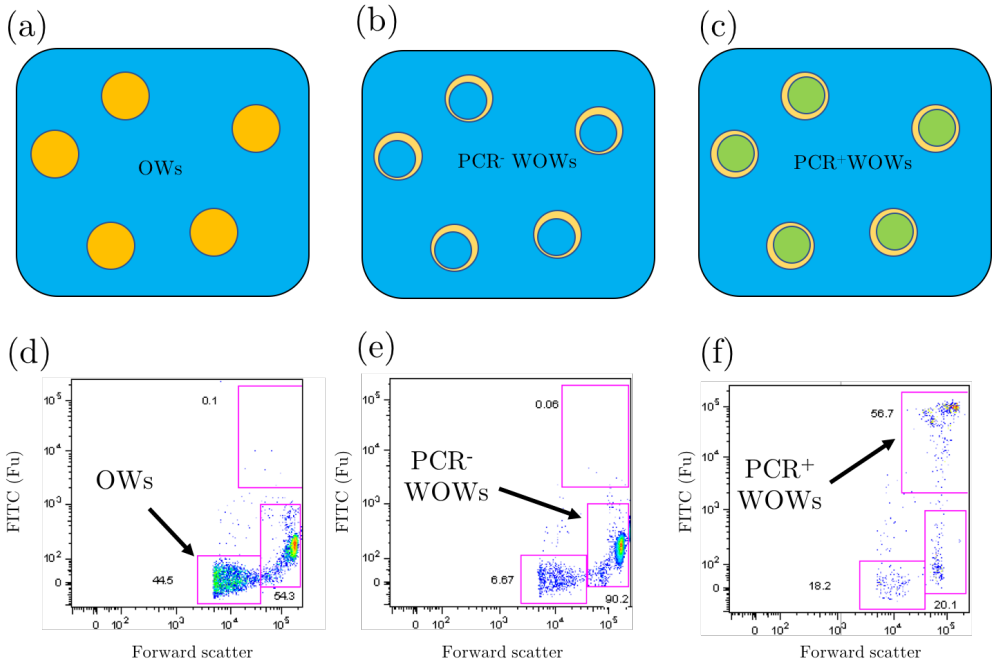


Figure 4.5: WOWs in flow cytometry(a-c) Schematics of the three droplet types obtainable in this experiment: WO, OW and WOW. (d) FACS plot showing the profile of the OW population. (e) FACS plot showing the profile of the PCR^- WOW population. (f) FACS plot showing the profile of the PCR^+ WOW population. Numbers on FACS plots indicate the percentage of cells falling within each gate

The LSRII flow cytometer was able to distinguish the three populations based on just the forward scatter (FSC) of the laser light and the intensity of the FITC

channel as can be observed in fig 4.5(d-f). The OW population although similar in size, produce less forward scatter than the two WOW populations. This is likely due to an increase in scatter as the laser light passes through the double interface present in the WOWs. The OW population as expected is $FITC^{low}$. Both WOW populations are found higher on the FSC axis and their similar positions suggest they are of a similar size and internal complexity. The only difference observed between these two populations is found on the FITC channel as expected. PCR^- WOWs are $FITC^{LOW}$ whilst PCR^+ WOWs are $FITC^{HIGH}$.

The FACS plots in 4.5(d-f) were all produced from populations that should in theory be pure populations. If this were the case then each plot should contain either 100% of the WOWs in the $FITC^{HIGH}$ gate or in the $FITC^{LOW}$ gate. As can be seen in figure 4.5.f, this is not the case in this experiment and $FITC^{LOW}$ WOWs are found in the PCR^+ sample. One possible explanation for this could be the fact that all three WOW populations were produced using the same double emulsion chip with oil flushed through inbetween each run. It is therefore possible that droplets from one condition may have not been flushed through and consequently been collected in the wrong tube. This is supported by the fact that very few $FITC^{HIGH}$ droplets are observed in either the OW or PCR^- WOW populations. As the PCR^+ WOWs were the last to be produced, they are unlikely to have been transferred into the incorrect collection tube.

An alternative explanation could be that due to the small size of the WOs, the fluorophore concentration may not be sufficient to disperse fluorophores throughout every droplet. If this were the case, a small number of PCR^+ WOWs may not produce a FITC signal and fall into the $FITC^{LOW}$ gate.

This experiment suggests it is possible to form double emulsions with PCR labelled droplets and analyse them in a conventional flow cytometer.

The main benefit of using a flow cytometer to analyse these types of droplets is the ease at which it is possible to interrogate multiple colours on a single event passing through the instrument. In this scenario, ideally it would be possible to couple the genomic PCR fluorescence signal with a series of cell markers that have

been assigned to different colours on the spectrum. For this to be successful it would be necessary for the fluorophores to survive thermocycling as the cells would have to be labelled prior to encapsulation and have to undergo the subsequent PCR conditions. To test the ability of common flow cytometry fluorophores to withstand thermocycling, four were chosen to be heated and subsequently encapsulated into WOWs for flow cytometric analysis. The four fluorophores selected were: FITC, PE, PE-Cy7 and APC. The four combined solutions were heated to 95℃ for 5 minutes and encapsulated, firstly into WOs and subsequently into WOWs ready for flow cytometric analysis on the LSRII.

Three populations were distinguishable when forward scatter and side scatter (SSC) were used to assess the WOWs (Fig 4.6.a):

- One FSC^{LOW}, SSC^{LOW} population that was consistent with the OW population seen in the previous experiment. Despite being at a low frequency, SSC^{HIGH} WOWs were present in this population. This may be evidence of some smaller WOs being encapsulated in WOWs resulting in a lower FSC and SSC signal.

- One FSC^{HIGH}, SSC^{LOW} population. It is not clear what this population represents as theoretically all WOWs should be SSC^{HIGH} as they contain the second interface in the droplet. One possible explanation is that this is evidence of flow instability during WOW production that has led to some polydispersity.

- One FSC^{HIGH}, SSC^{HIGH} population. This population is consistent with being WOWs that contain a WO with active fluorescence. Further support for this is added by the fact that a similar percentage of the WOWs are fluorescent in each of the different channels (table 4.3.)

Table 4.3 shows the frequency of fluorescent drops analysed for each channel in this experiment. Similar percentages of FITC, PE, APC and PE-$Cy7^{HIGH}$ droplets are observed, and at a similar frequency to the FSC^{HIGH}, SSC^{HIGH} WOWs in figure

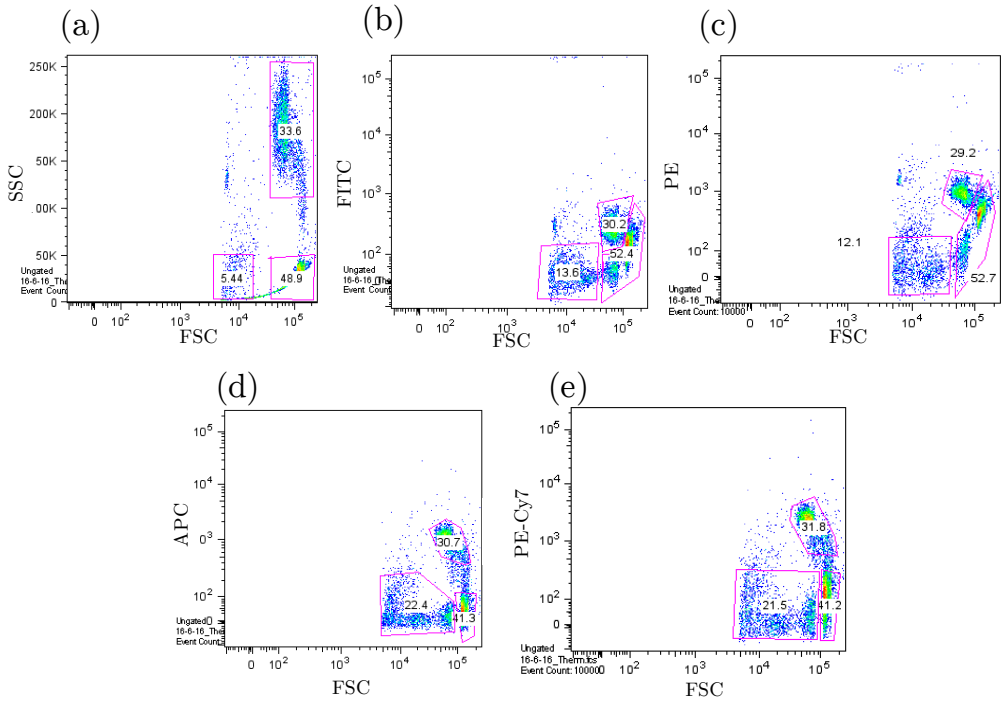


Figure 4.6: FACS fluorophores survive thermocycling Representative FACS plots from the WOW fluorophore experiment showing (a) FSC against SSC, (b) FSC against FITC, (c) FSC against PE, (d) FSC against APC, (e) FSC against PE-Cy7.

Intensity	FITC (%)	PE (%)	APC (%)	PE-Cy7 (%)
High	30.2	29.2	30.7	31.8
Low	52.4	52.7	41.3	41.2

Table 4.3: Population frequencies from the WOW fluorophore experiment

4.6.a. These data suggest that the fluorescent WOWs seen in the independent colour channels are the same population as are present in the FSC^{HIGH}, SSC^{HIGH} gate.

Further work is required in this area of the project to definitively establish which fluorophores can be used as markers to label cell surface proteins however in this proof of principle work, the data supports being able to integrate genome based sorting with multi colour biomarker analysis on a conventional flow cytometer using WOWs. This methodology would allow researchers in a variety of fields to be able to more thoroughly characterise cell subsets where a genomic signal is

required to distinguish a population.

4.2.4 Wetting in microfluidic devices

PDMS microfluidic devices have a number of diverse applications and as such are often required to be compatible with a range of fluids with different properties. Immediately following oxygen plasma treatment, polar functional groups are created on the exposed surface, thus changing the PDMS surface chemistry [189, 190]. This change renders the surface hydrophilic, the strength of this hydrophilicity however, is lost over time and the channels become more hydrophobic reverting to their natural state [201].

Flowing either organic or aqueous solutions successfully in a microfluidic device is dependent on having channels with the correct properties. In order to produce WOs it is essential to have hydrophobic channels that lift the aqueous phase from the walls. Conversely, for OW production the channels must remain hydrophilic. Inappropriate surface modifications will lead to wetting in the device due to interactions with the surface of the channel. Wetting in the context of droplet production is undesirable as the dispersed phase will not form droplets if there is an affinity to the channel surface.

In order to combat wetting, it is necessary to modify the channels to render them either hydrophobic or hydrophilic. A list of treatments and their respective properties can be found in table 4.4. The method of applying these coatings varies depending on the treatment. Hexamethyldisilazane (HMDS) is a volatile liquid, and as such, a small volume can be deposited in an enclosed container with the device overnight where the vapour penetrates the gas permeable PDMS to coat the channels of the device. This method is simple yet robust and has therefore been used predominantly throughout this body of work. Aquapel is a product commercialised for the automotive industry where, due to its hydrophobic properites, it is sold as a windscreen rain repellent. In order to apply Aquapel to a device it is necessary to flow the solution through the channels that are to be made hydrophobic. An additional flush step is required to remove any residual Aquapel

from the channels; failure to complete this step will result in crystalline deposits left as the Aquapel evaporates. Flushing is not always sufficient to remove the Aquapel and crystals that form inside the device reduce the channel dimensions and at least, lead to a modified flow-rate in the chip, if not cause a complete blockage. Silwet L-77 is a convenient solution as it is added directly to the PDMS mixture pre-curing and does not require any further application after baking. There are limitations however, as the glass slide that forms the base of the channels will not be hydrophobic and Silwet-77 channels are less hydrophobic than those treated with Aquapel or HMDS.

As both Aquapel and HMDS provide similar levels of hydrophobicity, HMDS was selected for this project for its ease of application.

Table 4.4: PDMS surface modifications

Treatment	Surface property	Application	Timespan
Aquapel	Hydrophobic	Liquid	Permanent
Hexamethyldisilazane	Hydrophobic	Vapour	Permanent
Silwet-77	Hydrophilic	Liquid	Permanent
Oxygen plasma	Hydrophilic	Vapour	Temporary

4.2.5 Chip geometry, flow-rates and surfactant concentration determine obtainable droplet sizes

As described above, droplets can be made by using a variety of different geometries and flow regimes. Further to this, by making small changes in the device geometry or the flow-rates being used, it is possible to control the size of the droplet being produced by the chip. A series of experiments were performed to establish the optimum geometry and flow-rates required to produce droplets of a suitable size for single-cell encapsulation.

4.2.5.1 FFJ chips produce smaller, more monodisperse droplets compared to T-junction devices

As described earlier, the literature suggests that FFJ chips are capable of forming more monodisperse droplets when compared to T-junction chips. To test this, droplets were produced with both chip types using a series of different flow-rates to determine both the size range achievable and the dispersity of those droplets. Average droplet size and the dispersity index ($\delta = \frac{\sigma}{d_{av}} \times 100\%$, where σ is the standard deviation and d_{av} is the mean droplet diameter) were measured for a range of flow-rate ratios. Flow-rate ratio was calculated using $\frac{Q_w(tot)}{Q_o}$ where $Q_w(tot)$ is the total flow-rate for all aqueous inlets and Q_o is the flow-rate for the oil inlet.

Figure 4.8.a highlights that for any given flow-rate ratio, the FFJ chips produce a smaller droplet on average when compared to T-junction chips. As the flow-rate ratio increases past 0.3 this difference begins to decrease. In addition, as the flow-rate ratio decreases, the droplet size increases from 130 µm at a flow-rate ratio of 0.1 to 184 µm at 0.5. The dispersity of the droplets produced using the FFJ is also much lower than it is for the T-shaped junction (Fig 4.8.a). At the highest flow-rate ratio (0.5), this effect is negligible. This is due to the increased efficiency in shearing caused by the dual continuous phase channels in the FFJ chips versus the single continuous phase in the T-junction chips. Following these tests, all further work was solely performed on the FFJ chips.

For this application to be successful, droplet size has to be tightly controlled during production. It was therefore necessary to determine the relationship between channel geometry and droplet size. Three aspects of the chip design were tested (W_c, L_G and W_R) as detailed below in fig 4.7.

The channel width at the FFJ and the flow-rate ratio are the key determinants for the size of droplets that can be produced by a microfluidic chip. The relationship between channel width and droplet size can be described by $WO_d = FFJ_W \pm 25\%$ where WO_d is the diameter of the water-in-oil droplet and FFJ_W is the width of the channel at the FFJ. A series of flow-rate ratios were used in the same chip to establish the range of droplet sizes that could be produced for a particular

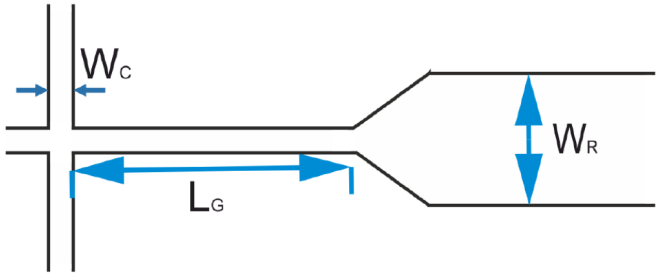


Figure 4.7: Chip geometry parameters Aerial schematic showing the parameters tested in relation to droplet size. W_c is the width of the channel at the FFJ, L_G is the length of the gap between the FFJ and the start of the reservoir and W_R is the width of the reservoir at its widest point.

geometry. Figure 4.8.c shows the range of sizes a 30 µm C_W chip is able to produce. For this chip, the $Q_w(tot)$ was 1 µl/minute. This chip was capable of producing droplets between ~14 µm and ~64 µm for Q_O rates between 8 µl/min and 2 µl/min respectively. The droplet stream became unstable with a Q_O of 10 indicated by the breadth of droplet sizes present for the 10 µl/min condition in figure 4.8.

A similar experiment was performed with a C_W = 100 µm chip. In order to fully explore the range of droplets achievable using this chip, an extended range of Q_O was used. A sharp decrease in droplet size for every incremental Q_O increase was observed for Q_O values $\leq$3 (fig 4.8.d). For example, a Q_O increase of 1 µl/min led to a 150 µm reduction in droplet diameter. In comparison for Q_O values >3, the same 1 µl increase in Q_O caused only a 7.7 µm decrease in droplet size. This biphasic relationship between C_W and Q_O means that it is increasingly difficult to make droplets that are $<C_W$. In addition, the higher Q_O required to make these smaller droplets is far less cost-efficient than simply using a chip where the C_W more closely matches the target droplet diameter. In light of this data, all future chips were designed with a C_W similar to the desired droplet diameter.

To establish the effect of the post-FFJ channel geometry on droplet size, a series of devices were designed with different L_G and W_R distances. As predicted, for chips where C_W = 100 µm, the mean droplet size was dependent on the geometry. An increase in $\frac{L_G}{W_R}$ results in smaller droplets being produced (fig 4.8.e). In this scenario, as channel length increases, an increase in hydraulic resistance is observed

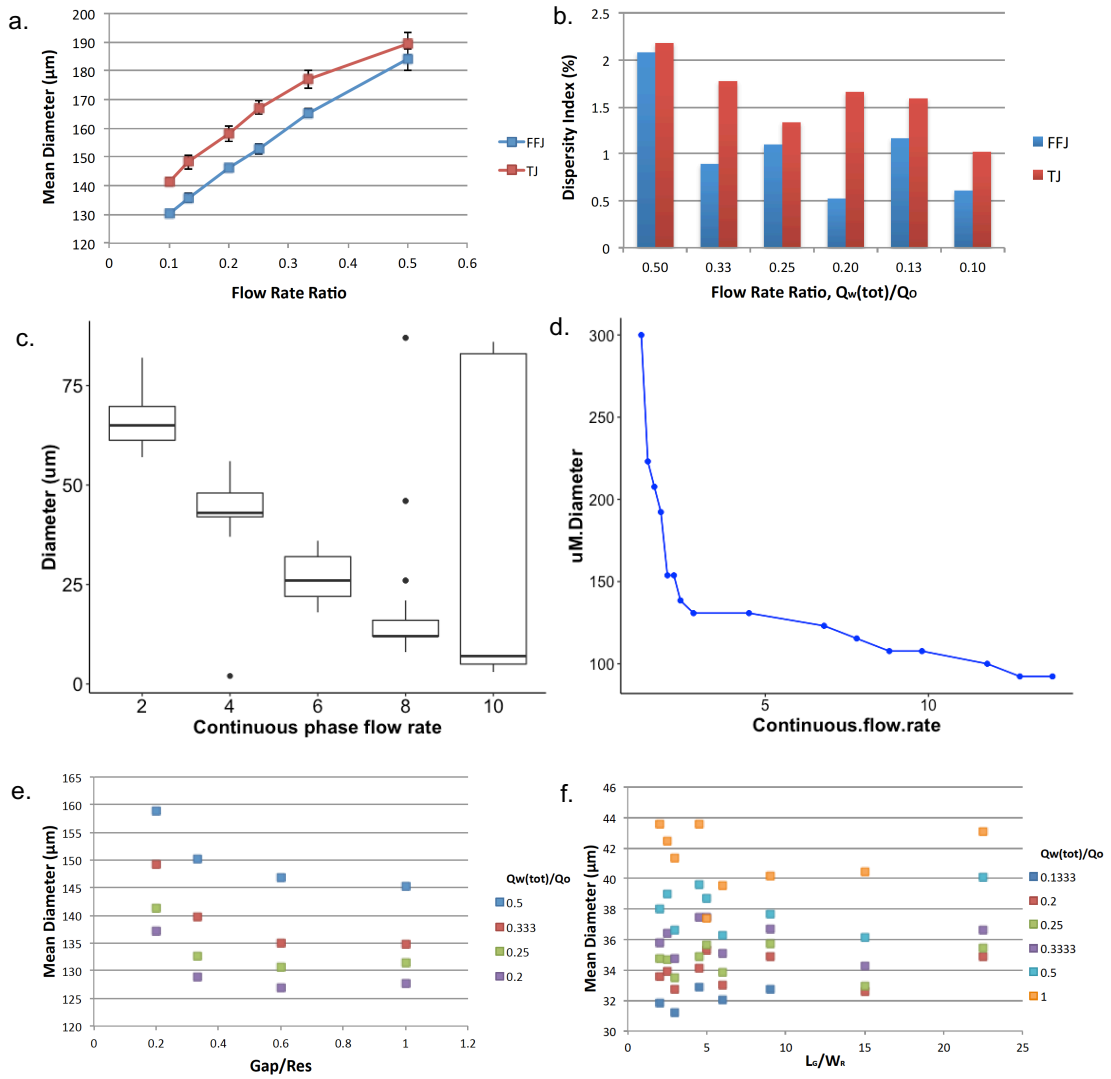


Figure 4.8: Flow-rate and geometry optimisation (a) (b,e,h) T-junction (c,f,i) Co-flow device.

and the droplet size before detachment is reduced. Interestingly, the same effect was not observed for chips where $C_W = 30\,\mu m$ (fig 4.8.f). In this chip, average droplet size was independent of the $\frac{L_G}{W_R}$ ratio and consistent droplet sizes were produced for all chips. This result may be due to the increased hydraulic resistance that is present in the narrower channels. It is possible that this resistance has reached a critical threshold where droplet formation occurs close to the FFJ and the L_G distance becomes obsolete.

4.2.6 Continuous phase and fluorosurfactant selection

The interface with the oil continuous phase is a thermodynamically unfavourable environment for an aqueous molecule within a WO as it loses a number of attractive interactions with its neighbouring molecules. It is this relationship that enforces the spherical shape of a WO as it produces the smallest surface area to volume ratio and reduces the thermodynamic free energy of the system. If we were to consider the dropletisation of a 25µL PCR reaction, the surface area of one 25 µl droplet will be orders of magnitude lower than the same volume in an emulsion of $1x10^6$ 25 nl droplets. It is this increase in the energy of the system that makes it thermodynamically unstable and promotes the separation of the original dispersed and continuous phase. Furthermore, dependent upon the phases used it is possible for molecules to pass freely between the dispersed and continuous phase resulting in cross-contamination between droplets. A fundamental requirement of this project is to produce droplets that are stable over time and impermissible to biological agents, therefore options for increasing droplet stability and biological permeability were investigated.

4.2.6.1 Fluorosurfactant selection

This project requires extensive thermocycling to allow for in-drop PCR to be performed. Unstabilised WOs left at any temperature for a period of time will begin to coalesce, hence the addition of a surfactant is required to stabilise the droplet. In the system proposed, fluorosurfactants are required to fulfil two separate roles [202]:

- Provide droplet stability and keep coalescence to a minimum
- Provide a biologically inert barrier at the interface between the dispersed and continuous phases

In a seminal study, Holtze et al [203] developed a novel fluorosurfactant to meet these criteria (E2K0660). A number of research groups have since adopted this fluorosurfactant for their own studies. EA-008 (an alternate name for E2K0660) is a triblock copolymer surfactant with twin PFPE tails that are fluorocarbon

oil-soluble. The length of the tails provides steric stabilisation of the emulsion and the inclusion of two tails increases the area the surfactant can cover on the fluorohydrocarbon side of the interface. The PEG component of EA-008 is water-soluble and aids in preventing the transfer of biologicals across the interface [204]. Increasing the size of the PFPE components aids droplet stability but comes as at a cost. The larger the molecule, the slower it will diffuse to the interface. When a new droplet buds off from the FFJ, the surfactant has to migrate to the interface in a scale of milliseconds in order to prevent coalescence and cross-contamination. The authors empirically derived an optimum PFPE block size of 6000 g mol^{-1}. Increasing the size of the PEG component has a similar trade-off, having to be large enough to anchor the molecule at the interface, but not so large as to affect the geometry of the droplets which can decrease emulsion stability. Using these parameters, a PEG block size of 600 g mol^{-1} was derived [203].

For a surfactant to be suitable for this pipeline it must be able to produce thermostable droplets. To assess thermostability, a concentration range of EA-008 was used to form droplets of different sizes. Subsequent emulsions were thermocycled and measured to assess thermostability. Figure 4.9 describes the relationship between: surfactant concentration, droplet size (determined by $\frac{Q_{W(TOT)}}{Q_O}$) and the change in either average droplet diameter (fig 4.9.a) or droplet diameter standard deviation(fig 4.9.b)

Increased droplet thermostability was observed at higher concentrations of fluorosurfactant. This is evident from the reduction in droplet diameter change observed with increasing concentrations of surfactant in fig 4.9.a. These data also suggest that droplet sizes produced with a $\frac{Q_{W(TOT)}}{Q_O}$ ratio >0.2 are in general less thermostable. For droplets produced with a concentration of fluorosurfactant >2% and a $\frac{Q_{W(tot)}}{Q_O}$ ratio <0.2 the observed diameter change post-thermocycling is near zero suggesting little to no coalescence was induced within these parameters. Confirming this, a similar relationship was observed between: surfactant concentration, $\frac{Q_{W(tot)}}{Q_O}$ ratio and the change in standard deviation. Here, a marginal increase in

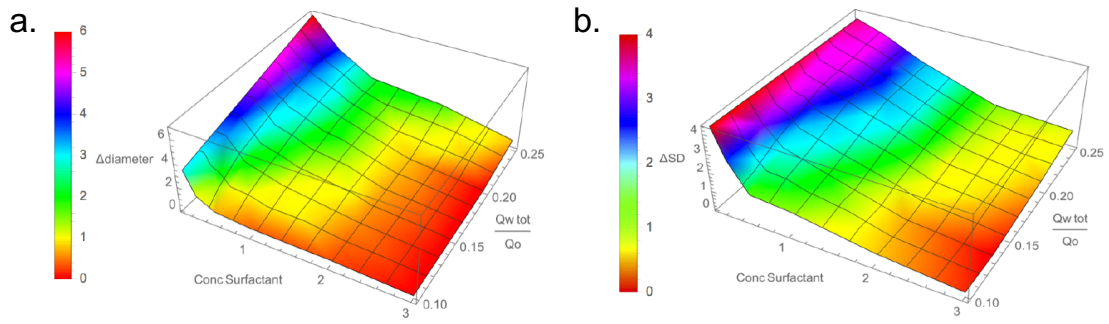


Figure 4.9: Wo thermostability is influenced by surfactant concentration and droplet size. (a) 3-D plot showing the relationship between surfactant concentration, droplet size ($\frac{Q_W(TOT)}{Q_O}$) and average change in droplet diameter post-thermocycling. (b) 3-D plot showing the relationship between surfactant concentration, droplet size ($\frac{Q_W(TOT)}{Q_O}$) and change in droplet diameter standard deviation post-thermocycling.

droplet diameter standard deviation was observed for conditions with a surfactant concentration of 2% and a $\frac{Q_W(tot)}{Q_O}$ ratio <0.15.

Based on the data shown here, a working fluorosurfactant concentration of $\geq$ 3% was proposed.

4.2.7 Single-cell encapsulation in droplets

One of the core aims of this project is to develop medium-high throughput assays that provide single-cell resolution sequencing data. In order to achieve this, it becomes necessary to encapsulate, with some degree of confidence, a single cell into an individual WO. Cell encapsulation in WOs, where there is no requirement to retain single-cell resolution is a straight forward process. A cell culture compatible media can be used as the dispersed phase and cells can be directly loaded into this and encapsulated using a standard FFJ junction with little optimisation (Fig 4.10). This will result in an emulsion that has droplets with varying levels of cell occupancy.

The distribution of cells within the dispersed phase is random, therefore, it is not possible to control the number of cells that will end up in one particular droplet. The average occupancy rate per droplet however, can be controlled and measured. In this system, it is possible to use the poisson distribution to express

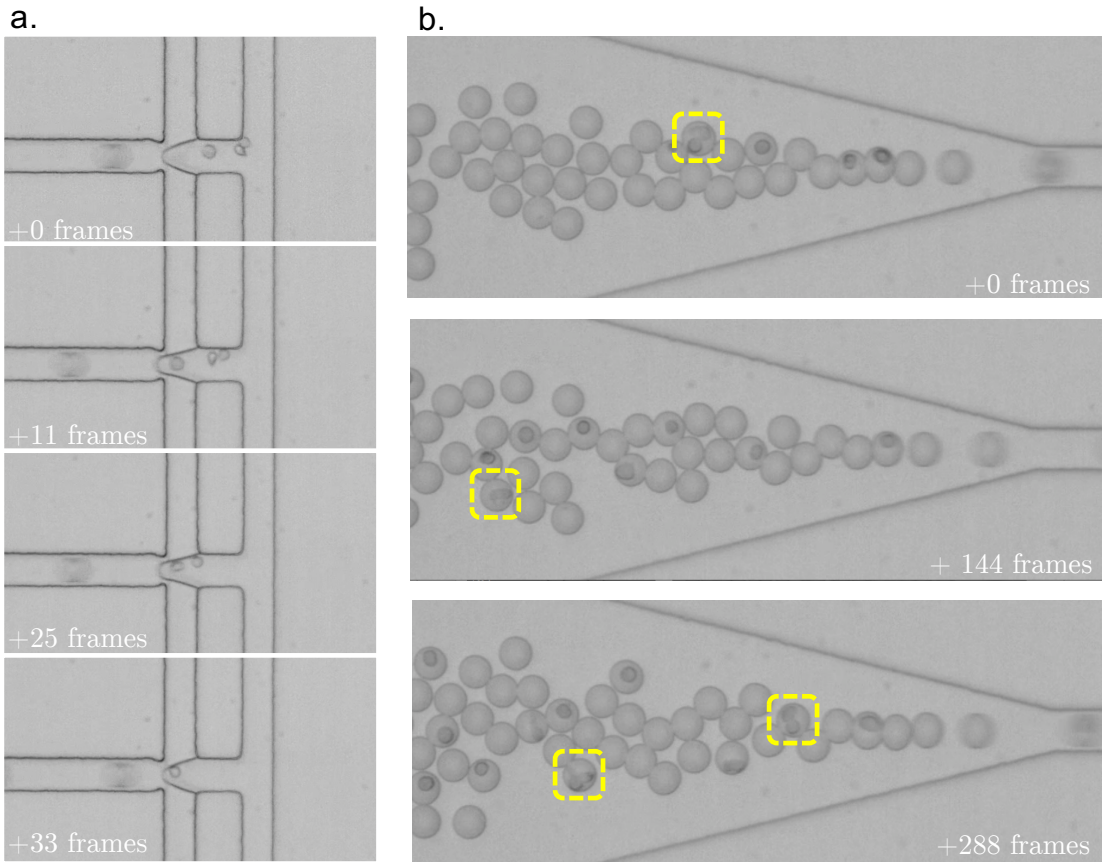


Figure 4.10: Cell encapsulation in a microfluidic device (a) Four still images taken from a GXR cell encapsulation video focusing on the FFJ (White text indicates the number of elapsed frames in the video). (b) Three still images taken from a GXR cell encapsulation video focusing on the reservoir (Dashed yellow boxes indicate droplets with multiple cell occupancy).

the probability of observing x cells in a fixed-size droplet for an expected frequency of events. The probability of having x cells per droplet can be calculated using the following equation where λ is the mean number of cells expected per droplet:

$$P(x, \lambda) = \frac{e^{-\lambda} \times \lambda^x}{x!}$$

Any change in the concentration of cells in the aqueous phase (ϕ_s) or the volume fraction of one cell in a droplet (ϕ_d) will modify λ; this relationship can be expressed as:

$$\lambda = \frac{\phi_s}{\phi_d}$$

To identify a suitable cell density to be loaded into the chip, experiments were performed using ArC beads (ThermoFisher) as cell surrogates. To assess the effect of bead sedimentation during the encapsulation process, two solutions were tested as shown in table 4.5

Table 4.5: Encapsulation solutions

	Bead conc. (Beads/mL)	Density matched	Density matching medium
Solution 1	$4x10^6$	No	N/A
Solution 2	$2x10^6$	Yes	Lymphoprep

Post-encapsulation, two images were taken per field of view. One image focused on the droplets and the other focused on the beads. Two images were necessary as over time, beads will sediment to the bottom of the droplet. Droplets and beads were analysed from their respective images and Mathematica was used to estimate the distribution parameter for a Poisson distribution.

In the solution with the higher concentration of beads in the dispersed phase (ϕ_s), an increase in the probability of encapsulation was observed. Despite having a higher chance of single-cell encapsulation (0.25 compared to 0.1), the higher cell concentration sample also gave a higher chance of encapsulating >1 cell at 0.12 compared with 0.01(fig 4.11). Single-cell resolution, in this project, is more important than encapsulation efficiency therefore a concentration of $4x10^6$ beads/mL was used as a starting point for future experiments. One disadvantage associated with this finding is a loss of throughput. Droplet formation is typically performed at a rate of 1kHz; if we were to assume that every droplet contains a single cell, this system is capable of encapsulating $3.6x10^6$ cells/h. With our new encapsulation efficiency of 0.1, it is only possible to encapsulate $3.6x10^5$ cells/h. In practical terms, this means an additional 12 hours to sort $5x10^6$ cells.

Extended periods of sorting lead to a further complication where beads will begin to sediment over time in a solution that is not being agitated. This sedimentation is caused by the density gradient between the media and the beads. In

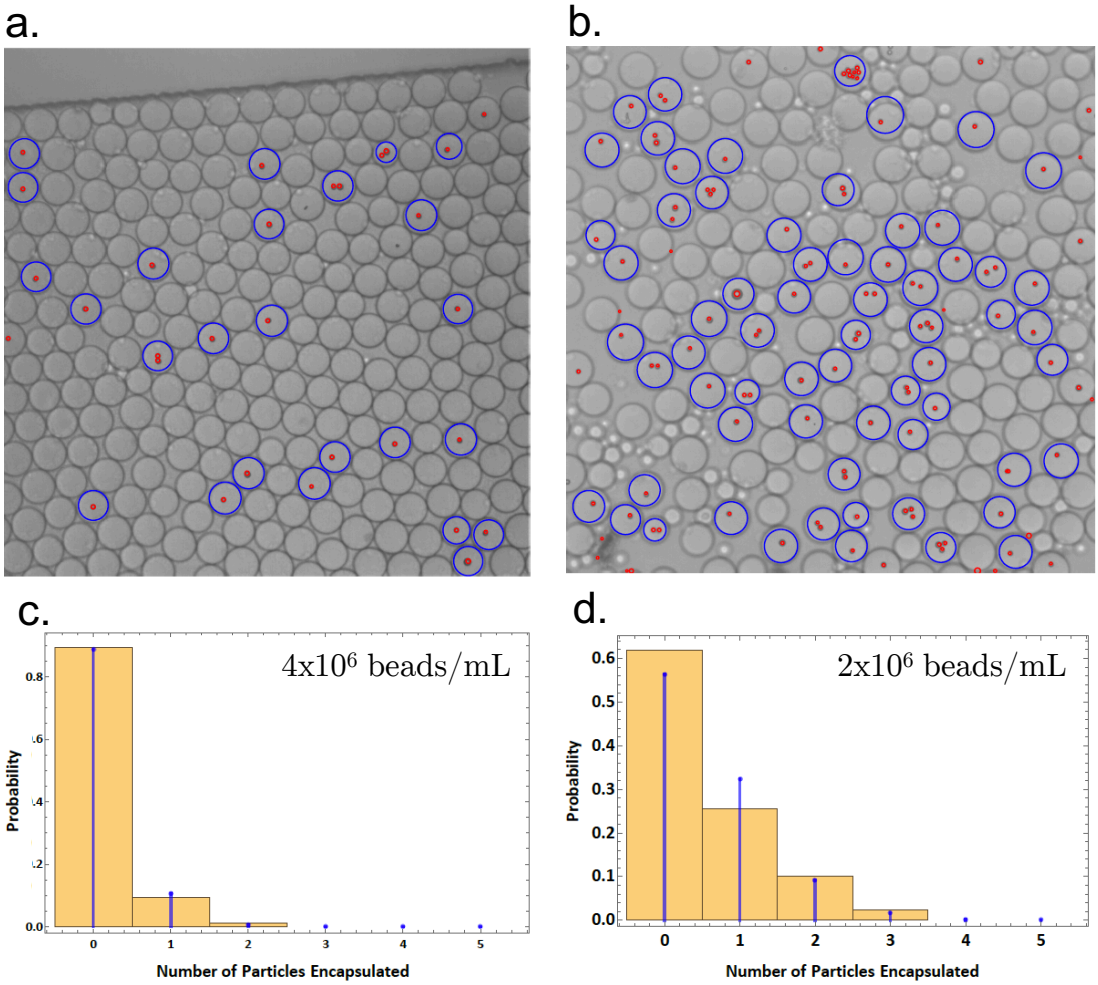


Figure 4.11: Bead distribution from encapsulation experiment (a) Example image showing particle analysis for solution 1 performed in Mathematica (Blue circles outline droplets containing a bead, red circles indicate beads)(b) Particle analysis for solution 2. (c,d) Histogram showing the distribution of encapsulated beads for solutions 1 and 2 respectively. Blue bars indicate the Mathematica Poisson distribution fit.

order to counter this, a density matching medium (Lymphoprep, commonly used for isolating PBMCs) was added to the solution. Figure 4.12 shows the change in bead concentration over time for solutions 1 and 2 (table 4.5). The concentration of solution 2 remains largely stable as the beads are density matched to the medium and hence remain homogenous throughout the sample. In solution 1, a steady increase is observed over time. This increase is caused by the sedimentation of the beads in the tubing. It is only when the immiscible oil phase passes through the tubing that the sedimented beads are then forced towards the chip (fig 4.12.b).

This data suggested that to accurately encapsulate single cells in WOs, it would be necessary to density match the medium to the cell. Furthermore, to limit the occurence of WOs with multiple cell occupancy, cell concentrations must be dilute, as to ensure only a single cell passes the FFJ at any one time.

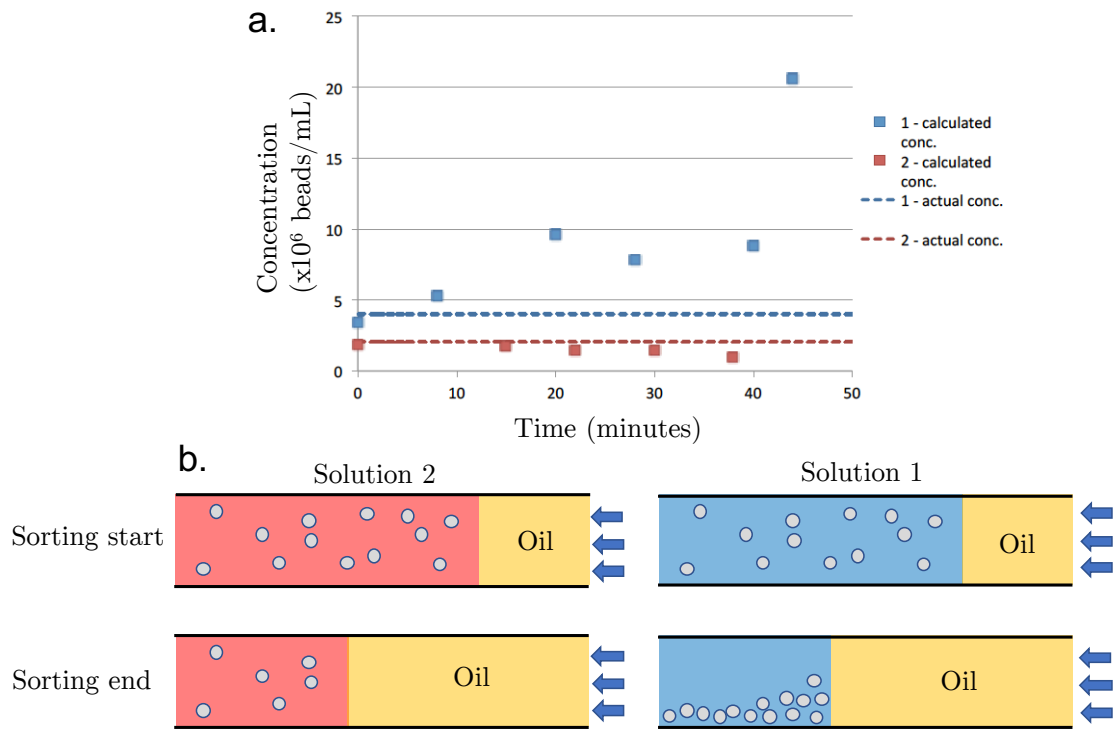


Figure 4.12: Particle sedimentation during sorting (a) Dot plot showing the concentration over time of solutions 1 (blue) and 2 (red). Dashed lines indicate the starting concentration. (b) Schematic drawing of sedimentation in the tubing.

4.2.8 Diluting lysis reagents within droplets

Performing single-cell PCR within the confines of a droplet requires two incompatible steps: complete cell lysis and PCR. Cell lysis can be performed using a variety of methods:

- Chemical lysis
- Mechanical lysis
- Electroporation lysis
- Thermal lysis
- Laser lysis
- Sonication lysis

Chemical lysis is the method that is most compatible with emulsion stability [205] and will form the basis of this section (lysing a cell within the confines of a

droplet via other means is discussed in more depth in chapter 5.3). Despite being compatible with emulsion stability, most agents capable of complete chemical cell lysis are incompatible with PCR as the polymerase enzyme is degraded by the lysis buffer [206, 207]. To negate this problem, protocols commonly have a series of wash steps to remove the lysis buffer before beginning the PCR. As it is a closed environment, wash steps are not possible inside a droplet. It is however, possible to dilute the contents of the droplet through either droplet merging or by injecting further reagents into the droplet.

4.2.8.1 Droplet merging

Droplet merging is the process by which two compatible droplets (in this case two WOs), are destabilised whilst in close proximity to each other, therefore forcing them to combine into one larger droplet. This ability to merge is critical in allowing droplets to be useful microreactors as a number of biological assays require the sequential addition of reagents. In order to merge two droplets, two physical conditions must be met:

- The droplets must be in contact with each other
- The droplets must overcome the interfacial tension stabilising them

In order to satisfy the first criteria (droplet contact), the most common method is to use a change in channel geometry to either slow or speed up the droplets. An example of slowing can be seen in figure 4.13. As droplet A passes the point where the channel widens, it slows as it no longer fills the channel and the continuous phase is able to drain around it. The draining of the continuous phase removes the thin film of oil separating the two droplets and as droplet B's speed is as yet unchanged, it catches and collides with droplet A. Using this method of widening and narrowing the channel widths of the chip, it is possible to perform multiple collisions within the same device without modifying any flow-rates [208].

For two droplets that are not stabilised by a surfactant, proximity is often adequate to overcome the interfacial tension and initiate merging. This process

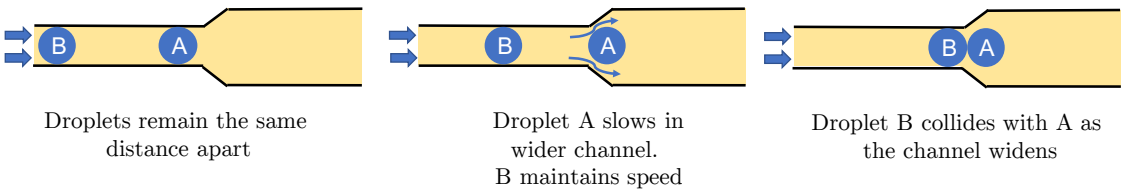


Figure 4.13: Schematic drawing of controlling droplet spacing through channel geometry

does not happen as one might logically expect, as a result of the droplets coming together, but of them moving apart again [209] (fig 4.14). Bremond et al showed that when two droplets were separating, two nipples form on the two adjacent interfaces caused by a drop in interstitial pressure. They hypothesise that this pair of nipples brings the two droplets close enough together to merge. Furthermore, they discovered that when droplets were stabilised by surfactants (in this case SPAN-80), they would not merge unless in the separation phase. This is likely due to the thinning of the surfactant as a result of the increased droplet surface area caused by the nipple. As the lead droplet then enters the narrower channel and elongates, its surface area increases during which it takes time for more surfactant molecules to migrate and saturate the interface. If another droplet is in contact in these few microseconds, the two are able to merge. This type of merging is described as passive and is very difficult to achieve for surfactant-stabilised droplets. In addition, when used with non-stabilised droplets, some uncontrolled coalescence will often occur outside the separation phase. As a result, a more common approach to merging droplets is to use active merging.

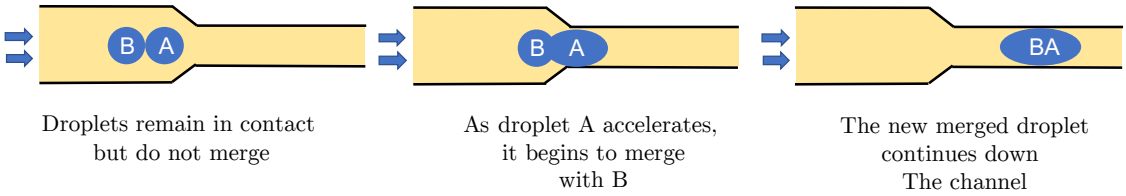


Figure 4.14: Schematic drawing of merging during droplet separation

For droplets that are stabilised by a surfactant, overcoming the interfacial tension is more difficult. As discussed earlier, the addition of a surfactant reduces the interfacial tension and makes the droplets more deformable. Increased

deformability means droplets are less likely to merge with one another, even when in close proximity. Active merging involves the use of an external force to stress the droplet, partially destabilising the interface and rendering it more amenable to merging. Electrocoalesence has most commonly been used for droplet stressing and involves passing the droplets through an electrical field. The presence of the field leads to the onset of attractive electrostatic forces between the droplets, which aid in merging. The closer together the droplets are in the field, the stronger the attractive force [210].

The droplets used in this system are stabilised by two surfactants to increase their thermostability, namely EA-008 and Igepal-CA630. The presence of two surfactants makes these droplets extremely deformable and therefore very difficult to merge. Droplets tested in this chapter were stabilised by both EA-008 and Igepal-CA630; initial designs for microfluidic merging chips based on the Bremond et al study (fig 4.15, devices 1 & 2) resulted in a poor merging success rate.

From left to right in device 1 from figure 4.15 there are the following elements:

- A droplet inlet from which large droplets are reinjected.
- A spacer-oil inlet for the purpose of spacing out the droplets in the channel
- A second droplet inlet from which smaller droplets are reinjected
- A series of channel expansions and contractions causing a change in droplet velocity whilst passing an electrode
- A droplet outlet

The design of device 1 allows large droplets to first be reinjected into the device where they are restricted by a tapering channel geometry into single file. As this train of single file droplets passes the spacer-oil inlet, a buffer of oil is introduced to increase the distance between each large droplet. The now spaced out droplets continue along the channel where they pass the second droplet inlet. Here, the smaller droplet is introduced to the channel. The rate of droplet reinjection at this site is controlled by flow-rate to ensure that no more than one small droplet

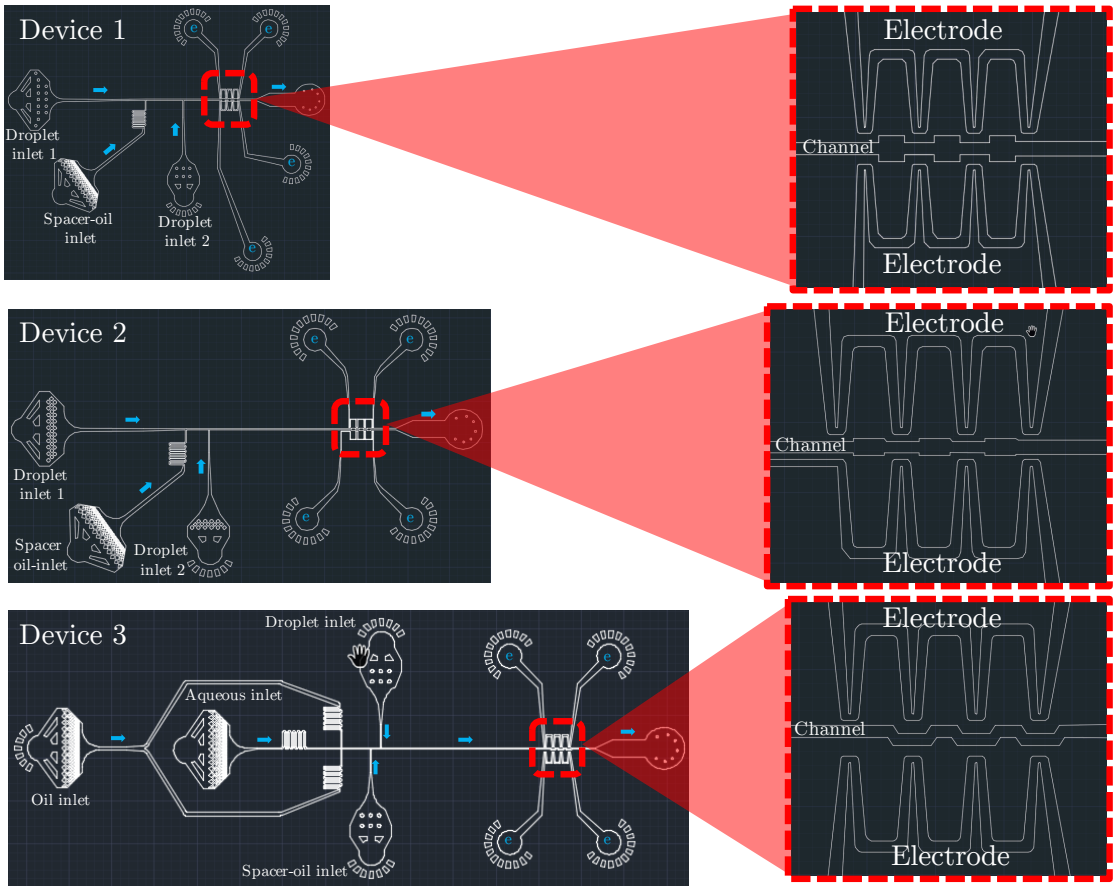


Figure 4.15: Soft lithography masks of merging devices. Copies of the soft lithography masks for merging devices 1, 2 and 3. Small e in blue = electrode. Areas surrounded in red dashed line are magnified on the right hand side of the figure.

is introduced for every large droplet that passes the inlet. Smaller droplets travel faster, therefore, given a long enough channel they catch up with the larger droplet and will be in close proximity as they enter the merging area (this sequence is pictorially represented in figure 4.16 and actual images are shown in figure 4.17).

The merging area for device 1 (shown magnified in the top right image of fig 4.15) consists of three compartments where droplet velocity is modified. A droplet pair consisting of one large, and one small droplet, enter this area in close proximity. An electric field is applied across the channel during the merging process and as per Bremond et al's findings, droplets should merge as they begin to separate. On testing this device, it was discovered that not all droplets were able to merge despite high voltages used for the electric field. It was noted that as the droplet pairs enter the chamber, as expected, the large droplets slow, however the small

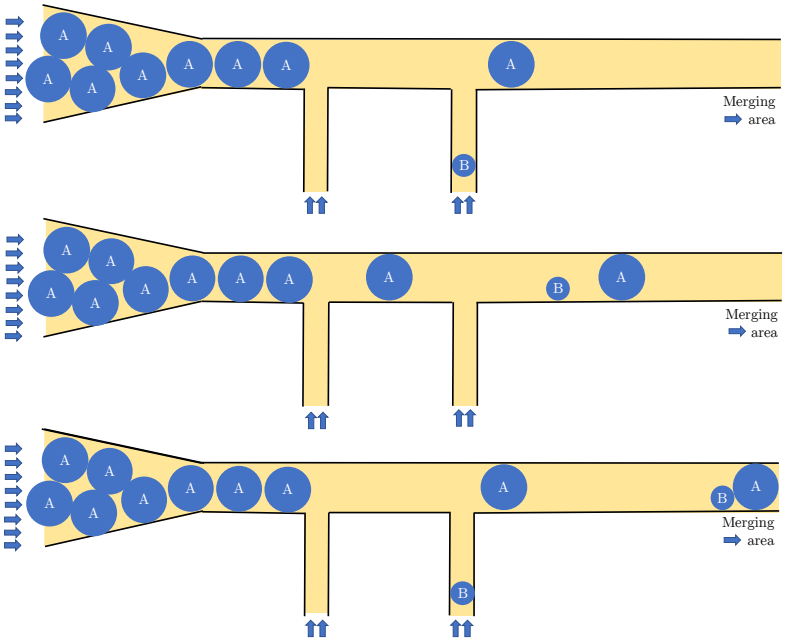


Figure 4.16: Schematic showing droplet re-injection and droplet catch-up. Droplets are shown being: confined to single file, being spaced out and smaller droplets catching up to the larger droplets.

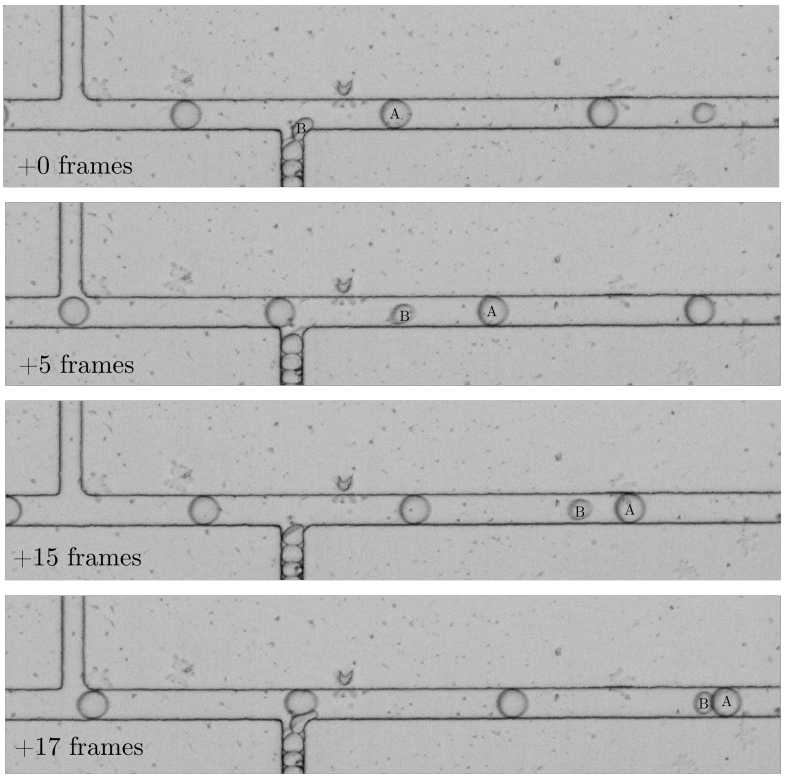


Figure 4.17: Brightfield images of droplet re-injection and droplet catch-up. Four images showing droplet catch-up in microfluidic devices. One drople pair (A&B) has been labelled to help track them through the images. Timestamp is measured in frames in bottom left.

droplets tended to try and travel past the large droplet in the draining continuous phase. Being carried with the continuous phase, the small droplet travels towards the sides of the channel and when the large droplet reaches the narrower exit to the chamber, the two droplets are often no longer in contact and fail to merge. When the droplets did merge, it was consistently when the smaller droplet was about to enter the channel restriction and at a ~45° angle to the large droplet. In keeping with the Bremond study, there was also little evidence of droplets merging during the initial collision as the large droplet enters the chamber and slows. It appeared that the failure to merge may have been due to a lack of confinement of the small droplet, a new device was therefore designed which sought to limit the movement allowed to the small droplet. In addition to the problems in the merging area of device 1, it was also noted that for certain droplet pairs, the small droplet would not catch the large droplet by the time they reached the merging area. This was also addressed in the design of device 2.

Device 2 contains many of the same features as device 1. There are two key differences between them. The first is in the length of the channel following droplet inlet 2. During the testing of device 1, it was observed that some smaller droplets did not have sufficient space to catch the large droplet infront of it. The ability of a smaller droplet to catch up was governed by the timing of its release into the channel in relation to when the large droplet had passed. The closer to the large droplet that the small droplet enters the channel, the greater the chance of the catch being made. As there is no penalty in doing so, the channel length between droplet inlet 2 and the merging area was increased from 1.5 mm to 4.3 mm for device 2.

The second difference in design 2 was in the geometry of the merging area. Small droplets were able to move with freedom around the large droplet in device 1 resulting in them often not being in contact when the droplets began to separate from each other. To counter this, the merging area width was reduced from 95 mm to 60 mm. This extra level of restriction placed on the droplet pair was designed to keep them in closer proximity as they passed through the merging chambers.

Small increases in merging efficiency were observed for device 2 and smaller droplets were no longer able to pass large droplets in the channel. Overall efficiency was still low however and merging was not reproducible.

Despite trying numerous different geometries of which two are shown above, it seemed unlikely to be able to merge the droplets stabilised with two surfactants. One possible solution to this problem was to attempt to merge the droplets before they were able to fully stabilise. Upon formation at a FFJ, droplets undergo a rearrangement phase as the surfactants contained within the two phases migrate and arrange themselves at the interface. The length of this phase is governed by both the viscosity of the phases and the size of the surfactant molecules used. Independent of these variables, the time scale is always in the scale of microseconds and therefore by the time they have left the device outlet, they are fully stabilised. Merging before they are stabilised therefore, requires droplet formation and merging to occur on the same chip. This is not a straightforward proposition as the flow-rates used will have to be compatible with both processes. Device 3 in figure 4.15 shows the chip designed for this purpose.

From left to right in device 3 there are the following elements:

- An oil inlet from which the continuous phase enters the chip.
- An aqueous inlet from which the dispersed phase enters the chip.
- A FFJ where WOs are formed
- A spacer-oil inlet for the purpose of spacing out the WOs in the channel
- A second droplet inlet from which smaller droplets are reinjected
- A serpentine merging section
- A droplet outlet

In this device, large droplets are formed at the FFJ, then spaced out using oil before smaller droplets are added to the channel to pair with them. Droplet pairs then flow into the merging area, here, the serpentine channel forces a 45°

angle between them facilitating merging (fig 4.18). Using this chip, the droplet pair arrives at the merging area before the surfactants in the large droplet have settled. This results in an interfacial tension conducive with merging.

From this point forward, pipelines requiring merging utilised the design labelled as device 3 in 4.15.

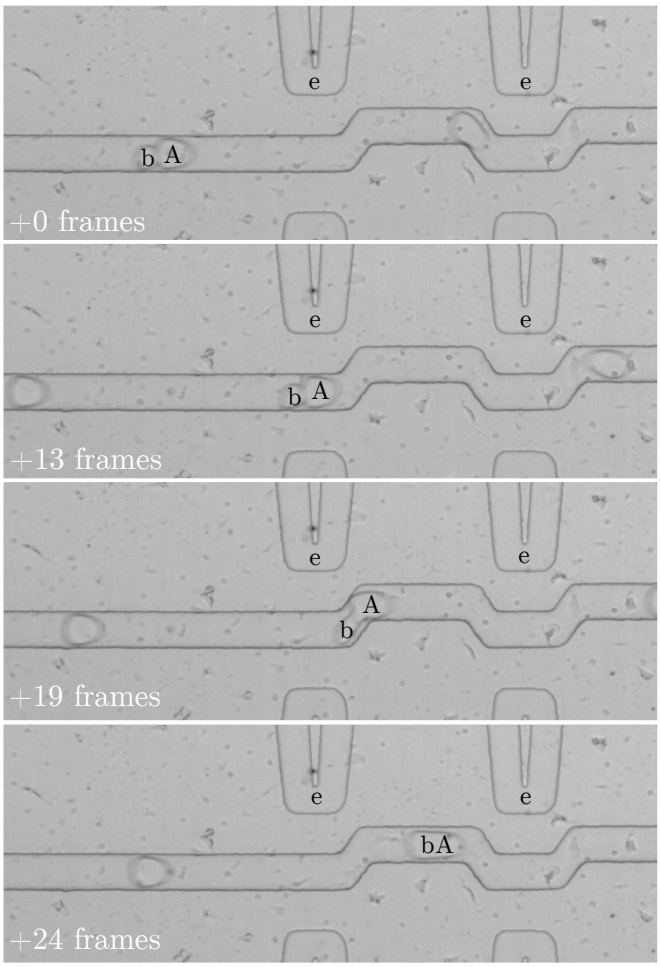


Figure 4.18: Still images showing droplet merging in serpentine channels Droplet pairs (A & b) are forced to 45° angle and merged under an electric field. e = electrode

4.2.8.2 Reagent picoinjection

The alternative to droplet merging as a means of adding new reagents or diluting droplet contents is droplet picoinjection. This process involves passing a droplet past a pressurised aqueous inlet containing the reagent to be added. This inlet is

situated within an electric field that destabilises the droplet interface and allows fluid from the aqueous inlet to enter the passing droplet. By switching the field on and off, it is also possible to selectively picoinject reagents into a subset of droplets based on an upstream criterion.

To test the ability of picoinjection to add PCR reagents into a fuly stabilised droplet, a chip was designed (fig 4.19) with the following features:

- Droplet inlet
- Spacer-oil inlet
- Pico-injection junction
- Droplet outlet

This chip was initally tested using the same motor driven syringe pumps that have been used for all other flow experiments in this project. Small step angle stepping motors like these are able to provide smooth, accurate flow-rates, however, they are designed to maintain that flow unless instructed to stop via the user interface. Using these syringe pumps with the picoinjection chip resulted in the pressure in the injection channel overcoming the pressure in the droplet channel and the injection reagent was released into the channel outside of a droplet. The injection reagent does not contain surfactant, and as a result causes wetting of the droplet channel as it is released. In addition, the frequency of droplets passing governed how much injection reagent was passed into the droplet. Over time, the pressure at the injection reagent junction increased until it was released by either injection into a droplet or overcame the pressure in the flow channel and entered the droplet channel independent of a droplet. In this system, the injection volume would be lower for droplets that pass more frequently as the pressure does not have as much time to build inbetween droplets. For this chip to work reproducibly, it was necessary to maintain an equilibrium of pressure at the injection reagent channel.

In contrast to motor-driven syringe pumps, pressure pumps operate by exerting a constant pressure on the fluid they are driving. Using a pressure pump should

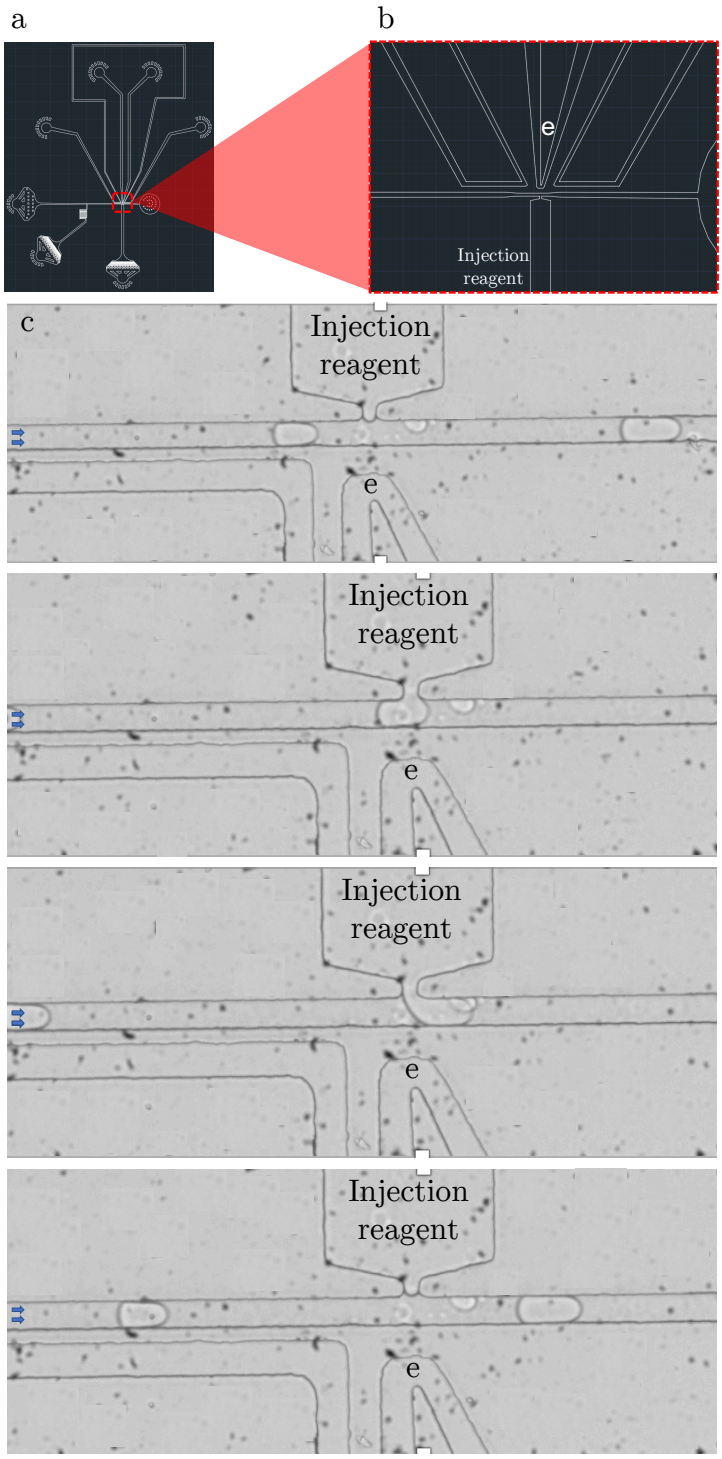


Figure 4.19: On-demand picoinjection of reagents into stabilised droplets (a) Soft lithography mask of the picoinjection chip. (b) Magnified image of the injection reagent junction (red dashed box in (a), e = electrode) (c) 4 still images taken from a high-speed movie showing (i) the droplet approaching the junction, (ii) interface destabilisation and onset of picoinjection, (iii) the final stages of picoinjection before pinch-off, (iv) the picoinjected larger droplet continuing down the channel.

therefore negate the undesirable effects observed using the picoinjection chip with the motor-driven pumps. To test this, an experiment was performed using a pressure pump in place of a motor-driven pump as reported above. Using a pressure pump it was possible to balance the pressure at the picoinjection junction and reproducibly inject the same volume of reagent into each passing droplet. Furthermore, by balancing this pressure, there were no instances where the pressure at the picoinjection junction exceeded that of the flow channel causing injection reagent to escape without a droplet. The picoinjection process can be seen in figure 4.19.c which consists of four still images extracted from a high-speed movie:

- The top still shows a droplet approaching the picoinjection junction; at this stage the droplet interface is intact and the injection reagent is being held at the picoinjection junction using the pressure pump to apply a positive pressure that does not exceed the pressure in the flow channel.

- The second still shows the droplet as it reaches the picoinjection junction. At this stage, the droplet is within the electric field and the interface has been destabilised. As a result of this destabilisation and the attractive forces between the aqueous injection reagent and the droplet contents, a bridge has formed between the two. As positive pressure is being applied to the injection reagent and the interfacial tension is no longer blocking its progress, reagent enters the droplet.

- The third still shows the newly enlarged droplet just before the pinchoff stage. Reagent is still able to flow into the droplet.

- The final still shows the newly picoinjected droplet continuing down the channel. It has now passed out of the electric field allowing the interface to restabilise. With no droplet passing the junction, the injection reagent is once again balanced against the pressure of the flow channel and the interfacial tension with the oil.

This is an elegant solution to adding reagents to droplets however has one major limitation that diminishes its utility in this project; the volume that can be injected into each droplet is limited. It has been shown that the volume injected is proportional to the injection velocity and the time the droplet stays at the junction [211]. Injection velocity is determined by the pressure of the injection reagent at the junction. Increasing the injection pressure, increases the injection velocity, there is a limit on this pressure however, as it must not exceed that of the flow channel coupled with the interfacial tension of the continuous phase. The time the droplet stays at the junction is determined by the flow-rates at the droplet inlet and the spacer-oil inlet. There is therefore, an inverse dependency of injection volume on spacer-oil flow-rate. If this were the only method available, it would be theoretically possible to increase the injection volume maximum by increasing the size of the injection channel. However, as droplet merging allowed for larger volumes to be reproducibly combined, merging was used from this point onwards in the thesis.

4.2.9 Droplet reinjection

Complex droplet microfluidic pipelines, such as the one proposed in this thesis, often contain processes that require either different channel coatings or different flow-rates. There have been attempts to integrate this into a single chip, however these complex chips often involve more sophisticated fabrication techniques or less reproducible methods. To simplify an already complicated pipeline, it was decided to split the entire process over three chips:

- Droplet production chip
- Droplet merging chip
- Droplet sorting chip

By splitting the pipeline, it is possible to thermocycle droplets off-chip allowing for better monitoring and quality control of the process.

By taking this direction, droplet transfer between chips becomes a critical step. Despite stabilisation by surfactants, WOs remain vulnerable to shear stress

induced coalescence. To reinject droplets into a new chip, it is necessary to first collect them from an existing droplet production/merging chip, further process them (if necessary) and then reinject them back into the next chip. Collection is performed using a length of PTFE tubing connecting the droplet outlet of the first chip to a reaction tube. As WOs cream, they will rest at the surface of the continuous phase unless disturbed. Droplets that require thermocycling will also have a mineral oil layer added above the droplet layer, this is a necessary step, however, complicates reinjection.

In order to reinject them into a new chip, droplets are backloaded into PTFE tubing attached to a syringe and pump at an aspiration rate of 5 µl/minute. Both water droplets, and droplets stabilised with EA-008 only will reside in the backloaded tubing without signs of coalescence. Droplets stabilised with both EA-008 and Igepal-CA630 however, will begin to coalesce after ~20 minutes in the PTFE tubing. At present, it is unclear what causes the additional stress on the droplets that drives coalescence in this situation. A temporary workaround has been implemented in which 20 µl of droplets are reinjected at a time and then replenished as required. As experiments can run for many hours and 20 µl will be used up in roughly 20 minutes, this has a negative impact on this working as a hands-off process. Of note, it is only post-thermocycling that this phenomenon occurs. Work is ongoing to elucidate the cause, which will hopefully lead to a more permanent solution to the problem.

4.2.10 Droplet sorting

The ability to sort droplets based on their fluorescent intensity underpins this entire project. Being able to isolate droplets that contain a copy of the HIV genome would allow researchers to investigate a pure population of latently infected cells for the first time. Performing a qPCR reaction targeting HIV within droplets containing primary cells, and subsequently sorting positive droplets has the potential to make this possible.

Droplet sorting based on fluorescent intensity has been discussed in a number of studies [212–215]. A number of methods have been utilised to actively sort droplets including [216]:

- Electric control
- Acoustic control
- Magnetic control
- Pneumatic control
- Thermal control

Of these different methods, the most commonly employed is electric control, in particular, dielectrophoretic sorting. The foundation of this method is the electric force that is generated from the interaction between a non-uniform electric field (propagated from electrodes embedded in the PDMS) and the dipole it induces within the droplet [216, 217]. The initial chip design tested can be seen in figure 4.20.

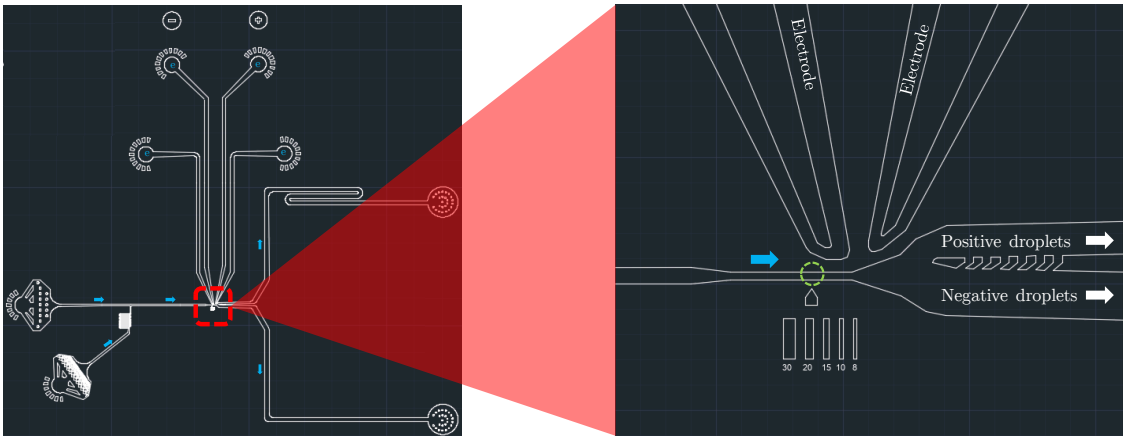


Figure 4.20: Example sorting chip (a) Soft lithography mask of the initial sorting chip, e = electrode. (b) Magnified image of the sorting junction (red dashed box in (a)), green dashed circle indicates the fluorescence detection area

From left to right in the sorting device there are the following elements:

- A droplet inlet

- A spacer-oil inlet
- A spacer-oil junction for the purpose of spacing out the WOs in the channel
- A fluorescent detection point
- A dielectrophoretic sorting junction
- Two droplet outlets, one for positive droplets and one for negative.

There are two key elements to a fluorescent sorting chip: being able to detect fluorescence inside the droplets and being able to actuate a process that isolates positive droplets from the negatives.

4.2.10.1 Fluorescence measurement

As all droplets will emit a low level of autofluorescence from the qPCR probe encapsulated within, the detection of a binary yes or no fluorescent signal is not sufficient to enable successful sorting. It is therefore necessary to design a system that is able to:

1. Measure the fluorescence emitted from a droplet
2. Pass that signal to a microcontroller
3. Determine if the signal is above a certain threshold
4. Actuate the sorting mechanism if required

The simplest and fastest method for detecting the emitted fluorescence is to use a photomultiplier tube (PMT). First proposed in the 1930s [218], PMTs are used throughout the sciences from agriculture to astronomy and their main purpose is to amplify weak photocurrents by up to 10^6x their initial signal. They are capable of ultra-fast responses upon registering photons and can have relatively large sensors.

The mechanism by which the signal is amplified utilises an initial photocathode and a series of in-line secondary dynodes. Each subsequent dynode holds a higher electrical potential than the previous one in the series. A photon colliding with

the photocathode displaces electrons, accelerating them into the first dynode, this in turn causes a greater number of electrodes to be accelerated towards the second dynode. This cycle continues throughout the series with a larger number of electrons being released with each further emission. The flux of electrons released from the last dynode in the series is then transmitted by an anode as the output signal of the PMT [219]. This voltage signal is proportional to the intensity of the emitted light of the droplet from which it emanated. Using a PMT allows a signal much larger than the initial input to be transmitted with relatively little noise. It is worth noting that a PMT is non selective and will detect any wavelength of light that collides with the photocathode. The fluorescent reporter of interest in this project is fluorescein with a maximal emission wavelength of 525 nm, therefore, it is necessary to install filters into the light path to ensure only light of this wavelength reaches the PMT. A diagram detailing the light path filters is shown in figure 4.22.a.

After the signal from the PMT has been amplified, it needs to be interpreted. This is achieved by the means of an Arduino Due (Arduino), a single board microcontroller. The Arduino receives as input, the voltage that is output from the PMT. This voltage is written to a .txt file by means of a custom python program uploaded to the Arduino via its 'Sketch' software. These fluorescence measurements can then be either monitored in real-time or written to memory via a USB peripheral for subsequent analysis.

4.2.10.2 Excitation source and fluorescence detection

Laser induced fluorescence (LIF) has historically been used for applications where sensitivity is expected to be a limiting factor. Advances in technology have also caused a surge in LED induced fluorescence (LEDIF) with LED light sources frequently displacing gas-discharge lamps due to their increased bulb lifespan and stable output [220]. Sensitivity is a particular problem in the microfluidics environment as targets are often travelling at relatively high speeds and therefore will only be aligned with the light source for a short period of time. In static imaging experiments, integration times are often in the region of seconds. For

fluorescence detection in this system, a droplet will be aligned with the light source for microseconds resulting in a limited signal output. At the outset of this project, the only light source available in the facility was an LED light source. Consequently, initial testing was performed using LEDIF to test the PMT signal and sorting performance of the workflow. Example PMT outputs from an LEDIF experiment can be seen in figure 4.21.e. The dense blue region below the red dashed line is the combined noise of the PMT and the continuous phase flowing through the LED spot. All of the peaks above the red dashed line indicate individual droplets that have passed the PMT. This experiment was performed on a population of negative droplets, therefore, the only signal should be a low level of autofluorescence. This autofluorescence should be consistent across the entire population as a homogenous mixture of probe was encapsulated in these droplets. The majority of droplets in this example do appear to be homogenous, however, four droplets (green ellipses) have significantly higher fluorescent intensities than the rest of the population. This frequency of droplets that report a higher degree of fluorescence is consistent over multiple experiments. Parallel analyses were performed on the same population of droplets by imaging them in a viewing chamber constructed by attaching a coverslip to a glass slide with double sided tape. Intensity measurements taken from these images show a modest increase in intensity for a subset of the droplets that were larger than the others. This increase is not reflective of greater fluorescence in the droplet however, but due to a greater z depth from the larger droplet as their shape is not constrained in the viewing chamber. This explanation, although valid for the droplets in the viewing chamber does not hold true for the PMT data. In this scenario, droplets are constrained in z by the dimensions of the channel, resulting in all droplets having the same depth as they pass the detection area. As the frequency of large droplets in the population seemed to broadly fit the frequency of high intensity droplets in the PMT data, it would be reasonable to hypothesise that droplet size may be having an effect on the PMT signal. To test this, a real-time video feed was used to view the fluorescent signal for small and large droplets as they passed the detection area.

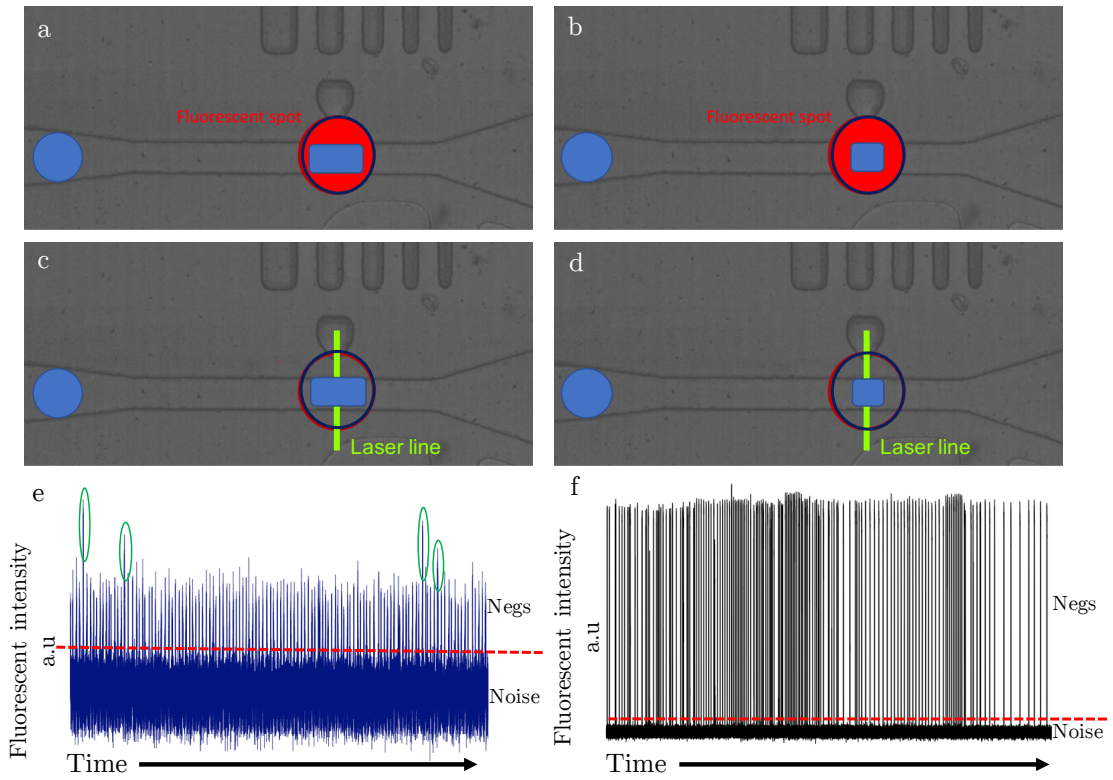


Figure 4.21: Excitation source testing (a) Schematic of a large droplet passing an LED spot excitation source (b) Schematic of a small droplet passing an LED spot excitation source (c) Schematic of a large droplet passing a laser line excitation source (d) Schematic of a small droplet passing a laser line excitation source. (e) Line plot showing the fluorescent intensity profile of negative droplets using an LED spot. (f) Line plot showing the fluorescent intensity profile of negative droplets using a laser line.

The microscope used for this project is a Leica DMIRE2, built into the system are a series of aperture shapes and sizes that allow the field of illumination to be restricted. Before beginning testing, the smallest aperture size available was selected, this happened to be a circle. The size of the fluorescent spot using this aperture was ~1 µm, in contrast, the mean droplet size for this experiment was ~1 µm. The result of this difference between fluorescent spot size and droplet size is that smaller droplets do not fill the detection area and will therefore have a lower total fluorescence value when compared to larger droplets. This effect can be seen more clearly in the schematics in figure 4.21.a&b. The larger droplet in 4.21.a has a larger surface area being activated by the LED when compared to the smaller droplet in figure 4.21.b. This results in a greater fluorescent emission from the larger

droplet despite it having the same mean intensity/volume. Droplet populations are produced to be monodisperse and in a perfect system, this confounding effect of droplet size on fluorescent intensity would not be a serious problem. A small fraction of droplet coalescence will however, always be present, due to the effects of thermocycling and droplet reinjection. If an LEDIF system were to be used in this case it would result in a considerable number of false positives being sorted and would negatively impact the validity of any findings. An alternative method of fluorescent detection was investigated with the aim of reducing the size of the fluorescent detection area.

LIF is most commonly used in droplet microfluidics platforms due to its ability to provide a greater intensity light source. The major obstacle to laser usage however is the set-up required. Installing a laser on to a microscope requires precise alignment of an open beam and additional safety measures. In order to improve the signal for this pipeline and allow greater control over the size of the fluorescent detection area, a laser was purchased and installed on the Leica DMIRE2. A series of lenses was installed to ensure the laser was focused on the detection area of the chip as a line, as in figure 4.21.c&d. The justification in using a line was to ensure the fluorescent area was always less than the size of a droplet in the channel. This should, in theory, remove any confounding effects caused by droplet size whist sorting. Once installed, the system was tested using droplets with a low level of uniform autofluorescence, made to be representative of qPCR negative droplets in our pipeline. As can be seen in figure 4.21.f, the fluorescent signal produced using LIF is much more reproducible when compared to the droplets analysed using LEDIF (fig 4.21.e.). LIF was used for all future sorting experiments.

4.2.10.3 Signal interpretation and actuation of sorting

With a reproducible signal now being detected and reported by the PMT, it was necessary to interpret this voltage signal, define sorting thresholds and trigger sorting on any droplets with fluorescent intensities above a set threshold. To achive this, a pair of microcontrollers were utilised; one connected to a desktop computer

to record droplet intensities and one connected to a function generator to effect sorting. The flow of data in this process is shown in figure 4.22.

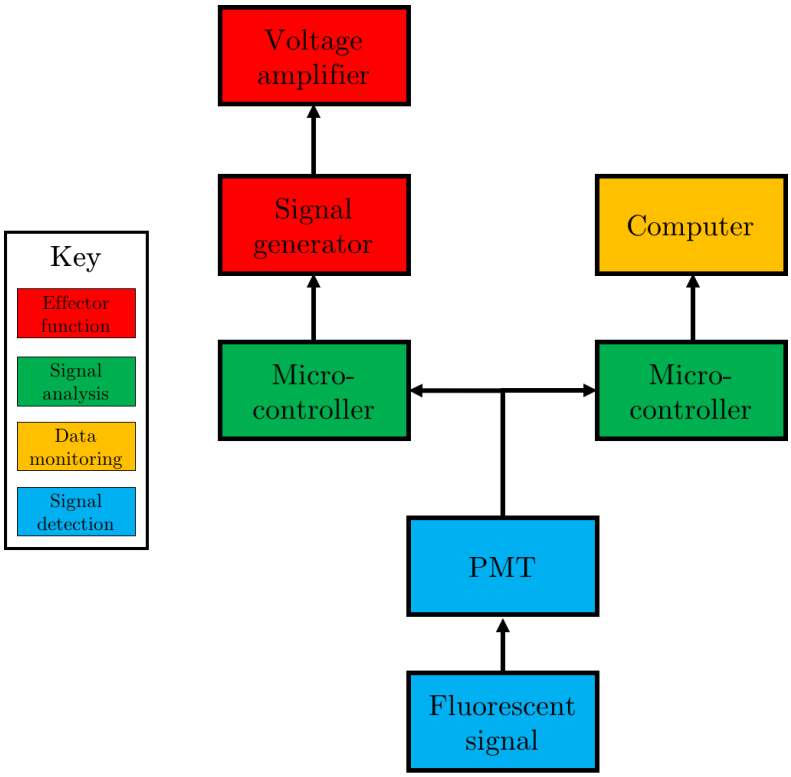


Figure 4.22: Flow of data processing for droplet sorting. Flow of data through the sorting pipeline, direction of flow is from bottom to top.

The fluorescent signal is first detected by the PMT where it is converted to a voltage proportional to its starting fluorescence. The PMT then outputs this voltage signal to two microcontrollers. One microcontroller is connected to a desktop computer where the signal can be observed in real-time and written to memory as a .txt file for subsequent analysis. Due to the RAM capacities of the microcontrollers, only 10^5 records can be stored at any one time before they have to be written to memory. Live monitoring of this feed allows the operator to detect, and compensate for any shift in baseline intensity that can be caused by small device movements on the microscope stage. It is not possible to record every droplet in an experiment using this method as the process of writing to memory takes computing time. Any time the microcontroller is writing data to memory, it is not able to receive new inputs.

The role of the second microcontroller is to trigger the sorting mechanism. When a sorting run is initially set up, a small run of droplets is analysed in order to determine a suitable threshold for qPCR positive droplets. This threshold is then entered into a custom python script and uploaded to the microcontroller. This script instructs the microcontroller to assess whether the signal for any given droplet is above the set threshold, and if so triggers the sorting cascade and adds 1 to a counter. If the signal for that droplet is below the threshold, no further action is taken by the microcontroller. An example plot of an initial droplet run used to set the positive threshold is shown in figure 4.23.

Figure 4.23.a shows the profile of ~3500 droplets that were analysed in order to determine a threshold above which a droplet would be considered qPCR positive. Each peak on the plot represents an individual droplet with ~75 datapoints/peak. In total, 175,000 datapoints are on display. Two populations are clearly visible: the qPCR negative droplets with a low level of autofluorescence (shaded in red) and the qPCR positive drolets with a significant shift in fluorescence (shaded in green). Both the negative and positive populations have consistent fluorescent intensities enabling easy thresholding. In figure 4.23.b, a smaller subset of readings is plotted, allowing a clearer view of individual peaks. The plot displays 10,001 individual datapoints representing 22 droplets. The red arrows in the two individual plots indicate the same droplet providing a reference point in 4.23.b to compare with 4.23.a. As an aside, the droplet frequency seen here is indicative of a feeding channel that is too wide to allow even spacing of the droplets, therefore two pass the detection area in very close succession. This is evident from the spacing of the peaks with two coming through close together, followed by a longer gap before another two are detected. Should they not fill the feeder channel, individual droplets are not separated by the spacer oil but instead, every pair of droplets is spaced causing the distinctive profile seen in 4.23.b. A schematic showing this flow regime can be seen in fig 4.24.

Once a sorting threshold is selected, it is uploaded to the microcontroller via a python script. The script analyses each reading that is passed from the PMT and

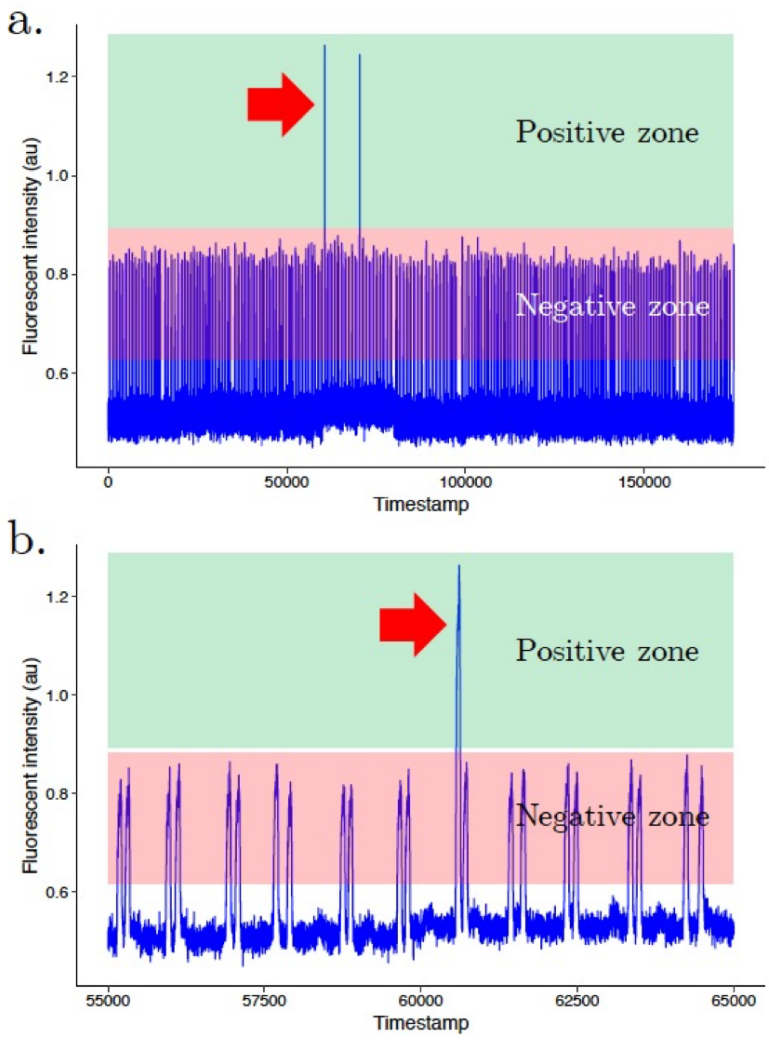


Figure 4.23: Example plot showing initial droplet threshold determination (a) Line plot showing 175,000 sequential fluorescence readings as measured by the PMT. Each peak represents an individual droplet. (b) Line plot showing a subset of 10,001 sequential fluorescence readings extracted from the fig 4.23.a dataset. In both panels, positive thresholded droplets are shaded in green and negative thresholded droplets shaded in red, red arrows indicate the same droplet shown as a reference point between the two plots.

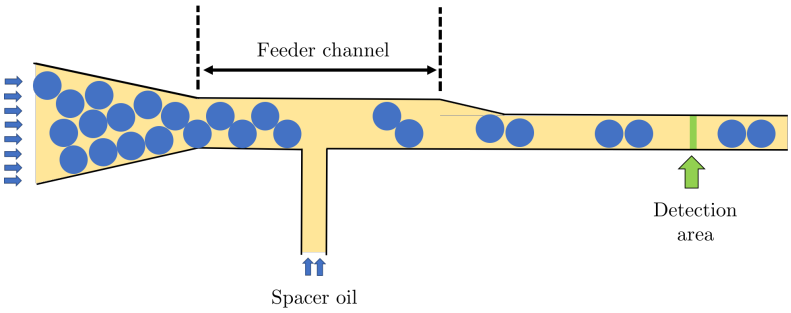


Figure 4.24: Schematic describing the paired droplet flow regime

determines whether it is above the pre-determined threshold. For readings above the threshold, the microcontroller triggers an AC voltage signal in the on-chip electrodes by means of an on-board I/O pin. AC voltage signals are generated using a function generator connnected to a voltage amplifier. This signal is transferred to the chip via salt-water electrodes containing 5M NaCl. The electrodes on the chip are situated close to the channel and have pointed edges in order to apply the largest force upon the droplets [221]. The AC voltage across the electrodes can be modified using the function generator. There is no one voltage that will work universally across different chip designs or droplet types. Voltages instead, have to be tailored to individual experiments based on a number of criteria including:

- Sorting channel flow rate
- Droplet constituents
- Surfactants used
- Electrode geometry
- Sorting junction geometry

This being the case, for the sorting chips used in this pipeline the following conditions shown in table 4.6 are a good starting point for optimisation.

Table 4.6: Sorting parameters

Frequency (kHz)	Voltage (V)	Delay (ms)	Wave	Cycle number
20	300	1	Sinusoid	170

4.2.10.4 Sorting geometries

A number of different chip geometries have been proposed for dielectrophoretic sorting [212, 213, 217, 222]. A common factor for all these chip geometries is a dependence on hydrodynamic resistance (R_H). Droplets are allowed to flow through a chip, past a detection point and towards two separate channels where one channel has a lower R_H than the other. Providing no external forces act upon the droplet, it will enter the channel with the lowest R_H (fig 4.25).

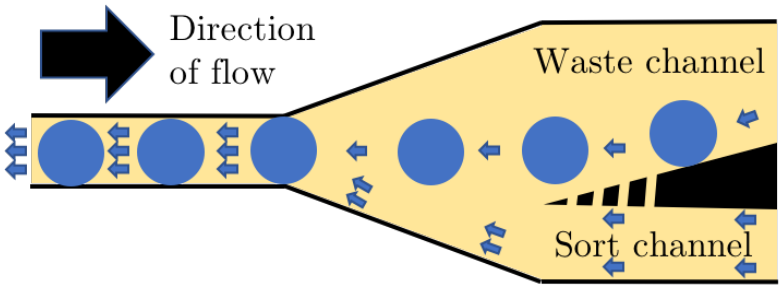


Figure 4.25: Hydrodynamic resistance Schematic showing the H_R in a sorting device. H_R (indicated by the number of blue arrows) is greater for flow moving towards the sort channel.

An imbalance in R_H between two channels can easily be introduced. For example, it can be increased by lengthening or narrowing a channel; this relationship is determined by:

$$R_H = \frac{12\eta L}{wh^3 F}$$

In this equation, L is the channel length, w the channel width, h the channel height, η the solution viscosity and F a geometric form factor related to the rectangular shape of the channel [223]. The introduction of a favoured channel for the droplet enables a sorting chip to only apply an external (sorting) force on droplets selected for isolation. In all the examples referenced above, the droplets selected for isolation are determined by fluorescent intensity and the external force used is dielectrophoresis. This principle is shown in figure 4.26.

Two types of sorting geometries (fig 4.27.a-d) have been tested in this project. Both rely on the dielectrophoretic attraction to sort fluorescent droplets, however,

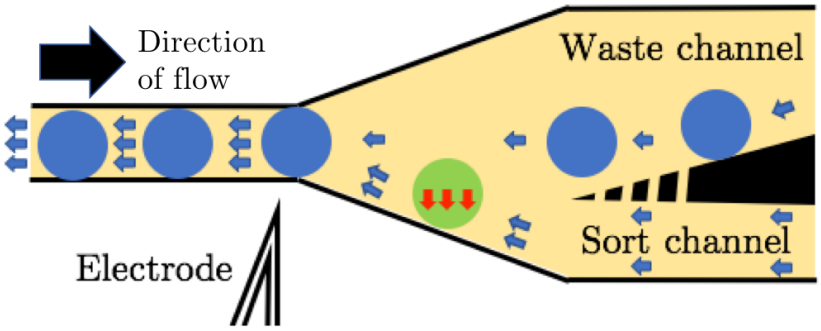


Figure 4.26: Dielectrophoretic sorting Schematic showing dilectrophoretic attraction overcoming the R_H in a sorting device. R_H (indicated by the number of blue arrows) dictates the trajectory for non-target droplets(blue circles) however for target droplets (green circles), the electric field is triggered and dielectrophoretic attraction (red arrows) overcomes the H_R to force the droplet through the sort channel.

whilst 4.27.a relies on R_H to direct negative droplets to the waste channel, 4.27.b relies on laminar flow.

The first design tested (fig 4.27.a), was based on a sorting chip published by Abate et al in 2015 [213]. From left to right in this sorting device there are the following elements:

- A droplet inlet
- A spacer-oil inlet
- A spacer-oil junction for the purpose of spacing out the WOs in the channel
- A fluorescent detection point
- A dielectrophoretic sorting junction
- Two droplet outlets, one for positive droplets (sort channel) and one for negative (waste channel).

Droplets enter the device, where they are evenly spaced by the spacer-oil before passing the fluorescence detection point. Droplets with fluorescence below the pre-determined threshold, do not trigger the function generator and therefore pass through the device along the path of least resistance, the waste channel. Droplets with fluorescence above the threshold, trigger the function generator and

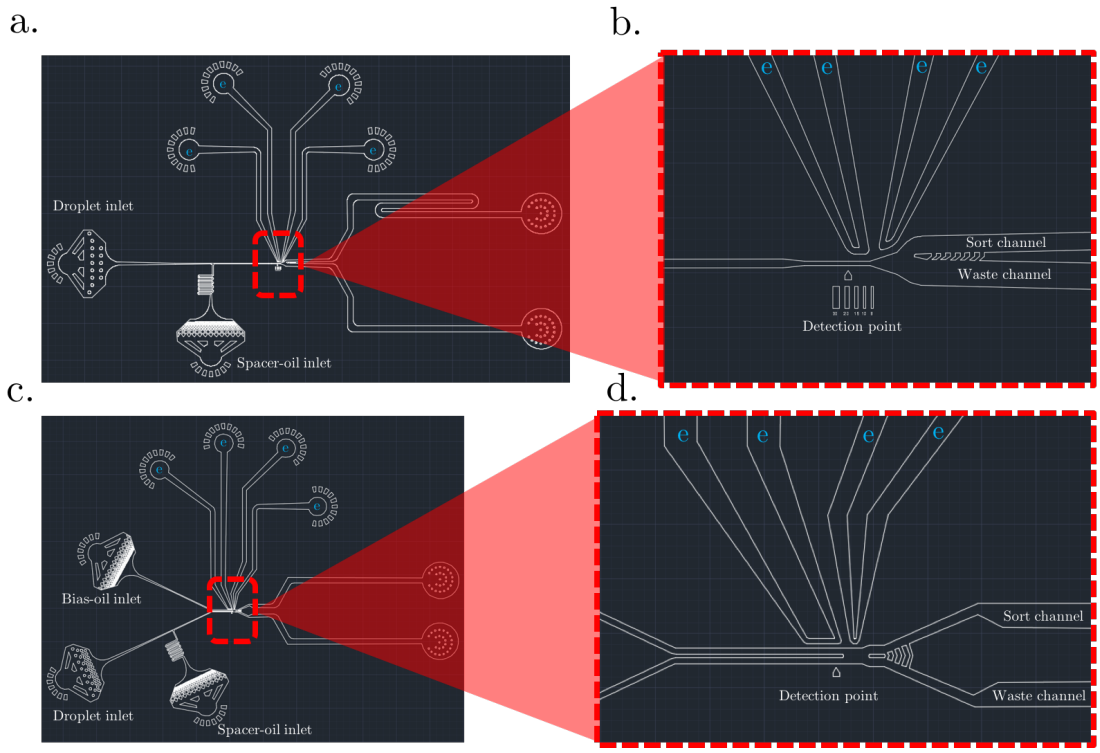


Figure 4.27: Lithography masks of sorting devices used in this project (a) A lithographic mask of device 1. (b) A magnified version of the device 1 sorting junction as indicated by the red dashed box. (c) A lithographic mask of device 2. (d) A magnified version of the device 2 sorting junction as indicated by the red dashed box. (e = electrode).

the electric field attracts them towards the sort channel overcoming the R_H. In this geometry, the electrical field has a narrow margin of error between being too weak and not overcoming the R_H and being too strong and causing droplet coalescence.

Another key consideration when using this geometry is its dependence on slowing the droplets as they are sorted. Once past the fluorescent detection point, the channel widens in order to slow the droplets and make them easier to sort. This is a critical dependence for this geometry and one that requires droplets to be well spaced. If two droplets are too close together as they pass through the chip, a collision will occur when the first droplet slows and the leading edge of the following droplet runs into the trailing edge of the preceding droplet. Collisions occuring at this point in the chip can cause one of the two droplets to be diverted into the sort channel. Mazutis et al, observed that to ensure these collisions do not occur the minimum spacing between two droplets should be 10 times the size of the droplet

[217]. Increasing the space between droplets increases the time that it takes to sort any given droplet number as a greater volume of spacer oil will pass the detector between sorting events. As this pipeline is attempting to sort extremely rare populations (as few as 1 in 1 million cells may be infected), even the most optimistic models of droplet sorting performance using this geometry predict multiple days of sorting in order to collect even 100 postive droplets. On average, a chip is capable of processing ~40 droplets/second. To address this problem, a second geometry that does not require slowing of the droplets was tested (fig 4.27.b).

From left to right in this second geometry there are the following elements:

- A droplet inlet
- A bias-oil inlet
- A spacer-oil inlet
- A spacer-oil junction for the purpose of spacing out the WOs in the channel
- A fluorescent detection point
- A dielectrophoretic sorting junction
- Two droplet outlets, one for positive droplets (sort channel) and one for negative (waste channel).

This second geometry is composed of two distinct channels that share an interface at the sorting junction where the channels are exposed to each other. At this interface, laminar flow of the two channels is maintained, therefore, droplets flowing in one channel will not cross to the other unless acted upon by an external force. As with the first design, if a droplet's fluorescence exceeds the pre-determined threshold, an electric field is triggered attracting the droplet towards the source (the electrode). As can be seen in figure 4.28, droplets that are attracted to the electrode are able to cross the laminar flow barrier and are directed to the collection channel.

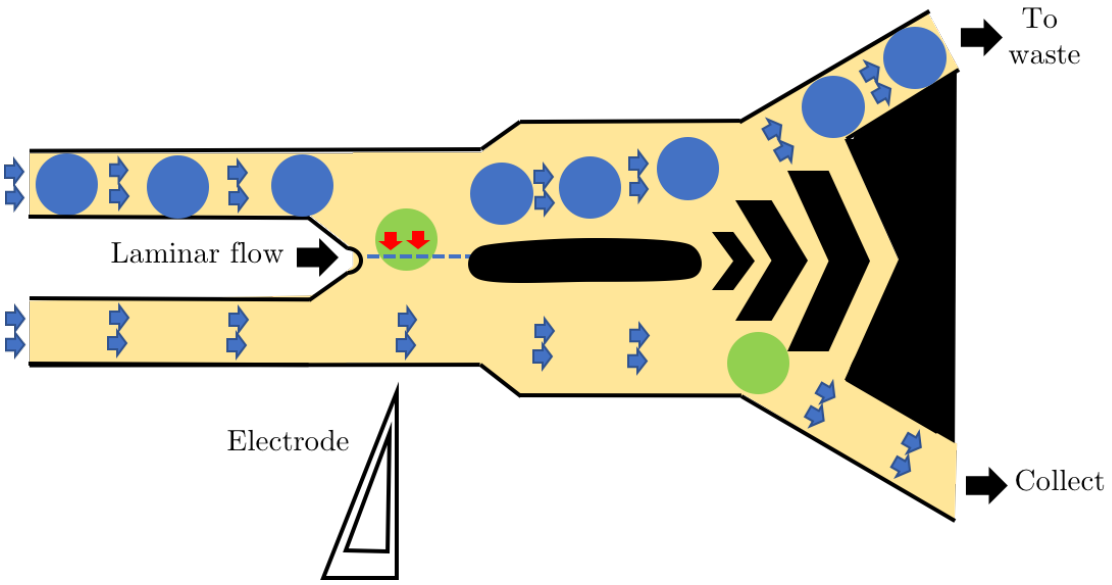


Figure 4.28: Schematic of sorting for second geometry

Using this second geometry, droplet deceleration is less pronounced as for the previous design and therefore can be separated by less spacer-oil. In addition, as the droplet only need cross the interface of the laminar flow before it will proceed along its new course, the length and strength of the dielectrophoretic field is reduced in this geometry, also positively impacting sort times. This chip design is capable of reproducibly sorting ~10 times faster than design 1 at ~400 droplets/second. Based on the data shown above, all future experiments were performed using the second design shown in figure 4.27.c & d.

4.3 Summary

This chapter catalogues an extensive body of work that has progressed from early attempts to produce basic microfluidics chips to the production of complex chips with integrated electrodes capable of generating, merging, picoinjecting and sorting droplets on a large scale. The major outcome of this chapter is that the laboratory is now able to generate a wide range of chips that can be tailored to a variety of different assays. In addition, information that has been generated whilst developing these methods will assist other researchers in pushing microfluidics

assays forward, and finding answers to questions that were previously beyond the technological frontier.

Chapter 5

Droplet microfluidics: In-drop PCR

Contents

5.1 Introduction 137
5.2 Droplet thermostability 138
5.2.1 Igepal CA-360 as a surprise surfactant 143
5.3 PCR compatible in-droplet lysis 147
5.3.1 PCR reagent optimisation 165

5.1 Introduction

A critical component underpinning the success of this project is the ability to perform a PCR reaction within a microfluidic droplet. Despite a wealth of information being available about the kinetics of PCR performed in bulk, it is not known how this translates to the pico-scale when performed inside droplets with volumes of tens of picolitres. Fluid dynamics in picolitre volumes are different to those observed in microlitre scale reactions commonly seen in bulk PCR. In addition, there is the potential for chemical interactions occurring at the droplet interface to consider. A number of studies have utilised in-drop RT-PCR to identify droplets containing RNA targets (transcripts) of interest, however few have attempted to amplify genomic DNA targets. This bias towards transcript targeting is largely a result of the added complication of complete nuclear lysis required for genomic

DNA targeting of single cells.

Detailed in this chapter are a number of the key challenges faced before successful in-drop PCR can be performed.

5.2 Droplet thermostability

In order to retain single-cell resolution throughout this pipeline it is imperative to ensure that droplets remain intact despite the harsh conditions under which they are placed. The droplets are most vulnerable to coalescence whilst being thermocycled; during this period of rapid temperature fluctuation, the volume of the aqueous fraction of the WOs continually shrinks and expands as the reaction cools and warms. These volume fluctuations cause movement of the surfactant molecules packed into the interface, in turn increasing the potential for transfer of the aqueous fraction between droplets. During thermocycling, the composition of the interface is a key determinant of whether the droplets will coalesce or remain intact.

As discussed previously, WOs made purely from water are not stable and are prone to coalescence during thermocycling. To test the stability of WOs containing a PCR mastermix core, control droplets with 1xPBS aqueous cores were compared with droplets containing a 1x PCR mastermix after either: 5, 10, 15, 25 or 35 thermal cycles between 95℃ and 60℃. Droplets containing only water in this experiment did not coalesce completely at any of the cycle numbers tested and a droplet layer was still visible in the tube (fig 5.1.a). Inversely, all droplets containing the PCR mastermix coalesced at every cycle number tested (fig 5.1). This is evidenced by the milky droplet layer that is visible in the water droplets and the coalesced aqueous layer seen in the PCR mix droplets.

The previous experiment showed that PCR mix droplets are not stable during thermocycling, however, as this mix has many components it was important to establish which element was causing the destabilisation in order to determine whether thermostability could be improved. The PCR mix being used is provided as a kit comprised of a 2x mastermix and a separate enzyme. To determine which of these was causing the coalescence, separate tubes of droplets were produced

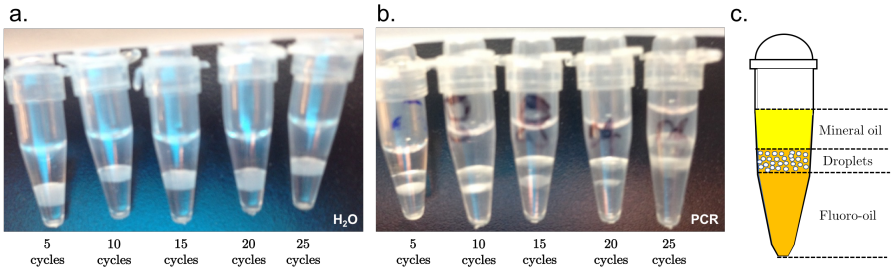


Figure 5.1: PCR mix droplet thermostability (a) Image of water droplets post-thermocycling (b) Image of PCR mix droplets post-thermocycling (c) Schematic drawing of a tube of droplets in a PCR tube with three distinct layers.

Table 5.1: PCR components coalescence

Droplet tube	Water	Enzyme	2x buffer	Coalescence observed
Water only	✔	✘	✘	None
No buffer	✔	✔	✘	Partial
No enzyme	✔	✘	✔	None
Standard PCR mix	✘	✔	✔	Complete

containing the components detailed in table 5.1 and thermocycled between 95℃ and 60℃ for a total of 45 cycles (fig 5.2). Table 5.1 shows the outcome of the coalescence testing for each of these conditions. For all subsequent experiments unless specified otherwise, the dropletised PCR mixes include DNA and a Taqman primer/probe set targeting the human albumin gene.

As expected, the water only droplets did not show any signs of coalescence during thermocycling and the standard PCR mix droplets showed complete coalescence. When the PCR mix was dropletised without the buffer, a partial recovery of coalescence was observed evidenced by a small number of droplets visible at the bottom of the droplet layer (fig 5.1.b). The remainder of the sample had however coalesced. When the PCR mix was dropletised without the enzyme, a complete recovery was observed and there was no evidence of coalescence (fig 5.2.b) This data suggests that it is the enzyme component of the PCR mix that triggers the coalescence of the droplets.

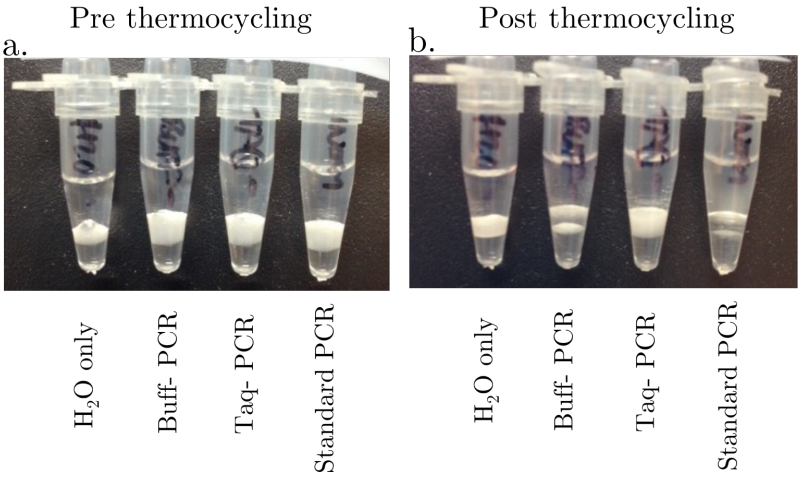


Figure 5.2: Removal of the PCR enzyme recovers droplet thermostability (a) Image of droplets pre-thermocycling (b) Image of droplets post-thermocycling.

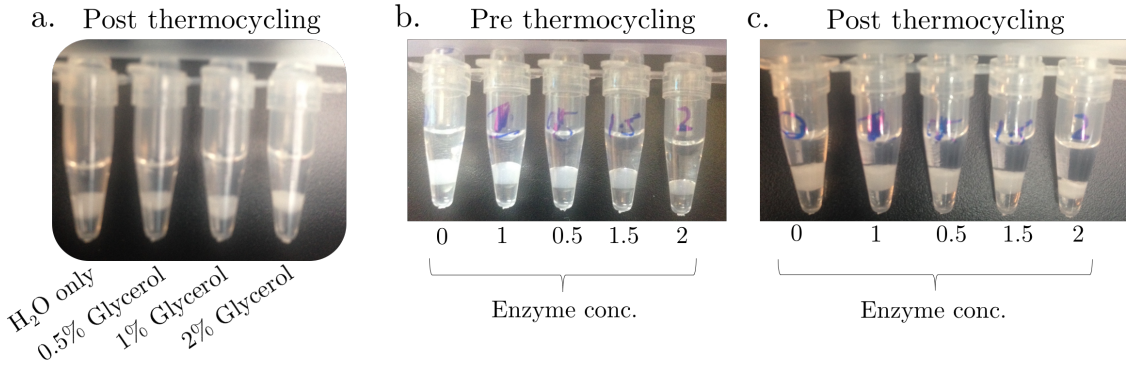


Figure 5.3: Glycerol and enzyme droplet thermostability (a) Image of glycerol titration droplets (b) Image of enzyme titration droplets pre-thermocycling. (c) Image of enzyme titration droplets post-thermocycling.

PCR enzymes are typically stored in ~50% glycerol to allow them to be kept at -20℃ without the reagent freezing and potentially damaging the enzyme [224]. With the exception of water, glycerol is the major diluent used in polymerase formulation and is therefore the agent most likely to be responsible for the increased coalescence observed in PCR mix droplets. To test this hypothesis, droplets were produced from a dilution series of glycerol diluted in water and subsequently thermocycled to evaluate the effect on coalescence. 0, 0.5, 1 & 2% glycerol in water dilutions were chosen as 2% represents the highest final glycerol concentration used in our assays. Droplets produced at all glycerol concentrations showed no signs of

coalescence suggesting that the glycerol within the formulation is not responsible for the coalescence observed previously. As glycerol did not destabilise the droplets, the enzyme was dropletised alone without any other PCR mix components and without any DNA or Taqman primer/probe. Droplets were produced from a serial dilution of the new enzyme in water. The dilutions tested were: 0, 0.5, 1, 1.5, & 2%. As was observed in the glycerol experiment, no coalescence was observed for any of the dilutions tested (fig 5.3.b & c) suggesting that there is an additive effect occuring when all of the PCR mix components are combined.

As the enzyme alone is not sufficient to cause coalescence, an experiment was designed to establish which combination of reagents resulted in coalescence and which did not. Droplets were produced with aqueous cores composed of the mixtures found in table 5.2 and thermocycled as per the previous experiments.

Table 5.2: PCR mix 2 components coalescence

Droplet tube	Water	Enzyme	2x buffer	Coalescence observed
Water only	✔	✘	✘	None
No buffer	✔	✔	✘	Limited
No enzyme	✔	✘	✔	Partial
Buffer & enzyme only	✔	✔	✔	Complete
Standard PCR mix	✘	✔	✔	Complete

As can be seen in the coalescence column of table 5.2 and fig 5.4, no coalescence was observed for the water only droplets and complete coalescence was observed for the combined PCR mix. Of note in this experiment, is the observation that although the enzyme only and buffer only conditions lead to limited or partial lysis respectively, when both are encapsulated together, complete lysis is observed (fig 5.4). This data confirms an additive effect of the buffer and the enzyme together as the cause for complete coalescence in these droplets. It was possible that the coalescence observed during thermocycling may have been unique to the PCR

mastermix used. As a result, another PCR mix was sourced and tested to determine whether it triggered coalescence.

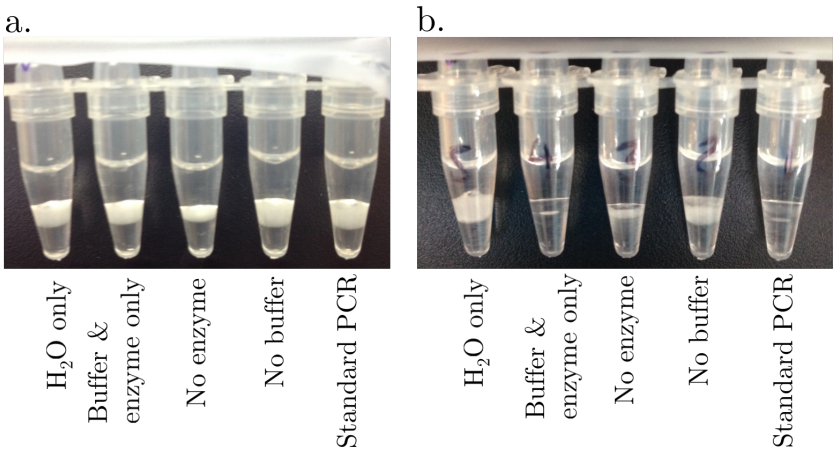


Figure 5.4: Droplet thermostability of different PCR components (a) Image of droplets pre-thermocycling (b) Image of droplets post-thermocycling.

Enzymes that are suspended in ~50% glycerol pose problems for liquid handling as their high viscosity makes them difficult to accurately aspirate and dispense. The widespread adoption of liquid handling robots being used for PCR plate set-up has driven a need for glycerol-free enzymes as the robots are unable to manage the pipetting required to reproducibly dispense the enzymes. As glycerol may be a contributing factor to the droplet coalescence, a glycerol free enzyme was selected to test in the droplets. An initial test was performed by dropletising a standard PCR reaction prepared as per the manufacturer's recommendations and thermocycling them between 95℃ and 60℃ for a total of 45 cycles. The results of this experiment are shown in figure 5.5. Using the new glycerol-free enzyme and buffer, no coalescence was observed post-thermocycling.

Despite appearing as if there was no coalescence in the tube as the droplets survived, upon closer examination under a microscope, evidence of coalescence was found. Figure 5.5.c shows a representative image of the droplet population visible in the H_2O only tube in 5.5.a. In this image, a polydisperse droplet population can be seen, indicating that although droplets survive the thermocycling process, there is still a low level of coalescence or droplet merging causing the droplets to become polydisperse.

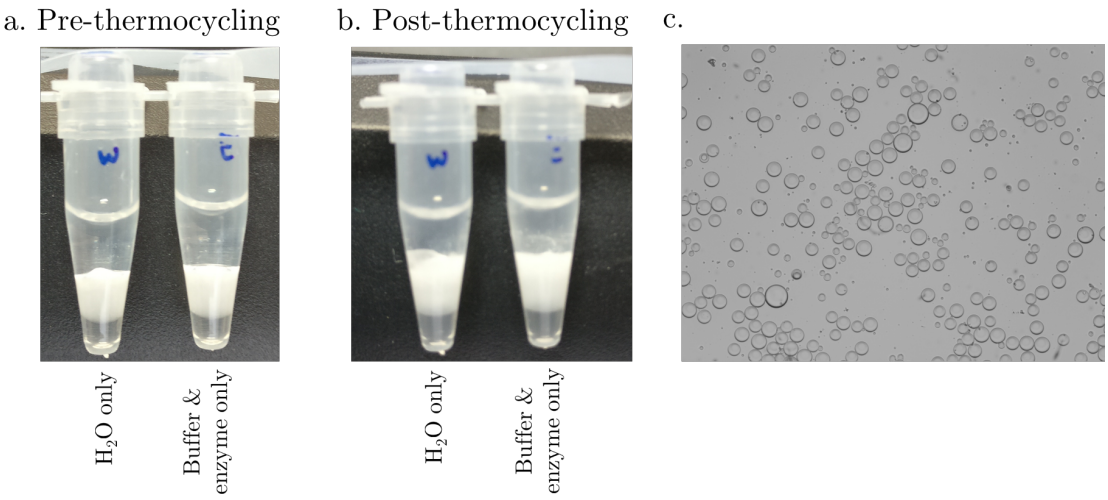


Figure 5.5: Glycerol-free enzyme does not cause coalescence (a) Image of droplets pre-thermocycling (b) Image of droplets post-thermocycling. (c) Magnified image of the 'H_2O only' droplets (10x magnification).

5.2.1 Igepal CA-360 as a surprise surfactant

This section of the project was performed in parallel with the cell-lysis experiments detailed in chapter 5.3. As part of the lysis experiments, the chemical Igepal CA-630 (Igepal), was added to the aqueous core of the droplet to facilitate lysis of the cell membrane. Whilst completing these experiments, it was noted that droplets containing Igepal appeared much more uniform in size post-thermocycling, when compared to similar droplets that did not contain Igepal. As can be seen in figure 5.6, the droplets containing Igepal remain unchanged post-thermocycling when compared to the water droplets. It is noteworthy that the comparison drawn here is with water droplets, which have previously acted as the positive control for thermostability. With that in mind, the addition of Igepal represents a step-change in the process with no evidence of coalescence present in any subsequent tests.

Igepal is a detergent and therefore a surfactant. The likely reason that Igepal stabilises the droplets so well compared to using other additional surfactants in combination with EA-008 is the size of the molecule. Igepal is considerably smaller than EA-008 and may therefore be able to fill the gaps between the EA-008 molecules and diffuse faster to fill gaps when the interface is changing during

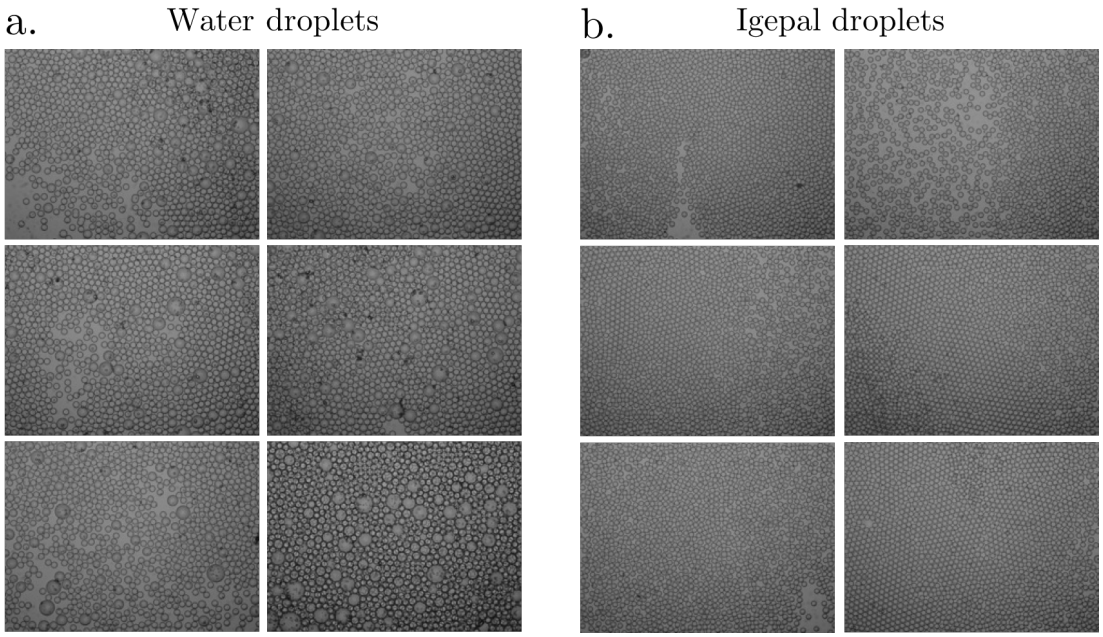


Figure 5.6: Igepal mediated droplet thermostability (a) Representative images of water droplets post-thermocycling (b) Representative images of PCR mix droplets spiked with Igepal post-thermocycling.

thermocycling. A simplified drawing of the Igepal stabilised droplet structure is shown in figure 5.7.

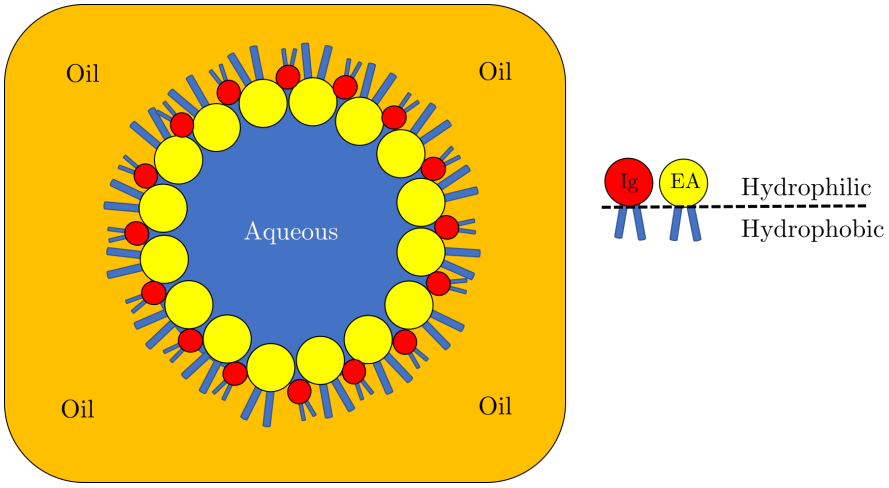


Figure 5.7: Schematic of an Igepal stabilised droplet.

Up until this point, the effect of Igepal-CA630 on droplet thermostability had only been qualitatively assessed by eye without any quantitative measurement. To ensure the observed effect was real and not an effect of investigator bias, the droplet populations were measured using the MATLAB 'findcircles' script. This

script detects circular shapes in images and using a graticule image to apply a global scale will output a diameter measurement for each circular shape detected. For figure 5.8, droplets were produced using the glycerol free PCR mix. One set of droplets were made with Igepal and one set were made without Igepal, droplets were subsequently thermocycled for 45 cycles between 95℃ and 60℃. Figure 5.8.a & b show examples of brightfield images from the with Igepal and without Igepal populations respectively. 5.8 c & d show the same two images with the MATLAB find circles detection algorithm overlaid. Red and white dashed circles indicate where the algorithm has identified a droplet. As predicted, the water droplets exhibit a higher range of polydispersity when compared to the Igepal droplets. This increase in polydispersity is caused by an increased frequency of larger droplets caused by coalescence. This is evidenced by the increase in the mean radius of the droplets without Igepal when compared to the population with Igepal (18.59 compared to 15.65). In addition, the standard deviation for the population without Igepal is 3.56 compared to 0.56 for the population wth Igepal. Summary statistics for the two populations as measured by the MATLAB 'findcircles' algorithm are shown in figure 5.3.

Table 5.3: The effect of Igepal on droplet radius post-thermocycling

	Droplet radii			
Droplet population	Mean	Median	Range	SD
No Igepal	18.59	17.93	87.9	3.56
With Igepal	15.65	15.64	26.1	0.56

The addition of Igepal, regardless of its impact on lysis for which it was originally tested, was a critical milestone towards the successful completion of this pipeline. The thermostability of droplets had become a serious obstacle to this project and although early testing with the glycerol free PCR mix showed promise, a moderate level of polydispersity remained post-thermocycling. Polydispersity, in particular large droplets, cause downstream problems in sorting as they are too wide for the

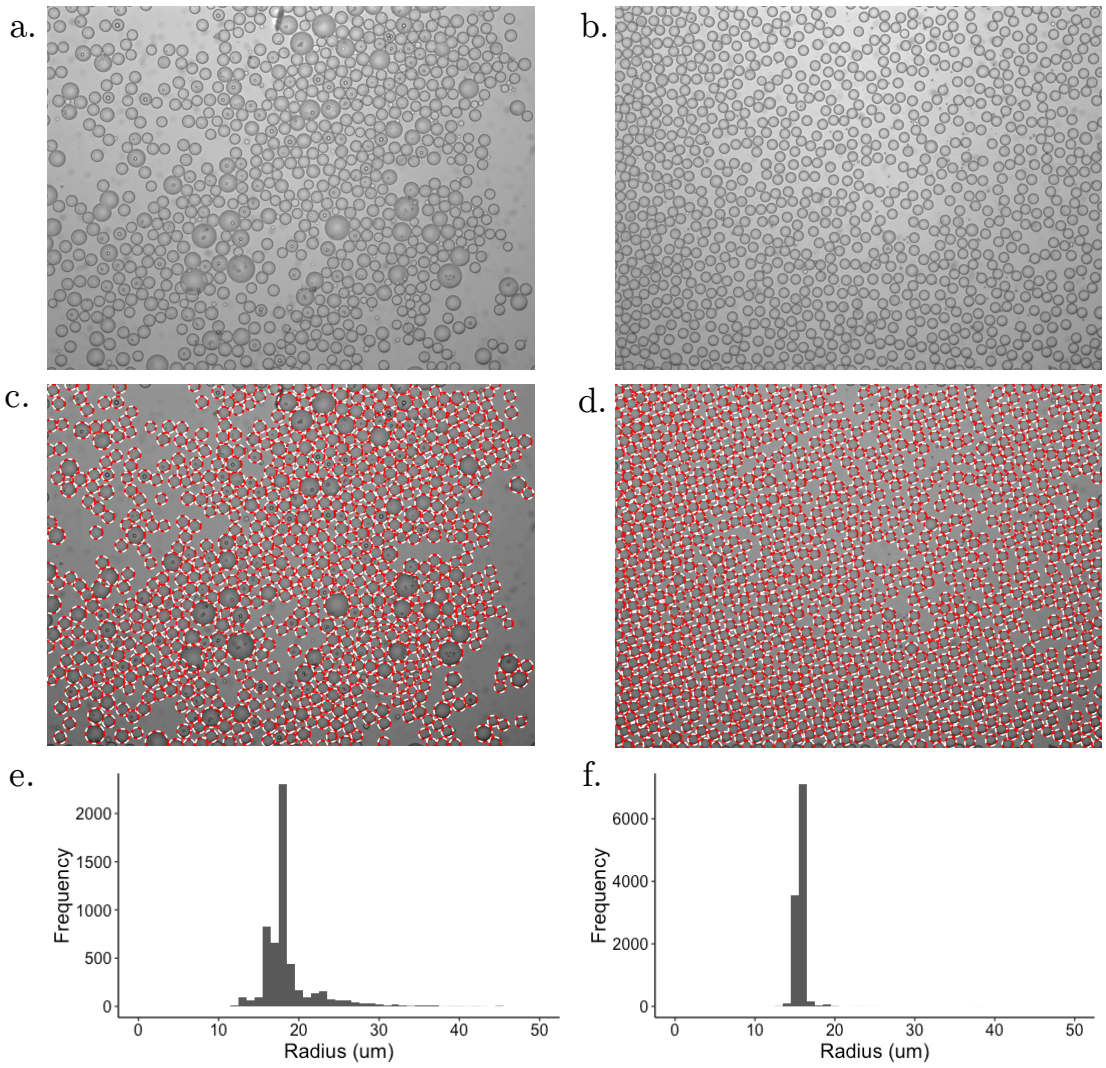


Figure 5.8: Droplet size analysis performed in MATLAB (a & b) Brightfield images for Igepal$^-$ and Igepal$^+$ droplets respectively. (c & d) Brightfield images for Igepal$^-$ and Igepal$^+$ droplets respectively with MATLAB findcircles masks overlaid. (e & f) Distribution plots of droplet size for Igepal$^-$ and Igepal$^+$ droplets respectively.

channels through which they are passing. As they are forced through the channels, they are deformed to the point of coalescence, at which point the aqueous core of the droplet can bond the channel surfaces of the chip and lead to wetting problems.

Before Igepal could be approved for this pipeline, it was necessary to show that it was compatible with successful PCR amplification. In order to test this, a qPCR experiment was designed to compare the standard albumin and total HIV assays with an equivalent reaction containing Igepal. Where, Igepal was added to a reaction mix, an equivalent volume of water was removed. Igepal was tested at

a final concentration of both 0.5% and 4%. 0.5% is the concentration at which Igepal stabilises droplets to be thermocycled and 4% is the concentration at which Igepal has lytic activity on the cell membrane. Results from this test can be seen in figure 5.9. Both the 0.5 and the 4% Igepal concentration showed a reduction in sensitivity as indicated by a higher Ct value for all concentrations tested. As a binary 'yes' or 'no' result is all that is required in order to sort positive droplets, and the effect of Igepal on droplet thermostability is so significant it was decided to continue with Igepal as the stabilising reagent.

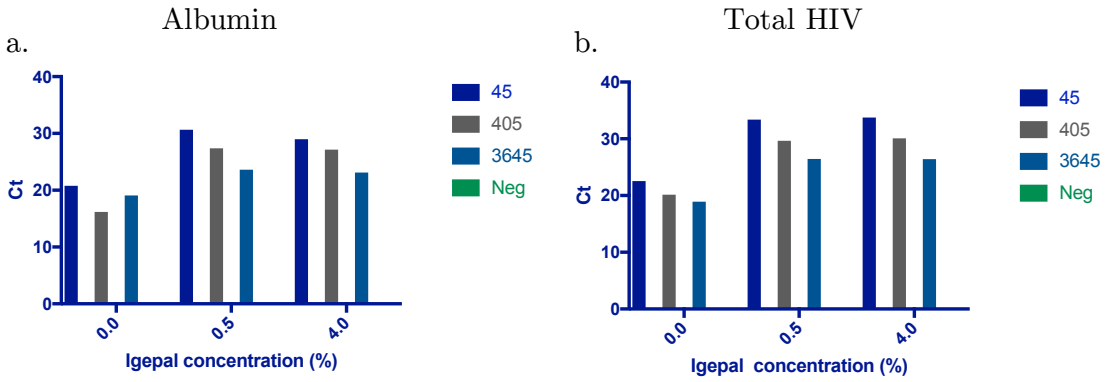


Figure 5.9: qPCR compatibility testing of Igepal (a) Bar plot showing the change in CT values for three known albumin standard concentrations (45, 405 & 3645 copies). (b) Bar plot showing the change in CT values for three known albumin standard concentrations.

As a result of this section of work, Igepal was used at a final percentage of 0.5% in all droplets that were to be thermocycled.

5.3 PCR compatible in-droplet lysis

To obtain single-cell resolution using this pipeline, it is necessary to first encapsulate intact individual cells within droplets. Once encapsulated, a PCR reaction must be performed inside the droplet, a prerequisite for this is that reagents must be able to interact with the genomic DNA within the cell. As intact cells are encapsulated, it is therefore necessary to perform intra-droplet cell lysis that is capable of disrupting: the cell membrane, the nuclear membrane and also opening up the coiled genomic DNA to allow the reagents access to the target.

The cell membrane consists of a phospholipid bilayer ~5 nm thick, studded with proteins that enable essential molecular transport in and out of the cell (fig 5.10.a) [225]. The nuclear envelope consists of two membranes: the inner nuclear membrane (INM) and the outer nuclear membrane (ONM), that are continuous with the endoplasmic reticulum (ER) [226]. Despite this continuity, the protein composition of each is unique and there is a shared structure between the INM and ONM. This nuclear pore complex is a ~40 nm aqueous channel which facilitates the bidirectional transfer of nucleic acids and proteins between the cytoplasm and nucleoplasm [227].

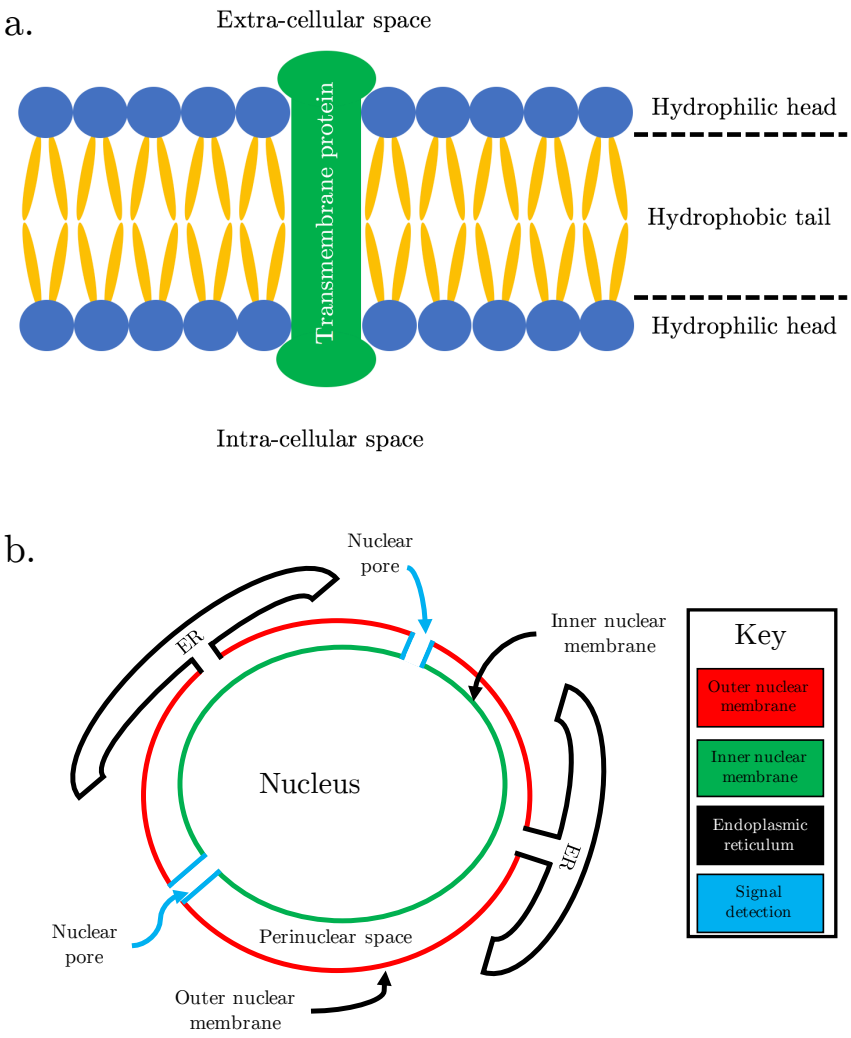


Figure 5.10: Schematic of cell membranes Schematic drawings of (a) the phospholipid bilayer cell membrane and (b) the double membraned nuclear envelope

The nuclear envelope, comprised of a double membrane is more resistant to lysis than the phospholipid bilayer that makes up the cell membrane. As a result, a different lysis method is required dependent upon whether RT-PCR from transcripts or PCR from genomic DNA is required.

Cell lysis is usually performed by either mechanical means or utilising aggressive lytic enzymes. Performing cell lysis in droplets places strict restrictions on the methods that can be used. Mechanical means of cell lysis, for example, are not suitable as any processing capable of disrupting the cell membrane is likely to cause droplet coalescence. Enzymes commonly used for cell lysis have similar limitations; this due to the fact that once lysis is complete, it is not possible to wash the DNA and remove the lytic enzyme as would be performed in bulk reactions. Any downstream applications therefore, have to be performed in the presence of the enzyme. The DNA polymerase used in PCR will be denatured by the same enzymes that cause cell lysis and is therefore not compatible with these methods.

One potentially droplet-friendly lysis method is thermal lysis where high temperatures damage the membrane by denaturing the proteins [228]. This method requires no mechanical agitation or the addition of extra reagents to the droplet making it an attractive proposition for this system. In order to test thermal lysis, it was first necessary to determine if heat alone was capable of lysing cells to the extent that genomic DNA would be made available for primer binding and subsequent PCR amplification. Thermal lysis is more effective when heat/freeze cycles are used; in this method, samples are moved between a hot waterbath and an isopropanol bath filled with dry-ice. The added efficiency over heating alone is due to the production of ice crystals during the freezing phase which rupture cell membranes as they form. The mechanical damage caused by these ice crystals would also lead to droplet instability in this system and therefore, heat alone was used to attempt to lyse the cells.

To establish proof of principle, an initial experiment was performed by heating the GXR cell-line to 98℃ over a timecourse and subsequently assessing cell

membrane integrity using the Muse cell analyser. The Count and Viability Assay provided with the Muse contains two DNA intercalating dyes:

- One membrane-permeant dye that will stain any nucleated cell in the sample
- One non-membrane-permeant dye that will only stain cells with compromised membranes.

By combining these two dyes it is possible to determine the fraction of cells in a sample that have had their membranes compromised using $\frac{C_T}{C_V}$ where C_T is the total cell number in the sample and C_V is the number of viable cells in the sample (cells with their membranes intact). Detailed in table 5.4 are the experimental results.

Table 5.4: Thermal lysis initial testing results

Time at 98℃(mins)	Events aquired	Total cells assayed	Cell viability (%)
0	1034	$7.46x10^5$	71.2
0.5	1054	$9.09x10^5$	0.4
1.5	1052	$9.39x10^5$	0.8
2.5	1074	$1.02x10^6$	0.8
3.5	1083	$1.04x10^6$	0.6
4.5	1064	$9.72x10^5$	0.8
5.5	1090	$9.77x10^5$	0.9
6.5	1070	$1.09x10^6$	0.9
7.5	1083	$7.79x10^5$	1.3
RT control	1035	$5.19x10^5$	69.6

Example plots from the Muse Count and Viability Assay can be seen in figure 5.11(a-f). Cell membranes were rapidly compromised when heated to 98℃. The earliest timepoint tested (30 seconds), yielded a cell viability of only 0.4% compared to an unheated cell viability of 71.2%. This suggests that after only 30 seconds at 98℃, both dyes were able to cross the cell membrane and bind the DNA within the cell. Longer durations of heat had no additional effect on cell viability (fig 5.11.g).

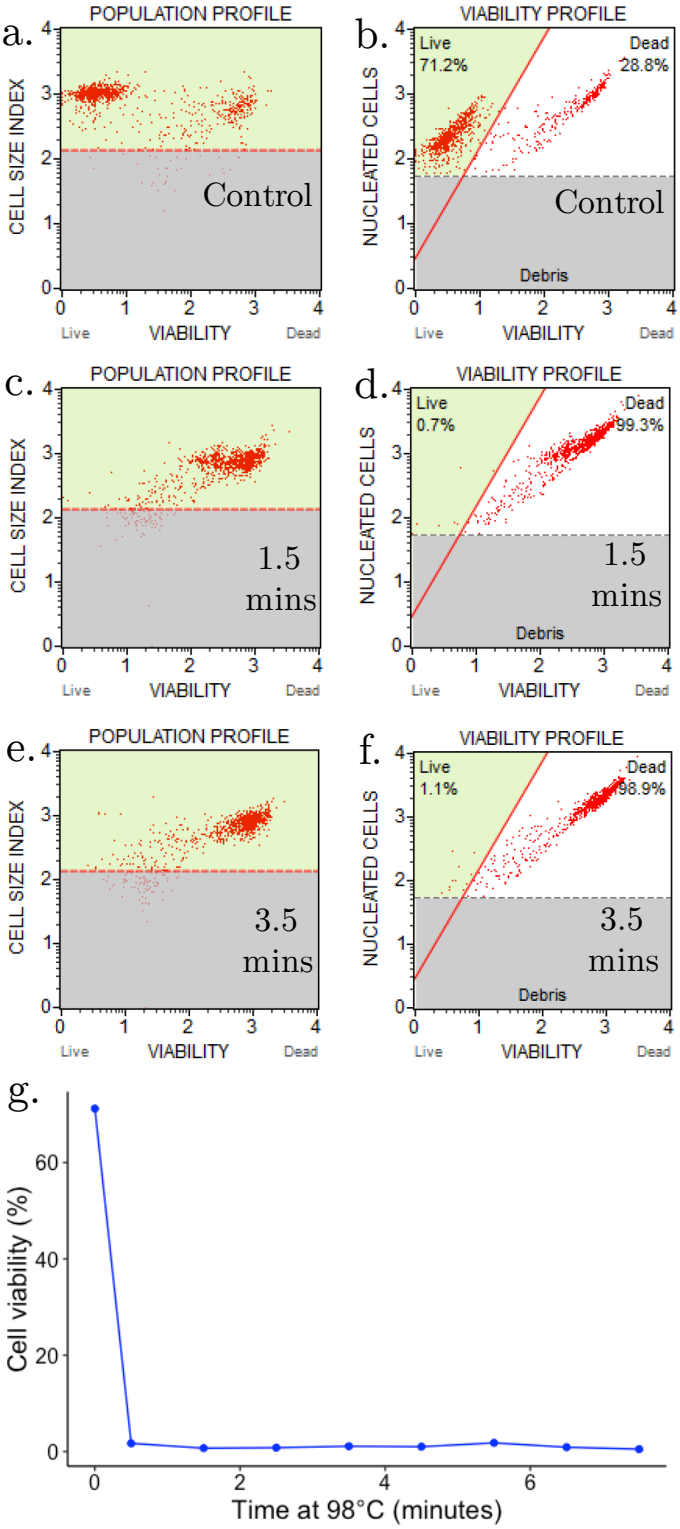


Figure 5.11: Thermal lysis (a,c & e) Example plots showing cell size indexing of three different conditions (Control, 1.5mins and 3.5mins). Cells falling within the grey shaded region were excluded as debris. (b, d & f) Example plots showing Live/Dead statistics for three conditions (Control, 1.5mins and 3.5mins). Cells falling within the green shaded region were demmed to be alive with intact cell membranes.

The previous experiment demonstrated that heat alone was sufficient to allow the dye access to double stranded DNA. In order for a successful PCR however, it is necessary to ensure the primers and probe in the PCR mastermix are able to mix fully with the genomic DNA to allow binding and subsequent rounds of amplification. To test whether this was the case for heat-lysed cells an albumin end-point PCR was used to assess the sensitivity of heat-lysed cells in comparison to extracted DNA from the same cell-line. For all concentrations tested, the PCR was successful, with the band of interest visible. For high cell concentrations $\geq$ 1x10^4/well, there was evidence of DNA becoming trapped in the loading well of the gel lane for the heat-lysed cell samples. As this is not evident for the equivalent DNA concentration, it is likely caused by the accumulation of cellular debris blocking the pores of the gel. Two dilution series were tested ranging from 50,000 cells/well down to 10 cells/well with consistent results between the two experiments (figs 5.12.a & b).

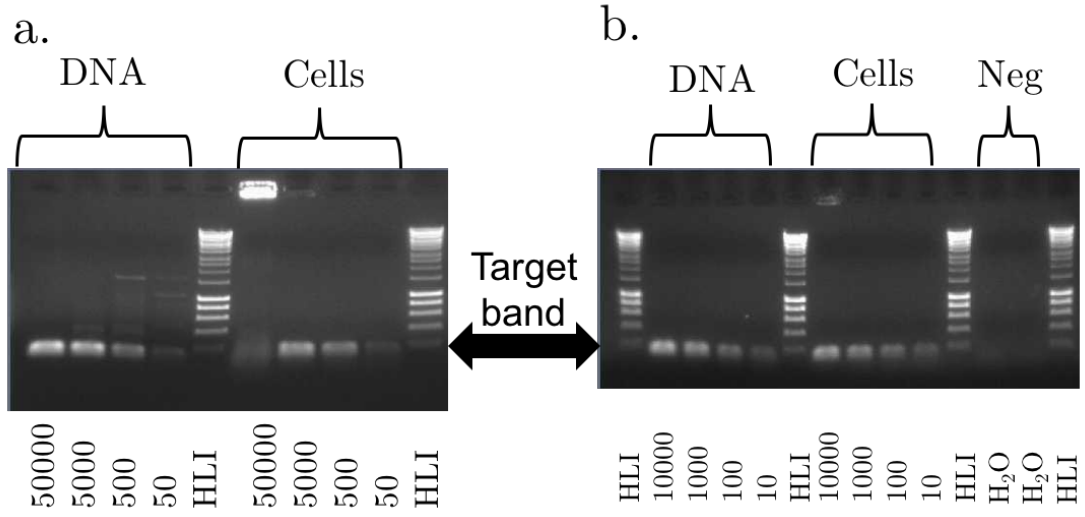


Figure 5.12: End-point albumin PCR on heat-lysed cells and extracted DNA Ethidium Bromide gel showing amplified bands from the heat-lysed cells albumin PCR when compared to extracted DNA. HLI indicates the well containing HyperladderI.

With the end-point PCR working effectively from a cellular input with heat lysis, it was necessary to dropletise the reaction to test the efficiency in-drop. To test this, cells were co-encapsulated with either the albumin or the total HIV qPCR

mastermixes and thermocycled for either 1 or 45 cycles at 98℃. As the cells used for this experiment were GXRs (an uninfected cell-line), the hypothesis was that the albumin reaction droplets should fluoresce and the total HIV reaction droplets should remain dim.

As can be seen in figure 5.13, there is a fluorescent signal detected, however, instead of being homogenous throughout the droplet as expected, it is localised to the cell. In addition, when the brightfield images were analysed (fig 5.13), there was evidence of a normal cell structure, suggesting the cell was still intact. This finding was contrary to the earlier experiment that found heating was sufficient to degrade cells to a point where the genomic DNA was available for PCR. In order to improve cell lysis beyond heating alone, it would appear that the use of an additional lysis buffer would be necessary.

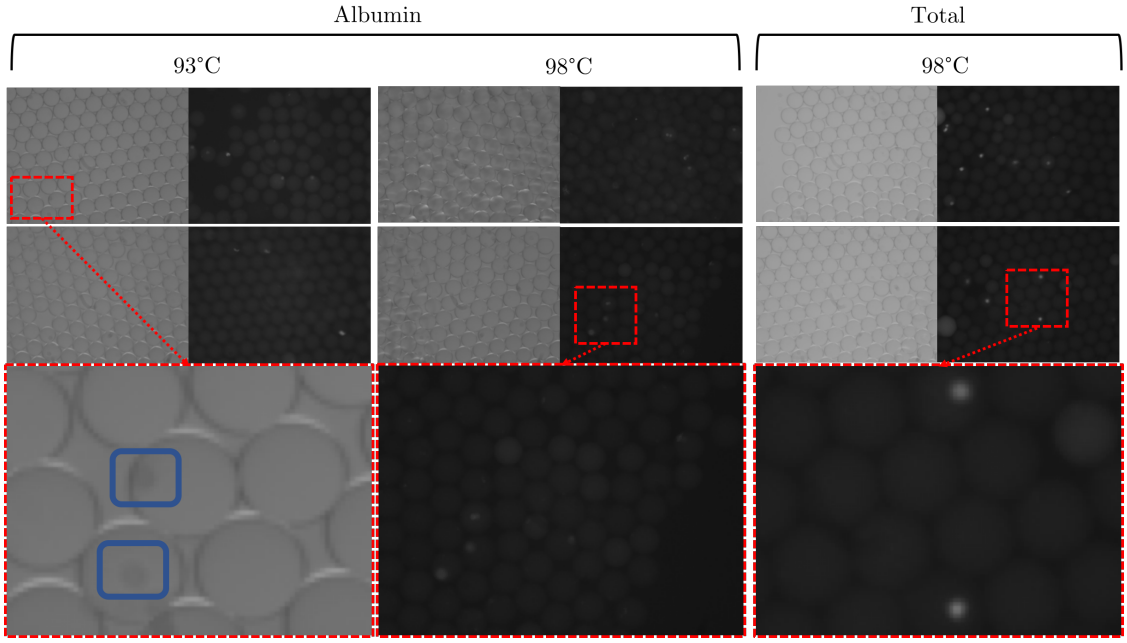


Figure 5.13: Test for albumin PCR amplification in-drop Brightfield (light background) and fluorescent (dark background) droplet images from three different conditions tested for albumin PCR amplification. Large red hashed boxes indicate magnified regins of the smaller red hashed boxes. Bright spots in the fluorescent images indicate localised qPCR probe. Solid blue boxes highlight cells encapsulated inside droplets.

As some lysis buffers are detergents, they share many of the properties of a surfactant and can therefore be drawn to the droplet interface reducing its in-drop

concentration. Lysis was first tested by co-encapsulating cells with lysis buffer in droplets and incubating at 98℃ to visually inspect lysis. An obvious reagent to use was the Phusion blood direct PCR kit (Phusion direct). This kit claims to amplify DNA from whole blood. The cell suspension being encapsulated in these droplets is a less complex sample and has less inhibitors. Theoretically therefore, this kit should be capable of handling the in-drop lysis and amplification required. To test this, an experiment was designed where Phusion direct was encapsulated with the cell suspension and thermocycled according to the manufacturer's recommendations as below:

Condition 1			Condition 2		
Temp (℃)	Time (mm:ss)	Cycles	Temp (℃)	Time (mm:ss)	Cycles
98	03:00	1	98	03:00	1
93	00:15	}40	98	00:15	}40
60	01:00		60	01:00	

Two conditions were tested as the kit requires an initial 98℃ step to activate the enzyme and aid lysis, in addition to this, the standard protocol returns to 98℃ with each subsequent cycle for the denaturation step. To avoid undue thermal stress on the droplets, an alternate condition was tested that only returned to 93℃, the hypothesis being that the enzyme would already be activated and a more conventional DNA denaturation temperature could be used. For this experiment, a working reaction mix was used including the Taqman albumin primer/probe mix, successful lysis and amplification should therefore yield a droplet with homogenous fluorescence.

For both conditions tested, cells appear intact after thermocycling suggesting that incomplete lysis has occurred. In addition, regions of high fluorescence are found colocalised in areas where a cell is present in brightfield images. From this data, it is not possible to deduce whether this increased fluoerescence is due to PCR amplification or whether the fluorescent qPCR probe merely has an affinity with the cell which causes this colocalisation. It is possible for example, that the

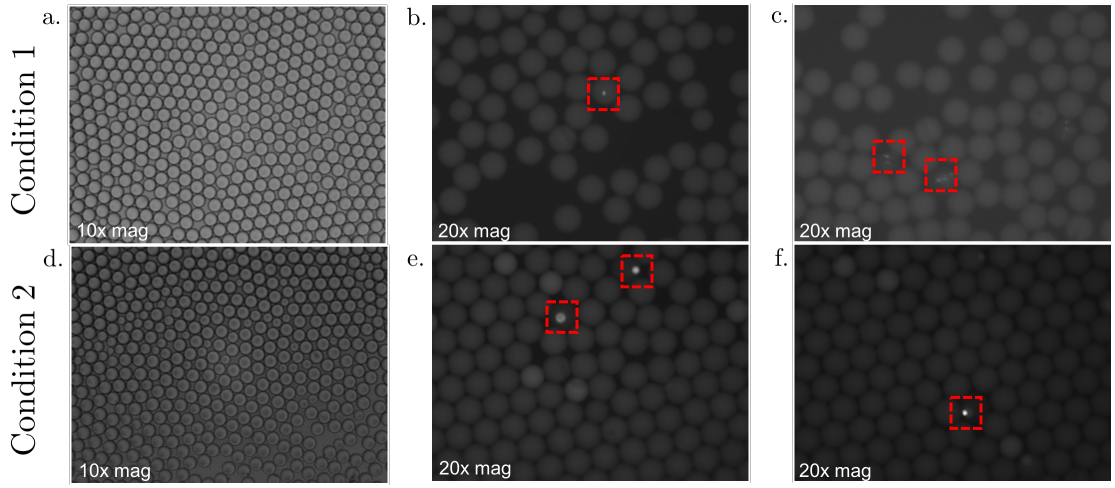


Figure 5.14: Test of direct lysis kit Brightfield (light background, a & d) and fluorescent (dark background, (b, c, e & f)) droplet images from two different thermocycling regimes tested for albumin PCR amplification. Red hashed boxes indicate droplets containing a cell. Bright spots in the fluorescent images indicate localised qPCR probe.

probe is able to enter the cell and has some weak affinity for cellular materials or is able to bind genomic DNA in the nucleus. As this material is densely packed within the cell, this would result in the localised fluorescent signal seen in these droplets. To confirm that the fluorescence observed is not a result of PCR amplification of the intended target, an experiment was designed using droplets incompatible with PCR amplification whilst still including the probe in the reaction mix. Table 5.5 shows the reagent mixtures encapsulated in droplets as well as their expected and observed amplification status post-thermocycling.

Amplification was deemed successful if droplets containing a cell had a homogenous fluorescent intensity. As hypothesised, no amplification was observed for any of the three conditions. This data shows that the localised fluorescence observed in droplets is not a result of amplification of the albumin target. All three conditions tested were designed to be incompatible with PCR amplification and all three had similar fluorescent intensities and a similar pattern of colocalisation with a cell (fig 5.15). This was assessed by creating a mask of the cell locations in the brightfield images (fig 5.15.a-c) and superimposing them over the paired fluorescent image of the same droplet population. A small amount of drift can be observed between the cell location mask and the bright spots in the fluorescent

Table 5.5: Mismatched primer/probe experiment design

Condition	Probe	Mastermix	Matched primer	Mismatched primer	Amplification expected	Amplification observed
Probe only	✔	✘	✘	✘	✘	✘
Probe & mastermix	✔	✔	✘	✘	✘	✘
Probe & mismatched enzyme	✔	✔	✘	✔	✘	✘

image, this is due to the drift of the carrier oil containing the droplets on the sample viewing chamber. The spatial pattern observed between the cell location mask and the fluorescent spots however remains consistent for all three samples. The data provided by this experiment indicates that the probe is becoming localised to the cell and the fluorescent spots are a result of fluorophore proximity and not on-target amplification. Cells appear to remain intact despite the thermocyling and the Phusion Direct reagents suggesting a more aggressive lysis solution is required to allow in-drop PCR to occur.

In order to gain a more thorough understanding of the extent of the cell lysis, an experiment was designed in which the GXR cell-line was stained with both a mitochondrial stain (surrogate for cell membrane lysis) and a nuclear stain prior to being exposed to a series of different lysis buffers. The goal of this experiment was to provide a clearer understanding of whether the cell membrane and nuclear envelope are being compromised during the different treatments. Each stain emits fluorescence on a different wavelength (mitochondria stain on green and nuclear stain on red); therefore, there are three possible outcomes from these experiments:

- **No lysis** - If both the cell membrane and the nuclear envelope are left intact during the treatment, both stains will remain localised to their target regions: the cytoplasm (mitochondria) and the nucleus.

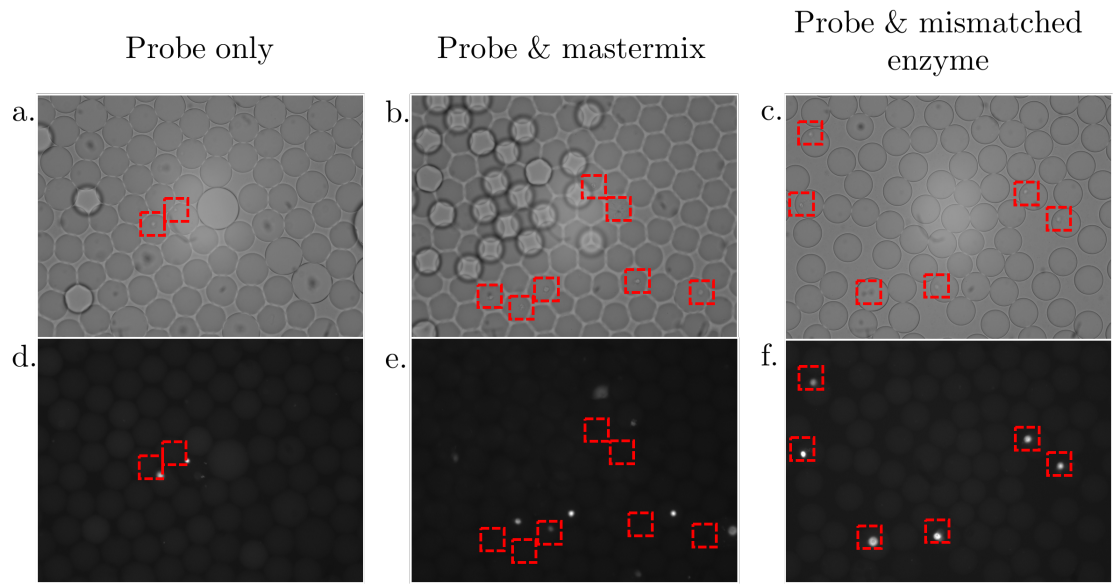


Figure 5.15: Localised fluorescence is not evidence of target amplification Paired, brightfield (light background, a-c) and fluorescent (dark background, d-f) droplet images from three different PCR incompatible albumin reaction mixes. Red hashed boxes in a-c indicate cells within droplets. Red hashed boxes in d-f are a spatial mask of the hashed boxes in a-c overlaid on the image to indicate colocalisation of fluorescent spots and cells. Bright spots in the fluorescent images indicate localised qPCR probe.

- **Partial lysis** - If only the cell membrane is lysed, the mitochondrial stain will be able to diffuse throughout the droplet resulting in a uniformly green droplet, however the nuclear stain remains trapped by the nuclear envelope and will hence stay localised to the nucleus.

- **Complete lysis** - If both the cell membrane and the nucleus are lysed, both stains will diffuse throughout the droplet resulting in uniform staining of both green and red upon imaging at their respective wavelengths.

The three potential outcomes are described pictorially in figure 5.16.

In addition to testing the lysis buffers ability to lyse cells, they were also tested for qPCR compatibility. The results of this experiment are detailed in table 5.6.

Surprisingly, complete cell lysis was difficult to achieve, with only a small subset of reagents capable of degrading both the cell membrane and the nuclear envelope. The reagents tested that were able to perfom complete cell lysis were:

- Proteinase-K and buffer AVL

Table 5.6: An overview of the reagents tested as part of the cell lysis experiments.

Lysis buffer	Concentration	Combined with	Cell membrane lysis	Nuclear lysis	qPCR compatible
RIPA	1x	N/A	✔	✔	✘
RIPA	0.5x	N/A	✔	✘	✘
Tween-20	4%	N/A	✘	✘	✔
Triton X-100	0.1%	N/A			
Buffer AVL	1x	✔	✔	✘	
Proteinase-K	1.5mg/ml	N/A	✔	✘	✘
Proteinase-K	1.5mg/ml	Sodium Dodecyl Sulfate	✔	✔	✘
Proteinase-K	1.5mg/ml	Buffer AVL (4%)	✔	✔	✘
Zygem	1x	N/A	✔	✘	✔
Igepal-CA630	0.5%	N/A	Partial	✘	✔
Igepal-CA630	1%	N/A	Partial	✘	✔
Igepal-CA630	4%	N/A	Partial	✘	✔
Phusion direct	1x	N/A	Untested	Untested	✔
Sodium Dodecyl Sulfate	0.5%	Tween-20 (4%)	✔	✘	Partial inhibition
Buffer RLT	1x	N/A	✔	✔	✘

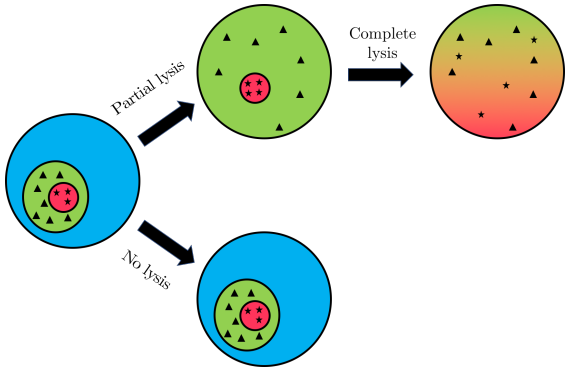


Figure 5.16: In-drop lysis outcomes. Schematic showing the three potential outcomes of the stained cell lysis experiment. Green = partial lysis, red = complete lysis, blue = no lysis.

- Proteinase-K and Sodium Dodecyl Sulfate
- Buffer RLT

Of these three reagent combinations, all caused complete PCR inhibition. As a result, an alternative solution was approached in order to perform PCR in droplets containing these reagents. Proteinase-K is a broad spectrum serine protease able to catalyse the cleavage of peptide bonds [229]. High temperatures and the presence of detergents increase the activity of Proteinase-K by exposing hydrophobic amino acid residues that would otherwise be protected within the core of the protein. This is supported by the fact that Proteinase-K alone is insufficient to cause complete lysis and that addition of a detergent such as Sodium Dodecyl Sulfate (SDS) initiates complete cell lysis. Buffer AVL and buffer RLT are two proprietary lysis mixtures supplied with Qiagen DNA extraction kits which also completely inhibit PCR, even at low concentrations (Fig 5.17). The proprietary nature of these products make them difficult to troubleshoot, therefore the Proteinase-K and SDS lysis buffer was progressed in order to improve its qPCR compatibility.

The optimum temperature for Proteinase-K is between 50 - 60℃. Above 70℃, Proteinase-K will quickly degrade and after ~10 minutes very little enzymatic activity will remain. This property of Proteinase-K provides a simple method to remove its inhibitory effect on PCR in droplets. By heat inactivating the enzyme at 95℃ for 5 minutes it is possible to remove nearly all enzymatic activity allowing

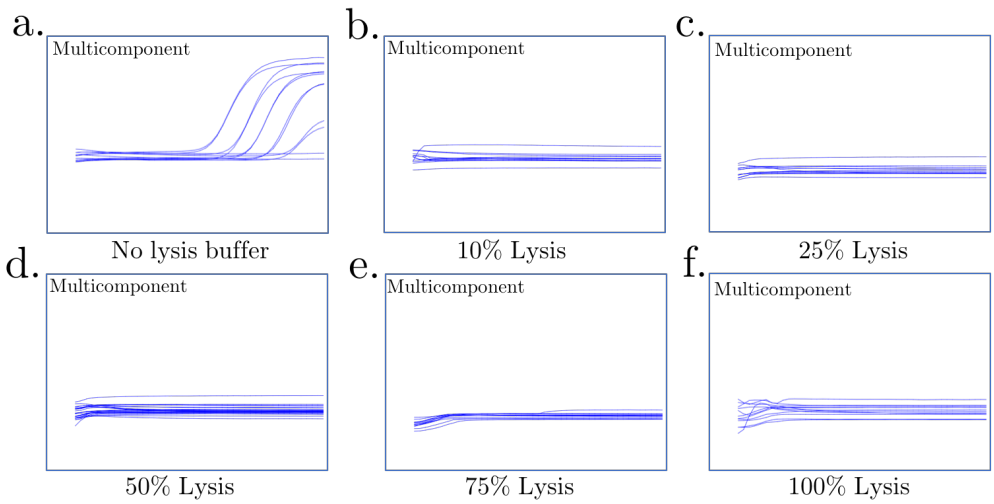


Figure 5.17: Qiagen lysis buffers completely inhibit PCR PCR multicomponent curves for increasing Qiagen lysis buffer concentrations (a) No lysis buffer, (b) 10% lysis buffer, (c) 25% lysis buffer, (d) 50% lysis buffer, (e) 75% lysis buffer, (f) 100% lysis buffer,

the DNA polymerase present in the PCR mix to retain its functionality throughout the PCR. SDS, which is to be added alongside Proteinase-K to aid cell lysis is also a known PCR inhibitor. To counter this inhibition it is possible to use a non-ionic detergent such as Tween-20, however at high concentrations these can also be inhibitory to PCR [230]. An alternative approach to removing the inhibitory effect of SDS is to dilute it within the qPCR reaction itself. By increasing the total volume whilst keeping the contribution from the cell lysate the same, it is possible to reduce the SDS concentration in the reaction. This concept of droplet dilution refers back to chapter 4.2.8 where two droplets can be successfully merged together; in this instance, a small droplet containing the cell lysate and lysis reagents and a second larger droplet containing the PCR mix that acts to dilute the lysis reagents. By combining a heat inactivation step of 95°C for Proteinase-K prior to merging, and a bulk dilution of SDS by droplet merging, it is possible to rescue succcessful qPCR amplification from cells lysed in-drop.

Figure 5.18.a-f shows the difference in cell lysis caused by a subset of different buffers. Figure 5.18.a& d show brightfield and fluorescent images respectively of cells that have been stained with Mitotracker Green and incubated in PBS. In this population, there is no evidence of cell lysis and as a result the stain remains

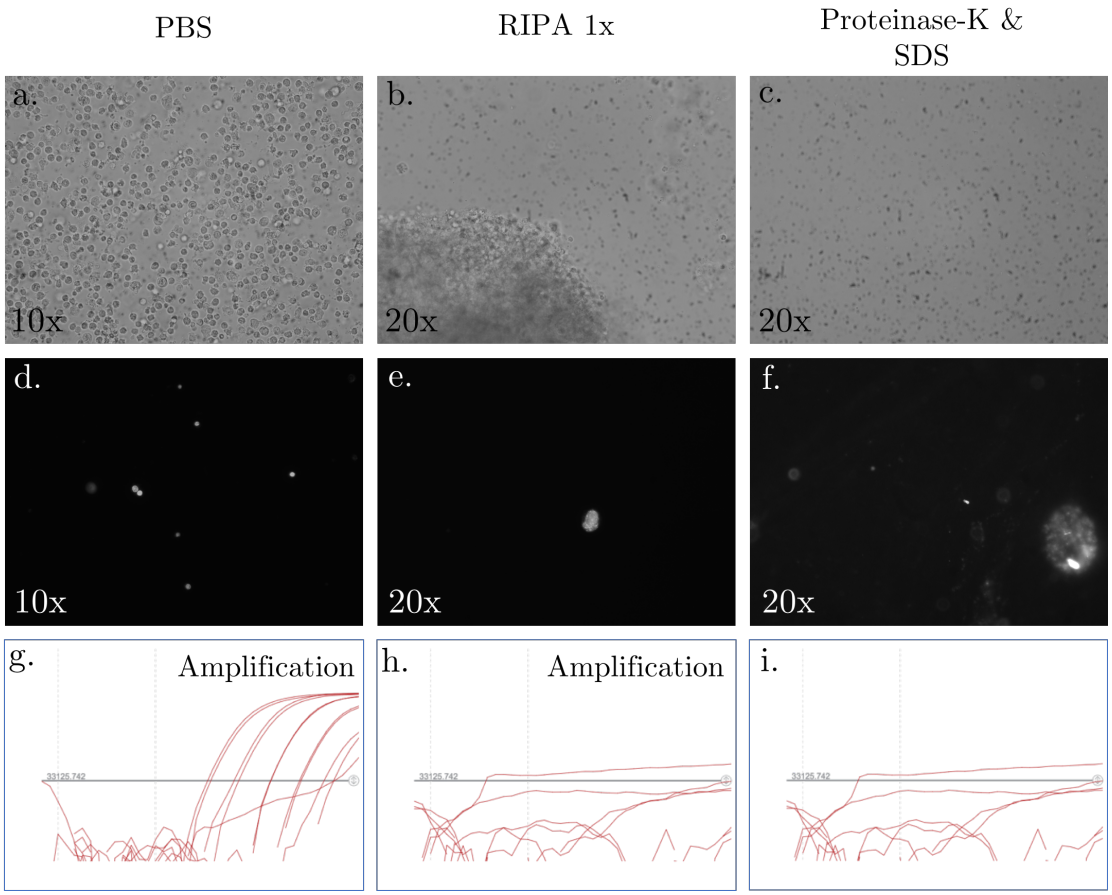


Figure 5.18: Lysis buffers able to lyse cells inhibit PCR (a-c) Brightfield images for cells suspended in PBS (a), 1x RIPA buffer(b) or Pro-K & SDS (c). (d-f) Fluorescent images showing mitochondrial stain for cells suspended in PBS (d), 1x RIPA buffer(e) or Pro-K & SDS (f). (g-i) qPCR amplification curves for standard qPCR reactins spiked with either PBS (g), 1x RIPA buffer(h) or Pro-K & SDS (i).

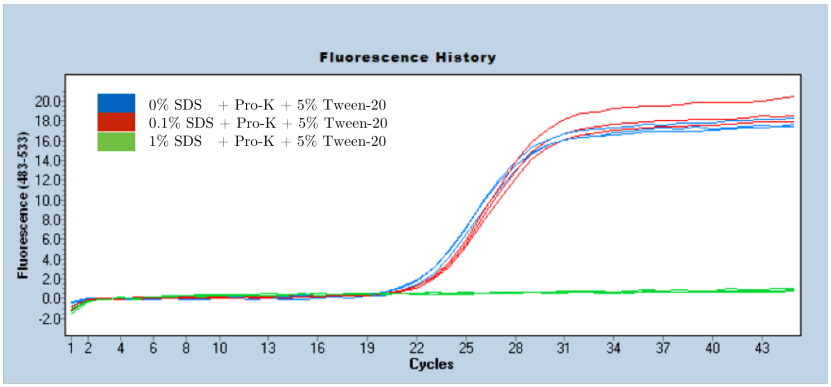
homogenous and colocalised with the cell coordinates. Figure 5.18.b & e show brightfield and fluorescent images respectively of cells that have been stained with Mitotracker Green and coencapsulated in droplets containing 1x RIPA buffer. In the brightfield image, there is an accumulation of cells in the lower left hand corner of the image where damage to the cell membrane has led to cell-cell adhesion and clumping. In the fluorescent image, one cell has been focussed on and cell staining appears more heterogenous when compared to the PBS control suggesting a level of membrane damage has occurred. Figure 5.18.c & f show brightfield and fluorescent images respectively of cells that have been stained with Mitotracker Green and coencapsulated in droplets containing Proteinase-K (1.5mg/ml) & 0.5% SDS. In

the brightfield image, no evidence of cellular structures can be seen and in the fluorescent image, the mitochondrial stain has spread beyond the size limit of a cell and is heterogenous in nature. These data indicate that RIPA has a limted effect on cell integrity when compared to the PBS control however a cocktail of Proteinase-K & SDS causes efficient cell lysis. Figure 5.18.g-i shows representative amplification plots of a standard curve assay targeting albumin performed in the presence of (g) PBS, (h) 1x RIPA buffer and (i) 1.5mg/ml Proteinase-K & 0.5% SDS. As is evident from the figure, standard PCR amplification curves fail to manifest in the presence of both RIPA buffer and the Proteinase-K and SDS cocktail when compared to the PBS control. This amplification efficiency can be recovered if the Proteinase-K is inactivated by heating the droplet prior to the PCR mix being added and if the SDS is either diluted to a level where it is no longer inhibitory or if it is neutralised by the addition of Tween-20. Figure 5.19.a shows the effect of adding either SDS, heat-inactivated Pro-K, Tween-20, or all three to a qPCR reaction. Here, all reaction mixes contained 5% Tween-20 in order to determine its effect on amplification efficiency. When added with heat-inactivated Pro-K, amplification performs well with an exponential curve present and a signal of ~17Fu. When added with heat-inactivated Pro-K and 0.1% SDS, the observed amplification efficiency is similar with an exponential phase present and a signal of ~17Fu. When 5% Tween-20 is added with heat-inactivated Pro-K and 1% SDS however, no amplification is observed. This data suggests that 5% Tween-20 is insufficient to counter the inhibitory effect of 1% SDS and leads to PCR failure.

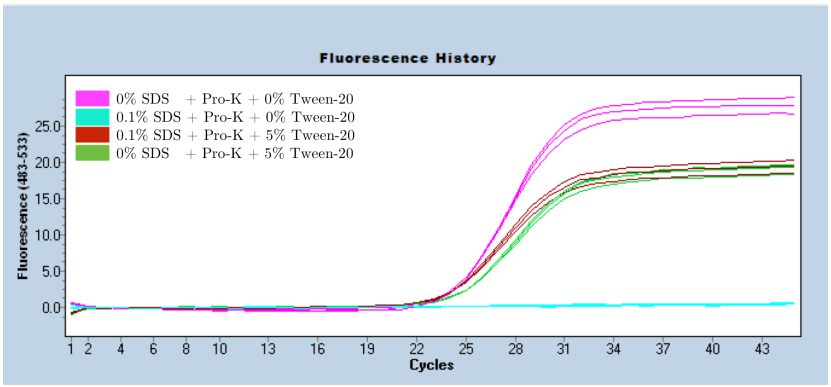
In order to further elucidate the relationship between SDS, Tween-20 and successful PCR amplification, a repeat experiment was performed with additional controls. The four conditions tested and their subsequent effect on PCR amplification are detailed in table 5.7. The graphical output of the amplification curves can be seen in 5.19.b.

This second experiment clarifies the relationship between Tween-20, SDS and PCR amplification. When added in isolation, both SDS and Tween-20 inhibit PCR, although the extent to which they do so varies. Tween-20 causes a decrease

a.



b.



c.

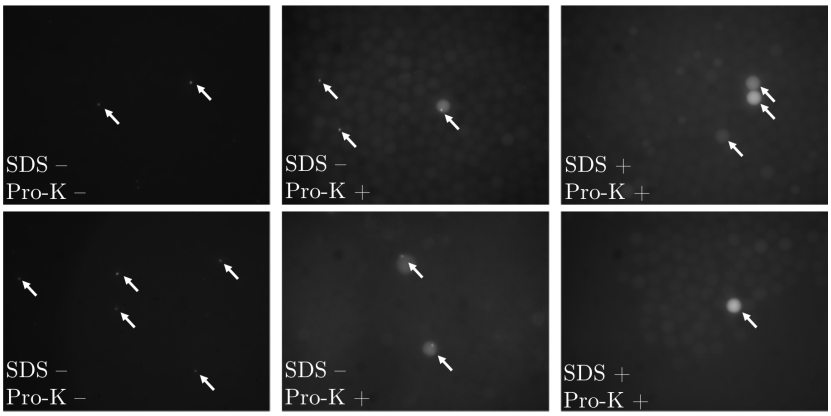


Figure 5.19: SDS PCR inhibition can be partially rescued by the addition of Tween-20 (a) Amplification plot for three different Pro-K/Tween mixture experiments. (b) Amplification plot for four confirmatory Pro-K/Tween mixture experiments.(c) Representative fluorescent images of the mitochondrial stain after incubation with either: PBS only, SDS^{-}/$Pro\text{-}K^{+}$, or SDS^{+}/$Pro\text{-}K^{+}$. White arrows indicate droplets containing cells and two images are shown per treatment.

Table 5.7: SDS and Tween-20 have an inhibitory effect on qPCR amplification.

Condition	SDS (0.5%)	Pro-K (0.5mg/ml)	Tween-20 (%)	Exponential curve	Fluorescent signal gain
Pro-K only	✘	✔	✘	✔	27
Pro-K & SDS	✔	✔	✘	✔	0
Pro-K, SDS & Tween	✔	✔	✔	✔	18
Pro-K & Tween	✔	✔	✔	✔	18

in the slope of the amplification curve and a reduction in the fluorescent signal gain (18Fu compared with 27Fu for the no-Tween control). This level of inhibition is not terminal for Tween-20's inclusion in this assay however, as discussed previously, a binary negative/positive signal is all that is required. Nevertheless, one concerning factor is that this reduced fluorecent signal in the bulk reaction will likely translate to the droplet assay where the strength of the positive droplet's fluorescence is key to enabling accurate sorting.

SDS alone however completely blocks the PCR amplification and no increase in fluorescence is detected. This amplification block is preventable by the addition of Tween-20, this in turn will lead to a reduced fluorescent signal due to Tween's previously described inhibitory effect on PCR. Also of note, the observed fluorescent signal increase and the shape of the amplification curve for both the Tween only and the Tween with SDS conditions were near identical suggesting that Tween completely recovers the SDS inhibition and it is only the Tween inhibition that leads to the decrease in signal intensity when both are added. The ability of Tween-20 to rescue PCR amplification from SDS inhibition has only been shown here to be effective at concentrations of SDS $\leq 0.1\%$. As a result of the observed Tween-based PCR inhibition, it is preferable to prevent SDS inhibition via droplet

merging and subsequent dilution of the reagent. In instances where this is not practicable, Tween is an imperfect yet plausible solution to the problem.

5.3.1 PCR reagent optimisation

At this stage of development, efficient in-drop lysis of cells was being achieved. It was not possible however, to detect whether PCR amplification was occurring using only droplet fluorescence as a marker of success. Experiments performed previously, where thermocycled PCR mix had been encapsulated within droplets and subsequently imaged to determine the difference between a negative droplet and a positive droplet, suggested there was a ~0.5x increase in signal intensity produced as result of successful amplification. Imaging droplets using a microscope has an inherent time cost associated with it; as a result of the positioning of the microscope stage, focussing on the sample and integrating enough light for a clean image, samples can be exposed to the light source for >10 seconds in some circumstances. The fluorophores used in this pipeline are not resistant to photobleaching and as such, a large reduction in fluorescence can be seen if the light source is kept on the sample for an extended period of time. All measures of PCR success have thus far been performed via microscope imaging, the photobleaching of the samples described above makes it difficult to reproducibly determine whether there is in fact an increase in fluorescence between two different droplet populations. To improve the signal to noise for this assay, and subsequently reduce potential bias in the analysis, a series of PCR reagent optimisations were tested (fig 5.20).

Probe concentration, enzyme concentration and magnesium concentration were all tested to determine optimum levels for in-drop PCR.

- **Probe concentration** A range of probe concentrations were tested between 200nM and 1400nM. For this assay, the 200nM exhibited the greatest increase in signal from the baseline of 18Fu. A concentration dependent increase was seen for the 600nM and 1000nM concentrations (16Fu and 11Fu respectively) with the top concentration, 1400nM causing total failure of the qPCR reaction 5.20.a.

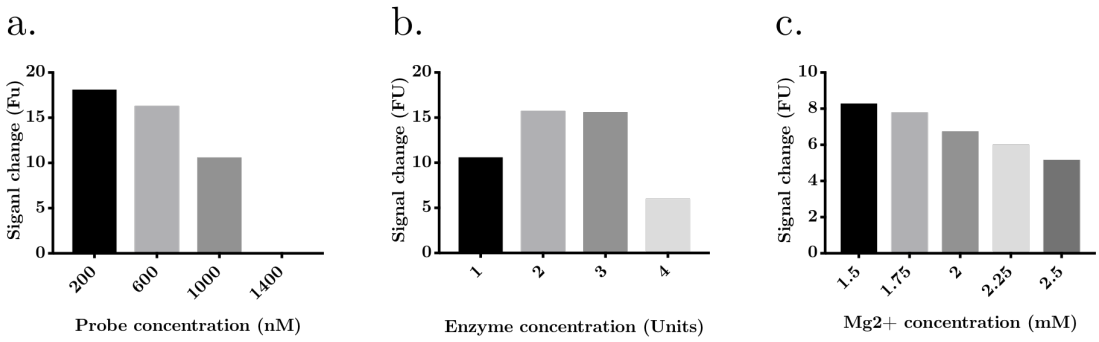


Figure 5.20: Optimisation of in-drop qPCR reagents (a) Barplot showing fluorescence signal increase observed using different probe concentrations (b) Barplot showing fluorescence signal increase observed using different enzyme concentrations (c) Barplot showing fluorescence signal increase observed using different $Mg2^{+}$ concentrations

- **Enzyme concentration** A range of enzyme concentrations were tested between 1 and 4 units. For this assay, both the 2unit & 3unit concentrations exhibited the greatest increase in signal from the baseline (15Fu) with the 1 and 4 unit concentrations underperforming in comparison (10Fu and 6Fu respectively).

- **Magnesium concentration** A range of total magnesium concentrations were tested between 1.5mM and 2.5mM. For this assay, the 1.5mM concentration exhibited the greatest increase in signal with a concentration dependent decrease observed for all higher concentrations.

Based on the data reported above, the reagent concentrations detailed in table 5.8 were used for all future experiments unless specified otherwise.

Table 5.8: Mismatched primer/probe experiment design

Reagent	Concentration
Probe	200mM
Enzyme	2 units
$Mg2^{+}$	1.5mM

The optimised reaction was subsequently tested within droplets to determine if the fluorescent signal intensity would be improved in-drop. This was achieved by

co-encapsulating PCR mix and DNA template from the 8e5 cell line into WOs. The newly formed WOs were thermocycled to complete the PCR reaction and finally imaged by microscopy to assess droplet fluorescence.

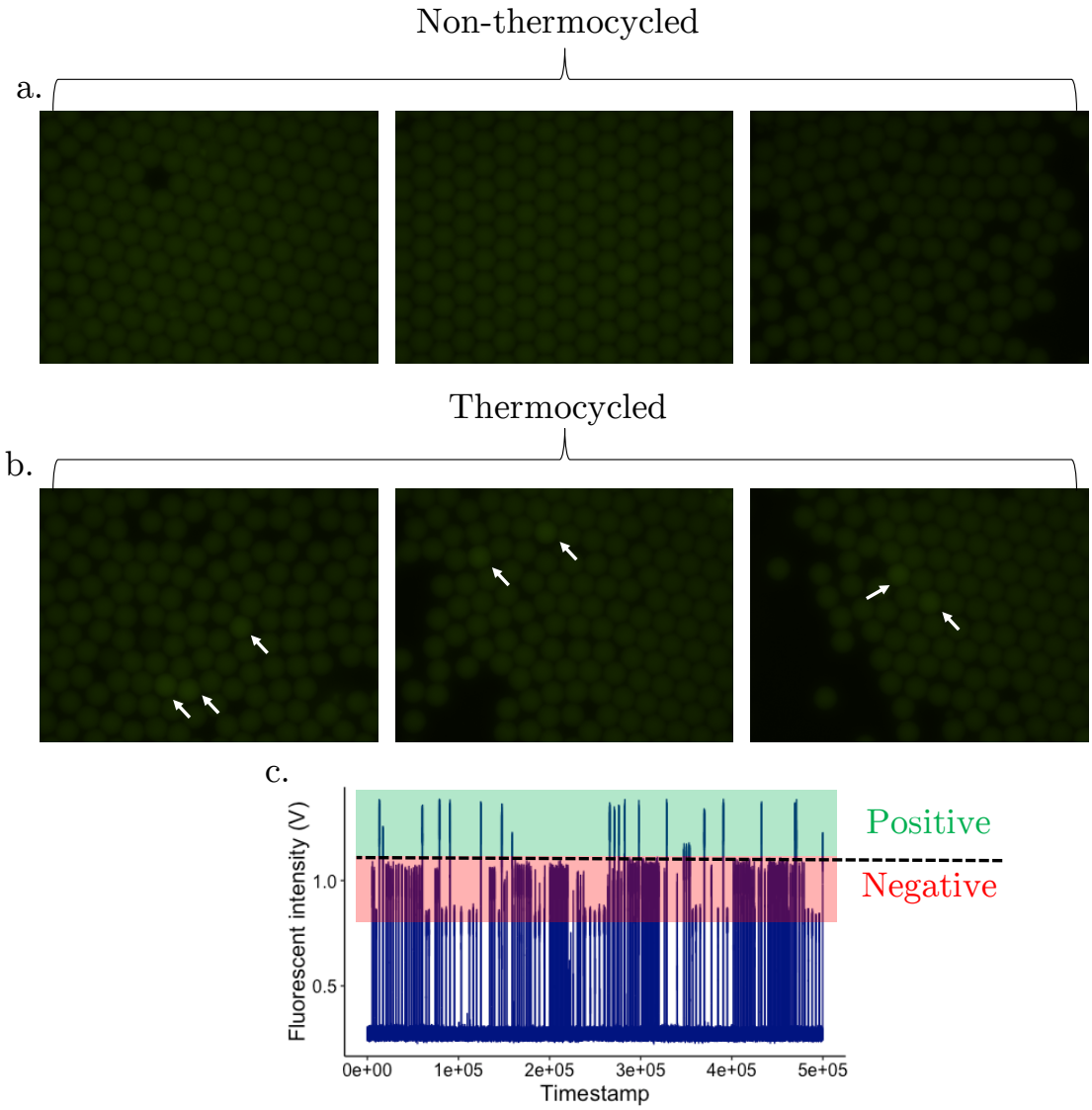


Figure 5.21: Fluorescence results from a positive population of albumin droplets as measured by a PMT. (a) Three representative images of DNA+ non-thermocycled droplets taken at 20x magnification on the FITC channel. (b) Three representative images of DNA+ thermocycled droplets taken at 20x magnification on the FITC channel (White arrows indicate droplets with a higher level of fluorescence). (c) Line plot showing the fluorescent intensities of the positive droplet population where each individual spike represents a droplet. Spikes that peak in the green shaded area are considered PCR+ and those that peak in the red shaded area are considered PCR-. Black dashed line indicates the positive threshold.

The optimised reaction did not have a significant effect on the signal increase

between the negative and positive droplets as can be observed in figure 5.21. Figure 5.21.a shows droplets that have PCR mix and DNA co-encapsulated however have not been thermocycled so should theoretically have no amplification and consequently no increase in fluorescence above background. Droplets in these three representative images appear uniform in fluorescent intensity as expected. The three images in figure 5.21.b show the same population of droplets post-thermocycling. At this stage, PCR amplification has occurred in droplets containing a fragment of target albumin DNA and consequently an increase in fluorescence is observed. In these three images a number of albumin-positive droplets can be seen to have a modest increase in fluorescence above the background level of the albumin-negative droplets. Figure 5.21.c shows the output from the PMT after assessing the fluorescent intensity of the thermocycled population. Spikes on this plot that fall within the red shaded area are droplets that do not contain a copy of albumin and therefore show no increase in fluorescent intensity. Spikes that peak within the green shaded region indicate droplets that contain a copy of albumin and therefore have successfully amplified and unquenched the fluorescent probe leading to an increase in fluorescence.

Despite the optimisation of the qPCR reaction described above in chapter 5.3.1, the difference in fluorescence observed between the negative and positive droplets is minimal, with the positive signal on average 1.5x that of the negative. A signal this small makes it difficult to accurately threshold the droplets as there can be some intra-population variation of fluorescent intensity. In these instances, a droplet towards the higher end of the negative fluorescence range can be difficult to separate from those towards the lower end of the positive fluorescence range. This concept is similar to the ddPCR 'rain' phenomenon discussed earlier in chapter 1.5.2.1 and may be due to a reduction in priming efficiency of the PCR due to nucleotide substitutions in the primer binding sites [140]. A major contributor to the lack of signal is the level of autofluorescence caused by the uncleaved qPCR probe that is present in the negative droplets. In an attempt to reduce the intensity of autofluorescence emitted by the probe, a different chemistry was tested including

two on-probe quenchers as opposed to one. The two different chemistries are described in figure 5.22.

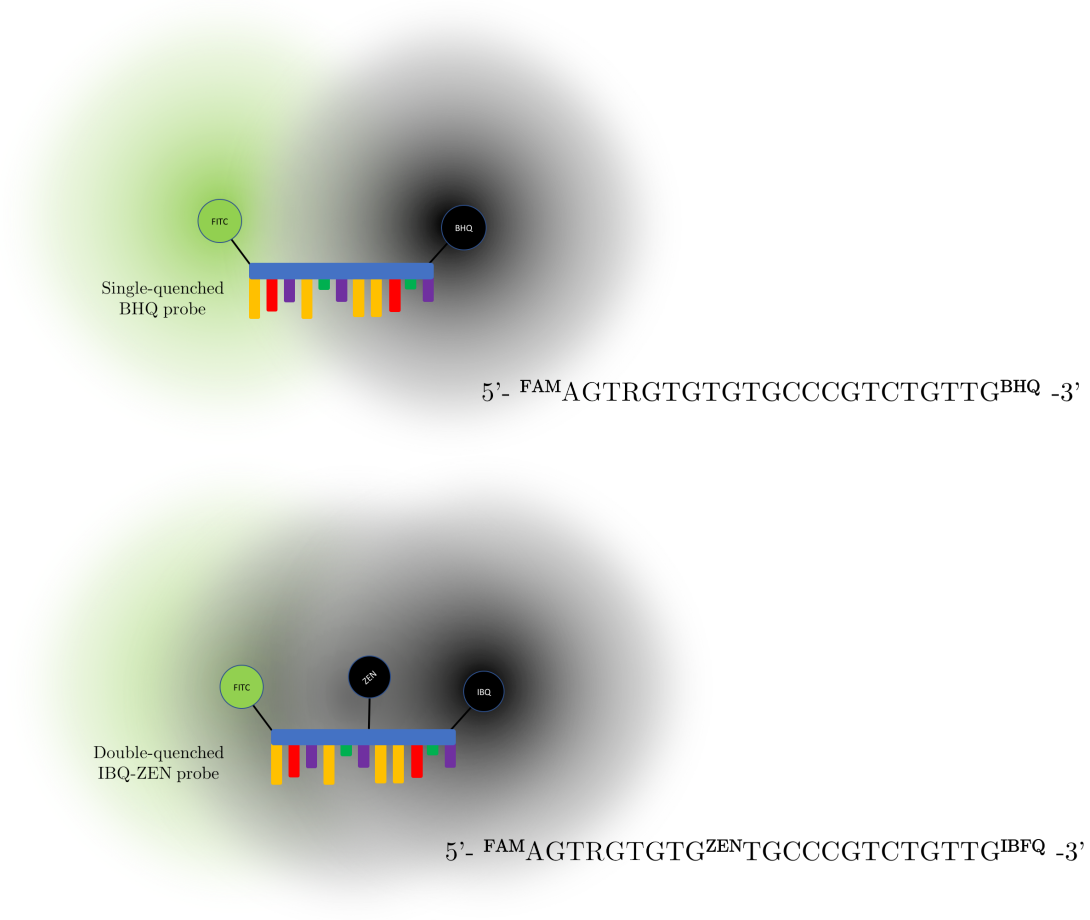


Figure 5.22: Comparison between single-quenched and double-quenched probes Schematic representation of how quenchers work on TaqMan probes. Green circles indicate fluorophores, black circles indicate quenchers. The transparency of each colour symbolises the strength of the effect (more transparent = weaker).

The single-quenched (black hole quencher (BHQ)) probe consists of a FITC fluorophore reporter on the 5' end of a linear nucleotide string and a BHQ on the 3' end. The double-quenched (Iowa Black Fluorescein Quencher (IBFQ) & ZEN quencher) probe consists of a FITC fluorophore reporter on the 5' end of the nucleotide string, an IBFQ on the 3' end and a ZEN quencher in the middle. The ability of a quencher to dissipate energy from the fluorophore diminishes as distance between the two is increased [231]. On a linear probe therefore, a quencher synthesised in the middle of a nucleotide string has a greater quenching

potential than one that is synthesised at the opposite end of the probe on the 3' end. Applying this knowledge to the droplet system being tested here, the double quenched probe should produce a reduced baseline level of fluorescence in the negative droplet population. In addition, as the FITC fluorophore reporter is similar between the two probes, the positive signal once cleaved should be comparable to the single quenched probe and therefore have a greater signal-background ratio.

The performance of the double-quenched probe was tested in-drop by co-encapsulating PCR mix and DNA from the 8e5 cell line in WOs. For this test, the single-quenched probe was replaced with the double-quenched probe to assess whether there was an increase in the signal generated in the positive population.

A reduction in fluorescence was observed for all droplets in the non-thermocycled control population and all negative droplets in the thermocycled positive population. The fluorescent intensity of the positive population remained consistent with the values observed with the single quenched probe suggesting that the double quenched probe increases signal by reducing the level of droplet autofluorescence as hypothesised. The signal generated by the positive droplets when using the double-quenched probe is ~2x that of the negative signal, this increase enables positive droplets to be clearly identified and allows for aggressive sorting thresholds to be set in order to minimise the number of false positives collected. Based on the data presented above, the double-quenched probe was used for all subsequent experiments unless otherwise specified.

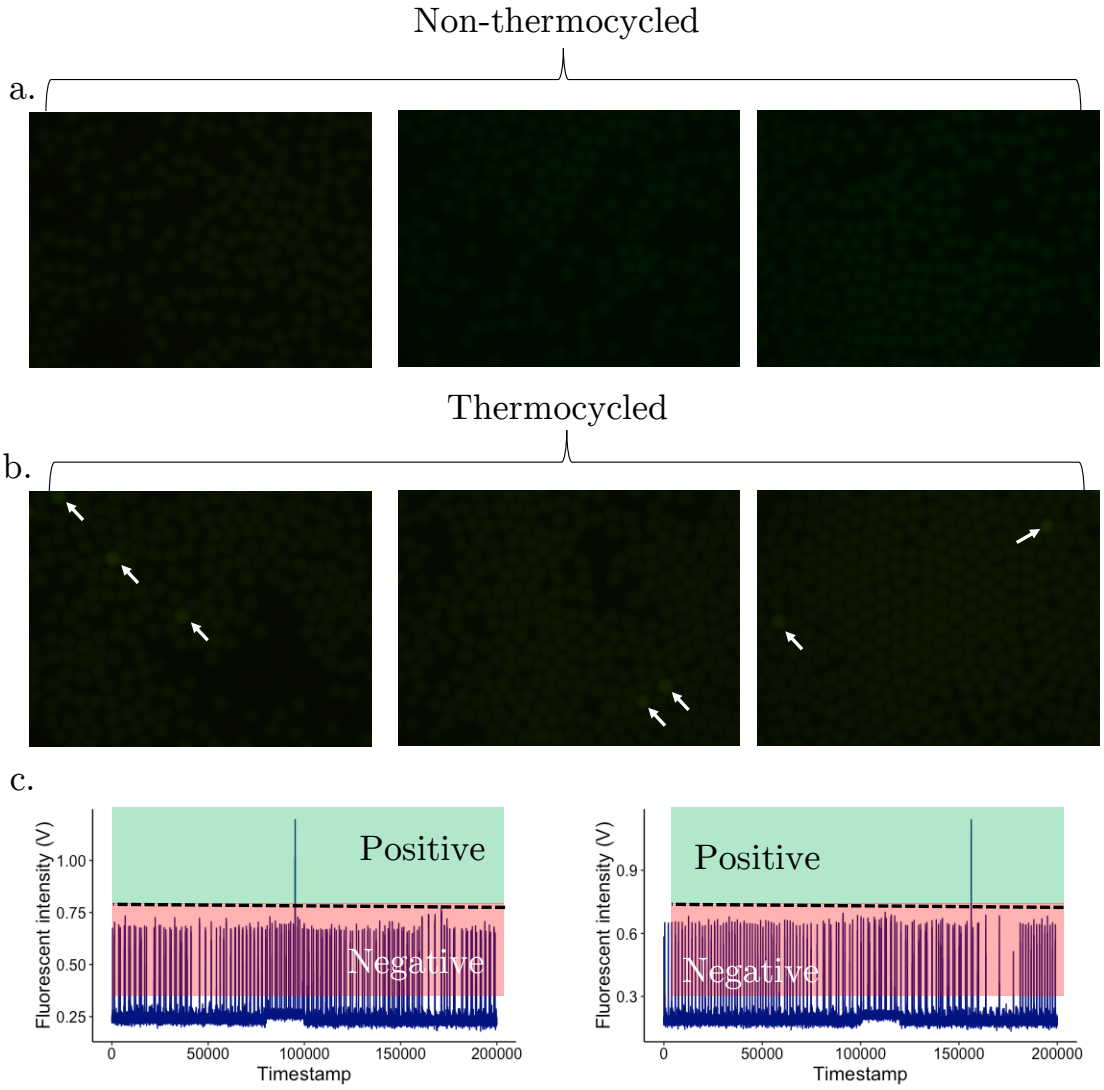


Figure 5.23: The double-quenched probe provides a greater fluorescent signal whe compared with te single-quenched probe (a) Three representative images of DNA+ non-thermocycled droplets with double-quenched probe taken at 20x magnification on the FITC channel. (b) Three representative images of DNA+ thermocycled droplets with double-quenched probe taken at 20x magnification on the FITC channel (White arrows indicate droplets with a higher level of fluorescence). (c) Line plot showing the fluorescent intensities of the positive droplet population where each individual spike represents a droplet. Spikes that peak in the green shaded area are considered PCR+ and those that peak in the red shaded area are considered PCR-. Black dashed line indicates the positive threshold.

Chapter 6

Droplet microfluidics: Results

Contents

6.1 Context 174
6.2 DNA enrichment pipeline 175
6.2.1 DNA encapsulation 175
6.2.2 DNA droplet thermocycling 176
6.2.3 DNA droplet sorting 178
6.2.4 Nucleic acid quantitation. 178
6.2.5 DNA shearing leads to proviruses being split between droplets. 182
6.2.6 DNA enrichment pipeline discussion 184
6.3 Cell enrichment pipeline testing 185
6.3.1 Testing the full cell enrichment pipeline using uninfected PBMCs. 185
6.3.2 Cell enrichment pipeline discussion 187
6.4 Conclusion 187

6.1 Context

Two distinct, yet closely related pipelines are being developed for use towards isolating, characterising and quantifying the latent reservoir in HIV infection, this chapter aims to provide a current snapshot of the droplet microfluidics branches of these two ongoing projects. Due to the complexity of the pipelines, they are not yet ready to process clinical samples. A simplified version of the two pipelines are detailed in figure 6.1 and indicate which parts of the pipeline are yet to be fully validated.

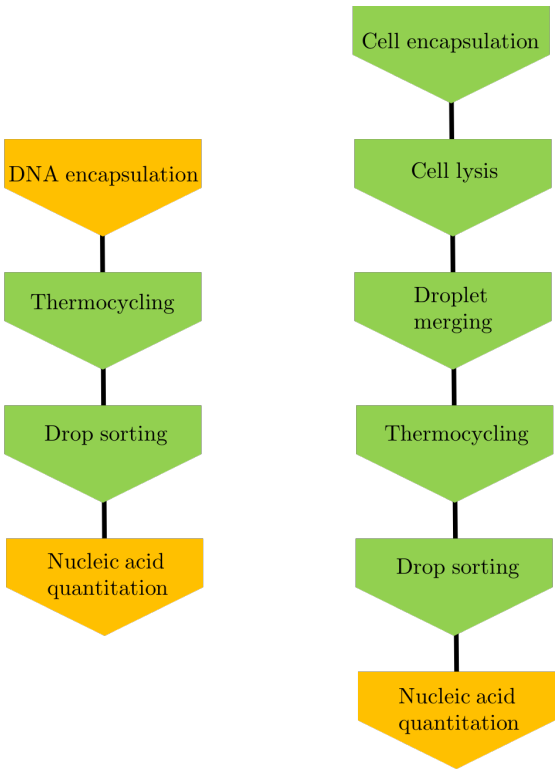


Figure 6.1: The droplet pipelines. Flow diagrams showing the various stages of the DNA droplet enrichment protocol (left) and the cell droplet enrichment protocol (right). Stages coloured green indicate a validated stage. Stages coloured orange indicate that further work is needed before processing clinical samples.

The remainder of this chapter will address each stage in turn using results obtained from cell-line experiments.

6.2 DNA enrichment pipeline

6.2.1 DNA encapsulation

DNA encapsulation involves resuspending extracted DNA in an aqueous medium (for compatibility with downstream processing) and forming WOs using the resuspended DNA as the dispersed phase. It was hypothesised that similar conditions to those used to form pure water droplets would be sufficient to form DNA droplets. Upon testing however this was not the case. As shown in figure 6.2.b, attempting to form droplets using high concentrations of DNA leads to jetting of the aqueous stream and unstable droplet production. This effect was recovered when a lower concentration of the same DNA sample was used to form the WOs (fig 6.2.a).

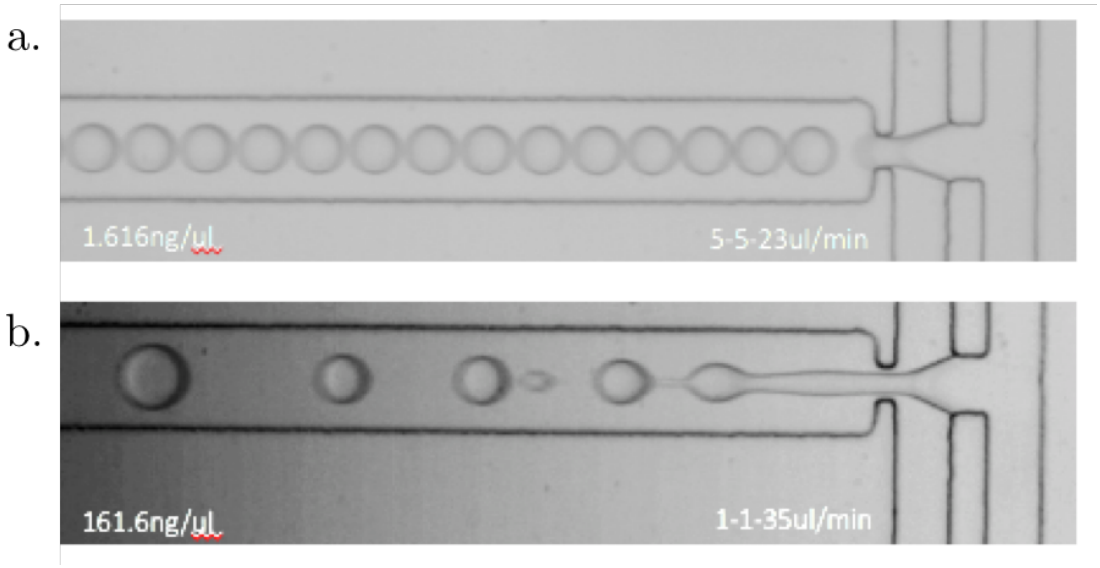


Figure 6.2: High DNA concentrations lead to jetting during WO production. (a) WO production using a low concentration of DNA (1.616 ng/µl). (b) WO production using a high concentration of DNA (161.6 ng/µl). Device flow rates during imaging are shown in the bottom right of each image in the format: aqueous inlet1 - aqueous inlet 2 - oil inlet.

The typical compensatory response to jetting of the aqueous phase is to either:

- Increase the oil (continuous) phase flow rate
- Decrease the aqueous (dispersed) phase flow rate

- Increase the oil (continuous) phase flow rate and decrease the aqueous (dispersed) phase flow rate

In the example shown, both the continuous and dispersed phase flow rate were modified to attempt to stabilise droplet production. For the lower DNA concentration tested the continuous phase flow rate used was 2.3x the dispersed phase flow rate. Whilst attempting to recover stable droplet production, a continuous phase flow rate 17.5x the dispersed phase flow rate was used with no observed increase in stability (fig 6.2.b). The highest concentration of resuspended DNA that has been successfully dropletised thus far is 8 ng/µl. This is the equivalent of encapsulating one cell's worth of DNA in every six droplets produced.

6.2.2 DNA droplet thermocycling

DNA droplet thermocycling is necessary to allow the PCR reaction to occur. Post-thermocycling, droplets containing a successful amplification are labelled as positive for the target of interest by the fluorophore attached to the PCR probe losing its quenchers and emitting fluorescence. Droplet thermocycling is dependent upon the thermostability of the droplets. A series of optimisations described in chapter 4 have resulted in minimal coalescence occuring during thermocycling. Table 6.1 shows droplet population statistics from a single set of droplets analysed both pre and post-thermocycling.

Table 6.1: Droplet thermostability

	Mean radius (µm)	Radius standard deviation	Number of droplets analysed
Pre-thermocycling	31.60	5.86	2420
Post-thermocycling	31.60	6.94	3218

The droplet size statistics reported in table 6.1 confirm there is no difference between the droplets pre and post-thermocycling (p-value = 0.989). Plots 6.3.a & b show the distribution of droplet size for the pre and post-thermocyling populations respectively.

The second critical component of the thermcycling stage of the pipeline is the ability to identify latently infected cells. Figure 6.3.c contains four fields of view from a thermocycled HIV in-drop PCR. Droplets containing target fragments of DNA are easily identifiable by their FITC^{HIGH} appearance.

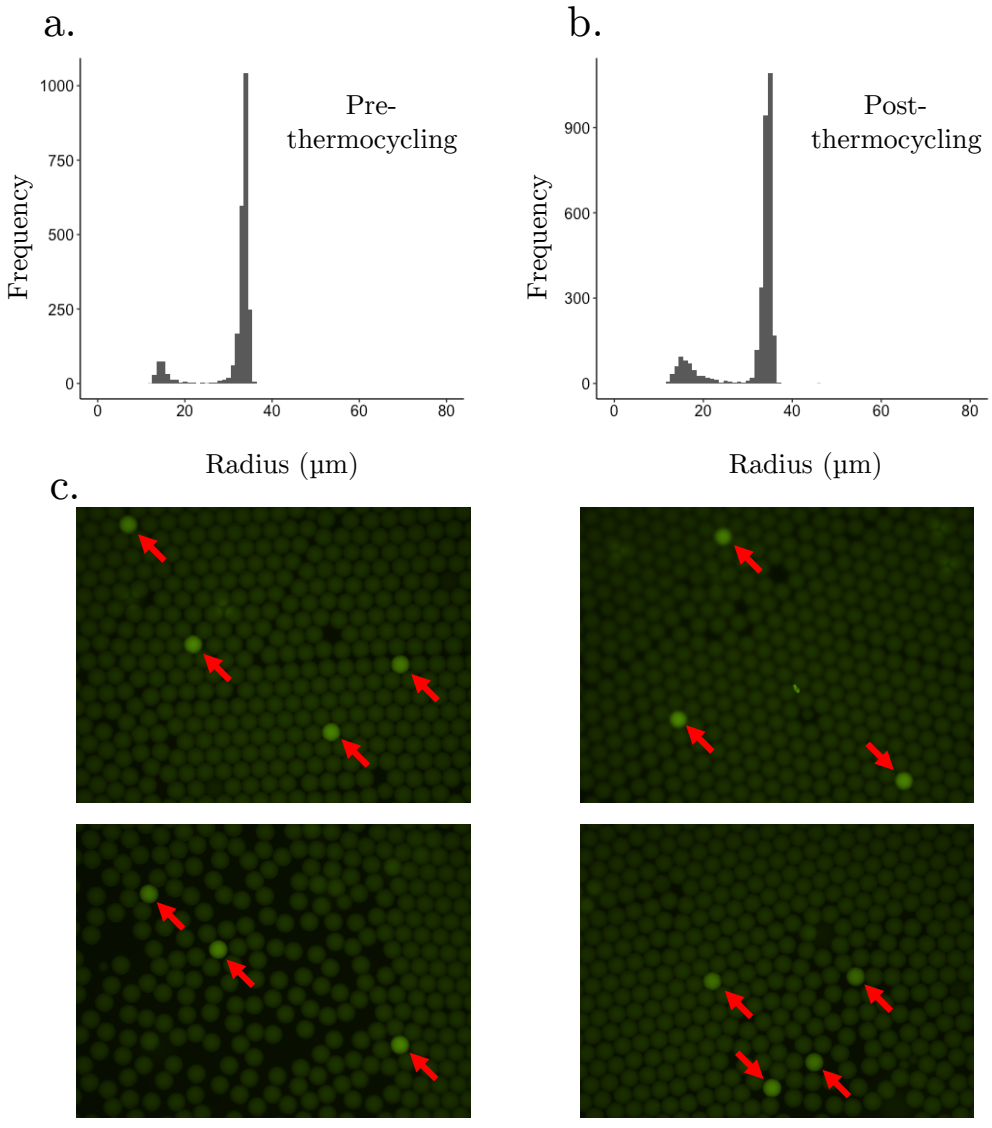


Figure 6.3: PCR can be performed within thermostable droplets. (a) Barplot showing the distribution of droplet size post-thermocycling. (b) Barplot showing the distribution of droplet size pre-thermocycling. (c) Four fields of view [10x] showing the droplet population post-thermocycling viewed under the FITC channel. Red arrows indicate PCR positive droplets.

6.2.3 DNA droplet sorting

Droplet sorting is required to isolate the PCR positive FITC^{HIGH} droplets from the PCR negative FITC^{LOW} droplets thereby enriching the sample for HIV proviral DNA. Sorting success is measured using live imaging of the microfluidic sorting junction. Prior to each individual droplet sorting run, a small subset of droplets is used to optimise the sorting conditions. These conditions can be influenced by:

- Droplet size
- Device wetting
- Channel imperfections unique to individual devices
- DNA concentration

Sucessful sorting is shown in figure 6.4 where a positive droplet is attracted to the sort channel via dielectrophoresis.

Following on-chip verification that sorting is working correctly at the beginning of each run, devices can be left unsupervised whilst sorting.

6.2.4 Nucleic acid quantitation.

In order to retrieve the cell lysate from within the droplet it is necsssary to break the emulsion. This process involves adding perfluorooctanol which acts to displace the surfactant from the water-oil interface. Droplet coalescence is followed by the separation of the aqueous and oil phases by centrifugation where nucleic acids will remain in the aqueous phase.

Quantitation of the recovered nucleic acids can be achieved by either double-stranded DNA staining or qPCR. Table 6.2 shows the DNA recovery for three independent experiments as mesaured by double-stranded DNA staining on the life Technologies Qubit instrument using the HS (High Sensitivity) DNA kit.

Three experiments were conducted encapsulating a DNA mix of 6.4 ng/µl and sorting based on the FITC status of the droplet. FITC^{HIGH} droplets were considered positive for the target fragment and FITC^{LOW} droplets were considered

htbp!

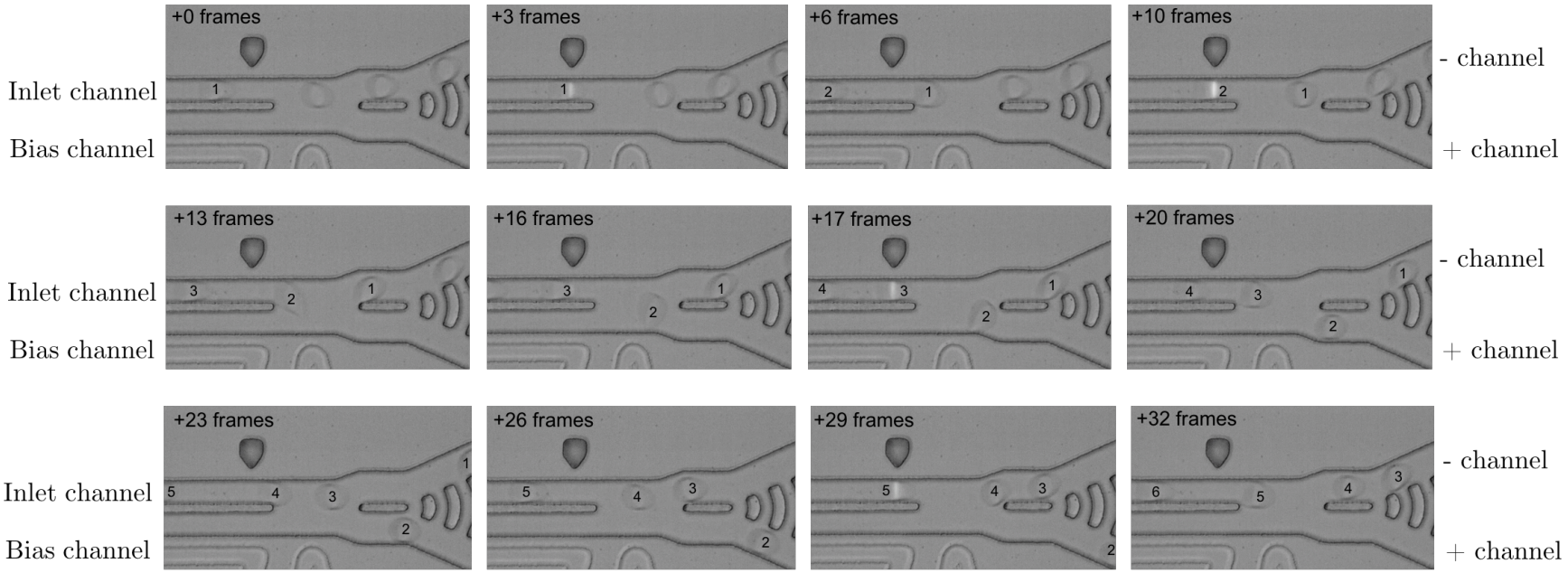


Figure 6.4: Still images of droplet sorting. A series of still images showing successful droplet sorting. Individual droplets are labelled with numbers of which droplet 2 is FITC^{HIGH} and droplets 1,3,4,5 & 6 are FITC^{LOW}. Timestamps are indicated by the frame count in the upper left corner of each still.

Table 6.2: Nucleic acid recovery

Channel	Experiment 1 (pg/µl)	Experiment 2 (pg/µl)	Experiment 3 (pg/µl)
Positive	Out of range	Out of range	Out of range
Negative	82	84	2000

negative for the target fragment. Emulsion breaking for experiments 1 and 2 was performed on equivalent numbers of droplets and experiment 3 was performed on ~2x the amount. Due to the low frequency of positive droplets in these enrichment assays (an expected ratio of 1:150 should be positive), considerably more were collected in the negative channel than the positive. DNA recovered from the negative channel, albeit at low concentrations, was consistently within the quantifiable range of the Qubit HS assay (10 pg/µl - 100 ng/µl) for all three experiments (0.082, 0.084 and 2.000 ng/µl respectively, table 6.2). DNA recovered from the positive channel however, was below the limit of quantitation of the assay for all three experiments. To improve the chances of detection, DNA from the first two experiments was concentrated by centrifugation and the HS DNA assay was repeated. Samples were concentrated ~10x compared to the original sample; both samples from the positive channel remained below the limit of quantitation for the HS DNA assay (table 6.3).

Table 6.3: Nucleic acid recovery after concentration of the DNA

Channel	Experiment 1 (pg/µl)	Experiment 2 (pg/µl)
Positive	Out of range	Out of range
Negative	322	288

Due to the greater sensitivity of qPCR when compared to the Qubit DNA stain, qPCR assays targeting albumin and the the HIV pol region were also performed on these samples. Both albumin and pol, in a ratio similar to the starting non-enriched sample were detected in DNA extracted from the negative channel. Neither albumin nor pol were detected in DNA extracted from the positive channel. If sorting had been successful, HIV pol should either: not have been present or

been present at extremely low levels in the negative channel DNA. A number of modifications were made to the emulsion breaking protocol at this stage in an attempt to improve droplet DNA recovery from the low frequency of positive droplets. These modifications included:

1. **Adding collected droplets to a buffer of water droplets** to reduce DNA loss to plasticware.

2. **Adding collected droplets to a buffer of BSA droplets** to reduce DNA loss to plasticware.

3. **Flushing the droplet transfer tubing with oil** to ensure all droplets were transferred to the collection eppendorf tube.

4. **Purifying DNA** with column and bead based protocols to remove potential PCR inhibitors carried over from the emulsion breaking reagents.

These modifications had minimal effects on the yield recovered from the samples. Specifically:

1. **Adding water droplets as a buffer** did not influence the yield of DNA or qPCR quantitation.

2. **Adding BSA droplets as a buffer** led to a small decrease in yield of DNA.

3. **Flushing the droplet transfer tubing** did not influence the yield of DNA or qPCR quantitation.

4. **Purifying DNA** led to an increase in the efficiency and maximum fluorescence of the qPCR reaction but at the cost of a small increase in yield due to inevitable losses through the purification steps.

6.2.5 DNA shearing leads to proviruses being split between droplets.

One potential explanation for the absence of pol HIV amplification in the positive channels and its presence in the negative channel is the possibility that the provirus has been sheared during upstream processing. DNA samples up to this stage had been extracted using either Qiagen DNA Mini or Qiagen DNA Midi columns. Qiagen report that their columns extract DNA with an expected fragment range between 25 and 100kilobases [Personal communication]. DNA fragments towards the lower end of this range (50Kb) have a ~20% chance of shearing in the middle of the 10Kb HIV genome. The qPCR assay used to identify droplets as containing HIV DNA targets the viral LTR which is located at the extremities of the genome. To ensure the LTR probe assay was successfully selecting for all droplets containing viral fragments, a multiplexed qPCR assay was utilised. To this end, the previously described LTR assay on the FITC channel was used in conjunction with the pol assay on the HEX channel. In this test, droplets can fit one of four different colour profiles as shown in table 6.4 and figure 6.5.

Table 6.4: Potential droplet types dependent upon provirus shearing.

Colour profile	HIV LTR present	HIV pol present	Classification
$FITC^{DIM}/HEX^{DIM}$	No	No	No viral DNA
$FITC^{HIGH}/HEX^{DIM}$	Yes	No	Sheared viral DNA
$FITC^{DIM}/HEX^{HIGH}$	No	Yes	Sheared viral DNA
$FITC^{HIGH}/HEX^{HIGH}$	Yes	Yes	Intact viral DNA

Droplets were deemed to be either 'intact', 'sheared' or 'no provirus' based on the criteria set out in figure 6.5. For the Qiagen column-sheared DNA used in this experiment, only 40% of droplets that were positive for one HIV target were also positive for the other and were therefore deemed intact. This data suggested that the extraction process was causing shearing of proviruses.

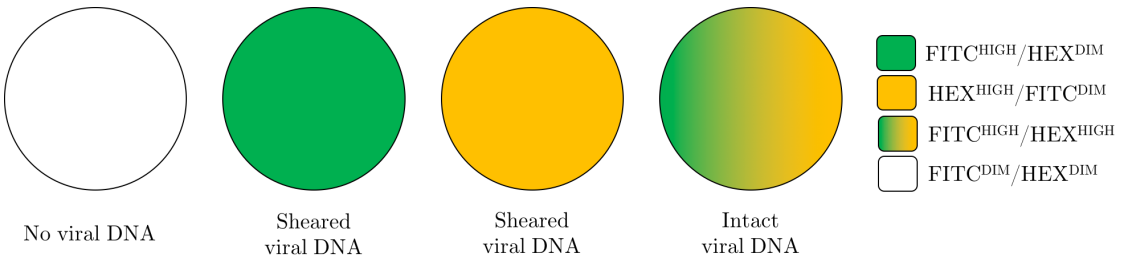


Figure 6.5: Potential droplet types dependent upon provirus shearing.

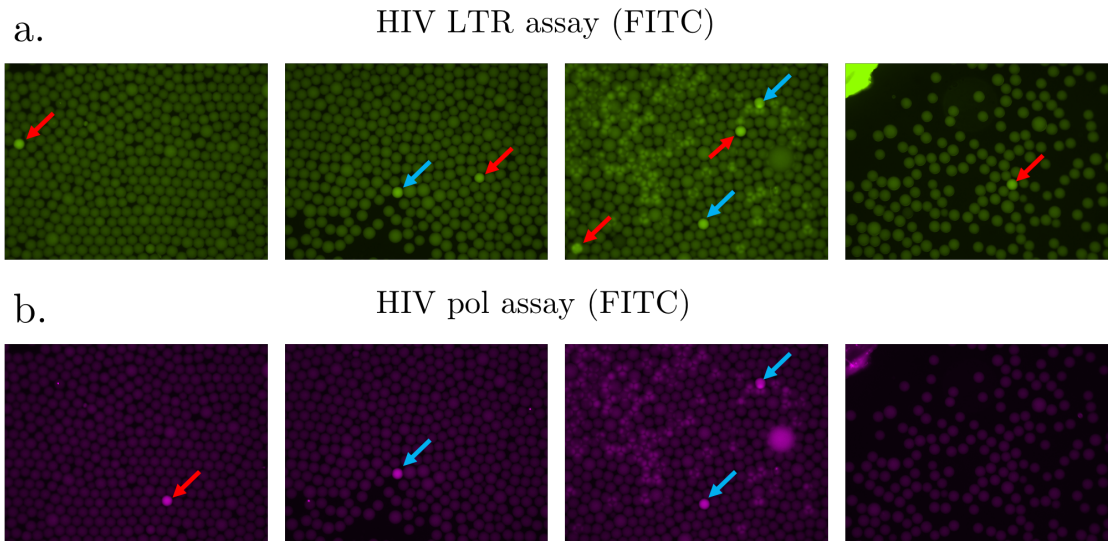


Figure 6.6: Colocalisation of HIV qPCR assay fluorescence for Qiagen extracted DNA. (a) Representative images of droplets viewed on thr FITC channel. (b) Representative images of droplets viewed on the HEX channel. Red arrows indicate droplets with sheared proviruses, blue arrows indicate droplets with intact proviruses.

In-general, column-based DNA extraction methods cause greater shearing than non-column-based methods. In order to test whether a more gentle extraction method would improve the DNA fragment size, the Qiagen Puregene method was tested. The Puregene method uses washes and centrifugation steps to lyse cells and pellet the DNA without the need for a column. A repeat of the multiplexed PCR assay experiment described above was performed on Puregene extracted DNA from the same cell-line. In this assay 75% of droplets that were positive for one HIV target were also positive for the other and were therefore deemed intact.

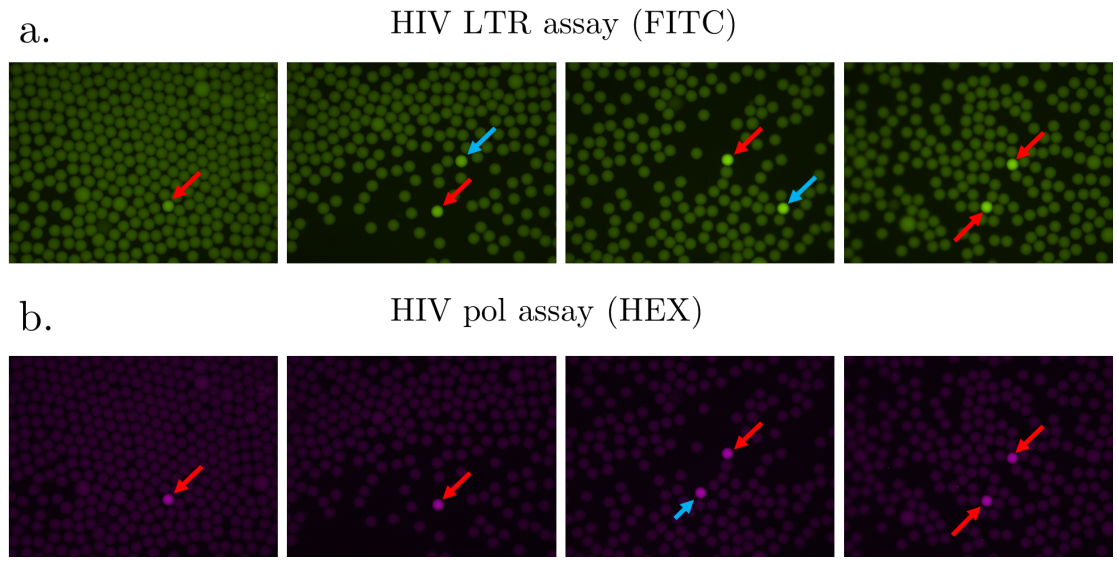


Figure 6.7: Colocalisation of HIV qPCR assay fluorescence for Qiagen extracted DNA. (a) Representative images of droplets viewed on thr FITC channel. (b) Representative images of droplets viewed on the HEX channel. Red arrows indicate droplets with shared proviruses, blue arrows indicate droplets with intact proviruses.

6.2.6 DNA enrichment pipeline discussion

As shown in figure 6.1, this pipeline is nearing completion and will likely soon be able to test clinical samples. Steps including: droplet production, DNA encapsulation, droplet thermocycling and droplet sorting have all been optimised for rare target enrichment. The remaining hurdle for this pipeline appears to be the potential shearing effects caused by genomic DNA extraction methods. This shearing results in a number of HIV proviral DNA fragments being split between multiple droplets and therefore negatively impacting enrichment. It has been shown that by replacing column-based methods with non-column based, it is possible to reduce this effect. Moving forward there are two possible solutions that will be tested:

1. Test a series of different DNA extraction methods to assess whether there are any technologies that allow for longer fragment retention. This would allow for longer DNA fragments to be processed in droplets and will reduce or eliminate the presence of sheared proviruses.

2. Utilise the multiplex qPCR assay in the droplet sorting stage of the pipeline. By using the same colour fluorophore on both the HIV LTR and the HIV pol assay it will be possible to sort droplets containing either of the two fragments therefore negating the requirement for intact proviruses. This could also be expanded to include further areas of the genome and may provide some protection against primer mismatch-induced loss where a particular provirus may not have the HIV LTR primer site intact and would therefore not be sorted.

6.3 Cell enrichment pipeline testing

6.3.1 Testing the full cell enrichment pipeline using uninfected PBMCs.

In order to test the constituent parts of the cell enrichment pipeline an experiment was performed on the GXR cell-line using a qPCR assay targeting albumin as the fluorescent readout. PBMCs were co-encapsulated with Pro-K and SDS using 10% Optiprep as a density matching medium (fig 6.8.a). Droplets were incubated for 30 minutes at 55℃. Following lysis, droplets were merged with larger droplets containing PCR mix as described in 4.2.8 and then thermocycled using the following program: 93℃ for 5mins followed by 45 cycles of 93℃ for 15secs and 60℃ for 1min. Example images of the droplets post-thermocycling can be seen in figure 6.8.b. Three different populations are created as a result of the merging process as can be seen in figure 6.8.c.

- **Unmerged PCR mix droplets.** This population consists of PCR mix droplets that were not paired with a cell lysis droplet in the merging chip. The performance of the initial merging run will determine the prevalence of these droplets in the bulk population. They are smaller than the other two droplets as a result of not merging and will have the same fluorescent intensity as the PCR mix droplets prior to merging ($FITC^{NORM}$).

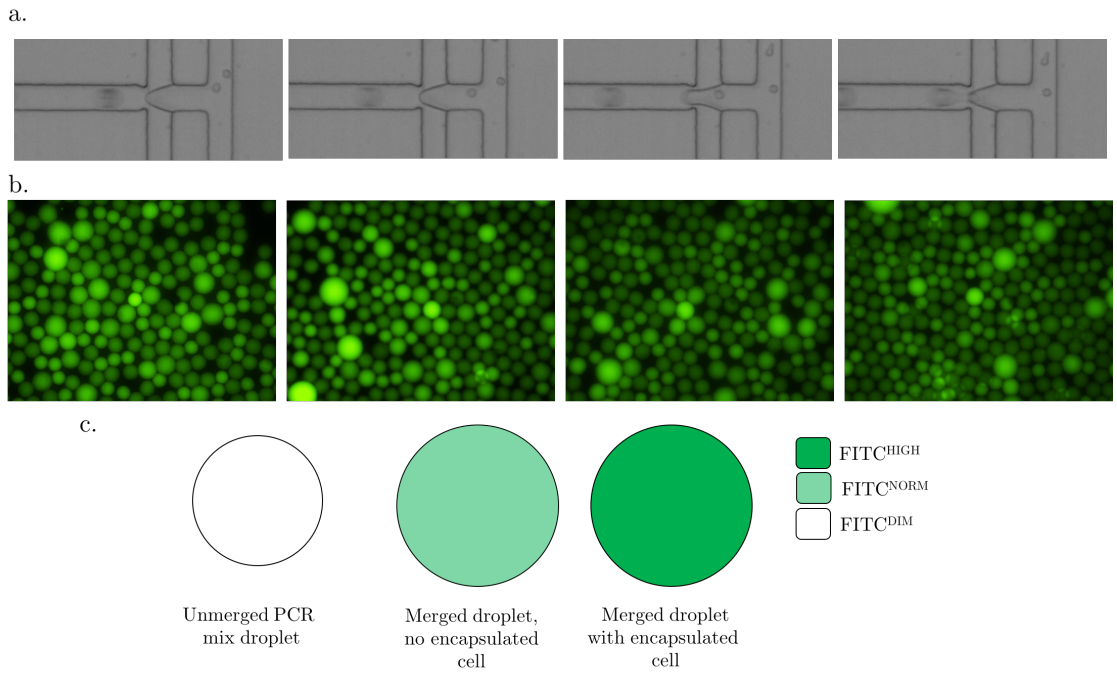


Figure 6.8: Colocalisation of HIV qPCR assay fluorescence for Qiagen extracted DNA. (a) Representative images of droplets viewed on thr FITC channel. (b) Representative images of droplets viewed on the HEX channel.

- **Merged droplets where the lysis droplet did not contain a cell.** This population consists of PCR mix droplets that have successfully merged with a cell lysis droplet but that particular droplet did not contain a cell. These droplets are larger than the unmerged droplets as they have the additional volume from the merged lysis droplet. They have a lower fluorescent intensity than the unmerged droplets as the PCR probe has been diluted by the contents of the cell lysis droplet (FITC^{LOW}).

- **Merged droplets containing an encapsulated cell lysate.** This population consists of PCR mix droplets that have successfully merged with a cell lysis droplet that contained a lysed cell. These droplets are larger than the unmerged droplets and the same size as the other merged population. As a result of successful PCR amplification of the albumin target DNA these droplets have the highest level of fluorescent intensity (FITC^{HIGH})

For this experiment ~4000 droplets were analysed for their fluorescent intensity

using a photomultiplier tube. Of these ~4000 droplets, 25 were positive for the albumin PCR amplification. This ratio of positive to negatives is consistent with the initial encapsulation rate of the experiment and the observed droplet pairing data during merging.

6.3.2 Cell enrichment pipeline discussion

This pipeline is further from being applied to clinical samples when compared to the DNA enrichment pipeline. At present, the pipeline is capable of identifying and sorting cells based on a genomic target of interest and the outstanding issues are in regards to time. In the experiment detailed above where albumin was used as a target the rate of positives was ~1:150. Taken separately, the encapsulation and merging stages of the pipeline took ~4 and ~3 hours respectively. To go on and sort the droplets, assuming an end-goal of 500 isolated positive cells, would take a further 18 hours of continuous sorting. If we were then to consider this in the context of HIV latency where one infected cell per 100 uninfected cells is more realistic, the processing time quickly becomes a barrier. With this in mind, there is a need to increase the efficiency of the process. This can be achieved by one or more of the following:

- **Increase the single cell encapsulation efficiency.** By improving the density matching of the cell suspension media and investigating further methods to prevent cell sedimentation it is possible to increase the density of the suspension and subsequently increase the number of cells encapsulated per hour.

- **Optimised device geometries** By investigating new device geometries it may be possible to increase cell encapsulation, droplet merging and droplet sorting speed. As an example, the current sorting device being used is ~5xfaster than the device that was initially used for this project.

6.4 Conclusion

The ability to identify and isolate a cell latently infected with HIV and potentially be able to retain the DNA and RNA at single-cell resolution is one with profound implications for the HIV cure field. At present, there are no methods able to reliably detect every cell that contains within it the ability to cause recrudescent HIV viraemia. The two pipelines described above, in particular, the cell enrichment pipeline have the potential to do so. Despite being technically challenging, persevering with these technologies may provide key breakthroughs in our understanding of the causes, maintenance and reactivation potential of the reservoir.

Chapter 7

General discussion

Contents

7.1 General discussion 190
7.1.1 Limitations of this work. 192
7.1.2 Implications of this thesis for the HIV cure field and beyond. 194
7.2 Publications 195
7.2.1 Papers in preparation. 195
7.3 Papers to follow 196
7.4 Papers published during the thesis period 196

7.1 General discussion

The HIV cure field is approaching a plateau in its understanding of factors that contribute to the creation, maintenance, and reactivation of the latent reservoir. Due to the latent nature of the resting cells that dominate this reservoir, they lack any cell surface expression that would label them as either: targets for immune clearance or targets by which researchers could isolate them for further study. The ongoing challenge to identify latently infected cells is hindered further by the fact that as many as 95% of proviruses resident within these cells may be replication incompetent and of no relevance to the recrudescent viraemia commonly observed in the absence of ART. This often highly heterogenous population of cells resident within both blood and tissue compartments that is able to become stochastically reactivated is likely to require isolation to be fully understood. As has been discussed previously in this thesis, a number of methods have been developed over the past 15 years attempting to enrich for and isolate this reservoir with varying degrees of success. By developing a combined microfluidics and DNA target enrichment pipeline, this thesis set out to develop a novel approach towards HIV reservoir isolation, quantification and characterisation.

In a departure from other assays used in reservoir research which focus largely on reactivation and the production of viral RNA transcripts, this pipeline aimed to target the integrated viral genome without which HIV viraemia can not rebound in the absence of ART.

The key findings of this thesis are:

- ☐ It is possible to perform full-length provirus deep sequencing on samples from PLWH on suppressive ART from relatively small bleed volumes using a biotinylated RNA bait target enrichment protocol.

- ☐ It is possible to identify provirus integration sites from PLWH on suppressive ART from relatively small bleed volumes using a biotinylated RNA bait target enrichment protocol.

- PCR primers commonly used in full-length and overlapping amplicon HIV sequencing assays contain a high level of mismatches that would likely bias results if applied to the samples from PLWH used in this study. The 120 nucleotide biotinylated RNA baits used in this thesis are able to capture a large diversity of sequence polymorphism across these sites.

- Using a combination of single-cell encapsulation, in-drop lysis, droplet merging and thermocycling with a double-quenched fluorescent probe, cells containing an integrated copy of HIV can be identified by their fluorescent intensity.

- Using a microfluidic droplet sorting chip, droplets can be dielectrophoretically sorted based on their fluorescent intensity.

- Fluorophores commonly used to label cell surface markers are able to withstand thermocycling and have the potential to be used in combination with in-drop PCR to characterise cells latently infected with HIV. By producing water-in-oil-in-water droplets these multi-labelled droplets can be analysed on a conventional LSRII flow cytometer.

Due to the technical nature of this thesis, the novelty of the concept and the numerous optimisations that have had to be tested and validated at each step of the pipeline, before this can be applied to a clinical cohort, there are still a number of small barriers to overcome. A proof of concept has been established for each step, however due to the scarcity of infected cells, huge numbers of droplets have to be screened in order to identify a relatively small fraction of infected cells. Consequently, further optimisations based primarily around increasing the efficiency of screening are planned. In addition, the issue regarding the fragmentation of DNA in the DNA enrichment pipeline requires resolution. A clear roadmap has been set for this as detailed in chapter 6.2.6.

The biotinylated RNA bait target enrichment pipeline is a stand-alone protocol in addition to its integration into the droplet pipeline. This pipeline has been

shown to enable full-length proviral sequencing from the CD4+ T-cells of PLWH on suppressive ART. In addition, through the development of a novel bioinformatic pileline has the ability to identify the proviral integration sites associated with these full-length sequences.

7.1.1 Limitations of this work.

7.1.1.1 Target enrichment

The modified target enrichment protocol presented in this thesis has some limitations when considered in the context for which it was designed: cohorts of PLWH on suppressive ART. Firstly, despite the significant enrichment observed using this method, the depth of sequencing produced for each sample is limited. It was initially hoped that it may be possible to produce a great enough depth of sequencing to permit haplotype reconstruction and achieve individual virus resolution of sequences. A similar consideration must also be made for proviral integration site identification. Despite the benefit of identifying integrants that may be missed by PCR based methods due to primer mismatch, the target enrichment pipeline proposed is only capable of identifying as many integrations as are present in the sample.

Both of these limitations stem from the dominant limiting factor in the target enrichment protocol: assay sample input. The 4 µg limit imposed by the enzymatic reactions at the beginning of the protocol and the subsequent loss through sequential ligations and washes results in a small number of proviruses being used in the hybridisation reaction. Outside the remit of this thesis but planned in the near future is the potential to pre-enrich proviral DNA using the DNA droplet microfluidics assay. This pre-enriched DNA will then be used as template for the target enrichment protocol and should allow for a greater number of proviruses to be included in the hybridisation reaction.

7.1.1.2 Droplet microfluidics DNA enrichment

The droplet enrichment protocol addresses a number of the key issues facing assays designed to target the reservoir, it does however have its limitations. In

its current guise the droplet protocol is realistically capable of isolating ~5000 positive droplets for the DNA pipeline and ~1000 positive droplets for the single cell pipeline in a reasonable timeframe. Although the prospect of being able to perform RNAseq on a pure population of 1000 latently infected cells is currently beyond what is possible, with further development the technology holds promise to isolate potentially 10x this number.

At present, droplets are identified as containing a provirus by the presence of one short 150bp region of the HIV genome. This identification method has already been shown to be a limiting factor in the DNA pipeline. Moving forward, it is likely that an identification method that involves a number of well conserved qPCR targets distributed throughout the HIV genome will be multiplexed within the droplet. Multiplexing qPCR assays within a droplet has already been successfully demonstrated within this thesis and by doing so, either the absence of one primer binding site or a critical nucleotide change in a primer binding site will not lead to the cell being incorrectly identified as uninfected.

7.1.1.3 The effect of compartmentalisation and sampling issues on HIV DNA reservoir studies

Another consideration that has not been addressed in detail in this thesis is that of compartmentalisation of the reservoir within the body. All of the work presented in this thesis has been performed on PBMCs. In order to gain a true representation of the viral populations present within the body it will likely be necessary to sample from potential drug sanctuary sites including the gut and lymphoid organs. A subset of participants from the HEATHER cohort have been enrolled in a gut sub study in which biopsies were taken at multiple time points from the rectum and terminal ileum. Following the final protocol optimisations on PBMCs, the aim is to modify the protocol to include tissue biopsy samples. By sequencing and quantifying these samples I hope to contribute to our understanding of reservoir dynamics over time across multiple anatomical compartments.

7.1.2 Implications of this thesis for the HIV cure field and beyond.

This thesis set out to develop a pipeline on the cutting edge of technology in an area of research that has struggled to account for the low frequency of latently infected cells present in samples from PLWH. This low frequency coupled with the absence of any cell associated markers that would label a cell as infected has continued to hinder research attempts aiming to perform an in-depth characterisation of the reservoir. The foundation that has been set down in this thesis has within it the potential to enable an entire suite of assays that can be used to interrogate the latent reservoir in HIV infection.

As discussed within the results chapters of this thesis, the following reservoir characteristics all have the potential to be addressed:

- Reservoir quantification.
- Comparative single-cell RNAseq between a population of pure latently infected cells and uninfected cells split from the same starting sample.
- Full-length proviral sequencing with associated provirus integration site identification.
- Analysis of cell-surface markers between a population of pure latently infected cells and uninfected cells split from the same starting sample as measured using WOW droplets on a conventional flow-cytometer.

In addition to those discussed in the body of this thesis there are further potential applications both within and beyond the HIV cure field:

- The ability to perform individual killing assays by co-encapsulating cells within droplets.
- Performing qVOA assays within individual droplets by stimulating and subsequently co-encapsulating them with uninfected target cells and measure p24 production.

- Latency reversal assays where primary resting CD4+ T-cells are encapsulated with LRAs and uninfected target cells and p24 production is measured.

- Outside the field of HIV cure, any condition that mimics a period of clinical latency as seen in HIV, such as circulating tumour cells in cancer patients, could also utilise the groundwork set out in this thesis to develop new assays to isolate and interrogate target cells.

In summary, the groundwork set out in this thesis will enable researchers, previously unfamiliar with microfluidics, to quickly develop the skills and knowledge necessary to implement these assays within their own laboratories. In addition to this, the work completed as part of this thesis has progressed alongside other microfluidics labs developing related assays. The information contained herein will therefore undoubtedly overlap with the goals of other laboratories who will be able to utilise it in their own research.

Finally, the overarching aim of this thesis was to develop an integrated suite of assays that would: enable an increased understanding of the latent reservoir, enable effective TI studies and potentially lead to the identification of novel therapeutic targets that might lead to the eradication of HIV. As detailed within this thesis, these assays have been conceived, developed, optimised and validated to the point where they will soon be in a position to be applied to clinical cohorts and begin providing some of these answers.

7.2 Publications

As the result of a successful patent application regarding the work contained within this thesis, there have been no direct peer-reviewed publications made to date. As the filing is now complete, the plan for publishing the work in this thesis is set out below.

7.2.1 Papers in preparation.

1. **Combined full-length HIV proviral sequencing and integration site identification: A target enrichment approach.**

   This paper details the work predominantly covered in chapters 2 and 3 regarding the development of the target enrichment protocol and its application to clinical samples.

2. **Breakalign: An integration site identification and visualisation tool for target enriched DNA sequencing data**

   This paper seeks to publish the Breakalign bioinformatics tool that allows for the easy identification and visualisation of integrations sites extracted from target enriched DNA sequencing data.

7.3 Papers to follow

1. **Enrichment of HIV proviral DNA through the utilisation of droplet microfluidics.**

   This paper will detail the DNA pipeline method development and its application to cell lines covered in chapters 4, 5 and 6 and also its application to clinical samples upon final validation of the pipeline.

2. **Single-cell HIV reservoir quantification and full-length sequencing through the utilisation of droplet microfluidics.**

   This paper will detail the single-cell pipeline method development as covered in chapters 4, 5 and 6 and also its application to clinical samples upon final validation of the pipeline.

7.4 Papers published during the thesis period

1. Foster C, Pace M, Kaye S, Hopkins E, **Jones M**, et al. (2017) Early antiretroviral therapy reduces HIV DNA following perinatal HIV infection. AIDS 31:1847-1851.

2. Murray LW, Satti I, Meyerowitz J, **Jones M**, Willberg CB, et al. (2018) Human Immunodeficiency Virus Infection Impairs Th1 and Th17 *Mycobacterium tuberculosis* Specific T-Cell Responses. The Journal of infectious diseases 217:1782-1792.

3. Gossez M, Martin GE, Pace, M, Ramjee G, Premraj A, Kaleebu P, Rees H, Inshaw J, Stohr W, Meyerowitz J, Hopkins E, **Jones M**, Hurst J, et al. (2019) Virological remission after antiretroviral therapy interruption in female African HIV seroconverters. AIDS 33(2):185-197.

Chapter 8

General methods

Contents

8.1 Cell lines . . . 199
8.2 General techniques for ex-vivo sample handling . . . 200
8.2.1 Processing and cryopreservation of blood samples . . . 200
8.2.2 CD4+ T-cell enrichment . . . 201
8.3 Nucleic acid extraction . . . 201
8.4 HIV reservoir quantification . . . 201
8.5 Statistical analysis . . . 201
8.6 Target capture and next-gen sequencing . . . 203
8.6.1 Sequencing analysis . . . 203
8.7 Droplet microfluidics . . . 204
8.7.1 Photolithography masks . . . 204
8.7.2 Silicon wafer master production . . . 204
8.7.3 PDMS chip production . . . 204
8.7.4 Running microfluidic devices . . . 205
8.7.5 Droplet merging . . . 205
8.7.6 Droplet sorting . . . 205
8.7.7 Emulsion breaking . . . 205
8.8 Clinical trials . . . 205
8.8.1 SPARTAC . . . 205
8.8.2 HEATHER . . . 206

8.1 Cell lines

- **8E5 cells** Containing a single proviral copy of HIV, cells were cultured in RPMI 1640 medium supplemented with 10% fetal calf serum, antibiotics and l-glutamine. This cell line was obtained through the AIDS Research

and Reference Reagent Program, Division of AIDS, NIAID, NIH from Dr. Thomas Folks [232].

- **MT-4 cells** Human T-cell isolated from a patient with adult T-cell leukemia, cells were cultured in RPMI 1640 medium supplemented with 10% fetal calf serum, antibiotics and l-glutamine.The following reagent was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: MT-4 from Dr. Douglas Richman [233].

- **GXR cells** A reporter cell-line used to express GFP upon HIV infection, cells were cultured in RPMI 1640 medium supplemented with 10% fetal calf serum, antibiotics and l-glutamine [234].

8.2 General techniques for ex-vivo sample handling

8.2.1 Processing and cryopreservation of blood samples

Blood samples in ACD tubes (BD) were centrifuged at 750rcf for 8 minutes and layered plasma was aspirated from the top of the tubes using a pasteur pipette. Plasma was immediately placed into cryotubes and stored at -80℃. The original blood volume was restored by adding 1x phosphate buffered saline (PBS, Gibco). Blood samples were then layered 2:1 over a cushion of density gradient medium (Stemcell technologies) and centrifuge separated at 900rcf for 24 minutes. The resulting PBMC layer was aspirated from the sample and subjected to two washes in R10 media (R10 media consists of RPMI-1640 with 10% heat-inactivated foetal calf serum (FCS), 2 mM L-glutamine, 100IU/mL penicillin and 0.1mg/ml streptomycin). Cell counts and viability were assessed using a cell analyser (Millipore) combined with a viability stain (Millipore). Cells were resuspended in aliquots of between 5-15 x 10^6cells/mL in a freezing medium consisting of 10% dimethylsulfoxide (DMSO, Sigma), 50% FCS and 40% RPMI-1640 (Invitrogen).

PBMCs were thawed by briefly placing the vial in a 37/celsius waterbath followed by dilution into 15mL of R10. Cells were washed once in R10 and resuspended in a media suitable for the downstream application.

8.2.2 CD4+ T-cell enrichment

PBMC samples were enriched for CD4+ cells using the Easysep Human CD4+ T-cell enrichment kit. The kit was used as per the manufacturer's instructions.

8.3 Nucleic acid extraction

DNA was extracted using either the Qiagen Blood mini, Qiagen Blood Midi or Qiagen Puregene extraction protocols as per the manufacturer's instructions including the RNase-H step. DNA was eluted using RNAse/DNAse free water and stored short term at 4℃ and longer term (>1 month) at -20℃.

RNA was extracted using the Qiagen RNeasy Mini kit as per manufacturer's instructions including the DNaseI step. RNA was eluted using RNAse/DNAse free water, aliquoted into volumes suitable for single use to avoid freeze/thaw cycles and stored at -80℃.

8.4 HIV reservoir quantification

Total HIV DNA was quantified as previously published in Jones et al [139]. HIV DNA was reported as total HIV copies/10^6 CD4+ T-cells. qPCR primers, probes and concentrations are shown in table 8.1.

Unless otherwise stated in the text, all qPCR reactions were perfomed using LC480 mastermix (Roche) at a 1x concentration with the following thermocycling conditions: one cycle of 95℃, 45 cycles of 95℃ for 15 seconds and 60℃ for 60 seconds. Data was either analysed using the LC480 instrument software or analysed using R.

Table 8.1: qPCR and in-drop PCR primer and probe sequences

FAM = Fluoroscein amidite, HEX = Hexachlorofluorescein succinimidyl ester, BHQ = Black-hole quencher 1 and ZEN = Internal Zen quencher. All primers and probes were procured from Sigma with the exception of the double quenched probes (Integrated DNA technologies).

Name	Type	Quenching	Sequence	Concentration used	Chapter
			Albumin		
Alb F	Primer	-	GCTGTCATCTCTTGTGGGCTGT	250 nM	2, 3
Alb R	Primer	-	AAACTCATGGGAGCTGCTGGTT	250 nM	2, 3
Alb Probe	Probe	Single	FAMCCTGTCATGCCCACACAAATCTCTCCBHQ	250 nM	2, 3
			Total HIV DNA		
LTR-OS	Primer	-	GRAACCCACTGCTTAASSCTCAA	500 nM	2, 3
LTR-AS	Primer	-	AAACTCATGGGAGCTGCTGGTT	500 nM	2, 3
LTR Probe	Probe	Single	FAMAGTRGTGTGTGCCCGTCTGTTGBHQ	500 nM	2, 3
			HIV iSCA pol assay		
iSCA F	Primer	-	TTTGGAAAGGACCAGCAAA	400 nM	2, 3
iSCA R	Primer	-	CCTGCCATCTGTTTTCCA	400 nM	2, 3
iSCAPrb	Probe	Single	HEXAAAGGTGAAGGGGCAGTAGTAATACABHQ	200 nM	5, 6
			Double quenched probes		
LTR DBL	Probe	-	FAMAGTRGTGTG-**ZEN**-TGCCCGTCTGTTGBHQ	400 nM	2, 3
iSCA DBL	Probe	-	FAMAAAGGTGAA-**ZEN**-GGGGCAGTAGTAATACABHQ	400 nM	2, 3

8.5 Statistical analysis

All data analysis was performed using R [235] or Prism (Graphpad). Figures were generated using either Prism or the ggplot2 R package [236].

8.6 Target capture and next-gen sequencing

The Agilent SureSelect XT protocol was used for target capture, however, a series of modifications were made to the standard protocol to improve probe specificity. The protocol in brief is as follows. Four micrograms of CD4+ T-cell DNA eluted in water were sheared using a Covaris M220 sonicator with a target fragment size of 150bp. End repair and sequencing-adapter ligation was performed as per manufacturers recommendations followed by a limited cycle amplification of sample. Between 1500 and 2000 nanograms of amplified DNA library was hybridised to the biotinylated probe library between 65řC - 71řC for 20hrs, after which, hybridised libraries were pulled down using streptavidin beads (Dynabeads T1, Life Technologies). Captured libraries were amplified using limited cycle PCR and clean-up performed using solid phase reversible immobilisation (SPRI) beads (Ampure XP, Beckman Coulter). Final libraries were quantified using either Picogreen, Qubit assays (Life technologies) or qPCR and subsequently sequenced using either the Illumina MiSeq V2 300bp cycle kit or the MiSeq V3 600bp cycle kit.

8.6.1 Sequencing analysis

Raw fastq files were first analysed with fastqc to ensure quality thresholds were met. Illumina and SureSelect adapters were then trimmed using the SureCall Trimmer (Agilent, USA). Post trimming, a repeat run of FastQC was performed to ensure any adapter and K-mer contamination had been removed from the raw reads. Trimmed reads were mapped to HIV-1 using either BWA-MEM or Novoalign. Either the HIV HXB2 reference sequence (Accession: K03455 M38432) was used or a multi fasta file consisting of 63 near full length HIV genomes downloaded from the Los Alamos HIV National Database [237]. Mapped SAM files were processed to indexed, sorted BAM files and PCR duplicates were removed using SamTools rmdup package.

8.6.1.1 LTR trimming

In order to determine HIV-1 integration sites, trimmed reads were first mapped to the HIV-1 HXB2 LTR reference sequence to look for potential mismatches and establish an LTR profile. Providing no mismatches were found, each read was subsequently interrogated for the first 12 nucleotides of the profiled LTR; only reads with 100% homology to this sequence were retained. All nucleotides downstream of these 12 bases were trimmed from the read thereby removing viral DNA and leaving only the adjacent host DNA. LTR trimmed reads were then mapped to the HG19 human reference genome in order to determine the viral insertion site.

8.7 Droplet microfluidics

8.7.1 Photolithography masks

Photolithography masks were produced using computer aided design software packages including CorelDraw and AutoCAD. Proofs were sent to Micro Lithography Services for printing.

8.7.2 Silicon wafer master production

Silicon wafers were first cleaned using nitrogen gas and then heated to 150℃ on a hotplate to remove any moisture on the surface. 5mL of degassed SU-8 (Microchem) was poured onto the centre of the wafer and spun according to the manufacters recommendations for the desired feature height. Once spin-coated the wafer was soft baked according to manufacters recommendations. The soft baked wafer was then exposed to UV light with the patterned lithographic mask between the light source and the wafer. Wafer was then hard baked and the uncrosslinked SU-8 removed chemically.

8.7.3 PDMS chip production

Sylgard 184 (Dow Corning) curing agent was mixed 1:10 with Sylgard 184 base elastomer, degassed and poured directly on to the silicon master. The master was then baked overnight at 65℃. Following baking, PDMS chips were peeled from

the master, inlet holes pierced, cleaned, and finally bonded to a glass slide using oxygen plasma. The bonded chip was silanised using Hexamethyldisilazane (Sigma) by placing 0.5ml in an histology staining jar with the chip overnight.

8.7.4 Running microfluidic devices

Inlet and outlet tubing was pushed into the corresponding holes on the chip and the other end of the tubing was attached to syringe pumps (Harvard Apparatus) or collection tubing. Flow was initiated and droplets were collected by inverting the outlet tubing over the collection tube.

8.7.5 Droplet merging

Droplets were loaded into an eppendorf tube with a PDMS plug above them and loaded into the merging chip by positive displacement using oil fed from a syringe pump. Merging was initiated by applying an electric field across the device using a function generator and voltage amplifier.

8.7.6 Droplet sorting

Droplets were loaded into an eppendorf tube with a PDMS plug above them and loaded into the merging chip by positive displacement using oil fed from a syringe pump. Sorting was controlled by a custom script written in Python operating through an Arduino Due microcontroller (Arduino). The electric field was generated using a function generator and voltage amplifier.

8.7.7 Emulsion breaking

Emulsions were broken by mixing emulsion with 2x the volume of perfluorooctanol and inverting ten times. Separation was achieved by centrifugation at 1000rcf for 1 minute and the aqueous fraction was aspirated from the oil/perfluorooctanol mix.

8.8 Clinical trials

8.8.1 SPARTAC

The SPARTAC trial was a multi-centre randomised controlled trial of short course ART during PHI, where PHI was determined to be within 6 months of diagnosis [238]. Participants were randomised to receive either 12 weeks of ART, 48 weeks of ART or to receive 'standard of care' which at the time of this study was to remain off of therapy until a CD4 count of <350 cell/mm^3. Participants on both the 12 and 48 week arms underwent a TI after the treatment period. The primary endpoint of the trial was a CD4 count of <350 cell/mm^3. PBMCs and plasma were collected at each visit. Follow-up was completed for SPARTAC in 2010. Following the introduction of the new WHO guidelines regarding therapy initation it is no longer deemed ethical to withhold treatment to allow an untreated trial arm. This makes the remaining SPARTAC trial samples, with their matched standard of care arm, a valuable resource.

Samples from the SPARTAC cohort to be used in this thesis were selected based on the proviral DNA burden and therapy status.

8.8.2 HEATHER

The HEATHER cohort is an observational study targeting PLWH who began ART during PHI. Heather participants belonging to the A group were prospectively enrolled prior to ART initiation and Heather B participants were enrolled whilst on effective ART if they were deemed to meet the inclusion criteria. PBMCs and plasma are collected at each visit and HEATHER is still actively recruiting.

Samples from the HEATHER cohort to be used in this thesis were selected based on the proviral DNA burden and the time since therapy initiation.

References

1. Barre-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220,** 868–871 (May 1983).
2. Gallo, R. *et al.* Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **220,** 865–867 (May 1983).
3. UNAIDS. *AIDS by the numbers 2015* Media release. 2015.
4. Loutfy, M. R. *et al.* Systematic review of HIV transmission between heterosexual serodiscordant couples where the HIV-positive partner is fully suppressed on antiretroviral therapy. *PloS one* **8,** e55747 (2013).
5. Michel, U. *Fast Track; ending the AIDS epidemic by 2030* July 2016. <http://www.unaids.org/sites/default/files/media_asset/JC2686_WAD2014report_en.pdf>.
6. Sidibé. M, L. L. & B, S. The UNAIDS 909090 target: a clear choice for ending AIDS and for sustainable health and development. *Journal of the International AIDS society.* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4947868/> (2016).
7. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. *Science* **271,** 1582–1586 (Mar. 1996).
8. Briggs, J. A. & Kräusslich, H.-G. The Molecular Architecture of HIV. *Journal of Molecular Biology* **410,** 491–500 (July 2011).
9. Fiorentini, S., Marini, E., Caracciolo, S. & Caruso, A. Functions of the HIV-1 matrix protein p17. *The new microbiologica* **29,** 1–10 (2006).
10. Wilen, C., Tilton, J. & Spring, D. .-.-. R. HIV: cell binding and entry. *Perspectives in medicine.* <http://perspectivesinmedicine.cshlp.org/content/2/8/a006866.short> (2012).
11. Arthos, J. *et al.* HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells. *Nature immunology* **9,** 301–9 (Mar. 2008).
12. Cameron, P. U. *et al.* Dendritic cells exposed to human immunodeficiency virus type-1 transmit a vigorous cytopathic infection to CD4+ T cells. *Science (New York, N.Y.)* **257,** 383–7 (July 1992).

13. Geijtenbeek, T. B. *et al.* DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100,** 587–97 (Mar. 2000).

14. Turville, S. G. HIV gp120 receptors on human dendritic cells. *Blood* **98,** 2482–2488 (Oct. 2001).

15. Tamamis, P. & Floudas, C. A. Molecular Recognition of CCR5 by an HIV-1 gp120 V3 Loop. *PLoS ONE* **9** (ed Paci, E.) e95767 (Apr. 2014).

16. Sattentau, Q. J. HIV gp120: double lock strategy foils host defences. *Structure* **6,** 945–949 (Aug. 1998).

17. Cooley, L. A. & Lewin, S. R. HIV-1 cell entry and advances in viral entry inhibitor therapy. *Journal of Clinical Virology* **26,** 121–132 (Feb. 2003).

18. Mamede, J. I., Cianci, G. C., Anderson, M. R. & Hope, T. J. Early cytoplasmic uncoating is associated with infectivity of HIV-1. *Proceedings of the National Academy of Sciences* **114,** E7169–E7178 (Aug. 2017).

19. Francis, A. C. & Melikyan, G. B. Single HIV-1 Imaging Reveals Progression of Infection through CA-Dependent Steps of Docking at the Nuclear Pore, Uncoating, and Nuclear Transport. *Cell Host & Microbe* **23,** 536–548.e6 (Apr. 2018).

20. Hu, W.-S. & Hughes, S. H. HIV-1 Reverse Transcription. *Cold Spring Harbor Perspectives in Medicine* **2,** a006882–a006882 (June 2012).

21. N, L. *et al.* Structural and functional studies of the HIV-1 pre-integration complex. *Retrovirology.* <https://retrovirology.biomedcentral.com/articles/10.1186/1742-4690-10-S1-P76> (2013).

22. Coffin JM, V. H. Hughes SH. *Retroviruses* (Cold Spring Harbor Laboratory Press, 1997).

23. Masuda, T. *et al.* Fate of HIV-1 cDNA intermediates during reverse transcription is dictated by transcription initiation site of virus genomic RNA. *Scientific Reports* **5.** <https://doi.org/10.1038/srep17680> (Dec. 2015).

24. Kharytonchyk, S. *et al.* Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proceedings of the National Academy of Sciences* **113,** 13378–13383 (Nov. 2016).

25. Egger, M. *et al.* Impact of new antiretroviral combination therapies in HIV infected patients in Switzerland: prospective multicentre study. *BMJ* **315,** 1194–1199 (Nov. 1997).

26. Michaels, S. H., Clark, R. & Kissinger, P. Declining Morbidity and Mortality among Patients with Advanced Human Immunodeficiency Virus Infection. *New England Journal of Medicine* **339,** 405–406 (Aug. 1998).

27. Collaboration, C. Determinants of survival following HIV-1 seroconversion after the introduction of HAART. *The Lancet* **362,** 1267–1274 (Oct. 2003).

28. HIV.gov. *High-Resolution Automated Electrophoresis of DNA, RNA, and Protein Samples* Online. Apr. 2019. <https://www.hiv.gov/hiv-basics/overview/data-and-trends/global-statistics>.

29. *Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV* 2019. <Guideline%20on%20when%20to%20start%20antiretroviral%20therapy%20and%20on%20pre-exposure%20prophylaxis%20for%20HIV>.

30. Major Clinical Outcomes in Antiretroviral Therapy (ART)Naive Participants and in Those Not Receiving ART at Baseline in the SMART Study. *The Journal of Infectious Diseases* **197,** 1133–1144 (Apr. 2008).

31. Cohen, M. S. *et al.* Prevention of HIV-1 Infection with Early Antiretroviral Therapy. *New England Journal of Medicine* **365,** 493–505 (Aug. 2011).

32. Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: a collaborative analysis of 18 HIV cohort studies. *The Lancet* **373,** 1352–1363 (Apr. 2009).

33. Group, I. S. S. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *New England Journal of Medicine.* <https://www.nejm.org/doi/full/10.1056/nejmoa1506816> (2015).

34. Jain, V. *et al.* Antiretroviral Therapy Initiated Within 6 Months of HIV Infection Is Associated With Lower T-Cell Activation and Smaller HIV Reservoir Size. *The Journal of Infectious Diseases* **208,** 1202–1211 (July 2013).

35. Archin, N. M. *et al.* Immediate antiviral therapy appears to restrict resting CD4+ cell HIV-1 infection without accelerating the decay of latent infection. *Proceedings of the National Academy of Sciences* **109,** 9523–9528 (May 2012).

36. Henrich, T. J. *et al.* HIV-1 persistence following extremely early initiation of antiretroviral therapy (ART) during acute HIV-1 infection: An observational study. *PLOS Medicine* **14** (ed Bekker, L.-G.) e1002417 (Nov. 2017).

37. Ananworanich, J. *et al.* HIV DNA Set Point is Rapidly Established in Acute HIV Infection and Dramatically Reduced by Early ART. *EBioMedicine* **11,** 68–72 (Sept. 2016).

38. Rodger, A. J. *et al.* Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy. *JAMA* **316,** 171 (July 2016).

39. Cohen, M. S. *et al.* Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *New England Journal of Medicine* **375,** 830–839 (Sept. 2016).

40. McCormack, S. *et al.* Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *The Lancet* **387,** 53–60 (Jan. 2016).

41. Molina, J.-M. *et al.* On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. *New England Journal of Medicine* **373,** 2237–2246 (Dec. 2015).

42. Baeten, J. M. *et al.* Antiretroviral Prophylaxis for HIV Prevention in Heterosexual Men and Women. *New England Journal of Medicine* **367,** 399–410 (Aug. 2012).

43. P, B. S. Maximizing the Benefits of HIV Preexposure Prophylaxis. *Topics in antiviral medicine* **25,** 138–142 (Apr. 2018).

44. Brown, T. R. I Am the Berlin Patient: A Personal Reflection. *AIDS Research and Human Retroviruses* **31,** 2–3 (Jan. 2015).

45. Hütter, G. *et al.* Long-Term Control of HIV byCCR5Delta32/Delta32 Stem-Cell Transplantation. *New England Journal of Medicine* **360,** 692–698 (Feb. 2009).

46. Liu, R. *et al.* Homozygous Defect in HIV-1 Coreceptor Accounts for Resistance of Some Multiply-Exposed Individuals to HIV-1 Infection. *Cell* **86,** 367–377 (Aug. 1996).

47. Gupta, R. K. *et al.* HIV-1 remission following CCR5 Delta32 haematopoietic stem-cell transplantation. *Nature* **568,** 244–248 (Mar. 2019).

48. Jensen, B. *et al.* ANALYTIC TREATMENT INTERRUPTION (ATI) AFTER ALLOGENEIC CCR5-D32 HSCT FOR AML IN 2013 (CROI Abstract 394). *Top Antivir Med* **26** (2018).

49. Churchill, M. J., Deeks, S. G., Margolis, D. M., Siliciano, R. F. & Swanstrom, R. HIV reservoirs: what, where and how to target them. *Nature Reviews Microbiology* **14,** 55–60 (Nov. 2015).

50. Darcis, G. *et al.* An In-Depth Comparison of Latency-Reversing Agent Combinations in Various In Vitro and Ex Vivo HIV-1 Latency Models Identified Bryostatin-1+JQ1 and Ingenol-B+JQ1 to Potently Reactivate Viral Gene Expression. *PLOS Pathogens* **11,** e1005063 (2015).

51. Lehrman, G. *et al.* Depletion of latent HIV-1 infection in vivo: a proof-of-concept study. *Lancet (London, England)* **366,** 549–55 (Jan. 2005).

52. Kline, C. *et al.* Persistence of viral reservoirs in multiple tissues after antiretroviral therapy suppression in a macaque RT-SHIV model. *PloS one* **8,** e84275 (Jan. 2013).

53. Kline, C. *et al.* Persistence of viral reservoirs in multiple tissues after antiretroviral therapy suppression in a macaque RT-SHIV model. *PloS one* **8,** e84275 (Jan. 2013).

54. Maldarelli, F. *et al.* HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science (New York, N.Y.)* **345,** 179–83 (July 2014).

55. Bailey, J. R. *et al.* Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. *Journal of virology* **80,** 6441–57 (July 2006).

56. Hashemi, F. B. *et al.* HIV provirus stably reproduce parental latent and induced transcription phenotypes regardless of the chromosomal integration site. *Journal of virology* (Mar. 2016).

57. Lorenzo-Redondo, R. *et al.* Persistent HIV-1 replication maintains the tissue reservoir during therapy. *Nature* **530,** 51–56 (Jan. 2016).

58. Brodin, J. *et al.* Establishment and stability of the latent HIV-1 DNA reservoir. *eLife* **5.** <https://doi.org/10.7554/elife.18889> (Nov. 2016).

59. Josefsson, L. *et al.* The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proceedings of the National Academy of Sciences* **110,** E4987–E4996 (Nov. 2013).

60. Crooks, A. M. *et al.* Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies. *Journal of Infectious Diseases* **212,** 1361–1365 (Apr. 2015).

61. Kandathil, A. J., Sugawara, S. & Balagopal, A. Are T cells the only HIV-1 reservoir? *Retrovirology* **13.** <https://doi.org/10.1186/s12977-016-0323-4> (Dec. 2016).

62. Laird, G., Eisele, E., Rabi, S., Lai, J. & Chioma, S. Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS ...* <http://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003398> (2013).

63. Ho, Y. C. *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155,** 540–551 (2013).

64. Churchill, M. J., Deeks, S. G., Margolis, D. M., Siliciano, R. F. & Swanstrom, R. HIV reservoirs: what, where and how to target them. *Nature Reviews Microbiology* **14,** 55–60 (Jan. 2016).

65. Brooks, D. G. & Zack, J. A. Effect of latent human immunodeficiency virus infection on cell surface phenotype. *Journal of virology* **76,** 1673–81 (Feb. 2002).

66. Garrigue, I. *et al.* Cell-associated HIV-1-DNA quantitation after highly active antiretroviral therapy-treated primary infection in patients with persistently undetectable plasma HIV-1 RNA. *Aids* **14,** 2851–2855 (2000).

67. T, F. *et al.* Susceptibility of normal human lymphocytes to infection with HTLV-III/LAV. *Journal of immunology* **135,** 4049–4053 (June 1986).

68. Dai, J. *et al.* Human Immunodeficiency Virus Integrates Directly into Naive Resting CD4+ T Cells but Enters Naive Cells Less Efficiently than Memory Cells. *Journal of Virology* **83,** 4528–4537 (Feb. 2009).

69. Finzi, D. *et al.* Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature Medicine* **5,** 512–517 (May 1999).

70. Ahmadzadeh, M., Hussain, S. F. & Farber, D. L. Heterogeneity of the Memory CD4 T Cell Response: Persisting Effectors and Resting Memory T Cells. *The Journal of Immunology* **166,** 926–935 (Jan. 2001).

71. Parisi, S. G. *et al.* Baseline cellular HIV DNA load predicts HIV DNA decline and residual HIV plasma levels during effective antiretroviral therapy. *Journal of clinical microbiology* **50,** 258–263 (2012).

72. Graf, E. H. & O'Doherty, U. Quantitation of integrated proviral DNA in viral reservoirs. *Current Opinion in HIV and AIDS* **8,** 100–105 (2013).

73. Murray, J., Zaunders, J., McBride, K. & Xu, Y. HIV DNA subspecies persist in both activated and resting memory CD4+ T cells during antiretroviral therapy. *Journal of ...* <https://jvi.asm.org/content/88/6/3516.full> (2014).

74. Siliciano, J. D. *et al.* Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nature Medicine* **9,** 727–728 (May 2003).

75. Blankson, J. *et al.* Biphasic Decay of Latently Infected CD4+T Cells in Acute Human Immunodeficiency Virus Type 1 Infection. *The Journal of Infectious Diseases* **182,** 1636–1642 (Dec. 2000).

76. Abrahams, M.-R. *et al.* The Replication-Competent HIV-1 Latent Reservoir is Primarily Established Near the Time of Therapy Initiation: Supplemental Material for Abrahams et al. <`https://doi.org/10.1101/512475`> (Jan. 2019).

77. Persaud, D. *et al.* Absence of Detectable HIV-1 Viremia after Treatment Cessation in an Infant. *New England Journal of Medicine* **369,** 1828–1835 (Nov. 2013).

78. Luzuriaga, K. *et al.* Viremic Relapse after HIV-1 Remission in a Perinatally Infected Child. *New England Journal of Medicine* **372,** 786–788 (Feb. 2015).

79. Kumari, G. & Singh, R. Highly Active Antiretroviral Therapy for treatment of HIV/AIDS patients: Current status and future prospects and the Indian scenario. *HIV & AIDS Review.* <`http://www.sciencedirect.com/science/article/pii/S1730127012000045`> (2012).

80. Hocqueloux, L. *et al.* Long-term immunovirologic control following antiretroviral therapy interruption in patients treated at the time of primary HIV-1 infection. *AIDS* **24,** 1598 (2010).

81. 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 pathogens* **9,** e1003211 (Mar. 2013).

82. 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,** e49–54 (Jan. 2016).

83. Fellay, J. *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science (New York, N.Y.)* **317,** 944–7 (Aug. 2007).

84. Goulder, P. J. & Watkins, D. I. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature reviews. Immunology* **8,** 619–30 (Aug. 2008).

85. Conway, J. M. & Perelson, A. S. Post-treatment control of HIV infection. *Proceedings of the National Academy of Sciences* **112,** 5467–5472 (Apr. 2015).

86. Martin, G. E. *et al.* Post-treatment control or treated controllers? Viral remission in treated and untreated primary HIV infection. *AIDS* **31,** 477–484 (Feb. 2017).

87. Anderson, E. M. & Maldarelli, F. The role of integration and clonal expansion in HIV infection: live long and prosper. *Retrovirology* **15.** <`https://doi.org/10.1186/s12977-018-0448-8`> (Oct. 2018).

88. Dornadula, G. Residual HIV-1 RNA in Blood Plasma of Patients Taking Suppressive Highly Active Antiretroviral Therapy. *JAMA* **282,** 1627 (Nov. 1999).

89. Palmer, S. *et al.* New Real-Time Reverse Transcriptase-Initiated PCR Assay with Single-Copy Sensitivity for Human Immunodeficiency Virus Type 1 RNA in Plasma. *Journal of Clinical Microbiology* **41,** 4531–4536 (Oct. 2003).

90. Maldarelli, F. *et al.* ART Suppresses Plasma HIV-1 RNA to a Stable Set Point Predicted by Pretherapy Viremia. *PLoS Pathogens* **3,** e46 (2007).

91. Van Zyl, G., Bale, M. J. & Kearney, M. F. HIV evolution and diversity in ART-treated patients. *Retrovirology* **15.** <https://doi.org/10.1186/s12977-018-0395-4> (Jan. 2018).

92. Solas, C. *et al.* Discrepancies between Protease Inhibitor Concentrations and Viral Load in Reservoirs and Sanctuary Sites in Human Immunodeficiency Virus-Infected Patients. *Antimicrobial Agents and Chemotherapy* **47,** 238–243 (Jan. 2003).

93. Kearney, M. F. *et al.* Ongoing HIV Replication During ART Reconsidered. *Open Forum Infectious Diseases* **4.** <https://doi.org/10.1093/ofid/ofx173> (2017).

94. Sedaghat, A. R., Siliciano, J. D., Brennan, T. P., Wilke, C. O. & Siliciano, R. F. Limits on Replenishment of the Resting CD4+ T Cell Reservoir for HIV in Patients on HAART. *PLoS Pathogens* **3,** e122 (2007).

95. Joos, B. *et al.* HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proceedings of the National Academy of Sciences* **105,** 16725–16730 (Oct. 2008).

96. Evering, T. H. *et al.* Absence of HIV-1 Evolution in the Gut-Associated Lymphoid Tissue from Patients on Combination Antiviral Therapy Initiated during Primary Infection. *PLoS Pathogens* **8** (ed Malim, M. H.) e1002506 (Feb. 2012).

97. Simonetti, F. R. *et al.* Clonally expanded CD4+T cells can produce infectious HIV-1 in vivo. *Proceedings of the National Academy of Sciences* **113,** 1883–1888 (Feb. 2016).

98. Reeves, D. B. *et al.* A majority of HIV persistence during antiretroviral therapy is due to infected cell proliferation. *Nature Communications* **9.** <https://doi.org/10.1038/s41467-018-06843-5> (Nov. 2018).

99. Chomont, N. *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature Medicine* **15,** 893–900 (June 2009).

100. Maldarelli, F. *et al.* Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345,** 179–183 (June 2014).

101. Wagner, T. A. *et al.* Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345,** 570–573 (July 2014).

102. Vandergeeten, C. *et al.* Interleukin-7 promotes HIV persistence during antiretroviral therapy. *Blood* **121,** 4321–4329 (Apr. 2013).

103. Bosque, A., Famiglietti, M., Weyrich, A. S., Goulston, C. & Planelles, V. Homeostatic Proliferation Fails to Efficiently Reactivate HIV-1 Latently Infected Central Memory CD4+ T Cells. *PLoS Pathogens* **7** (ed Emerman, M.) e1002288 (Oct. 2011).

104. Rasmussen, T. A., Tolstrup, M. & Søgaard, O. S. Reversal of Latency as Part of a Cure for HIV-1. *Trends in Microbiology* **24,** 90–97 (Feb. 2016).

105. J, K. H.-J. & C, B. S.-C. Histone deacetylase inhibitors: molecular mechanisms of action and clinical trials as anti-cancer drugs. *American journal of translational research* **3,** 166–79 (Feb. 2011).

106. Elliott, J. H. *et al.* Activation of HIV Transcription with Short-Course Vorinostat in HIV-Infected Patients on Suppressive Antiretroviral Therapy. *PLoS Pathogens* **10** (ed Cullen, B. R.) e1004473 (Nov. 2014).

107. Søgaard, O. S. *et al.* The Depsipeptide Romidepsin Reverses HIV-1 Latency In Vivo. *PLOS Pathogens* **11** (ed Siliciano, R. F.) e1005142 (Sept. 2015).

108. Wightman, F., Elliot, J., Solomon, A. & Fromentin, R. Multidose vorinostat in HIV-infected individuals on effective ART leads to an increase in regulatory T-cells but no change in in-ducible virus or HIV-specific T-cells. <http://www.iasociety.org/Web/WebContent/File/HIV_Cure_Symposium_2014/Posters/PE38.pdf> (2014).

109. Bullen, C. K., Laird, G. M., Durand, C. M., Siliciano, J. D. & Siliciano, R. F. New ex vivo approaches distinguish effective and ineffective single agents for reversing HIV-1 latency in vivo. *Nature Medicine* **20,** 425–429 (Mar. 2014).

110. Laird, G. M. *et al.* Ex vivo analysis identifies effective HIV-1 latencyreversing drug combinations. *Journal of Clinical Investigation* **125,** 1901–1912 (Mar. 2015).

111. EN, B. AIDS research: charting new directions. *Public Health Reports.* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1424632/> (1984).

112. Rgp120 HIV Vaccine Study Group, T. Placebo-Controlled Phase 3 Trial of a Recombinant Glycoprotein 120 Vaccine to Prevent HIV-1 Infection. *The Journal of Infectious Diseases* **191,** 654–665 (Mar. 2005).

113. Pitisuttithum, P. *et al.* Randomized, Double-Blind, Placebo-Controlled Efficacy Trial of a Bivalent Recombinant Glycoprotein 120 HIV-1 Vaccine among Injection Drug Users in Bangkok, Thailand. *The Journal of Infectious Diseases* **194,** 1661–1671 (Dec. 2006).

114. Rerks-Ngarm, S. *et al.* Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *The New England journal of medicine* **361,** 2209–20 (Dec. 2009).

115. Wang, W. *et al.* CCR5 Gene Disruption via Lentiviral Vectors Expressing Cas9 and Single Guided RNA Renders Cells Resistant to HIV-1 Infection. *PLoS ONE* **9** (ed Jiang, S.) e115987 (Dec. 2014).

116. Li, C. *et al.* Inhibition of HIV-1 infection of primary CD4+ T-cells by gene editing of CCR5 using adenovirus-delivered CRISPR/Cas9. *Journal of General Virology* **96,** 2381–2393 (Aug. 2015).

117. Holt, N. *et al.* Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nature Biotechnology* **28,** 839–847 (July 2010).

118. Spragg, C., Feelixge, H. D. S. & Jerome, K. R. Cell and gene therapy strategies to eradicate HIV reservoirs. *Current Opinion in HIV and AIDS* **11,** 442–449 (July 2016).

119. Bruner, K. M. *et al.* A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* **566,** 120–125 (Jan. 2019).

120. Williams, J. P. *et al.* HIV-1 DNA predicts disease progression and post-treatment virological control. *eLife* **3.** <https://doi.org/10.7554/elife.03821> (Sept. 2014).

121. Wang, Z., Simonetti, F. R., Siliciano, R. F. & Laird, G. M. Measuring replication competent HIV-1: advances and challenges in defining the latent reservoir. *Retrovirology* **15.** <https://doi.org/10.1186/s12977-018-0404-7> (Feb. 2018).

122. "Creative Commons". *Creative commons license for adaptation of Wang et al* Online. Apr. 2019. <https://creativecommons.org/licenses/by/4.0/legalcode>.

123. Finzi, D., Hermankova, M., Pierson, T. & Carruth, L. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. ... <http://science.sciencemag.org/content/278/5341/1295.short> (1997).

124. 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. *Human Retrovirus Protocols: Virology and Molecular Biology,* 3–15 (2005).

125. Chun, T.-W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387,** 183–188 (May 1997).

126. Soriano-Sarabia *et al.* Quantitation of Replication-Competent HIV-1 in Populations of Resting CD4+ T Cells. *Journal of Virology* **88,** 14070–14077 (2014).

127. Procopio, F. *et al.* A Novel Assay to Measure the Magnitude of the Inducible Viral Reservoir in HIV-infected Individuals. *EBioMedicine* **2,** 872–81 (2015).

128. Jones, M. *et al.* Low copy target detection by Droplet Digital PCR through application of a novel open access bioinformatic pipeline, "definetherain". *Journal of virological methods* **202,** 46–53 (June 2014).

129. Strain, M. *et al.* Highly precise measurement of HIV DNA by droplet digital PCR. *PloS one* **8,** e55943 (2013).

130. Malatinkova, E. *et al.* Impact of a decade of successful antiretroviral therapy initiated at HIV-1 seroconversion on blood and rectal reservoirs. *eLife* **4** (2015).

131. Imamichi, H. *et al.* Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proceedings of the National Academy of Sciences* **113,** 8783–8788 (July 2016).

132. Eriksson, S. *et al.* Comparative Analysis of Measures of Viral Reservoirs in HIV-1 Eradication Studies. *PLoS Pathogens* **9,** e1003174 (2013).

133. Bruner, K. M. *et al.* Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nature Medicine* **22,** 1043–1049 (Aug. 2016).

134. Simmonds, P., Balfe, P., Peutherer, J. & of, L. .-.-. C. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *Journal of.* <https://jvi.asm.org/content/64/2/864.short> (1990).

135. Pinheiro, L. B. *et al.* Evaluation of a Droplet Digital Polymerase Chain Reaction Format for DNA Copy Number Quantification. *Analytical Chemistry* **84,** 1003–1011 (Dec. 2011).

136. Hindson, B. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical chemistry* **83,** 8604–10 (2011).

137. Trypsteen, W. & Kiselinova, M. Diagnostic utility of droplet digital PCR for HIV reservoir quantification. *Journal of virus.* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4967968/> (2016).

138. TJ, H., S, G., JZ, L. & F, P. Low-level detection and quantitation of cellular HIV-1 DNA and 2-LTR circles using droplet digital PCR. *Journal of virological.* <https://www.sciencedirect.com/science/article/pii/S0166093412002984> (2012).

139. Jones, M. *et al.* Low copy target detection by Droplet Digital PCR through application of a novel open access bioinformatic pipeline, definetherain. *Journal of virological methods* **202,** 46–53 (June 2014).

140. Trypsteen, W. *et al.* ddpcRquant: threshold determination for single channel droplet digital PCR experiments. *Analytical and Bioanalytical Chemistry* **407,** 5827–5834 (May 2015).

141. Abram, M. E. *et al.* Mutations in HIV-1 Reverse Transcriptase Affect the Errors Made in a Single Cycle of Viral Replication. *Journal of Virology* **88,** 7589–7601 (Apr. 2014).

142. Coffin, J. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* **267,** 483–489 (Jan. 1995).

143. Metzner, K. J. HIV Whole-Genome Sequencing Now: Answering Still-Open Questions. *Journal of Clinical Microbiology* **54** (ed Tang, Y.-W.) 834–835 (Jan. 2016).

144. Nijhuis, M. *et al.* Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy. *Proceedings of the National Academy of Sciences* **95,** 14441–14446 (Nov. 1998).

145. Baxter, A. E., O'Doherty, U. & Kaufmann, D. E. Beyond the replication-competent HIV reservoir: transcription and translation-competent reservoirs. *Retrovirology* **15.** <https://doi.org/10.1186/s12977-018-0392-7> (Feb. 2018).

146. Schroder, A. R. *et al.* HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell* **110,** 521–529 (Aug. 2002).

147. Gonçalves, J. *et al.* Integration of HIV in the Human Genome: Which Sites Are Preferential? A Genetic and Statistical Assessment. *International Journal of Genomics* **2016,** 1–6 (2016).

148. Marini, B. *et al.* Nuclear architecture dictates HIV-1 integration site selection. *Nature* **521,** 227–231 (Mar. 2015).

149. Worobey, M. *et al.* Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455,** 661–664 (Oct. 2008).

150. Jones, B. R. *et al.* Phylogenetic approach to recover integration dates of latent HIV sequences within-host. *Proceedings of the National Academy of Sciences* **115,** E8958–E8967 (Sept. 2018).

151. Whitesides, G. M. The origins and the future of microfluidics. *Nature* **442,** 368–373 (July 2006).

152. Beebe, D. J., Mensing, G. A. & Walker, G. M. Physics and Applications of Microfluidics in Biology. *Annual Review of Biomedical Engineering* **4,** 261–286 (Aug. 2002).

153. AgilentTechnologies. *High-Resolution Automated Electrophoresis of DNA, RNA, and Protein Samples* Online. Apr. 2019. <https://www.agilent.com/en/product/bioanalyzer-automated-electrophoresis>.

154. Panaro, N. J. *et al.* Evaluation of DNA fragment sizing and quantification by the agilent 2100 bioanalyzer. *Clinical chemistry* **46,** 1851–1853 (2000).

155. Fluidigm. *Fluidigm system for rapid, reliable, and cost effective SNP genotyping* Online. Apr. 2019. <https://www.fluidigm.com/binaries/content/documents/fluidigm/marketing/fluidigm-system-for-rapid-reliable-and-cost-effective-genotyping>.

156. Simpson, C., Lee, S. S., Lee, C.-S. & Yamauchi, Y. Microfluidics: an Untapped Resource in Viral Diagnostics and Viral Cell Biology. *Current Clinical Microbiology Reports* **5,** 245–251 (Dec. 1, 2018).

157. Duncombe, T. A., Tentori, A. M. & Herr, A. E. Microfluidics: reframing biological enquiry. *Nature Reviews Molecular Cell Biology* **16,** 554–567 (Sept. 2015).

158. Jacobs, K. Introduction to Microfluidics. By Patrick Tabeling. *Angewandte Chemie International Edition* **45,** 7875–7875 (Dec. 2006).

159. Sharp, K., Adrian, R., Santiago, J. & Molho, J. English (US). in *MEMS* 10–1–10–46 (CRC Press, Jan. 1, 2005).

160. Raamanand, C. *Stabilising Emulsions Using Food Grade Complexes in Microfluidics Devices* PhD thesis (University of Oxford, 2013).

161. *Emulsions* (ed Schramm, L. L.) <https://doi.org/10.1021/ba-1992-0231> (American Chemical Society, May 1992).

162. Fluidigm. *Fluidigm spotlights* Online. Apr. 2019. <https://www.fluidigm.com/spotlights#8cc9e1eb-d5e6-40b6-8c38-7a65cab18638>.

163. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics* **17,** 175–188 (Jan. 2016).

164. Bedard, P. L., Hansen, A. R., Ratain, M. J. & Siu, L. L. Tumour heterogeneity in the clinic. *Nature* **501,** 355–364 (Sept. 2013).

165. Pellegrino, M. *et al.* High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. *Genome Research* **28,** 1345–1352 (Aug. 2018).

166. Eastburn, D. J., Sciambi, A. & Abate, A. R. Ultrahigh-Throughput Mammalian Single-Cell Reverse-Transcriptase Polymerase Chain Reaction in Microfluidic Drops. *Analytical Chemistry* **85,** 8016–8021 (Aug. 2013).

167. Kintses, B. *et al.* Picoliter Cell Lysate Assays in Microfluidic Droplet Compartments for Directed Enzyme Evolution. *Chemistry & Biology* **19,** 1001–1009 (Aug. 2012).

168. Rane, T. D., Zec, H. C., Puleo, C., Lee, A. P. & Wang, T.-H. Droplet microfluidics for amplification-free genetic detection of single cells. *Lab on a Chip* **12,** 3341 (2012).

169. Yucha, R. W. *et al.* High-throughput Characterization of HIV-1 Reservoir Reactivation Using a Single-Cell-in-Droplet PCR Assay. *EBioMedicine* **20,** 217–229 (June 2017).

170. Zhu, Z. *et al.* Highly sensitive and quantitative detection of rare pathogens through agarose droplet microfluidic emulsion PCR at the single-cell level. *Lab on a Chip* **12,** 3907 (2012).

171. Samorodnitsky, E. *et al.* Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Human Mutation* **36,** 903–914 (July 2015).

172. Bonsall, D. *et al.* ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *F1000Research* **4,** 1062 (Oct. 2015).

173. Kebschull, J. M. & Zador, A. M. Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Research,* gkv717 (July 2015).

174. Aigrain, L., Gu, Y. & Quail, M. A. Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing. *BMC Genomics* **17.** <https://doi.org/10.1186/s12864-016-2757-4> (June 2016).

175. Wilson, R. H. *et al.* Engineered DNA ligases with improved activities in vitro. *Protein Engineering Design and Selection* **26,** 471–478 (June 2013).

176. Owczarzy, R., Moreira, B. G., You, Y., Behlke, M. A. & Walder, J. A. Predicting Stability of DNA Duplexes in Solutions Containing Magnesium and Monovalent Cations. *Biochemistry* **47,** 5336–5353 (May 2008).

177. Desai, A. *et al.* Identification of Optimum Sequencing Depth Especially for De Novo Genome Assembly of Small Genomes Using Next Generation Sequencing Data. *PLoS ONE* **8** (ed Zhang, S.-D.) e60204 (Apr. 2013).

178. Institute, B. *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM* Online. Jan. 2013. <https://arxiv.org/abs/1303.3997>.

179. HIV.gov. *High-Resolution Automated Electrophoresis of DNA, RNA, and Protein Samples* Online. Jan. 2019. <http://www.novocraft.com/products/novoalign/>.

180. Sunshine, S. *et al.* HIV Integration Site Analysis of Cellular Models of HIV Latency with a Probe-Enriched Next-Generation Sequencing Assay. *Journal of virology* **90,** 4511–9 (May 2016).

181. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29,** 24–26 (Jan. 2011).

182. Park, G. *et al.* Characterization of background noise in capture-based targeted sequencing data. *Genome Biology* **18.** <https://doi.org/10.1186/s13059-017-1275-2> (July 2017).

183. Chung, J. *et al.* The minimal amount of starting DNA for Agilent's hybrid capture-based targeted massively parallel sequencing. *Scientific Reports* **6.** <https://doi.org/10.1038/srep26732> (May 2016).

184. Bodi, K. *et al.* Comparison of Commercially Available Target Enrichment Methods for NextGeneration Sequencing. *Journal of Biomolecular Techniques* **24,** 73–86 (July 2013).

185. Bernath, K. *et al.* In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Analytical Biochemistry* **325,** 151–157 (Feb. 2004).

186. Clark, M. J. *et al.* Performance comparison of exome DNA sequencing technologies. *Nature Biotechnology* **29,** 908–914 (Sept. 2011).

187. Zheng, Z. *et al.* Titration-free massively parallel pyrosequencing using trace amounts of starting material. *Nucleic Acids Research* **38,** e137–e137 (Apr. 2010).

188. Kang, W.-J., Rabe, E., Kopetz, S. & Neyer, A. Novel exposure methods based on reflection and refraction effects in the field of SU-8 lithography. *Journal of Micromechanics and Microengineering* **16,** 821–831 (Mar. 2006).

189. Tang, K. C. *et al.* Evaluation of bonding between oxygen plasma treated polydimethyl siloxane and passivated silicon. *Journal of Physics: Conference Series* **34,** 155–161 (Apr. 2006).

190. Liston, E., Martinu, L. & Wertheimer, M. Plasma surface modification of polymers for improved adhesion: a critical review. *Journal of Adhesion Science and Technology* **7,** 1091–1127 (Jan. 1993).

191. Roumpea, E. *et al.* Experimental studies on droplet formation in a flow-focusing microchannel in the presence of surfactants. *Chemical Engineering Science* **195,** 507–518 (Feb. 2019).

192. Christopher, G. F., Noharuddin, N. N., Taylor, J. A. & Anna, S. L. Experimental observations of the squeezing-to-dripping transition in T-shaped microfluidic junctions. *Physical Review E* **78.** <https://doi.org/10.1103/physreve.78.036317> (Sept. 2008).

193. Fu, T., Ma, Y., Funfschilling, D. & Li, H. Z. Bubble formation and breakup mechanism in a microfluidic flow-focusing device. *Chemical Engineering Science* **64,** 2392–2400 (May 2009).

194. Thorsen, T., Roberts, R. W., Arnold, F. H. & Quake, S. R. Dynamic Pattern Formation in a Vesicle-Generating Microfluidic Device. *Physical Review Letters* **86,** 4163–4166 (Apr. 2001).

195. Cubaud, T., Tatineni, M., Zhong, X. & Ho, C.-M. Bubble dispenser in microfluidic devices. *Physical Review E* **72.** <https://doi.org/10.1103/physreve.72.037302> (Sept. 2005).

196. Fu, T., Wu, Y., Ma, Y. & Li, H. Z. Droplet formation and breakup dynamics in microfluidic flow-focusing devices: From dripping to jetting. *Chemical Engineering Science* **84,** 207–217 (Dec. 2012).

197. Nisisako, T., Torii, T. & Higuchi, T. Droplet formation in a microchannel network. *Lab Chip* **2,** 24–26 (2002).

198. Tarchichi, N., Chollet, F. & Manceau, J.-F. New regime of droplet generation in a T-shape microfluidic junction. *Microfluidics and Nanofluidics* **14,** 45–51 (July 2012).

199. Seemann, R., Brinkmann, M., Pfohl, T. & Herminghaus, S. Droplet based microfluidics. *Reports on Progress in Physics* **75,** 016601 (Dec. 2011).

200. Tran, T. M., Lan, F., Thompson, C. S. & Abate, A. R. From tubes to drops: droplet-based microfluidics for ultrahigh-throughput biology. *Journal of Physics D: Applied Physics* **46,** 114004 (Feb. 2013).

201. Fritz, J. L. & Owen, M. J. Hydrophobic Recovery of Plasma-Treated Polydimethylsiloxane. *The Journal of Adhesion* **54,** 33–45 (Nov. 1995).

202. Zhang, C. & Xing, D. Single-Molecule DNA Amplification and Analysis Using Microfluidics. *Chemical Reviews* **110,** 4910–4947 (Aug. 2010).

203. Holtze, C. *et al.* Biocompatible surfactants for water-in-fluorocarbon emulsions. *Lab on a Chip* **8,** 1632 (2008).

204. Jeon, S., Lee, J., Andrade, J. & Gennes, P. D. Proteinsurface interactions in the presence of polyethylene oxide. *Journal of Colloid and Interface Science* **142,** 149–158 (Mar. 1991).

205. Cchová, M., Proková, M., Tóthová, L., Sántha, H. & Mayer, V. On-line cell lysis of bacteria and its spores using a microfluidic biochip. *Open Life Sciences* **7.** <https://doi.org/10.2478/s11535-012-0005-8> (Jan. 2012).

206. Abolmaaty, A., El-Shemy, M., Khallaf, M. & Levin, R. Effect of lysing methods and their variables on the yield of Escherichia coli O157: H7 DNA and its PCR amplification. *Journal of Microbiological Methods* **34,** 133–141 (Nov. 1998).

207. Lee, H. J. *et al.* Electrochemical cell lysis device for DNAextraction. *Lab Chip* **10,** 626–633 (2010).

208. Gu, H., Duits, M. H. G. & Mugele, F. Droplets Formation and Merging in Two-Phase Flow Microfluidics. *International Journal of Molecular Sciences* **12,** 2572–2597 (Apr. 2011).

209. Bremond, N., Thiam, A. R. & Bibette, J. Decompressing Emulsion Droplets Favors Coalescence. *Physical Review Letters* **100.** <https://doi.org/10.1103/physrevlett.100.024501> (Jan. 2008).

210. Harpur, I. *et al.* Destabilisation of water-in-oil emulsions under the influence of an A.C. electric field: Experimental assessment of performance. *Journal of Electrostatics* **40-41,** 135–140 (June 1997).

211. Abate, A. R., Hung, T., Mary, P., Agresti, J. J. & Weitz, D. A. High-throughput injection with microfluidics using picoinjectors. *Proceedings of the National Academy of Sciences* **107,** 19163–19166 (Oct. 2010).

212. Eastburn, D. J., Sciambi, A. & Abate, A. R. Identification and genetic analysis of cancer cells with PCR-activated cell sorting. *Nucleic Acids Research* **42,** e128–e128 (July 2014).

213. Lim, S. W., Tran, T. M. & Abate, A. R. PCR-Activated Cell Sorting for Cultivation-Free Enrichment and Sequencing of Rare Microbes. *PLOS ONE* **10** (ed Jayaraman, A.) e0113549 (Jan. 2015).

214. Wu, L., Chen, P., Dong, Y., Feng, X. & Liu, B.-F. Encapsulation of single cells on a microfluidic device integrating droplet generation with fluorescence-activated droplet sorting. *Biomedical Microdevices* **15,** 553–560 (Feb. 2013).

215. Eastburn, D. J. *et al.* Microfluidic droplet enrichment for targeted sequencing. *Nucleic Acids Research* **43,** e86–e86 (Apr. 2015).

216. Xi, H.-D. *et al.* Active droplet sorting in microfluidics: a review. *Lab on a Chip* **17,** 751–771 (2017).

217. Mazutis, L. *et al.* Single-cell analysis and sorting using droplet-based microfluidics. *Nature Protocols* **8,** 870–891 (Apr. 2013).

218. Zworykin, V., Morton, G. & Malter, L. The Secondary Emission Multiplier-A New Electronic Device. *Proceedings of the IRE* **24,** 351–375 (Mar. 1936).

219. Lubsandorzhiev, B. On the history of photomultiplier tube invention. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **567,** 236–238 (Nov. 2006).

220. Prashant, P. *Fluorescence Microscopy: Optimizing the imaging performance of LED light engines* Online. Apr. 2019. <`https://www.bioopticsworld.com/articles/print/volume-8/issue-3/features/fluorescence-microscopy-optimizing-the-imaging-performance-of-led-light-engines.html`>.

221. Ahn, K. *et al.* Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices. *Applied Physics Letters* **88,** 024104 (Jan. 2006).

222. Caen, O. *et al.* High-throughput multiplexed fluorescence-activated droplet sorting. *Microsystems & Nanoengineering* **4.** <`https://doi.org/10.1038/s41378-018-0033-2`> (Oct. 2018).

223. Godwin, L. A., Deal, K. S., Hoepfner, L. D., Jackson, L. A. & Easley, C. J. Measurement of microchannel fluidic resistance with a standard voltage meter. *Analytica Chimica Acta* **758,** 101–107 (Jan. 2013).

224. Bradbury, S. L. & Jakoby, W. B. Glycerol as an Enzyme-Stabilizing Agent: Effects on Aldehyde Dehydrogenase. *Proceedings of the National Academy of Sciences* **69,** 2373–2376 (1972).

225. Alberta, B. *et al.* Essential Cell Biology. Fourth Edition. By Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. New York: Garland Science (Taylor & Francis Group). 151.00. xxiii + 726 p. + A:1 - A:58: G:1 - G:26: I:1 - I:28 (index). ISBN: 978-0-8153-4454-4. 2014. *The Quarterly Review of Biology* **90,** 343–343 (Sept. 2015).

226. Watson, M. L. The nuclear envelope: its structure and relation to cytoplasmic membranes. *The Journal of Cell Biology* **1,** 257–270 (May 1955).

227. Hetzer, M. W. The Nuclear Envelope. *Cold Spring Harbor Perspectives in Biology* **2,** a000539–a000539 (Feb. 2010).

228. Islam, M., Aryasomayajula, A. & Selvaganapathy, P. A review on macroscale and microscale cell lysis methods. *Micromachines* **8,** 83 (2017).

229. Mótyán, J., Tóth, F. & Tzsér, J. Research Applications of Proteolytic Enzymes in Molecular Biology. *Biomolecules* **3,** 923–942 (Nov. 2013).

230. Lorenz, T. C. Polymerase Chain Reaction: Basic Protocol Plus Troubleshooting and Optimization Strategies. *Journal of Visualized Experiments.* <https://doi.org/10.3791/3998> (May 2012).

231. Aldrich, S. *qPCR technical guide* Online. Apr. 2017. <https://www.gene-quantification.de/SIAL-qPCR-Technical-Guide.pdf>.

232. Folks, T. M. Biological and biochemical characterization of a cloned Leu-3- cell surviving infection with the acquired immune deficiency syndrome retrovirus. *Journal of Experimental Medicine* **164,** 280–290 (July 1986).

233. Harada, S., Koyanagi, Y. & Yamamoto, N. Infection of HTLV-III/LAV in HTLV-I-carrying cells MT-2 and MT-4 and application in a plaque assay. *Science* **229,** 563–566 (1985).

234. Ogunshola, F. *et al.* P15.03 Development of a new cem reporter t-cells (gxr-cells) viral inhibition assay (via) for elucidating the role of class-i-hla alleles on the inhibitory capacity of hiv-1-specific cd8+t-cells. *Sexually Transmitted Infections* **91,** A210–A210 (2015).

235. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2017). <https://www.R-project.org/>.

236. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* <http://ggplot2.org> (Springer-Verlag New York, 2016).

237. *Los AlamosNational Database* Apr. http://www.hiv.lanl.gov/content/index. <http://www.hiv.lanl.gov/content/index>.

238. "SPARTAC trial Investigators". Short-Course Antiretroviral Therapy in Primary HIV Infection. *New England Journal of Medicine* **368,** 207–217 (Jan. 2013).

Chapter 9

Full length contig reconstruction

9.1 Full-length contig reconstruction

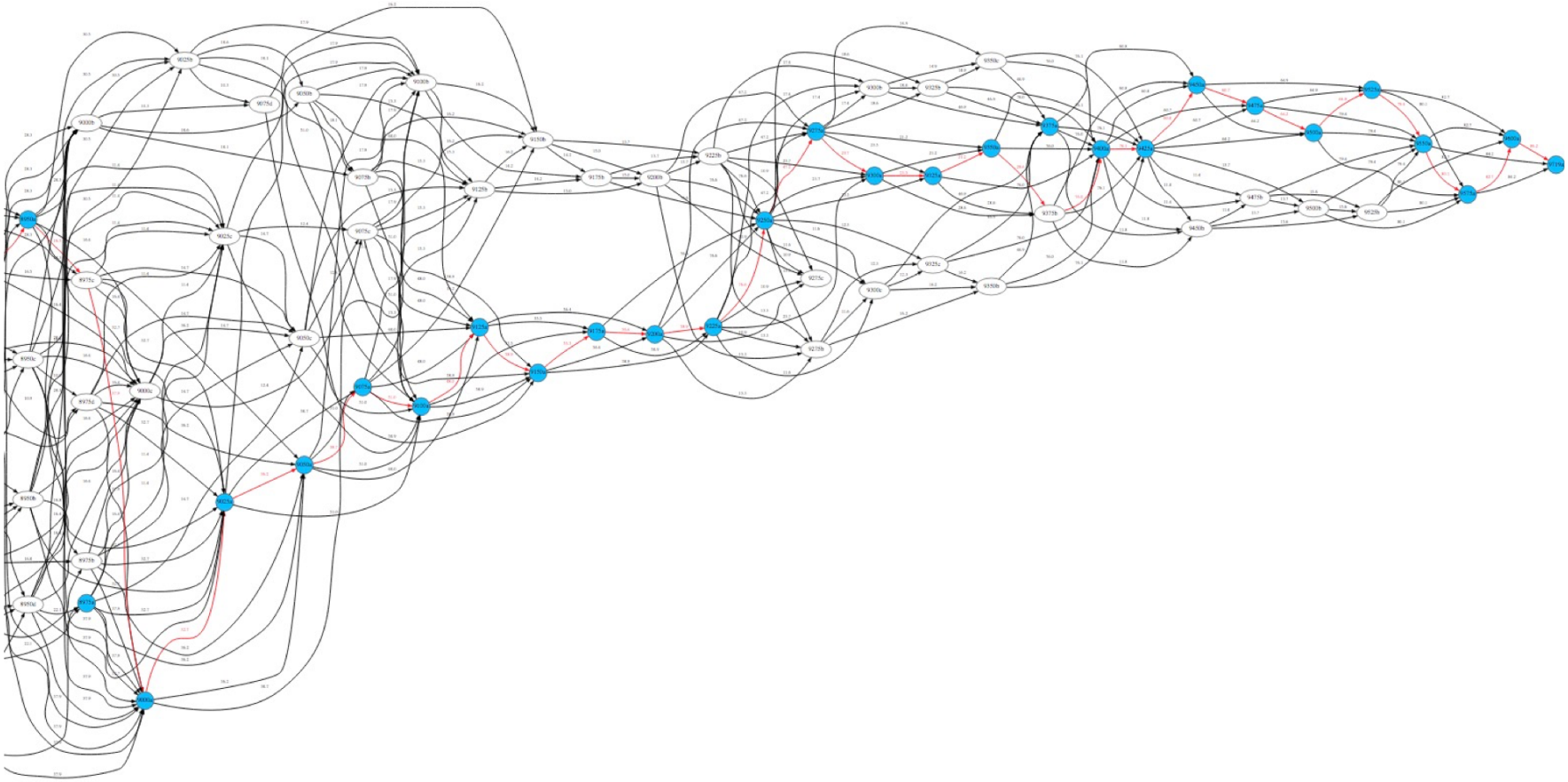


Figure 9.1: Full-length contig - view 1 Example reconstruction of the HIV genome from one Sureselect sequenced sample. Blue nodes indicate the concensus path with the greatest confidence. White nodes indicate alternate paths based on coverage data and minority variants at each position.

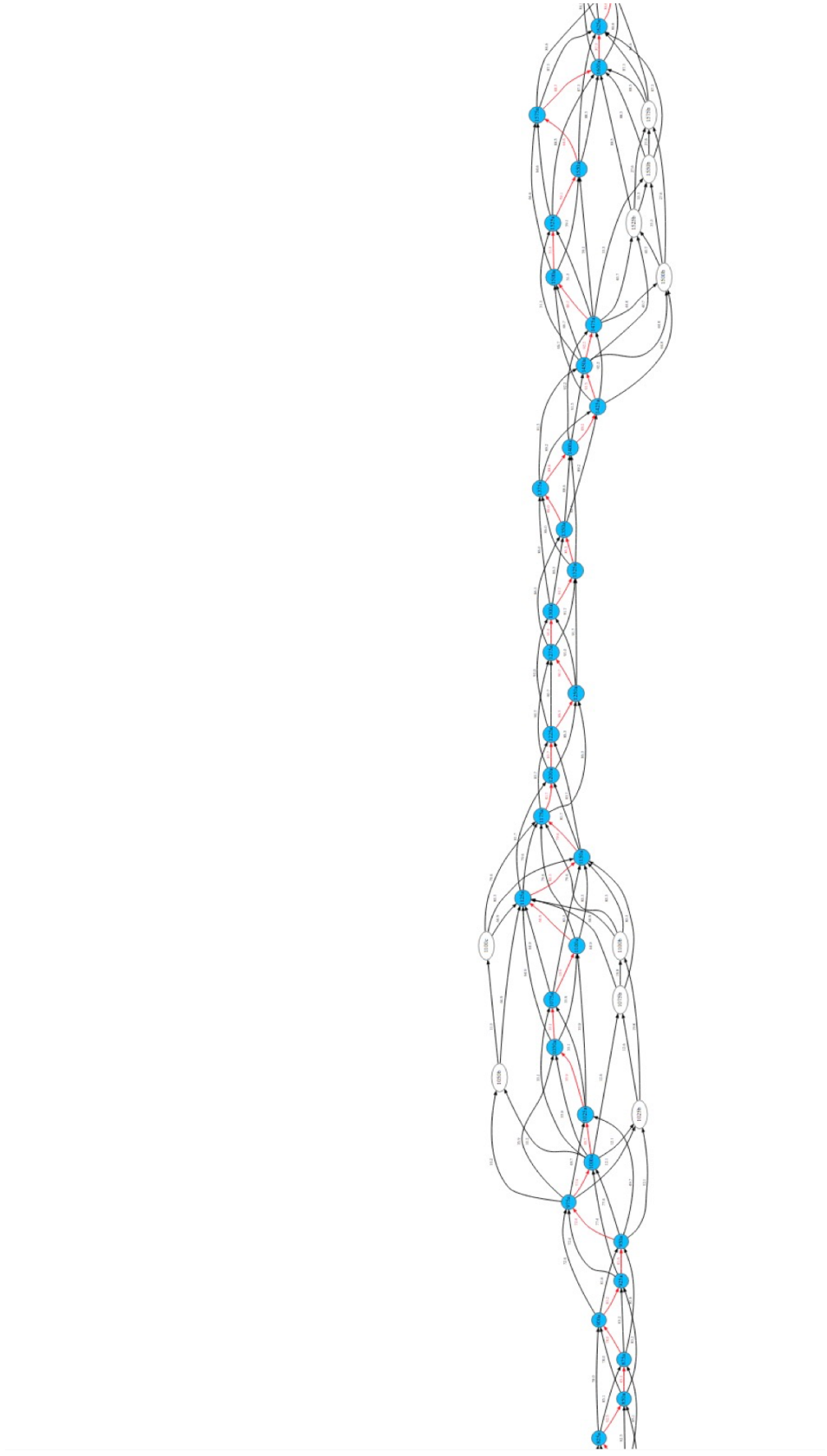


Figure 9.2: Full-length contig - view 2

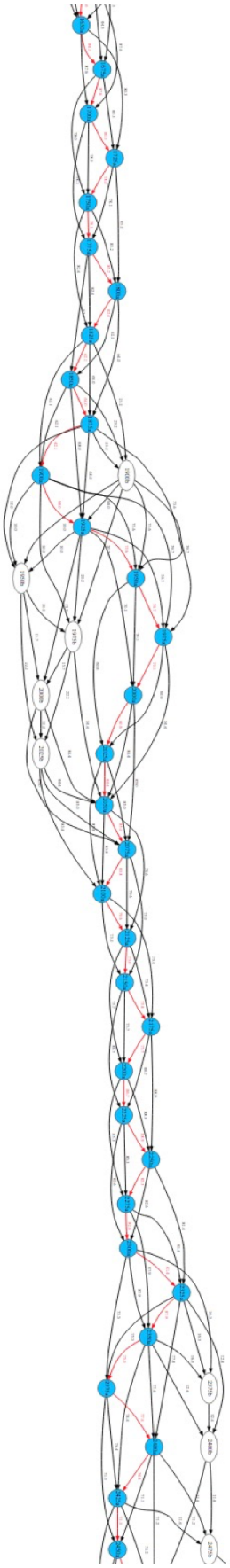


Figure 9.3: Full-length contig - view 3

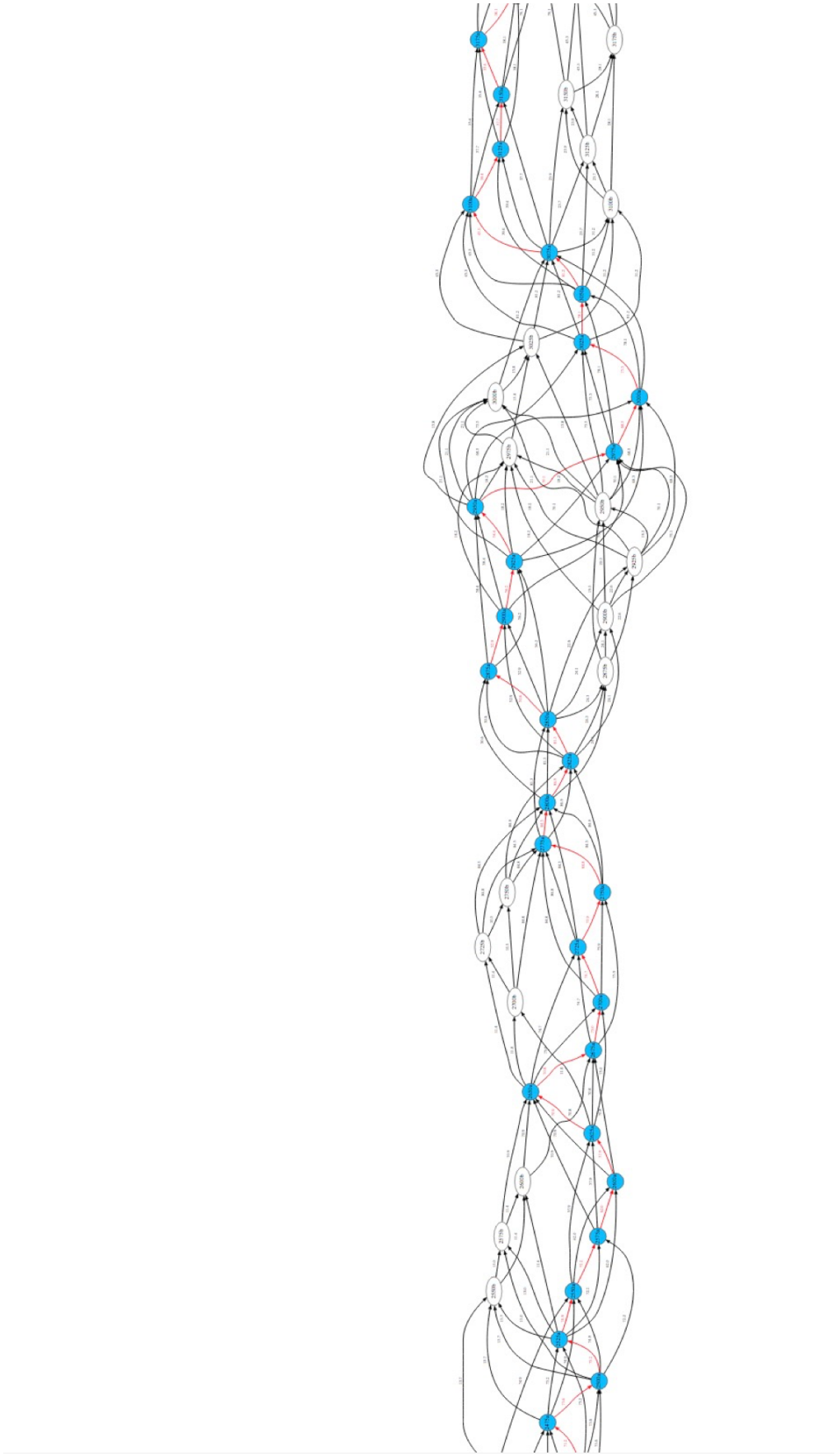


Figure 9.4: Full-length contig - view 4

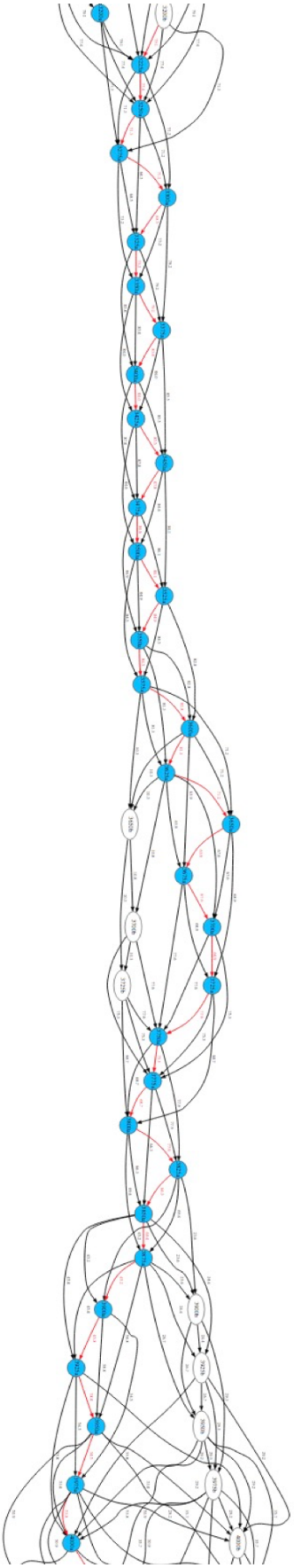


Figure 9.5: Full-length contig - view 5

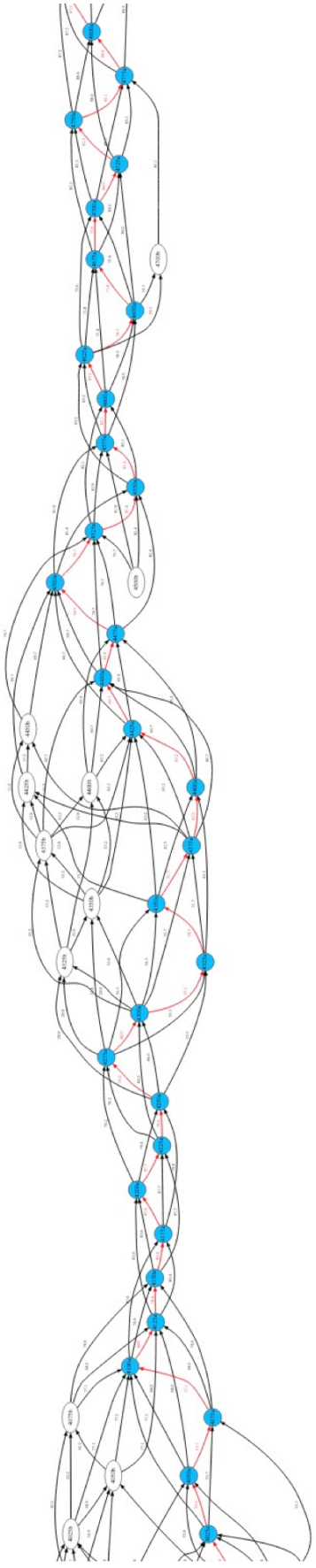


Figure 9.6: Full-length contig - view 6

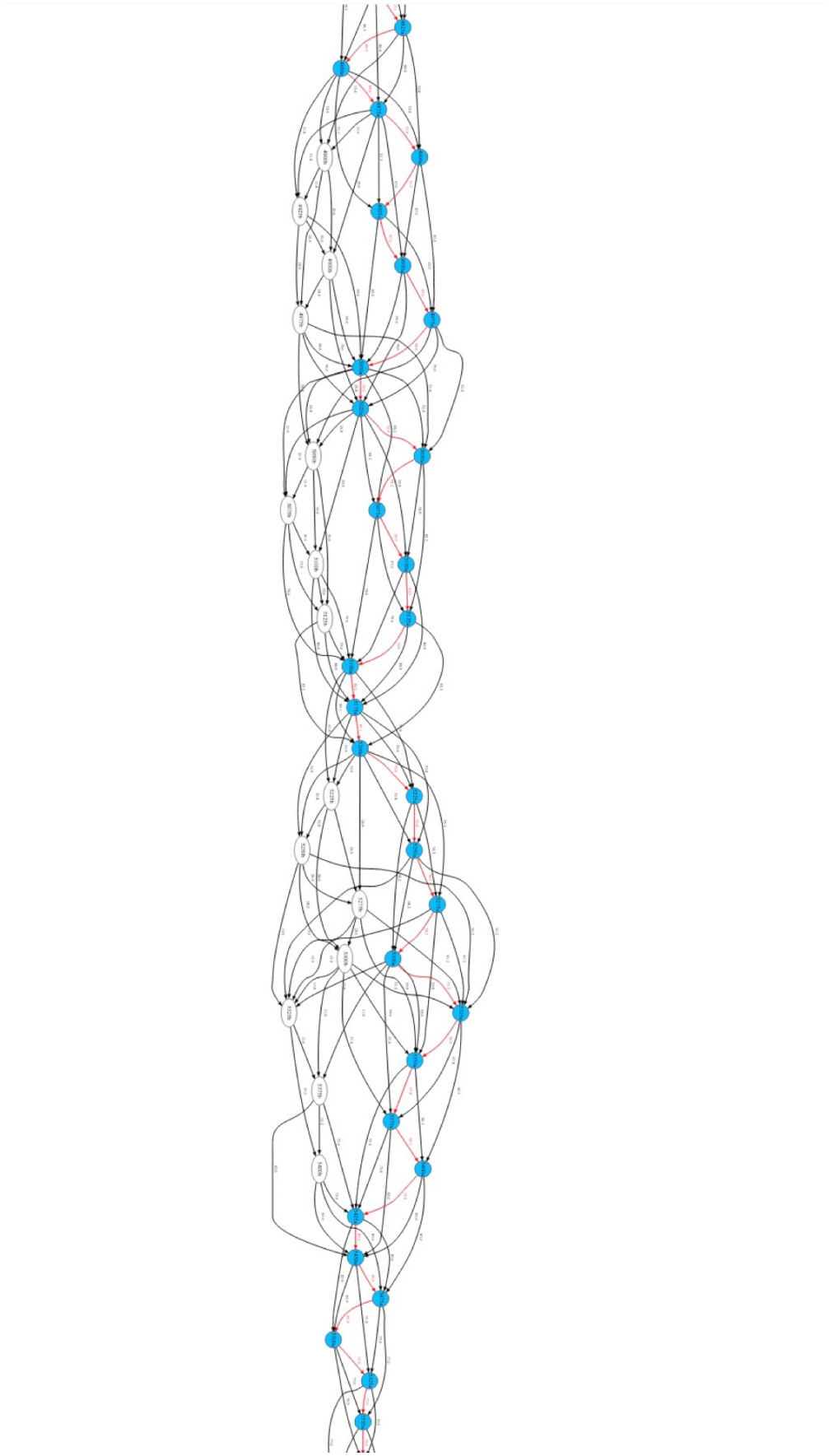


Figure 9.7: Full-length contig - view 7

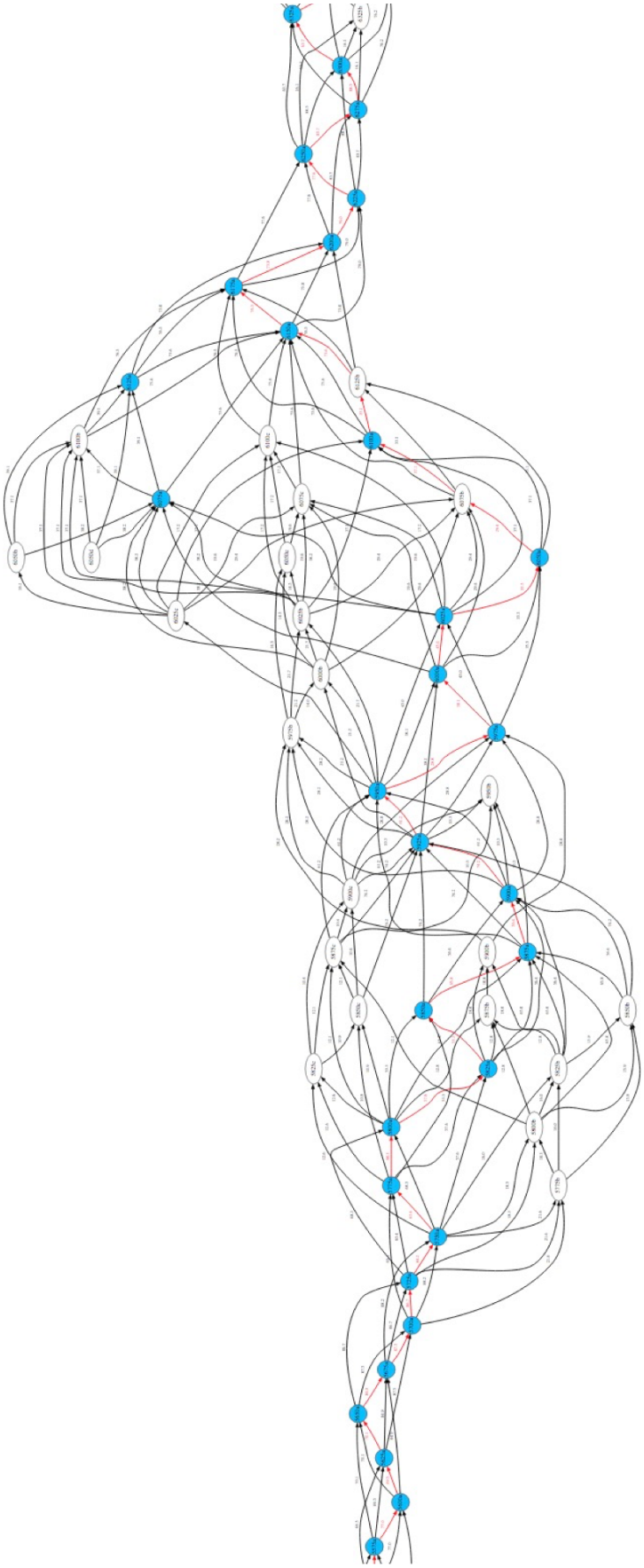


Figure 9.8: Full-length contig - view 8

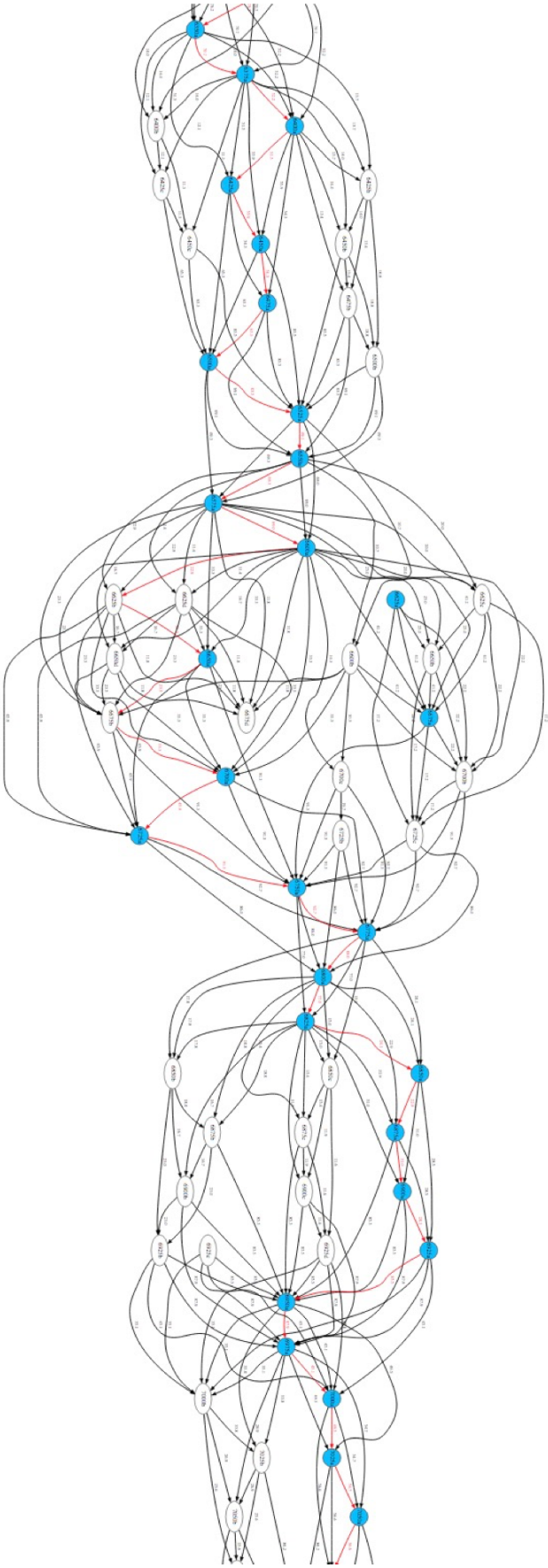


Figure 9.9: Full-length contig - view 9

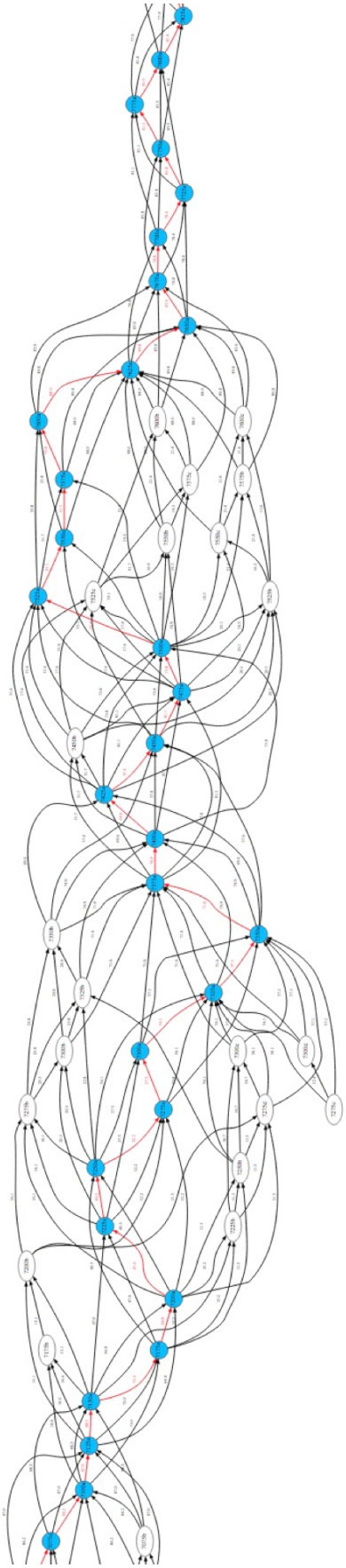


Figure 9.10: Full-length contig - view 10

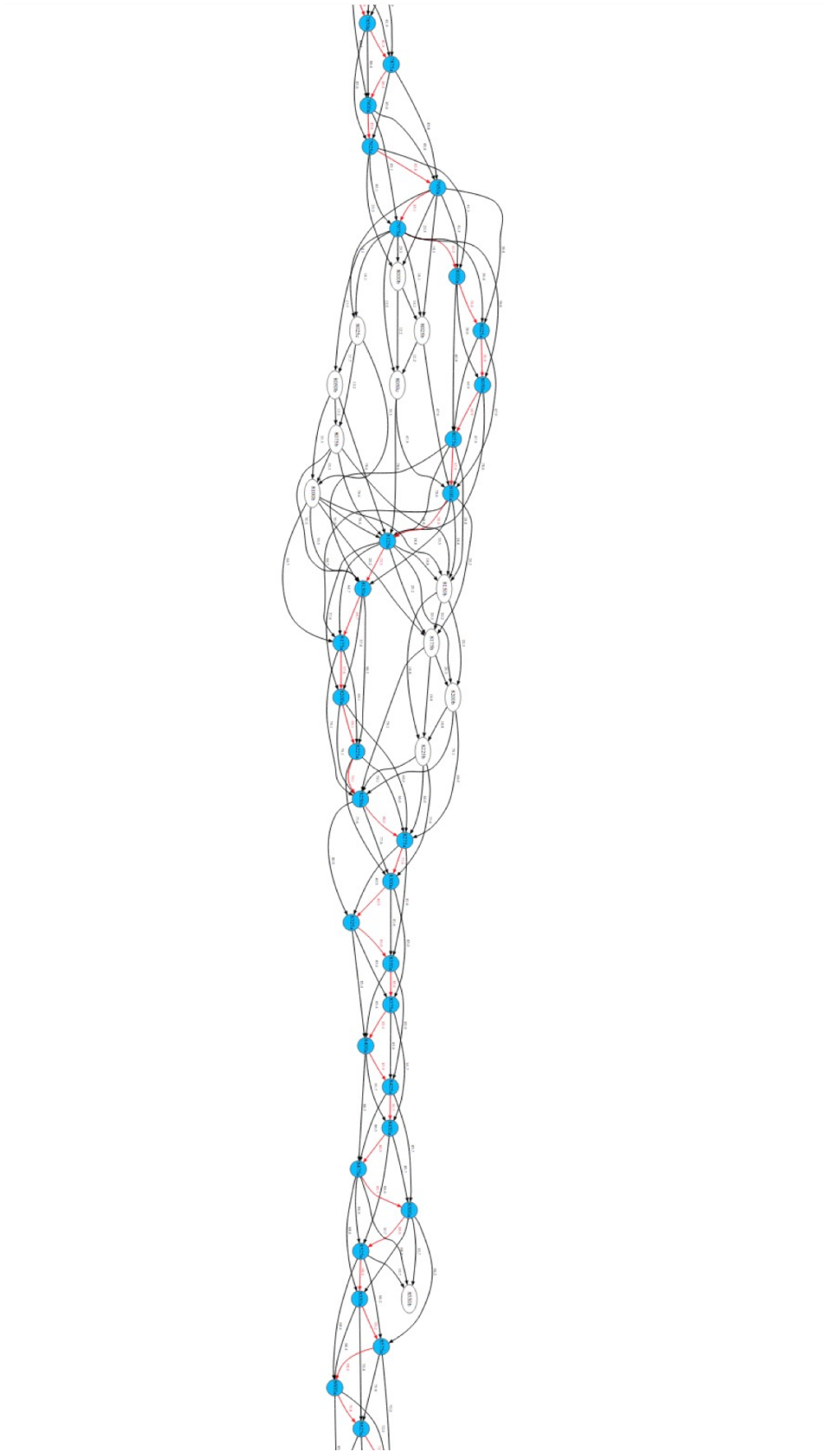


Figure 9.11: Full-length contig - view 11

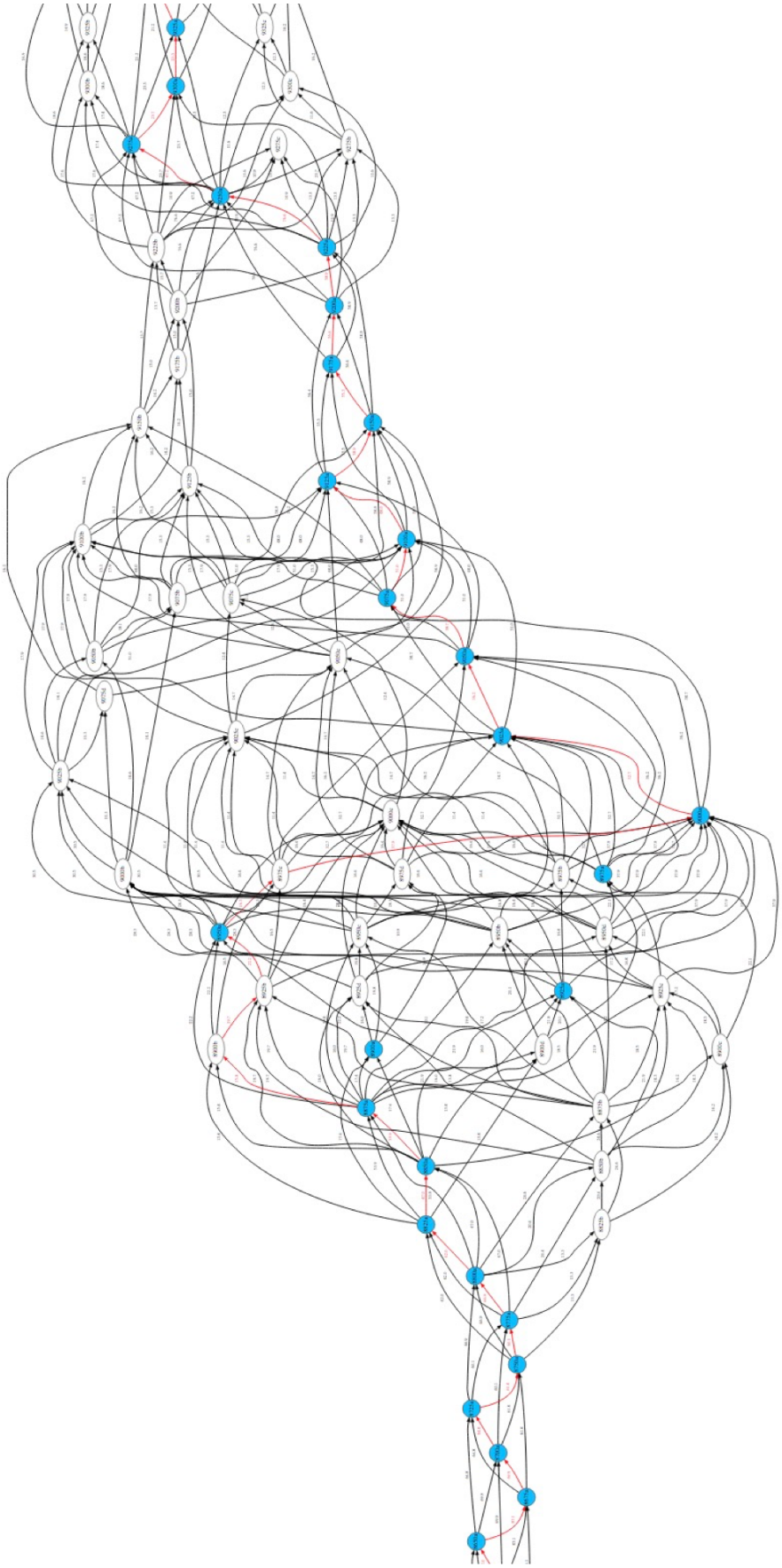


Figure 9.12: Full-length contig - view 12

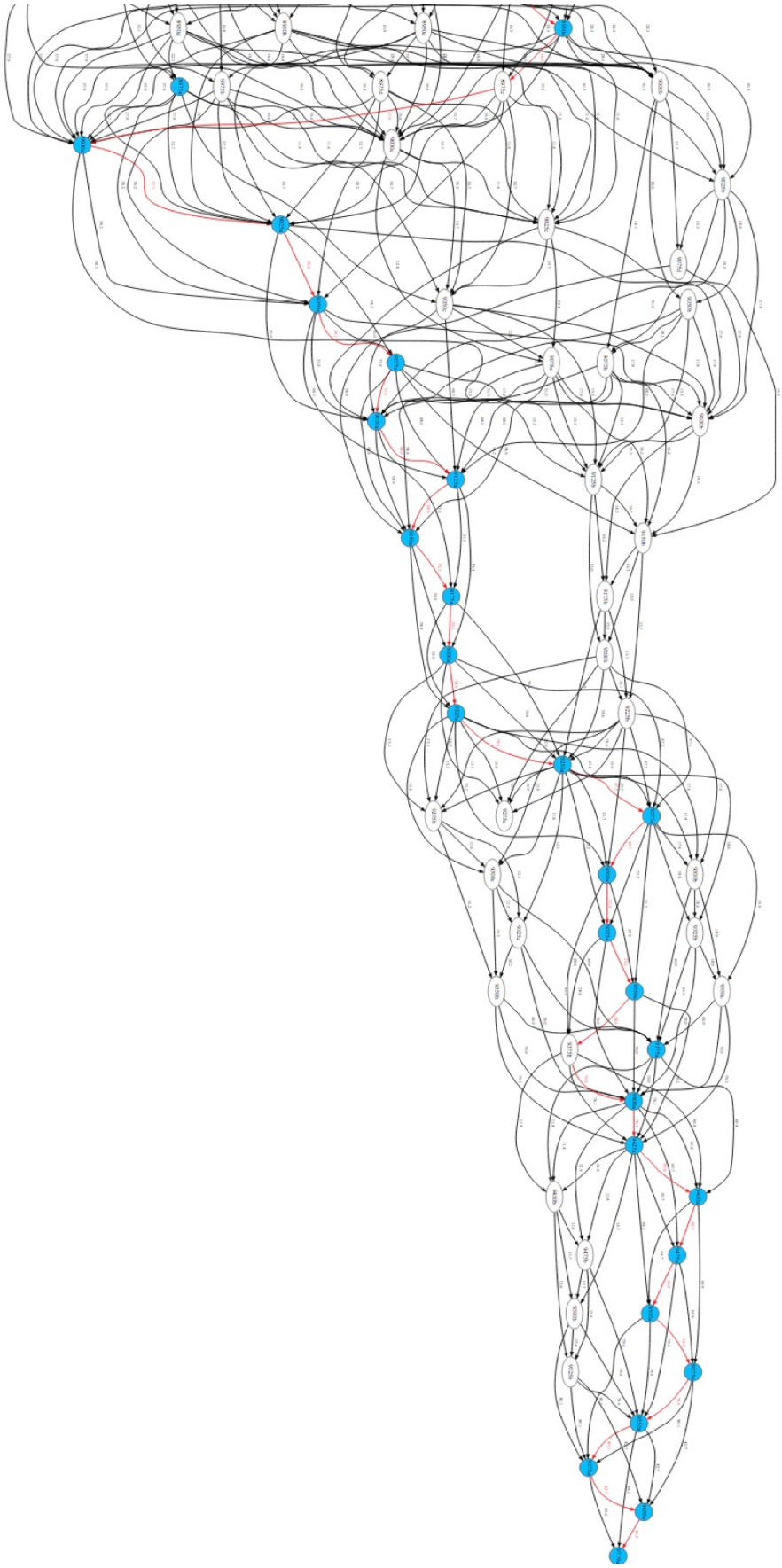


Figure 9.13: Full-length contig - view 13

Chapter 10

Copyright permission

Contents

9.1 Full-length contig reconstruction 223

10.1 Springer Nature copyright permission

This Agreement between Mathew Jones ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4561890846041
License date	Apr 04, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Medicine
Licensed Content Title	Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4 T cells
Licensed Content Author	Janet D Siliciano, Joleen Kajdas, Diana Finzi, Thomas C Quinn, Karen Chadwick et al.
Licensed Content Date	May 18, 2003
Licensed Content Volume	9
Licensed Content Issue	6
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Novel Methods Towards HIV Eradication: A combined microfluidic and target enrichment approach
Institution name	University of Oxford
Expected presentation date	May 2019
Portions	Figure 1.a