Using the natural variation in iPSCs reprogrammed from HIV-1 infected patients to define HIV-1 regulatory network

Acors, Sam

Awarding institution:
King's College London

Using the natural variation in iPSCs reprogrammed from HIV-1 infected patients to define HIV-1 regulatory networks

Submitted for the degree of Doctor of Philosophy by

Sam James Acors

Department of Infectious Diseases
School of Immunology & Microbial Sciences
King's College London

January 2024

The work presented in this thesis is my own

Acknowledgements

First and foremost, I would like to thank my primary supervisor, Michael Malim, for his support and guidance over the course of my PhD. His mentorship, both scientifically and personally, has been outstanding and greatly appreciated. I would also like to thank my secondary supervisor, Davide Danovi, and my day-to-day point of contact, Luis Apolonia, for providing insightful ideas and continuingly responding to my many queries.

I would also like to thank my thesis committee members, Malcolm Logan and Subhankar Mukhopadhyay, for helping direct and elevate my project.

This project would not have been possible without the collaboration and support from Northwestern University, particularly Steven Wolinsky, Eun-Young Kim, and Patricia Otto.

It has been a pleasure to work with everyone in the Malim lab and the Department of Infectious Diseases, both past and present. A special thank you goes to Nathalia Almeida dos Santos, who not only assisted in the analysis of some of my data, but was also an invaluable friend throughout my PhD, alongside Adrian Signell, Daniel Cox, and Hataf Kahn. I would also like to thank Stelios Papaioannou for support for all things laboratory; Neophytos Kouphou for developing the iPSC infection pipeline; members of the Stem Cell Hotel, particularly Thomas Williams and Lazaros Fotopoulos, for their immense knowledge of the Operetta CLS, and Robert Carton for his assistance in processing the raw RNA sequencing data.

I would also like to express my gratitude to the Medical Research Council and Orchard Therapeutics for funding my PhD, and to Bobby Gaspar and Pervinder Sagoo who inspired me to undertake this PhD journey.

Finally, I would like to thank my family and friends. To Trevor and Sarah for being the best parents a son could ask for; to Dani and Abi for continuingly pushing me further to remain the favourite child; to Dexie and Sully for always being there when a fluffy cuddle was needed, and to all my friends who continue to make my life an exciting adventure.

Abstract

Human immunodeficiency virus type 1 (HIV-1) is the causative agent of a progressive failure in the host immune system, leading to acquired immunodeficiency syndrome (AIDS). Antiretroviral therapies effectively treat the disease but fail to eradicate the virus from the patient, presenting the need for further knowledge of virus-host interactions to reach the ultimate goal of a vaccine or cure. Previous work has either taken binary approaches, involving the knockdown or overexpression of specific genes, to finding host-encoded dependency and restriction factors that interact with the virus and modulate replication, or have sought associated genetic polymorphisms (e.g. *HLA-B* and *CCR5*) that correlate with outcomes of natural infection.

This project utilises the natural genetic variation between patients to identify cellular HIV-1 regulatory factors by conducting a permissivity screen in characterised induced pluripotent stem cells (iPSCs) reprogrammed from samples of peripheral blood mononuclear cells (PBMCs) from the multicentre AIDS cohort study (MACS) – a prospective study of men at high risk of HIV-1 infection whose clinical histories have been described in great detail. The consistency in generation and culture of the iPSCs means that differences between iPSC lines are predominantly attributable to the genetic variation of the patient from which they were derived. This, along with their ability to self-renew, makes iPSCs an excellent resource to utilise for a permissivity screen and has the potential to result in more reproducible permissivity phenotypes when compared to the use of other cell types such as PBMCs from a patient's peripheral blood. Clinical data describing disease outcomes of patients in the MACS can then be evaluated in the context of *in vitro* iPSC permissivity phenotypes, genome sequence and transcriptomic profiles.

An optimised reprogramming pipeline was established and enabled the successful generation of 50 MACS-derived iPSC lines from 18 MACS participants who demonstrated extreme HIV-1 disease progression clinical phenotypes. The *in vitro* permissivity phenotypes of the iPSCs were then identified and comparative transcriptomic analysis of those exhibiting extreme opposite phenotypes allowed for the identification of cellular factors, pathways and networks

that possibly influence HIV-1 infection and pathogenesis. Numerous candidate genes were identified and require validation of their influence on HIV-1 infection as either potential restriction or dependency factors. As well as suggesting novel host factors, the work in this thesis demonstrates a functioning pipeline that utilises host natural variation to identify host factors involved in HIV-1 infection that can serve as a foundation for future work. Defining new viral regulatory networks will build on our understanding of HIV-1-host interactions whilst providing clues to potential therapeutic targets.

COVID-19 Impact Statement

As with many recent post-graduate students, my PhD timeline was severely impacted by the COVID-19 pandemic; as an estimate, ~9 months of research time was disrupted. Hence, I present this statement which contains information on the disruptions faced as well as the research I contributed to during this time that is not shown in this thesis.

In March 2020, as a result of the United Kingdom's national lockdown in response to the COVID-19 pandemic, all HIV-related lab-based research was paused. Having only commenced my PhD in October 2019 and with the wet-lab heavy nature of my project, my work from home capacity was very limited. Thus, during this time, I was fortunate to receive the opportunity to second in Professor Katie Doores' lab, supporting their research on the antibody response to the causative agent of the COVID-19 pandemic, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Over the next series of months, we developed and employed assays for quantitatively assessing the binding and neutralisation capacities of antibodies produced by patients infected with SARS-CoV-2. Over this period of time I also volunteered at the Centre for Clinical Infection and Diagnostics Research at St Thomas' hospital where I assisted the banking of clinical samples. As a result of these two secondments, I contributed to several peer-reviewed scientific publications, including one co-first author article (Carter et al., 2020; Cheetham et al., 2023; Dupont et al., 2021; El-Sayed Moustafa et al., 2022; Graham et al., 2021; Pickering et al., 2020; Seow et al., 2020; Sweeney et al., 2021; Wells et al., 2020).

Upon the resumption of HIV-based research, challenges such as supply chain issues and reduced access to external research services continued to impact the progression of my research. Despite the effects of these disruptions on my PhD project, I hope the work presented in this thesis, coupled with my involvement in COVID-19 research, demonstrates my scientific competence to the level required for the attainment of a PhD.

Abbreviations

AdDMEM	Advanced DMEM F12
AIDS	Acquired immunodeficiency syndrome
ALCAM	Activated cell adhesion molecule
APOBEC3G/F	Apolipoprotein B mRNA-editing complex polypeptide 3G/F
ART	Antiretroviral therapy
BSA	Bovine Serum Albumin
BST-2	Bone marrow stromal cell antigen 2
CDK	Cyclin dependent kinase
cGAMP	cyclic GMP–AMP
cGAS	cGAMP synthase
CL3	Containment level 3
CNS	Central nervous system
CNV	Copy-number variations
CPSF6	Cleavage and polyadenylation specific factor 6
CRF	Circulating recombinant form
CRM1	Chromosome regional maintenance 1
CycT1	Cyclin T1
CypA	Cyclophilin A
DMEM	Dulbecco's modified Eagle's medium
dNTP	Deoxynucleoside triphosphate
dNTPase	dNTP triphosphohydrolase
DSIF	DRB sensitivity-inducing factor
E. coli	*Escherichia coli*
EC	Elite controller
EDTA	Ethylenediaminetetraacetic acid
EM	Expansion media
Env	Envelope
ESC	Embryonic stem cell
ESCRT	Endosomal sorting complex required for transport
FBS	Fetal bovine serum
FDA	Food and Drug Administration
FGF	Fibroblast growth factor
Gag	Group specific antigen
GCDR	Gentle Cell Dissociation Reagent
GWAS	Genome-wide association studies
HAART	Highly active antiretroviral therapy
HAND	HIV-1-associated neurocognitive disorder
HEK	Human embryonic kidney
HipSci	Human Induced Pluripotent Stem Cells Initiative
HIV-1	Human immunodeficiency virus type-1

HLA	Human leukocyte antigen
IFITM	IFN-induced transmembrane protein
IFN	Interferon
IFU	Infectious unit
II	Integrase inhibitor
IL	Interleukin
iMacs	iPSC-derived macrophages
iPSC	Induced pluripotent stem cells
IRF	Interferon regulatory factor
ISG	IFN-stimulated gene
JAK1	Janus kinase 1
LB	Luria-Bertani
LDL-R	Low density lipoprotein receptor
LEDGF	Lens epithelium-derived growth factor
LTNP	Long-term non-progressor
LTR	Long terminal repeat
MACS	Multicentre AIDS cohort study
MagACS	Magnet-activated cell sorting
MEF	Mouse embryonic fibroblast
MOI	Multiplicity of infection
MSM	Men who had sex with men
MX2	Myxovirus resistance protein 2
NEAA	Non-essential amino acids
Nef	Negative factor
NF-κB	Nuclear factor κB
NK	Natural killer
NELF	Negative elongation factor
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NP	Non-progressor
NPC	Nuclear pore complex
NRTI	Nucleoside reverse transcriptase inhibitor
NUP	Nucleoporin
OCT-4	Octamer-binding transcription factor 4
ORF	Open reading frame
P-TEFb	Positive transcription elongation factor b
P/S	Penicillin-streptomycin
PAMP	Pathogen-associated molecular pattern
PBMC	Peripheral blood mononuclear cell
pbs	Primer binding site
PBS	Phosphate buffered saline
PEI	Polyethyleneimine
PFA	Paraformaldehyde
PI	Protease inhibitor

PIC	Pre-integration complex
Pol	Polymerase
ppt	Polypurine tract
PRR	Pattern recognition receptor
PSC	Pluripotent stem cells
qPCR	Quantitative PCR
Rev	Regulator of expression of virion proteins
rIL-2	Recombinant interleukin-2
RLR	RIG-I-like receptor
RNA-seq	RNA-sequencing
RNAi	RNA-interference
ROCKi	Rho-associated protein kinase inhibitor
RP	Rapid progressor
RPK	Reads per kilobase
RRE	Rev response element
rRNA	Ribosomal RNA
SAMHD1	Sterile alpha motif and HD-domain-containing protein
SCF	Stem cell factor
SCNT	Somatic cell nuclear transfer
SeV	Collective Sendai Virus
shRNA	Small hairpin RNA
SIV	Simian immunodeficiency virus
SLC35B2	Solute carrier family 35 member B2
SNV	Single nucleotide variations
SP1	Spacer peptide 1
SP2	Spacer peptide 2
STAT	Signal transducer and activator of transcription proteins
STING	Stimulator of interferon genes
TAR	Transactivation region
Tat	Trans-activator of transcription
TLR	Toll-like receptor
TPM	Transcripts per million
TPST2	Tyrosylprotein sulfotransferase 2
TRIM5α	Tripartite-containing motif 5α
TYK2	Tyrosine kinase 2
URF	Unique recombinant form
VC	Viraemic controller
VC – Sooty	Sooty-like VC
Vif	Viral infectivity factor
Vpr	Viral protein R
Vpu	Viral protein U
Vpx	Viral protein X
VSV-G	Glycoprotein of the vesicular stomatitis virus

Table of Contents

Acknowledgements 3

Abstract 4

COVID-19 Impact Statement 6

Abbreviations 7

Table of Contents 10

List of Figures 14

List of tables 16

Chapter 1: Introduction 17

1.1 HIV-1 & AIDS – A brief account of humans vs virus 18

1.2 Human Immunodeficiency Virus type 1 21

1.2.1 Classification and origin 21

1.2.2 Genome organisation 22

1.2.3 Virion Structure 22

1.3 HIV-1 life cycle 25

1.3.1 Entry 25

1.3.2 Capsid uncoating 25

1.3.3 Reverse Transcription 26

1.3.4 Nuclear import 27

1.3.5 Integration 27

1.3.6 Transcription 28

1.3.7 mRNA processing and export 28

1.3.8 Translation 29

1.3.9 Assembly 29

1.3.10 Budding 30

1.3.11 Maturation 30

1.4 HIV-1 infection: The cell's response 32

1.4.1 Sensing and interferon production 32

1.4.2 Dependency factors 33

1.4.3 Restriction factors 36

1.4.4 Identification of host factors that interact with HIV-1 41

1.5 HIV-1 infection: A clinical outlook 43

1.5.1 Transmission 43

1.5.2 A clinically typical HIV-1 infection 43

1.5.3 Establishment of a reservoir 46

1.5.4 The host response and why it fails 46

1.5.5 Variation in HIV-1 clinical outcomes 47

1.5.6 The natural genetic variation in individuals infected with HIV-1 48
1.6 The Multicentre AIDS cohort study 52
1.7 Induced pluripotent stem cells 53
1.7.1 Discovery and applications 53
1.7.2 Reprogramming 55
1.7.3 The effect of reprogramming on the cell 58
1.7.4 Use as a tool for a permissivity to HIV-1 screen 61
1.8 Thesis aims 63
Chapter 2: Materials & Methods 64
2.1 Plasmids and bacteria 65
2.1.1 Plasmids 65
2.1.2 Plasmid production in competent bacteria 65
2.2 Cell culture 67
2.2.1 Cell lines 67
2.2.2 PBMC isolation, thawing, sorting, and activation 68
2.2.3 Reprogramming 69
2.2.4 Trilineage differentiation 70
2.2.5 G-banded karyotype analysis 71
2.3 Lentivirus production and quantification 72
2.3.1 Production of VSV-G pseudotyped HIV-1 based GFP vector 72
2.3.2 Production of VSV-G pseudotyped full length HIV-1, strain NL4-3 72
2.3.3 Purification and concentration of lentivirus particles 72
2.3.4 Quantification of lentivirus stocks 72
2.4 Lentiviral infections 73
2.4.1 Infection of activated CD4+ T-cells 73
2.4.2 iPSC permissivity to HIV-1 infection pipeline 73
2.5 Flow cytometry 76
2.5.1 Equipment and software 76
2.5.2 Assessment of pluripotency marker expression 76
2.5.3 Quantification of intracellular HIV-1 Gag 76
2.5.4 PBMC phenotyping 76
2.6 Microscopy 78
2.6.1 Equipment and software 78
2.6.2 Trilineage differentiation assay 78
2.6.3 iPSC permissivity to HIV-1 infection data acquisition 78
2.6.4 Operetta analysis pipelines 79
2.7 Nucleic acid extraction and manipulation 80
2.7.1 gDNA extraction 80
2.7.2 RNA extraction 80
2.7.3 cDNA synthesis 80
2.8 qPCR 81
2.8.1 Detection of HIV-1 provirus 81

2.8.2 Detection of Sendai reprogramming vectors 81
2.9 iPSC permissivity to HIV-1 data analysis 83
2.9.1 Linear regression model and normalisation to the CTRs 83
2.9.2 Logistic curve fitting analysis 83
2.9.3 Z-score normalisation 84
2.10 RNA sequencing 85
2.10.1 Sequencing 85
2.10.2 PluriTest analysis 85
2.10.3 Processing of RNA-seq data 85
2.10.4 Differentiational gene expression analysis 86
2.11 Statistical tests 87
Chapter 3: Establishment of a reprogramming pipeline for generating iPSCs from HIV-1 infected PBMCs 88
3.1 Introduction 89
3.2 Results 91
3.2.1 Comparison of feeder-dependent and feeder-free reprogramming protocols for generating iPSCs from PBMCs 91
3.2.2 Evaluating the use of fresh and cryopreserved PBMCs for reprogramming 96
3.2.3 Trialling of the reprogramming protocols with 10-fold fewer cell numbers 98
3.2.4 Evaluation of using freeze/thawed Sendai viral vectors for reprogramming 98
3.2.5 Validation of the optimised reprogramming protocol 98
3.2.6 Optimisation of CD4+ cell depletion from a PBMC population using MagACS 100
3.2.7 Generation of HIV-1 provirus-free iPSCs from PBMCs containing HIV-1 infected cells 104
3.3 Discussion 108
Chapter 4: Reprogramming of the MACS PBMC samples 111
4.1 Introduction 112
4.2 Results 114
4.2.1 MACS participant selection based upon their clinical history in response to HIV-1 infection 114
4.2.2 Phenotyping of the MACS PBMC samples 119
4.2.3 Reprogramming of MACS PBMC samples to iPSCs 122
4.2.4 Pluripotency marker expression in MACS iPSCs 124
4.2.5 Use of a trilineage differentiation assay to further validate the pluripotency of the MACS iPSCs 126
4.2.6 Pluripotency analysis of MACS iPSCs by PluriTest 129
4.2.7 Generation of Sendai viral vector-free iPSCs from the MACS PBMC samples 132
4.2.8 Detection of HIV-1 provirus in the MACS generated iPSCs 134
4.2.9 G-banded karyotypic analysis of the MACS iPSCs to detect chromosomal abnormalities 136
4.3 Discussion 138

Chapter 5: Permissivity to HIV-1 infection screen on the MACS iPSCs and the correlation of phenotypes to transcriptomes 142

5.1 Introduction 143

5.2 Results 146

5.2.1 Set-up of the Operetta CLS imaging and analysis pipelines 146

5.2.2 Examination and cleaning of the infection percentage data produced by the infectivity and late gene expression assays 148

5.2.3 Analysis of the permissivity of the MACS iPSCs to the VSV-G pseudotyped HIV-1 based GFP vector using a linear regression model 151

5.2.4 Analysis of the permissivity of the MACS iPSCs to the VSV-G pseudotyped HIV-1 based GFP vector using a logistic curve fitting model 154

5.2.5 Interpretation of the linear regression and logistic curve fitting model analyses of the iPSC infectivity phenotyping assay data 157

5.2.6 Analysis of the permissivity of the MACS iPSCs to the VSV-G pseudotyped full length HIV-1, strain NL4-3, using a linear regression model 160

5.2.7 Analysis of the permissivity of the MACS iPSCs to the VSV-G pseudotyped full length HIV-1, strain NL4-3, by normalising to virus volume 163

5.2.8 Interpretation of the linear regression model and virus volume normalised analysis of the late gene expression phenotyping assay data 166

5.2.9 Hierarchical clustering of RNA-seq samples from the screened iPSC lines 169

5.2.10 Comparison of the transcriptomes of the two MACS iPSC lines identified to exhibit opposite extreme infectivity phenotypes by the infectivity assay. 172

5.2.11 Comparison of the transcriptomes of the two MACS iPSC lines identified to exhibit opposite extreme late gene expression phenotypes from the late gene expression assay. 178

5.3 Discussion 184

Chapter 6: Discussion 196

6.1 Summary 197

6.2 Future directions 200

6.2.1 Validation of candidate genes 200

6.2.2 Correlation of permissivity phenotypes to genomic data of the MACS participants 200

6.2.3 Correlation to other screens utilising natural variation between individuals 201

6.2.4 Future uses of the MACS iPSCs 204

6.3 Impact 205

References 206

Supplementary Files 239

List of Figures

Figure 1.1: HIV-1 genome in proviral form. 24

Figure 1.2: HIV-1 life cycle. 31

Figure 1.3: Dependency and restriction factors in HIV-1 infection. 40

Figure 1.4: Typical trend of an individual's HIV-1 infection over time. 45

Figure 2.1: iPSC permissivity to HIV-1 infection screen pipeline. 75

Figure 3.1: iPSC quality control checks on a representative iPSC clonal line reprogrammed from PBMCs using the feeder-free protocol. 93

Figure 3.2: Growth overtime of iPSC colonies post Sendai viral vector transduction during the 2nd reprogramming experiment. 95

Figure 3.3: Representative iPSC lines generated from either fresh or cryopreserved PBMCs analysed for their OCT-4 expression. 97

Figure 3.4: CD4+ cell depletion of PBMCs using MagACS. 101

Figure 3.5: Optimisation of CD4+ cell depletion using MagACS. 103

Figure 3.6: Optimisation of the detection of HIV-1 provirus by qPCR. 105

Figure 3.7: Generation of HIV-1 provirus-free iPSCs from a population of PBMCs containing HIV-1 infected cells. 107

Figure 4.1: Longitudinal analysis comparing the molecular based clinical data of two MACS participants with opposite clinical outcomes to HIV-1 infection. 118

Figure 4.2: Phenotyping of the MACS PBMC samples. 121

Figure 4.3: Representative image of a MACS PBMC sample-derived iPSC line demonstrating typical iPSC morphological characteristics. 123

Figure 4.4: Expression of pluripotency markers, OCT-4 and NANOG, in the MACS iPSC lines, with particular focus on a representative line, 20480A. 125

Figure 4.5: Directed differentiation of a representative MACS iPSC line, 20480A, into the three germ layers and analysis of OCT-4 expression before and after differentiation. 128

Figure 4.6: PluriTest analysis results for the MACS iPSCs and the CTR iPSC line. 131

Figure 4.7: Detection of the Sendai viral vectors in MACS iPSCs by qPCR. 133

Figure 4.8: Detection by qPCR of HIV-1 provirus in the MACS iPSC lines and validation of the assay's sensitivity. 135

Figure 4.9: G-banded karyotyping analysis of the MACS iPSC line 20236A. 137

Figure 5.1: Operetta analysis pipeline used to calculate percentage infections for the infectivity and late gene expression assays. 147

Figure 5.2: Representative data of the infection curves generated from the infectivity and late gene expression assays. 150

Figure 5.3: Vector infection data from the infectivity assay on the MACS iPSCs analysed by a linear regression model. 153

Figure 5.4: B-values calculated by logistic curve fitting of the percentage infection data from the infectivity assay. 156

Figure 5.5: Interpretation of the infectivity assay data when analysed by the linear regression and logistic curve fitting models. 159

Figure 5.6: Virus infection data from the late gene expression assay on the MACS iPSCs analysed by a linear regression model. 162

Figure 5.7: Virus infection data from the late gene expression assay on the MACS iPSCs analysed by normalisation to virus volume. 165

Figure 5.8: Interpretation of the late gene expression assay data when analysed by the linear regression and virus volume normalised models. 168

Figure 5.9: Hierarchical clustering of all RNA-seq samples from the iPSC lines run in the permissivity to HIV-1 screen. 171

Figure 5.10: Pearson's correlation coefficient analysis of the transcriptomes of two MACS iPSC lines exhibiting opposite extreme permissivity phenotypes in the infectivity assay. 174

Figure 5.11: Top 50 protein coding genes exhibiting the greatest ΔTPM between 20318A and 20543B. 175

Figure 5.12: Comparison of the transcriptomes of 20318A and 20543B considering fold change in expression. 176

Figure 5.13: Pearson's correlation coefficient analysis of the transcriptomes of two MACS iPSC lines exhibiting opposite permissivity phenotypes in the late gene expression assay. 180

Figure 5.14: Top 50 protein coding genes exhibiting the greatest ΔTPM between 20318A and 20722A. 181

Figure 5.15: Comparison of the transcriptomes of 20318A and 20722A considering fold change in expression. 182

Figure 5.16: Candidate genes identified as a result of the two assays of the permissivity to HIV-1 infection screen. 194

Figure 6.1: Top 50 ΔTPM genes from two different screens exploiting the natural variation between individuals 203

List of tables

Table 2.1: Primers and probe used for the detection of HIV-1 provirus by qPCR. 82

Table 3.1: Summary of the three preliminary reprogramming experiments performed to optimise the reprogramming of PBMCs to iPSCs. 99

Table 4.1: Clinical phenotypes of participants from the MACS relevant to this project and the criteria used to define them. 116

Table 4.2: Details of the MACS PBMC samples selected to be reprogrammed into iPSCs. 116

Table 5.1: MACS iPSC lines tested, along with the CTR iPSC line, in each of the screens. 145

Table 5.2: The top 5 most resistant MACS iPSC lines from the infectivity assay when analysed by either the linear regression or logistic curve fitting model. 158

Table 5.3: The top 5 most permissive MACS iPSC lines from the infectivity assay when analysed by either the linear regression or logistic curve fitting model. 158

Table 5.4: The top 5 most resistant cell lines from the late gene expression assay when analysed by either the linear regression or virus volume normalised model. 167

Table 5.5: The top 5 most permissive cell lines from the late gene expression assay when analysed by either the linear regression or virus volume normalised model. 167

Table 5.6: Candidate gene list identified by comparing the transcriptomes of two MACS iPSC lines, 20318A and 20543B, that exhibited opposite extreme phenotypes in the infectivity assay.....177

Table 5.7: Candidate gene list identified by comparing the transcriptomes of two MACS iPSC lines, 20318A and 20722A, that exhibited opposite extreme phenotypes in the late gene expression assay. 183

Chapter 1: Introduction

1.1 HIV-1 & AIDS – A brief account of humans vs virus

Since its emergence in the early 1980s, human immunodeficiency virus type-1 (HIV-1) has posed a substantial and ongoing challenge to global public health. By the end of 2022, along with a closely related lentivirus, HIV-2, HIV-1 had infected a predicted 85.6 million people worldwide. These HIV infections had led to an estimated 40.4 million deaths as a result of the development of acquired immunodeficiency syndrome (AIDS) (World Health Organisation, 2022).

HIV-1 is the most prevalent causative agent of AIDS, which is characterised by a progressive failure in the host immune system. This syndrome was first characterised in 1981 when several men who had sex with men (MSM) presented with *Pneumocystis carinii* pneumonia, a disease common in individuals with a weakened immune system, at hospitals in Los Angeles, California (Gottlieb et al., 1981). The virus responsible for this syndrome was first isolated from the peripheral blood lymphocytes of AIDS patients in 1983 and was initially termed either lymphadenopathy-associated virus (LAV) or human T-lymphotropic virus type-3 (HTLV-III) (Barré-Sinoussi et al., 1983; Gallo et al., 1983). Just two years later, the spread of the virus was manifest, with at least one case of the now named HIV-1 having been reported in each region of the world.

In 1987, the first drug for treating AIDS, the nucleoside reverse transcriptase inhibitor (NRTI) azidothymidine, was approved by the Food and Drug Administration (FDA). Its success at not only treating AIDS but also delaying its onset, albeit with serious side effects, encouraged the development of additional NRTIs which were FDA approved in the early 1990s (Merigan et al., 1991). Despite these efforts, HIV-1 infections were still increasing worldwide and by 1992, AIDS had become the leading cause of death in males aged 25-44 years in the United States (CDC, 1993). It became apparent that the single-target drug treatment regimens, although somewhat efficacious, were not sufficient to stop HIV-1, as the virus's extraordinary ability to evolve meant it was rapidly developing drug resistance (Concorde Coordinating Committee, 1994; Vella et al., 1994). A two-drug combination was shown to be more effective than the single therapies, however, still provided limited duration of efficacy (Hammer et al., 1996). It

wasn't until a triple combination of drugs was trialled in 1996 that a major breakthrough in the treatment of HIV-1 was made. This three-drug therapy, termed highly active antiretroviral therapy (HAART), was far superior at suppressing viral replication, whilst providing a greater genetic barrier for HIV-1 to overcome and develop drug resistance (Cole & Stuart, 2010). The success of HAART was also partly attributable to the development of two new classes of antiretroviral drugs: protease inhibitors (PIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs); even further increasing the necessary number of mutations to achieve drug resistance. The following year after the introduction of HAART, a 47% decrease in AIDS-related deaths was observed in the United States (CDC, 1998). Despite this advancement, HAART came with adherence challenges and often major side effects demonstrating the need for optimisations in the therapy. Over the next many years, the persistent development of new generations of anti-retroviral drugs, as well as a new class of anti-retroviral drug, integrase inhibitors (IIs), produced continued progression in the treatment of HIV. Today, HAART requires just a single daily tablet and grants an HIV-infected individual a life expectancy comparable to that of an uninfected person, with minimal side effects (Trickey et al., 2017). HAART is now also approved by the FDA for prophylactic use, both pre-exposure (PrEP) and post-exposure (PEP), for uninfected individuals at high risk of infection in an attempt to stop the spread of the virus (Holmes, 2012; Rey, 2011). Despite the successes of HAART, its limitations lie with its inability to cure the patient of HIV, requiring life-long adherence. There are only two peer-reviewed instances of an HIV-1 infected patient being cured of infection, and unfortunately they were achieved by a method not feasible for the majority of the population. Both cases occurred after the patient received an allogenic haemopoietic stem-cell transplant with donor cells not expressing the C-C chemokine receptor 5 (CCR5Δ32/Δ32) in response to a previous diagnosis of blood cancer (Gupta et al., 2020; Hütter et al., 2009). A different strategy taken to combat HIV-1 is to generate a prophylactic vaccine, however, due to an array of factors, none have proven successful enough thus far (Hargrave et al., 2021).

Despite the advancements in the treatments against HIV-1, it still remains a significant burden on global public health. In 2022, 39 million people were still living with HIV, of which 1.3 million were newly infected that year, and 630,000 individuals died of AIDS-related illnesses. This illustrates the need for further research to discover the elusive HIV-1 cure and with over

130 HIV cure-related clinical trials active in 2019, the ongoing efforts are substantial (Barr & Jefferys, 2020).

1.2 Human Immunodeficiency Virus type 1

1.2.1 Classification and origin

HIV-1 is a member of the *Lentivirus* genus, within the *Retroviridae* family (International Committee on Taxonomy of Viruses, 2022). *Lentiviruses* are characterised by their lengthy incubation periods resulting in chronic and life-threatening illnesses in many mammalian species, including humans. Members of the *Retroviridae* family contain genomic RNA that is reverse transcribed, by the virus's own reverse transcriptase enzyme, to produce DNA which is then integrated into the host cell genome.

HIV-1 and the closely related less prevalent, transmissible and virulent lentivirus, HIV-2, originated after the zoonotic transfer of simian immunodeficiency viruses (SIVs) from non-human primates to humans. HIV-1 is classified into 4 groups, M, N, O and P, which are each the result of an independent zoonotic transmission of SIVs from African apes. Group M and N are believed to have originated from SIVs in chimpanzees (SIVcpz) whilst groups O and P are thought to have evolved from SIVs in gorillas (SIVgor) (Keele et al., 2006; Plantier et al., 2009; Takehisa et al., 2009). Group M is the most prevalent, responsible for perhaps 98% of all HIV-1 infections, and is found worldwide, whereas the other groups are still restricted to areas of Africa. Its increased transmission paired with HIV's intrinsic ability to rapidly evolve, means group M requires further categorisation, based upon genetic similarity, into the sub-types: A1, A2, A3, A4, B, C, D, F1, F2, G, H, J and K (Taylor et al., 2008). HIV-1 is also capable of homologous recombination, generating unique recombinant forms (URFs) and circulating recombinant forms (CRFs) that have been observed in two or at least three epidemiologically unlinked individuals respectively (Burke, 1997). These strains occur in geographical regions where multiple variants of HIV-1 are circulating and as a result, a single cell can be coinfected by two different HIV-1 strains. These URFs and CRFs (e.g. CRF01_AE) offer a threat to public health as they can be associated with increased disease progression and multi-drug resistance (Chu et al., 2017). HIV-2, which is found mainly in west Africa, is classified into two groups, A and B, that are thought to have evolved from two independent zoonotic transfers to humans from SIVs in sooty mangabeys (SIVsmm) (Damond et al., 2004; Hirsch et al., 1989).

1.2.2 Genome organisation

Two identical positive-sense single-stranded copies of RNA make up the genome of HIV-1. The ~9.7kb long strands of RNA contain 9 partially overlapping open reading frames (ORFs) which encode for 15 proteins (**figure 1.1A**). For production of these proteins, the RNA must first be reverse transcribed into DNA and integrated into the host genome. This integrated HIV-1 DNA is termed the HIV-1 provirus.

Long terminal repeats (LTRs) are present at both the 5' and 3' ends of the provirus. The 5' LTR acts as a promoter for the transcription of the viral DNA as well as providing binding sites for many cellular factors, whilst the 3' LTR acts as the site of 3'-end formation of RNA (Pereira et al., 2000; Shah et al., 2014). Of the nine viral genes encoded by the HIV-1 genome, three, *gag*, *pol*, and *env*, encode for its structural proteins and are conserved across all retroviruses. The *gag* ORF encodes for a polyprotein, p55, which is later cleaved by the viral protease to produce the main structural proteins of the virus: matrix, capsid, nucleocapsid, p6, and spacer peptides 1 and 2 (SP1&SP2). Polymerase (Pol) is generated as a $pr160^{gag\text{-}pol}$ precursor, due to a ribosomal frameshift during translation, and is later cleaved by the viral protease to produce the viral enzymes: integrase, protease, reverse transcriptase and RNase H (Jacks et al., 1988). The *env* ORF also encodes for a polyprotein, p160, which this time is cleaved by furin, a cellular protease, into gp120 and gp41 (Hallenberger et al., 1992). The remaining six genes encode for regulatory proteins. *Tat* and *rev* encode for essential regulatory proteins that help instigate viral replication, whilst *nef*, *vif*, *vpr*, and *vpu* encode for accessory proteins that aid viral replication.

1.2.3 Virion Structure

A mature HIV-1 particle is spherical and measures ~145nm in diameter (Briggs et al., 2003). The external layer of the virion is made of the lipid bilayer from the membrane of the host cell coated with ~14 spikes comprised of envelope (Env) proteins. Each spike consists of a trimer of the transmembrane protein, gp41, bound to a trimer of surface protein, gp120 (Gelderblom, 1991). Matrix lines the inside of the lipid bilayer envelope. At the centre of the virion, a conical shaped capsid core encases the two genomic strands of RNA, which are bound to several molecules of integrase, reverse transcriptase, and RNaseH. Protease and viral

protein R (Vpr), the remaining packaged proteins, are also incorporated into the virion (Bachand et al., 1999).

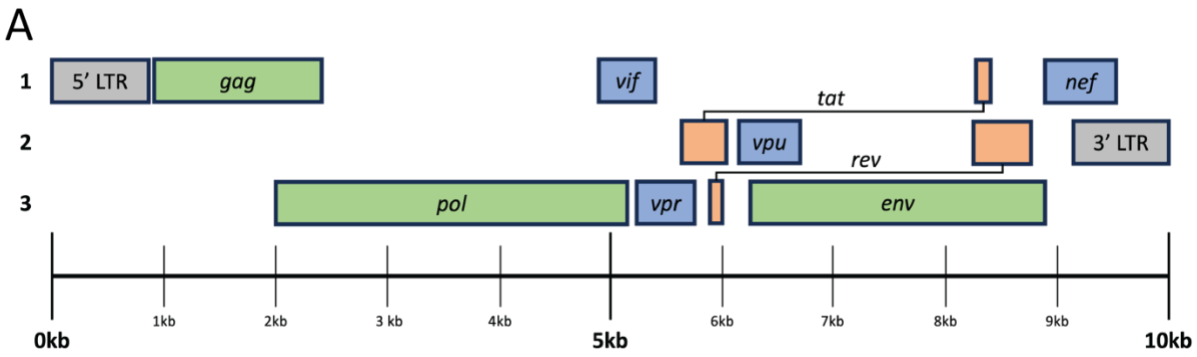


Figure 1.1: HIV-1 genome in proviral form.
A) 5' to 3' organisation of the HIV-1 genome in proviral form split by the 3 ORFs. Grey = LTRs. Green = canonical retroviral structural protein encoding genes: *gag*, *pol* and *env*. Orange = essential transcriptional regulator encoding genes: *tat* and *rev*. Blue = viral accessory protein encoding genes: *vif*, *vpr*, *vpu* and *nef*. Image generated by myself in PowerPoint.

1.3 HIV-1 life cycle

The HIV-1 life cycle consists of a sequence of intricate and highly coordinated events (**figure 1.2A**). Throughout its life cycle, HIV-1 requires interactions with its virally encoded proteins and host cellular factors for the successful production of infectious progeny.

1.3.1 Entry

For entry, HIV-1 requires the cellular receptor CD4 (Dalgleish et al., 1984). gp120, in the spike on the surface of HIV-1, binds to CD4 initiating energetically favourable structural alterations in Env. These changes uncover the co-receptor binding site on gp120, resulting in binding to one of or both of two co-receptors, CCR5 and CXCR4, depending on the HIV-1 strain (Alkhatib et al., 1996; Deng et al., 1996; Dragic et al., 1996; Feng et al., 1996). CCR5 is used by R5 tropic viruses, CXCR4 is used by X4 tropic viruses, and both can be used by R5/X4 tropic viruses (Berger et al., 1998). Uncoupling of gp120 reveals the fusion peptide domain of gp41 which inserts into the membrane of the target cell (Chan et al., 1997). A six-helix bundle forms through alterations to gp41 that brings together the viral and host cell membranes. Lipids from the two membranes mix and a fusion pore that allows for the virus's contents to enter the cell cytoplasm is generated (Melikyan, 2008).

1.3.2 Capsid uncoating

After entry, the HIV-1 capsid core is released into the cytoplasm of the host cell through the fusion pore. The capsid core consists of hexamers and pentamers of capsid arranged to form a fullerene-like closed conical shape (Ganser et al., 1999). ~250 capsid hexamers make up the bulk of the core while 12 pentamers, 5 located at the narrow end and 7 located at the wide end, give the capsid its shape (Zhao et al., 2013). Contained within the capsid are: the two strands of viral RNA genome, nucleocapsid; viral enzymes integrase and reverse transcriptase, and accessory protein, Vpr. Early theories hypothesise that the capsid disassembles, releasing its contents, shortly after entry into the cytoplasm (Fassati & Goff, 2001). However, more recent indications propose that the capsid remains intact to help transport its contents across the cell cytoplasm to the nucleus, whilst providing protection from antiviral factors and host cellular sensing (Toccafondi et al., 2021). Throughout its passage, capsid has been shown to

interact with host cell factors including tripartite-containing motif 5α (TRIM5α), nucleoporin (NUP) 358 and cyclophilin A (CypA) (Bichel et al., 2013; Jimenez-Guardeño et al., 2019; Y. Li et al., 2009). Although still a contentious topic, a dynamic model where capsid may disassemble in the cytoplasm, at the nuclear pore complex or in the nucleus, depending on its interactions with host cellular factors, seems currently favourable.

1.3.3 Reverse Transcription

HIV-1's positive-sense single-stranded RNA viral genome is converted into double-stranded DNA via the process of reverse transcription, which is mediated by the viral enzyme reverse transcriptase. Reverse transcriptase is comprised of a heterodimer of two subunits, p66 and p51. The p66 subunit includes the two catalytic domains of the enzyme, one with polymerase activity and the other with RNase H activity, while the p51 subunit provides structural integrity (Kohlstaedt et al., 1992). Reserve transcription takes places inside of the capsid core and has been shown to trigger capsid uncoating (Cosnefroy et al., 2016; Rankovic et al., 2017).

The viral genomes, a pair of positive-sense single-strands of RNA, act as templates for DNA synthesis. Cellular $tRNA^{Lys3}$ binds to the 5' LTR of the viral genome at the primer binding site (pbs) and signifies the initiation of reverse transcription. The $tRNA^{Lys3}$ acts as a primer for the synthesis of the negative-sense single-stranded DNA via the polymerase activity of reverse transcriptase (Isel et al., 1996). During elongation, the resultant DNA-RNA hybrid acts as a substrate for the RNase H domain of reverse transcriptase, which degrades the 5' end of the RNA template (Purohit et al., 2007). This releases the newly produced minus-strand strong stop DNA which contains a region, known as 'R', that then anneals to an identical R region at the 3' end of the viral RNA template (Panganiban & Fiore, 1988). Synthesis of the negative-sense single-stranded DNA continues shadowed by the degradation of the RNA template excluding a region resistant to RNase H-mediated degradation, the polypurine tract (ppt) (Smith & Roth, 1992). This ppt acts as the primer for the synthesis of the positive-sense strand DNA which extends to the far end of the $tRNA^{Lys3}$, creating the pbs. Once elongation of the negative-sense strand DNA has also synthesised the pbs region, the two complementary pbs regions anneal. Synthesis then continues from the 3' end of the negative-sense strand DNA, which displaces the 5' end of the same negative-sense strand DNA from the positive sense strand DNA, finalising the first LTR. Extension also occurs from the pbs of the positive-sense

strand DNA resulting in formation of the second LTR and completion of the double stranded DNA.

1.3.4 Nuclear import

For HIV-1 to achieve successful integration into the host cell genome, the viral genomic DNA and integrase require access to the inside of the target cell nucleus. For this to be possible in non-dividing cells, the pre-integration complex (PIC) must be conveyed through the nuclear pore complex (NPC). Nuclear import is a capsid-dependent process and involves capsid interacting with numerous cellular factors such as NUP153, NUP358, cleavage and polyadenylation specific factor 6 (CPSF6) and β-karyopherin transportin 3 (TNPO3) (Brass et al., 2008; De Iaco et al., 2013; Di Nunzio et al., 2012; Matreyek & Engelman, 2011; Yamashita & Emerman, 2004). The precise details of HIV-1 nuclear import, in terms of its temporal relationship with capsid uncoating and reverse transcription, are still disputed. Though, new evidence, such as capsid being imaged inside the nucleus and reverse transcription continuing post-nuclear entry, suggests that these three events could function concurrently rather than successively (Dharan et al., 2020; Ingram et al., 2021; Zila et al., 2021).

1.3.5 Integration

Once the viral genomic DNA has entered the nucleus of the target cell, it needs to be integrated into the host genome. The viral enzyme that catalyses this process, integrase, consists of three domains: An N-terminal domain that forms dimers, possessing a Zn^{2+} binding capacity via HCCH motifs; a central catalytic core domain exhibiting RNase H-like activity; and a C-terminal domain involved in multimerisation and facilitating binding to DNA (Engelman et al., 1993).

Integrase multimerises on the newly synthesised viral double-stranded DNA, creating an intasome (Hare et al., 2010). Integrase reveals a hydroxyl group at both 3’ blunt ends of the viral DNA by cleaving a dinucleotide. A nucleophilic attack by the nascent 3’ hydroxyl groups on the phosphodiester bond of host DNA, that has been cut by integrase, results in the viral DNA and host DNA linking. Gaps in the DNA formed as a result of integration are mended by host cell DNA repair machinery resulting in the completed HIV-1 provirus.

Interactions between integrase and host cellular factors influence the integration site of HIV-1. These cellular factors include: lens epithelium-derived growth factor (LEDGF)/p75, CPSF6, PC4 and SFRS1 Interacting Protein 1 (PSIP1), and SWI/SNF Related, matrix associated, actin dependent regulator of chromatin, subfamily B, member 1 (SMARCB1) (Kalpana et al., 1994; Llano et al., 2006; Marshall et al., 2007; Sowd et al., 2016). As a result of these interactions, the viral DNA is integrated into the host genome near the nuclear periphery in areas of transcriptionally active chromatin (Albanese et al., 2008; Schröder et al., 2002).

1.3.6 Transcription

The 5' LTR of the HIV-1 provirus serves as the promoter for transcription which begins at the R region. Upstream of R, modulatory and enhancer elements possess cellular transcription factor binding sites for NFκB, nuclear factor of activated T-cells (NFAT) and SP1 (Pereira et al., 2000). The binding of cellular transcription factors commences transcription of the HIV-1 provirus by recruitment of RNA polymerase II, producing mRNA encoding for trans-activator of transcription (Tat) and regulator of expression of virion proteins (Rev). This transcription is short lived without Tat production as repressive proteins, including DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF), are recruited to the transcription site (Berkhout et al., 1989; Kao et al., 1987; Raha et al., 2005). This short mRNA product adopts a looped structure termed the transactivation region (TAR). Tat binds to the TAR and recruitments the cellular multiprotein positive transcription elongation factor b (P-TEFb), consisting of a cyclin dependent kinase (CDK) 9 subunit and a cyclin T1 (CycT1) subunit (P. Wei et al., 1998; Y. Zhu et al., 1997). The CDK9 subunit of P-TEFb phosphorylates RNA polymerase II, DSIF, and NELF stimulating transcription, thus, creating a positive feedback loop of Tat production and efficient HIV-1 transcription (Wada, 1998; M. Zhou et al., 2000).

1.3.7 mRNA processing and export

Seeing as the HIV-1 genome is relatively small, ~9.7kb, the produced transcripts must undergo considerable alternative splicing to generate the 15 viral proteins (Purcell & Martin, 1993). Numerous splice donor and acceptor sites within the proviral mRNA make this possible and the extent of splicing is demonstrated by at least 109 distinctive transcripts having been detected in a single HIV-1 isolate (Ocwieja et al., 2012). The full array of mRNAs produced

consists of unspliced transcripts (for the HIV-1 genome, and synthesis of group specific antigen (Gag) and Gag-Pol), partially spliced transcripts (for synthesis of Env, Vpr, viral infectivity factor (Vif) and viral protein U (Vpu)), and completely spliced transcripts (for synthesis of negative factor (Nef), Rev and Tat).

The export of HIV-1 mRNAs occurs via two mechanisms (Blissenbach et al., 2010). Initially, only the completely spliced transcripts, Nef, Rev, and Tat, are exported from the nucleus via the same process as cellular mRNAs, whereas the unspliced and partially spliced transcripts are retained. To export these remaining transcripts, the translation of Rev protein is required. By binding to the Rev response element (RRE) in the Env coding region of unspliced and partially spliced mRNA, Rev promotes their nuclear export via the NPC (Malim et al., 1989). This occurs by Rev binding to an exportin called chromosome regional maintenance 1 (CRM1), which transports the complex to the cytoplasm in a process relying on Ran-GTP (Askjaer et al., 1998).

1.3.8 Translation

For translation of its viral proteins, HIV-1 hijacks the host cell translation machinery. Cap-dependent scanning is utilised for most of the transcripts via eukaryotic translation initiation factor recognising the 7-methyl-guanosine cap at the 5' end of the mRNA (de Breyne et al., 2012). For the production of Pol protein, a ribosomal frameshift, programmed by an upstream slippery sequence, after translating the Gag protein results in the formation of the GagPol polyprotein (Brakier-Gingras et al., 2012; Jacks et al., 1988). HIV-1 can also drive translation via a cap-independent mechanism, where an internal ribosome entry site in the unspliced mRNAs recruits the 40S ribosomal subunit (Brasey et al., 2003).

1.3.9 Assembly

Assembly of HIV-1 involves the organised formation of its structure and packaging of the viral contents. The nucleocapsid domain of the polyprotein, Gag, interacts with the core packaging signal (ψ) of a dimerised genomic viral RNA pair forming a complex that is transited to the periphery of the cell via an interaction between matrix and phosphatidylinositol-4,5-bisphosphate at the inner layer of the cell membrane (Ono et al., 2004). Production of a

curved protein lattice begins at the plasma membrane as Gag-RNA structures drive multimerisation via their capsid domain. Heterodimeric trimers of gp120 and gp41, Env, arrive at the plasma membrane from the Golgi and are incorporated into the forming virions. The mechanism of action of Env incorporation is not fully understood and may occur through passive means; indirect means where Env and Gag are linked by a host protein; or co-targeting where both Env and Gag are directed to microdomains in the host cell membrane known as lipid rafts (Freed, 2015). Conversely, a recent study has shown that the heterodimeric trimers, via an interaction between their cytoplasmic tail and the matrix domain of Gag, are integrated directly into the Gag protein lattices (Pezeshkian et al., 2019).

1.3.10 Budding

For budding, HIV-1 hijacks the cellular endosomal sorting complex required for transport (ESCRT) machinery. To do this, the tumour susceptibility gene 101 (TSG101) and ALG2-interacting protein X (Alix) are recruited by interactions with two separate late domains of the P6 domain of Gag (Fujii et al., 2009; Martin-Serrano et al., 2001). As a result, the ESCRT machinery are recruited. The ESCRT machinery consists of ESCRT-I and ESCRT-II, which are vital for bud creation, ESCRT-III, which is necessary for release of the bud, and Vps4, which recovers the ESCRT proteins after the completion of budding (Hurley & Hanson, 2010).

1.3.11 Maturation

After successful budding and detachment from the host cell, maturation of the newly produced virion is necessary for its infectivity. The small amount of GagPol polyprotein present is proteolytically cleaved by the viral enzyme protease at numerous sites (Pettit et al., 1994). This initiates a structural remodelling of nucleocapsid, capsid and matrix within the virion. Matrix proteins gather to line the inside of the virus's lipid bilayer envelope and capsid proteins build the fullerene-like conical shaped capsid core that encases the dimerised RNA genome that is in complex with nucleocapsid, reverse transcriptase and integrase (Zhao et al., 2013).

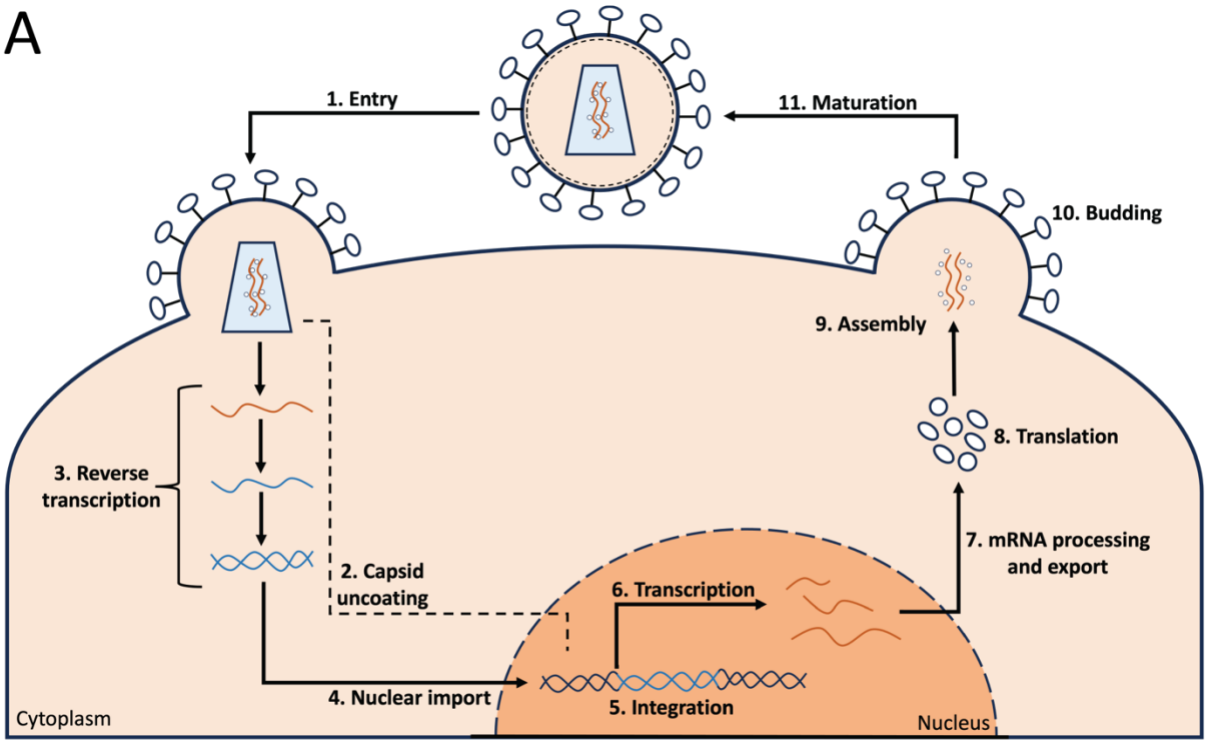


Figure 1.2: HIV-1 life cycle.
A) Diagrammatic representation of the HIV-1 life cycle. HIV-1 enters the cell via an interaction between Env, CD4, and either CCR5 or CXCR4. The viral contents are then released into the cell cytoplasm, and the viral genomic RNA is reverse transcribed into double-stranded DNA. This is imported into the nucleus and the viral double-stranded DNA is integrated into the host genome. It is still debated at which point between entry and integration that capsid uncoating transpires and it may occur in tandem with reverse transcription and nuclear import. Transcription of the HIV-1 provirus produces viral mRNAs that are processed and exported from the nucleus. These are translated into viral proteins, which assemble at the cell plasma membrane. Budding of the newly assembled virion releases it from the host cell, followed by the maturation process to produce a new fully infectious virion. Image generated by myself in PowerPoint.

1.4 HIV-1 infection: The cell's response

1.4.1 Sensing and interferon production

The host's initial defence against viral infection is the innate immune system. A host cell detects foreign molecular material, pathogen-associated molecular patterns (PAMPs), via pattern recognition receptors (PRRs), triggering a downstream signalling cascade that results in an antiviral response. Various groups of PRRs have been identified as detectors responsible for identifying components of HIV-1, including RIG-I-like receptors (RLRs), toll-like receptors (TLRs), and cytosolic DNA sensors (Yin et al., 2020). Recognition of HIV-1 nucleic acid derived PAMPs via these PRRs results in a release of proinflammatory cytokines and upregulation of interferons (IFNs) via multiple signalling pathways. RLRs can identify HIV-1 single-stranded RNA resulting in downstream signalling via nuclear factor κB (NF-κB) (Solis et al., 2011; M. Q. Wang et al., 2015). Members of the TLR family, TLR 7 and TLR 8, can also recognise the viral single-stranded RNA and trigger interferons via interferon regulatory factor (IRF) 3 (Lepelley et al., 2011). As the HIV-1 single-stranded RNA is reverse transcribed, the nascent double-stranded DNA can be detected by the cytosolic DNA sensor cyclic GMP–AMP (cGAMP) synthase (cGAS). Detection by cGAS leads to cGAMP production, triggering activation of stimulator of interferon genes (STING), resulting in downstream signalling via NF-κB and IRF3 (Gao et al., 2013). STING can also be activated by the sensing of single-stranded DNA by the PRR IFN-γ inducible protein 16 (IFI16) (Jakobsen et al., 2013). The HIV-1 capsid lattice is also a PAMP and is recognised by cellular TRIM5 acting as a PRR and promoting innate immune signalling (Pertel et al., 2011). After sensing capsid, TRIM5 interacts with ubiquitin-conjugating enzyme UEV1A to catalyse the production of unattached K63-linked ubiquitin chains. These chains stimulate the transforming growth factor β (TGFβ)-activated kinase 1 (TAK1) kinase complex which activates activator protein 1 (AP-1) and NF-κB signalling.

The subsequent production of IFNs is an important part of the host's defence against viral infection, including in the context of HIV-1 (Goujon, Schaller, et al., 2013). IFN-α binds to the type I IFN-α receptor triggering activation of Janus kinase 1 (JAK1) and tyrosine kinase 2 (TYK2). JAK1 and TYK2 subsequently phosphorylate signal transducer and activator of transcription proteins (STAT) 1 and 2. Phosphorylated homodimers of STAT1 and

phosphorylated heterodimers of STAT1 and STAT2 then bind to gamma-activated sequences in the host DNA and IRF9 respectively. Binding to IRF9 results in establishment of the IFN-stimulated gene (ISG) factor 3 transcriptional activator which binds to IFN-stimulated response elements. As a result of this IFN-mediated pathway, a family of genes, named ISGs, who play a crucial role in the host's antiviral response, are upregulated (Doyle et al., 2015).

1.4.2 Dependency factors

To successfully replicate, HIV-1 relies on multiple host cell factors. These factors help aid HIV-1 during its infection of a cell, and thus, are termed co-factors or dependency factors. Currently identified HIV-1 dependency factors act at multiple points of the viral life cycle and include: the cellular receptors CD4, CCR5, and CXCR4, CypA, NUP358, CPSF6, LEDGF/p75, P-TEFb, CRM1, the ESCRT machinery (**figure 1.3A**). Additionally, more recently identified dependency factors are: tyrosylprotein sulfotransferase 2 (TPST2), solute carrier family 35 member B2 (SLC35B2), and activated cell adhesion molecule (ALCAM). Research to identify more HIV-1 dependency factors and their mechanisms of action is ongoing.

1.4.2.1 CD4, CCR5 and CXCR4

Necessary for entry into the cell, the cellular receptor CD4 is essential for HIV-1 infection (Dalgleish et al., 1984). It acts as a high affinity receptor for HIV-1 surface molecule gp120, initiating the fusion process between the virus and host cell membranes. HIV-1 may also utilise the CD4 receptor to infect a new cell via cell-to-cell transmission by gp120 on the HIV-infected cell interacting with the CD4 receptor on an uninfected cell (Bracq et al., 2018). After the binding of gp120 to the CD4 receptor, CCR5 and/or CXCR4, depending on the viral strain, are also necessary to advance membrane fusion. Typically, transmitted strains utilise the CCR5 co-receptor and shifting of co-receptor usage to CXCR4 is associated with accelerated disease progression (Connor et al., 1997). The significance of CCR5 as an HIV-1 dependency factor is demonstrated by a 32-base pair deletion mutation in the CCR5 gene (CCR5Δ32) conveying partial resistance and slower disease progression or high resistance to HIV-1 infection in heterozygous or homozygous individuals respectively (Dean et al., 1996; Martinson et al., 1997).

1.4.2.2. CypA

CypA is a ubiquitously expressed protein with peptidyl prolyl isomerase activity that is incorporated into HIV-1 particles via its interaction with Gag during viral assembly. CypA was shown to be required in the early steps of the HIV-1 life cycle at a time between membrane fusion and reverse transcription through interactions with capsid (Braaten et al., 1996; Hatziioannou et al., 2005). More recently, CypA is thought to stabilise the HIV-1 capsid core by binding to it in a distinct pattern where a single CypA molecule bridges two subunits from separate capsid hexamers (C. Liu et al., 2016). This binding to capsid by CypA has also been shown to help protect the capsid core from action of the HIV-1 restriction factor, TRIM5α (K. Kim et al., 2019; Selyutina et al., 2020).

1.4.2.3. NUP358

NUP358 is a cytoplasmic NPC protein that plays a vital role in HIV-1 nuclear import. The cyclophilin binding loop of the HIV-1 capsid has been shown to bind directly to NUP358 and helps mediate HIV-1 capsid import and uncoating (Schaller et al., 2011). However, the exact mechanism of this interaction remains unknown.

1.4.2.4 CPSF6

CPSF6 is an mRNA-processing protein first shown to be essential for HIV-1 nuclear import in primary macrophages (Bejarano et al., 2019). In the absence of CPSF6, HIV-1 is blocked at the NPC but trafficking or reverse transcription up to this point are unaffected. CPSF6 has also been described to play a role in determining the integration site of the viral double-stranded DNA. Its interaction with capsid conveys the PIC towards the nuclear interior towards gene dense, transcriptionally active regions of the host genome, bypassing the peripheral heterochromatin (Achuthan et al., 2018; Sowd et al., 2016).

1.4.2.5 LEDGF/p75

LEDGF/p75 is a ubiquitously expressed nuclear transcription coactivator that has been suggested to aid HIV-1 replication by tethering PICs to chromatin. It has been shown to precisely co-localise with integrase and in the absence of LEDGF/p75, integrase loses its nuclear localisation and chromosomal targeting (Cherepanov et al., 2003; Maertens et al., 2003). LEDGF/p75's action strongly influences the location at which the HIV-1 reverse

transcription product is integrated into the host genome, while also protecting integrase from degradation (Llano et al., 2006).

1.4.2.6 P-TEFb

P-TEFb is a multiprotein complex crucial in RNA polymerase II transcriptional regulation. The transcriptional elongation effect of Tat is P-TEFb-dependent and is suggested to function via P-TEFb's increased action on the HIV-1 LTR (Y. Zhu et al., 1997). P-TEFb is made up of CDK9 and CycT1. It is suggested that because the affinity and specificity of Tat binding to TAR is improved by the interaction between Tat and CycT1, Tat may utilise CycT1 to direct the CDK9 to RNA polymerase II, upregulating transcription (P. Wei et al., 1998).

1.4.2.7 CRM1

CRM1 facilitates the nuclear export of numerous RNAs and proteins. The HIV-1 protein Rev binds to the RRE elements in HIV-1 RNA transcripts that require nuclear export. It has been shown, *in vitro*, that Rev binds directly to CRM1 (Askjaer et al., 1998). CRM1, in a Ran-GTP dependent manner, then exports the Rev-transcript complex via the nuclear export signal on Rev.

1.4.2.8 ESCRT

The ESCRT machinery is comprised of ESCRT-I, ESCRT-II, ESCRT-III, and accessory proteins that are all evolutionarily conserved across eukaryotes. The ESCRT proteins collaborate to remodel the cell membrane causing it to curve away from the cytosol, playing crucial roles in cellular abscission, multivesicular body biogenesis, as well as viral budding (Hurley & Emr, 2006; Wollert & Hurley, 2010). During retroviral assembly, ESCRT-I and ESCRT-II begin the budding process by directing the membrane away from the cytoplasm (Hurley & Hanson, 2010). Once accumulation of Gag is complete, ESCRT-III is recruited to cleave the neck of the bud, releasing the virion from the cell (Jouvenet et al., 2011).

1.4.2.9 TPST2 and SLC35B2

TPST2 is a type II transmembrane Golgi-resident enzyme that is involved in the post-translational modifications of membrane or secretory proteins by catalysing the sulphation of tyrosines. SLC35B2 carriers the donor of the sulphate, 3'-phosphoadenosine-5'-

phosphosulfate, from the cytoplasm to the Golgi where it can be utilised by TPST2. A loss of either TPST2 or SLC35B2 has been shown to result in a lack of sulphation of CCR5, which is necessary for enabling its interactions with the gp120 molecule on the surface of HIV-1. As a result, viral entry is blocked, justifying both TPST2 and SLC35B2 as crucial HIV-1 dependency factors (Park et al., 2017).

1.4.2.10 ALCAM

ALCAM is expressed on activated T-cells, monocytes and dendritic cells and functions as a cell adhesion molecule. It has been shown that loss of ALCAM conveys a loss in immune cell aggregation, thus, a reduction in cell-to-cell contacts being formed. As a result, cell-to-cell transmission of HIV-1 in the absence of ALCAM was strongly restricted, demonstrating its function as an HIV-1 dependency factor for this mechanism of HIV-1 spread (Park et al., 2017). However, ALCAM is expressed at lower levels in primary CD4+ T-cells compared to the cell line model used in this study, suggesting a limit to its functionality in a physiological setting.

1.4.3 Restriction factors

As well as dependency factors that help the virus to replicate, the host cell also contains factors that are detrimental to HIV-1 replication, termed restriction factors. These factors function by targeting part of the viral life cycle and include: SERINCs, IFN-induced transmembrane proteins (IFITMs), TRIM5α, apolipoprotein B mRNA-editing complex polypeptide 3G/F (APOBEC3G/F), sterile alpha motif and HD-domain-containing protein (SAMHD1), myxovirus resistance protein 2 (MX2), and bone marrow stromal cell antigen 2 (BST-2) (**figure 1.3A**). In certain instances, HIV-1 has adapted to overcome the effect of these restriction factors through the use of viral accessory proteins. As with dependency factors, research continues to attempt to uncover novel HIV-1 restriction factors and the mechanisms by which they function.

1.4.3.1 SERINCs

SERINCs are transmembrane proteins whose exact cellular function remains unknown. In particular, SERINC 3 and SERINC 5 are potent restriction factors against HIV-1. They localise to lipid rafts, the site of virus assembly, and are incorporated into the membrane of progeny virus (Schulte et al., 2018). Although the mechanism of action is not fully understood, the restriction

is thought to occur via a block on viral entry by inhibiting the membrane fusion process. Recent studies have suggested that SERINCs may disable Env trimers preventing establishment of the fusion pore (Beitari et al., 2017; Sood et al., 2017). Further, SERINCs may also increase IFN and proinflammatory cytokine production via innate immune signalling resulting in heightened cellular antiviral activities (Zeng et al., 2021). The viral accessory protein, Nef, counters the HIV-1 restriction by SERINCs (Rosa et al., 2015; Usami et al., 2015). Nef targets SERINCs for lysosomal degradation, resulting in decreased SERINC presence at the membrane, and thus, less incorporation into newly produced virions.

1.4.3.2 IFITMs

IFITMs are a family of proteins that inhibit enveloped viruses, including HIV-1, by preventing their fusion with the host cell plasma membrane (Lu et al., 2011). Within the IFITM family, IFITM1, IFITM2, and IFITM3 exhibit antiviral activity. Like SERINCs, IFITMs are incorporated into the membranes of viral progeny during assembly (Compton et al., 2014; Yu et al., 2015). Also akin to SERINCs, it is thought that IFITMs inhibit HIV-1 fusion with the host cell, though, the exact mechanism of their antiviral activity is unknown. The most agreed mechanism of action is that IFITMs increase the rigidity of the virus's membrane, reducing the movement of Env in the membrane, lessening its ability to form meaningful interactions with the necessary host cell receptors (Amini-Bavil-Olyaee et al., 2013; Desai et al., 2014). While IFITMs inhibit HIV-1 infection, resistance to them in transmitted founder strains of HIV-1 has been shown (Foster et al., 2016). This study also found that R5 viruses were inhibited by IFITM1 and X4 viruses were inhibited by IFITM2 and IFITM3.

1.4.3.3 TRIM5α

The TRIM family of proteins exhibit a wide range of cellular functions. Only the alpha isoform of TRIM5, TRIM5α, has been linked with antiviral activity against HIV-1. This isoform contains a PRYSPRY domain which enables its interactions with retroviral capsids (James et al., 2007; Sawyer et al., 2005). Via this interaction with HIV-1 capsid and the ubiquitin ligase activity of TRIM5α, the capsid core is prematurely disassembled and synthesis of viral cDNA is restricted (Stremlau et al., 2004). However, the human TRIM5α-mediated restriction of HIV-1 was not realised until more recently. It was shown that stimulation of the immunoproteasome via IFN-α increases TRIM5α turnover and allows TRIM5α to effectively restrict HIV-1 DNA synthesis by

binding to HIV-1 capsid lattices (Jimenez-Guardeño et al., 2019). As mentioned previously, the TRIM5α restriction of HIV-1 can be alleviated by CypA's action as an HIV-1 dependency factor (K. Kim et al., 2019; Selyutina et al., 2020).

1.4.3.4 APOBEC3G/F

Members of the APOBEC3 family are polynucleotide cytidine deaminases that edit cytosines in single-stranded DNA to uracils by hydrolysis in a zinc-dependent manner. Two members of this family, APOBEC3G and APOBEC3F, are restrictors of HIV-1, with APOBEC3G being the most potent and well-studied. APOBEC3G functions by deaminating the negative-strand single-stranded DNA synthesised during the reverse transcription process (Malim, 2009; Mangeat et al., 2003; Sheehy et al., 2002; H. Zhang et al., 2003). To do this, APOBEC3G is incorporated into virions during assembly, thus, carrying out its function during the sequential round of infection. These cytosine-to-uracil mutations in the negative-stranded single-stranded DNA, result in guanine-to-adenine base changes in the positive-strand. This hypermutation leaves the viral genome dysfunctional, resulting in its cleavage and degradation in a uracil base excision repair-dependent manner (Yang et al., 2007). A recent study has shown that APOBEC3G can exert anti-HIV-1 activity by directly binding and inhibiting reverse transcriptase (Pollpeter et al., 2017). HIV-1 accessory protein, Vif, inhibits APOBEC3G-mediated restriction of HIV-1 (Conticello et al., 2003; Sheehy et al., 2003). After binding to APOBEC3G, Vif targets it for proteasomal degradation by recruitment of a cellular ubiquitin ligase (Mehle et al., 2004).

1.4.3.5 SAMHD1

SAMHD1 regulates the quantity of deoxynucleoside triphosphate (dNTP) present in a cell by hydrolysing them into deoxynucleosides and inorganic triphosphates via dNTP triphosphohydrolase (dNTPase) activity in a GTP-dependent manner (Goldstone et al., 2011). dNTPs are fundamental components of DNA, and thus, HIV-1 requires cellular dNTPs to synthesise viral DNA during reverse transcription. In non-cycling cells dNTPs are less required, and therefore, SAMHD1 activity depletes them to low levels, strongly restricting or entirely blocking HIV-1 reverse transcription. This occurs in resting CD4+ T-cells and monocytes, making them highly resistant to infection by HIV-1 (Descours et al., 2012; Hrecka et al., 2011; Laguette et al., 2011). CDK1 and CDK2 phosphorylate SAMHD1 to inactivate its dNTPase

activity in cycling cells, such as activated CD4+ T-cells, that require high amounts of dNTPs in order to replicate their genetic information. This elevates the level of dNTPs present in the cell, rendering it more susceptible to HIV-1 infection because reverse transcription can occur (Cribier et al., 2013). SAMHD1 has also been suggested to restrict HIV-1 infection via nuclease activity on single-stranded nucleic acids, including the HIV-1 RNA genome (Beloglazova et al., 2013; Ryoo et al., 2014; White et al., 2013). While HIV-1 doesn't encode for a viral factor that inhibits SAMHD1-dependent restriction of infection, the closely related lentivirus HIV-2, has evolved the viral protein X (Vpx) which targets SAMHD1 for proteasomal degradation (Lim et al., 2012).

1.4.3.6 MX2

MX2 belongs to the dynamin superfamily and exhibits guanosine triphosphatase (GTPase) activity. It has been shown to be a potent restriction factor against HIV-1 infection, inhibiting replication between reverse transcription and integration (Goujon, Moncorgé, et al., 2013; Kane et al., 2013). MX2 is thought to inhibit the nuclear import of the PIC via direct binding to capsid. Of the two translated isoforms of MX2, antiviral activity is only demonstrated by the long isoform (Betancor et al., 2019). The antiviral activity of MX2 has been shown to be a result of a 25-amino acid long domain at its N-terminal (Goujon et al., 2014). This N-terminal domain binds to capsid via a triple arginine motif and also interacts with nuclear transport receptor protein 1 (TNPO1) and NUP214, localising it to the nuclear envelope (Dicks et al., 2018; Goujon et al., 2015). Interestingly the MX2-dependent restriction of HIV-1 is inhibited without the presence of CypA.

1.4.3.7 BST-2

BST-2, also known as tetherin, is a type II transmembrane protein that inhibits the release of fully formed HIV-1 virions (Neil et al., 2008; Van Damme et al., 2008). The N-terminal domain of BST-2 present at the site of viral budding is incorporated into the membrane of the budding virus and the C-terminal domain is anchored to the host cell membrane. As a result, release of the new virus post-scission is ineffective as it's tethered to the cell surface. This BST-2-mediated restriction of HIV-1 infection is inhibited by the HIV-1 accessory protein, Vpu. BST-2 is bound by Vpu, restricting its function and diminishing its abundance in the cell (Dubé et al., 2011; Masuyama et al., 2009; Mitchell et al., 2009).

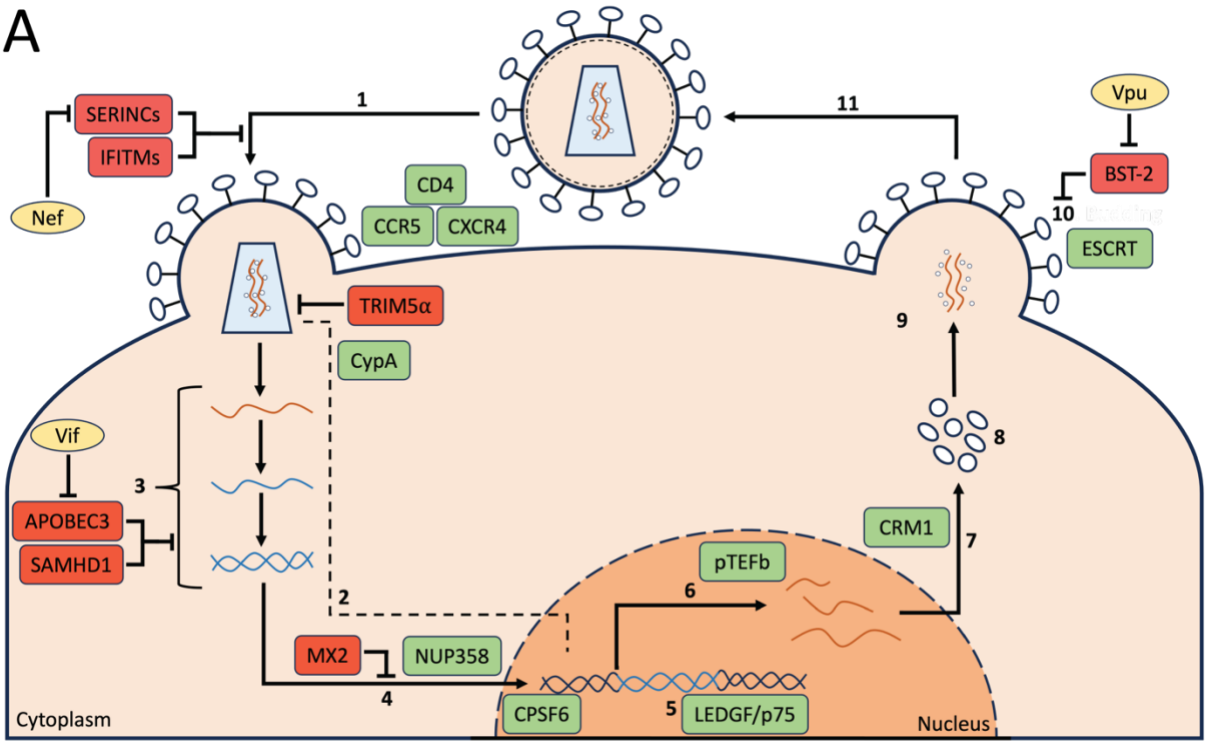


Figure 1.3: Dependency and restriction factors in HIV-1 infection.
A) Examples of identified host cellular factors that interact with HIV-1 at different points throughout its life cycle. Green = dependency factors. Red = restriction factors. Yellow = HIV-1 viral proteins that inhibit particular restriction factors. Numbers refer to stages of the viral life cycle: 1 = entry, 2 = capsid uncoating, 3 = reverse transcription, 4 = nuclear import, 5 = integration, 6 = transcription, 7 = mRNA processing and export, 8 = translation, 9 = assembly, 10 = budding, 11 = maturation. Image generated by myself in PowerPoint.

1.4.4 Identification of host factors that interact with HIV-1

These aforementioned cellular factors that either promote or suppress HIV-1 replication play an influential role in HIV-1 infection. The equilibrium between these dependency or restriction factors can be the difference between successful and unsuccessful HIV-1 replication. Thus, manipulating this balance by suppressing the function of dependency factors or boosting the effectiveness of restriction factors could significantly reduce viral replication and transmission. This makes these factors potential therapeutic targets that could result in a substantial alteration in disease progression. Only a relatively small number of these factors have been identified thus far, leaving much scope for the identification of new host proteins or mechanisms that interact with HIV-1 throughout its life cycle.

Many of the previously mentioned factors, including MX2, APOBEC3, and BST-2, were discovered using binary methods where certain genes were knocked down or ectopically expressed, and any changes to viral replication were observed (Goujon, Moncorgé, et al., 2013; Neil et al., 2008; Sheehy et al., 2002). Genome-wide RNA-interference (RNAi) screens have also been previously conducted in an attempt to uncover novel factors. However, these four RNAi based screens failed to produce cross-comparative results, with no gene being identified as a 'hit' in all four screens and very few being identified across even three (Brass et al., 2008; König et al., 2008; Yeung et al., 2009; H. Zhou et al., 2008). Intrinsic disadvantages of RNAi-based methodologies and dissimilarities in the immortalised cell-line models employed have been postulated as responsible for the variation seen between these studies (Bushman et al., 2009; Pache et al., 2011).

Recently, two screens, one of which identified the dependency factors TPST2, SLC35B2, and ALCAM, employed a pooled screening approach utilising CRISPR-Cas9 technology (OhAinle et al., 2018; Park et al., 2017). Limitations of these screens were the use of immortalised cell-line models, as well as less influential factors, denoting more mild phenotypes, being drowned out by stronger factors as a result of the use of a pooled methodology.

More recently, a study has attempted to avoid many of the previously mentioned limitations by using CRISPR-Cas9 ribonucleoproteins technology in primary CD4+ T-cells (Hiatt et al., 2022). This study focused on previously identified virus-host protein-protein interactions in

primary cells and yielded promising results; identifying and validating several new HIV-1 host factors. However, once again, this study came with some drawbacks, including the limited availability of primary cells and a bias due to focusing on a previously generated protein interactions dataset.

While screens have proved somewhat successful, experimental design can heavily impact findings when looking on a whole genome level. All of these previously mentioned screens lack the ability to assess factors that are critical to cell viability, due to the inability to knock down these genes without causing cell death. Due to these techniques only knocking out a single gene, they also lack the potential to uncover pathways or networks involved in regulating HIV-1 infection. Therefore, further screen-based studies analysing the virus-host interaction on a genome-wide level via different methodologies could lead to the discovery of new mechanisms and explanations for differing disease outcomes.

1.5 HIV-1 infection: A clinical outlook

1.5.1 Transmission

HIV-1 can pass between individuals by numerous routes of transmission, with the most common being sexual transmission via the exchange of virus containing bodily fluids at mucosal surfaces. Other transmission routes include: vertical transmission from an infected mother to the foetus, needle sharing during injection drug use, and exposure to blood during transfusions. The viral load, defined by HIV-1 RNA copies/ml, present in an infected individual is a major factor in the likelihood of transmission. A 10-fold greater viral load raises the risk of transmission by as much as 2.45 times (Fideli et al., 2001; Quinn et al., 2000). As a result, transmission risk from individuals in their acute phase of infection, where viral load is at its peak, is higher than from individuals with an established infection (Brenner et al., 2007; Wawer et al., 2005).

Despite a highly genetically diverse population of HIV-1 viruses being present in an individual, only one (or very few) variants, termed transmission founder strains, are usually responsible for a new infection (Wolfs et al., 1992; Wolinsky et al., 1992; L. Q. Zhang et al., 1993; T. Zhu et al., 1993). This founder strain tends to be an R5 tropic virus, due to CD4+ T-cells with high expression of CCR5 being abundant in the mucosal surfaces (Connor et al., 1997; Keele et al., 2008).

1.5.2 A clinically typical HIV-1 infection

After transmission, HIV-1 infection, in terms of viral load and CD4+ T-cell count, follows a typical trend (**figure 1.4A**). HIV-1 infection in a newly infected individual begins with a period of clinical silence, before viral load can be detected in the blood, called the eclipse phase, which typically lasts 7-21 days. During this phase, based upon studies performed in SIV models, HIV-1 is first thought to replicate within the localised mucosa in partially activated CD4+ T-cells before beginning to propagate in more susceptible, fully activated CD4+ T-cells (Z.-Q. Zhang et al., 1999). Migration of HIV-1, via draining lymph nodes, towards the gut-associated lymphoid tissue then occurs and marks the end of the eclipse phase as levels of replication erupt, coupled with a sharp depletion of CD4+ T-cells (Brenchley et al., 2004;

Guadalupe et al., 2003; Mehandru et al., 2004) As the virus becomes detectable in the blood, the acute phase of the infection begins. Due to a lack of adaptive immune response to the virus at this point, the substantial increase in viral replication continues and viral load reaches its peak, up to 1 x 10^7 RNA copies/ml, ~21-28 days after infection (Little et al., 1999; Piatak et al., 1993). A high variety of clinical symptoms can be experienced at this point with varying degrees of severities. The CD4+ T-cell depletion seen at this time point is thought to be a result of direct killing by the virus as well as the induction of apoptosis in bystander cells (Q. Li et al., 2005; Mattapallil et al., 2005).

After this peak in viral load during the acute phase of infection, the adaptive immune system begins to respond (discussed further in **1.5.4**). This immune response results in a substantial decrease in viral load, ~100-fold, and an increase in CD4+ T-cell counts over the next several months. The viral load then plateaus, reaching a stable level, termed the viral set point. This viral set point can vary vastly between infected individuals, as a result of influence by both viral and host factors, from a small number of HIV-1 RNA copies/ml to 1 x 10^6 copies/ml (Lyles et al., 2000; Richardson et al., 2003). A higher viral set point is strongly associated with faster disease progression (Mellors et al., 1996).

Failure of the host to effectively clear the virus results in a gradual destruction of the immune system, typically characterised by a decrease in CD4+ T-cell count over several years. The exact mechanisms by which CD4+ T-cells are lost still remain unclear but are thought to be a result of both an increase in destruction and an impairment in production (McCune, 2001). An advancement in disease progression is often correlated with a change in co-receptor usage by the virus, from CCR5 to CXCR4 (Connor et al., 1997). Over time, immunodeficiencies emerge leaving the individual vulnerable to previously preventable diseases, leading to AIDS, and eventual death. AIDS diagnosis can be defined in different ways, but is typically either at the diagnosis of an AIDS defining infection or malignancy, or when a CD4+ T-cell count of <200 cell/µl is observed. The expected period of time from HIV-1 infection to death is ~10 years, however, depending on the individual this can vary greatly.

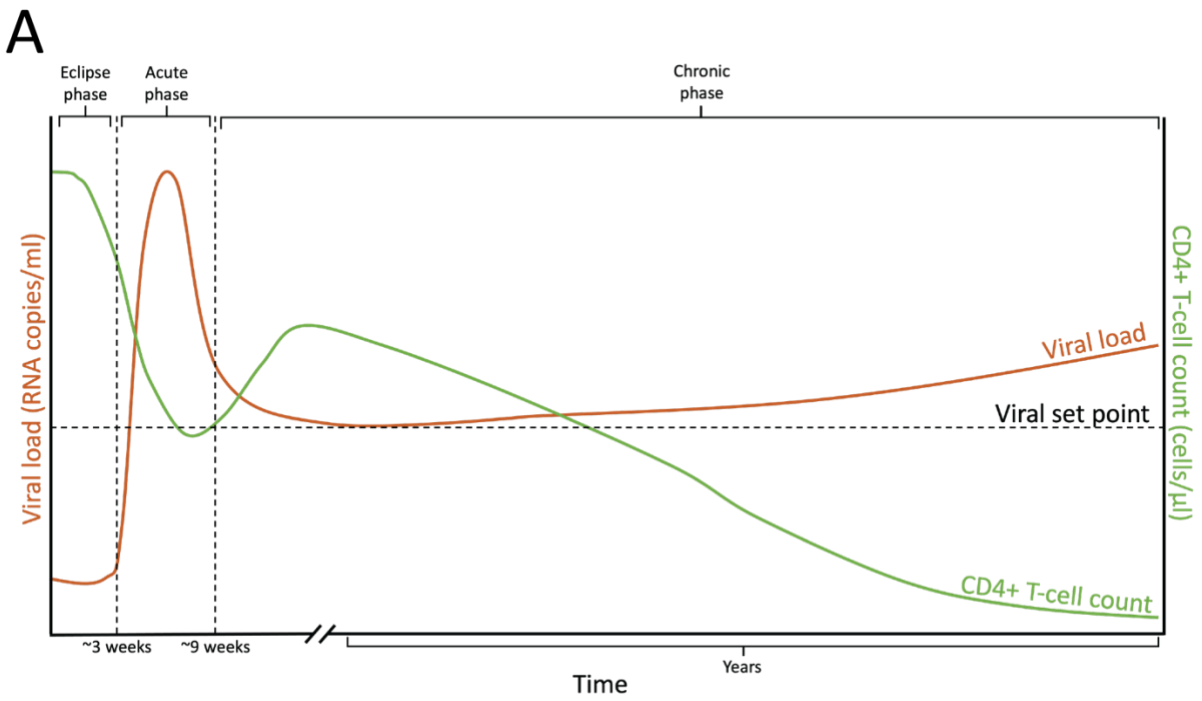


Figure 1.4: Typical trend of an individual's HIV-1 infection over time.
A) In a newly acquired HIV-1 infection, viral load remains undetectable during the eclipse phase. After ~3 weeks, viral replication explodes causing a huge peak in viral load and a steep decline in CD4+ T-cell count. As the adaptive immune system responds, viral load decreases and CD4+ T-cell numbers recover. The viral load is then maintained at a stable level, called the viral set point, for an extended period of time. This viral set point can vary greatly between individuals and is a good indication of disease prognosis. Disease progression then continues, eventually decreasing the CD4+ T-cell count to levels low enough to render the individual susceptible to AIDS-related illnesses. Green line = CD4+ T-cell count over time. Orange line = HIV-1 viral load over time. Vertical dotted lines = beginning/end of a phase of infection. Horizontal dotted line = viral set point. Image generated by myself in PowerPoint.

1.5.3 Establishment of a reservoir

Establishment of the HIV-1 reservoir in an infected individual remains the greatest challenge for curing HIV-1 infection. This reservoir has been shown, in an SIV model, to be established in as little as three days post exposure to the virus (Whitney et al., 2014). Due to the virus integrating its genome into the host genome, viral replication can commence as long as that cell remains viable. The reservoir is formed by latent infection of a long-lived cell, such as a memory or resting CD4+ T-cell, thus, remaining hidden to the host immune system (Chomont et al., 2009; Chun et al., 1998). Despite HAART being clinically effective and almost completely inhibiting viral replication, it cannot eliminate an already infected cell and, as a result, cessation in HAART results in a reinitiation of viral replication. The reservoir is believed to be stored within the lymph nodes due to a lack of antiretroviral therapy (ART) penetration and host clearance here, creating a sanctuary for the virus (Fletcher et al., 2014; Fukazawa et al., 2015). Even with zero HIV-1 replication under completely effective HAART, projections still estimate numerous decades for elimination of the viral reservoir via natural deterioration (Siliciano et al., 2003). Therefore, attempts to accelerate the natural decay or purge the viral reservoir via a 'shock and kill' strategy, are areas of focus in the search for an HIV-1 cure (Hamer, 2004).

1.5.4 The host response and why it fails

Following the virus successfully overcoming the innate immune system and establishing a persistent infection, the adaptive immune system attempts to eliminate the HIV-1 infection. This stimulation of the adaptive immunity is responsible for the ~100-fold decrease in viral load after the acute phase of infection. Most of this response is attributable to HIV-1-specific CD8+ cytotoxic T-cells that are able to recognise HIV-1 infected cells and destroy them. Additionally, roughly 12 weeks after infection, B-cells begin to produce HIV-1 neutralising antibodies. The detection of antibodies in the blood is termed as the point of seroconversion. The initial success of the adaptive immune response is thought to be due to the viral transmission event being a genetic bottleneck for the virus as only the transmission founder strains manage to establish a successful infection. These founder viruses tend to be more sensitive to CD8+ cytotoxic T-cell detection and neutralisation by antibodies, attributable to a lack of glycosylation on their gp120 (Derdeyn et al., 2004). The effectiveness of these early

responses contribute to the viral set point and the longevity for which this stable state is maintained.

Only in rare cases does this adaptive response manage to control HIV-1 infection sustainably. The selective pressure on the virus caused by the adaptive immune response, predominantly by the CD8+ cytotoxic T-cells, coupled with the high mutational rate of HIV-1, drives viral evolution away from the founder strains (Goonetilleke et al., 2009; Keele et al., 2008; Salazar-Gonzalez et al., 2009). This HIV-1 evolution leads to mutations in the epitopes recognised by CD8+ T-cells, and thus, escape from their cytotoxic effector functions. The virus evolves Env with higher glycosylation levels, forming 'glycan shields', and longer loops resulting in antibody neutralisation-resistant variants (Richman et al., 2003; X. Wei et al., 2003). Also, chronic infection resulting in prolonged immune stimulation causes immune cell exhaustion and dysfunction (Day et al., 2006; Trautmann et al., 2006).

1.5.5 Variation in HIV-1 clinical outcomes

The disease progression of an HIV-1 infection can vary significantly between individuals. After being infected with HIV-1, a small minority of individuals, <1%, may never progress to AIDS. Termed elite controllers (ECs), these individuals maintain their viral load at undetectable levels (<50 HIV-1 RNA copies/ml) in the absence of ART (Deeks & Walker, 2007). These ECs have received vast amounts of scientific interest by those attempting to uncover the influencing factors responsible for their impressive ability to control HIV-1. In contrast, some individuals progress to AIDS within a small number of years if untreated, these individuals are termed rapid progressors (RPs). The huge disparity observed between these clinical phenotypes is contributed to by both viral and host factors, many of which remain elusive.

Two population-based studies, looking at either transmissions between partners or transmissions from mothers to infants, have shown that transmission pairs display correlative viral set points (Hecht et al., 2010; Obimbo et al., 2009). As mentioned previously, the viral set point that is established in an infected individual correlates well to disease prognosis, thus, these studies suggest that viral genotype has some influence on HIV-1 clinical outcome. Two further studies, suggested that viral genotype is responsible for ~30-50% of the variation in

viral set point, denoting the remaining variation attributable to the host (Alizon et al., 2010; Blanquart et al., 2017; Fraser et al., 2014).

Host factors also determine some of the variation seen in disease progression between individuals. For example, an SIV-based model suggested that the timing, magnitude and location of the HIV-specific host adaptive immune response were contributors to the disease progression (Q. Li et al., 2009). Also, as discussed earlier, the cellular balance between HIV-1 restriction and dependency host factor can play a major role in the rate of HIV-1 infection. The natural genetic variation between infected individuals has also been shown to play a large role in the clinical outcome of HIV-1 infection and this will be discussed further in section **1.5.6**. Uncovering more of these viral and host factors remains a target of much HIV-1 scientific research as their influence on clinical outcome could provide essential insights into potential therapeutic avenues.

1.5.6 The natural genetic variation in individuals infected with HIV-1

A significant portion of the variation in disease progression of HIV-1 infected individuals is attributable to the natural genetic variation between them. This inherited variation, in the genes or pathways that interact with HIV-1 or influence the hosts immune system, may lead to a better response to infection, a lower viral set point, and thus, a better disease prognosis. Due to advancements in the understanding of the human genome and developments in the technologies surrounding it, diversity in human genetics can be more readily analysed and assessed for factors that may influence an individual's response to HIV-1 infection.

Candidate gene studies, where genes were targeted based upon their known biological function, were the initial studies used to correlate differences in disease outcomes to genetic variability. Genes confirmed to be or thought to be involved in host immune modulation or HIV-1 replication were selected. Multiple studies assessing independent cohorts in this manner generated some limited success, identifying two major genetic variations: CCR5Δ32 and human leukocyte antigen (HLA)-B alleles B*57 and B*27 (Dean et al., 1996; Huang et al., 1996; Kaslow et al., 1996; R. Liu et al., 1996).

Genome-wide association studies (GWAS) have been used to survey vast quantities of genetic polymorphisms within the human genome and assess them for their potential involvement in the variation in HIV-1 disease progression. Up to now, GWAS have predominantly been focused on single nucleotide polymorphisms (SNPs). The variation in viral set point between individuals during the chronic phase of HIV-1 infection was analysed by many early HIV-1-related GWAS, revealing and cross-validating that discrepancies in the HLA class I region were partly responsible for the clinical phenotypic differences (Fellay et al., 2007, 2009; Pelak et al., 2010). Further GWAS on HIV-1 viraemic controllers (VCs), individuals who can control their viral load <2000 HIV-1 RNA copies/ml, also exposed the same HLA class I region variations as significant in determining disease outcome (McLaren et al., 2012; The International HIV Controllers Study, 2010). Notably, the consistent findings between the VC GWAS and the general population GWAS indicated a lack of specific genetic factors influencing viral set point that are responsible for the ability of an individual to control the virus effectively. Therefore, this successful control of HIV-1 is more likely due to many genetic traits contributing minor advantages rather than a few genetic factors denoting major impacts (Khera et al., 2018). A study focusing on the protein coding sequences in the human genome was also unsuccessful in identifying clinical phenotype affecting variations outside of the HLA class I region (McLaren et al., 2017). GWAS looking for genetic traits responsible for a resistance to becoming infected with HIV-1 initially only identified variations in CCR5 as significant (Lane et al., 2013; McLaren et al., 2013). However, CD101 and ubiquitin-conjugating enzyme E2 V1 (UBE2V1) were suggested by a more recent GWAS, but these remain to be further validated (Mackelprang et al., 2017). An even more recent GWAS identified a genetic variation, specific to populations of African descent, near *CHD1L* that correlates to a ~0.3$\log_{10}$-transformed copies/ml lower viral set point (McLaren et al., 2023). They also showed that knockdowns of CHD1L *in vitro* led to increased HIV-1 replication. Despite limited success thus far, GWAS do suggest there are genetic factors in addition to CCR5 playing a role in HIV-1 susceptibility, justifying the need for larger, more extensive genomic studies.

These discovered variations in CCR5 and HLA are suggested to be responsible for ~10-15% of the variation in HIV-1 disease progression, thus, suggesting that a significant proportion of host genetic factors that influence HIV-1 disease outcome remain undiscovered (Fellay et al., 2007; Fraser et al., 2014).

1.5.6.1 Variations in CCR5

Although HIV-1 utilises two co-receptors for entry, CCR5 is the co-receptor typically utilised by HIV-1 when establishing an infection in a new individual. As a result, those homozygous for a loss-of-function allele in the gene encoding for CCR5 show high levels of resistance to HIV-1. This deletion mutation results in an inability of HIV-1 to enter the target cell due to no functional CCR5 being present on the cell surface. This was first identified via a candidate gene study on a group of MSM that, despite several HIV-1 exposures, remained seronegative (Dean et al., 1996). Though other CCR5 loss-of-function variants do exist, the most common is CCR5Δ32 (Hladik et al., 2005). With no notable frequency in most demographics, this CCR5Δ32 variation only exists in ~10% of individuals from European decent, with only ~1% presenting as homozygous (Novembre et al., 2005). These individuals heterozygous for CCR5Δ32 are not as resistant as their homozygous counterparts, but do present with a lower viral set point and slower disease progression than those not exhibiting this variation (Dean et al., 1996; Huang et al., 1996; Michael et al., 1997). Another variation at the CCR5 locus was found to have a large impact on viral set point and disease outcome by integrating direct CCR5Δ32 genotyping with GWAS data (Fellay et al., 2009; McLaren et al., 2015). Variation rs1015164, caused decreased production of CCR5-AS, an antisense long non-coding RNA, found to inhibit the degradation of CCR5 mRNA, resulting in less CCR5 expression (Kulkarni et al., 2019). Other alternate variations in the CCR5 locus may also provide some form of protection against HIV-1 but remain to be found.

1.5.6.2 HLA class I region

The HLA genes lie within an extremely genetically diverse loci, the human major histocompatibility complex (MHC). Two classes of alleles make up the HLA genes: class I, comprised of *HLA-A*, *HLA-B*, and *HLA-C*, and class II, containing *HLA-DR*, *HLA-DQ*, and *HLA-DP*. The class I HLA genes encode proteins involved in the presentation of epitopes on infected cells for CD8+ cytotoxic T-cell recognition. Variations in these class I alleles have been demonstrated to have a substantial impact on the clinical outcome of an individual infected with HIV-1, and are often present in individuals within the EC clinical phenotype group who demonstrate excellent CD8+ cytotoxic T-cell-mediated control of the virus (Ferre et al., 2009; Migueles et al., 2008).

As described above, candidate gene studies and GWAS identified variants of HLA as playing a role in determining the viral set point and disease progression of an HIV-1 infection. Individuals carrying a specific class I HLA variation, B*57:01, were found to display an average reduction in viral load of ~0.8log10 RNA copies/ml (McLaren et al., 2015). Other variations in this allele were also found to correlate with either a decrease in the viral set point (B*27:05, B*13:02, B*14:02, C*06:02, C*08:02, and C*12:02) or an increase in the viral set point (B*07:02, B*08:01, C*07:01, C*07:02, and C*04:01). The effects caused by these variations were shown to be attributed to a small number of amino acid changes. One paper identifies four amino acid changes in the HLA encoded proteins responsible for the variations in viral set point, all of which occur in the peptide-binding groove of the protein, demonstrating the significant influence of peptide presentation on the immune control of HIV-1 (Luo et al., 2021). The proposed mechanism of action of these HLA variants is by pressuring for viral evolutionary adaptations that reduce the virus's fitness (Kelleher et al., 2001; Martinez-Picado et al., 2006; Schneidewind et al., 2007). This beneficial change in epitope presentation has also been shown to increase CD8+ cytotoxic T-cell proliferation and function (Almeida et al., 2007; Horton et al., 2006; Migueles et al., 2002). HLA allele diversity within an individual has also been shown to effect disease progression, with those heterozygous for different alleles on all classes of HLA genes presenting an increased time to AIDS (Arora et al., 2020; Carrington et al., 1999). This suggests an advantageous effect of the ability to present a wider range of HIV-1 epitopes and was further supported by a recent *in silico* study (Arora et al., 2019). However, it is not just the broadness of epitope presentation that plays a role, but also the specificity of which epitopes can be uniquely presented by the HLA variations. Another *in silico* study showed that the beneficial HLA isoforms for an individual to possess, favourably select to present more critical HIV-1 epitopes (Gaiha et al., 2019). These HLA allele variations have also been suggested to alter HIV-1 disease progression via interactions with HLA-E and killer cell immunoglobulin-like receptor (KIR) proteins, which are both involved in natural killer (NK) cell activation and inhibition (Martin et al., 2002; Ramsuran et al., 2018).

1.6 The Multicentre AIDS cohort study

The multicentre AIDS cohort study (MACS) was a prospective study of men at high risk of HIV-1 infection whose clinical histories were described in great detail. The study began in 1984 and conducted clinical research at four sites distributed through the United States. Although the study recently closed, in 2019, the data and specimens collected throughout, including samples of peripheral blood mononuclear cells (PBMCs), are still available for use.

Beginning in a pre-ART era allowed the study to follow the natural progression of an individual's HIV-1 infection, without the intervention of ART, which would now be unethical to do so. This meant that differences in the natural disease progression of the MACS participants could be monitored and analysed. Participants were selected based upon epidemiological and geographic attributes, consisting of men either already infected or at high risk of infection with HIV-1. Over 7000 MSM enrolled and attended the study every 6 months to provide biological and behavioural data, as well as clinical samples. MSM were the focus of the study because of their increased risk of acquiring HIV-1 as a result of them regularly participating in high risk practices, such as receptive anal intercourse, and an increased number of exposures (Kaslow et al., 1987). The biological data collected included longitudinal analysis of CD4+ T-cell count (cells/µl) and viral load (HIV-1 RNA copies/ml), and any AIDS diagnoses. These biological data were subsequently used to group the patients based upon their clinical phenotype in response to HIV-1 infection. These clinical phenotypes ranged from those able to control HIV-1 to undetectable levels, ECs, to those who progressed to AIDS within just a few years, RPs.

The impact of the MACS on the field of HIV/AIDS has been substantial: broadening the understanding of the virus and disease; influencing clinical strategies and treatments, and contributing to more than 1400 peer reviewed scientific publications. Some of the most notable contributions of the MACS are: identifying the protective nature of the loss-of-function mutations in CCR5; uncovering the hazardous extent of receptive anal intercourse in viral transmission, and correlating the use of HAART with an increased incidence of diabetes mellitus (Brown, 2005; Dean et al., 1996; Kingsley et al., 1987).

1.7 Induced pluripotent stem cells

1.7.1 Discovery and applications

Cellular differentiation was initially thought to be a unidirectional event, consisting of a less differentiated cell developing into a more specialised cell. However, in 2006, with the addition of just four transcription factors, OCT-4, SOX2, KLF4, and c-MYC, it was demonstrated that terminally differentiated mouse fibroblast cells could be reprogrammed into cells with characteristics similar to that of an embryonic stem cell (ESC) (Takahashi & Yamanaka, 2006). The resultant cells were termed induced pluripotent stem cells (iPSCs) and the four transcription factors were named the 'Yamanaka factors'. The following year, this phenomenon was demonstrated in human cells resulting in the generation of the first human iPSCs (Takahashi et al., 2007; Yu et al., 2007).

iPSCs possess certain signatures and abilities that allow for their identification and validation after being generated. Expression of one of the Yamanka factors, octamer-binding transcription factor 4 (OCT-4), is the first typically looked for protein when assessing the pluripotency of a newly generated iPSC line. OCT-4, encoded by the *POU5F1* gene, is an indispensable regulator of pluripotency in ESCs (Boyer et al., 2005). NANOG, a second pluripotency marker of iPSCs, is an important transcription factor that represses cell fate determining genes to maintain pluripotency (Heurtier et al., 2019). Further validation of newly generated iPSCs is conducted by evaluating their ability to differentiate into all three germ layers: ectoderm, mesoderm and endoderm. The most widely regarded method for this is the teratoma assay, where pluripotent stem cells (PSCs) are injected into immuno-compromised mice. These PSCs self-renew and differentiate into non-malignant tumours, termed teratomas, which can be analysed for their make-up of cells from the different germ layers (Wesselschmidt, 2011). However, this assay is expensive, time consuming, and requires the use of animals. Another approved technique for proving pluripotency is the *in vitro* directed differentiation of PSCs into the three germ layers. After a period of differentiation, the cells can be analysed for expression of a lineage specific marker. Common markers assessed for the validation of ectoderm, mesoderm, and endoderm differentiation are PAX-6, Brachyury, and

SOX-17 respectively, which are all key transcription factors involved in the regulation of genes for the development of that particular lineage.

Being of human origin, holding the potential to self-renew, possessing the ability to differentiate into almost any cell type in the body, and the lack of ethical issues surrounding them compared to the use of ESCs make iPSCs an invaluable scientific resource. As a result, the application potential of iPSCs is vast and ever growing, spanning fields such as regenerative medicine, disease modelling, and drug discovery. iPSCs' potential for use in regenerative medicine has been demonstrated by their use in numerous clinical trials, the first of which, initiated in 2014, utilised them to treat macular degeneration (Kimbrel & Lanza, 2015).

The use of iPSCs in the field of HIV-1 has been relatively limited, with most efforts focused on the development of HIV-1 cellular therapies or modelling HIV-1 infection. Several studies have developed iPSC-derived macrophages (iMacs) that are resistant to HIV-1 after knocking out an HIV-1 dependency factor, such as CCR5 or CDK2, via either small hairpin RNA (shRNA) or CRISPR-Cas9 technology (Jerebtsova et al., 2012; Kang et al., 2015; Teque et al., 2020; Ye et al., 2014). iMacs generated from patients naturally exhibiting the CCR5Δ32 mutation were also highly resistant to HIV-1 (Chen et al., 2023). Other studies have demonstrated the potent activity of TRIM5α as an HIV-1 restriction factor by showing its knockdown or over expression in iMacs results in increased or decrease permissivity to HIV-1 infection respectively (Iwamoto et al., 2021; Kambal et al., 2011). Alternative cell therapy techniques utilising iPSCs has been to generate iPSC-derived immune cells that suppress HIV-1 infection. One study demonstrated that the introduction of iPSC-derived CD8+ T-cells specific for HIV-1 Gag epitope SL9 significantly reduced the degree of infection by a chimeric HIV-1 in a mouse model (Haque et al., 2021). This study also saw that this CD8+ T-cell action decreased expression of programmed cell death protein 1 (PD-1) on CD4+ T-cells, reducing their apoptosis and inhibition, and resulted in generation of anti-HIV-1 memory T-cells. Another study and its follow up study found that iPSC-derived NK cells can supress HIV-1 infection *in vitro* and that modification of the NK cells to express a molecule with the extracellular portion of the CD4 receptor increased the efficacy of the viral suppression (Ni et al., 2011, 2014). A potential gene therapy for HIV-1 is to target the HIV-1 promoter by shRNAs and this technique was shown to

restrict HIV-1 replication *in vitro* via the mechanism of transcriptional gene silencing (Higaki et al., 2018). iPSCs have also been suggested for use as a means of vaccinating against HIV-1. A significant antigen-specific response was exhibited when iPSCs encoding for gp160 were introduced into mice (Yoshizaki et al., 2011). The study of HIV-1-associated neurocognitive disorder (HAND) has also been advanced by utilising iPSCs and their ability to form organoids. HAND occurs as a result neuroinflammation and glial activation induced by the presence of HIV-1 in the central nervous system (CNS), leading to neurodegeneration. This has been previously difficult to study due to the lack of a good model of an HIV-1 infected human brain, however, the development of an iPSC-derived 3D brain organoid which inherently produces microglia, the target cell of HIV-1 in the CNS, provides a good foundation for its study (Ormel et al., 2018). A few labs have utilised this 3D organoid model in the context of HIV-1, showing that HIV-1 infected microglia induce additional inflammation compared to uninfected microglia via elevated production of tumour necrosis factor α (TNF-α) and interleukin (IL) 1β (dos Reis et al., 2020; Gumbs et al., 2022; Z. Wei et al., 2023). Finally, a recent pre-print has described a mouse model for studying the acute HIV-1 infection of microglia and its association with HIV-1 crossing the blood-brain barrier by utilising iPSC-derived human microglia (Min et al., 2023). Despite these progressions in the use of iPSCs in the HIV-1 field for clinical aspirations or understanding HIV-1 infection on a disease level, little has been described regarding their use in understanding the interaction between HIV-1 and its host on a molecular level.

1.7.2 Reprogramming

The introduction of the Yamanaka factors into somatic cells to induce their reprogramming into iPSCs has been conducted via multiple methodologies. These methodologies, discussed below, each have advantages and disadvantages in striving for higher reprogramming efficiencies and the derivation of footprint-free iPSCs.

1.7.2.1 Integrative viral vectors

The first reprogramming vectors used were retroviruses due to their existing usage for transferring genetic material. However, due to being silenced by epigenetic modifications shortly after integration into the host genome, they only provided temporal gene expression, thus impacting the quality of the iPSCs produced (Matsui et al., 2010; Stadtfeld et al., 2008).

Lentiviral vectors were then utilised because of their ability to infect both dividing and non-dividing cells. With a reprogramming efficiency of 0.1-1%, lentiviral vectors remain one of the most efficient reprogramming vector systems (Stadtfeld & Hochedlinger, 2010). However, both of these methods involve integrating of the transgenes into the host genome, creating the potential for mutagenesis, which raises significant concerns for the generated iPSCs, particularly those for use in a clinical setting. Attempts to generate lentiviral vectors that leave no genomic footprint, using an excisable vector, were successful but still left traces of impact on the host genome, therefore, are unfavourable to use (Papapetrou & Sadelain, 2011).

1.7.2.2 Nonintegrative viral vectors

Adenoviruses showed great promise for producing iPSCs with no genomic footprint, however, their reprogramming efficiency is very low, at 0.0002% (W. Zhou & Freed, 2009). This low efficiency makes adenoviruses a poor choice as a reprogramming vector and the use of a 'double hit' adenovirus system to increase reprogramming efficiencies is unfavoured.

Sendai virus is an RNA virus that produces large quantities of protein and doesn't require entry into the nucleus in order to replicate, making it a potentially excellent candidate for generating iPSCs. Reprogramming using sendai vectors can be conducted on both fibroblasts and blood cells (e.g. PBMCs) at reasonably high efficiencies of ~1% and ~0.1% respectively (Fusaki et al., 2009; Seki et al., 2010). A further advantage of using Sendai virus as the vector for reprogramming is that commercial kits are available, allowing for reliable and reproducible reprogramming experiments on demand. A limitation of Sendai virus use is that multiple passages of the generated iPSCs are required for their passive removal, however, use of temperature-sensitive Sendai vectors can reduce this time period (Ban et al., 2011). These points make Sendai virus-mediated generation of iPSCs a popular choice within the field.

1.7.2.3 Nonviral vectors

Introduction of mRNA to generate iPSCs has shown potential but still suffers from some limitations. High reprogramming efficiencies, 1.4-4.4%, can be attained, but initial disadvantages such as the labour intensive requirement of multiple rounds of transfection and a restriction to fibroblasts limited their use (Warren et al., 2010). Recent efforts have alleviated

this labour intensity greatly, however, reprogramming of PBMCs still remains a challenge when using mRNA based systems (Warren & Lin, 2019).

Reprogramming by microRNAs (miRNAs) has been shown by three separate studies to efficiently reprogramme human somatic cells into iPSCs in the absence of the Yamanaka factors (Anokye-Danso et al., 2011; Miyoshi et al., 2011; Subramanyam et al., 2011). The studies all utilised different methodologies and miRNA combinations and none of the studies have yet been reproduced, making it difficult to define any of these approaches as reliable for reprogramming.

Utilising a PiggyBac creates genetic footprint-free iPSCs. PiggyBac is a transposon that can integrate into the host genome, complete its function, and then be removed without leaving any trace. This method produces a slightly low reprogramming efficiency of 0.02-0.05% and has mostly been demonstrated thus far in murine cells, with little work completed on human mesenchymal stem cells (Kaji et al., 2009; Mali et al., 2010; Woltjen et al., 2009).

Minicircle vectors, vectors consisting of just a promoter and the cDNA to be expressed, can also be used to generate iPSCs. However, this method has only been demonstrated in adipose stromal cells and neonatal fibroblasts, both with very low reprogramming efficiencies (Jia et al., 2010; Narsinh et al., 2011).

The transient transfection of OriP/EBNA-based plasmids, which are expressed for longer than standard plasmids, is a another methodology for reprogramming somatic cells into iPSCs. A single round of transfection can generate plasmid and transgene free iPSCs with no genomic footprint (Yu et al., 2009). Episomal transfections using non-Yamanaka factor-based reprogramming factors have also been demonstrated (Okita et al., 2011). The episomal plasmids are passively lost from the cells over divisions, eventually leading to complete removal. This cost effective method remains a popular choice for reprogramming cells into iPSCs.

1.7.3 The effect of reprogramming on the cell

The reprogramming process or the extensive long-term culturing of iPSCs has been shown to induce genetic and epigenetic alterations in iPSCs compared to their somatic cell predecessors. For example, the aforementioned clinical study utilising iPSCs for treating macular degeneration had to be ceased early due to identification of copy-number variations (CNVs) and single nucleotide variations (SNVs) in the iPSCs that were not present in the original somatic cells (Garber, 2015).

1.7.3.1 Genetic alterations

iPSC lines are derived from a population of stromal cells that may contain slight genetic difference between each other. Therefore, a clonal iPSC line, that is derived from one stromal cell, may contain genetic differences when compared to the other stromal cells present in that initial population. This renders it difficult to decipher whether any genetic abnormalities between the iPSCs and stromal cells were pre-existing variations within the stromal cell population, were produced during the reprogramming process or were created during the culturing of the iPSCs. One study showed that the older the donor of the somatic cells was, the more mutations there were present within the stromal cell population, and therefore, a greater genetic variability in generated iPSC clones was observed (Lo Sardo et al., 2017).

Other genetic differences between iPSC lines and the stromal cells from which they were derived may have occurred as a result of the reprogramming process. To assess this, a study compared CNVs in generated iPSCs to CNVs in the parental fibroblast lines (Hussein et al., 2011). Numerous CNVs were formed as a result of reprogramming, however, cells exhibiting these CNVs were less fit, and thus, were diluted out of the population over prolonged passaging. This trend was seen across different methods of reprogramming, with non-integrating methods creating fewer CNVs (Bhutani et al., 2016; Cheng et al., 2012; Schlaeger et al., 2015). One study also showed that the reprogramming process causes hundreds of point mutations in the resultant iPSCs, although, accurately distinguishing between pre-existing mutations and those induced by reprogramming is difficult (Sugiura et al., 2014).

Extended culture periods can also induce genetic variations in iPSCs. In culture, p53 mutations can be spontaneously acquired and become more prevalent in the population over passages

(Merkle et al., 2017). These mutations can result in extensive DNA lesions developing throughout the iPSC line due to the loss of p53 functions. Therefore, iPSCs with lower passage numbers are preferred.

1.7.3.2 Epigenetic alterations

The reprogramming process must induce large epigenetic changes, such as changes in DNA methylation patterns, in order to reprogramme a cell from a differentiated state into a pluripotent state. During this process any unusual modifications to DNA or histones are called epimutations and can have a large impact on the fate of the affected cell. As with genetic mutations, epimutations can develop from the pre-existing epigenetic state, during the reprogramming process, or as a result of prolonged culture.

Reprogrammed cells can often maintain some of the epigenetic signature of the somatic cell from which they were derived, called epigenetic memory. This may result in some lineage-specific genes remaining expressed or pluripotent genes remaining repressed causing a bias or limitation in the differentiation potential of the iPSCs (Bar-Nur et al., 2011; Ghosh et al., 2010; Horvath, 2013; K. Kim et al., 2010; Marchetto et al., 2009). Age of the somatic cell donor can also play a role in this epigenetic memory signature as some regions in the genome, particularly those near genes related to differentiation and development, can become hypermethylated with increasing age (Benayoun et al., 2015; Lo Sardo et al., 2017). These hypermethylated spots may remain during the reprogramming process, meaning older cells are more difficult to fully reprogramme.

The reprogramming process can also have an impact on the cell's epigenetics aside from the expected changes required to induce pluripotency. As mentioned above, traits of the epigenetic signature from the somatic cell can be retained by the iPSCs after the reprogramming process. The reprogramming method can affect this, as cells reprogrammed by somatic cell nuclear transfer (SCNT), where a somatic cell nucleus is inserted into an unfertilised oocyte, can result in complete reprogramming, generating PSCs highly comparable to an ESC, with no epigenetic memory from the somatic cell remaining (Ma et al., 2014). This may be a result of specific factors present within the oocyte (Han et al., 2015). New changes to DNA methylation that are unique to the reprogrammed cell, called

differentially methylated regions, can also occur during the reprogramming process. These regions, that typically exhibit a reduction in methylation, are often matched in unrelated iPSC lines, demonstrating a preference of this in certain regions of the genome (Lister et al., 2011).

An iPSC's epigenetic memory acquired from the somatic cell it was reprogrammed from diminishes with extended culture periods (Nishino et al., 2011). However, also with increased culture times, epigenetic abnormalities can increase, an example of which is the loss in X chromosome inhibition often observed in iPSCs derived from female donors (Silva et al., 2008; Tchieu et al., 2010).

1.7.3.3 Imprinting misregulation

An imprinted gene is a gene where one of its two copies is switched off via DNA methylation. This is determined by the parents during gametogenesis and is maintained throughout life (Okae et al., 2014). Imprinting allows tight regulation by monoallelic expression and its loss can result in biallelic expression, which can be potentially damaging. Within iPSCs, loss of imprinted genes is more common in those received paternally and occurs favourably in certain genes, including *IGF2*, *H19*, and *MEG3* (Bar et al., 2017; K.-P. Kim et al., 2007; Pick et al., 2009; Rugg-Gunn et al., 2007).

As with the genetic or epigenetic mutations in iPSCs, changes to imprinting may already be pre-existing in the original somatic cell (Bar et al., 2017). Demonstrated by reprogramming of fibroblasts from patients with imprinting diseases, Prader Willi syndrome and Angelman syndrome, these changes to imprinting within the somatic cell are maintained in the generated iPSCs (Chamberlain et al., 2010).

The process of reprogramming can also result in loss of imprinting, as iPSCs show a higher level when compared to ESCs (Johannesson et al., 2014; Pick et al., 2009). As with epigenetic alterations, the reprogramming method can affect this, with cells reprogrammed using SCNT displaying less imprinting modifications than cells reprogrammed with the Yamanaka factors (Ma et al., 2014).

Imprinting stability has been shown to be affected by the *in vitro* culture of iPSCs (Bar et al., 2017). Additionally, imprinted genes are often those associated with regulators of cell growth (e.g. *IGF2*), thus, loss of imprinting on these genes grants a selective advantage to these cells within the culture (Rugg-Gunn et al., 2007).

1.7.4 Use as a tool for a permissivity to HIV-1 screen

As mentioned previously, the limited identification of host cell factors and host genetic factors that influence HIV-1 replication leaves much scope for the discovery of new factors. The use of iPSCs as a tool for an HIV-1 based screen holds the potential to uncover some of these elusive factors. Firstly, iPSCs possess the potential to self-renew. This allows for the creation of a stable cell line giving a virtually inexhaustible source of cells for use in the screening assays. A major advantage of iPSCs is that the majority of differences between two cell lines will be dictated by the natural variation between the donors from which they were derived (Vigilante et al., 2019). The generation and maintenance of iPSC lines is performed in a fairly consistent manner, resulting in iPSC lines being in a mostly similar pluripotent state. Therefore, iPSCs are largely comparable to each other and have the potential to generate more reliable and reproducible data. Whereas, for example, assays performed on a single patient's PBMCs can display greater variation each time a new sample is collected. iPSCs have also been shown to inherently express high levels of ISGs, similar to that of an HIV-1 target cell after its stimulation by IFN as a result of viral infection (Wu et al., 2018). Innately high expression of these genes may allow easier identification of any unidentified ISGs. Previously conducted HIV-1 screens have predominantly adopted strategies that manipulate the genetics of the cell via knockouts or over expression, restricting the study of genes that are critical for cell survival or may be toxic if over expressed. These techniques also tend to limit the screen's output to an independent gene that impacts HIV-1 replication. In contrast, by utilising the natural variation within iPSCs derived from different donors, all genes and whole regulatory pathways or networks can be analysed.

Since iPSCs do not express the cellular receptor, CD4, they are not a naturally targeted cell by HIV-1. Therefore, for their use in an HIV-1 based permissivity screen, a pseudotyped HIV-1 must be utilised. The glycoprotein of the vesicular stomatitis virus (VSV-G) is commonly used to pseudotype viruses since its mechanism of entry utilises a commonly expressed molecule,

the low density lipoprotein receptor (LDL-R). The use of a VSV-G pseudotyped HIV facilitates its entry into iPSCs, as they express the LDL-R, however, it does divert the mechanism of entry away from that normally used by HIV-1. After binding, the VSV-G pseudotyped HIV-1 gains access to the cell via a clathrin-mediated endocytic pathway (Cureton et al., 2009; Johannsdottir et al., 2009). The viral and endocytic vesicle membranes then fuse, facilitated by the vesicle's acidic environment, releasing the HIV-1 capsid into the cell cytoplasm which then follows its usual procedure to successfully establish a productive infection within the iPSC.

1.8 Thesis aims

Current ARTs against HIV-1 still fail to eradicate the virus from the patient. Therefore, further understanding of how HIV-1 avoids the host immune system and negotiates host cell proteins to replicate resides as an important hurdle for the HIV/AIDS field to overcome.

This study uses a novel approach in an attempt identify HIV-1 regulatory factors, pathways and networks. PBMCs from participants of the MACS will be reprogrammed into iPSCs which will then be assessed for variable levels of permissivity to a VSV-G pseudotyped HIV-1 based GFP vector and a VSV-G pseudotyped replication competent HIV-1, strain NL4-3. By correlating the permissivity phenotypes of each MACS iPSC line to its transcription profile, candidate genes potentially responsible for the differing permissivity phenotypes to HIV-1 will be identified. The permissivity phenotypes can also be validated against the clinical phenotypes of the MACS participants and candidate genes can then validated and assessed for their mechanism of involvement in HIV-1 infection.

The aims of this study are:

Aim 1: To establish and employ a reprogramming pipeline for the production of validated iPSCs from MACS PBMC samples.

Aim 2: To address differing levels of permissivity to HIV-1 by utilising natural variation between MACS-derived iPSC lines.

Aim 3: To use comparative transcriptomic analysis to identify HIV-1 regulatory factors responsible for the extreme permissivity phenotypes of MACS iPSC lines.

Chapter 2: Materials & Methods

2.1 Plasmids and bacteria

2.1.1 Plasmids

The plasmids used to generate replication-incompetent lentivirus particles were: p8.91, packaging plasmid encoding virion proteins Gag-Pol, Tat and Rev (pCMV-ΔR8.91), pGFP, an engineered gRNA encoding GFP under transcriptional control of the SFFV promoter (pHRsin.SFFV.eGFP), and pVSV-G, envelope plasmid that expressed VSV-G (pMD2.G).

The plasmids used to generate replication-competent lentivirus particles were: pNL4-3, wild type provirus NL4-3 cloned into a plasmid, and pVSV-G.

2.1.2 Plasmid production in competent bacteria

Plasmids were propagated in competent *Escherichia coli* (*E. coli*) strain Stbl 2 (ThermoFisher). Liquid cultures used sterile Luria-Bertani (LB) broth (1% sodium chloride (w/v), 1% tryptone (w/v) and 0.5% yeast extract (w/v) in Type 1 ultrapure water) and solid-phase cultures used sterile LB-Agar (20g/L in Type 1 ultrapure water). Culture media was supplemented with 100µg/ml Ampicillin (Sigma-Aldrich) as necessary.

Transformation was conducted by addition of 1ng of plasmid DNA to 50µl of Stbl 2 *E.coli* followed by incubation for 30 minutes on ice. Bacteria were then heat-shocked for 45 seconds at 42°C and recovered with a 2 minute incubation on ice. 450µl of LB broth was then added and the culture was incubated for 45 minutes at 37°C (30°C for proviral plasmids) in a shaking incubator at 200rpm. 100µl of this culture was then plated on an LB-agar plate containing ampicillin overnight at 37°C (30°C for proviral plasmids). Individual colonies were selected and used to inoculate 5ml of LB broth containing ampicillin which was incubated for 6-8 hours at 37°C (30°C for proviral plasmids) in a shaking incubator at 200rpm. This culture was then added to 300ml of LB broth containing ampicillin and incubated overnight at 37°C (30°C for proviral plasmids) in a shaking incubator at 200rpm.

The following day, the cultures were pelleted by a 15-minute centrifugation at 6000*xg* at 4°C. The supernatant was removed and purification of plasmid DNA followed using the

Nucleobond Xtra Maxi kit (Machery-Nagel), according to the manufacture's guidelines. Purified plasmid DNA was reconstituted to 1µg/µl in nuclease-free water. Plasmid integrity was confirmed by an appropriate restriction enzyme digestion.

2.2 Cell culture

2.2.1 Cell lines

Human embryonic kidney (HEK) 293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) with GlutaMAX (ThermoFisher) supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 1% penicillin-streptomycin (P/S) (Sigma). When near confluency, adherent HEK293Ts were passaged using TrypLE Express with phenol red (ThermoFisher).

ACH-2 cells were cultured in RPMI-1640 medium with 2mM L-Glutamine (ThermoFisher) supplemented with 10% heat-inactivated FBS, 10mM HEPES (Gibco), 1X MEM non-essential amino acids (NEAA) (Gibco) and 1% P/S. Suspension ACH-2 cells were passaged every 3-4 days to a concentration of 1 x 10^6 cells/ml.

U1 cells were cultured in RPMI-1640 medium with 2mM L-Glutamine supplemented with 10% heat-inactivated FBS, 1X MEM NEAA, and 1% P/S. Suspension U1 cells were passaged every 3-4 days to a concentration of 1 x 10^6 cells/ml.

iPSCs generated by feeder-dependent culture were maintained on a confluent layer of mouse embryonic fibroblasts (MEFs) (Life Technologies), plated according to the manufacturer's protocol. The cells were maintained in K+/FGF media consisting of advanced DMEM F12 (AdDMEM) with 2mM L-Glutamine (ThermoFisher) supplemented with 20% knockout serum replacement (Life technologies), 0.1 mM 2-mercaptoethanol (Life technologies), 6ng/ml of fibroblast growth factor (FGF) (Qkine) and 1% P/S. Feeder-dependent iPSCs were passaged using dispase (Gibco) and collagenase IV (Invitrogen) at 1g/L in AdDMEM and K+/FGF respectively.

iPSCs generated by feeder-Free culture were maintained on plates coated with vitronectin (VTN-N) (ThermoFisher) following the manufacture's protocol. The iPSCs were maintained in complete Essential 8 medium (Gibco) supplemented with 1% P/S (E8+) from Monday-Friday and in complete StemFlex medium (Gibco) supplemented with 1% P/S from Friday-Monday (SF+). iPSCs were passaged onto freshly coated plates when colonies began to touch using

0.5mM ethylenediaminetetraacetic acid (EDTA) (Life technologies) after one wash with phosphate buffered saline (PBS).

All cells were cultured at 37°C at 5% CO_2 in a humidified atmosphere. Temperature was raised to 38.5°C for 5 days to aid the removal of the Sendai viral vectors from newly generated iPSCs.

2.2.2 PBMC isolation, thawing, sorting, and activation

PBMCs used in the optimisation of reprogramming experiments were isolated from the venous whole blood of healthy volunteers obtained via the King's College London BioBank. PBMCs were isolated by density gradient centrifugation separation using SepMate-50 centrifuge tubes (STEMCELL Technologies) and Lymphoprep density medium (STEMCELL Technologies) according to the manufacturer's instructions. PBMCs were either used immediately or cryopreserved in FBS + 10% dimethyl sulfoxide (DMSO) and stored in liquid nitrogen.

Upon thawing, all PBMCs, including after sorting, were maintained in RPMI-1640 medium with 2mM L-Glutamine supplemented with 10% heat-inactivated FBS, 1% P/S, and 10µg/ml DNase I (Roche) for a 30 minute rest period prior to commencing the reprogramming pipeline.

CD4+ cells were depleted from the bulk PBMC population by magnet-activated cell sorting (MagACS) using CD4 Microbeads, human (Miltenyi Biotec) following the manufacturer's instructions.

CD4+ T-cells were isolated from PBMCs by negative MagACS using the CD4+ T-cell Isolation Kit, human (Miltenyi Biotec) according to manufacturer's instructions. CD4+ T-cells were activated by supplementing the culture media, RPMI-1640 medium with 2mM L-Glutamine supplemented with 10% heat-inactivated FBS and 1% P/S, with Dynabeads Human T-Activator CD3/CD28 beads at a 1:1 cell to bead ratio (ThermoFisher), and 30U/ml recombinant interleukin-2 (rIL-2) (Roche) for 48 hours prior to infection.

2.2.3 Reprogramming

Reprogramming of PBMCs was carried out using two different protocols. Both protocols utilised the CytoTune-iPS 2.0 Sendai Reprogramming Kit (ThermoFisher) which consisted of three Sendai viral vectors, KOS, KLF4, and c-Myc, that encode for the four Yamanka factors. The feeder-dependant protocol was provided by the Wellcome Sanger Institute and the feeder-free protocol was conducted according to the manufacturer's protocol; both are outlined below.

2.2.3.1 Feeder-dependant protocol

9 days prior to transduction with the reprogramming Sendai vectors, 5 x 10^6 PBMCs, either fresh or cryopreserved, were seeded at 1 x 10^6 cells/ml in wells of a 6-well plate in expansion media (EM) consisting of StemSpan H3000 media (STEMCELL Technologies) supplemented with 50µg/ml ascorbic acid (Sigma), 100ng/ml stem cell factor (SCF) (ThermoFisher), 10ng/ml IL-3 (ThermoFisher), 3U/ml erythropoietin (R&D Systems), 40ng/ml insulin-like growth factor 1 (IGF-1) (Miltenyi Biotech), 1µM dexamethasone (Sigma), and 1% P/S. Spent media was exchanged 6 and 3 days prior to transduction. On the day of transduction, EM was prepared with the addition of polybrene (Millipore) at 8µg/ml. A virus mix containing the three Sendai reprogramming viral vectors, KOS, KLF4, and c-Myc, was prepared in EM + polybrene at multiplicity of infections (MOIs) of 5, 3, and 5 respectively. 3 x 10^5 cells were then incubated overnight in 300µl of virus mix in a 12-well plate. 24 hours post-transduction, the cells were washed and reseeded in 2ml of EM. A further 48 hours later, the transduced cells were transferred onto a 10cm plate seeded with a layer of MEFs. Cellular adaptation to iPSC media was then begun day 5 post-transduction by transferring the transduced cells into K+ media supplemented with the growth factors of EM. After another 48 hours, K+ with EM growth factors was replaced with K+/FGF to form the complete iPSC media. Spent K+/FGF was replaced 48 hours later and changed daily thereafter until 15-21 days post-transduction. When colonies had grown to an appropriate size, individual colonies were picked and expanded as separate clones on a feeder-dependent system.

2.2.3.2 Feeder-free protocol

4 days prior to transduction, PBMCs, cryopreserved or fresh, and CD4+ cell depleted or not depleted, were seeded at 5 x 10^5 cells/ml in 1ml of PBMC media in wells of a 24-well plate.

PBMC medium consisted of StemPro-34 media (ThermoFisher) supplemented with 2mM L-Glutamine, 100ng/ml SCF, 100ng/ml FLT-3 (ThermoFisher), 20ng/ml IL-3, 20ng/ml IL-6 (ThermoFisher), and 1% P/S. On each of the next three days, 500µl from all wells was exchanged for fresh PBMC media. On transduction day, 3 x 10^5 cells were added to a 14ml round-bottom tube (Scientific Laboratory Supplies Ltd) for transduction with the three Sendai reprogramming vectors. The volume of each vector, KOS, KLF4, and c-Myc, was added to 1ml of PBMC media to give MOIs of 5, 3 and 5 respectively. The virus and cells were mixed and then centrifuged at 1000*xg* for 30 minutes at room temperature. On completion of centrifugation, an additional 1ml of PBMC media was added and the cells were re-suspended. The cell-virus mix was then transferred to one well of a 12-well plate and incubated overnight. 24 hours post-transduction, the cells were washed and resuspended in 500µl of PBMC media in one well of a 24-well plate. A further 48 hours later, the cells were plated on a vitronectin coated 6-well plate in 2 ml of StemPro-34 medium without cytokines at varying densities between 1 x 10^4-1 x 10^5 cells per well. After 2 days, 1ml of spent media was exchanged for fresh StemPro-34 medium without cytokines. 7 days post-transduction, 1ml of spent StemPro-34 medium was replaced by 1ml of E8+ to start the cell adaptation to iPSC media. The following day, and every day thereafter, the full spent media was replaced with fresh E8+. When colonies had grown to an appropriate size, individual colonies were picked and expanded as separate clones.

2.2.4 Trilineage differentiation

Pluripotency of the MACS iPSC lines was confirmed using the STEMdiff Trilineage Differentiation Kit (STEMCELL Technologies). On day 0, high quality iPSC line cultures were dissociated to single cells using Gentle Cell Dissociation Reagent (GCDR) (STEMCELL Technologies). Cells were centrifuged, had their supernatant removed and were resuspended in E8+ supplemented with 1X RevitaCell (ThermoFisher). Cells were plated at densities depending on their differentiation route in wells of a 96-well plate that were coated with Matrigel (Corning) according to the manufacturer's instructions. 16,000 cells were plated per well for those to remain as iPSCs and those being differentiated into the mesoderm. 64,000 cells were plated per well for those being differentiated into either the ectoderm or endoderm. Media was replaced after 24 hours and every day thereafter with 200µl of the appropriate iPSC/differentiation medium. iPSCs, mesoderm differentiated cells, and

endoderm differentiated cells were fixed on day 4, and ectoderm differentiated cells were fixed on day 6 using 200µl of 4% paraformaldehyde (PFA) and incubated for 15 minutes at room temperature. This was then replaced with 200µl of MagACS buffer (PBS supplemented with 0.5% Bovine serum albumin (BSA) and 2mM EDTA), and the plates were wrapped in parafilm and stored at 4°C until used later for assessment of lineage specific marker expression by confocal microscopy, as described in section **2.6.2**.

2.2.5 G-banded karyotype analysis

G-banded karyotype analysis of the MACS iPSC lines was outsourced to Cell Guidance Systems. However, due to the MACS iPSCs being maintained in a containment level 3 (CL3) laboratory, the cells were prepared and fixed prior to sending, following the company's protocol that is available online.

2.3 Lentivirus production and quantification

2.3.1 Production of VSV-G pseudotyped HIV-1 based GFP vector

HEK293Ts were seeded on 10cm plates to be 90-95% confluent the following day. The transfection mix was prepared in reduced serum medium Opti-MEM (ThermoFisher) containing 4µg p8.91, 4µg pGFP, 0.5µg pVSV-G, and polyethyleneimine (PEI) (Poly Sciences) at a ratio to plasmid DNA amount optimised for that batch. The transfection mix was vortexed thoroughly and incubated for 20 minutes at room temperature, before being added to the near confluence plate of HEK293Ts. Media was exchanged 18 hours later and the lentiviral vector-containing supernatant was harvested 48 hours post-transfection.

2.3.2 Production of VSV-G pseudotyped full length HIV-1, strain NL4-3

The production was conducted as in section **2.3.1** except the transfection mix contained 10µg of pNL4-3 and 0.5µg of pVSV-G as well as PEI.

2.3.3 Purification and concentration of lentivirus particles

The harvested lentivirus particle-containing supernatant was filtered through a 0.45µm filter and gently layered above a 20% sucrose in PBS solution in 36ml ultracentrifuge tubes. The samples were then ultracentrifuged for 75 minutes at 150,000*xg* at 4°C with no brake. The supernatant was then removed and purified lentivirus particles were resuspended in Opti-MEM, aliquoted and stored at -80°C.

2.3.4 Quantification of lentivirus stocks

VSV-G pseudotyped HIV-1 based GFP vector was quantified by a titration experiment on HEK293T cells. 1 x 10^4 HEK293T cells were seeded in a 96-well plate and transduced the following day with the concentrated vector using a 1:4 serial dilution. Cells were harvested 48 hours post-transduction and the percentage of cells expressing GFP, analysed by flow cytometry, was used to determine the vector titre.

VSV-G pseudotyped full length HIV-1, strain NL4-3, was quantified using the HIV p24 high sensitivity AlphaLISA Detection Kit (PerkinElmer) following the manufacturer's instructions.

2.4 Lentiviral infections

2.4.1 Infection of activated CD4+ T-cells

Activated CD4+ T-cells were collected, pelleted and resuspended in fresh medium supplemented with rIL-2, and seeded at a concentration of 4 x 10^6 cells/ml in 500µl in a 24 well-plate. A desired amount of virus stock was added to the cell suspension and mixed before proceeding with spinoculation for 120 minutes at 2,000*xg* at 30°C. 500µl of fresh medium was then added to cells before overnight incubation.

2.4.2 iPSC permissivity to HIV-1 infection pipeline

The screen consisted of two arms designed to test the permissivity of iPSCs to either a VSV-G pseudotyped HIV-1 based GFP vector or a VSV-G pseudotyped full length HIV-1, strain NL4-3, in an infectivity or late gene expression assay respectively (**figure 2.1A**). For the set-up of both assays in parallel, high quality iPSCs ready for passaging were dissociated to single cells using GCDR and pelleted. The singlet iPSCs were resuspended in E8+ supplemented with 1X RevitaCell. 3 x 10^4 cells of each iPSC line were then seeded in 300µl of E8+ and 1X RevitaCell in 12 wells of a 48-well plate that was coated with a double concentration of vitronectin and then incubated for 16 hours. The 12 seeded wells for each assay were used for two biological replicates of a 5-fold dilution of the appropriate virus, plus an uninfected control well.

For the infectivity assay, VSV-G pseudotyped HIV-1 based GFP vector was diluted in SF+ medium so that there was 0.206µl (which equated to 3.3 x 10^5 transducing units) of vector per 300µl of media. 1 in 4 serial dilutions of this were then prepared, resulting in 5 distinct dilutions of vector. The supernatant was then removed from all wells and 300µl of the appropriate vector dilution was added to each well. SF+ containing no vector was added to the uninfected control well. 48 hours post-infection, the supernatant was aspirated and 200µl of 4% PFA was added to the cells for 15 minutes at room temperature. This was then removed and replaced with 200µl of MagACS buffer. The plates were then wrapped in parafilm and stored at 4°C until analysed for percentage of cells expressing GFP, as described in section **2.6.3** and **2.6.4**.

For the late gene expression assay, VSV-G pseudotyped full length HIV-1, strain NL4-3, was diluted in E8+ medium so that there was 0.015µl (which equated to 8ng of p24) of virus per 300µl of media. 1 in 2 serial dilutions of this were then prepared, resulting in 5 distinct dilutions of virus. The supernatant was then removed from all wells and 300µl of the appropriate virus dilution was added to each well. E8+ containing no virus was added to the uninfected control well. 24 hours post-infection the supernatant was aspirated, each well was washed once with 200µl PBS and 200µl of E8+ was added to the cells. After a further 24 hours, the supernatant was removed and 200µl of 4% PFA was added to the cells for 15 minutes at room temperature. This was then removed and replaced with 200µl of MagACS buffer. The plates were then wrapped in parafilm and stored at 4°C until analysed for percentage of cells expressing HIV-1 Gag, as described in section **2.6.3** and **2.6.4**.

Note: Only one production batch of each virus was used for all iPSC infection experiments.

In parallel to seeding for infection, 1 x 10^6 cells of each iPSC line were also seeded into double concentration vitronectin coated 6-well plates. These cells were harvested after 16 hours, and stored as a dry pellet at -80°C for later RNA extraction.

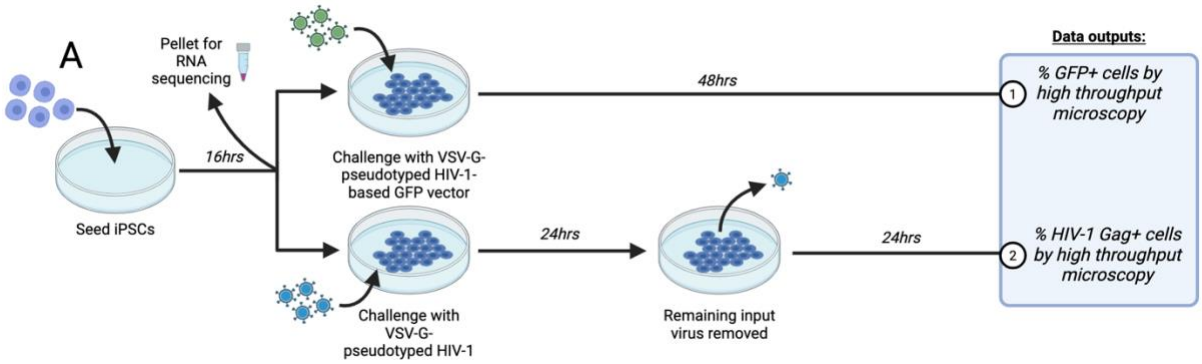


Figure 2.1: iPSC permissivity to HIV-1 infection screen pipeline.
A) iPSCs are seeded for 16 hours prior to infection by VSV-G pseudotyped HIV-1 based GFP vector and VSV-G pseudotyped full length HIV-1, strain NL4-3, and for RNA analysis. iPSCs infected with the GFP vector are incubated for 48 hours before being analysed for percentage of cells positive for GFP expression. iPSCs infected with the replication competent virus are incubated for 24 hours. The virus input is then removed and the iPSCs are incubated for a further 24 hours before being analysed for percentage of cells positive for HIV-1 Gag expression. Image generated by myself using BioRender.

2.5 Flow cytometry

2.5.1 Equipment and software

All flow cytometry-based assays were performed using a BD FACSCanto II and analysed using FlowJo (v10.9).

2.5.2 Assessment of pluripotency marker expression

High quality iPSC line cultures were dissociated to single cells using GCDR. 2.5 x 10^5 cells were pelleted and then fixed using a 4% PFA solution in PBS. The cells were then permeabilised using 100µl of 0.1% Triton-X for 15 minutes at room temperature. Following permeabilisation, the cells were pelleted and blocked using 100µl of 2% BSA in PBS for 30 minutes at room temperature. The cells were pelleted once more and resuspended in 100µl of 2% BSA in PBS containing a 1:100 dilution of either PE-conjugated anti-OCT4 antibody (BioLegend) or PE-conjugated anti-NANOG antibody (Invitrogen) and incubated for 30 minutes in the dark at room temperature. Cells were then washed 3 times in MagACS buffer, before being run through a flow cytometer. Percentage expression of the pluripotency marker was quantified by comparing iPSC samples to a stained HEK 293T sample and unstained iPSCs.

2.5.3 Quantification of intracellular HIV-1 Gag

The day after infection, activated CD4+ T-cells were pelleted and resuspended in 100µl TrypLE Express with phenol red and incubated for 10 minutes at 37°C. The cells were then fixed and permeabilised as in section **2.5.2** before being stained with 50µl of a 1:200 dilution of PE-conjugated anti-HIV-1 Gag antibody, clone KC57-RD1 (Beckman Coulter) in 2% BSA in PBS for 20 minutes in the dark at room temperature. Cells were then washed 3 times in MagACS buffer, before being run through a flow cytometer. Percentage of cells positive for HIV-1 Gag expression was quantified by comparison to a stained, uninfected sample of activated CD4+ T-cells.

2.5.4 PBMC phenotyping

Upon thawing, PBMC samples were phenotyped using the 8-colour Immunophenotyping Kit, anti-human REAfinity (Miltenyi Biotec) according to the manufacturer's protocol and assessed

by flow cytometry. The kit detected the expression the following cell surface markers: CD3, CD4, CD8, CD14, CD16, CD19, CD45, and CD56.

2.6 Microscopy

2.6.1 Equipment and software

Imaging of the trilineage differentiation assay was performed using a Leica TCS SP8 Confocal laser scanning microscope and analysed using ImageJ2 (v2.14.0/1.54f).

The samples produced by the iPSC permissivity to HIV-1 screens were imaged and analysed by the Operetta CLS high-content screen system using Harmony high-content analysis software (v4.8) to calculate percentage infections.

2.6.2 Trilineage differentiation assay

The MagACS buffer was removed from the cells and replaced with 100µl of 0.1% Triton-X and incubated for 20 minutes at room temperature. This was then replaced with 100µl of 2% BSA in PBS and incubated for 1 hour at room temperature. The supernatant was removed and the appropriate primary antibody was added to the cells at the following dilutions in 2% BSA in PBS and incubated overnight in the dark at 4°C. Rabbit anti-OCT-4 antibody, 1:500 (Abcam), mouse anti-PAX6 antibody (Abcam), 1:200, goat anti-Brachyury, 1:40 (Bio-Techne), and goat anti-SOX-17 antibody, 1:250 (Bio-Techne). The cells were then washed 3 times with 200µl PBS. For OCT-4 stained samples, 100µl of anti-rabbit Alexa Fluor 488-conjugated secondary antibody (ThermoFisher), diluted 1:500 in 2% BSA in PBS was then added. For the other primary antibodies, 100µl of the appropriate species Alexa Fluor 647-conjugated secondary antibody (ThermoFisher), diluted 1:500 in 2% BSA in PBS was added. DAPI (ThermoFisher) was also added to all samples at this point and they were incubated for 1 hour in the dark at room temperature. The cells were then washed another 3 times with 200µl PBS and stored in 200µl of 2% BSA in PBS for imaging. OCT-4 expression was judged in comparison to differentiated cells stained for OCT-4. Lineage specific marker expression was judged based on comparison to undifferentiated iPSCs stained for the same lineage-specific marker.

2.6.3 iPSC permissivity to HIV-1 infection data acquisition

The MagACS buffer from samples produced by both the infectivity and late gene expression assays was replaced with PBS supplemented with 1X Perm/Wash (BioLegend), and 10% FBS,

and incubated for 1 hour at room temperature. Following this incubation, samples from either assay were stained with the appropriate antibody mixes.

Samples from the infectivity assay were stained with 200µl of 2% BSA in PBS containing rabbit anti-GFP antibody , 1:100 (Invitrogen), 1X CellMask Deep Red Actin Tracking Stain (Invitrogen), and DAPI overnight in the dark at 4°C. Samples were then washed twice with 200µl PBS before addition of 200µl of 2% BSA in PBS containing anti-rabbit Alexa Fluor 488-conjugated secondary antibody, 1:500 (Invitrogen) and incubated for 30 minutes in the dark at room temperature. The cells were then washed 3 times with PBS and stored in 200µl of 2% BSA in PBS for imaging by the Operetta CLS using a 20X water lens.

Cells from the late gene expression assay were stained with 200µl of 2% BSA in PBS containing PE-conjugated anti-HIV-1 Gag antibody, clone KC57-RD1, 1:200, 1X CellMask Deep Red Actin Tracking Stain, and DAPI overnight in the dark at 4°C. The cells were then washed 3 times with PBS and stored in 200µl of 2% BSA in PBS for imaging by the Operetta CLS using a 20X water lens.

2.6.4 Operetta analysis pipelines

After imaging, percentage infection of the iPSC lines was calculated using an optimised multi-step algorithm, for each of the assays, embedded in the Harmony software. In short, this algorithm was optimised to accurately identify the nucleus, cytoplasm and the expression of either GFP or Gag of each imaged cell. Within a sample, the total number of cells and the number of either GFP or Gag positive cells was then used to calculate the infection percentage of that sample. The full detailed version of the Operetta algorithms used for the infectivity assay and the late gene expression assay are attached to this thesis (**Supplementary file 2.1 and 2.2 respectively**).

2.7 Nucleic acid extraction and manipulation

2.7.1 gDNA extraction

To obtain gDNA for use in quantitative PCR (qPCR), the desired cells were harvested, pelleted and stored at -80°C until extraction. Extraction was conducted using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. Extracted gDNA was resuspended in 100µl of nuclease-free water and stored at -20°C.

2.7.2 RNA extraction

To obtain RNA for use in qPCR or RNA sequencing applications, the desired cells were harvested, pelleted and stored at -80°C until extraction. Extraction was conducted using Monarch Total RNA Miniprep Kit (New England Biolabs) following the manufacturer's instructions, including an optional on-column DNase treatment step to eliminate contaminating genomic DNA. Extracted RNA was resuspended in 100µl of nuclease-free water, aliquoted and stored at -80°C.

2.7.3 cDNA synthesis

1µg of extracted total RNA was converted to cDNA prior to use in qPCR using the High-Capacity cDNA Reverse Transcription Kit (ThermoFisher) in a total volume of 100µl, following the manufacturer's instructions. The cDNA produced was stored at -20°C.

2.8 qPCR

2.8.1 Detection of HIV-1 provirus

An HIV-1 provirus detection qPCR master mix was prepared, consisting of 5µl TaqMan Universal PCR Master Mix (ThermoFisher), 1µl HIV-1 provirus detection forward primer (10µM), 1µl HIV-1 provirus detection reverse primer (10µM), and 1µl HIV-1 provirus detection probe (2µM) per sample. See **table 2.1** for primer and probe sequences. As a loading control, an RNaseP detection qPCR master mix was also prepared consisting of 5µl TaqMan Universal PCR Master Mix, 0.5µl TaqMan RNase P Detection Reagent (ThermoFisher), and 2.5µl nuclease-free water. The 8µl per sample of both master mixes prepared were distributed in separate wells of a MicroAmp Optical 384-well reaction plate (ThermoFisher). gDNA samples to be tested were diluted accordingly prior to use and 2µl was added to each reaction. All samples were tested in triplicate and the mean CT value was taken. Cycling temperatures used were as recommended by the TaqMan Universal PCR Master Mix User Guide for 40 cycles.

2.8.2 Detection of Sendai reprogramming vectors

A separate qPCR master mix was prepared for detection of collective Sendai Virus (SeV) and each of the 3 individual Sendai reprogramming vectors, KOS, KLF4, and c-Myc. 8µl of master mix consisting of 5µl TaqMan Universal PCR Master Mix, 0.5µl of the TaqMan assay specific to either SeV, KOS, KLF4 or c-Myc (ThermoFisher), and 2.5µl of water was prepared for each sample. As a loading control, a GAPDH detection qPCR master mix was also prepared, consisting of 5µl TaqMan Universal PCR Master Mix, 0.5µl TaqMan assay GAPDH (ThermoFisher), and 2.5µl nuclease-free water. The 8µl per sample of all master mixes prepared were distributed in separate wells of a MicroAmp Optical 384-well reaction plate. 2µl of the cDNA samples to be tested were added directly to each reaction. All samples were tested in triplicate and the mean CT value was taken. Cycling temperatures used were as recommended by the TaqMan Universal PCR Master Mix User Guide for 40 cycles and all primer/probe combinations utilised a FAM probe with a minor groove binder non-fluorescent quencher system.

Table 2.1: Primers and probe used for the detection of HIV-1 provirus by qPCR.

Primer name	Primer sequence (5'-3')
HIV-1 provirus detection forward	CAGATGCTGCATATAAGCAGCTG
HIV-1 provirus detection reverse	GAGGGATCTCTAGNYACCAGAGT
HIV-1 provirus detection probe	FAM - CCTGTACTGGGTCTCTCTGG - TAMRA

2.9 iPSC permissivity to HIV-1 data analysis

2.9.1 Linear regression model and normalisation to the CTRs

For the infectivity assay, samples with infection percentage values between 2-60% were used in the calculation of vector titre. The calculation utilised was:
Vector titre (Infectious units (IFU)/µl) = Number of cells * ((Percentage infection/100) / Vector volume))

For the late gene expression assay, samples with infection percentage values between 1-20% were used in the calculation of vector titre. The calculation utilised was:
Virus titre (IFU/µl) = Number of cells * ((Percentage infection/100) / Virus volume))

For the analysis of both assays conducted by the linear regression model, variation between screens was accounted for by normalisation to a well characterised iPSC line, CTRs. This normalisation was performed by dividing the vector/virus titre of a MACS iPSC line by the vector/virus titre of the CTRs conducted within the same screen.

2.9.2 Logistic curve fitting analysis

Analysis of the infectivity assay data by a logistic curve fitting model was performed in Python using the curve_fit function from the scipy.optimize (v1.11.3) module, where x equals vector volume and y equals percentage of GFP-positive cells. The logistic model was fitted to the data with initial parameter (p0) guesses using a = 100, b = 1, c = 10, which provided starting points for the optimisation process, and the fitting process allowed for a maximum of 10,000 iterations. The log function utilised was as below, and the b-value was taken as the measurement to assess permissivity to HIV-1 infection.

$f(x) = a/(1+e^{-b(x-c)})$

Variation across screens was accounted for by normalisation using z-score, as described in section **2.9.3**. This data analysis was conducted with assistance from Nathalia Almeida dos Santos.

2.9.3 Z-score normalisation

Z-score normalisation was completed using scikit-learn library's StandardScaler module (v1.3.2). This standardises the data by first subtracting the mean of the parameter of interest from a measured value of the parameter of interest, and then scaling to parameter of interest's variance by dividing by the standard deviation. This method of normalisation was used to account for variation between screens when indicated and in the virus volume normalised analysis model used for the late gene expression assay data. This data analysis was conducted with assistance from Nathalia Almeida dos Santos.

2.10 RNA sequencing

2.10.1 Sequencing

Extracted total RNA samples from iPSCs collected during the permissivity to HIV-1 infection pipeline were outsourced for ribosomal RNA (rRNA) depletion, library preparation and sequencing to Azenta. Depletion of rRNA was conducted by polyA selection for mRNA species (eukaryotic). Libraries were sequenced paired-end (2x150bp) on the Illumina NovaSeq sequencing platform with a target sequencing depth of 20-30 million reads per library. Raw data was received in FASTQ format.

2.10.2 PluriTest analysis

Raw FASTQ RNA sequencing (RNA-seq) data files are run through a standardized data processing and analysis pipeline made available at https://www.pluritest.org/. The files are pre-processed and projected against a model of pluripotency based upon an RNA-seq dataset of hundreds of validated pluripotent and somatic cell lines. This projection allows for the examination of the pluripotency signature of a cell line (pluripotency score) and the uniquity of its signature compared to the database (novelty score).

2.10.3 Processing of RNA-seq data

Processing of the RNA-seq data was achieved via a custom bioinformatics pipeline designed and performed by Dr Robert Carton.

Raw data quality control was rendered by Validatefastq (v0.1.1), Trimmomatic (v0.39) and FastQC (v0.11.9). Alignments to the hg38 reference genome were achieved with Burrows-Wheeler Aligner (v0.7.17) and post-alignment processing was conducted via Samtools (v1.13). Counts were generated from the alignments using the FeatureCounts function of the Rsubread package (v2.8.2). Counts were then annotated using the biomaRt package (v3.18) facilitated by the Ensembl database. Counts were normalised using the transcripts per million (TPM) method using R. This TPM normalisation process involved dividing raw counts of each gene by the length of that gene to give reads per kilobase (RPK). The RPKs of that

sample were totalled and divided by a scaling factor of 1 x 10^6, resulting in TPM values normalised for transcript length and library size.

All samples were analysed using this custom bioinformatics pipeline on King's College London's CREATE high-performance computer. King's College London. (2022). King's Computational Research, Engineering and Technology Environment (CREATE). Retrieved March 2, 2022, from https://doi.org/10.18742/rnvf-m076.

2.10.4 Differentiational gene expression analysis

Differential gene expression analysis between MACS iPSC lines identified as exhibiting extreme opposite permissivity phenotypes to HIV-1 was achieved using ΔTPM and log_2-fold change in expression approaches. Ensembl gene IDs were first annotated using BiomaRt (Durinck et al., 2009). The RNA-seq data was then filtered for protein coding genes. Samples were compared by building a correlation matrix between the TPM-normalised libraries, performed via the cor function in R to confirm sample correlation. Euclidean distances between sequenced libraries were analysed by hierarchical clustering. Hierarchical clustering analysis was performed using BAM counts with the DESeq2 package (v1.40.2) (Love et al., 2014). The TPM count matrix was normalised using gene-wise z-score normalisation (z-score described in section **2.9.3**) to scale gene expression across all genes, ensuring highly expressed genes would not affect the analysis. The absolute difference in gene expression counts between two iPSC lines, ΔTPM, was then calculated, and genes were listed from largest to smallest. TPM counts were also converted into log_2 and fold change in expression was calculated between the two iPSC lines of interest. A threshold of >1.5log_2-fold change was set for a gene to be considered differentially expressed. Differential gene expression analysis was conducted with assistance from Nathalia Almeida dos Santos.

2.11 Statistical tests

Statistical tests used in data analysis are indicated in the figure legends. Statistical tests employed were: unpaired t test with Welch's correction and Mann-Whitney U test. Asterix (*) represent statistical significance: * = $p<0.05$, *** = $p<0.0005$, **** = $p<0.0001$.

Chapter 3: Establishment of a reprogramming pipeline for generating iPSCs from HIV-1 infected PBMCs

3.1 Introduction

Since their discovery in 2006, iPSCs have become a revolutionary tool used in modern scientific research (Takahashi & Yamanaka, 2006). Their applications span a wide variety of areas such as disease modelling, drug screening, and the study of infectious diseases, including HIV-1 (Shi et al., 2017). In the field of HIV-1, studies have utilised iPSCs for clinical aspirations. For example, the generation of iMacs that are CCR5 deficient, and thus, are resistant to HIV-1, or the formation of an iPSC-derived 3D mini-brain model to study HIV-1 infection in the CNS (Iwamoto et al., 2021; Kambal et al., 2011; Z. Wei et al., 2023; Ye et al., 2014). However, little has been reported about the use of iPSCs as a tool to discover more about HIV-1 and its interactions with its host on a molecular level.

iPSCs have been shown to intrinsically express ISGs and be highly resistant to viral infection (Wu et al., 2018). Thus, if used as a screening tool, could uncover novel cellular proteins, pathways or networks that interact with HIV-1 to either aid or restrict its replication. This chapter describes the establishment and optimisation of the reprogramming pipeline to be used for generating iPSCs from PBMCs collected in the MACS. The resultant iPSCs are to be employed in a permissivity to HIV-1 screen, although, being generated from a clinically interesting cohort of patients, these iPSCs could prove to be a valuable scientific resource for other applications in the future.

Due to the finite nature of the MACS PBMC samples, it was important to ensure an established and reproducible reprogramming protocol was in place. As a result, numerous preliminary experiments were carried out utilising different reprogramming protocols, whilst also altering conditions in an attempt to mimic working with the MACS samples. Completing these optimisations significantly increased the likelihood of success when generating iPSCs from these precious samples.

After three practice reprogramming experiments, it was then learned that 14 of the 18 MACS PBMC samples were collected from patients at a time when they were already infected with HIV-1. This produced two extra criteria required for consideration when finalising the

reprogramming pipeline. Firstly, all subsequent work was to be completed in a CL3. Secondly, all of the generated iPSC lines needed to be free of HIV-1 provirus, as its presence could potentially impact findings when conducting the permissivity to HIV-1 phenotyping screen. HIV-1 provirus-free iPSCs have been generated from an HIV-1 infected patient's PBMCs previously, however, the patients in this study had been receiving HAART at the point of PBMC collection (Ye et al., 2020). Here, we faced the additional challenge of the patients not receiving any ART at the PBMC collection time point, therefore, a higher prevalence of HIV-1 infected cells was expected.

3.2 Results

3.2.1 Comparison of feeder-dependent and feeder-free reprogramming protocols for generating iPSCs from PBMCs

The optimisation of generating iPSCs from PBMCs began by trialling two reprogramming protocols: feeder-dependent and feeder-free. The two protocols shared a similar workflow: an initial expansion of a specific cell population, followed by transduction with three Sendai viral vectors encoding for the four 'Yamanaka factors' and finished with an expansion culture of the reprogrammed cells. Individual iPSC colonies, that had grown from a single successfully reprogrammed cell, were then picked and expanded separately as clonal iPSC lines. There were two main differences between the protocols. The first dissimilarity was the additional five days of cell expansion in the feeder-dependent protocol. The second, and probably more significant, difference was the cell culture matrix the reprogrammed cells were maintained on from 3 days post-transduction. The feeder-dependent protocol utilised MEFs as its matrix compared to a vitronectin matrix used in the feeder-free protocol. Both protocols were carried out simultaneously in two distinct run throughs, with each run using PBMCs from a different donor (donor numbers = 1 and 2). In the first reprogramming experiment on PBMCs from donor 1, only the feeder-free protocol was successful in generating iPSCs. For the feeder-dependent protocol, viable cells were present for the transduction with the Sendai reprogramming vectors, however, no iPSC colonies formed on the MEF layer. 22 feeder-free generated colonies were picked and expanded as individual clones. Roughly one third of these clones, n = 8, endured the picking process and avoided spontaneous differentiation. After successful expansion of the clones, it was necessary to conduct quality control checks on each clone to validate the cells as iPSCs. **Figure 3.1** shows the iPSC quality control results from this first reprogramming experiment, using one clonal line as a representative. **Figure 3.1A** demonstrates that the representative iPSC line exhibits the classical morphological characteristics expected in an iPSC line of high quality: cells are compacted and individual cells are difficult to distinguish; the colonies are rounded with smooth edges, and there is minimal, if any, spontaneous differentiation present. The generated iPSCs were also tested for their expression of the pluripotency marker OCT-4 by flow cytometry. **Figure 3.1C** shows that 88.3% of the cells in the representative iPSC line were positive for OCT-4; granting further validation

that the generated cells are iPSCs. It is reported that pluripotency markers, such as OCT-4, should be present in >70% of the cells in the culture to be deemed an iPSC line, therefore, this cell line exceeds this threshold (Sullivan et al., 2018). All 8 surviving iPSC clonal lines generated here exhibited the typical iPSC morphological characteristics and were shown to contain a percentage of OCT-4 expressing cells beyond the above mentioned threshold.

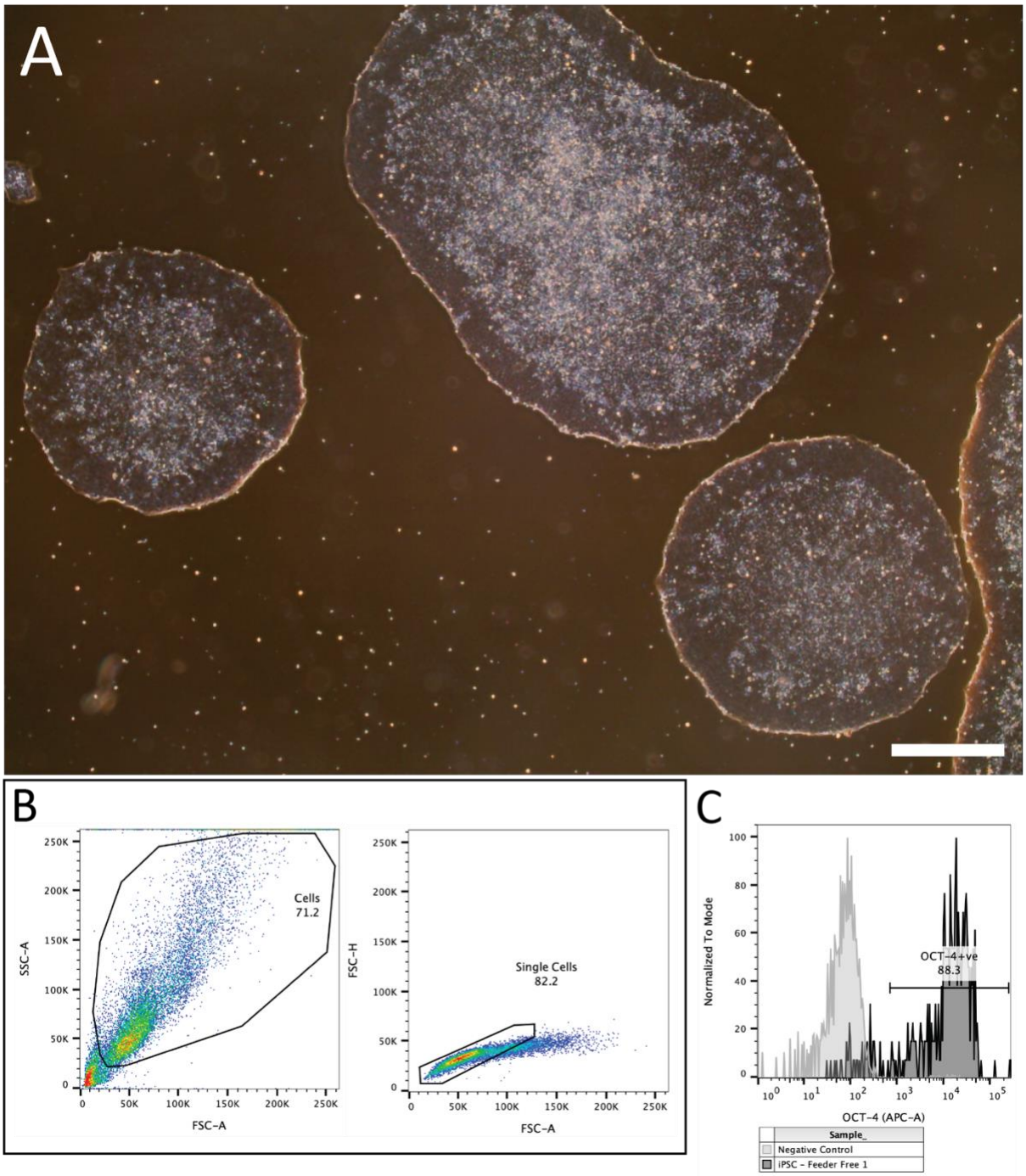


Figure 3.1: iPSC quality control checks on a representative iPSC clonal line reprogrammed from PBMCs using the feeder-free protocol.

A) Light microscopy image showing an example iPSC line generated in the first practice reprogramming experiment. Image taken at 4X magnification and the scale bar represents 500μm. **B)** Example of the gating strategy used when analysing iPSCs by flow cytometry for pluripotency marker expression. **C)** OCT-4 expression in the representative iPSC line. The light grey and dark grey peaks represent a stained negative control (HEK293Ts) and the stained iPSC line respectively.

As a result of the feeder-dependent protocol failing to produce iPSCs in the first attempt, a slight modification was made to the protocol. During the expansion culture of the reprogrammed cells, the MEFs were replenished midway through to ensure a high enough confluency to support iPSC growth and maintenance. With the addition of this adjustment, both protocols were carried out again for the second reprogramming experiment on PBMCs from donor 2. This time, both the feeder-dependent and feeder-free protocols successfully generated iPSC colonies and their growth post-transduction was imaged (**Figure 3.2**). **Figures 3.2D+H** show that the iPSC colonies generated from both protocols exhibited the morphological features expected of a high quality iPSC culture. From the feeder-dependent protocol, 6 colonies were picked and grown as individual clones on fresh MEF plates, however, none were successfully expanded as clonal lines. All 6 clones suffered large amounts of spontaneous differentiation and, as a result, were not maintained. A further 6 feeder-dependent colonies were picked and seeded on vitronectin-coated plates, of which 2 survived and avoided spontaneous differentiation. This demonstrated a clear superiority, in my hands, for the use of vitronectin over MEFs as a matrix for culturing iPSCs. From the feeder-free protocol, 18 iPSC colonies were picked. 8 of these colonies successfully expanded to healthy iPSC lines, exhibiting the expected morphological characteristics and containing a high percentage of cells expressing the pluripotency marker OCT-4 (as demonstrated in section **3.2.1**). As a result of the two reprogramming runs discussed above, the feeder-free variation of the protocol was chosen as the more reliable and reproducible method to generate iPSCs from PBMCs and was used in all subsequent reprogramming experiments.

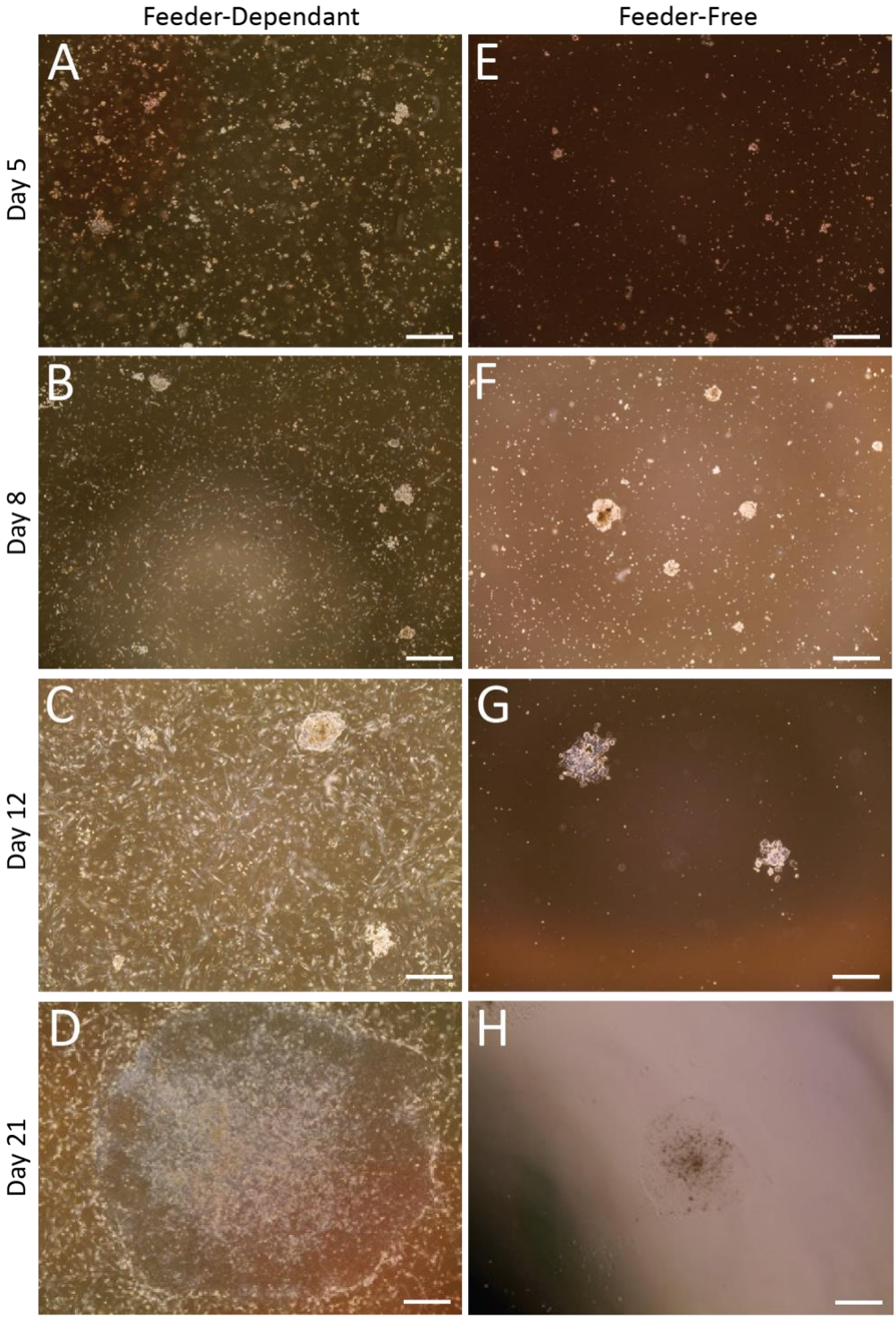


Figure 3.2: Growth overtime of iPSC colonies post Sendai viral vector transduction during the 2nd reprogramming experiment.
PBMCs were reprogrammed to iPSCs using the two different protocols in parallel. **A-D+E-H)** Growth of iPSCs after Sendai viral vector transduction for the feeder-dependent protocol and feeder-free protocol respectively. **A+E)** 5 days post-transduction. **B+F)** 8 days post-transduction. **C+G)** 12 days post-transduction. **D+H)** 21 days post-transduction. Images were taken at 4X magnification and scale bars represent 500μm.

3.2.2 Evaluating the use of fresh and cryopreserved PBMCs for reprogramming

During the first reprogramming experiment, both protocols were conducted on fresh and cryopreserved PBMCs from donor 1 in parallel. Despite the feeder-dependent protocol not working for either of the PBMC conditions, there was no noticeable difference observed in the number of iPSC colonies produced from fresh or cryopreserved PBMCs by the feeder-free protocol; with both conditions producing plenty of colonies. 11 colonies were picked from each condition, of which 4 from each were successfully grown up as individual clones all exhibiting excellent iPSC morphological characteristics. The cell lines generated here were also analysed for their OCT-4 expression by flow cytometry. **Figure 3.3** shows the OCT-4 expression of two representative clones of iPSC lines, one generated from fresh PBMCs (**Figure 3.3A**) and one produced from cryopreserved PBMCs (**Figure 3.3B**). For both cell lines, >95% of the cells were positive for OCT-4 expression, demonstrating that whether the iPSCs were generated from fresh or cryopreserved PBMCs had no effect on the percentage of cells in the iPSC line expressing OCT-4. All remaining 6 cell lines, 3 from fresh PBMCs and 3 from cryopreserved PBMCs, showed comparable percentages of OCT-4 expression. As a results, all subsequent reprogramming runs were conducted using only cryopreserved PBMCs, as this was how the PBMC samples from the MACS would be received.

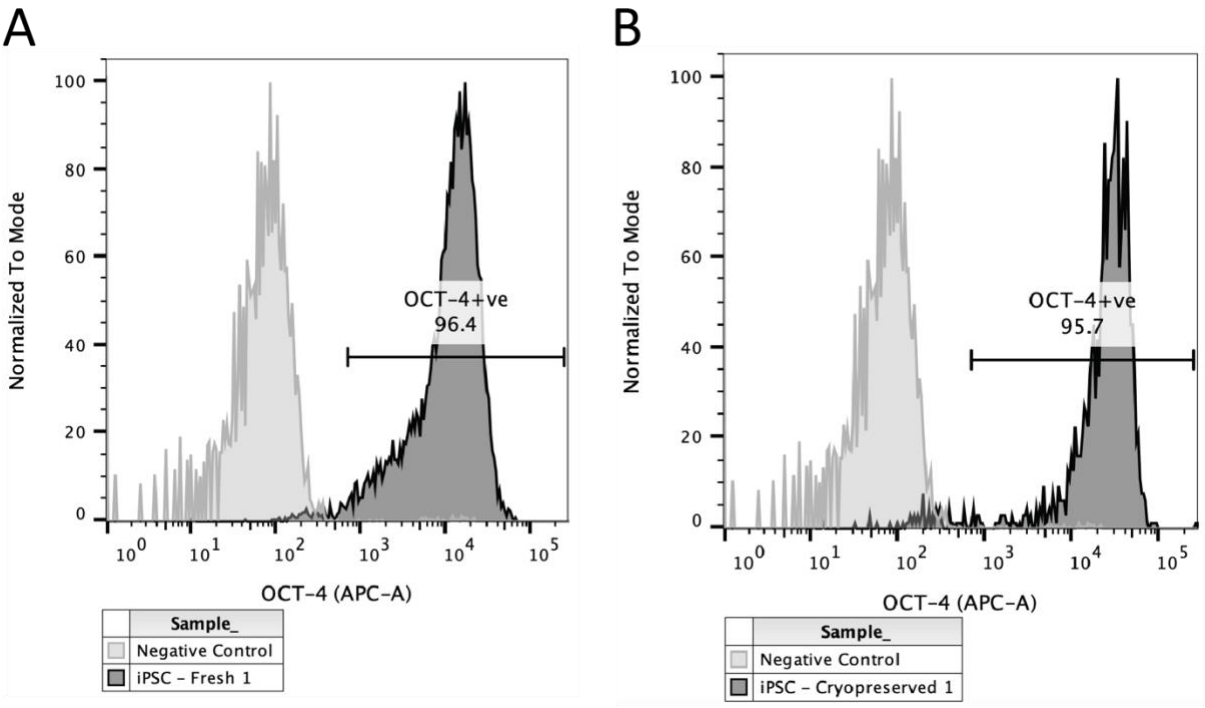


Figure 3.3: Representative iPSC lines generated from either fresh or cryopreserved PBMCs analysed for their OCT-4 expression.

iPSC lines generated in the first reprogramming run were stained for OCT-4 expression and analysed by flow cytometry. **A+B)** OCT-4 staining for iPSCs (dark grey) reprogrammed from fresh and cryopreserved PBMCs respectively compared to a stained negative control cell line, HEK293Ts (light grey).

3.2.3 Trialling of the reprogramming protocols with 10-fold fewer cell numbers

Due to the extended period of time that the MACS samples had been cryopreserved for, preparations were made for a potentially poor viable cell recovery. In an attempt to simulate this, both the feeder-dependent and the feeder-free protocols were performed scaled down 10-fold in cell number in parallel to the standard second reprogramming practice run. Using just 5 x 10^5 viable PBMCs as starting material for the expansion period, followed by just 3 x 10^4 viable cells at transduction, iPSC colonies were still generated by both protocols, albeit ~10-fold fewer. All iPSC clonal lines generated here were confirmed to be iPSCs via morphological characteristics and OCT-4 expression as described in section **3.2.1**.

3.2.4 Evaluation of using freeze/thawed Sendai viral vectors for reprogramming

Also in parallel to the second reprogramming experiment, transduction was performed using remaining Sendai viral vectors from the first reprogramming experiment. To account for a potential loss in viral titre during the extra freeze/thaw process, double the volume of each virus was used on the day of transduction. Apart from this, the two reprogramming protocols were followed as standard. No iPSCs were generated, from either protocol, from the cells transduced with the recycled virus. This demonstrated that a new batch of the CytoTune-iPS 2.0 Sendai Reprogramming Kit was required for each new reprogramming experiment.

3.2.5 Validation of the optimised reprogramming protocol

To further the confidence in the optimised reprogramming protocol, a third reprogramming experiment was conducted using the feeder-free protocol and cryopreserved PBMCs. This time, 3 new donors (donor numbers = 3, 4 and 5) were reprogrammed simultaneously to validate scaling up of the protocol and to also account for any donor variability. Plenty of iPSC colonies were generated for all three donors and 10 colonies from each donor were picked to be expanded as clones. Of the 10 picked colonies, 6, 6, and 5 colonies avoided spontaneous differentiation and were successfully expanded to clonal iPSC lines from donor 3, 4 and 5 respectively. All iPSC clonal lines generated here were confirmed to be iPSCs via their morphological characteristics and OCT-4 expression (as described in section **3.2.1**). This successful generation of iPSC lines from multiple donors simultaneously further validates the

optimised reprogramming protocol for use on the MACS PBMC samples. The reprogramming experiments conducted so far are summarised in **Table 3.1**.

Table 3.1: Summary of the three preliminary reprogramming experiments performed to optimise the reprogramming of PBMCs to iPSCs.

Reprogramming experiment	Protocol	Donor	Protocol Modifications	iPSC Colonies Generated?	Expansion Modifications	iPSC clones successfully expanded?
1	Feeder-dependent	1	Fresh PBMCs	×	N/A	×
			Cryopreserved PBMCs	×	N/A	×
	Feeder-free		Fresh PBMCs	✓	None	✓
			Cryopreserved PBMCs	✓	None	✓
2	Feeder-dependent	2	N/A	✓	None	×
				✓	Transferred to vitronectin	✓
			10-fold fewer cell number	✓	None	×
			Recycled Sendai Vector	×	N/A	×
	Feeder-free		N/A	✓	None	✓
			10-fold fewer cell number	✓	None	✓
			Recycled Sendai Vector	×	N/A	×
3	Feeder-free	3	None	✓	None	✓
		4	None	✓	None	✓
		5	None	✓	None	✓

3.2.6 Optimisation of CD4+ cell depletion from a PBMC population using MagACS

It was learned at this point that the majority of the MACS PBMC samples (14 out of 18) to be received were from donors that were HIV-1+ at the time of sample donation. Because iPSCs already containing HIV-1 provirus may respond differently to incoming virus when exposed during the permissivity screen (to be discussed in **Chapter 5**), it was important to minimise the possibility of generating HIV-1+ iPSC lines. To do this, depleting any cells expressing the CD4 receptor from the initial PBMC sample, as HIV-1 can only infect cells expressing this receptor, would, in theory, eliminate any HIV-1 infected cells. With the remaining CD4- cell population, the reprogramming process would then be continued as previously optimised.

To deplete the CD4+ cell population MagACS was used. A trial depletion was conducted to test the efficiency of removing the CD4+ cells from a non-HIV-1 infected PBMC population using MagACS (**Figure 3.4**). As shown in **figure 3.4C**, roughly half of the starting PBMCs from this donor (donor number = 6) expressed the CD4 receptor. The vast majority of this CD4+ population consisted of either CD4+ T-cells, characterised by $CD4+_{HI}$, or monocytes, characterised by $CD4+_{LO}$ and SSC_{HI}. **Figure 3.4E** shows there is still a fair proportion of CD4+ cells, 32.2%, still remaining in the CD4- fraction after depletion. The flow data indicates that the CD4+ T-cells are efficiently removed, however, a large proportion of the monocyte population still remains. **Figure 3.4F** is the CD4+ fraction of cells after depletion and confirms the above by presenting a large density of CD4+ T-cells and a smaller population of monocytes.

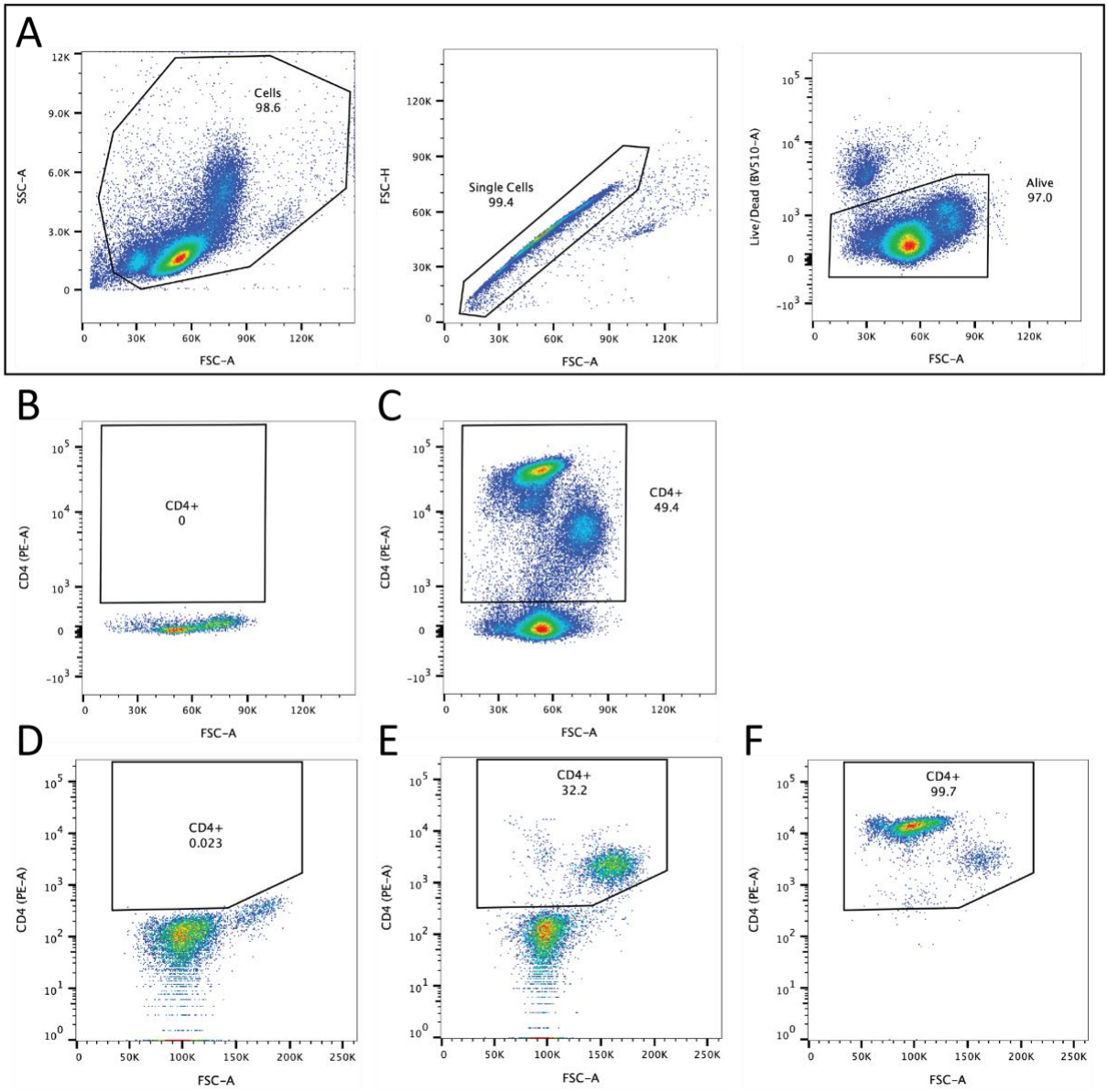


Figure 3.4: CD4+ cell depletion of PBMCs using MagACS.
Flow cytometry plots to show PBMCs before and after the depletion of CD4+ cells using MagACS. **A)** Gating strategy taken when analysing PBMCs by flow cytometry. **B+C)** Pre-depletion PBMC population unstained and stained for CD4 expression respectively. **D+E)** CD4- fraction of cells after the CD4+ cell depletion unstained and stained for CD4 respectively. **F)** CD4+ fraction of cells after the CD4+ cell depletion stained for CD4.

In an attempt to increase the efficiency of the removal of CD4+ cells, a second trial was conducted. This time, as well as the standard CD4+ cell depletion using MagACS, a sample was run through an additional round of CD4+ depletion by MagACS (**Figure 3.5**). Again, as shown in **figure 3.5B**, roughly half of the cells in the starting PBMC sample from this donor (donor number = 7) expressed the CD4 receptor, however, this donor had a higher proportion of CD4+ T-cells to monocytes compared to donor 6 (**figure 3.4C**). **Figure 3.5C** displays the CD4- fraction of cells after one round of CD4+ cell depletion and demonstrates there is a significant decrease in the proportion of CD4+ cells remaining, from 51.8% to 9.20%. However, it demonstrates once again, that whilst MagACS is successful at removing the CD4+ T-cell population, it is less effective at removing the monocyte population. This is confirmed by looking at the CD4+ fraction of cells after the first round of depletion in **figure 3.5D**, which demonstrates a large amount of CD4+ T-cells and a smaller number of monocytes. **Figure 3.5E** shows the CD4- fraction after the two rounds of CD4+ cell depletion and demonstrates a further decrease in the proportion of CD4+ cells remaining in the sample from 9.20% to 5.91%. **Figure 3.5F** displays the CD4+ fraction of cells after the second round of depletion and demonstrates that the vast majority of the cell depleted here were monocytes, as well as a smaller proportion of CD4+ T-cells. This suggests that while the first round of depletion is effective at depleting the CD4+ T-cells and some monocytes, using a second round of depletion allows for a higher monocyte removal efficiency as well as removing some remaining CD4+ T-cells. **Figure 3.5G** shows the CD4+ and CD4- cell numbers present before and after the CD4+ cell depletions. After one round of depletion there is a large decrease in the amount of CD4+ cells present in the sample, from 2.59 x 10^6 to 2.40 x10^5, while maintaining a stable number of CD4- cells. After the second round of depletion, the CD4+ cell population was reduced by a further ~1 x 10^5 cells, at the sacrifice of very few CD4- cells. Therefore, in the interest of minimising the probability of generating iPSCs from HIV-1 infected cells, it was decide that all subsequent reprogramming experiments would be preceded with a double CD4+ cell depletion using MagACS.

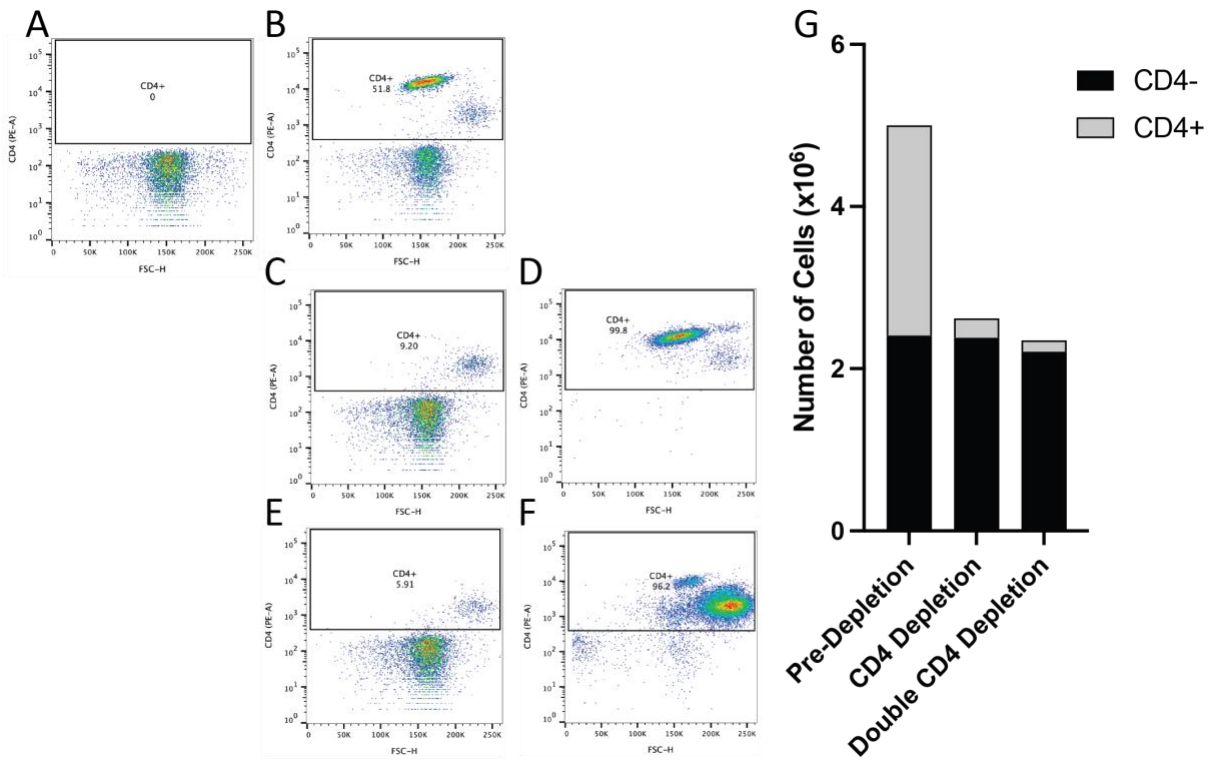


Figure 3.5: Optimisation of CD4+ cell depletion using MagACS.
Flow cytometry plots and a bar graph to show the efficiency of removing CD4+ cells from a PBMC sample using one or two rounds MagACS. **A+B)** PBMC sample from donor 7 before depletion either unstained or stained for CD4 respectively. **C+D)** CD4- and CD4+ cell fraction stained for CD4 after one round of CD4+ cell depletion respectively. **E+F)** CD4- and CD4+ cell fraction stained for CD4 after a second round of CD4 depletion respectively. **G)** Number of cells present before depletion, after one round of depletion and after two rounds of depletion. CD4+ cells are represented by the grey part of the bars and CD4- cells are represented by the black part of the bars.

3.2.7 Generation of HIV-1 provirus-free iPSCs from PBMCs containing HIV-1 infected cells

A final reprogramming optimisation experiment was completed to test whether the iPSC clonal lines generated from a PBMC population containing HIV-1 infected cells that had been double depleted of CD4+ cells using MagACS would be free of HIV-1 provirus. Firstly, to detect HIV-1 provirus, a qPCR was optimised using a primer-probe combination that targeted a highly conserved region in the 5' LTR of the HIV-1 provirus (**Figure 3.6**). Using two cell lines, ACH2 and U1, both known to contain 1 integrated copy of HIV-1 per cell, as positive controls, the amount of input gDNA used for the qPCR was optimised by titration. **Figure 3.6A** shows that as the amount of input gDNA is decreased, the detection of HIV-1 provirus also decreases (demonstrated by an increasing mean Ct value), whilst remaining undetected at all gDNA amounts for the negative control cell line, HEK293Ts. As a result, the highest amount, 50ng, of gDNA was chosen to be added to all future HIV-1 provirus detection qPCR experiments.

When testing samples for the presence or absence of the HIV-1 provirus, it is important to run a parallel qPCR detecting an endogenous gene as a loading control. As shown in **figure 3.5B**, RNaseP was tested for use as the endogenous control in this assay. The presence of RNaseP was detected in both the HIV-1 positive control and HIV-1 negative control while remaining negative when H_20 was added instead of gDNA, making it an ideal loading control for this HIV-1 provirus detection assay. Going forward, all HIV-1 provirus detection qPCR experiments were conducted using 50ng of input gDNA and looked for the presence of both HIV-1 and RNaseP.

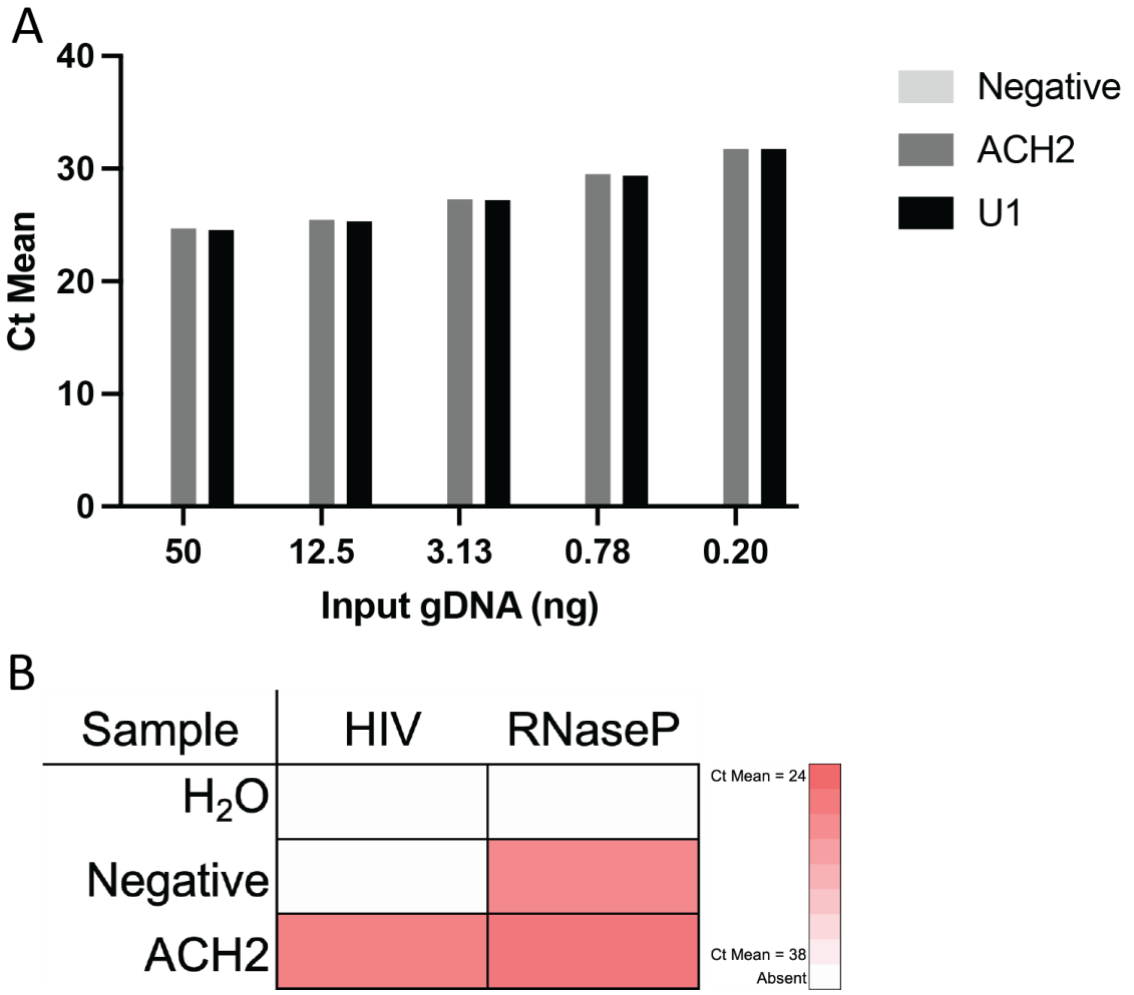


Figure 3.6: Optimisation of the detection of HIV-1 provirus by qPCR.
qPCR results to show the detection of HIV-1 provirus or RNaseP present in the gDNA of cell lines. **A)** Bar graph to show HIV-1 provirus detection when using of differing amounts of input gDNA. **B)** Heat map to demonstrate HIV-1 provirus detection when conducting a parallel qPCR to detect an endogenous gene, RNaseP, used as a loading control for when samples are identified as negative for HIV-1. As indicated in the heat map scale bar, the darker the red box equates to a lower mean Ct value and, therefore, a higher quantity of the PCR product detected in the sample. White boxes indicate an undetected Ct mean value and, therefore, an absence of the PCR product. All results shown are a Ct mean result of three Ct values.

To mimic working with an infected population of PBMCs, for the final reprogramming optimisation experiment CD4+ T-cells were isolated from a population of PBMCs of a new donor (donor number = 8), activated for 48 hours and then infected with the NL4-3 strain of HIV-1. The infection was quantified and infected CD4+ T-cells were added at differing ratios (1%, 0.1% and 0.01%) to uninfected PBMCs also from donor 8. This now HIV+ PBMC sample was then double depleted for CD4+ cells using MagACS and then continued through the previously optimised reprogramming protocol to generate iPSCs. The presence of HIV-1, either by intracellular staining for Gag or detection of provirus by qPCR, was analysed at different points throughout the process (**Figure 3.7**). Confirmation of the CD4+ T-cell infection is shown in **figure 3.7A** by intracellular staining for Gag. 97.7% of the infected CD4+ T-cells were HIV-1+ and this value was used to add 50,000, 5000, and 500 infected CD4+ T-cells to 3 separate pools of 5 x 10^6 uninfected PBMCs, resulting in 3 different PBMC populations which were made up of 1%, 0.1% and 0.01% HIV-1 infected cells respectively. All three populations were then carried forward for double CD4+ cell depletion and iPSC generation. **Figure 3.7B** presents detection of HIV-1 via qPCR before the CD4+ cell depletion, after the depletion and after expansion of the clonal iPSC lines. It demonstrates, as expected, the amount of HIV-1 provirus present pre-CD4+ cell depletion decreased as fewer HIV-1 infected cells were added to the uninfected cell population. However, the results for post-depletion did not show the expected decrease in HIV-1 provirus for the CD4- cell fraction. Instead, after the double CD4+ cell depletion, both the CD4+ cell fraction and the CD4- cell fraction of cells showed similar levels of HIV-1 provirus present across all 3 PBMC samples. This result may be explained by HIV-1's ability to downregulate surface expression of CD4 on infected cells, and therefore, many of the HIV-1 infected cells would not be removed by CD4+ cell depletion. Despite this suboptimal result, all of the resultant iPSC clones that were generated from all three samples, 3 clones from the 1% sample, 2 clones from the 0.1% sample and 3 clones from the 0.01% sample, were shown to be negative for HIV-1 provirus. Therefore, it was concluded that iPSC lines free from HIV-1 provirus could be successfully reprogrammed from a population of PBMCs containing HIV-1 infected cells that had been double depleted for CD4+ cells by MagACS.

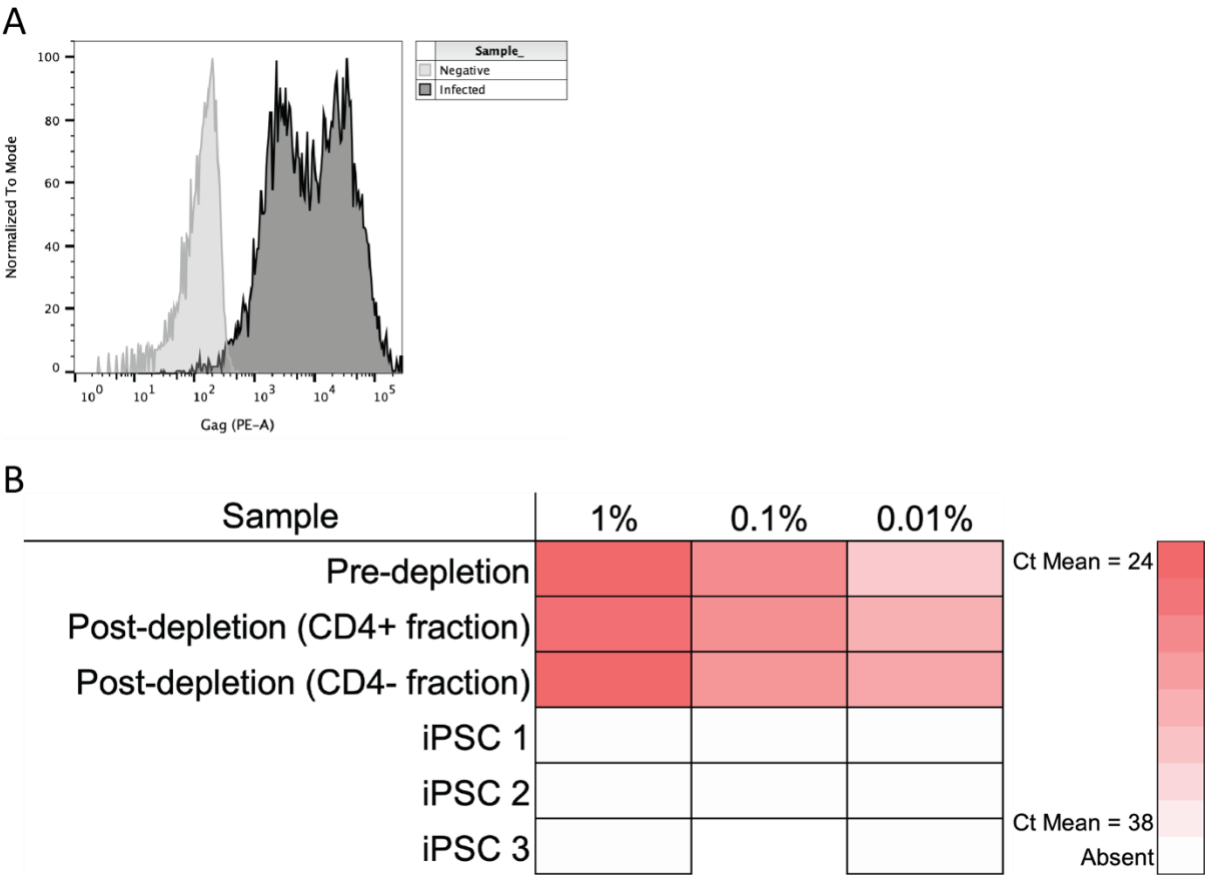


Figure 3.7: Generation of HIV-1 provirus-free iPSCs from a population of PBMCs containing HIV-1 infected cells.

Detection of HIV-1 by either intracellular HIV-1 Gag staining or provirus detection by qPCR at different points throughout the generation of iPSCs from a population of PBMCs containing HIV-1 infected cells. **A)** Uninfected and infected CD4+ T-cells stained for intracellular Gag expression respectively. **B)** qPCR results for detection of HIV-1 provirus at different timepoints throughout the reprogramming of PBMCs containing HIV-1 infected cells to iPSCs. As indicated in the heat map scale bars, the darker the red box equates to a lower mean Ct value and, therefore, a higher amount of the HIV-1 provirus present in the sample. White boxes indicate an undetected Ct mean value and, therefore, an absence of the HIV-1 provirus. All results shown are a Ct mean result of three Ct values.

3.3 Discussion

This chapter sought to establish a protocol for the derivation of iPSCs from PBMCs. The protocol was then further developed to be optimised for the generation of HIV-1 free iPSCs from PBMCs containing HIV-1 infected cells. It was important to optimise this process to ensure its efficiency and consistency in preparation for working on the precious MACS PBMC samples.

To begin with, two different protocols were tested for generating iPSCs from PBMCs, a feeder-dependent protocol and a feeder-free protocol. The protocols were largely similar but crucially differed in their matrix use; MEFs were used for the feeder-dependent protocol and vitronectin was used for the feeder-free protocol. Despite both protocols successfully generating iPSCs, the feeder-free protocol reigned superior in terms of consistency and ease of use. Also, after the generation of the initial iPSC colonies, the expansion of the iPSC clonal lines proved far more successful using vitronectin. Therefore, it was concluded that the feeder-free protocol would be used to generate iPSCs from PBMCs. This evaluation seems to correspond with the iPSC field, where there is a general shift from using feeder-dependent to feeder-free culture systems for various reasons. Firstly, the use of animal-derived feeder cells can exert unknown xenogeneic effects on the cultured cells while also compromising the purity and safety for iPSCs intended for clinical aspirations. Utilising human-derived feeder cells can evade these issues, however, the labour-intensive nature of feeder cell preparation limits their scalability. Furthermore, there is still a limited understanding regarding the precise mechanism by which the feeder cells apply their function. On the other hand, feeder-free culture systems offer a defined, qualified and non-animal-derived alternative supporting enhanced consistency and scalability, while reducing labour demands and costs.

Since the MACS samples were cryopreserved, it was necessary to demonstrate that iPSCs could be generated from cryopreserved PBMCs as well as fresh PBMCs. To do this, both fresh and cryopreserved PBMCs from the same donor were run through the two reprogramming protocols to assess if there was any variability in the success of iPSC generation. No obvious

variation was found, and thus, it was concluded that iPSCs could be generated from either fresh or cryopreserved PBMCs.

Reprogramming experiments using either 10-fold fewer cell numbers or recycled Sendai viral vectors were also conducted. Using 10-fold fewer cell numbers showed that this protocol could still be performed confidently to generate iPSCs if a low viable cell number was revived from a MACS PBMC sample. Conversely, using recycled Sendai viral vectors did not yield any iPSCs. As a result, it was concluded that a new CytoTune 2.0 Sendai Reprogramming Kit was required for each reprogramming experiment.

The optimisation of the reprogramming protocol thus far was validated by conducting another reprogramming experiment on 3 previously untested donors simultaneously. The successful generation of iPSCs from all 3 donors demonstrated the scale up feasibility of the protocol and also minimised any concerns surrounding donor variability.

Due to the majority of the MACS samples being collect from HIV-1+ donors, optimisation of generating HIV-1 provirus-free iPSCs from a PBMC population infected with HIV-1 was required. To decrease the chances of generating iPSC clonal lines that were reprogrammed from HIV-1+ cells, the PBMC sample was depleted of any cells expressing the CD4 receptor, and therefore, could be HIV-1 infected. The remaining CD4- cell population are uninfectable by HIV-1 and so, any iPSCs generated from these cells would not contain any HIV-1 provirus. Utilising MagACS technology, it was found that a double CD4+ cell depletion was optimal for the removal of the CD4+ cells, whilst retaining the majority of the CD4- population. The CD4- cell population would then be reprogrammed into iPSCs using the previously optimised protocol.

To detect if any HIV-1 provirus was present in cells, it was important to have a robust, sensitive and specific assay in place. The optimisation of a qPCR-based detection assay is shown in this chapter. The use of 50ng of input gDNA for the qPCR allowed for a strong signal when detecting the provirus whilst remaining undetected in the negative controls. The detection of the endogenous gene, RNaseP, was used as a loading control. This gives confirmation of gDNA being present if the sample is analysed as HIV-1 provirus negative.

Despite the approach taken to eliminate any HIV-1+ cells from a sample of PBMCs, it was demonstrated that after a double CD4+ cell depletion, HIV-1 provirus could still be detected by qPCR in the CD4- cell fraction. This outcome may be explained by downregulation of the cell surface receptor CD4 on HIV-1 infected cells via the action of the viral proteins Nef and Vpu (Lindwasser et al., 2007). Nef establishes connections between mature CD4 and elements involved in clathrin-dependent trafficking pathways at the cell surface, and potentially within intracellular regions. This results in the internalisation of CD4 and its transportation to lysosomes for degradation. Vpu engages with newly-synthesised CD4 within the endoplasmic reticulum, forming a link between CD4 and the Skp1-Cullin-F-box E3 ubiquitin ligase, thereby promoting CD4's entry into the endoplasmic reticulum-associated degradation pathway. These mechanisms are essential for HIV-1 replication because, if present, CD4 can sequester Env on progeny viruses lowering their infectivity (Lama et al., 1999). Although HIV-1 provirus was still detectable after the depletion, it was demonstrated that, by using the previously mentioned optimisations, iPSCs not containing any HIV-1 provirus could be generated from a population of PBMCs containing HIV-1 infected cells. This successful generation of provirus-free iPSCs is likely due to only a small number of HIV+ cells remaining within the CD4- cell population, and therefore, there is only a low chance of one those cells being reprogrammed and selected as an iPSC clonal line. Furthermore, it is unclear which cell type within the heterogenous PBMC population is reprogrammed, thus, the cells potentially HIV-1 infected, such as CD4+ T-cells and monocytes, may not be the ones that are reprogrammed.

In summary, results in this chapter show the establishment, optimisation and validation of a reproducible protocol for reprogramming an HIV-1 infected PBMC population to iPSCs. Having this in place allowed for progression onto reprogramming the MACS samples with the upmost confidence of success.

Chapter 4: Reprogramming of the MACS PBMC samples

4.1 Introduction

The results in the previous chapter demonstrate the establishment of an optimised pipeline for reprogramming an HIV-1 infected PBMC sample into HIV-1 provirus-free iPSCs. Developing a robust protocol was critical due to the nature of the MACS PBMC samples that were to be reprogrammed next. The MACS (further reviewed in section **1.6**) is a prospective study, started in the U.S in 1981, of homosexual and bisexual men at high risk of HIV-1 infection whose natural and treated histories of HIV-1 infection were tracked in great detail (Kaslow et al., 1987). PBMC samples were collected from participants throughout their study attendance. As a result of the nature of the MACS and the high scientific demand for such samples, the PBMC samples are a finite resource that are extremely valuable from a scientific view point.

The MACS patients to have their PBMCs reprogrammed into iPSCs were selected based upon their clinical data in response to HIV-1 infection. Despite all being infected with HIV-1, there is a broad range of how well the study participants respond to the virus in terms of disease progression. In some patients, the virus causes a rapid physiological decline, leading to AIDS quickly, whereas, other patients are able to control the virus and never progress to AIDS; even in the absence of ART. Therefore, host factors must play a significant role in determining the outcome of HIV-1 infection (Blanquart et al., 2017). Polymorphisms within CCR5 and HLA have been previously identified as host factors responsible for differences in disease progression, however, these only account for a small proportion of the variation seen (An & Winkler, 2010; Dean et al., 1996; Huang et al., 1996; International HIV Controllers Study et al., 2010). It is the MACS patients displaying extreme clinical phenotypes that are of particular interest to investigate further as they present the greatest disparity in disease progression and so are mostly likely to show variation in host factors that strongly influence the success of HIV-1 replication. Thus, it was these participant's PBMC samples that were chosen to be reprogrammed into iPSCs.

This chapter presents some of the clinical data (provided by our collaborators at Northwestern university) behind the selection of the 18 MACS participant to have their PBMCs

reprogrammed into iPSCs. It also shows the results from the generation of the iPSCs and the validations conducted on them to ensure their quality.

4.2 Results

4.2.1 MACS participant selection based upon their clinical history in response to HIV-1 infection

The selection of the MACS participants to derive iPSCs from was based upon the comprehensive clinical data, including both clinical outcome data and molecular data, collected throughout their period of enrolment in the study. The clinical outcome data consisted of information regarding clinical milestones, such as if/when the patient developed an AIDS defining opportunistic infection or malignancy, or when any ART was started. Whereas the molecular data were comprised of clinically relevant molecular signals, such as CD4+ T-cell count or HIV-1 viral load. The raw data discussed in section **4.2.1** were provided to me by our collaborators at Northwestern University, though the analysis and presentation of it was completed by myself.

Based on their clinical data in response to HIV-1 infection, participants in the MACS were categorised into groups, defined as their clinical phenotype. For the purpose of this investigation, the more extreme clinical phenotypes were of the greatest interest as these presented the greatest potential for identifying host factors that influence the rate of HIV-1 replication. These extreme clinical phenotypes and the criteria used to define them are outlined in **table 4.1**. At one end of the clinical phenotype spectrum is the RPs. This group of individuals is characterised by a very poor ability to control viral replication and, as a result, the patient succumbs to the virus's pathological effects rapidly. For this study, we used two RP identifying criteria, where if a MACS participant matched either of them they were classified as a RP. The first was based on clinical outcome data and stated that if the patient developed an AIDS defining opportunistic infection or malignancy within 5 years of seroconversion they were classified as a RP. The second was based on molecular clinical data and specified that a CD4+ T-cell decline of >25 cells/µl/month over the first two years after seroconversion classified that donor as a RP. At the other extreme of the clinical phenotype scale, ECs are characterised by their ability to control the virus extremely well. These individuals are defined by maintaining their viral load at undetectable levels (<50 HIV RNA copies/ml) on 2 or more occasions within 1.5 years in the absence of therapy. The non-progressor (NP) clinical

phenotype sits towards the same end of the spectrum as the ECs. They are defined by the study participant not developing AIDS over an extended period of time despite not receiving any ART. If this period of time reaches 15 years, the NP is reclassified as a long-term non-progressor (LTNP). Individuals can match the criteria for multiple clinical phenotypes, so, for example, an EC could also be a LTNP. Another clinical phenotype of relevance to this project is the VC. VCs also have the ability to control the virus well, however, not to quite the same level as the ECs. These individuals keep their viral load at <2000 HIV RNA copies/ml on 2 or more occasions within 1.5 years in the absence of therapy. The final clinical phenotype is a variation of the VC phenotype, adapted by our collaborators at Northwestern University called Sooty-like VCs (VC – Sootys). These individuals are a particularly clinically fascinating cohort, defined by presenting a high viral load (>10,000 HIV RNA copies/ml) but maintaining a CD4+ T-cell level >300 cell/µl and not developing an AIDS defining illness over a time period of >8 years. This group are named so after a species of primate, sooty mangabeys, that despite exhibiting a high SIV viral load, do not experience the pathological effects of the virus and never progress to AIDS (Silvestri et al., 2003).

Taking into account these clinical phenotypes and PBMC sample availability, 18 MACS participants were chosen to have their cryopreserved PBMCs reprogrammed into iPSCs. The list of donors and details about the PBMC samples is displayed in **table 4.2**. Out of the 18 donors, there were 4 RPs, 4 ECs (2 of which were also LTNPs), 3 NPs, 4 LTNPs, 1 VC and 2 VC – Sootys. The date that the PBMC samples were collected ranged from 1985 to 2014 and were collected on a wide variety MACS visit numbers. The majority of the samples, 14 out of 18, were collected at a time point when the participant was seropositive for HIV-1.

Table 4.1: Clinical phenotypes of participants from the MACS relevant to this project and the criteria used to define them.

Clinical phenotype	Abbreviation	Criteria
Rapid progressor	RP	AIDS defining opportunistic infection or malignancy within 5 years of seroconversion Or CD4+ T-cell decline >25 cells/µl/month over first 2 years after seroconversion
Elite controller	EC	<50 HIV RNA copies/ml on 2 or more occasions within 1.5 years in the absense of therapy
Non-progressor	NP	Remaining AIDS-free in the absense of therapy
Long-term non-progressor	LTNP	Remaining AIDS-free in the absense of therapy for >15 years
Viraemic controller	VC	<2,000 HIV RNA copies/ml on 2 or more occasions within 1.5 years in the absense of therapy
Sooty-like viraemic controller	VC - Sooty	>8 years follow up since seroconversion with the absense of therapy Never diagnosed with AIDS >10,000 HIV RNA copies/ml at all points within time period >300 CD4+ T-cell/µl at all points within time period

Table 4.2: Details of the MACS PBMC samples selected to be reprogrammed into iPSCs.

MACS ID	Clinical phenotype	Sample date	MACS visit	Post HIV-1 infection?
20032	RP	02/07/1990	13	Yes
20318	RP	12/04/1999	31	Yes
20480	RP	22/02/1999	32	Yes
20543	RP	02/04/1985	45	No
24020	EC	03/09/2003	39	Yes
24208	EC + LTNP	16/04/2003	39	Yes
24312	EC + LTNP	10/04/2012	57	Yes
25501	EC	18/02/2014	50	Yes
20152	NP	10/06/1986	5	No
20446	NP	09/02/1993	8	No
20722	NP	19/11/1998	30	Yes
20176	LTNP	25/04/1989	11	Yes
20316	LTNP	18/07/2001	35	Yes
20451	LTNP	12/09/2006	45	Yes
20777	LTNP	01/05/2007	47	Yes
20236	VC	21/01/1988	8	No
21052	VC - Sooty	03/07/1986	5	Yes
41624	VC - Sooty	05/06/1992	17	Yes

Since the clinical data were collected at all participant visits to the MACS, the molecular based clinical data for each participant can be plotted longitudinally. Displaying the data in this manner highlights the variation in patient responses to HIV-1 between the clinical phenotypes. **Figure 4.1** shows representative molecular clinical data from 2 of the 18 donors chosen from the MACS. Exhibited here are the CD4+ T-cell counts and HIV-1 viral loads for a donor from each end of the clinical outcome spectrum; donor 20032, a RP, and donor 25501, an EC. **Figure 4.1A** demonstrates a steep decline of donor 20032's CD4+ T-cell count after HIV-1 infection, despite initiation of ART after ~2 years. In correlation with this, **figure 4.1B** shows a fluctuating but overall increasing viral load for the same patient. Donor 20032 was then diagnosed with AIDS ~4 years after seroconversion and died within 1 year as a result. In contrast, after HIV-1 infection, donor 25501, an EC, maintained a steady CD4+ T-cell count in the absence of any ART (**Figure 4.2C**). Additionally, as displayed in **figure 4.2D**, participant 25501 maintains undetectable levels of HIV-1 viral load for almost 6 years.

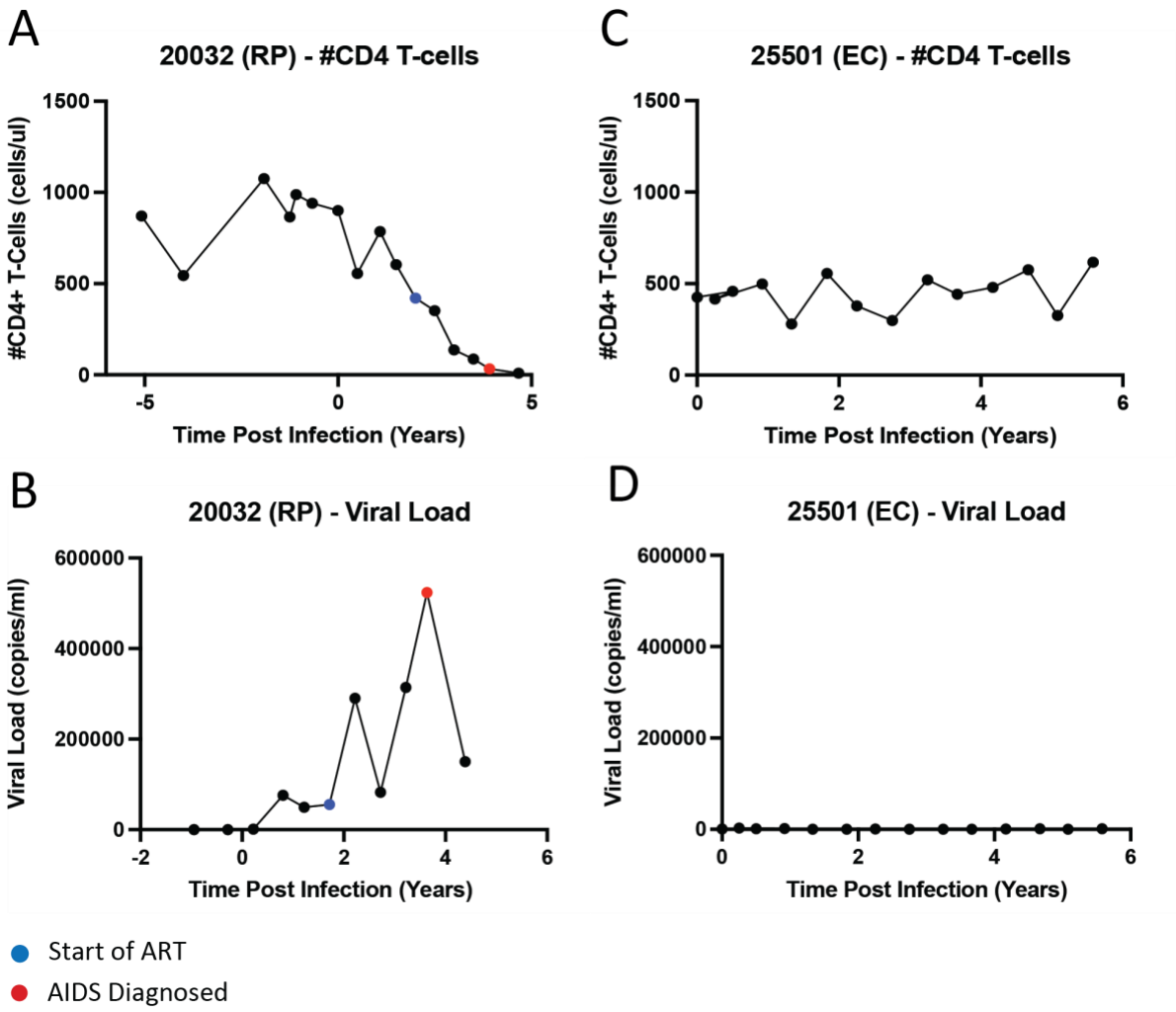


Figure 4.1: Longitudinal analysis comparing the molecular based clinical data of two MACS participants with opposite clinical outcomes to HIV-1 infection.
A+B) Longitudinal data for donor 20032, a RP, plotting their CD4+ T-cell count and viral load overtime respectively. **C+D)** Longitudinal data for donor 25501, an EC, plotting their CD4+ T-cell count and viral load overtime respectively. Blue data points represent the time point when ART was initiated. Red data points indicate the time point of an AIDS diagnosis.

4.2.2 Phenotyping of the MACS PBMC samples

Upon thawing of the MACS PBMC samples, a small proportion was analysed to identify the proportion of different cell types present in the sample. The samples were stained, using a cocktail of antibodies formulated to distinguish the different immune cell types, fixed and then analysed by flow cytometry (**figure 4.2**). The gating strategy taken to identify live PBMCs was as shown in **figure 3.4A**. The live PBMCs were then examined for their distinguishing cell surface marker expression to identify the proportions of each immune cell type within the full PBMC population. CD8+ T-cells were characterised by co-expression of CD3 and CD8 (**Figure 4.2A**). CD4+ T-cells were identified by co-expression of CD3 and CD4 (**Figure 4.2C**). NK cells were distinguished by being CD3- and CD56+ (**Figure 4.2E**). Monocytes were identified by expression of CD14 (**Figure 4.2G**). B-cells were distinguished by expression of CD19 (**Figure 4.2I**).

The proportions of the immune cell types in each donor's PBMC samples are also displayed in **figure 4.2**, where the dotted lines in each graph represent the upper and lower boundaries for proportions of that cell type expected in a healthy donor. For the proportion of CD8+ T-cells, 11 of the 18 PBMC samples lie within the healthy range (**figure 4.2B**). The remaining 7 samples had a higher proportion of CD8+ T-cells than expected in a healthy donor. 3 of these samples were <10% out of the normal range and the other 4 samples were between 10-20% outside of the healthy range. **Figure 4.2D** displays the proportions of CD4+ T-cells present in the MACS PBMC samples. Generally, the proportion of CD4+ T-cells present was low, with none of the PBMC samples containing a CD4+ T-cell proportion in the upper half of the standard healthy range. Additionally, just 8 of the 18 PBMC samples showed levels within the usual range, with all remaining 10 displaying CD4+ T-cell proportions below the expected range. This could be explained by the majority of the donors being HIV-1 seropositive at the time of PBMC collection, and therefore, their CD4+ T-cells have been somewhat depleted. In support of this explanation, of the 8 donors that had CD4+ T-cell proportion within the healthy range, 4 were collected at a time when the donors were HIV- and the remaining 4 were from donors exhibiting clinical phenotypes that suggest good control of their HIV-1 infection, and therefore, a better maintenance of CD4+ T-cell count (2 ECs and 2 LTNPs). The proportion of NK cells showed variation across the samples, with 1 donor containing a higher proportion than the healthy range and 8 donors presenting a lower proportion than the normal range

(**figure 4.2F**). The proportion of monocytes tended to be on the higher side of the healthy donor range (**figure 4.2H**). 15 of the MACS PBMC samples had a proportion of monocytes above the anticipated range. The PBMC sample from donor 20480 exhibited a monocyte proportion of almost 4X the upper boundary of the healthy range. The proportion of B-cells in the PBMC samples fell mostly within or close to the healthy range. Exceptions to this were donor 24312 who displayed a proportion of B-cells almost 2.5X that of the upper boundary of the healthy range, and donors 24020 and 24208 who had only half the expected relative proportion of B-cells (**figure 4.2J**).

For all of the MACS PBMC samples, at least a small proportion of each cell type was present and generally, the proportions of each immune cell type was within or close to the expected range of a healthy PBMC donor. Part of the reason for some samples lying outside of the expected range is probably attributable to the HIV-1+ nature of the patient when the PBMCs were collected. This is supported by all 4 of the PBMC samples that were collected at a time point when the donor was HIV- being within or very close to the expected healthy range for all immune cell types. The PBMC phenotyping confirmed that the MACS PBMC samples contained a diverse mix of the expected cell types despite being HIV+ and cryopreserved for an extended period of time. Therefore, the MACS PBMC samples were considered to be of reasonably high quality to be taken forward for reprogramming.

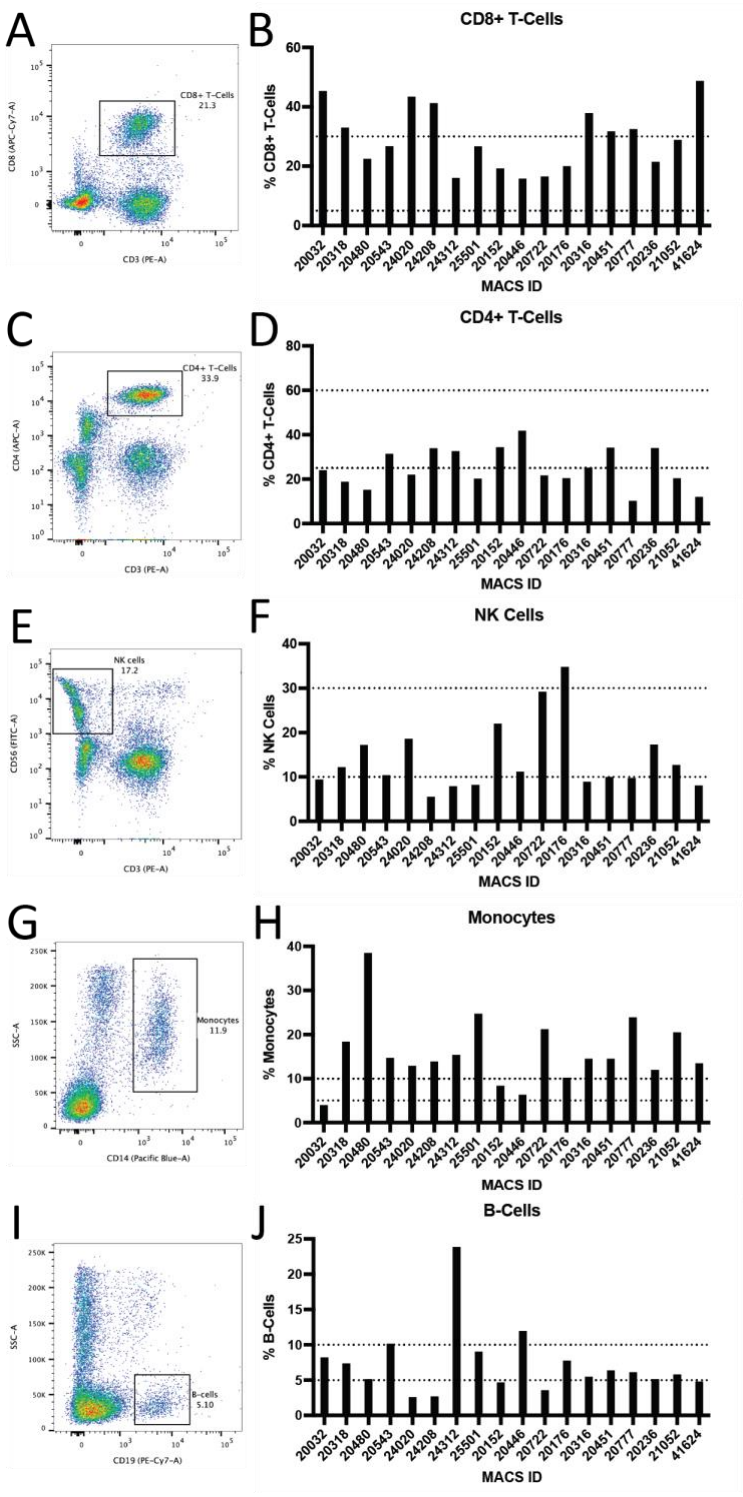


Figure 4.2: Phenotyping of the MACS PBMC samples.
A+C+E+G+I) Gating strategy taken on live PBMCs to identify CD8+ T-cells, CD4+ T-cells, NK cells, monocytes and B-cells respectively. **B+D+F+H+J)** Proportion of the full PBMC population from each MACS donor made up of CD8+ T-cells, CD4+ T-cells, NK cells, monocytes and B-cells respectively. Dotted lines demonstrate the upper and lower boundaries of each cell type expected in a healthy donor.

4.2.3 Reprogramming of MACS PBMC samples to iPSCs

Over the series of 5 reprogramming experiments, all 18 MACS donor PBMC samples, listed in **table 4.2**, were successfully reprogrammed into iPSCs, termed MACS iPSCs. Despite some samples being collected before the donor was infected with HIV-1, in the interest of consistency, all samples were double depleted for CD4+ cells prior to the reprogramming process. For each donor, the aim was to generate 3 clonal iPSC lines and the nomenclature of the generated iPSC lines is the donor number followed by either A, B or C depending on the clone (For example, the second clonal iPSC line generated from the PBMCs of donor 20543 is named '20543B'). 3 clonal lines were successfully expanded and banked for 14 of the MACS donors, however, only two lines were able to be expanded and banked for the remaining 4 donors. The donors with just two iPSC lines are 20543, 24208, 24312 and 41624. Hence, a total of 50 iPSC lines were derived from 18 MACS donors.

A light microscopy image of a representative MACS iPSC line, 20480A, is shown in **figure 4.3**. This cell line exhibits the expected morphological characteristics of an iPSC line of high quality. The colonies in the culture are rounded, with smooth, defined edges, and are comprised of compacted cells. There are minimal signs of overgrowth and little to no differentiation present. All 49 other derived MACS iPSC lines also exhibited these typical morphological characteristics of an iPSC line giving a good indication that all 50 MACS derived iPSC lines are of high quality (**Supplementary file 4.1**).

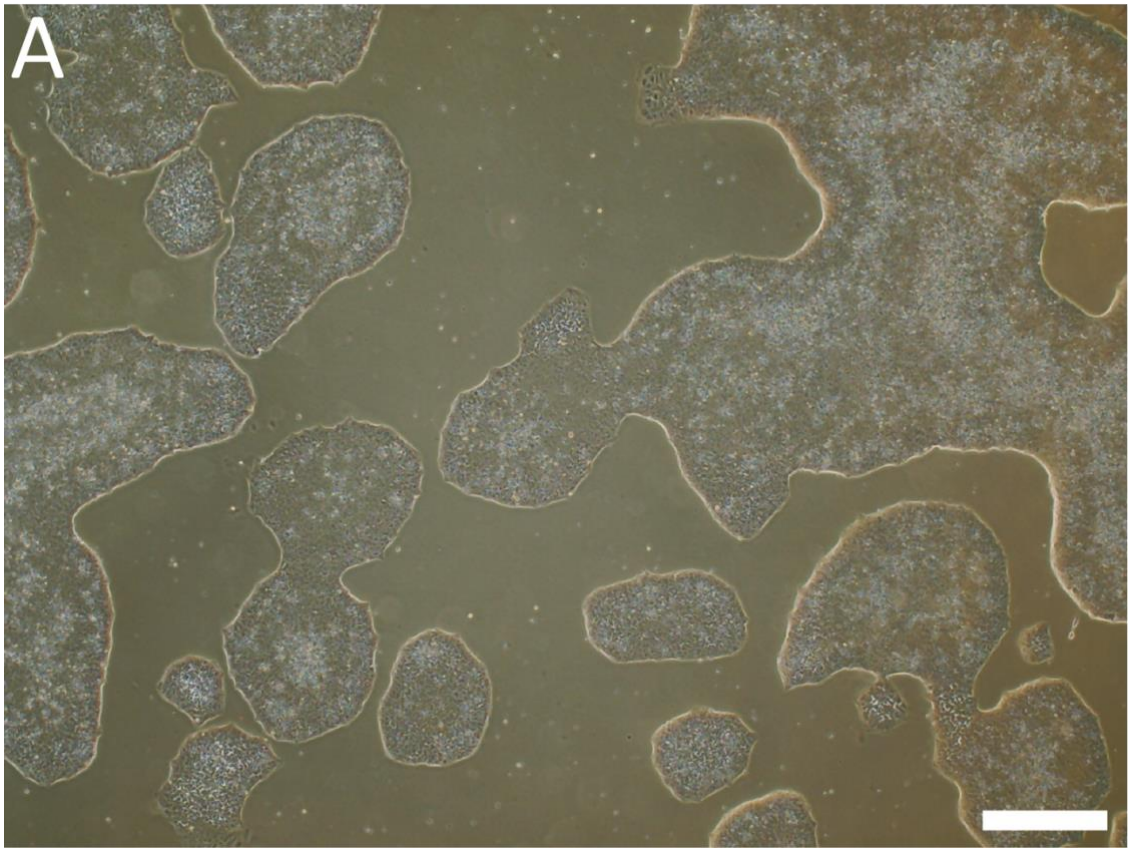


Figure 4.3: Representative image of a MACS PBMC sample-derived iPSC line demonstrating typical iPSC morphological characteristics.
A) Light microscopy image of MACS iPSC line 20480A. Image taken at 4X magnification and scale bar represents 500μm.

4.2.4 Pluripotency marker expression in MACS iPSCs

Although the generated iPSC lines demonstrated excellent iPSC morphology, further examinations were required to validate them as iPSCs. The MACS iPSCs were analysed for their expression of known pluripotency markers. As discussed previously in section **3.2.1**, OCT-4 is a well-defined pluripotency marker to use for this. Additionally, in the case of the MACS generated iPSCs, a second pluripotency marker, NANOG, was used for extra validation of pluripotency. The results for the expression of OCT-4 and NANOG in the MACS iPSC lines, with extra focus on a representative line, 20480A, are shown in **figure 4.4**.

The analysis of OCT-4 and NANOG expression in the representative iPSC line, 20480A, is shown in **figure 4.4A+B** respectively. The gating strategy taken was as shown in **figure 3.1B**. In 20480A, 96.2% of the cells were positive for OCT-4 and 89.6% of the cells were positive for NANOG. Hence, percentage expression of both of these markers far exceeded the aforementioned threshold of 70% for a cell line to be deemed pluripotent (Sullivan et al., 2018). **Figure 4.4C** demonstrates the OCT-4 and NANOG expression as a percentage when analysed by flow cytometry for all of the MACS generated cell lines. 49 out of 50 of the MACS iPSC lines contained >90% of cells that were OCT-4 positive. The one remaining cell had a lower percentage OCT-4 expression, 82.4%, however, this still surpassed the aforementioned threshold. Expression of NANOG was slightly more variable and tended to be lower than OCT-4 expression. Despite this, all the MACS iPSC lines still had a percentage expression higher than the threshold, with the lowest line showing 74.8% of cells expressing NANOG. OCT-4 expression was also analysed by immunofluorescence via confocal microscopy in 30 of the MACS iPSC lines. **Figure 4.4D** exhibits the representative line, 20480A, when analysed for OCT-4 expression by microscopy. The OCT-4 expression is very strong and its localisation correlates highly with the cell nuclei, identified by DAPI staining. This is as expected since OCT-4 is expressed in the nucleus of iPSCs. Analysing the MACS iPSCs for two pluripotency markers and by two technologically distinct methods grants further validation that the generated cell lines are iPSCs.

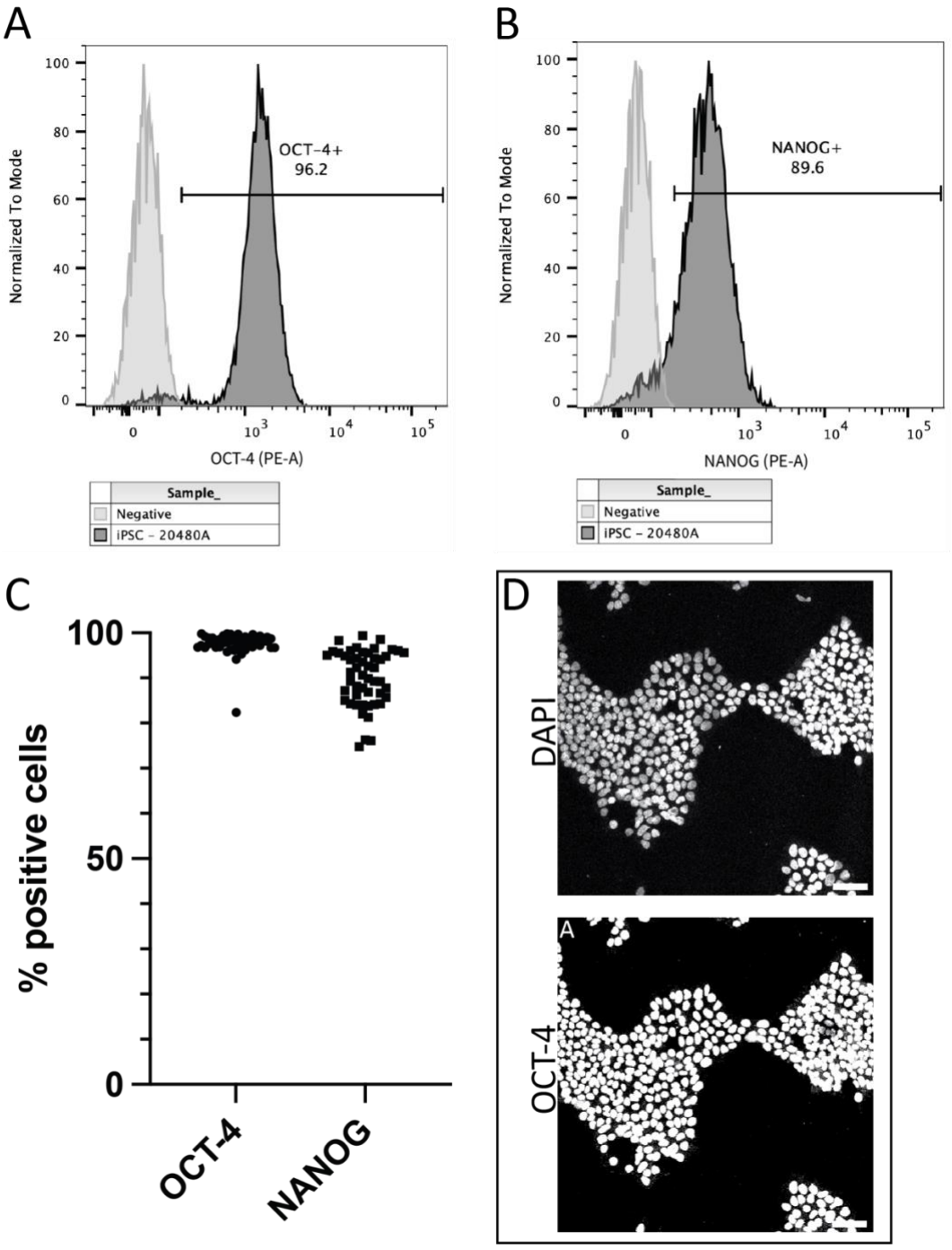


Figure 4.4: Expression of pluripotency markers, OCT-4 and NANOG, in the MACS iPSC lines, with particular focus on a representative line, 20480A.

A+B) Expression of OCT-4 and NANOG, respectively, in the representative iPSC line, 20480A. The light grey and dark grey peaks represent a stained negative control (HEK293Ts) and 20480A respectively. **C)** Percentage expression of OCT-4 and NANOG for all MACS iPSCs when analysed by flow cytometry. **D)** Confocal microscopy images to demonstrate expression of OCT-4 in 20480A. The top image shows the nucleus via DAPI staining and the bottom image presents detection of OCT-4. Confocal images taken at 20X magnification and scale bars represent 100μm.

4.2.5 Use of a trilineage differentiation assay to further validate the pluripotency of the MACS iPSCs

To further validate the iPSCs for their pluripotency potential, a trilineage differentiation assay was performed. This assay was used to demonstrate the ability of the cells to differentiate into all three germ layers (ectoderm, mesoderm and endoderm), and therefore, be classified as pluripotent. For each iPSC line, the directed differentiation into each lineage was conducted independently. To confirm the successful differentiation of the cell lines, the lineage-specific differentiated cells were analysed for expression of a marker exclusive to that lineage in comparison to undifferentiated iPSCs, acting as a negative control. This assay was conducted for 30 out of the 50 MACS iPSC lines, and the results for a representative line, 20480A, are displayed in **figure 4.5**.

Figure 4.5A demonstrates the ability of MACS iPSC line 20480A to differentiate into the ectoderm. To show this, both undifferentiated iPSCs and iPSCs that had undergone directed differentiation to ectoderm were analysed for their expression of an ectoderm specific marker, Pax-6. DAPI was used to identify the cell nuclei and Pax-6 expression in the ectoderm differentiation cells correlated well with this. Although detection of expression for this marker was low, the signal was nuclear specific and wasn't detected in the undifferentiated iPSCs also stained for Pax-6. The lineage exclusive marker for mesoderm determination was Brachyury and the results for its detection are shown in **figure 4.5B**. There is a clear signal for the expression of Brachyury in the mesoderm differentiated cells that correlates with the cell nuclei. This is compared to no signal in the undifferentiated iPSCs when stained for Brachyury. To determine if the cells could differentiate into the endoderm, expression of SOX-17 was analysed (**figure 4.5C**). The undifferentiated iPSCs showed no expression of SOX-17, whereas the endoderm differentiated cells showed high level of SOX-17 expression that correlated with the nuclei location. However, for the endoderm differentiated cells, not all of the cells expressed SOX-17 and therefore only some can be confirmed as having the potential to differentiate to the endoderm lineage. However, since the iPSC lines are clonal, this is more likely to be a result of poor differentiation, as opposed to an inability to differentiate.

Alongside the trilineage differentiation assay, the cells were also analysed for their OCT-4 expression before and after differentiation (**Figure 4.5D**). This was conducted as another

means of validating the differentiation ability of these cells. When the iPSCs undergo differentiation, their expression of the pluripotency marker, OCT-4, should diminish. As previously shown in **figure 4.4**, the undifferentiated iPSC line 20480A shows a high level of OCT-4 expression. After directed differentiation into either the ectoderm or mesoderm lineage, the expression of OCT-4 is completely lost. In contrast, for the endoderm differentiated cells, OCT-4 expression remains for many of the cells in the culture, albeit at a lower level of expression. This decreased amount of differentiation, demonstrated by maintenance of some OCT-4 expression, reinforces the above explanation of poor differentiation for why only some of the endoderm differentiated cells expressed the endoderm marker SOX-17.

The confirmation of the ability of the iPSC line to differentiate into all three germ layers, along with the loss of OCT-4 expression during differentiation, provides strong evidence that the 30 tested MACS iPSCs are pluripotent. Immunofluorescence images for the other 29 MACS iPSC lines are shown in **supplementary file 4.2**.

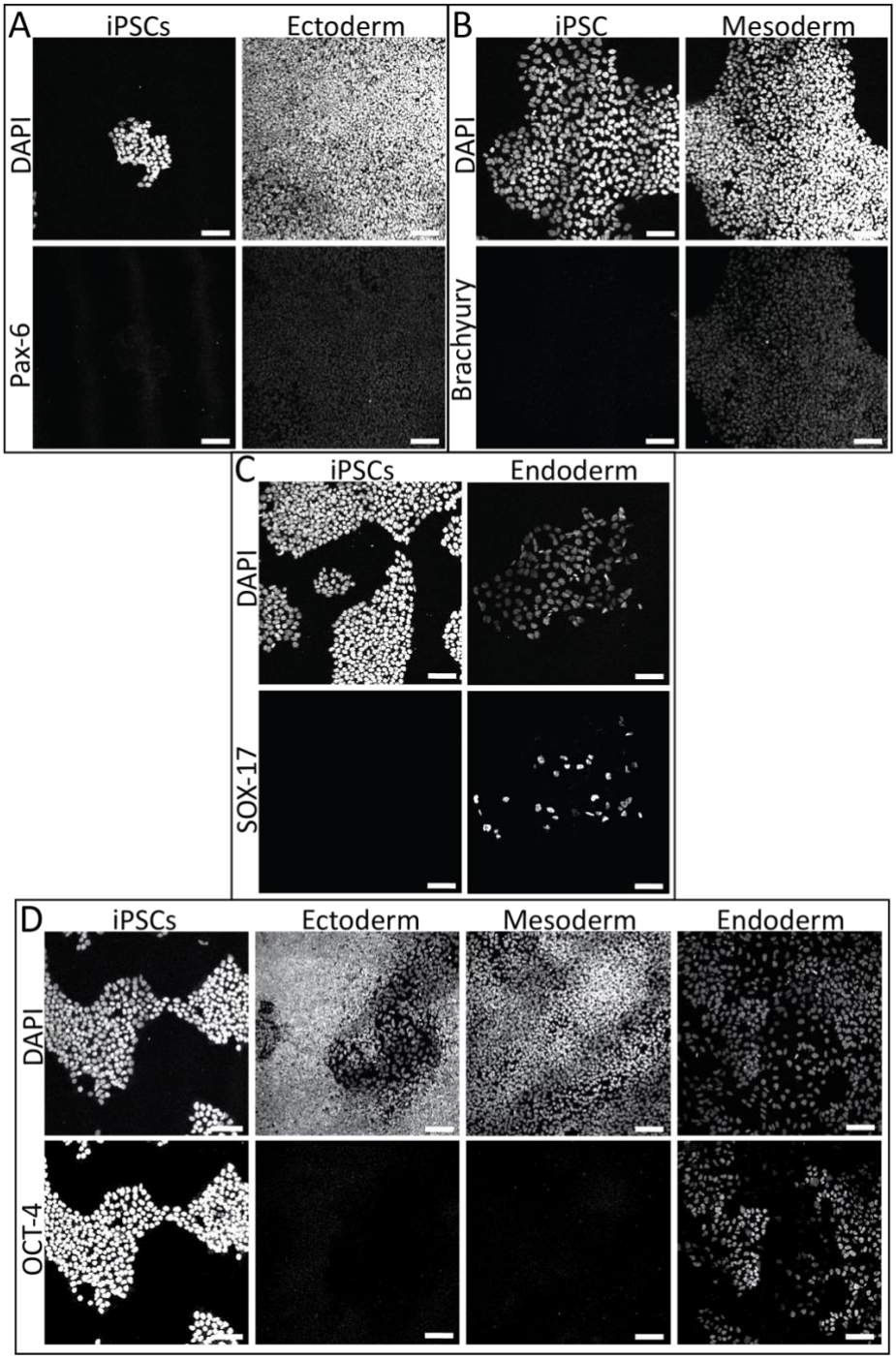


Figure 4.5: Directed differentiation of a representative MACS iPSC line, 20480A, into the three germ layers and analysis of OCT-4 expression before and after differentiation.
A+B+C) Immunofluorescence images demonstrating the ability of an iPSC line to differentiate into ectoderm, mesoderm and endoderm lineages respectively. Top left and right images show undifferentiated iPSCs and lineage differentiated cells stained with DAPI respectively. Bottom left and right images display the same undifferentiated iPSCs and lineage differentiated cells stained for that lineage's specific marker respectively. **D)** Immunofluorescence images to show OCT-4 expression before and after lineage specific differentiation. Images taken at 20X magnification and scale bars represent 100μm.

4.2.6 Pluripotency analysis of MACS iPSCs by PluriTest

Although the pluripotency of 30 of the MACS iPSCs had been confirmed by the trilineage differentiation assay, an extra bioinformatic-based pluripotency assay was trialled, on the same 30 cell lines, for additional confirmation. This assay, named 'PluriTest', utilises raw RNA-seq data of a cell line and compares it to the transcription profiles of numerous confirmed pluripotent and somatic cell lines using a data-driven model of pluripotency to determine whether or not the tested cell line is in a pluripotent state (Müller et al., 2011). The PluriTest generates two outputs: a pluripotency score and a novelty score. The pluripotency score represents the likelihood of a cell line to be pluripotent; the higher the score, the better pluripotent signature that cell line exhibits. The novelty score represents how well the measured signal relates to the other iPSC lines in the dataset; the lower the value, the more closely related to the model that cell line's transcriptome is.

The RNA-seq data analysed here was generated from RNA harvested from iPSCs collected during the infection pipeline as described in section **2.4.2**. The results of the PluriTest analysis of this RNA-seq data for the 30 tested MACS iPSC lines are shown in **figure 4.6A**. Firstly, the MACS iPSC lines, indicated by black data points, cluster together signifying similarity between their transcriptomes and, therefore, good consistency in their derivation and maintenance. All of the MACS iPSC lines fell below the pluripotency empirical threshold (threshold = 20), represented by the horizontal dashed line, which is set to give high sensitivity and specificity for iPSC detection based upon the model used. iPSCs typically display a pluripotency score of >20, partially reprogrammed cells demonstrate a pluripotency score of roughly -50, and somatic cells typically present a pluripotency score of between -50 and -100 (Müller et al., 2011). The green, orange, and red data points in **figure 4.6A** represent where other iPSCs, partially reprogrammed cells, and fibroblasts would typically plot respectively. In contrast to the partially reprogrammed and somatic cells, all of the MACS iPSC lines display a pluripotency score of >0, thus, demonstrating that although they don't surpass the threshold, they exhibit a reasonably strong pluripotency signature. A potential explanation for the pluripotency scores not being above the threshold may be the time point at which the cells were harvested for RNA analysis. Due to the primary purpose of the RNA-seq data for each of these cell lines being to compare their transcriptomes at the time of HIV-1 infection, a population of each cell line was collected for RNA analysis after being seeded in the same conditions as for the

infectivity screen pipeline. This meant the cells were not in their optimal pluripotent condition as they had been dissociated to single cells; they had been passaged only ~16 hours prior, and they were in the presence of a media supplement, RevitaCell. These factors would have a large impact on an iPSC lines transcriptome and could greatly reduce its pluripotency signature in comparison to an iPSC line growing as colonies, 4-6 days after a standard passage with no additional media supplements. This explanation is reinforced by the PluriTest results of an extremely well-characterised, known iPSC line, the CTRs. Cells from this iPSC line were also harvested for RNA analysis under the same conditions as the MACS iPSC lines and also displayed a pluripotency score below the threshold (11.52), represented by the grey data point in **figure 4.6A**. This considerably reduces any concerns over the pluripotency scores of the MACS iPSC lines and indicates that the time point when the RNA was extracted is a critical factor. Therefore, the MACS iPSC lines should be reanalysed by PluriTest using RNA-seq data generated from an RNA sample taken at a time point when the cell lines are of high quality, exhibiting classic iPSC colony morphology, during routine culture.

The novelty scores of the MACS iPSC lines and the CTR line all clustered around the novelty empirical threshold, represented by the vertical dashed line. A high novelty score can indicate a cell line may have epigenetic or genetical abnormalities, or unwanted differentiation. Since the pluripotency score is below the threshold for all the tested MACS iPSCs, the novelty score is not necessary to analyse as the cells will not closely relate to the iPSCs used in the model.

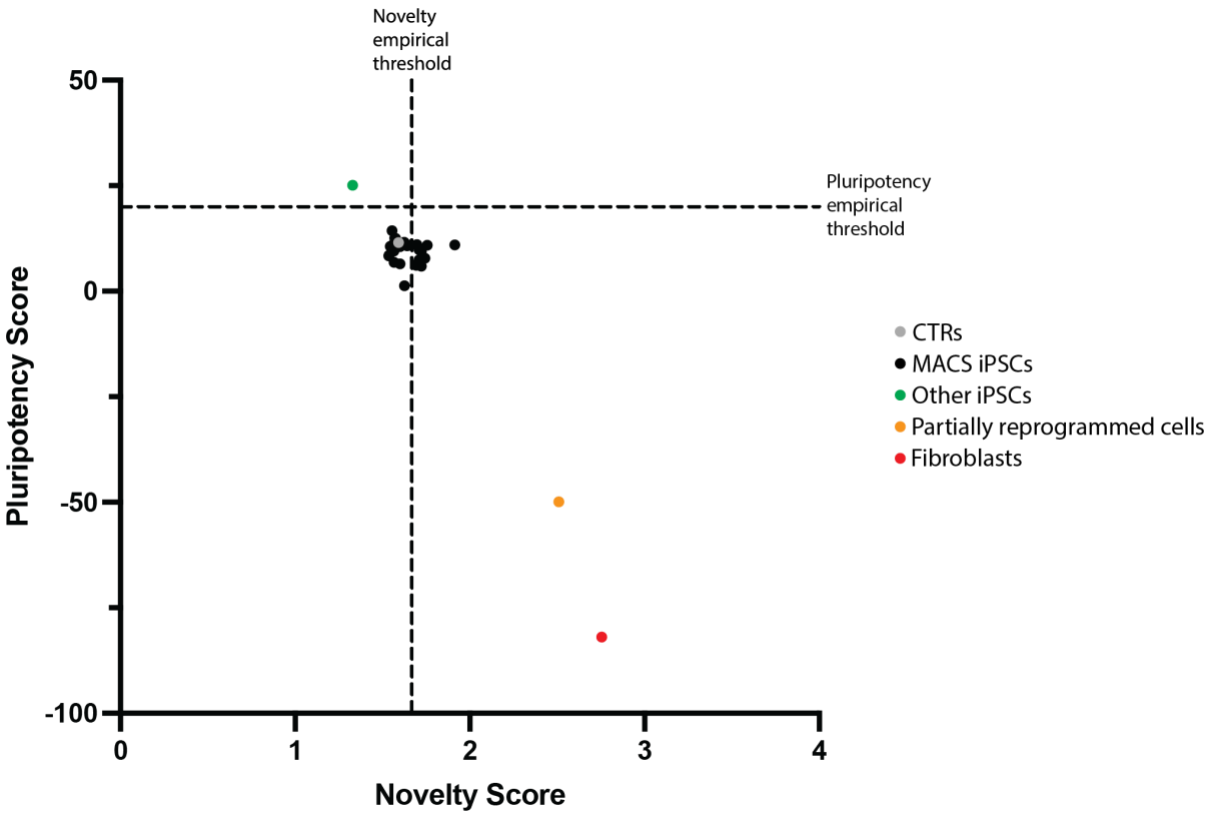


Figure 4.6: PluriTest analysis results for the MACS iPSCs and the CTR iPSC line.
A) PluriTest analysis results of 30 MACS iPSC lines and a known iPSC line, the CTRs. MACS iPSCs, CTRs, other iPSCs, partially reprogrammed cells, and fibroblasts are represented by black, grey, green, orange and red data points respectively. The vertical and horizontal dashed lines represent the novelty empirical threshold and the pluripotency empirical threshold respectively.

4.2.7 Generation of Sendai viral vector-free iPSCs from the MACS PBMC samples

The introduction of the four 'Yamanaka factors' that induce reprogramming of PBMCs into iPSCs was achieved using Sendai viral vectors. These non-integrating vectors remain in the cytoplasm to provide their function, before being passively removed by dilution over cell divisions. However, these vectors can often persist longer than necessary inside the newly generated iPSCs. It is important to remove these remaining vectors so the iPSCs contain no foreign DNA that could prove problematic in future work. For example, when conducting the HIV-1 permissivity phenotyping screen (to be discussed in **chapter 5**), it is important that the cells contain no foreign DNA, especially that from viruses, as it could influence the cellular response to new incoming virions.

To detect remaining Sendai virus in the MACS iPSC lines, RNA was extracted and converted to cDNA. qPCR was then used to detect the presence of SeV and each of the individual Sendai viral vectors (KOS, Klf4 and c-Myc) in the cDNA sample (**Figure 4.7A**). GAPDH was used as a loading control. All of the MACS iPSC lines were initially analysed at passage ~10. At this time point, SeV was still detectable in 43 of the MACS iPSC lines, therefore, only 7 lines had efficiently removed all three of the Sendai viral vectors to undetectable levels. For the cell lines containing persisting SeV, in general, the KOS vector was detected at moderate levels and the c-Myc vector was detected at high levels. Importantly, the Klf4 vector had been removed from all the cell lines (except very low detection in 20446A).

The significance of efficient removal of the Klf4 vector was due to the KOS and c-Myc Sendai viral vectors both containing temperature sensitive mutations. Hence, an increase in the culture temperature of the MACS iPSC lines containing these vectors, from 37°C to 38.5°C, should aid their removal. The cells were cultured at this increased temperature for 5 days and then passaged twice before being analysed for Sendai viral vector presence again. After this treatment, as shown in **figure 4.7A**, SeV was undetectable in all but 4 of the MACS iPSC lines. For the 4 remaining lines, SeV was still detectable but at lower levels than before the treatment. To completely remove the remaining SeV, these 4 cell lines underwent a second round of the 38.5°C treatment. SeV was then undetectable. As a result of this, all 50 MACS iPSC lines were now confirmed to be free of any Sendai viral vectors.

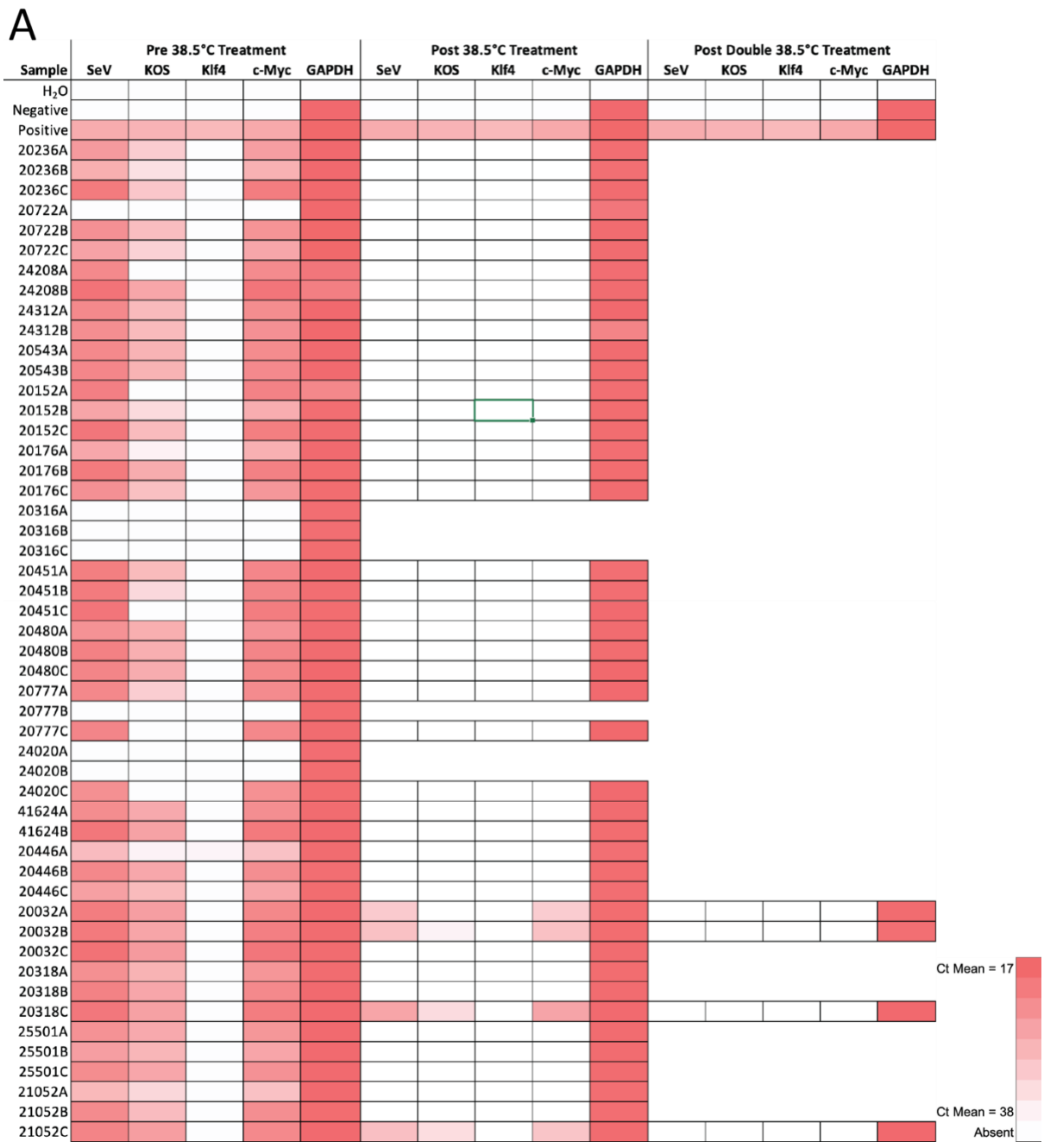


Figure 4.7: Detection of the Sendai viral vectors in MACS iPSCs by qPCR.
A) qPCR results for detection of the Sendai virus reprogramming vectors in the MACS iPSC lines before 38.5°C treatment, after 38.5°C treatment and after double 38.5°C treatment. SeV = detection of universal Sendai Virus. KOS, Klf4 and c-Myc = detection of each Sendai vector independently. GAPDH = Loading control. As indicated in the heat map scale bars, the darker the red box equates to a lower mean Ct value and, therefore, a higher amount of the Sendai viral vector present in the sample. White boxes indicate an undetected Ct mean value and, therefore, an absence of the sendai viral vector. All results shown are a Ct mean result of three Ct values.

4.2.8 Detection of HIV-1 provirus in the MACS generated iPSCs

Despite precautions being taken to minimise the probability that the generated iPSCs would contain HIV-1 provirus, it was important to confirm the absence of provirus in all of the MACS iPSC lines. To do this, the optimised qPCR based HIV-1 detection assay, discussed in section **3.2.7**, was used (**figure 4.8**). As shown in **figure 4.8A**, HIV-1 provirus was not detected in any of the 50 MACS iPSC lines. RNaseP was used as a loading control.

For confirmation that this qPCR based HIV-1 provirus detection assay is able to detect the particular HIV-1 strain found in the MACS participant PBMC samples, gDNA from the depleted CD4+ cell fraction from some of the samples was analysed. **Figure 4.8B** shows the detection of the presence of HIV-1 provirus in the CD4+ cell fractions of donors 20032, 20318 and 20480. Although the detection is lower than the ACH-2 positive control, this is as expected as not all of the cells in the CD4+ fraction, <1%, would contain HIV-1 provirus (Coffin & Hughes, 2021).

To give further confidence in the negative HIV-1 provirus results for the MACS iPSCs, the ACH-2 cell line was mixed with uninfected HEK293T cells at different ratios from neat to 1:1,000,000. As shown in **figure 4.8C**, HIV-1 provirus was still detectable at ACH2:HEK293T = 1:1000. This demonstrated that the qPCR based assay can detect 1 cell containing an HIV-1 provirus for every 1000 uninfected cells. This detection threshold provides an appropriate level of leeway for its application due to the clonal nature of the iPSC lines. Hence, any iPSC line that was reprogrammed from an HIV-1 infected cell, would be made up completely of cells containing at least one HIV-1 provirus, and thus, would certainly be detected by this assay.

The undetectable HIV-1 provirus results for all of the MACS iPSCs, alongside the extra validations of the assays sensitivity give a very strong indication that the MACS iPSCs are free from HIV-1 provirus.

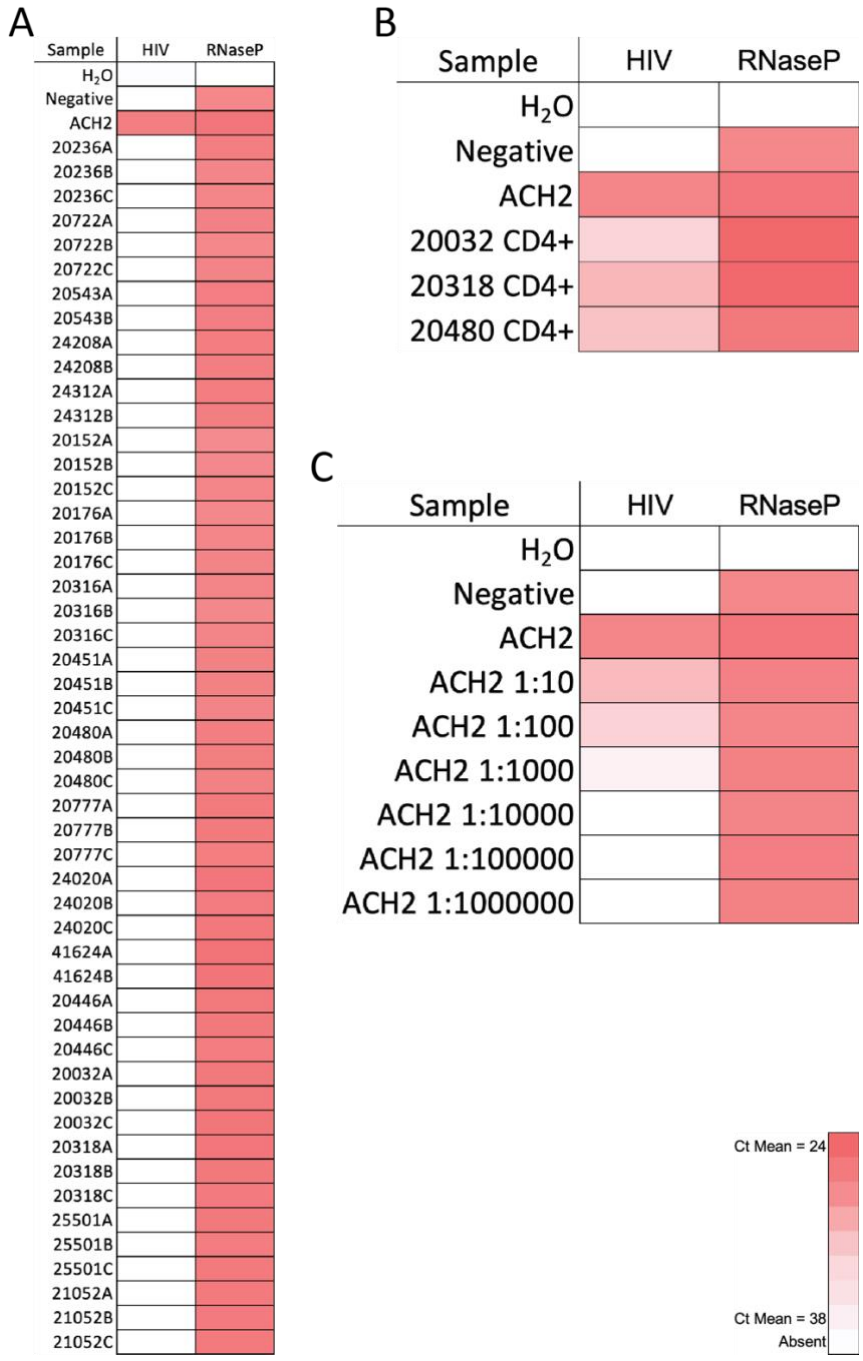


Figure 4.8: Detection by qPCR of HIV-1 provirus in the MACS iPSC lines and validation of the assay's sensitivity.

A) qPCR results for detection of the HIV-1 provirus in the MACS iPSC lines. **B)** qPCR results for detection of HIV-1 provirus in the CD4+ cell fraction of the MACS PBMC sample after double CD4+ cell depletion by MagACS. **C)** qPCR results for detection of HIV-1 provirus in an increasingly diluted population of cells known to contain an HIV-1 provirus. RNaseP = Loading control. As indicated in the heat map scale bar, the darker the red box equates to a lower mean Ct value and, therefore, a higher amount of the HIV-1 provirus present in the sample. White boxes indicate an undetected Ct mean value and, therefore, an absence of HIV-1 provirus. All results shown are a Ct mean result of three Ct values.

4.2.9 G-banded karyotypic analysis of the MACS iPSCs to detect chromosomal abnormalities

To ensure the genomic stability of the MACS iPSC lines has been maintained over *in vitro* culture, G-banded karyotyping was carried out to analyse chromosomal integrity. Due to time constraints, only 2 of the MACS iPSC lines have been analysed by this method. **Figure 4.9A** shows a representative karyotype, 1 of 20 analysed events, of MACS iPSC line 20236A at passage 20. The figure demonstrates a normal male diploid karyotype 46, XY. MACS iPSC line 20722A also exhibited the same karyotype phenotype (**supplementary file 4.3**). All remaining MACS iPSCs should be tested to ensure their chromosomal integrity.

A

Figure 4.9: G-banded karyotyping analysis of the MACS iPSC line 20236A.
A) Representative image of the karyotype for MACS iPSC line, 20236A, showing a normal male diploid karyotype 46,XY for 20 examined cells at passage 20.

4.3 Discussion

This chapter showed the reprogramming of samples of PBMCs from patients in the MACS into iPSCs. 50 iPSC lines were derived from 18 different MACS participants and quality control analyses were conducted on them to validate their pluripotency, Sendai viral vector negativity, HIV-1 provirus negativity, and chromosomal integrity.

The MACS patients whose PBMC samples were to be reprogrammed into iPSCs were chosen based upon their clinical phenotypes in response to being infected with HIV-1. The clinical phenotypes of the patients were defined by clinical outcome factors, such as the time to AIDS diagnosis, and molecular signs, such as CD4+ T-cell count and HIV-1 viral load. The patients towards the extremities of the clinical phenotype spectrum were selected as they displayed the most disparity in disease progression, and thus, posed the greatest potential of possessing host factors that strongly influence HIV-1 infection.

The MACS PBMC samples were analysed shortly after being thawed for their composition of different immune cell types. This was performed as a quality control step to test the cell composition of each of the PBMC samples. Although some of the samples showed proportions of CD8+ T-cells, CD4+ T-cells, NK cells, monocytes, and B-cells outside of the expected healthy donor range, this could largely be explained by the patients being HIV-1+ at the time of sample collection. The presence of HIV-1 infection will cause inflammatory responses that drive activation, recruitment and proliferation of immune cells, resulting in the distortion of immune cell proportions compared to a healthy uninfected individual. Interestingly, the proportion of CD4+ T-cells appeared to be low for most of the donors. This may also be explained by the HIV+ state of many of the donors. The CD4+ T-cells in an HIV-1 infected individual are depleted by several mechanisms: direct targeting and elimination by the virus; chronic inflammation leading to activation-induced cell death or apoptosis, and chain reactions of pyroptosis (Cooper et al., 2013; Doitsh et al., 2010; Vidya Vijayan et al., 2017; Yates et al., 2007). Nevertheless, all 18 MACS PBMC samples contained at least some of all the expected cell types, indicating a reasonable level of quality despite HIV-1 infection and cryopreservation for an extended period of time.

The 18 PBMC samples were reprogrammed into 50 iPSC lines, either 2 or 3 clonal lines from each donor, which were checked for their quality. The generated MACS iPSC lines all demonstrated excellent iPSC morphology and expressed high levels of the two analysed pluripotency markers, OCT-4 and NANOG, when analysed by flow cytometry. The expression of OCT-4 was also confirmed using immunofluorescence in 30 of the MACS iPSC lines. Further validation of pluripotency on 30 of the 50 MACS iPSC lines was completed via a trilineage differentiation assay. The ability of the 30 MACS iPSC lines to differentiate into the ectoderm, mesoderm and endoderm was confirmed by detection of Pax-6, brachyury and SOX-17 respectively after a short period of directed differentiation. Although expression of Pax-6 and brachyury were low, these markers are very specific for their respective lineages, and therefore, still confirm successful ectoderm and mesoderm differentiation respectively. The expression of SOX-17 was higher, however, not all of the cells in the differentiation culture presented expression. This was shown to most likely be a result of insufficient differentiation, as opposed to an inability to differentiate, due to the detection of persisting OCT-4 expression in some of the endoderm differentiated cells. Despite this assay being effective at proving pluripotency, it is not considered the most stringent assay available. The teratoma assay, where iPSCs are introduced into immunodeficient mice with the intention of forming tumours containing a diverse mixture of distinct tissues from all three germ layers, is widely accepted as the most rigorous test for confirming the pluripotency of human iPSCs (Damjanov, 2005). However, this method is time consuming, expensive and poses the ethical and technical issues of working with mice. Therefore, for this project, the trilineage differentiation assay sufficed as a method of confirming pluripotency. The PluriTest was developed with the idea of replacing the teratoma assay as an easy, open access bioinformatic-based alternative. Since RNA-seq data had already been gathered from 30 of the MACS iPSCs as part of the infectivity phenotyping screen process (to be discussed in **chapter 5**), the PluriTest analysis could be conducted at no extra expense. Unfortunately, the tested MACS iPSC lines, and a known iPSC line (CTRs), fell below the pluripotency empirical threshold. This, however, is most likely a result of the time point at which the cells were harvested for RNA-seq. As the primary purpose of the RNA-seq data was to analyse the iPSC's transcriptomes at the point of infection, the cells were harvested after being passaged as single cells ~16 hours before. This is a suboptimal time to analyse pluripotency as passaging of iPSCs can induce cell stress and apoptosis (Rivera

et al., 2020; Takahashi et al., 2022). Further, due to the single cell nature of the passage, a media supplement, RevitaCell, was used which also impacts the cell's transcriptome (Liao et al., 2007). As a result, it is anticipated that the MACS iPSC lines harvested at this time point would not present a pluripotency signature comparable to an iPSC line that had been harvested at an ideal time point for pluripotency, like the ones used to generate the data in the PluriTest model. Therefore, to effectively use the PluriTest to analyse the MACS iPSCs, the cells should be harvested for RNA-seq analysis when at peak condition during routine culture.

Several different delivery methods of the four 'Yamanaka factors' can be used for the generation of iPSCs such as lentiviruses, adenoviruses and episomal vectors (Karami et al., 2022). Here we utilised Sendai viral vectors to deliver the 'Yamanaka factors' due to their non-integrative nature and the need for only one round of infection. Although non-integrative, the Sendai viral vectors can persist in the cytoplasm of cells, so it is still important to ensure the removal of them from newly generated iPSCs. The elimination happens passively over passages, as with 7 of the 50 MACS iPSC lines, but often persistent Sendai viral vectors require the removal efficiency to be enhanced. To help aid the removal, the two persistent Sendai viral vectors from the CytoTune-iPS 2.0 Sendai Reprogramming Kit, KOS and c-Myc, both contain point mutations in their polymerase-related genes that increase their temperature sensitivity (Ban et al., 2011). By increasing the culture temperature of the MACS iPSCs to 38.5°C for 5 days (either once or twice), the remaining Sendai viral vectors were removed to undetectable levels by qPCR.

Since many of the MACS iPSCs were reprogrammed from HIV+ PBMCs, confirmation of the absence of any HIV-1 provirus in the cell lines was necessary to define them as HIV-1 negative. The optimised HIV-1 qPCR based detection assay was used to do this. Due to the high mutational capacity of HIV-1, primers targeting a highly conserved region within the LTR of the HIV-1 genome were used. The primers also contained ambiguous bases in an attempt to increase the array of HIV-1 variants the assay could detect. HIV-1 provirus was undetectable in all 50 of the MACS iPSC lines. To demonstrate the assay's ability to detect the specific HIV-1 strain that would be present in the MACS participant's cells (likely clade B due to their geographical location), gDNA was isolated from the depleted CD4+ fraction of cells from the MACS PBMC samples. HIV-1 provirus was detected in these samples, indicating the assay's

sensitivity to these particular HIV-1 strains, thus, granting further confidence in the HIV-1 provirus-free nature of the generated MACS iPSCs. The assay was also confirmed to be sensitive enough to detect 1 cell containing an HIV-1 provirus in 1000 uninfected cells, therefore, these clonal iPSCs are comfortably within the detection limit of the assay.

Only two MACS iPSC lines were analysed for their genomic integrity by G-banded karyotyping due to time constraints. Both lines exhibited a normal male diploid karyotype, 46, XY, in all 20 analysed events. All remaining MACS iPSC lines should have their karyotypic phenotype tested to ensure genetic abnormalities haven't occurred during reprogramming or culture.

In summary, results in this chapter report the derivation of 50 new iPSC lines from 18 donors from the MACS, as well as the quality control checks conducted on them. These MACS iPSCs can now be used for their intended function, in the HIV-1 permissivity phenotyping screen. Moreover, they could also prove a useful resource for many future scientific studies in the field of HIV.

Chapter 5: Permissivity to HIV-1 infection screen on the MACS iPSCs and the correlation of phenotypes to transcriptomes

5.1 Introduction

Following the successful generation and validation of the MACS iPSCs, the permissivity to HIV-1 infection screen could be conducted. iPSCs exhibit certain characteristics that make them a desirable cell type to conduct such a screen on. Firstly, it has been reported that due to the consistency in generation and maintenance of iPSC lines, differences between them are predominantly attributable to the genetic variation of the patient from which they were derived (Vigilante et al., 2019). This, along with their ability to self-renew, makes iPSCs an excellent resource to utilise for the permissivity to HIV-1 infection screen and has the potential to result in more reproducible phenotypes when compared to the use of other cell types, such as PBMCs from a patient's peripheral blood. iPSCs also demonstrate inherent high levels of expression of numerous ISGs typically only stimulated in specialised cells during viral infection after the production of interferon (Wu et al., 2018). This could mean any undiscovered ISGs that interact with HIV-1 would more readily present themselves. As a result of this ISG expression, iPSCs are in a heightened anti-viral state and are typically ~10-40X less infectible than the permissive immortalised cell line HEK293Ts. Further, by exploiting the natural variation between iPSC lines, there is no need to knockdown genes and thus, all expressed genes or pathways in a cell can be studied. Previous studies have been restricted by this as utilising knockouts limits the ability to study genes vital for cell viability. These attributes make iPSCs a valuable asset for examining variations in permissivity to HIV-1 infection in order to uncover novel host factors that influence the rate of HIV-1 replication.

The permissivity to HIV-1 infection screen was employed to identify MACS iPSC lines with extreme opposite permissivity phenotypes to infection by HIV-1 (screen pipeline described in section **2.4.2**). The screen consisted of two separate assays run in parallel: an assay to assess infectivity and an assay to assess late gene expression. The infectivity assay utilised a VSV-G pseudotyped HIV-1 based GFP vector and assessed early events in HIV-1's life cycle, up to and including integration. For this, each iPSC line was challenged with five 1 in 4 serial dilutions of this vector and the percentage of cells positive for GFP expression after 48 hours was measured via high-throughput microscopy and analysis using an Operetta CLS. The percentage of GFP-positive cells identified in a sample equates to the percentage infection for that iPSC

line at that vector volume. The late gene expression assay employed a VSV-G pseudotyped full length HIV-1, strain NL4-3, and assessed later stages in the virus's life cycle, up to production of the viral proteins. Cells from each iPSC line were challenged with five 1 in 2 serial dilutions of virus and 24 hours later the input virus was removed and replaced with fresh media. After a further 24 hours, the cells were measured for the percentage of cells positive for Gag expression, also via high-throughput microscopy and analysis using an Operetta CLS. The percentage of cells positive for Gag expression is referred to as the percentage infection for that cell line at that virus volume. These two data outputs, one from the infectivity assay and one from the late gene expression assay, can then be used to assign a permissivity to HIV-1 infection phenotype to each MACS iPSC line. Comparative transcriptomic analysis of the iPSC lines exhibiting opposite extreme permissivity phenotypes could then be conducted in an attempt to identify host factors responsible for the differing phenotypes. Importantly, cells were harvested for RNA-seq analysis immediately before being infected. This meant that the RNA-seq data would relate as closely as possible to that cell line's transcriptome at the exact time point of infection, and would not, therefore, monitor changes induced by infection.

Due to practicality and the labour intensiveness of the infectivity and late gene expression assays, just 11 iPSC lines could be tested each week. In an attempt to account for experimental variation between the screening assays conducted on different weeks, a well-defined and consistent iPSC line, the CTRs, was used in every week as a method for normalisation. This allowed ten MACS iPSC lines to be examined per week, which could be normalised against the CTRs ran in that same week. Each week's experiments are referred to as a separate screen and the cell line composition of each screen is shown in **table 5.1**. Within each screen, each cell line was tested in duplicates, referred to as replicates 1 and 2. Each MACS iPSC line was also run in two consecutive screens, resulting in a total of four replicates per line.

Table 5.1: MACS iPSC lines tested, along with the CTR iPSC line, in each of the screens.

Screens 6 + 7	Screens 8 + 9	Screens 10 + 11
20176A	20176C	20032A
20176B	20480C	20032B
20480A	20543A	20236A
20480B	20543B	20236B
20777A	20777C	20236C
20777B	24020C	20318A
24020A	24208A	20318B
24020B	24208B	20318C
41624A	24312A	20722A
41624B	24312B	20722B

5.2 Results

5.2.1 Set-up of the Operetta CLS imaging and analysis pipelines

Percentage infection for both the infectivity and late gene expression assays was analysed and calculated using an Operetta CLS (**figure 5.1**). To demonstrate the process by which the Operetta conducts this, an example field of image from both the infectivity and late gene expression assays are displayed in **figure 5.1A** and **figure 5.1B** respectively. After completion of the infection pipeline for both assays, the cells were fixed and stained for DAPI and F-actin. Additionally, the vector transduced cells, from the infectivity assay, were stained with an anti-GFP antibody (used to increase the GFP signal, making GFP-positive and GFP-negative cells more distinct), and the virus infected cells, from the late gene expression assay, were stained with an anti-HIV-1 Gag antibody that detects Gag. The cells were then imaged by an Operetta CLS using a 20X water lens. After imaging, a multi-step algorithm on the Operetta was used to calculate the percentage of infected cells present in each sample. Firstly, the Operetta uses the intensity, shape and size of the DAPI staining to identify the nuclei present in the sample. Next, the F-actin staining supports the Operetta in identifying the area of cytoplasm that corresponds to each nucleus. Finally, the Operetta was trained on a sample population for which cells are positive or negative for the expression of either GFP or Gag. Using machine learning, the Operetta then sets a threshold value that differentiates positive and negative cells and applies this to all fields imaged. The number of positive cells and the number of total cells are then used to generate a percentage infection for that sample. The Operetta conducts this algorithm in a consistent high-throughput manner for all samples, from both assays, across all screens. The full detailed version of the Operetta algorithms used for the infectivity assay and the late gene expression assay are attached to this thesis (**Supplementary file 2.1 and 2.2 respectively**).

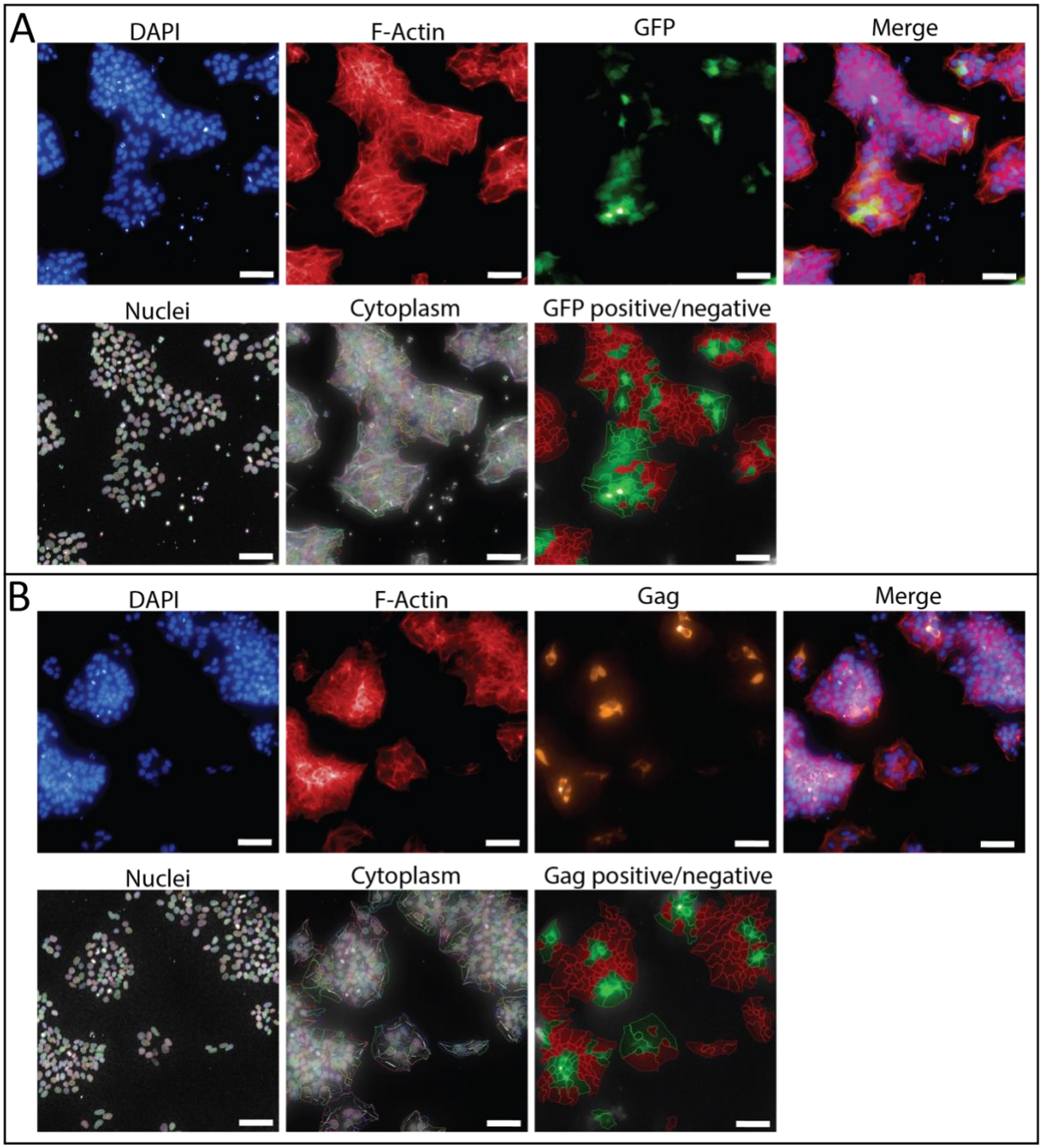


Figure 5.1: Operetta analysis pipeline used to calculate percentage infections for the infectivity and late gene expression assays.

A+B) A one field example of the pipeline used to calculate percentage infections in the infectivity and late gene expression assays respectively. DAPI staining is utilised to identify the nuclei present in the image. F-actin staining, paired with nuclei identification, allows for the identification of each cell cytoplasm. GFP or Gag expression is then detected and used to decipher which cells have been infected by the vector or virus respectively. Based upon this, a percentage infection is then calculated by the Operetta. Images were taken at 20X magnification and scale bars represent 100μm.

5.2.2 Examination and cleaning of the infection percentage data produced by the infectivity and late gene expression assays

The infectivity assay on the iPSC lines utilised five 1 in 4 serial dilutions of the VSV-G pseudotyped HIV-1-based GFP vector (**figure 5.2**). These vector volumes were previously optimised to produce infection percentages in 30,000 iPSCs that captured the linear part of the infection curve. The highest vector volume utilised 0.206µl of vector, the second used 0.0515µl, the third used 0.0129µl, the fourth used 0.00322µl and the fifth used 0.000805µl. A sixth well, where no vector was added, was used as a negative control for infection.

When conducting the late gene expression assay, each iPSC line was infected with 5 different volumes of a VSV-G pseudotyped full length HIV-1 virus, strain NL4-3, at a 1 in 2 serial dilution. These virus volumes had also been previously optimised to achieve infection percentages in 30,000 iPSCs that corresponded to the linear phase of the infection curve. The virus quantities used were: 0.015µl for the highest volume of virus, 0.0075µl for the second, 0.00375µl for the third, 0.00188µl for the fourth, and 0.00094µl for the fifth. As with the vector set-up, a sixth well, where no virus was added, served as a negative control for infection.

The percentage infection for each iPSC line was measured as the percentage of cells that were GFP or Gag-positive in the infectivity assay or late gene expression assay respectively. The percentage infections when infecting with the vector, produced a sigmoidal curve when plotted against vector volume on a $\log_{10}$ scale. **Figure 5.2A** shows the percentage infection data of a representative MACS iPSC line, 20722A, when infected with differing volumes of vector in the infectivity assay. As vector volume increases at the lower amounts, the infection percentage increases slowly demonstrated by a more gently sloped phase of the curve. As vector volume increases further, a linear relationship between vector volume and percentage infection is captured, shown by a steep section of the curve. As infection percentages start to saturate at even higher vector volumes, the curve begins to plateau at ~75-100% depending on the iPSC line. When infected with virus, during the late gene expression assay, the percentage infection of an iPSC line demonstrates a more linear relationship with virus volume when plotted on a $\log_{10}$ scale. The infection percentages of the same representative MACS iPSC line, 20722A, when infected with different volumes of virus is displayed in **figure 5.2B**. As opposed to the infectivity assay data, a linear relationship is seen between percentage

infection and virus volume even at low amounts of virus. Similarly to the vector data, the curve begins to plateau at the higher virus volumes as infection percentages saturate. Interestingly, the infection percentages in the late gene expression assay saturate at lower levels, ~25-60% depending on the iPSC line, than in the infectivity assay.

For both the infectivity and late gene expression assays, as percentage infection reaches its maximum value, the iPSCs, depending on the cell line, may succumb to cytotoxic effects of the vector/virus, resulting in increased cell death and sometimes a decrease in percentage infection. An example MACS iPSC line, 24208A, displays a sigmoidal infection curve exhibiting this decrease in infection percentage when infected with the higher amounts of vector (**figure 5.2C**). To avoid the effects of this cell death having an impact on calculations used to determine a cell line's permissivity to HIV-1 phenotype, the data collected from the screens were cleaned. This cleaning process involved removal of a data point if its infection percentage was lower compared to the preceding vector/virus volume's infection percentage (**figure 5.2D**). Additionally, further cleaning was required for the late gene expression assay data due to an increased amount of cell death observed as a result of viral infection compared to vector infection. Fewer cells survived viral infection, and therefore, less cells were present for analysis by the Operetta pipeline. If the Operetta analysed fewer than 100 cells in a sample, that sample was also excluded as a data point and not processed in the models used to determine the late gene expression phenotypes of the iPSC lines.

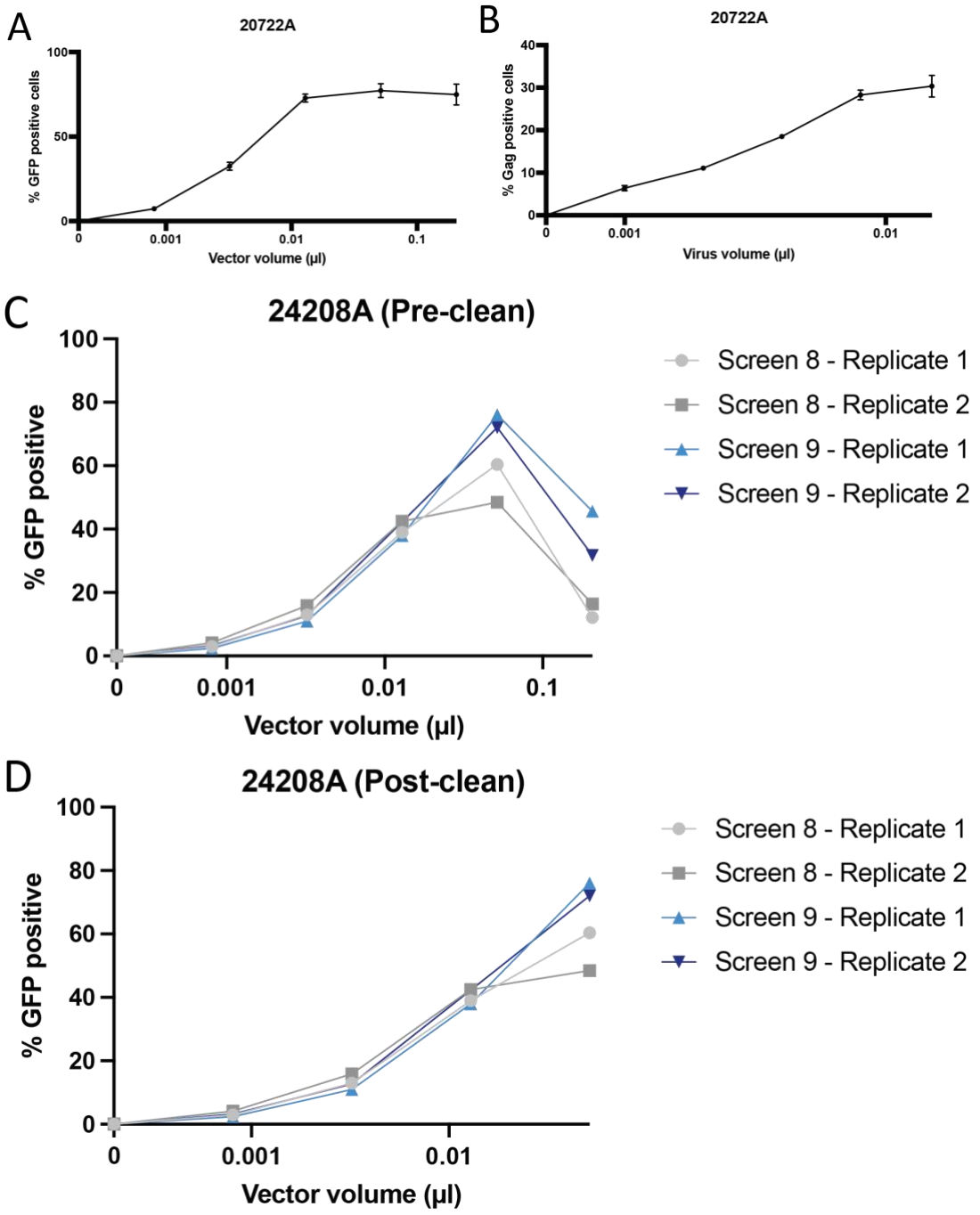


Figure 5.2: Representative data of the infection curves generated from the infectivity and late gene expression assays.
A+B) Vector (VSV-G pseudotyped HIV-1 based GFP vector) and virus (VSV-G pseudotyped full length HIV-1, strain NL4-3) infection data of an example MACS iPSC line, 20722A, tested in the infectivity and late gene expression phenotyping assays respectively. Error bars represent standard error of the mean (SEM). **C+D)** Vector infection data of an example MACS iPSC line, 24208A, tested in the infectivity assay before and after data cleaning respectively. The light grey and dark grey lines represent the two replicates from the first time 24208A was run through the screen pipeline (screen 8). The light blue and dark blue lines represent the two replicates from the second time 24208A was run through the screen pipeline (screen 9).

5.2.3 Analysis of the permissivity of the MACS iPSCs to the VSV-G pseudotyped HIV-1 based GFP vector using a linear regression model

After the data produced by the infectivity assay were cleaned, the data were analysed by two models to determine the infectibility phenotypes of the iPSC lines to the VSV-G pseudotyped HIV-1-based GFP vector. The first analysis method used was a linear regression model which assumed a linear relationship between vector volume and infection percentage (**figure 5.3**). The second analysis method utilised a logistic curve fitting model and will be discussed in section **5.2.4**. As mentioned above, the sigmoidal curve generated by the infection of an iPSC line with a titration of vector is only linear for a particular phase of the curve. To select for the linear part of the curve in a consistent way across all tested iPSC lines, only infection percentage values falling between 2-60% were included in the calculation to determine the infectibility of an iPSC line. The values fitting this criteria were then inputted into the following calculation to generate a vector titre for each of the tested iPSC lines:

Vector titre (IFU/µl) = Number of cells * ((Percentage infection/100) / Vector volume)

The vector titres for each iPSC line generated at the different volume of vector were averaged to give a vector titre for that cell line in that replicate. Each cell line had 2-4 replicates. The average of these replicates for each of the tested iPSC lines were subsequently ordered, left to right, from least infectible to most infectible (**figure 5.3A**). Replicates from within each screen are represented by data points of the same shape. Circle data points represent average titres generated from the first screen involving that particular cell line (either screen 6, 8 or 10) and square data points represent average titres generated in the second screen involving that particular cell line (either screen 7, 9 or 11). The replicates generated from within each screen show high levels of consistency, whereas greater variability is observed between the replicates conducted across screens. By utilising **table 5.1** to identify which MACS iPSC lines were tested in each screen, the cell line order appears to cluster based upon screen number, with the cell lines run in screens 6 + 7 exhibiting the highest vector titres. The cause of this seems to be the circle data points (collected in screen 6) having far higher titres than the rest of the data points. Therefore, variation in the replicates of each cell line between the screens seems to have a big effect on vector titres. The high titres detected in screen 6 does, however, correlate with the CTRs analysed in screen 6 also having a high vector titre compared to the

CTRs analysed in the other screens. Thus, normalising each of the MACS iPSC line's vector titres to the CTR's vector titre of that screen should help alleviate some this variation.

The vector titres of the iPSC lines normalised to the CTRs within each screen are shown in **figure 5.3B**. The cell lines are ordered left to right from least to most infectible. After this normalisation, the order the MACS iPSC lines now seems to be less clustered based on screen number. The fold difference between the least infectible iPSC line and the most infectible iPSC lines are 4.5 or 3.3 when including or excluding the CTR iPSC line respectively. Interestingly, some of the different clones from the same donors (e.g. 20176A, B and C) group well. However, other donor's clones (e.g. 20722A and B) do not cluster together. Also displayed in **figure 5.3B**, indicated by the colour of the circle below each MACS iPSC line's name, is the clinical phenotypes of the donors from which each MACS iPSC line was derived. If the infectibility of the MACS iPSC lines correlated perfectly with the clinical phenotype of their donor, it would be expected to see the red circles, resembling RPs, grouped towards the right hand side of the graph. Despite the most infectible iPSC line being derived from a RP, the other iPSC lines derived from a RP donor are scattered throughout the graph, and two are even situated near the hardest to infect extreme. Unfortunately, this seems to indicate that the infectivity phenotype of the MACS iPSC lines is not consistent with the clinical phenotype of the donor they were generated from. The inconsistent clustering between iPSC clones from the same donor and the lack of correlation to the donor's clinical phenotypes may be attributable to the potential change undergone by the cells during the reprogramming process from PBMCs to iPSCs. Also, since PBMCs are a mixed population of cells, the generated iPSC lines may have originated from varying cell types within the PBMC population, resulting in varying susceptibilities to HIV-1 infection.

Analysing the infection percentage data from the infectivity assay by a linear regression model results in the three iPSC lines most resistant to transduction by the VSV-G pseudotyped HIV-1 based GFP vector being 24208B, 20318A, and 20032B and the three cell lines most permissive to the vector being 20543B, 20722A, and 20543A.

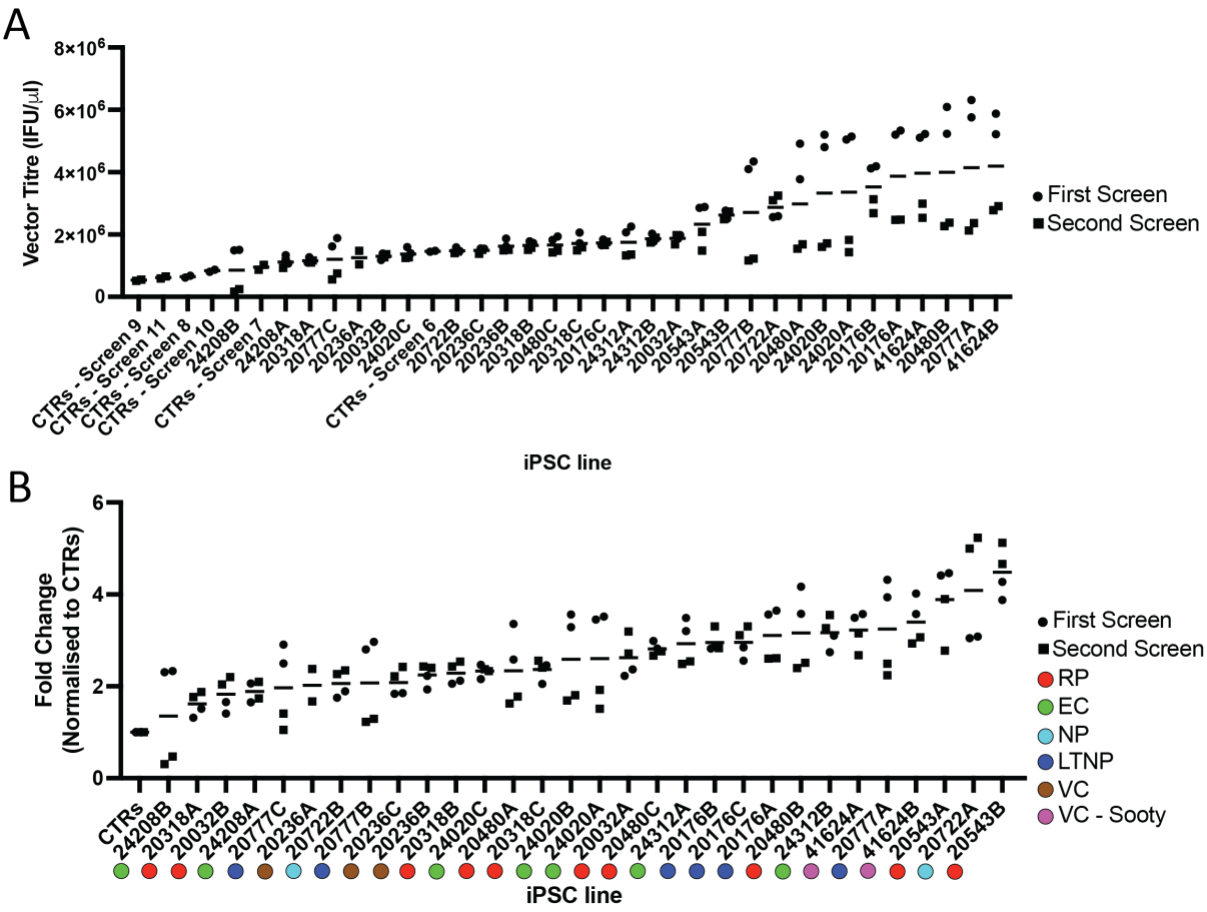


Figure 5.3: Vector infection data from the infectivity assay on the MACS iPSCs analysed by a linear regression model.

A) iPSC lines ordered left to right by vector titre from least permissive to most permissive. **B)** iPSC lines ordered left to right by fold change when normalised to the vector titre of the CTRs within the corresponding screen from least permissive to most permissive. The line represents the mean value for that cell line. Circle and square data points represent vector titres calculated from the first and second screen of that cell line respectively. Coloured circles correspond to the clinical phenotype of the donor from which that cell line was derived (Red = RP, Green = EC, Cyan = NP, Blue = LTNP, Brown = VC, Pink = VC - Sooty).

5.2.4 Analysis of the permissivity of the MACS iPSCs to the VSV-G pseudotyped HIV-1 based GFP vector using a logistic curve fitting model

The second model used to analyse the cleaned vector data produced by the infectivity assay was a logistic curve fitting (**figure 5.4**). The logistic curve fitting was used for this analysis as it typically forms an 'S' shaped curve, which begins with a gentle incline before gradually increasing its steepness, reaching a period of steep incline, and finally levelling off as it reaches the maximum value. This curve matches the sigmoidal shape of the vector infection curves generated by this screen pipeline, as discussed in section **5.2.2**, making it a good model to use to analyse this data. The logistic function is:

$$f(x) = a/(1+e^{-b(x-c)})$$

The 'b' value in the logistic function is of most interest to this study as it represents the speed of curve transition from the gentle to steep incline as well as the steepness of the curve. For the context of this infectivity phenotyping screen, a higher b-value represents a greater change in percentage infections over the same change in vector volume. Therefore, an iPSC line with a higher b-value is considered more permissive to HIV-1 infection than an iPSC line with a lower b-value. The b-values calculated as a result of logistic curve fitting of the vector data collected in screen 11 are shown in **figure 5.4A**. Within screen 11, the least and most permissive iPSC lines to infection by the VSV-G pseudotyped HIV-1-based GFP vector are 20318A (not including the CTRs) and 20236A respectively.

Due to experimental variation between the weekly screens, normalisation is required to effectively compare iPSC lines run in the different screens. To complete this normalisation for the logistic curve fitting analysis of the vector data, the z-score statistical measurement of each screen was employed. The z-score represents how a specific data point differs from the mean value of the screen it was generated in using standard deviations. The mean b-value of a screen is subtracted from each iPSC line's b-value within that screen, which is then divided by the standard deviation of the b-values of that screen. This scaling allows for the comparison of b-values between different screens. The vector infectibility data of all iPSC lines before and after z-score standardisation are displayed in **figure 5.4B+C** respectively. Scaling results in highly variable b-values (ranging from ~50 to ~1200) being standardised to b-values

normalised to z-scores spanning around zero (ranging from ~-2 to ~3). Before scaling, screens 7 and 9 exhibit consistently low b-values compared to the other screens, which is most likely a result of experimental variation in those screens as opposed to these cell lines actually exhibiting lower b-values. Therefore, scaling allows normalisation for this variation. After standardisation, screens 7 and 9 display b-values normalised to z-score that appear much more consistent to the other screens.

After scaling of the data, the b-values of all the iPSC lines across the vector infection screens can be compared. **Figure 5.4D** shows the infectibility of the iPSC lines, ordered left to right, from least permissive to most permissive to infection by a VSV-G pseudotyped HIV-1-based GFP vector. As with the linear regression analysis model results, some of the different clones from a single donor cluster together well whereas other do not, and the clinical phenotypes of the donors from which the MACS iPSC lines were derived don't seem to correlate to the cell line's infectivity phenotype.

Fitting the infectivity assay data to a logistic curve model and normalising to the z-score of each screen results in the three MACS iPSC lines most resistant to infection being 24208B, 20318A, and 20777B, and the three MACS iPSC lines most permissive to infection being 20236A, 41624B and 20722A.

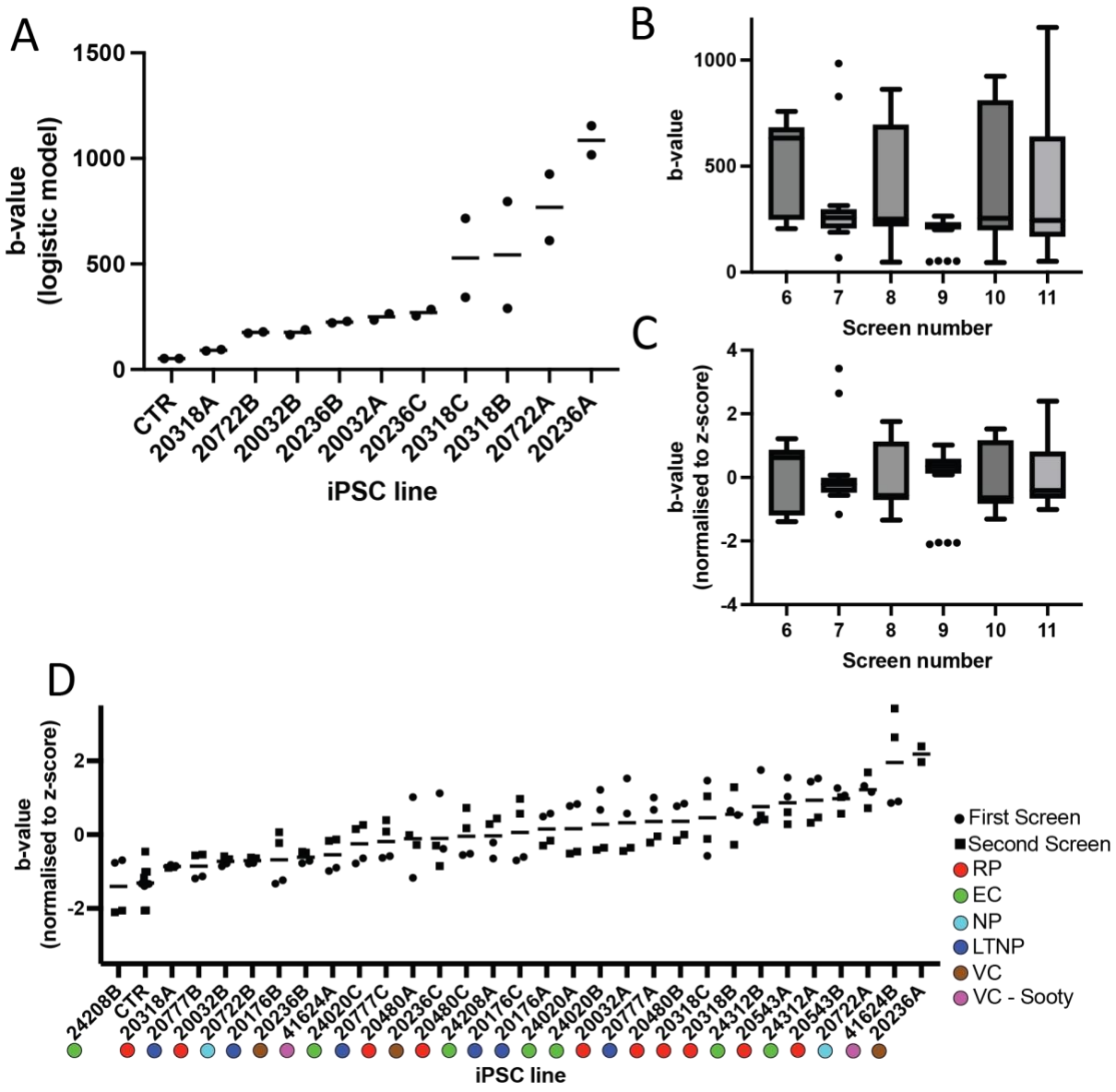


Figure 5.4: B-values calculated by logistic curve fitting of the percentage infection data from the infectivity assay.

A) B-values for the vector infectivity of the iPSC lines run in screen 11. The line represents the mean value for that cell line. Data points represent replicates 1 and 2 of each iPSC line in screen 11. **B+C)** B-values of all iPSC lines from all screens displayed as box and whisker plots before and after scaling of the data using z-score normalisation respectively. **D)** iPSC lines ordered by b-value normalised to the z-score from least permissive to most permissive. The line represents the mean value for that cell line. Circle and square data points represent vector titres calculated from the first and second screen of that cell line respectively. Coloured circles correspond to the clinical phenotype of the donor from which that cell line was derived (Red = RP, Green = EC, Cyan = NP, Blue = LTNP, Brown = VC, Pink = VC - Sooty).

5.2.5 Interpretation of the linear regression and logistic curve fitting model analyses of the iPSC infectivity phenotyping assay data

The ranks of the 5 MACS iPSC lines exhibiting the most resistant and most permissive infectivity phenotypes to the VSV-G pseudotyped HIV-1-based GFP vector when analysed by both analysis models are shown in **table 5.2** and **table 5.3** respectively. The level of correlation seen between the two analysis models, in terms of an iPSC line's resistance to HIV-1 infection ranking, is displayed in **figure 5.5A**. An R^2 value of 0.4655, after the fitting of a simple linear regression, demonstrates a moderate correlation. As a result, the rank of a MACS iPSC line determined by the linear regression model was multiplied by its rank determined by the logistic curve fitting model. This generated a collective rank for each cell line, thus, favouring the MACS iPSC lines that exhibit extreme infectivity phenotypes in both analysis models.

The most resistant MACS iPSC line, with the lowest collective rank, for the infectivity assay was 24208B. This cell line was analysed to be rank 1 most resistant MACS iPSC line by both the linear regression model and logistic curve fitting model. However, this cell line only had one RNA-seq sample, so any comparison of its transcriptome to another cell line would be made with low statistical power. Therefore, the MACS iPSC line with the next lowest collective rank was chosen, 20318A. 20318A was identified to be the rank 2 most resistant MACS iPSC line by both analysis models. This iPSC line had 3 RNA-seq samples associated with it, 2 from one screen and 1 from the other.

The most permissive MACS iPSC line was determined to be 20543B. This cell line was identified as rank 1 most permissive by the linear regression model and rank 4 most permissive by the logistic curve fitting model. This gave MACS iPSC line 20543B a collective rank for the infectivity assay of 4. 20543B also had 3 RNA-seq samples associated with it, thus, making it a good candidate to have its transcriptome compared to.

The comparison of the infectivity data between these MACS iPSC lines of interest, 20318A and 20543B, when analysed by the linear regression model and the logistic curve fitting model are displayed in **figure 5.5B** and **figure 5.5C** respectively. The comparison of the transcriptomes of these MACS iPSC lines will be further discussed in section **5.2.10**.

Table 5.2: The top 5 most resistant MACS iPSC lines from the infectivity assay when analysed by either the linear regression or logistic curve fitting model.

Rank	Linear Regression	Logistic
1	24208B	24208B
2	20318A	20318A
3	20032B	20777B
4	24208A	20032B
5	20777C	20722B

Table 5.3: The top 5 most permissive MACS iPSC lines from the infectivity assay when analysed by either the linear regression or logistic curve fitting model.

Rank	Linear Regression	Logistic
1	20543B	20236A
2	20722A	41624B
3	20543A	20722A
4	41624B	20543B
5	20777A	24312A

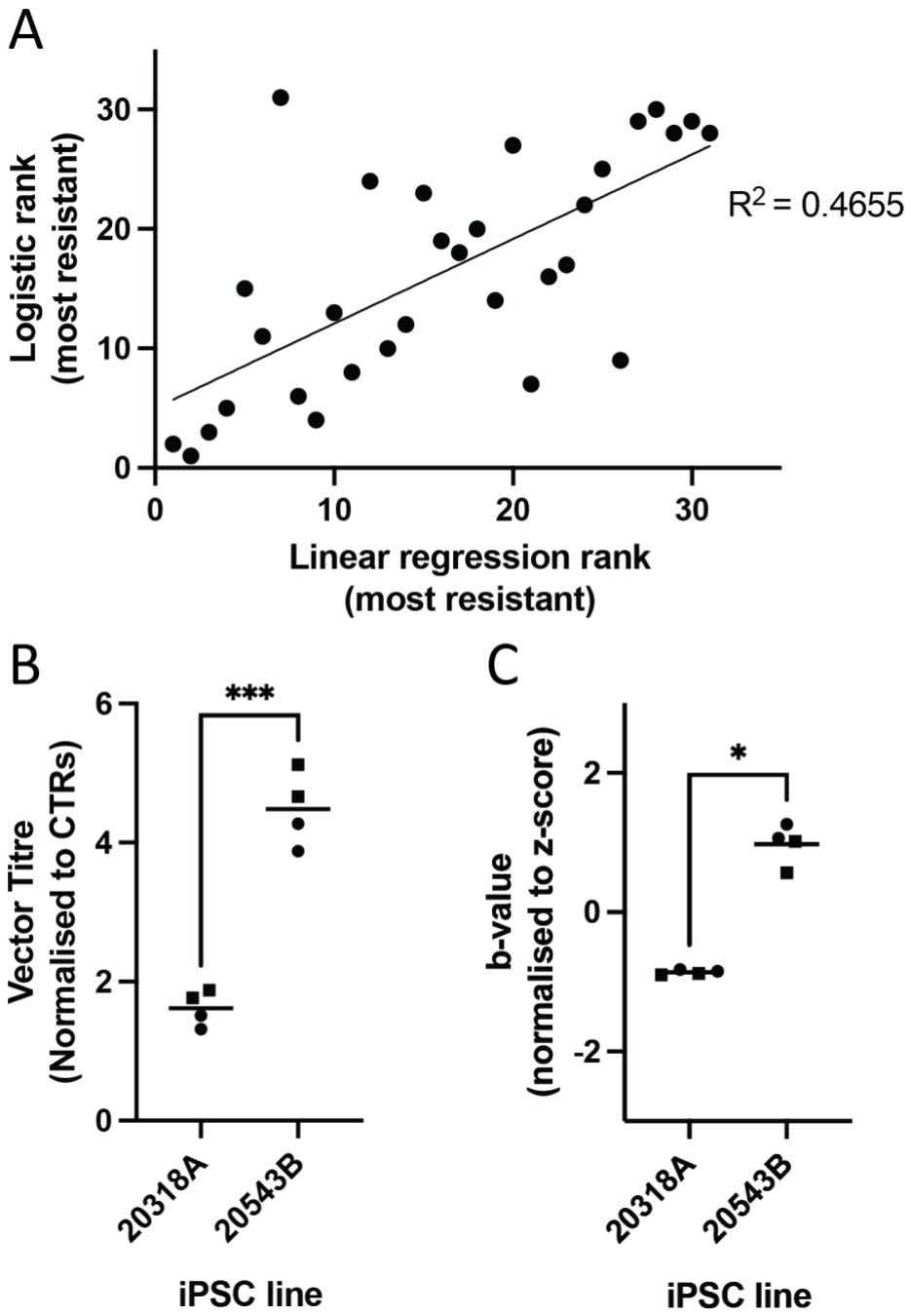


Figure 5.5: Interpretation of the infectivity assay data when analysed by the linear regression and logistic curve fitting models.

A) Plot displaying the correlation between both infectivity assay analysis models in terms of all iPSC line's resistance to HIV-1 rank. A simple linear regression is fitted, and the R^2 value is displayed. **B)** Comparison of the second most resistant MACS iPSC line, 20318A, to the most permissive MACS iPSC line, 20543B, after analysis by the linear regression model and normalisation to the CTR iPSC line. Statistical test is unpaired t test with Welch's correction (*** = $p<0.0005$). **C)** Comparison of the second most resistant MACS iPSC line, 20318A, to the most permissive MACS iPSC line, 20543B, after analysis by the logistic curve fitting model and normalisation to the z-score. Statistical test is Mann Whitney U test (* = $p<0.05$).

5.2.6 Analysis of the permissivity of the MACS iPSCs to the VSV-G pseudotyped full length HIV-1, strain NL4-3, using a linear regression model

Post cleaning, the infection percentage data generated by the late gene expression assay of the permissivity to HIV-1 infection screen was analysed by two models to produce the late gene expression phenotypes of the MACS iPSCs. As with the infectivity assay data, the first model used was linear regression (**figure 5.6**). This required a linear relationship between the infection percentage of the iPSCs by the virus and the input virus volume. As discussed in section **5.2.2**, the relationship between these two variables appears to be linear at low and medium virus volumes before plateauing at the higher virus volumes. As a result, any infection percentages between 1-20% were used in the below calculation to generate the virus titre for that iPSC line at that virus volume.

Virus titre (IFU/µl) = Number of cells * ((Percentage infection/100) / Virus volume)

The virus titres from within that replicate of virus serial dilutions were averaged. Each MACS iPSC line was infected with 2-4 replicates, split across two screens, resulting in 2-4 virus titres. The iPSC lines were consequently ordered left to right based upon the average of these virus titres from most resistant to most permissive to the virus (**figure 5.6A**). The replicates produced within a screen, represented by matching shapes for an iPSC line, displayed excellent consistency, whereas, the replicates between screens showed greater variation. Similarly to the infectivity assay results, when analysed by a linear regression model before normalisation to the CTRs, the 8 iPSC lines most permissive to infection in the late gene expression assay were lines tested in screens 6 and 7. This again correlated with the virus titres observed in the CTRs being highest in screen 6 and 7 compared to the other screens. Therefore, normalisation of an iPSC line's virus titre to the virus titre of the CTRs in that screen appeared to be a good way to standardise the data between screens for more reliable comparison.

After normalisation to the CTRs within each screen, the cell lines were ordered left to right from least to most permissive to infection by the VSV-G pseudotyped full length HIV-1 virus (**figure 5.6B**). As a result of normalisation, the iPSC lines grouped less according to screen number. The most permissive iPSC line, 20722A, was 10.35-fold more permissive to HIV-1

infection than the CTR line and 5.06-fold more permissive to HIV-1 infection than the least permissive MACS iPSC line, 20318A. Similarly to the infectivity assay results, some clones from the same donor cluster closely (41624A and B), whereas other clones from the same donor do not (20722A and B). The clinical phenotypes of the donors from which the iPSC lines were generated, once again, appear not to have an effect on the permissivity to HIV-1 infection phenotype of the MACS iPSC lines. The MACS iPSCs generated from the RPs, indicated by the red coloured circles, display permissivity phenotypes at both ends of the spectrum.

Analysing the late gene expression assay data using a linear regression model and normalisation to the CTRs within each screen resulted in 20722A, 20236A, and 20236C being identified as the most permissive cell lines to infection by HIV-1 and 20318A, 24208B, and 20777B being identified as the most resistant to infection by HIV-1.

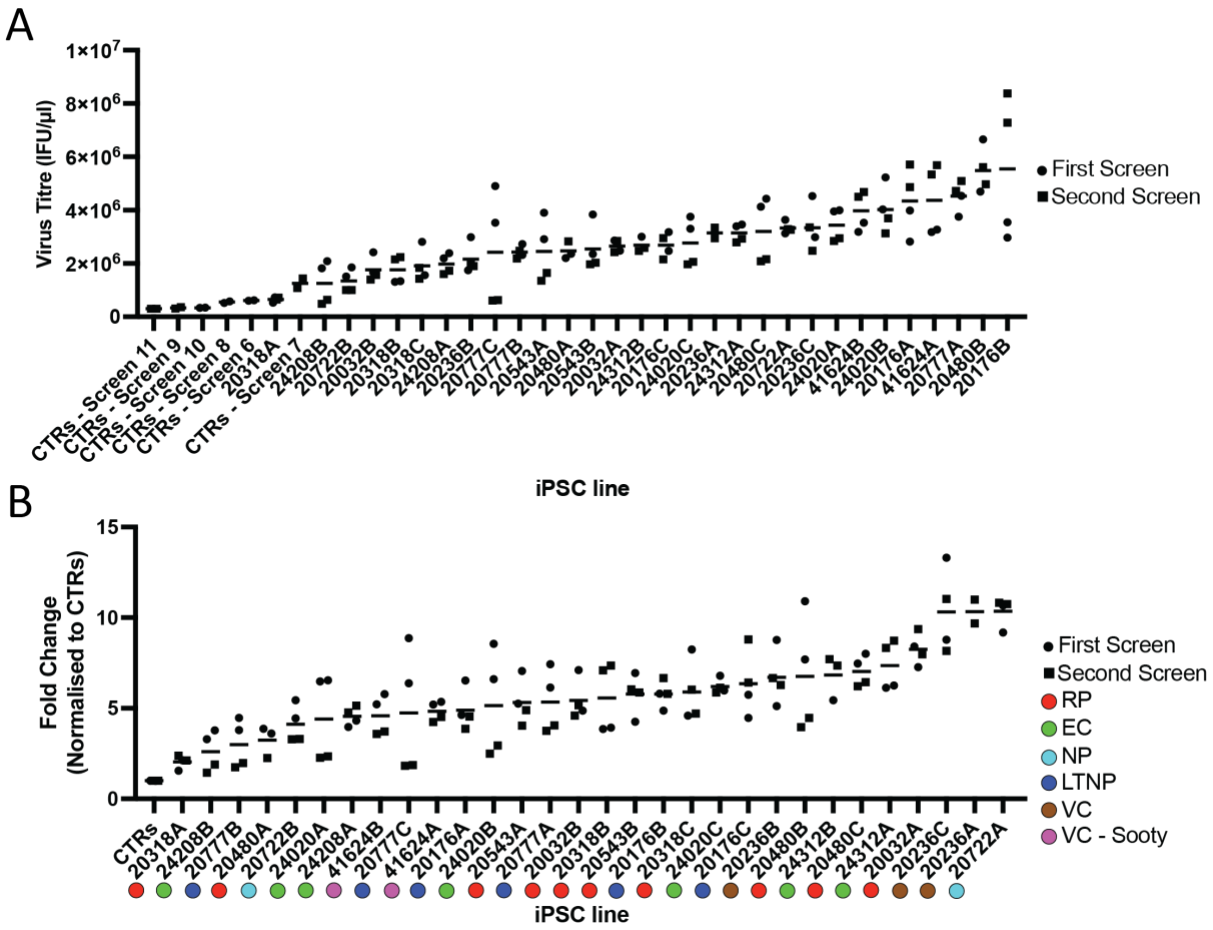


Figure 5.6: Virus infection data from the late gene expression assay on the MACS iPSCs analysed by a linear regression model.

A) iPSC lines ordered left to right by average virus titre from least permissive to most permissive. **B)** iPSC lines ordered left to right by fold change when normalised to the virus titre of the CTRs within the corresponding screen from least permissive to most permissive. The line represents the mean value for that cell line. Circle and square data points represent vector titres calculated from the first and second screen of that cell line respectively. Coloured circles correspond to the clinical phenotype of the donor from which that cell line was derived (Red = RP, Green = EC, Cyan = NP, Blue = LTNP, Brown = VC, Pink = VC - Sooty).

5.2.7 Analysis of the permissivity of the MACS iPSCs to the VSV-G pseudotyped full length HIV-1, strain NL4-3, by normalising to virus volume

The second method of analysis used on the cleaned data produced by the late gene expression assay was conducted by normalising to virus volume using z-score normalisation (**figure 5.7**). Since the linear regression model only utilised the data points within the linear part of the infection curve in a consistent way for all iPSC lines, some potentially relevant data points, such as the point at which the curve began to plateau, were overlooked. This point of plateau, representing the maximum percentage infection of an iPSC line, may prove an interesting characteristic of that cell line regarding its permissivity to HIV-1. For example, an iPSC line may exhibit low increases in infection percentages with the addition of increasing amounts of virus, but, may reach a higher maximum infection percentage compared to another iPSC line. Using the normalising to virus volume model means all data points of that iPSC line are used in the assessment of its relative permissivity to HIV-1 infection.

As mentioned previously, the cleaned late gene expression assay data demonstrates a linear relationship between percentage infection and virus volume, until a saturation in percentage infection is reached at the higher virus volumes. Demonstrating this relationship, **figure 5.7A** plots the infection percentage of every iPSC line at the different virus volumes. To utilise all of these data points by normalisation by virus volume, the z-score statistical measurement was used. The virus volume normalised value of an iPSC line at a particular virus volume was calculated by subtracting the mean percentage infection value of all iPSC lines within the same screen at that virus volume from the percentage infection of that iPSC line at that virus volume. This was then divided by the standard deviation of all iPSC lines within the same screen at that virus volume. Scaling of the data in this way allows for the direct comparison of virus volume normalised values across different virus volumes (**figure 5.7B**). As a result, virus volume could be excluded as a variable, thus, within each screen an iPSC line had 10 data points, consisting of 2 replicates for each of the 5 virus volumes (if none were removed during data cleaning). Since scaling occurs in a consistent manner for all virus volumes for all screens, this normalised data can also be compared across screens, giving each iPSC line a potential 20 data points. The average of these virus volume normalised values for each iPSC line can then be calculated to give one value representing the permissivity of that cell line to HIV-1 infection. **Figure 5.7C** demonstrates the comparison of this average value for all iPSC lines tested by the

late gene expression assay. Again, the clinical phenotypes of the patients from which the MACS iPSC lines were derived didn't correlate well with the permissivity phenotypes of the iPSC line as the different clinical phenotypes were distributed throughout the graph with minimal grouping observed. Also consistent with earlier, some iPSC clones from the same donor cluster closely whereas others don't.

By this method of analysis, the three cell lines most permissive to infection by the VSV-G pseudotyped full length HIV-1 virus were: 20480B, 20236A, and 20236C, and the three cell lines most resistant to infection were: 20318A, 24208B, and 20777B.

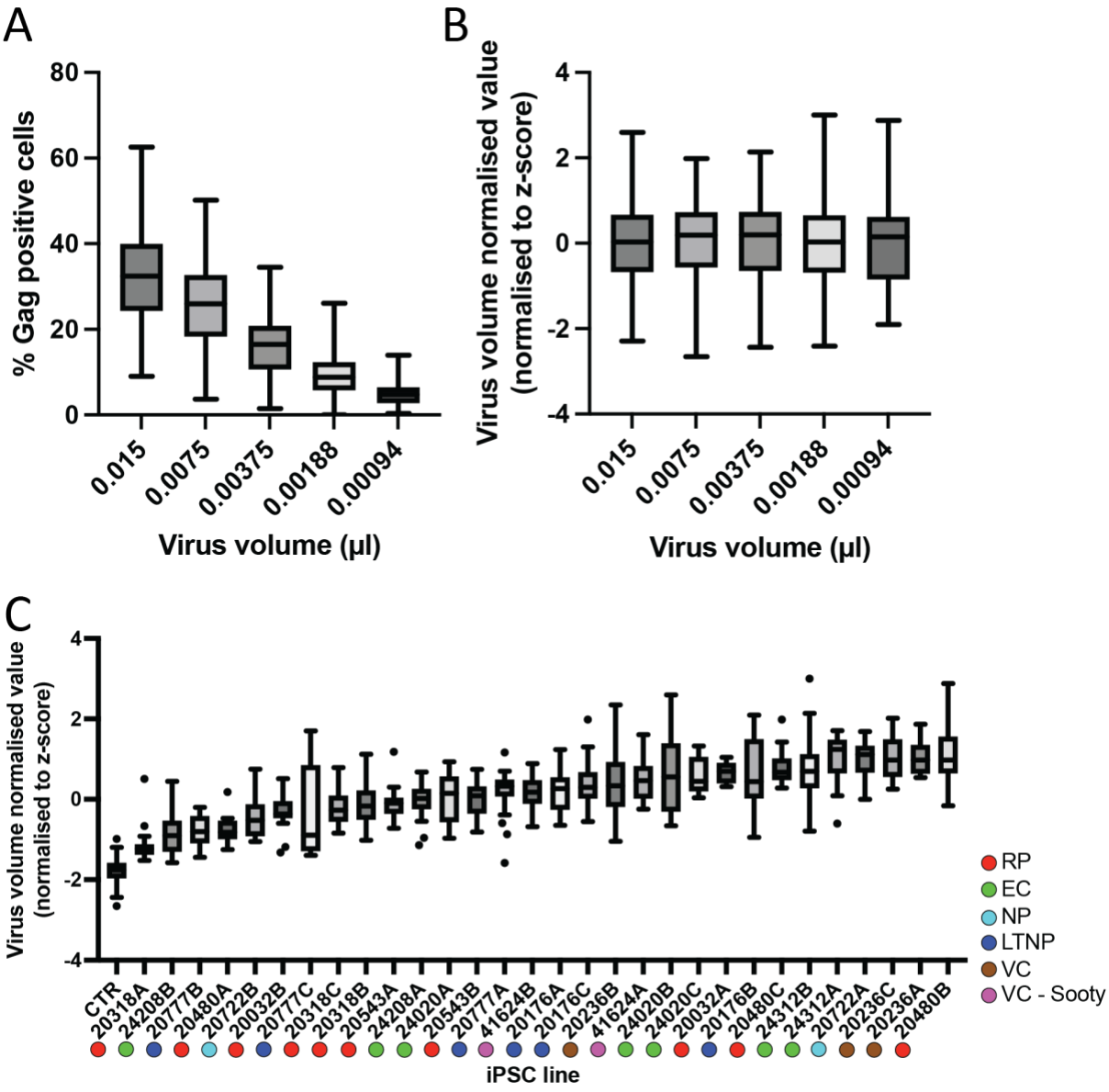


Figure 5.7: Virus infection data from the late gene expression assay on the MACS iPSCs analysed by normalisation to virus volume.

A) Box and whisker plot of percentage infection plotted against virus volume for all iPSC lines screened in the late gene expression assay. **B)** Box and whisker plot of percentage infection for iPSC lines after normalisation to virus volume using z-score. **C)** Box and whisker plot to show the iPSC lines ordered left to right by mean virus volume normalised value, when normalised for virus volume and screen number, from least permissive to most permissive. Coloured circles correspond to the clinical phenotype of the donor from which that cell line was derived (Red = RP, Green = EC, Cyan = NP, Blue = LTNP, Brown = VC, Pink = VC - Sooty).

5.2.8 Interpretation of the linear regression model and virus volume normalised analysis of the late gene expression phenotyping assay data

The 5 most resistant and permissive MACS iPSC lines to VSV-G pseudotyped full length HIV-1, strain NL4-3, when analysed by the two methods of analysis are ranked in **table 5.4** and **table 5.5** respectively. The level of correlation observed between the two analysis models, in terms of an iPSC line's resistance to HIV-1 infection ranking, is displayed in **figure 5.8A**. An R^2 value of 0.7397, after the fitting of a simple linear regression demonstrates a strong correlation. As with the infectivity assay iPSC line's ranks, to decipher which MACS iPSC lines were the most consistently resistant and permissive cell lines across both models of analysis, the collective rank for each MACS iPSC line was calculated.

For both the linear regression and virus volume normalised models, the MACS iPSC line, 20318A, was rank 1, making it undoubtably the most resistant MACS iPSC line to the virus. As mentioned earlier, this line has 3 RNA-seq samples so can provide reasonable statistical power when comparing its transcriptome to that of another cell line.

When identifying the most permissive MACS iPSC lines to the virus, two lines, 20722A and 20236A, shared the same collective rank. MACS iPSC line 20722A was found to be rank 1 when analysed by the linear regression model and rank 4 when analysed by the virus volume normalisation model. MACS iPSC line 20236A was found to be rank 2 for both analysis models. This gave a collective rank of 4 for both of these cell lines. To decipher between them, their late gene expression assay data analysed by both methods were compared to that of the most resistant MACS iPSC line, 20318A (**Figure 5.8**). When comparing 20722A and 20236A to 20318A after analysis with the linear regression model, the difference between 20722A and 20318A showed a higher statistical significance than the difference between 20236A and 20318A (**figure 5.8B**). This is due to 20236A having fewer data points as a result of a problem with this line during screen 10, leading to a less reliable permissivity phenotype. **Figure 5.8C** shows the comparison on the cell lines after analysis with the virus volume normalised model. Due to the nature of this model producing more data points, the statistical significance here was very high when comparing both highly permissive MACS iPSC lines to 20318A. Both 20722A and 20236A had two RNA-seq samples associated with them, so no preference could be made from this. Thus, as a result of the higher significance in the difference between MACS

iPSC lines 20722A and 20318A when analysed with the linear regression model, 20722A was taken forward to have its transcriptome compared to that of 20318A (discussed in section **5.2.11**).

Table 5.4: The top 5 most resistant cell lines from the late gene expression assay when analysed by either the linear regression or virus volume normalised model.

Rank	Linear Regression	Virus volume normalised
1	20318A	20318A
2	24208B	24208B
3	20777B	20777B
4	20480A	20480A
5	20722B	20722B

Table 5.5: The top 5 most permissive cell lines from the late gene expression assay when analysed by either the linear regression or virus volume normalised model.

Rank	Linear Regression	Virus volume normalised
1	20722A	20480B
2	20236A	20236A
3	20236C	20236C
4	20032A	20722A
5	24312A	24312A

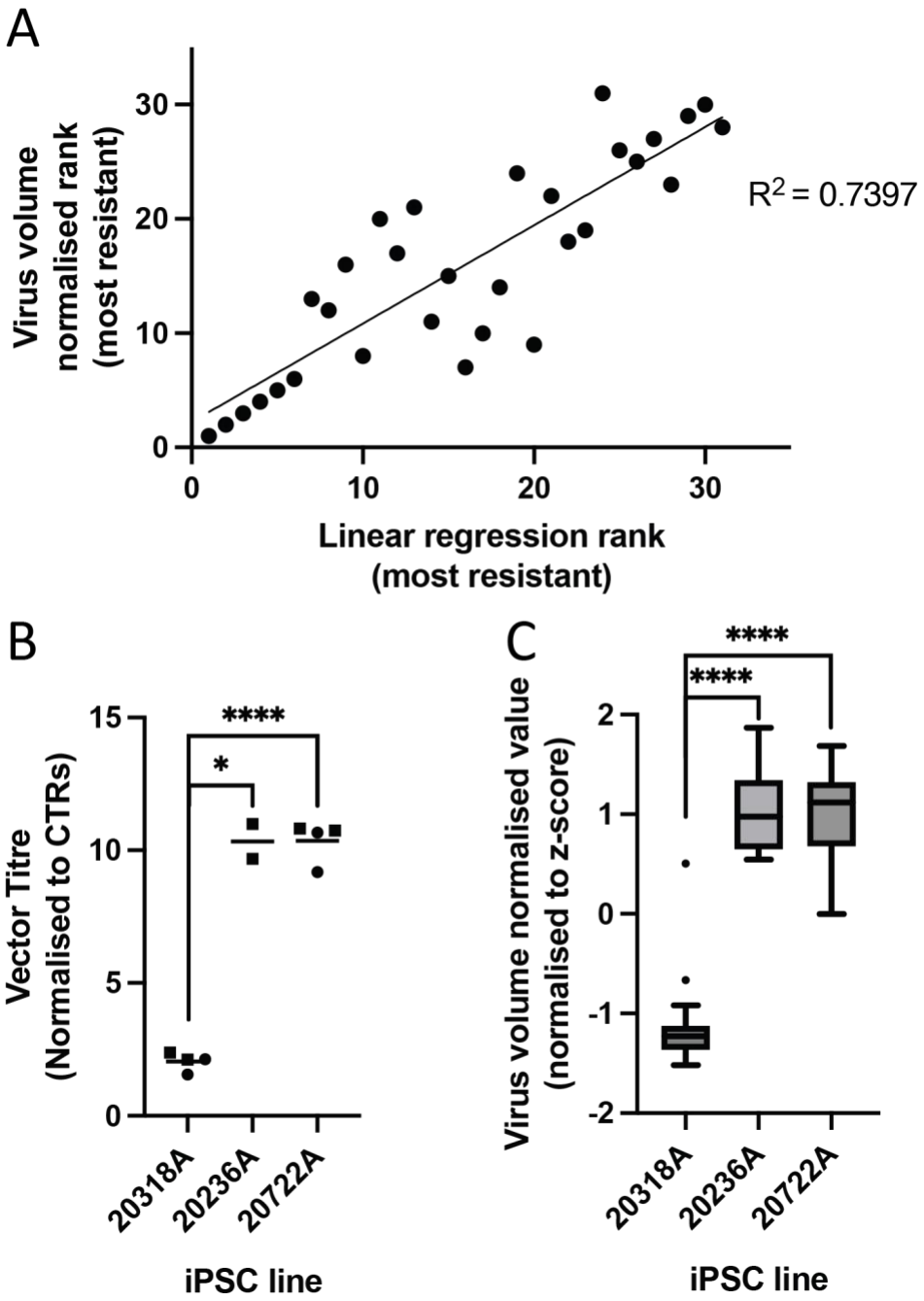


Figure 5.8: Interpretation of the late gene expression assay data when analysed by the linear regression and virus volume normalised models.

A) Plot displaying the correlation between both late gene expression assay analysis models in terms of all iPSC line's resistance to HIV-1 ranks. A simple linear regression is fitted, and the R^2 value is displayed. **B)** The most resistant MACS iPSC line, 20318A, compared to the two most permissive MACS iPSC lines, 20236A and 20722A, after analysis by the linear regression model and normalisation to the CTR iPSC line. Statistical test is unpaired t test with Welch's correction (* = $p<0.05$, **** = $p<0.0001$). **C)** The most resistant MACS iPSC line, 20318A, compared to the two most permissive MACS iPSC lines, 20236A and 20722A, after analysis by normalisation to virus volume and screen number. Statistical test is Mann Whitney U test (**** = $p<0.0001$).

5.2.9 Hierarchical clustering of RNA-seq samples from the screened iPSC lines

In order to compare the transcription profile of the screened iPSC lines, RNA samples for each of the lines were sent for next generation sequencing by Illumina Novaseq. PolyA selection was used to select for mRNA within the sample. Due to time constraints, RNA samples were sent before identification of the iPSC lines with extreme permissivity phenotypes. Therefore, 71 samples, consisting of at least one sample from each of the 30 screened MACS iPSC lines (plus the CTRs) were analysed rather than focusing on just those with extreme phenotypes. This somewhat compromised the replicate numbers of the following RNA-seq data analysis, however, this analysis shown in this thesis functions as preliminary to demonstrate a working pipeline. Further samples are available to be sent if necessary, and further analysis could be conducted. For the majority of the screened iPSC lines, an RNA-seq sample is available for each of the weekly screens it was conducted in. Additionally, some iPSC lines have a second replicate (denoted as 'rep2') from within one weekly screen. This allowed for the comparison of the RNA-seq data between cell lines as well as between the same cell line in different screens.

The similarity between all of the RNA-seq samples was visualised by creation of a dendrogram via hierarchical clustering using the DESeq2 package (**figure 5.9A**). This package conducts its own RNA-seq normalisation, and thus, utilises raw mapped counts to cluster samples based upon their differences and similarities in an unbiased technique assessing differential gene expression. The difference between each sample was measured as Euclidean distance, with a lower score representing less difference between those two samples. Reassuringly, all second replicates of RNA-seq data of a cell line from within a screen clustered directly next to the first replicate, demonstrating good consistency between RNA-seq replicates within screens. The transcriptome of the CTR line run in all the screens cluster together well, demonstrating their consistency and validating their use as a tool for normalisation between each screen. For some iPSC lines (e.g. 20722A, 20722B, 20236B, 24208A), the RNA-seq data collected from the different screens of that cell line clustered next to each other, showing a good level of consistency in the transcriptome of that cell line. In contrast, other cell lines' transcriptome results collected from different screens clustered poorly (e.g. 41624B, 24312A, 24020A, 20777C). This demonstrates undesired heterogeneity for these cell lines, suggesting that they exhibit a greater degree of viability in their transcriptome throughout culturing.

Importantly, the cell lines discussed in sections **5.2.5** and **5.2.8** to be taken forward for comparisons of their transcriptomes, 20318A, 20543B, and 20722A, clustered fairly well. These cell lines are indicated by coloured rectangles in **figure 5.9A**. A green rectangle highlights the RNA-seq samples for 20318A, the MACS iPSC line shown as highly resistant in both the infectivity assay and the late gene expression assay. Its two replicate samples collected in screen 10 cluster next to each other, and the sample taken from screen 11 clusters closely. Orange rectangles highlight the RNA-seq samples of the most permissive cell line identified by the infectivity assay, 20543B. Again, the replicate samples from within one screen cluster directly next to each other, and the sample collected from the other screen clusters reasonably close. Finally, the RNA-seq samples of the MACS iPSC line identified as the most permissive cell line by the late gene expression assay, 20722A, are highlighted by a red rectangle. Although no replicates are present from within a screen for this cell line, the replicates from the different screen cluster directly next to each other.

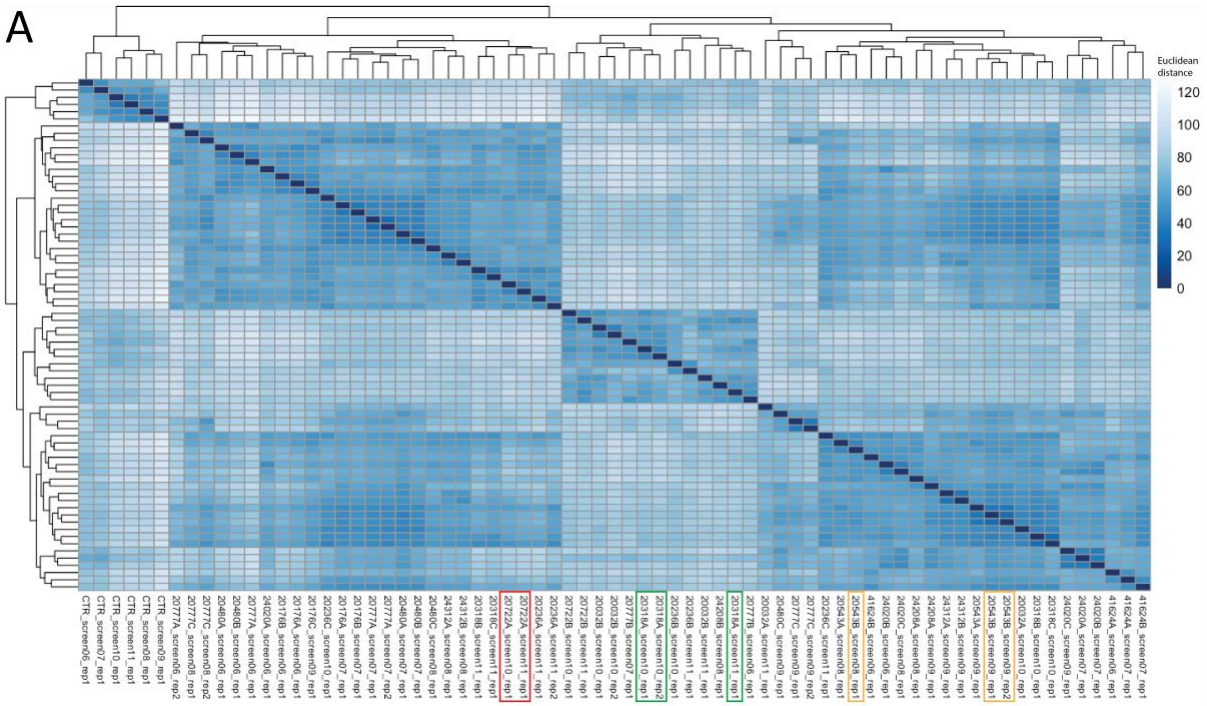


Figure 5.9: Hierarchical clustering of all RNA-seq samples from the iPSC lines run in the permissivity to HIV-1 screen.

A) Dendrogram demonstrating the differences and similarities between RNA-seq samples collected from iPSCs run in the permissivity screen. RNA-seq data was analysed by the DESeq2 package and cell lines were clustered unbiasedly based on their Euclidean distance to each other. A lower Euclidean distance, represented by a darker blue colour, demonstrates a higher similarity between those two samples. Coloured rectangles highlight the RNA samples used for the comparison of transcriptomes. Green = 20318A, iPSC line exhibiting an extreme resistant phenotype in both the infectivity and late gene expression assays. Orange = 20543B, iPSC line exhibiting an extreme permissive phenotype in the infectivity assay. Red = 20722A, iPSC line exhibiting an extreme permissive phenotype in the late gene expression assay. Figure generated with assistance from Nathalia Almeida dos Santos.

5.2.10 Comparison of the transcriptomes of the two MACS iPSC lines identified to exhibit opposite extreme infectivity phenotypes by the infectivity assay.

The raw RNA sequencing files were processed as described in section **2.10.3**, followed by z-score normalisation of the TPMs of each gene for each sample. This normalisation allows for the comparison of the different RNA-seq samples. **Figure 5.10A** shows the use of Pearson's correlation coefficient for multiple comparisons of two TPM-normalised sequence libraries. The figure displays the comparison of 6 RNA-seq samples, 3 from each of the two MACS iPSC lines, 20318A and 20543B, identified as having opposite extreme permissivity phenotypes by the infectivity assay. The RNA-seq samples from one of the cell lines shows higher levels of positive correlation to each other than when compared to RNA-seq samples from the other cell line. Also, as expected from **figure 5.9A**, the replicate RNA-seq samples of a cell line collected within a screen show higher positive correlation to each other than the same cell line's RNA-seq sample collected during the other screen. This correlation data demonstrates that both these cell line's transcriptomes are more similar to themselves than to the other cell line, granting validation for their use for transcriptome comparison.

The RNA-seq samples were then filtered for the ~20,000 protein coding genes within the human genome. Since a good level of similarity between the three RNA-seq samples of both cell lines was identified, an average TPM count for each cell line for each gene was calculated. To control for highly expressed genes, the average TPM counts were normalised using the z-score against all RNA-seq samples. The absolute difference for the z-score normalised average TPM counts of all protein coding genes between 20318A and 20543B was then calculated and termed ΔTPM. The 50 genes showing the greatest ΔTPMs are displayed in **figure 5.11A**.

Since the ΔTPM plot only accounts for the absolute difference in expression levels of a gene, a second analysis method considering fold change in expression was also conducted to compare the transcriptomes of these two MACS iPSC lines. The fold-change in expression of a gene between the two cell lines was plotted on an MA plot (**figure 5.12A**). To generate the MA plot, all average TPM counts were converted into $\log_2$. The $\log_2$-fold change between the two cell lines for each gene was then calculated and plotted on the y-axis. The x-axis signifies the mean expression of that gene. For this method of analysis, genes with a $\log_2$-fold change of >1.5 or <-1.5 (indicated by the horizontal dotted lines) have been identified as significantly

differentially expressed. The 50 genes identified in **figure 5.11A** are labelled on **figure 5.12A**. Many of these labelled genes do not present as significantly differentially expressed in terms of log_2-fold change on the MA plot. As a result, only genes that are identified as differentially expressed by both the ΔTPM and the MA analysis have been taken forward for the final gene candidate list shown in **table 5.6**. This table represents the genes identified by this project as potentially responsible for the differing infectivity phenotypes of two cell lines from the infectivity assay. Those genes more expressed in the resistant MACS iPSC line are potential HIV-1 restriction factors and the genes that are more expressed in the permissive MACS iPSC line are potential HIV-1 dependency factors. These 13 genes must now be validated for their influence on HIV-1 replication in downstream assays.

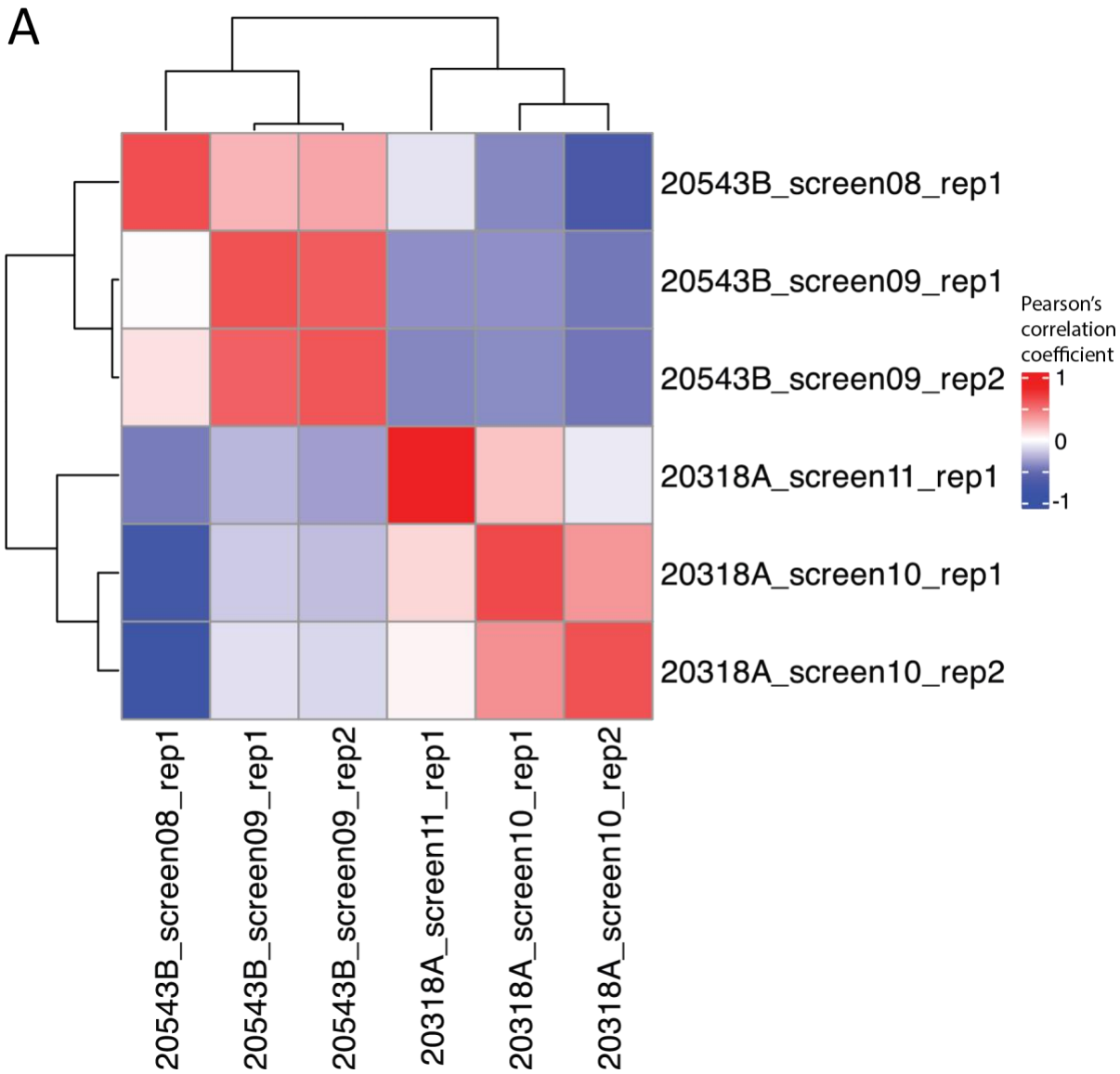


Figure 5.10: Pearson's correlation coefficient analysis of the transcriptomes of two MACS iPSC lines exhibiting opposite extreme permissivity phenotypes in the infectivity assay.
A) Heat map displaying the correlation between RNA-seq samples of a highly resistant MACS iPSC line, 20318A, and the most permissive MACS iPSC line, 20543B, when normalised to z-score. A Pearson's correlation coefficient of 1 represents perfect positive correlation, a Pearson's correlation coefficient of 0 denotes no correlation, and a Pearson's correlation coefficient of -1 symbolises perfect negative correlation between the samples. Figure generated with assistance from Nathalia Almeida dos Santos.

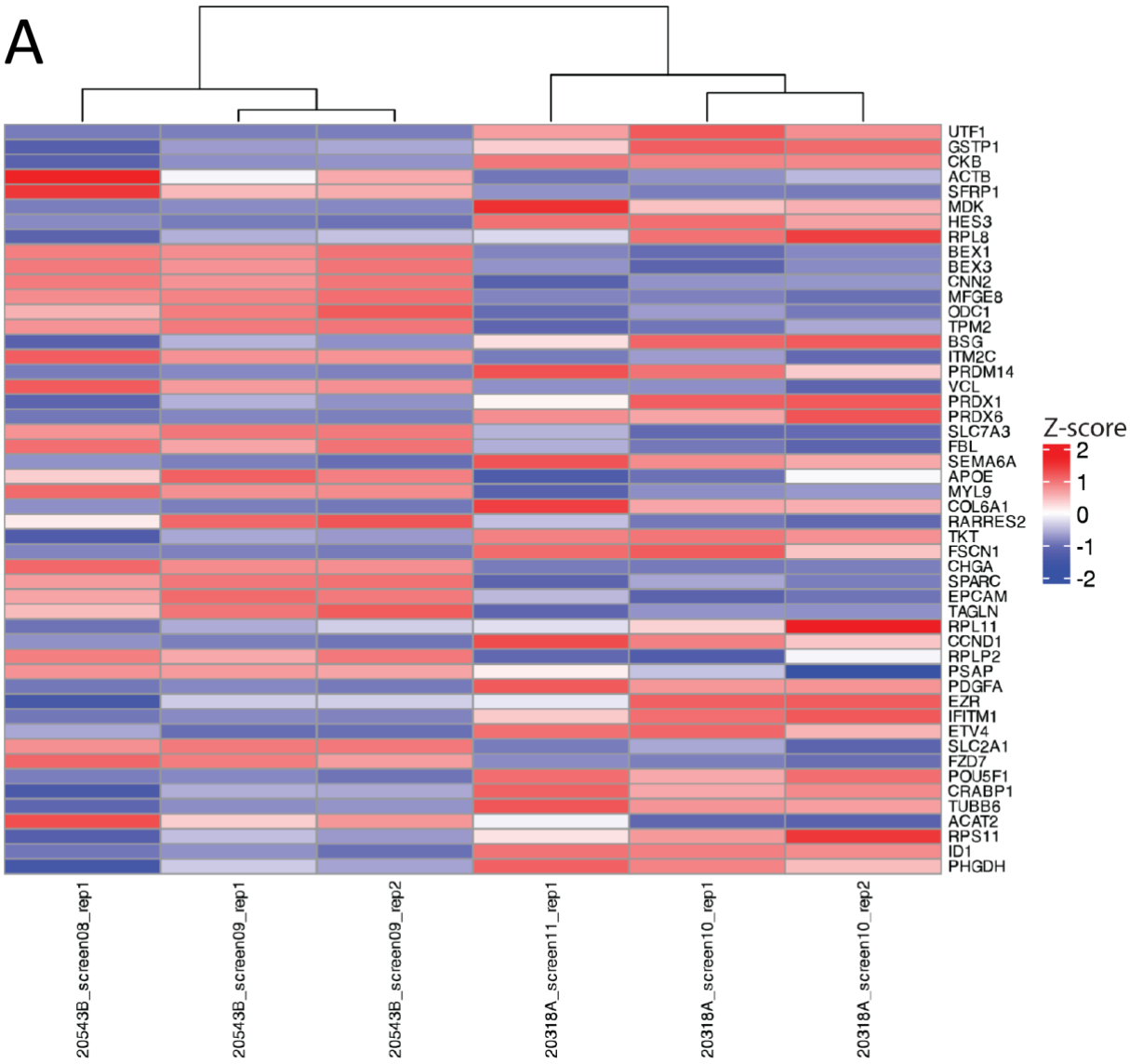


Figure 5.11: Top 50 protein coding genes exhibiting the greatest ΔTPM between 20318A and 20543B.

A) Heat map displaying the z-score normalised average ΔTPMs of the RNA-seq samples from 20318A to 20543B. Genes are ranked from top to bottom in terms of greatest average ΔTPM of all the RNA-seq samples from that cell line and data for each individual sample are shown. A higher z-score denotes a higher expression of that gene and a lower z-score indicates a lower expression of that gene. Figure generated with assistance from Nathalia Almeida dos Santos.

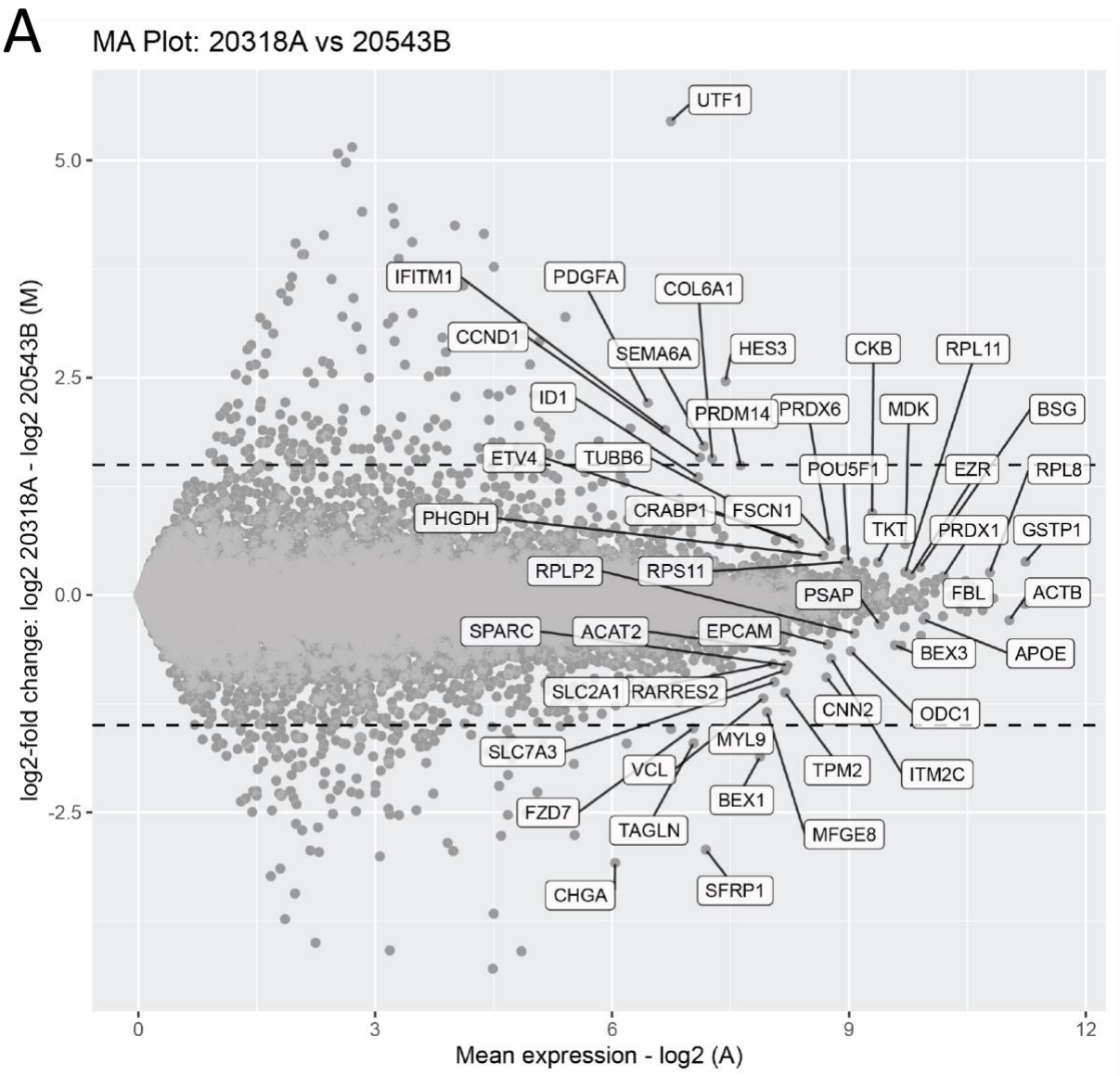


Figure 5.12: Comparison of the transcriptomes of 20318A and 20543B considering fold change in expression.

A) MA plot to show differential gene expression in the transcriptomes of two MACS iPSC lines, 20318A and 20543B. TPM values are converted to log_2 and each dot represents a different gene. The y-axis shows the log_2-fold change in expression between the two cell lines and the x-axis displays the mean expression level of that gene. The top 50 genes identified by ΔTPM analysis are labelled. Horizontal dotted lines represent the threshold set to be determined as a significant difference in log_2-fold change in expression (either >1.5 or <-1.5). Figure generated with assistance from Nathalia Almeida dos Santos.

Table 5.6: Candidate gene list identified by comparing the transcriptomes of two MACS iPSC lines, 20318A and 20543B, that exhibited opposite extreme phenotypes in the infectivity assay.

Gene	Full name	Potential dependency/restriction factor?
SEMA6A	Semaphorin 6A	Restriction
CCND1	Cyclin D1	Restriction
UTF1	Undifferentiated Embryonic Cell Transcription Factor 1	Restriction
HES3	Hes Family BHLH Transcription Factor 3	Restriction
IFITM1	Interferon Induced Transmembrane Protein 1	Restriction
PDGFA	Platelet Derived Growth Factor Subunit A	Restriction
COL6A1	Collagen Type VI Alpha 1 Chain	Restriction
CHGA	Chromogranin A	Dependency
MYL9	Myosin light chain 9	Dependency
SFRP1	Secreted Frizzled Related Protein 1	Dependency
BEX1	Brain Expressed X-Linked 1	Dependency
TAGLN	Transgelin	Dependency
FZD7	Frizzled Class Receptor 7	Dependency

5.2.11 Comparison of the transcriptomes of the two MACS iPSC lines identified to exhibit opposite extreme late gene expression phenotypes from the late gene expression assay.

The RNA-seq data for the two MACS iPSC lines, 20318A and 20722A, identified as exhibiting extreme opposite permissivity phenotypes in late gene expression assay were treated in the same way as the cell lines in section **5.2.10**. The z-score normalised TPM count libraries from each RNA-seq sample from the two cell lines were analysed for their level of similarity by Pearson's correlation coefficient (**figure 5.13A**). This time, 5 RNA-seq samples were compared, 3 from the most resistant MACS iPSC line, 20318A, and 2 from the most permissive MACS iPSC line, 20722A. Once again, the RNA-seq samples displayed a strong positive correlation when compared to an RNA-seq sample from the same cell line. These samples also demonstrated strong negative correlation when compared to a sample from the opposing cell line. These correlation trends are exhibited more strongly between these iPSC lines analysed in **figure 5.13A** than between the cell lines analysed in **figure 5.10A**. This high level of positive correlation within a cell line and high level of negative correlation between cell lines gives validation for using these samples for comparative transcriptomic analysis.

The RNA-seq samples were then filtered for the human genome's ~20,000 protein coding genes. As a result of the high level of correlation between samples from the same iPSC line, demonstrated by Pearson's correlation coefficient, the average TPM count across the samples of each cell line was calculated for each gene. The average TPM values were then normalised to z-score to control for highly expressed genes. The ΔTPM for each gene was then measured as the absolute difference between the z-score normalised average TPM values of 20318A and 20722A. **Figure 5.14A** displays the top 50 genes exhibiting the largest ΔTPM.

The MA plot was also conducted for these two cell lines as a second analysis method that considered fold-change in expression as opposed to absolute change in expression (**figure 5.15A**). Following the same procedure as with the infectivity assay identified cell lines, the TPM counts for each gene expressed in these cell lines were converted to $\log_2$ and plotted assessing $\log_2$-fold change in expression between the two cell lines and mean expression level. The top 50 genes displaying the greatest ΔTPM value identified in **figure 5.14A** are labelled in **figure 5.15A**. Of these 50 genes, 21 also exceeded the set threshold of either a $>1.5\log_2$ or $<-1.5\log_2$-fold change in expression. These 21 genes are taken forward as the candidate genes

potentially responsible for the differing permissivity phenotypes of 20318A and 20722A identified by the late gene expression assay. The list of these genes is displayed in **table 5.7**. The genes more expressed in the resistant cell line are potential HIV-1 restriction factors and the genes more expressed in the permissive cell line are potential HIV-1 dependency factors. These candidate genes must now have their potential influence on HIV-1 replication validated by downstream assays.

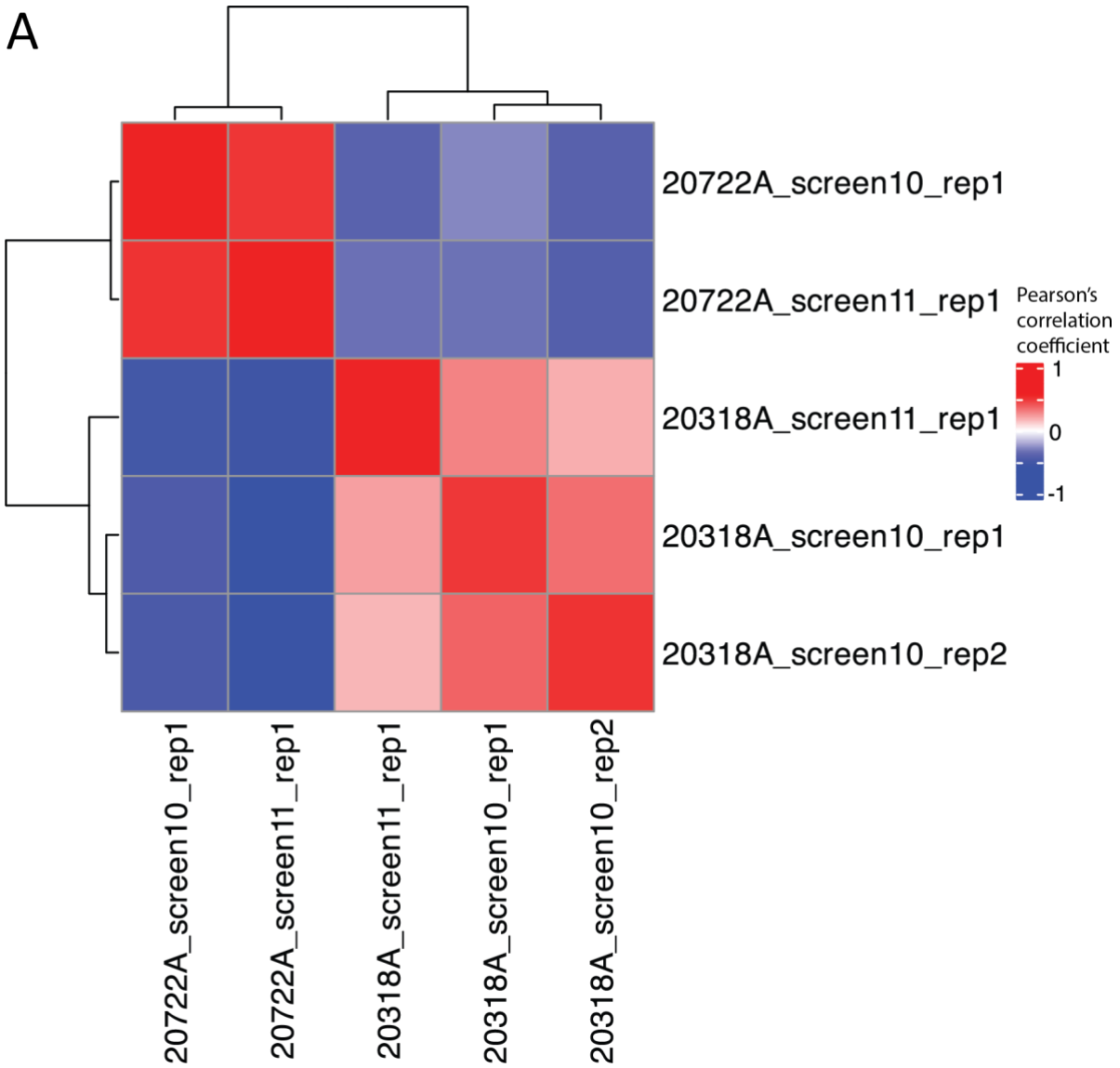


Figure 5.13: Pearson's correlation coefficient analysis of the transcriptomes of two MACS iPSC lines exhibiting opposite permissivity phenotypes in the late gene expression assay.
A) Heat map displaying the correlation between RNA-seq samples of the most resistant MACS iPSC line, 20318A, and the most permissive MACS iPSC line, 20722A, when normalised to z-score. A Pearson's correlation coefficient of 1 represents perfect positive correlation, a Pearson's correlation coefficient of 0 denotes no correlation, and a Pearson's correlation coefficient of -1 symbolises perfect negative correlation between the samples. Figure generated with assistance from Nathalia Almeida dos Santos.

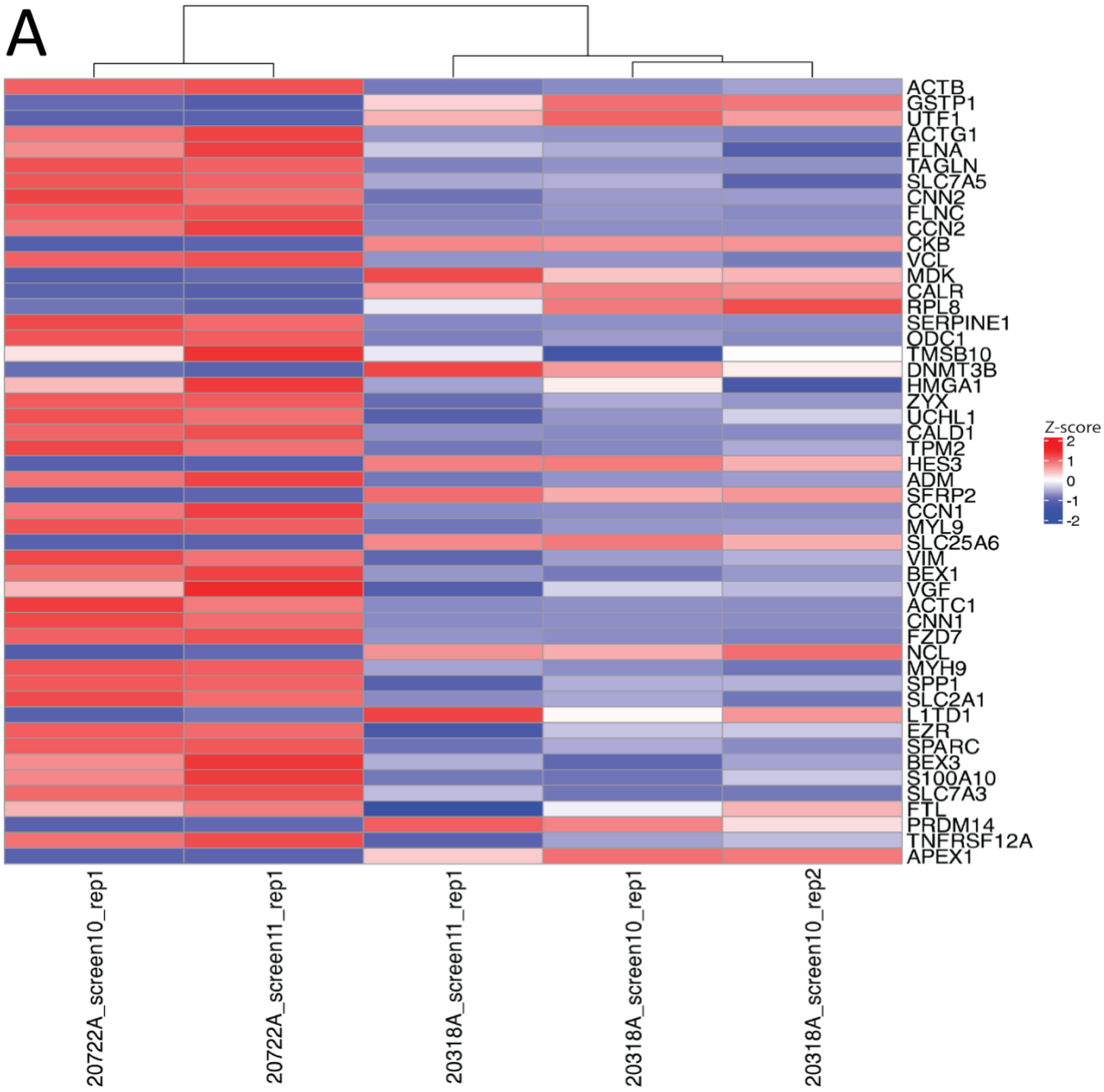


Figure 5.14: Top 50 protein coding genes exhibiting the greatest ΔTPM between 20318A and 20722A.

A) Heat map displaying the z-score normalised average ΔTPMs of the RNA-seq samples from 20318A to 20772A. Genes are ranked from top to bottom in terms of greatest average ΔTPM of all the RNA-seq samples from that cell line and data for each individual sample are shown. A higher z-score denotes a higher expression of that gene and a lower z-score indicates a lower expression of that gene. Figure generated with assistance from Nathalia Almeida dos Santos.

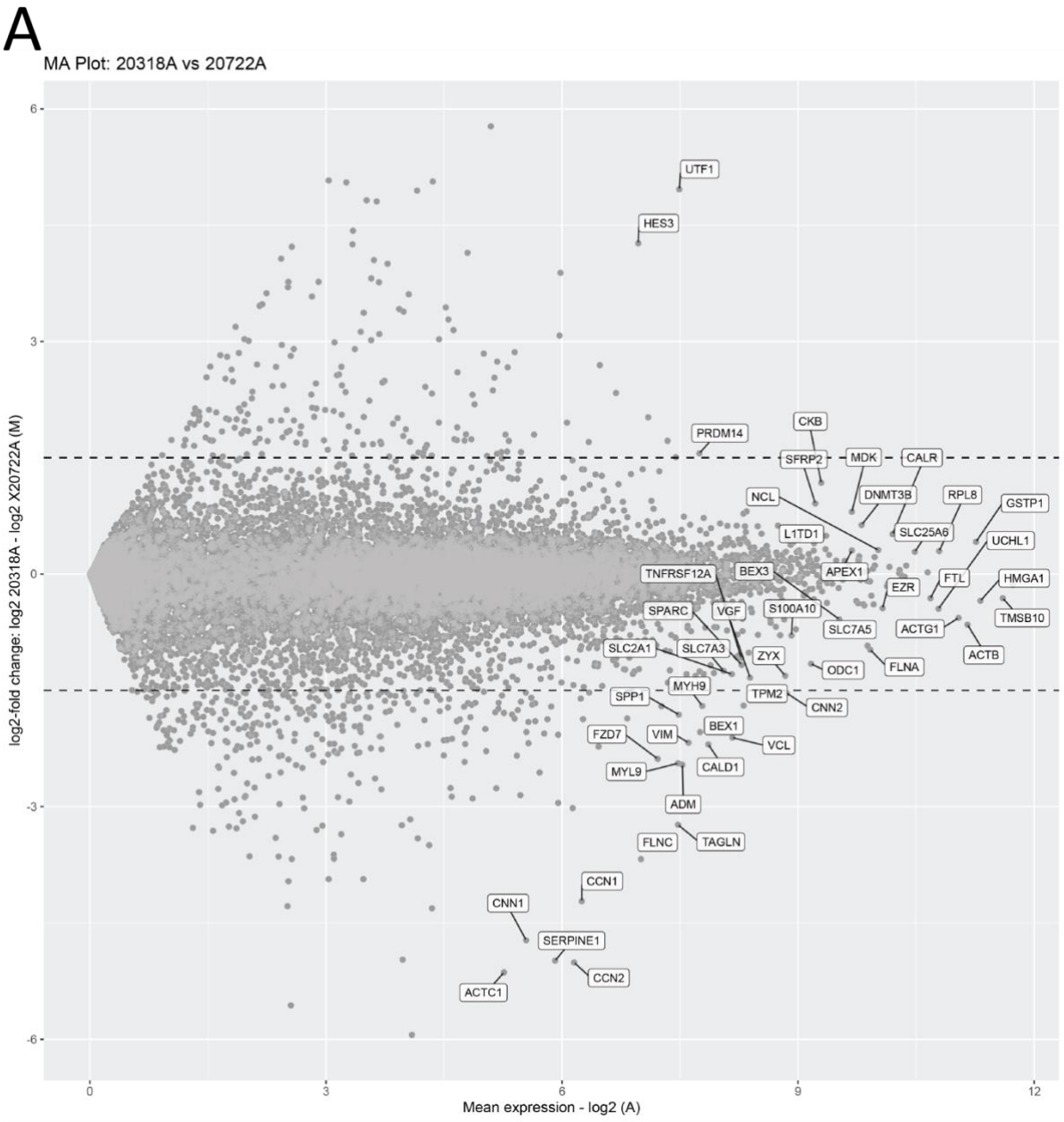


Figure 5.15: Comparison of the transcriptomes of 20318A and 20722A considering fold change in expression.

A) MA plot to show differential expression in the transcriptomes of two MACS iPSC lines, 20318A and 20722A. TPM values are converted to into log_2 and each dot represents a different gene. The y-axis shows the log_2-fold change in expression between the two cell lines and the x-axis displays the mean expression level of that gene. The top 50 genes identified by ΔTPM analysis are labelled. Horizontal dotted lines represent the thresholds set to be determined as a significant difference in log_2-fold change in expression (either >1.5 or <-1.5). Figure generated with assistance from Nathalia Almeida dos Santos.

Table 5.7: Candidate gene list identified by comparing the transcriptomes of two MACS iPSC lines, 20318A and 20722A, that exhibited opposite extreme phenotypes in the late gene expression assay.

Gene	Full name	Potential dependency/restriction factor?
PRDM14	PR/SET Domain 14	Restriction
UTF1	Undifferentiated Embryonic Cell Transcription Factor 1	Restriction
HES3	Hes Family BHLH Transcription Factor 3	Restriction
VIM	Vimentin	Dependency
VCL	Vinculin	Dependency
CNN2	Calponin 2	Dependency
MYH9	Myosin Heavy Chain 9	Dependency
MYL9	Myosin Light Chain 9	Dependency
SERPINE1	Serpin Family E Member 1	Dependency
CCN2	Cellular Communication Network Factor 2	Dependency
SPP1	Secreted Phosphoprotein 1	Dependency
CALD1	Caldesmon 1	Dependency
FLNC	Filamin C	Dependency
CNN1	Calponin 1	Dependency
BEX1	Brain Expressed X-Linked 1	Dependency
CCN1	Cellular Communication Network Factor 1	Dependency
ADM	Adrenomedullin	Dependency
TAGLN	Transgelin	Dependency
FZD7	Frizzled Class Receptor 7	Dependency
ACTC1	Actin Alpha Cardiac Muscle 1	Dependency
TPM2	Tropomyosin 2	Dependency

5.3 Discussion

In this chapter I sought to identify MACS iPSC lines that exhibited opposite extreme permissivity phenotypes to a VSV-G pseudotyped HIV-1-based GFP vector and a VSV-G pseudotyped full length HIV-1, strain NL4-3, by employing them in the infectivity and late gene expression assays respectively. For both assays, the transcriptomes of a highly resistant and highly permissive MACS iPSC line were compared to generate two lists of candidate genes potentially responsible for the differing permissivity phenotypes.

As previously discussed, the use of iPSCs as a tool for an HIV-1-based permissivity phenotyping screen poses many potential advantages: iPSCs predominantly differ by natural variation, they hold the ability to self-renew, and they intrinsically express high levels of many ISGs (Vigilante et al., 2019; Wu et al., 2018). Also, in contrary to past screens that have used gene knockouts, and therefore, cannot investigate genes essential for cell viability, studying the natural variation between different iPSC lines allows for the consideration of all genes, pathways, and networks. However, there are also some drawbacks of using undifferentiated iPSCs. Firstly, they do not express the cellular receptor CD4, an essential requirement for the entry of HIV-1 into a cell. To alleviate this problem, VSV-G pseudotyping of the HIV-1-based vector and replication-competent HIV-1, strain NL4-3, was necessary. This pseudotyping facilitates entry of the virus in to the iPSCs, however, it alters the entry mechanism used. When the Env-CD4 interaction is utilised, HIV-1 enters a cell by fusion with the plasma membrane and release of its contents into the cytoplasm. Whereas, with VSV-G pseudotyped viruses, the virus enters the cell via a clathrin-mediated endocytic pathway (Cureton et al., 2009; Johannsdottir et al., 2009). As a result of this modified entry mechanism, host cellular factors that interact with these VSV-G pseudotyped viruses during entry may not interact with natural HIV-1. Thus, factors identified by this screen suspected to interact with HIV-1 at this point of its life cycle must be carefully considered and validated using a virus that enters as HIV-1 typically would. Another shortcoming of the use of iPSCs is that they are not the primary target cell of HIV-1. HIV-1 naturally targets CD4 expressing highly specialised immune cells, such as CD4+ T-cells or macrophages. Although iPSCs do intrinsically express many of the ISGs that these HIV-1 target cells express upon stimulation with interferon, the dissimilarities between these cell

types are vast. Even between the relatively closely related different HIV-1 target cell types, some HIV-1 dependency or restriction factors play a more prominent role in some cell types than in others. For example, macrophages seem to exhibit higher resistance to HIV-1 infection than CD4+ T-cells due to high expression of the restriction factor SAMHD1 (Koppensteiner et al., 2012). Therefore, any factors identified by this screen as responsible for a differing permissivity phenotypes in iPSCs may not be physiologically relevant in the HIV-1 target cells. A final disadvantage of working with iPSCs for this project lies with their cell culture requirements. iPSCs are a cell type that can be problematic to work with and under slightly suboptimal conditions, can die or undergo undesired spontaneous differentiation. Therefore, particular expertise and techniques are necessary to maintain them in a viable undifferentiated state of high quality. One problematic requirement of iPSCs is their need to remain in contact with other iPSCs as 'small clumps' during passaging. This generated an obstacle for the permissivity screen as cells within these clumps would be less accessible to the viruses. This would result in the difference in clump sizes of different iPSC lines being a variable influencing their permissivity phenotypes rather than just their natural intrinsic variation. To remove this variable, single cell passaging of iPSCs was necessary. This is only possible in the presence of a media supplement, rho-associated protein kinase inhibitor (ROCKi), that blocks the apoptosis of solitary iPSCs. Unfortunately, the effects of ROCKi are not absolute, and different iPSC lines react with varying severities to single cell passaging, even in the presence of ROCKi. Therefore, some iPSC lines display greater survival rates than others, resulting in dissimilar cell numbers present prior to infection. To minimise this, RevitaCell, a more specific ROCKi with additional free radical scavenger and antioxidant properties, was used to increase iPSC survival. Despite this, variations in viable cell numbers present the following day after seeding the iPSC lines still persisted. This difference in survival after seeding coupled with varying survival rates of the different iPSC lines following infection with the viruses led to different cell numbers being present for analysis at the end of the pipeline. It is unclear how much of an influence varying cell numbers had on the percentage infection of each iPSC line. Potentially, a lower cell number may result in a greater percentage infection due to a greater virus to cell ratio. The cell number present at the end of the infection pipeline was quantified by the Operetta CLS during data acquisitions and so further analysis can be conducted in an attempt to decipher how impactful cell number was on percentage infection. As well as varying cell numbers, another limitation of using a ROCKi-like supplement is that it

blocks the ROCK1 and ROCK2-mediated cellular pathways, causing an unnatural change in the cell, that could potentially impact factors or pathways that interact with HIV-1. This also restricts the ability of this project to study any factors involved in these pathways. However, the presence of RevitaCell was taken into account for when collecting the RNA used to analyse transcriptomes, as the iPSCs that were harvested for this were cultured in conditions that matched the infection conditions.

Further potential limitations of the two permissivity to HIV-1 assays lie with the viruses used. Firstly, for the VSV-G pseudotyped HIV-1 based GFP vector, as well as integrated DNA, unintegrated DNA can also result in GFP expression (Yáñez-Muñoz et al., 2006). The two are indistinguishable by this assay, thus, any host factors identified could be interacting with unintegrated DNA rather than integrated DNA. Despite this, the unintegrated DNA will have progressed through all steps in the virus's life cycle up to and including reverse transcription, therefore, still providing plenty of opportunity for identification of host cell factor interactions. For the late gene expression assay, the lab adapted full length virus strain NL4-3 was employed and pseudotyped with VSV-G. The use of an HIV-1 infectious molecular clone may have proved more physiologically relevant, however, NL4-3 was utilised due to it being a widely used, well characterised provirus that shares all the same fundamental replication steps as clinical HIV-1 isolates. Future virus comparative studies could be completed on the MACS iPSCs using infectious molecular clones. Furthermore, any candidate genes identified by the assay conducted using NL4-3 will be screened against a panel of HIV-1 isolates.

The infection percentage data collected from the infectivity and late gene expression assays of the permissivity to HIV-1 screen were both analysed by two different models to identify iPSC lines that demonstrated consistent extreme permissivity phenotypes, regardless of the analysis model. A linear regression model followed by normalisation to the CTR iPSC line within each screen was used for both the infectivity and late gene expression data. This model assumes a linear relationship between infection percentage and vector/virus volume. Both the infectivity and late gene expression assays generated infection curves for each iPSC line where at least a section was linear, making this a good model to use for these data sets. However, this model can be susceptible to variation depending on the decision for which data points of each iPSC line's infection curve corresponded to the linear phase of the curve, and

thus, are included in the calculation of vector/virus titre. For an unbiased method for choosing which data points to include that was consistent across all the iPSC lines, the criteria typically employed by this lab when working out vector/virus titres was used. This involved using any data points from the infectivity assay presenting an infection percentage between 2-60%. For the late gene expression assay, this range was set as 1-20%, as the iPSC lines in this assay saturated at a lower infection percentage. To maintain the unbiased nature of this analysis, all points presenting an infection percentage within the range were utilised in the titre calculations and would have an effect on the final generated titre (unless removed by data cleaning). To control for variation between screens when employing the linear regression model, normalisation to the CTR iPSC line was used. This method of normalisation was validated by an observed correlation where the screens displaying the highest percentage infected cell lines also contained the highest infected CTRs (screen 6 for the infectivity assay and screen 7 for the late gene expression assay). The results of the dendrogram (**figure 5.9A**) also further validated their use for normalisation as the CTRs showed a consistent far greater similarity to itself between screens than to the MACS iPSCs. However, this method of normalisation does assume the CTRs are completely consistent in their permissivity phenotype, and that the variation observed between screens is only a result of unintended differences in the set-up of each screen.

The second method of analysis used for the infectivity assay data was the b-value from the logistic curve fitting. The b-value represents the rate of transition from the gentle to steep part of the infection curve as well as the steepness of the curve. The logistic curve typically generates a sigmoidal shape and the steepness of the curve represents the rate of change in infection percentage with variations in vector volume, making it a good model to use for analysis to determine a cell line's infectivity phenotype. Using this model negates the limitation of choosing individual data points experienced with the linear regression model as it uses a curve fitting. For the logistic model, normalisation between the screens was conducted using the z-score statistical measurement. Z-score normalisation is a useful tool that allows for data comparisons across screens, however, it does possess some limitations. Firstly, it assumes normal distribution in the data from within screens and thus, is susceptible to significant impact by outliers. Secondly, it also assumes normal distribution between

screens, meaning if there is a skewed proportion of resistant or permissive iPSC lines in one screen compared to others, this skewing would be lost by normalisation using z-score.

A possibly interesting permissivity characteristic of the iPSC lines not taken into account by either the linear regression or the b-value from the logistic curve fitting models is the infection percentage saturation point reached by an iPSC line. A lower saturation percentage may indicate that more of the bulk population of that iPSC line are resistant to HIV-1 infection. Since the RNA-seq data are collected from the bulk population of the iPSC line, the responsible factors or pathways for this could be studied. To take this peak saturation point into account, further data analysis could be conducted. For example, use of the 'a' value from the logistic function could be used as this would represent the maximum infection percentage value reached.

The final analysis method employed was normalisation to virus volume using the z-score statistical measurement. Use of this method meant all data points from the infection curve would be utilised, greatly increasing the statistical power of the results while also taking into account the maximum infection percentage value of an iPSC line. The limitations seen by this analysis method lie with aforementioned weaknesses in the use of z-score normalisation.

In general, for all the different analysis models used on the two assays, the replicates from within a screen grouped together well, whereas more variation was observed between the replicates of a cell line conducted in the different screens. This could be a result of unintended variations in the screen set-up each week or real changes in the permissivity of an iPSC line between screens. Although the iPSC lines all exist in a pluripotent state, slight variations within this state may be responsible for this unanticipated and undesired heterogeneity in permissivity phenotype of a cell line between screens. For example, the quality of the cells within the culture may affect how permissive that iPSC line is because the more pluripotent a cell line is, the more resistant it is to infection (Wu et al., 2018). This cell quality may be affected by several factors such as cell confluency before seeding the iPSCs, slight differences during passages or extensive time periods experienced not at 37°C. Variation may also arise from differing growth rates of the iPSC lines. This could play a role in the permissivity of the iPSC lines, as faster growing lines may be more susceptible to HIV-1 infection. However, if this

variation was present, the cellular factors responsible for the faster growth rate should be elucidated by the comparative transcriptomic analysis of the iPSC lines exhibiting differing permissivity phenotypes. This would provide useful insight into the relationship between HIV-1 and the cell cycle.

From both of the analysis methods used in both assays, the clinical phenotype of the donor from which the MACS iPSC lines were derived seemed to have little impact on their permissivity phenotypes. Also, different iPSC clones generated from the same donor's PBMC sample showed inconsistent clustering of permissivity phenotypes. Some clones generated from a donor clustered fairly well, whereas clones from other donors clustered poorly. This lack of clinical correlation and donor clonal consistency were somewhat expected due to the unsought changes in the genetics, epigenetics and imprinting induced by the reprogramming process and extended culture periods of the iPSCs. Further, PBMCs are a heterogeneous population of cells, and therefore, clonal iPSC lines from the same donor may have originated from different cell types, resulting in iPSC clonal populations with different epigenetic signatures. iPSCs have been shown to be reprogrammed from many of the different cell types that are present within a population of PBMCs, including CD4+ T-cells, CD8+ T-cells, monocytes, B-cells, and CD34+ haematopoietic stem cells (Bueno et al., 2016; Iizuka-Koga et al., 2017; Isogai et al., 2018; Loh et al., 2009; Nishimura et al., 2013). The CytoTune-iPS 2.0 Sendai Reprogramming Kit does state that it targets a specific cell population during the pre-transduction culture of the PBMCs, however, it does not disclose exactly what cell type this is. The lack of correlation to clinical phenotypes may also be due to the use of iPSCs in an *in vitro* system which differs significantly to the *in vivo* setting these clinical phenotypes were established in.

Despite using two different analysis models for both the infectivity and late gene expression assay data, a good level of correlation in the identification of the MACS iPSC lines exhibiting extreme permissivity phenotypes was observed. Utilising two models allowed for the identification of MACS iPSC lines that consistently demonstrated extreme permissivity phenotypes, regardless of the analysis method. The infectivity assay identified the most resistant MACS iPSC line to be 24208B, followed by 20318A, via both analysis models. The most permissive MACS iPSC line was judged by collective ranking to be 20543B. This MACS

iPSC line was ranked 1st and 4th most permissive when analysed by the linear regression model and logistic curve fitting model respectively. The late gene expression assay identified 20318A as the most resistant MACS iPSC line in both methods of analysis. This was highly consistent with the infectivity assay, indicating a good overall resistance of this MACS iPSC line to HIV-1 infection. The most permissive MACS iPSC lines identified by the late gene expression assay were 20722A and 20236A, which both presented a collective rank of 4.

The MACS iPSC lines taken forward for transcriptome analysis for the infectivity assay were 20318A, due to more RNA-seq data being available, and 20543B. For the late gene expression assay, the two MACS iPSC lines used for transcriptome comparison were 20318A and 20722A, due to 20722A exhibiting a more statistically significant extreme late gene expression phenotype than 20236A when compared to 20318A.

Hierarchical clustering using the DESeq2 package demonstrated the similarities and dissimilarities between the RNA-seq data of the iPSC lines tested in the different screens. It showed that RNA-seq replicates from the same iPSC line within the same screen were highly consistent as they clustered directly next to each other, validating the reliability of the RNA-seq data. RNA-seq samples from the same iPSC line collected from different screens showed mixed consistency, some clustered very closely, whereas others clustered less well. This demonstrated an unanticipated and undesired heterogeneity in the transcriptome of some iPSC lines between different screens. This heterogeneity may be explained by the same aforementioned factors potentially responsible for some iPSC lines exhibiting large variations in their permissivity phenotype between the two screens. For example, MACS iPSC lines 20777C and 24020A show this variation in permissivity phenotypes when analysed by the linear regression model in both the infectivity and the late gene expression assay data. The RNA-seq samples of these two cell lines from within a screen also cluster very distantly from the RNA-seq samples of the same cell line collected from the other screen. This suggests that these two cell lines show large amounts of transcriptional and phenotypic variation during their culture period. Therefore, it may provide fruitful to compare the transcriptomes of these iPSC lines to their own transcriptome collected during the other screen.

Pearson's correlation coefficient was used to assess the similarities and difference between the RNA-seq samples of the cell lines identified as exhibiting extreme opposite permissivity phenotypes. For the cells lines identified by both the infectivity assay and the late gene expression assay, correlation was shown to be greater between samples from the same cell line compared to samples from the other cell line. This validated the use of these RNA-seq samples for comparative transcriptomic analysis.

The comparison of the transcriptomes of the MACS iPSCs identified as exhibiting extreme permissivity phenotypes by both the infectivity and late gene expression assays was conducted using two analysis methods: ΔTPM and log_2-fold change. ΔTPM investigated the absolute difference in expression of a gene, whereas log_2-fold change considered the fold change in expression between the two extreme opposite cell lines. Coupling both of these analysis methods generated a more succinct gene list, consisting of just genes that had a high ΔTPM and a significant log_2-fold change. However, genes identified by only one of the methods, with a high ΔTPM and a low log_2-fold change or visa-versa, may prove to be physiologically relevant and could also be considered for further validation. Further, more genes could be judged as significantly differentially expressed by reducing the stringency of the thresholds set. More than just the top 50 ΔTPM genes could be examined and a decreased log_2-fold change threshold could be applied.

The candidate gene list produced as a result of the comparison of the transcriptomes of MACS iPSC lines 20318A and 20543B, consisted of 13 genes. 6 of the genes were more expressed in the highly resistant cell line, 20318A, making them potential restriction factors, and 5 genes were more expressed in the permissive cell line, 20543B, making them potential dependency factors. The identified genes span a wide array of proposed functions, including plasma membrane proteins, structural proteins, and involvement in cell signalling. Of note, one potential restriction factor, *HES3*, and one potential dependency factor, *BEX1*, have been predicted to encode for proteins that enable the binding of negative or positive regulators to RNA polymerase II respectively. Thus, these factors may influence the transcription of the HIV-1 provirus. *CCND1*, which encodes for cyclin D1, was also identified as a potential restriction factor by the infectivity assay. Cyclin D1 modulates transition from G1 to S phase in the cell cycle via its regulation of CDK4 and CDK6. Since HIV-1 is thought to preferentially infect cycling

cells, this may make Cyclin D1's action as a restriction factor seem paradoxical, however, Cyclin D1 has been previously suggested to restrict SARS-CoV-2 and influenza infection (Fan et al., 2017; Gupta & Mlcochova, 2022). Therefore, further work into Cyclin D1's action as an HIV-1 restriction factor could prove rewarding. Interestingly, *IFITM1*, which encodes for an already recognised HIV-1 restriction factor that inhibits viral entry, IFITM1, was identified by this analysis (Compton et al., 2014; Yu et al., 2015). While this gives validation for the screen's ability to pick up host factors that impact HIV-1 replication, IFITM1 may be solely responsible for the differing infectivity phenotypes between these two MACS iPSC lines, thus, other cell lines may need to be examined for the identification of novel host factors. However, IFITM1 is primarily expressed at the cell plasma membrane and is usually associated with the restriction of R5 viruses (Foster et al., 2016). This makes the identification of IFITM1 in this screen somewhat surprising due to the endosomal route of entry used by the VSV-G pseudotyped GFP vector.

The second candidate gene list, generated by the comparison of the MACS iPSC lines 20318A and 20722A, was made up of 21 genes, 18 of which were potential dependency factors. Once again, these genes spanned a wide variety of functions, with an intriguing amount being involved in the structural organisation of the cell cytoskeleton and the formation or binding of actin. HIV-1 has been documented to utilise the cytoskeleton during both early and late stages of the virus's life cycle (Jolly et al., 2007; Stolp & Fackler, 2011). Two of the proposed genes, *VIM* and *SERPINE1*, have already been suggested to have viral interactions. *VIM*, encodes for a type III intermediate filament protein, vimentin, that is cleaved by the HIV-1 protease resulting in an altered nuclear architecture in the cell (Shoeman et al., 1990, 2001). Vimentin's role as an HIV-1 dependency factor has been validated by an observed restriction of HIV-1 replication after it was blocked with a synthetic peptide (Fernández-Ortega et al., 2016). Further, vimentin has also been shown to be involved in the trafficking of Gag to the plasma membrane, however, this function would not be picked up by the late gene expression assay discussed here (Q. Wang et al., 2016). *SERPINE1* encodes for SERPINE1, which has been previously described as part of the innate antiviral immunity and has been found to prevent influenza A virus maturation (Dittmann et al., 2015). However, here it is suggested as an HIV-1 dependency factor. In the context of HIV-1 clinical infection, a pre-print article indicated that increased SERPINE1 levels leads to increased amyloid beta deposition, resulting in amyloid

pathology (András et al., 2023). However, its potential function as an HIV-1 dependency factor remains to be discovered. The identification of genes already known to interact with HIV-1 validates this screens methodology and provides some confidence that the remaining genes not yet linked to HIV-1 will yield some positive outputs.

6 of the genes identified as potential HIV-1 restriction or dependency factors, including the previously mentioned *HES3* and *BEX1*, feature in both candidate gene lists (**figure 5.16A**). This demonstrates a good level of consistency in the gene hits produced by the screen, even though they are the result of two different assays. The use of the same highly resistant MACS iPSC line, 20318A, in both transcriptome comparisons does give a greater chance for this to occur, but still provides additional confidence that these genes are responsible for this cell line's extreme permissivity phenotype. One of these repeated genes, *UTF1*, is expressed almost exclusively in pluripotent stem cells. The high expression of this gene in the highly resistant MACS iPSC line compared to both highly permissive MACS iPSC lines suggests a differing level of pluripotency at the time of infection. As discussed previously, higher levels of pluripotency has been described to correlate to higher levels of resistance to viral infection (Wu et al., 2018). As a result, this difference in pluripotency may explain the difference in infectivity phenotypes, thus, it may be more beneficial to analyse other MACS iPSC lines exhibiting opposite extreme permissivity phenotypes that show similar expression levels of pluripotency genes.

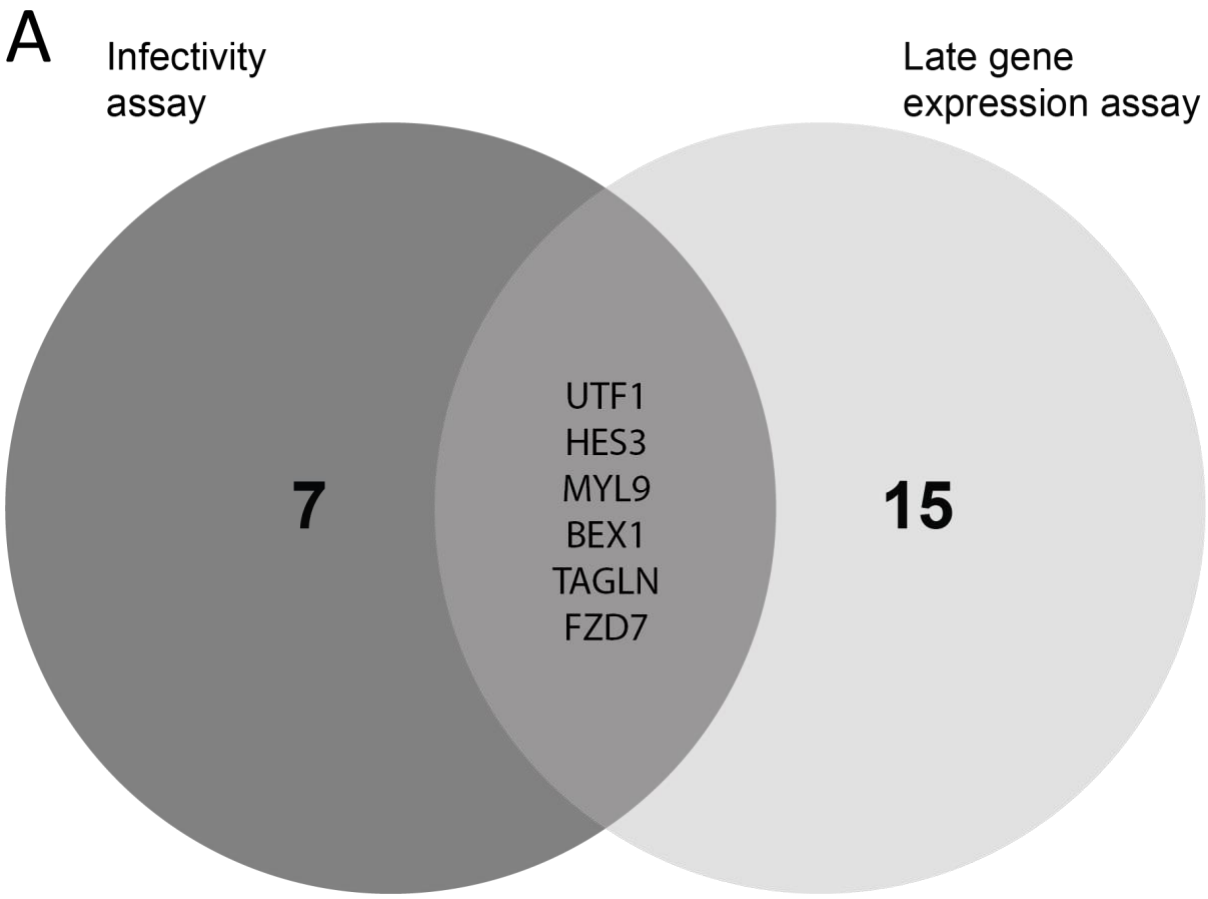


Figure 5.16: Candidate genes identified as a result of the two assays of the permissivity to HIV-1 infection screen.

A) Venn diagram displaying the overlap of candidate genes produced by the comparative transcriptome analysis of the cell lines identified as exhibiting extreme opposite permissivity phenotypes in the infectivity assay and the late gene expression assay.

As aforementioned, this transcriptome comparison analysis is preliminary and only shows the production of two candidate gene lists. Different transcriptome analytical approaches that utilise these identified iPSC line infectivity phenotypes could lead to the generation of many more candidate gene lists. Comparison of the transcriptomes of different MACS iPSC lines exhibiting extreme permissivity phenotypes (e.g. 20236A and 24208B) would yield different gene candidates or further validate already proposed candidates. Additionally, the CTR iPSC line consistently displayed very high levels of resistance to HIV-1 infection. It was not taken forward for transcriptome comparison here as the MACS iPSCs were the primary focus, however, it could possess a valuable transcriptome to compare to. Another potential transcriptome comparison suggestion would be to compare the RNA-seq samples of a single MACS iPSC line that exhibited large amounts of variation in permissivity phenotype between screens (e.g. 20777C). This may produce an overall more consistent transcriptome between the samples, thus, highlighting any particular genes responsible for the change in permissivity phenotype. Comparisons of the transcriptomes of different MACS iPSC clones derived from the same donor that exhibited opposite permissivity phenotypes could also produce some interesting findings (e.g. 20722A+B), as these lines will be very similar, if not identical, genetically. The transcriptomes of these cells may be more similar (they cluster fairly well in **figure 5.9A**), and therefore, the differences responsible for the differing infectivity phenotypes may be more easily uncovered. Perhaps differences between these clones may be a result of genetic, epigenetic or imprinting changes in genes that influence HIV-1 infection that were produced during the reprogramming process. Finally, to increase the power and confidence behind the RNA-seq data collected, more RNA sample replicates of the iPSC lines of interest could be sent for sequencing.

In summary, this chapter presents a novel screen-based method for uncovering host factors, pathways or networks that potentially influence HIV-1 infection. The results described here identify two lists of candidate genes that require validation for their possible interactions with HIV-1. Also, the permissivity phenotyping of the MACS iPSCs can act as a foundation for further transcriptomic analysis and the generation of additional candidate gene lists.

Chapter 6: Discussion

6.1 Summary

More than 40 years of HIV-1 research have highlighted a growingly intricate interaction between the virus and its host. Discovery of numerous cellular factors that the virus depends on and other factors that restrict viral replication, as well as virally encodes proteins that inhibit the action of these restriction factors, underline the complexity of this interaction. The balance between these factors can determine whether HIV-1 replication is successful or fails, thus, influencing transmission and pathogenesis. Several of these factors were identified through binary techniques involving the manipulation of specific genes, either by knocking them down or over expressing them artificially. More recently, latest developments in omics-driven techniques have empowered scientists to explore the effects of HIV-1 infection extensively, offering high-throughput and comprehensive genomic-wide resolution. Screens utilising RNAi or CRISPR-Cas9 technologies, candidate gene studies, and GWAS have been somewhat successful in identifying host factors that influence HIV-1 disease outcome, however, much of these virus-host interactions remain to be found. Hence, the overall aim of this thesis was to establish an HIV-1 permissivity screen that exploited the natural variation between individuals utilising iPSCs reprogrammed from participants of the MACS. The MACS grants an unprecedented opportunity to access PBMC samples from individuals whose natural HIV-1 disease progressions were studied in great detail.

First, it was important to establish a reliable protocol for the successful generation of iPSCs from PBMCs, due to the MACS PBMC samples being a finite resource. The feeder-free protocol was concluded to be the method of choice for this after demonstrating increased reproducibility and ease of use. Additional complications associated with the MACS PBMC samples were then made apparent as 14 of the 18 samples were collected at a time point when the donor was HIV+. To maximise the probability that the resultant iPSC lines would not contain any HIV-1 provirus, a CD4+ cell depletion was conducted prior to the reprogramming protocol. A double CD4+ cell depletion utilising MagACS technology was shown to be sufficient for the generation of HIV-1 provirus-free iPSCs from a PBMC sample containing HIV-1 infected cells.

After the development of this reprogramming pipeline, the MACS iPSCs could then be generated with the upmost confidence of success. The 18 MACS donors were selected based upon their extreme clinical phenotypes in response to HIV-1 infection. Notably, 4 ECs, who maintain their HIV-1 viral load below detectable levels in the absence of therapy, and 4 RPs, who show an extraordinarily rapid physiological decline after HIV-1 infection, were included in the selected population. 50 MACS iPSC lines, 2-3 clones per donor, were reprogrammed from the 18 MACS donors. These iPSCs were validated for their pluripotency by expression of indicative markers, OCT-4 and NANOG, and their ability to differentiate into all three germ layers. The MACS iPSCs were also confirmed to have been depleted of any residual Sendai viral vectors, that were employed to deliver the Yamanka factors, and were also verified to contain no HIV-1 provirus. Unfortunately, due to time restraints, only 2 of the MACS iPSCs were tested for their chromosomal integrity via G-banded karyotype analysis, but both of these lines displayed a normal diploid 46, XY karyotype.

With the successful generation and validation of the MACS iPSCs, the permissivity to HIV-1 infection screen could then be completed. The aim of this screen was to identify MACS iPSC lines that exhibited extreme permissivity phenotypes when challenged with either a VSV-G pseudotyped HIV-1-based GFP vector or a VSV-G pseudotyped full length HIV-1, strain NL4-3, in an infectivity assay or late gene expression assay respectively. The factors responsible for these natural permissivity phenotypic differences could then potentially be identified by analysing the transcriptomes of the iPSCs at the point of infection. Infectivity phenotypes of the MACS iPSC lines were identified after analysis of the infectivity assay data by a linear regression model and a logistic curve fitting model, and late gene expression phenotypes of the MACS iPSC lines were determined after analysis of the late gene expression assay data by a linear regression model and a normalisation to virus volume model. Finally, the transcriptomes of the MACS iPSC lines identified as exhibiting opposite extreme permissivity phenotypes in each of the assays were compared, considering both ΔTPM and $\log_2$-fold change in expression, in an attempt to discover the factors responsible for the differences in permissivity phenotypes. Two candidate gene lists were generated, one from each assay, that contained potential dependency or restriction factors that might influence HIV-1 infection. The genes in both lists spanned a wide range of functions and some had been linked previously to HIV-1. These results validate the methodology undertaken in this thesis to uncover genes

involved in HIV-1 infection, offer a list of new candidate genes that require confirmation of their interaction with HIV-1, and provide a platform for further work to be conducted.

6.2 Future directions

6.2.1 Validation of candidate genes

The candidate genes identified in this thesis require confirmation for their influence on HIV-1 infection, and if successfully validated, their mechanism of HIV-1 regulation can be studied. To do this, the genes are first targeted for knockdown or over expression, and any effect on HIV-1 replication is observed. Gene knockdown could be achieved via RNAi or CRISPR inactivation and gene over expression could be produced with the use of cDNA or CRISPR activation. These assays could be conducted on a variety of cell types. Immortalised cell lines of HIV-1 natural target cells, such as T-cell lines including Jurkat, CEM-SS or SUP-T1, or macrophage lines such as phorbol 12-myristate 13-acetate (PMA) differentiated THP-1 cells, provide a good model for studying HIV-1 infection. However, these cell lines can differ greatly from their primary cell counter parts and may exhibit a loss in expression of key genes. iPSCs could also be utilised, however, as previously discussed, these cells are not the natural target of HIV-1, and thus, factors deemed as influential on HIV-1 replication in this cell line may not be physiologically relevant. Primary CD4+ T-cells pose as potentially the best model for validating candidate factors that interact with HIV-1, however, their limitations involve cell number availability and an increased difficultly to genetically manipulate. Therefore, the employment of each of these cell models could prove useful. After the validation of a candidate factor, the mechanism by which that gene provides its aid or restriction on HIV-1 infection can be deduced with further experiments.

6.2.2 Correlation of permissivity phenotypes to genomic data of the MACS participants

As well as utilising the transcriptomic data of the MACS iPSC lines, the genomic data of the MACS participants are also available. The genomic data can be coupled with the MACS iPSC permissivity phenotypes to identify possible human SNVs, or more substantial polymorphisms, affecting HIV-1 infection. As mentioned earlier, different clones of MACS iPSCs generated from the same MACS participant that exhibited opposite extreme permissivity phenotypes (e.g. 20722) may be of particular interest to study further. Resequencing the genomes of the different clones from a donor may highlight genetic mutations, either a result of reprogramming of different original cells, a consequence of the

reprogramming process or acquired during culturing, that could potentially be responsible for the contrasting permissivity phenotypes. For example, one study has shown that a high proportion of blood-derived iPSC lines obtained BCOR mutations that have functional consequences, thus, HIV-1 infection influencing mutations could also arise (Rouhani et al., 2022).

6.2.3 Correlation to other screens utilising natural variation between individuals

The Human Induced Pluripotent Stem Cells Initiative (HipSci) is a bank of over 950 high-quality well characterised iPSC lines (Kilpinen et al., 2017). In contrast to the MACS iPSCs, the majority of these lines are generated from healthy donors, but these lines also have extensive data sets associated with them, including genomics and transcriptomics. A similar screen to the one discussed in this thesis is currently being conducted in our lab using the HipSci lines by Neophytos Kouphou, thus, any candidate genes identified here can be cross-referenced to that work. Further, the use of the HipSci bank opens this study to a epidemiologically different population. Since the majority of MACS participants worked with in this thesis are of North American descent, the use of the HipSci consortium, which is made up of >90% Europeans, will deliver a larger scope for identifying potentially influential SNVs (Nehme & Barrett, 2020).

Another study of interest to this thesis being conducted in our lab is being performed by Nathalia Almeida dos Santos. This study also utilises the HipSci bank of iPSCs but with the additional step of differentiating them into a natural target cell of HIV-1, macrophages. These iMacs are then employed in a similar HIV-1 permissivity screen to the one described here. This additional differentiation step adds a level of undesired variation between the cell lines, however, it alleviates the issues of no CD4 expression and not being a natural target cell of HIV-1, associated with infecting iPSCs. As a result, this study can monitor spreading infections of non-pseudotyped HIV-1 viruses, and therefore, is potentially a more physiologically relevant model for assessing HIV-1 infection *in vivo*. Cross-referencing the results from the late gene expression assay of this MACS iPSC HIV-1 permissivity screen with the late gene expression assay results from the similar screen conducted on the iMacs yielded some encouragingly reproducible results. Examining the top 50 ΔTPM genes from both screens, found 10 genes in common (**figure 6.1A**). These 10 genes are: TMSB10, FTL, ACTG1, SERPINE1, SPARC, S100A10, TPM2, CALR, EZR, UCHL1. Although only two of these genes, SERPINE1 and

TPM2, appear in late gene expression assay candidate gene list (**table 5.7**), due to the extra consideration of log_2-fold change in expression, this substantiates these other 8 genes as potential influencers of HIV-1 replication.

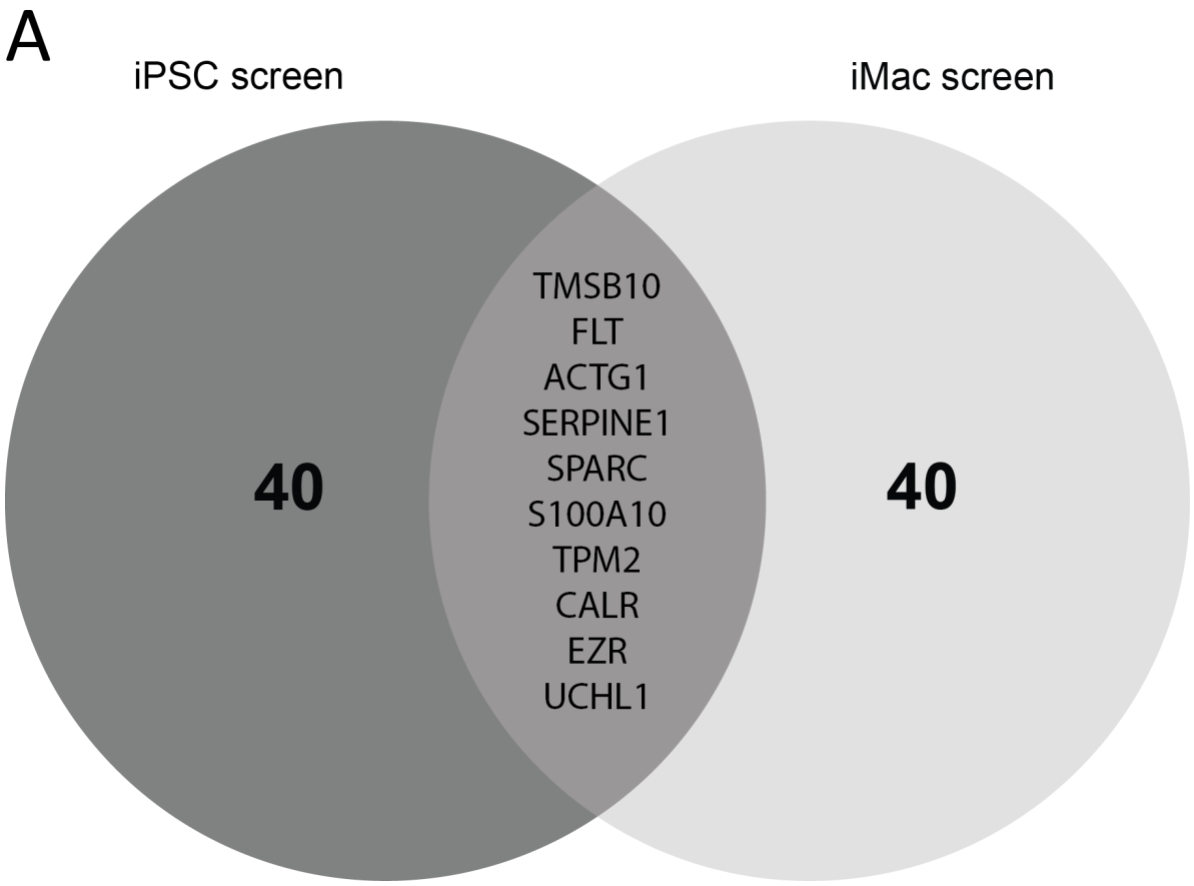


Figure 6.1: Top 50 ΔTPM genes from two different screens exploiting the natural variation between individuals

A) Venn diagram displaying the overlap of the top 50 ΔTPM identified genes by the late gene expression assay of the screen described in this thesis (iPSC screen) and a similar screen utilising iMacs (iMac screen).

6.2.4 Future uses of the MACS iPSCs

The generation of the MACS iPSCs by this project offers a new scientific resource to the field of HIV-1 research. Being derived from individuals that not only displayed an array of clinically interesting disease outcomes, but also had their clinical histories mapped in great detail, means the MACS iPSCs are a unique bank of virtually inexhaustible cells, with the potential to differentiate into any cell type, whose future applications have vast potential.

An example use of these MACS iPSCs would be to harness their differentiation potential to produce large quantities of the natural target cells of HIV-1, such as CD4+ T-cells or macrophages. Unfortunately, an *in vitro* methodology to generate CD4+ T-cells has not yet been defined. However, an effective and reproducible method to generate iMacs has been available since 2013 (Wilgenburg et al., 2013). Similar to the aforementioned study on HipSci derived iMacs, iMacs could be derived from the MACS iPSCs and tested in a permissivity to HIV-1 screen. In fact, Nathalia Almeida dos Santos has already successfully generated iMacs from one of the MACS iPSC lines, 20480A, and analysed its permissivity to HIV-1. Upon the establishment of an iPSC-derived CD4+ T-cell producing protocol, a similar HIV-1 permissivity screen could also be conducted on this cell type.

6.3 Impact

The work presented in this thesis demonstrates a novel HIV-1 permissivity screen utilising the natural variation in iPSCs derived from participants in the MACS, in an attempt to uncover new host factors that influence HIV-1 infection. The results provide several candidate genes that require validation for their interactions with HIV-1, while also acting as a foundation for future analysis that could be conducted on the data generated. Further, the MACS-derived iPSCs will be deposited and made publicly available for others to use to pursue their research interests.

References

Achuthan, V., Perreira, J. M., Sowd, G. A., Puray-Chavez, M., McDougall, W. M., Paulucci-Holthauzen, A., Wu, X., Fadel, H. J., Poeschla, E. M., Multani, A. S., Hughes, S. H., Sarafianos, S. G., Brass, A. L., & Engelman, A. N. (2018). Capsid-CPSF6 Interaction Licenses Nuclear HIV-1 Trafficking to Sites of Viral DNA Integration. *Cell Host & Microbe*, *24*(3), 392-404.e8. https://doi.org/10.1016/j.chom.2018.08.002

Albanese, A., Arosio, D., Terreni, M., & Cereseto, A. (2008). HIV-1 Pre-Integration Complexes Selectively Target Decondensed Chromatin in the Nuclear Periphery. *PLoS ONE*, *3*(6), e2413. https://doi.org/10.1371/journal.pone.0002413

Alizon, S., von Wyl, V., Stadler, T., Kouyos, R. D., Yerly, S., Hirschel, B., Böni, J., Shah, C., Klimkait, T., Furrer, H., Rauch, A., Vernazza, P. L., Bernasconi, E., Battegay, M., Bürgisser, P., Telenti, A., Günthard, H. F., & Bonhoeffer, S. (2010). Phylogenetic Approach Reveals That Virus Genotype Largely Determines HIV Set-Point Viral Load. *PLoS Pathogens*, *6*(9), e1001123. https://doi.org/10.1371/journal.ppat.1001123

Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M., & Berger, E. A. (1996). CC CKR5: A RANTES, MIP-1α, MIP-1β Receptor as a Fusion Cofactor for Macrophage-Tropic HIV-1. *Science*, *272*(5270), 1955–1958. https://doi.org/10.1126/science.272.5270.1955

Almeida, J. R., Price, D. A., Papagno, L., Arkoub, Z. A., Sauce, D., Bornstein, E., Asher, T. E., Samri, A., Schnuriger, A., Theodorou, I., Costagliola, D., Rouzioux, C., Agut, H., Marcelin, A.-G., Douek, D., Autran, B., & Appay, V. (2007). Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *The Journal of Experimental Medicine*, *204*(10), 2473–2485. https://doi.org/10.1084/jem.20070784

Amini-Bavil-Olyaee, S., Choi, Y. J., Lee, J. H., Shi, M., Huang, I.-C., Farzan, M., & Jung, J. U. (2013). The Antiviral Effector IFITM3 Disrupts Intracellular Cholesterol Homeostasis to Block Viral Entry. *Cell Host & Microbe*, *13*(4), 452–464. https://doi.org/10.1016/j.chom.2013.03.006

An, P., & Winkler, C. A. (2010). Host genes associated with HIV/AIDS: advances in gene discovery. *Trends in Genetics : TIG*, *26*(3), 119–131. https://doi.org/10.1016/j.tig.2010.01.002

András, I. E., Serrano, N., Djuraskovic, I., Fattakhov, N., Sun, E., & Toborek, M. (2023). Extracellular vesicle-Serpine-1 affects neural progenitor cell mitochondrial functions and synaptic density: modulation by amyloid beta and HIV-1. *Research Square*. https://doi.org/10.21203/rs.3.rs-2551245/v1

Anokye-Danso, F., Trivedi, C. M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P. J., Epstein, J. A., & Morrisey, E. E. (2011). Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency. *Cell Stem Cell*, *8*(4), 376–388. https://doi.org/10.1016/j.stem.2011.03.001

Arora, J., McLaren, P. J., Chaturvedi, N., Carrington, M., Fellay, J., & Lenz, T. L. (2019). HIV peptidome-wide association study reveals patient-specific epitope repertoires associated with HIV control. *Proceedings of the National Academy of Sciences*, *116*(3), 944–949. https://doi.org/10.1073/pnas.1812548116

Arora, J., Pierini, F., McLaren, P. J., Carrington, M., Fellay, J., & Lenz, T. L. (2020). HLA Heterozygote Advantage against HIV-1 Is Driven by Quantitative and Qualitative

Differences in HLA Allele-Specific Peptide Presentation. *Molecular Biology and Evolution*, *37*(3), 639–650. https://doi.org/10.1093/molbev/msz249

Askjaer, P., Jensen, T. H., Nilsson, J., Englmeier, L., & Kjems, J. (1998). The Specificity of the CRM1-Rev Nuclear Export Signal Interaction Is Mediated by RanGTP. *Journal of Biological Chemistry*, *273*(50), 33414–33422. https://doi.org/10.1074/jbc.273.50.33414

Bachand, F., Yao, X. J., Hrimech, M., Rougeau, N., & Cohen, E. A. (1999). Incorporation of Vpr into human immunodeficiency virus type 1 requires a direct interaction with the p6 domain of the p55 gag precursor. *The Journal of Biological Chemistry*, *274*(13), 9083–9091. https://doi.org/10.1074/jbc.274.13.9083

Ban, H., Nishishita, N., Fusaki, N., Tabata, T., Saeki, K., Shikamura, M., Takada, N., Inoue, M., Hasegawa, M., Kawamata, S., & Nishikawa, S.-I. (2011). Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. *Proceedings of the National Academy of Sciences*, *108*(34), 14234–14239. https://doi.org/10.1073/pnas.1103509108

Bar, S., Schachter, M., Eldar-Geva, T., & Benvenisty, N. (2017). Large-Scale Analysis of Loss of Imprinting in Human Pluripotent Stem Cells. *Cell Reports*, *19*(5), 957–968. https://doi.org/10.1016/j.celrep.2017.04.020

Bar-Nur, O., Russ, H. A., Efrat, S., & Benvenisty, N. (2011). Epigenetic Memory and Preferential Lineage-Specific Differentiation in Induced Pluripotent Stem Cells Derived from Human Pancreatic Islet Beta Cells. *Cell Stem Cell*, *9*(1), 17–23. https://doi.org/10.1016/j.stem.2011.06.007

Barr, L., & Jefferys, R. (2020). A landscape analysis of HIV cure-related clinical research in 2019. *Journal of Virus Eradication*, *6*(4), 100010. https://doi.org/10.1016/j.jve.2020.100010

Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., & Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (New York, N.Y.)*, *220*(4599), 868–871. https://doi.org/10.1126/science.6189183

Beitari, S., Ding, S., Pan, Q., Finzi, A., & Liang, C. (2017). Effect of HIV-1 Env on SERINC5 Antagonism. *Journal of Virology*, *91*(4). https://doi.org/10.1128/JVI.02214-16

Bejarano, D. A., Peng, K., Laketa, V., Börner, K., Jost, K. L., Lucic, B., Glass, B., Lusic, M., Müller, B., & Kräusslich, H.-G. (2019). HIV-1 nuclear import in macrophages is regulated by CPSF6-capsid interactions at the nuclear pore complex. *ELife*, *8*. https://doi.org/10.7554/eLife.41800

Beloglazova, N., Flick, R., Tchigvintsev, A., Brown, G., Popovic, A., Nocek, B., & Yakunin, A. F. (2013). Nuclease Activity of the Human SAMHD1 Protein Implicated in the Aicardi-Goutières Syndrome and HIV-1 Restriction. *Journal of Biological Chemistry*, *288*(12), 8101–8110. https://doi.org/10.1074/jbc.M112.431148

Benayoun, B. A., Pollina, E. A., & Brunet, A. (2015). Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nature Reviews Molecular Cell Biology*, *16*(10), 593–610. https://doi.org/10.1038/nrm4048

Berger, E. A., Doms, R. W., Fenyö, E.-M., Korber, B. T. M., Littman, D. R., Moore, J. P., Sattentau, Q. J., Schuitemaker, H., Sodroski, J., & Weiss, R. A. (1998). A new classification for HIV-1. *Nature*, *391*(6664), 240–240. https://doi.org/10.1038/34571

Berkhout, B., Silverman, R. H., & Jeang, K.-T. (1989). Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell*, *59*(2), 273–282. https://doi.org/10.1016/0092-8674(89)90289-4

Betancor, G., Dicks, M. D. J., Jimenez-Guardeño, J. M., Ali, N. H., Apolonia, L., & Malim, M. H. (2019). The GTPase Domain of MX2 Interacts with the HIV-1 Capsid, Enabling Its Short Isoform to Moderate Antiviral Restriction. *Cell Reports*, *29*(7), 1923-1933.e3. https://doi.org/10.1016/j.celrep.2019.10.009

Bhutani, K., Nazor, K. L., Williams, R., Tran, H., Dai, H., Džakula, Ž., Cho, E. H., Pang, A. W. C., Rao, M., Cao, H., Schork, N. J., & Loring, J. F. (2016). Whole-genome mutational burden analysis of three pluripotency induction methods. *Nature Communications*, *7*(1), 10536. https://doi.org/10.1038/ncomms10536

Bichel, K., Price, A. J., Schaller, T., Towers, G. J., Freund, S. M., & James, L. C. (2013). HIV-1 capsid undergoes coupled binding and isomerization by the nuclear pore protein NUP358. *Retrovirology*, *10*(1), 81. https://doi.org/10.1186/1742-4690-10-81

Blanquart, F., Wymant, C., Cornelissen, M., Gall, A., Bakker, M., Bezemer, D., Hall, M., Hillebregt, M., Ong, S. H., Albert, J., Bannert, N., Fellay, J., Fransen, K., Gourlay, A. J., Grabowski, M. K., Gunsenheimer-Bartmeyer, B., Günthard, H. F., Kivelä, P., Kouyos, R., … Fraser, C. (2017). Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *PLOS Biology*, *15*(6), e2001855. https://doi.org/10.1371/journal.pbio.2001855

Blissenbach, M., Grewe, B., Hoffmann, B., Brandt, S., & Überla, K. (2010). Nuclear RNA Export and Packaging Functions of HIV-1 Rev Revisited. *Journal of Virology*, *84*(13), 6598–6604. https://doi.org/10.1128/JVI.02264-09

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., & Young, R. A. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell*, *122*(6), 947–956. https://doi.org/10.1016/j.cell.2005.08.020

Braaten, D., Franke, E. K., & Luban, J. (1996). Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. *Journal of Virology*, *70*(6), 3551–3560. https://doi.org/10.1128/jvi.70.6.3551-3560.1996

Bracq, L., Xie, M., Benichou, S., & Bouchet, J. (2018). Mechanisms for Cell-to-Cell Transmission of HIV-1. *Frontiers in Immunology*, *9*. https://doi.org/10.3389/fimmu.2018.00260

Brakier-Gingras, L., Charbonneau, J., & Butcher, S. E. (2012). Targeting frameshifting in the human immunodeficiency virus. *Expert Opinion on Therapeutic Targets*, *16*(3), 249–258. https://doi.org/10.1517/14728222.2012.665879

Brasey, A., Lopez-Lastra, M., Ohlmann, T., Beerens, N., Berkhout, B., Darlix, J.-L., & Sonenberg, N. (2003). The Leader of Human Immunodeficiency Virus Type 1 Genomic RNA Harbors an Internal Ribosome Entry Segment That Is Active during the G_2/M Phase of the Cell Cycle. *Journal of Virology*, *77*(7), 3939–3949. https://doi.org/10.1128/JVI.77.7.3939-3949.2003

Brass, A. L., Dykxhoorn, D. M., Benita, Y., Yan, N., Engelman, A., Xavier, R. J., Lieberman, J., & Elledge, S. J. (2008). Identification of Host Proteins Required for HIV Infection Through a Functional Genomic Screen. *Science*, *319*(5865), 921–926. https://doi.org/10.1126/science.1152725

Brenchley, J. M., Schacker, T. W., Ruff, L. E., Price, D. A., Taylor, J. H., Beilman, G. J., Nguyen, P. L., Khoruts, A., Larson, M., Haase, A. T., & Douek, D. C. (2004). CD4+ T Cell Depletion during all Stages of HIV Disease Occurs Predominantly in the Gastrointestinal Tract. *The Journal of Experimental Medicine*, *200*(6), 749–759. https://doi.org/10.1084/jem.20040874

Brenner, B. G., Roger, M., Routy, J., Moisi, D., Ntemgwa, M., Matte, C., Baril, J., Thomas, R., Rouleau, D., Bruneau, J., Leblanc, R., Legault, M., Tremblay, C., Charest, H., & Wainberg, M. A. (2007). High Rates of Forward Transmission Events after Acute/Early HIV-1 Infection. *The Journal of Infectious Diseases*, *195*(7), 951–959. https://doi.org/10.1086/512088

Briggs, J. A. G., Wilk, T., Welker, R., Kräusslich, H.-G., & Fuller, S. D. (2003). Structural organization of authentic, mature HIV-1 virions and cores. *The EMBO Journal*, *22*(7), 1707–1715. https://doi.org/10.1093/emboj/cdg143

Brown, T. T. (2005). Antiretroviral Therapy and the Prevalence and Incidence of Diabetes Mellitus in the Multicenter AIDS Cohort Study. *Archives of Internal Medicine*, *165*(10), 1179. https://doi.org/10.1001/archinte.165.10.1179

Bueno, C., Sardina, J. L., Di Stefano, B., Romero-Moya, D., Muñoz-López, A., Ariza, L., Chillón, M. C., Balanzategui, A., Castaño, J., Herreros, A., Fraga, M. F., Fernández, A., Granada, I., Quintana-Bustamante, O., Segovia, J. C., Nishimura, K., Ohtaka, M., Nakanishi, M., Graf, T., & Menendez, P. (2016). Reprogramming human B cells into induced pluripotent stem cells and its enhancement by C/EBPα. *Leukemia*, *30*(3), 674–682. https://doi.org/10.1038/leu.2015.294

Burke, D. (1997). Recombination in HIV: An Important Viral Evolutionary Strategy. *Emerging Infectious Diseases*, *3*(3), 253–259. https://doi.org/10.3201/eid0303.970301

Bushman, F. D., Malani, N., Fernandes, J., D'Orso, I., Cagney, G., Diamond, T. L., Zhou, H., Hazuda, D. J., Espeseth, A. S., König, R., Bandyopadhyay, S., Ideker, T., Goff, S. P., Krogan, N. J., Frankel, A. D., Young, J. A. T., & Chanda, S. K. (2009). Host Cell Factors in HIV Replication: Meta-Analysis of Genome-Wide Studies. *PLoS Pathogens*, *5*(5), e1000437. https://doi.org/10.1371/journal.ppat.1000437

Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J. J., Kaslow, R., Buchbinder, S., Hoots, K., & O'Brien, S. J. (1999). *HLA* and HIV-1: Heterozygote Advantage and *B*35 - Cw*04* Disadvantage. *Science*, *283*(5408), 1748–1752. https://doi.org/10.1126/science.283.5408.1748

Carter, M. J., Fish, M., Jennings, A., Doores, K. J., Wellman, P., Seow, J., Acors, S., Graham, C., Timms, E., Kenny, J., Neil, S., Malim, M. H., Tibby, S. M., & Shankar-Hari, M. (2020). Peripheral immunophenotypes in children with multisystem inflammatory syndrome associated with SARS-CoV-2 infection. *Nature Medicine*, *26*(11), 1701–1707. https://doi.org/10.1038/s41591-020-1054-6

CDC. (1993). Update: Mortality Attributable to HIV Infection Among Persons Aged 25-44 Years -- United States, 1991 and 1992. *Morbidity and Mortality Weekly Report (MMWR)*, *42*(45), 869–872.

CDC. (1998). *AIDS falls from top ten causes of death; teen births, infant mortality, homicide all decline*.

Chamberlain, S. J., Chen, P.-F., Ng, K. Y., Bourgois-Rocha, F., Lemtiri-Chlieh, F., Levine, E. S., & Lalande, M. (2010). Induced pluripotent stem cell models of the genomic imprinting disorders Angelman and Prader–Willi syndromes. *Proceedings of the National Academy of Sciences*, *107*(41), 17668–17673. https://doi.org/10.1073/pnas.1004487107

Chan, D. C., Fass, D., Berger, J. M., & Kim, P. S. (1997). Core Structure of gp41 from the HIV Envelope Glycoprotein. *Cell*, *89*(2), 263–273. https://doi.org/10.1016/S0092-8674(00)80205-6

Cheetham, N. J., Kibble, M., Wong, A., Silverwood, R. J., Knuppel, A., Williams, D. M., Hamilton, O. K., Lee, P. H., Bridger Staatz, C., Di Gessa, G., Zhu, J., Katikireddi, S. V., Ploubidis, G. B., Thompson, E. J., Bowyer, R. C., Zhang, X., Abbasian, G., Garcia, M. P., Hart, D., … Steves, C. J. (2023). Antibody levels following vaccination against SARS-CoV-2: associations with post-vaccination infection and risk factors in two UK longitudinal studies. *ELife*, *12*. https://doi.org/10.7554/eLife.80428

Chen, G., Calcaterra, F., Ma, Y., Ping, X., Pontarini, E., Yang, D., Oriolo, F., Yu, Z., Cancellara, A., Mikulak, J., Huang, Y., Della Bella, S., Liu, Y., Biesecker, L. G., Harper, R. L., Dalgard, C. L., Boehm, M., & Mavilio, D. (2023). Derived myeloid lineage induced pluripotent stem as a platform to study human C-C chemokine receptor type 5Δ32 homozygotes. *IScience*, *26*(11), 108331. https://doi.org/10.1016/j.isci.2023.108331

Cheng, L., Hansen, N. F., Zhao, L., Du, Y., Zou, C., Donovan, F. X., Chou, B.-K., Zhou, G., Li, S., Dowey, S. N., Ye, Z., Chandrasekharappa, S. C., Yang, H., Mullikin, J. C., & Liu, P. P. (2012). Low Incidence of DNA Sequence Variation in Human Induced Pluripotent Stem Cells Generated by Nonintegrating Plasmid Expression. *Cell Stem Cell*, *10*(3), 337–344. https://doi.org/10.1016/j.stem.2012.01.005

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., & Debyser, Z. (2003). HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/p75 Protein in Human Cells. *Journal of Biological Chemistry*, *278*(1), 372–381. https://doi.org/10.1074/jbc.M209278200

Chomont, N., El-Far, M., Ancuta, P., Trautmann, L., Procopio, F. A., Yassine-Diab, B., Boucher, G., Boulassel, M.-R., Ghattas, G., Brenchley, J. M., Schacker, T. W., Hill, B. J., Douek, D. C., Routy, J.-P., Haddad, E. K., & Sékaly, R.-P. (2009). HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature Medicine*, *15*(8), 893–900. https://doi.org/10.1038/nm.1972

Chu, M., Zhang, W., Zhang, X., Jiang, W., Huan, X., Meng, X., Zhu, B., Yang, Y., Tao, Y., Tian, T., Lu, Y., Jiang, L., Zhang, L., & Zhuang, X. (2017). HIV-1 CRF01_AE strain is associated with faster HIV/AIDS progression in Jiangsu Province, China. *Scientific Reports*, *7*(1), 1570. https://doi.org/10.1038/s41598-017-01858-2

Chun, T.-W., Engel, D., Berrey, M. M., Shea, T., Corey, L., & Fauci, A. S. (1998). Early establishment of a pool of latently infected, resting CD4 $^{+}$ T cells during primary HIV-1 infection. *Proceedings of the National Academy of Sciences*, *95*(15), 8869–8873. https://doi.org/10.1073/pnas.95.15.8869

Coffin, J. M., & Hughes, S. H. (2021). Clonal Expansion of Infected CD4+ T Cells in People Living with HIV. *Viruses*, *13*(10), 2078. https://doi.org/10.3390/v13102078

Cole, S. R., & Stuart, E. A. (2010). Generalizing Evidence From Randomized Clinical Trials to Target Populations: The ACTG 320 Trial. *American Journal of Epidemiology*, *172*(1), 107–115. https://doi.org/10.1093/aje/kwq084

Compton, A. A., Bruel, T., Porrot, F., Mallet, A., Sachse, M., Euvrard, M., Liang, C., Casartelli, N., & Schwartz, O. (2014). IFITM Proteins Incorporated into HIV-1 Virions Impair Viral Fusion and Spread. *Cell Host & Microbe*, *16*(6), 736–747. https://doi.org/10.1016/j.chom.2014.11.001

Concorde Coordinating Committee. (1994). Concorde: MRC/ANRS randomised double-blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection. *The Lancet*, *343*(8902), 871–881. https://doi.org/10.1016/S0140-6736(94)90006-X

Connor, R. I., Sheridan, K. E., Ceradini, D., Choe, S., & Landau, N. R. (1997). Change in Coreceptor Use Correlates with Disease Progression in HIV-1–Infected Individuals. *The Journal of Experimental Medicine*, *185*(4), 621–628. https://doi.org/10.1084/jem.185.4.621

Conticello, S. G., Harris, R. S., & Neuberger, M. S. (2003). The Vif Protein of HIV Triggers Degradation of the Human Antiretroviral DNA Deaminase APOBEC3G. *Current Biology*, *13*(22), 2009–2013. https://doi.org/10.1016/j.cub.2003.10.034

Cooper, A., García, M., Petrovas, C., Yamamoto, T., Koup, R. A., & Nabel, G. J. (2013). HIV-1 causes CD4 cell death through DNA-dependent protein kinase during viral integration. *Nature*, *498*(7454), 376–379. https://doi.org/10.1038/nature12274

Cosnefroy, O., Murray, P. J., & Bishop, K. N. (2016). HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology*, *13*(1), 58. https://doi.org/10.1186/s12977-016-0292-7

Cribier, A., Descours, B., Valadão, A. L. C., Laguette, N., & Benkirane, M. (2013). Phosphorylation of SAMHD1 by Cyclin A2/CDK1 Regulates Its Restriction Activity toward HIV-1. *Cell Reports*, *3*(4), 1036–1043. https://doi.org/10.1016/j.celrep.2013.03.017

Cureton, D. K., Massol, R. H., Saffarian, S., Kirchhausen, T. L., & Whelan, S. P. J. (2009). Vesicular Stomatitis Virus Enters Cells through Vesicles Incompletely Coated with Clathrin That Depend upon Actin for Internalization. *PLoS Pathogens*, *5*(4), e1000394. https://doi.org/10.1371/journal.ppat.1000394

Dalgleish, A. G., Beverley, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F., & Weiss, R. A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*, *312*(5996), 763–767. https://doi.org/10.1038/312763a0

Damjanov, I. (2005). The road from teratocarcinoma to human embryonic stem cells. *Stem Cell Reviews*, *1*(3), 273–276. https://doi.org/10.1385/SCR:1:3:273

Damond, F., Worobey, M., Campa, P., Farfara, I., Colin, G., Matheron, S., Brun-Vézinet, F., Robertson, D. L., & Simon, F. (2004). Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification. *AIDS Research and Human Retroviruses*, *20*(6), 666–672. https://doi.org/10.1089/0889222041217392

Day, C. L., Kaufmann, D. E., Kiepiela, P., Brown, J. A., Moodley, E. S., Reddy, S., Mackey, E. W., Miller, J. D., Leslie, A. J., DePierres, C., Mncube, Z., Duraiswamy, J., Zhu, B., Eichbaum, Q., Altfeld, M., Wherry, E. J., Coovadia, H. M., Goulder, P. J. R., Klenerman, P., … Walker, B. D. (2006). PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. *Nature*, *443*(7109), 350–354. https://doi.org/10.1038/nature05115

de Breyne, S., Chamond, N., Décimo, D., Trabaud, M., André, P., Sargueil, B., & Ohlmann, T. (2012). *In vitro* studies reveal that different modes of initiation on HIV-1 mRNA have different levels of requirement for eukaryotic initiation factor 4F. *The FEBS Journal*, *279*(17), 3098–3111. https://doi.org/10.1111/j.1742-4658.2012.08689.x

De Iaco, A., Santoni, F., Vannier, A., Guipponi, M., Antonarakis, S., & Luban, J. (2013). TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm. *Retrovirology*, *10*(1), 20. https://doi.org/10.1186/1742-4690-10-20

Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J. J., Buchbinder, S. P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R.,

Saah, A., Rinaldo, C., Detels, R., & O'Brien, S. J. (1996). Genetic Restriction of HIV-1 Infection and Progression to AIDS by a Deletion Allele of the *CKR5* Structural Gene. *Science*, *273*(5283), 1856–1862. https://doi.org/10.1126/science.273.5283.1856

Deeks, S. G., & Walker, B. D. (2007). Human Immunodeficiency Virus Controllers: Mechanisms of Durable Virus Control in the Absence of Antiretroviral Therapy. *Immunity*, *27*(3), 406–416. https://doi.org/10.1016/j.immuni.2007.08.010

Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Marzio, P. Di., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R., & Landau, N. R. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, *381*(6584), 661–666. https://doi.org/10.1038/381661a0

Derdeyn, C. A., Decker, J. M., Bibollet-Ruche, F., Mokili, J. L., Muldoon, M., Denham, S. A., Heil, M. L., Kasolo, F., Musonda, R., Hahn, B. H., Shaw, G. M., Korber, B. T., Allen, S., & Hunter, E. (2004). Envelope-Constrained Neutralization-Sensitive HIV-1 After Heterosexual Transmission. *Science*, *303*(5666), 2019–2022. https://doi.org/10.1126/science.1093137

Desai, T. M., Marin, M., Chin, C. R., Savidis, G., Brass, A. L., & Melikyan, G. B. (2014). IFITM3 Restricts Influenza A Virus Entry by Blocking the Formation of Fusion Pores following Virus-Endosome Hemifusion. *PLoS Pathogens*, *10*(4), e1004048. https://doi.org/10.1371/journal.ppat.1004048

Descours, B., Cribier, A., Chable-Bessia, C., Ayinde, D., Rice, G., Crow, Y., Yatim, A., Schwartz, O., Laguette, N., & Benkirane, M. (2012). SAMHD1 restricts HIV-1 reverse transcription in quiescent CD4+T-cells. *Retrovirology*, *9*(1), 87. https://doi.org/10.1186/1742-4690-9-87

Dharan, A., Bachmann, N., Talley, S., Zwikelmaier, V., & Campbell, E. M. (2020). Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. *Nature Microbiology*, *5*(9), 1088–1095. https://doi.org/10.1038/s41564-020-0735-8

Di Nunzio, F., Danckaert, A., Fricke, T., Perez, P., Fernandez, J., Perret, E., Roux, P., Shorte, S., Charneau, P., Diaz-Griffero, F., & Arhel, N. J. (2012). Human Nucleoporins Promote HIV-1 Docking at the Nuclear Pore, Nuclear Import and Integration. *PLoS ONE*, *7*(9), e46037. https://doi.org/10.1371/journal.pone.0046037

Dicks, M. D. J., Betancor, G., Jimenez-Guardeño, J. M., Pessel-Vivares, L., Apolonia, L., Goujon, C., & Malim, M. H. (2018). Multiple components of the nuclear pore complex interact with the amino-terminus of MX2 to facilitate HIV-1 restriction. *PLOS Pathogens*, *14*(11), e1007408. https://doi.org/10.1371/journal.ppat.1007408

Dittmann, M., Hoffmann, H.-H., Scull, M. A., Gilmore, R. H., Bell, K. L., Ciancanelli, M., Wilson, S. J., Crotta, S., Yu, Y., Flatley, B., Xiao, J. W., Casanova, J.-L., Wack, A., Bieniasz, P. D., & Rice, C. M. (2015). A Serpin Shapes the Extracellular Environment to Prevent Influenza A Virus Maturation. *Cell*, *160*(4), 631–643. https://doi.org/10.1016/j.cell.2015.01.040

Doitsh, G., Cavrois, M., Lassen, K. G., Zepeda, O., Yang, Z., Santiago, M. L., Hebbeler, A. M., & Greene, W. C. (2010). Abortive HIV infection mediates CD4 T cell depletion and inflammation in human lymphoid tissue. *Cell*, *143*(5), 789–801. https://doi.org/10.1016/j.cell.2010.11.001

dos Reis, R. S., Sant, S., Keeney, H., Wagner, M. C. E., & Ayyavoo, V. (2020). Modeling HIV-1 neuropathogenesis using three-dimensional human brain organoids (hBORGs) with HIV-1 infected microglia. *Scientific Reports*, *10*(1), 15209. https://doi.org/10.1038/s41598-020-72214-0

Doyle, T., Goujon, C., & Malim, M. H. (2015). HIV-1 and interferons: who's interfering with whom? *Nature Reviews Microbiology*, *13*(7), 403–413. https://doi.org/10.1038/nrmicro3449

Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P., & Paxton, W. A. (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*, *381*(6584), 667–673. https://doi.org/10.1038/381667a0

Dubé, M., Paquay, C., Roy, B. B., Bego, M. G., Mercier, J., & Cohen, É. A. (2011). HIV-1 Vpu Antagonizes BST-2 by Interfering Mainly with the Trafficking of Newly Synthesized BST-2 to the Cell Surface. *Traffic*, *12*(12), 1714–1729. https://doi.org/10.1111/j.1600-0854.2011.01277.x

Dupont, L., Snell, L. B., Graham, C., Seow, J., Merrick, B., Lechmere, T., Maguire, T. J. A., Hallett, S. R., Pickering, S., Charalampous, T., Alcolea-Medina, A., Huettner, I., Jimenez-Guardeño, J. M., Acors, S., Almeida, N., Cox, D., Dickenson, R. E., Galao, R. P., Kouphou, N., … Doores, K. J. (2021). Neutralizing antibody activity in convalescent sera from infection in humans with SARS-CoV-2 and variants of concern. *Nature Microbiology*, *6*(11), 1433–1442. https://doi.org/10.1038/s41564-021-00974-0

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, *4*(8), 1184–1191. https://doi.org/10.1038/nprot.2009.97

El-Sayed Moustafa, J. S., Jackson, A. U., Brotman, S. M., Guan, L., Villicaña, S., Roberts, A. L., Zito, A., Bonnycastle, L., Erdos, M. R., Narisu, N., Stringham, H. M., Welch, R., Yan, T., Lakka, T., Parker, S., Tuomilehto, J., Seow, J., Graham, C., Huettner, I., … Small, K. S. (2022). ACE2 expression in adipose tissue is associated with cardio-metabolic risk factors and cell type composition—implications for COVID-19. *International Journal of Obesity*, *46*(8), 1478–1486. https://doi.org/10.1038/s41366-022-01136-w

Engelman, A., Bushman, F. D., & Craigie, R. (1993). Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *The EMBO Journal*, *12*(8), 3269–3275. https://doi.org/10.1002/j.1460-2075.1993.tb05996.x

Fan, Y., Mok, C. K.-P., Chan, M. C. W., Zhang, Y., Nal, B., Kien, F., Bruzzone, R., & Sanyal, S. (2017). Cell Cycle-independent Role of Cyclin D3 in Host Restriction of Influenza Virus Infection. *The Journal of Biological Chemistry*, *292*(12), 5070–5088. https://doi.org/10.1074/jbc.M117.776112

Fassati, A., & Goff, S. P. (2001). Characterization of Intracellular Reverse Transcription Complexes of Human Immunodeficiency Virus Type 1. *Journal of Virology*, *75*(8), 3626–3635. https://doi.org/10.1128/JVI.75.8.3626-3635.2001

Fellay, J., Ge, D., Shianna, K. V., Colombo, S., Ledergerber, B., Cirulli, E. T., Urban, T. J., Zhang, K., Gumbs, C. E., Smith, J. P., Castagna, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Günthard, H. F., Mallal, S., Mussini, C., Dalmau, J., Martinez-Picado, J., … Goldstein, D. B. (2009). Common Genetic Variation and the Control of HIV-1 in Humans. *PLoS Genetics*, *5*(12), e1000791. https://doi.org/10.1371/journal.pgen.1000791

Fellay, J., Shianna, K. V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Francioli, P., Mallal, S., Martinez-Picado, J., Miro, J. M., Obel, N., Smith, J. P., … Goldstein, D. B. (2007). A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1. *Science*, *317*(5840), 944–947. https://doi.org/10.1126/science.1143767

Feng, Y., Broder, C. C., Kennedy, P. E., & Berger, E. A. (1996). HIV-1 Entry Cofactor: Functional cDNA Cloning of a Seven-Transmembrane, G Protein-Coupled Receptor. *Science*, *272*(5263), 872–877. https://doi.org/10.1126/science.272.5263.872

Fernández-Ortega, C., Ramírez, A., Casillas, D., Paneque, T., Ubieta, R., Dubed, M., Navea, L., Castellanos-Serra, L., Duarte, C., Falcon, V., Reyes, O., Garay, H., Silva, E., Noa, E., Ramos, Y., Besada, V., & Betancourt, L. (2016). Identification of Vimentin as a Potential Therapeutic Target against HIV Infection. *Viruses*, *8*(6), 98. https://doi.org/10.3390/v8060098

Ferre, A. L., Hunt, P. W., Critchfield, J. W., Young, D. H., Morris, M. M., Garcia, J. C., Pollard, R. B., Yee, H. F., Martin, J. N., Deeks, S. G., & Shacklett, B. L. (2009). Mucosal immune responses to HIV-1 in elite controllers: a potential correlate of immune control. *Blood*, *113*(17), 3978–3989. https://doi.org/10.1182/blood-2008-10-182709

Fideli, Ü. S., Allen, S. A., Musonda, R., Trask, S., Hahn, B. H., Weiss, H., Mulenga, J., Kasolo, F., Vermund, S. H., & Aldrovandi, G. M. (2001). Virologic and Immunologic Determinants of Heterosexual Transmission of Human Immunodeficiency Virus Type 1 in Africa. *AIDS Research and Human Retroviruses*, *17*(10), 901–910. https://doi.org/10.1089/088922201750290023

Fletcher, C. V., Staskus, K., Wietgrefe, S. W., Rothenberger, M., Reilly, C., Chipman, J. G., Beilman, G. J., Khoruts, A., Thorkelson, A., Schmidt, T. E., Anderson, J., Perkey, K., Stevenson, M., Perelson, A. S., Douek, D. C., Haase, A. T., & Schacker, T. W. (2014). Persistent HIV-1 replication is associated with lower antiretroviral drug concentrations in lymphatic tissues. *Proceedings of the National Academy of Sciences*, *111*(6), 2307–2312. https://doi.org/10.1073/pnas.1318249111

Foster, T. L., Wilson, H., Iyer, S. S., Coss, K., Doores, K., Smith, S., Kellam, P., Finzi, A., Borrow, P., Hahn, B. H., & Neil, S. J. D. (2016). Resistance of Transmitted Founder HIV-1 to IFITM-Mediated Restriction. *Cell Host & Microbe*, *20*(4), 429–442. https://doi.org/10.1016/j.chom.2016.08.006

Fraser, C., Lythgoe, K., Leventhal, G. E., Shirreff, G., Hollingsworth, T. D., Alizon, S., & Bonhoeffer, S. (2014). Virulence and Pathogenesis of HIV-1 Infection: An Evolutionary Perspective. *Science*, *343*(6177). https://doi.org/10.1126/science.1243727

Freed, E. O. (2015). HIV-1 assembly, release and maturation. *Nature Reviews Microbiology*, *13*(8), 484–496. https://doi.org/10.1038/nrmicro3490

Fujii, K., Munshi, U. M., Ablan, S. D., Demirov, D. G., Soheilian, F., Nagashima, K., Stephen, A. G., Fisher, R. J., & Freed, E. O. (2009). Functional role of Alix in HIV-1 replication. *Virology*, *391*(2), 284–292. https://doi.org/10.1016/j.virol.2009.06.016

Fukazawa, Y., Lum, R., Okoye, A. A., Park, H., Matsuda, K., Bae, J. Y., Hagen, S. I., Shoemaker, R., Deleage, C., Lucero, C., Morcock, D., Swanson, T., Legasse, A. W., Axthelm, M. K., Hesselgesser, J., Geleziunas, R., Hirsch, V. M., Edlefsen, P. T., Piatak, M., … Picker, L. J. (2015). B cell follicle sanctuary permits persistent productive simian immunodeficiency virus infection in elite controllers. *Nature Medicine*, *21*(2), 132–139. https://doi.org/10.1038/nm.3781

Fusaki, N., Ban, H., Nishiyama, A., Seaji, K., & Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy, Series B*, *85*(8), 348–362. https://doi.org/10.2183/pjab.85.348

Gaiha, G. D., Rossin, E. J., Urbach, J., Landeros, C., Collins, D. R., Nwonu, C., Muzhingi, I., Anahtar, M. N., Waring, O. M., Piechocka-Trocha, A., Waring, M., Worrall, D. P.,

Ghebremichael, M. S., Newman, R. M., Power, K. A., Allen, T. M., Chodosh, J., & Walker, B. D. (2019). Structural topology defines protective CD8 $^{+}$ T cell epitopes in the HIV proteome. *Science*, *364*(6439), 480–484. https://doi.org/10.1126/science.aav5095

Gallo, R. C., Sarin, P. S., Gelmann, E. P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V. S., Mann, D., Sidhu, G. D., Stahl, R. E., Zolla-Pazner, S., Leibowitch, J., & Popovic, M. (1983). Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science (New York, N.Y.)*, *220*(4599), 865–867. https://doi.org/10.1126/science.6601823

Ganser, B. K., Li, S., Klishko, V. Y., Finch, J. T., & Sundquist, W. I. (1999). Assembly and Analysis of Conical Models for the HIV-1 Core. *Science*, *283*(5398), 80–83. https://doi.org/10.1126/science.283.5398.80

Gao, D., Wu, J., Wu, Y.-T., Du, F., Aroh, C., Yan, N., Sun, L., & Chen, Z. J. (2013). Cyclic GMP-AMP Synthase Is an Innate Immune Sensor of HIV and Other Retroviruses. *Science*, *341*(6148), 903–906. https://doi.org/10.1126/science.1240933

Garber, K. (2015). RIKEN suspends first clinical trial involving induced pluripotent stem cells. *Nature Biotechnology*, *33*(9), 890–891. https://doi.org/10.1038/nbt0915-890

Gelderblom, H. R. (1991). Assembly and morphology of HIV: potential effect of structure on viral function. *AIDS (London, England)*, *5*(6), 617–637.

Ghosh, Z., Wilson, K. D., Wu, Y., Hu, S., Quertermous, T., & Wu, J. C. (2010). Persistent Donor Cell Gene Expression among Human Induced Pluripotent Stem Cells Contributes to Differences with Human Embryonic Stem Cells. *PLoS ONE*, *5*(2), e8975. https://doi.org/10.1371/journal.pone.0008975

Goldstone, D. C., Ennis-Adeniran, V., Hedden, J. J., Groom, H. C. T., Rice, G. I., Christodoulou, E., Walker, P. A., Kelly, G., Haire, L. F., Yap, M. W., de Carvalho, L. P. S., Stoye, J. P., Crow, Y. J., Taylor, I. A., & Webb, M. (2011). HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature*, *480*(7377), 379–382. https://doi.org/10.1038/nature10623

Goonetilleke, N., Liu, M. K. P., Salazar-Gonzalez, J. F., Ferrari, G., Giorgi, E., Ganusov, V. V., Keele, B. F., Learn, G. H., Turnbull, E. L., Salazar, M. G., Weinhold, K. J., Moore, S., Letvin, N., Haynes, B. F., Cohen, M. S., Hraber, P., Bhattacharya, T., Borrow, P., Perelson, A. S., … McMichael, A. J. (2009). The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *Journal of Experimental Medicine*, *206*(6), 1253–1272. https://doi.org/10.1084/jem.20090365

Gottlieb, M., Schanker, H., Fan, P., Saxon, A., & Weisman, J. (1981). Epidemiologic Notes and Reports: Pneumocystis Pneumonia --- Los Angeles. *Morbidity and Mortality Weekly Report (MMWR)*, *30*(21), 1–3.

Goujon, C., Greenbury, R. A., Papaioannou, S., Doyle, T., & Malim, M. H. (2015). A Triple-Arginine Motif in the Amino-Terminal Domain and Oligomerization Are Required for HIV-1 Inhibition by Human MX2. *Journal of Virology*, *89*(8), 4676–4680. https://doi.org/10.1128/JVI.00169-15

Goujon, C., Moncorgé, O., Bauby, H., Doyle, T., Barclay, W. S., & Malim, M. H. (2014). Transfer of the Amino-Terminal Nuclear Envelope Targeting Domain of Human MX2 Converts MX1 into an HIV-1 Resistance Factor. *Journal of Virology*, *88*(16), 9017–9026. https://doi.org/10.1128/JVI.01269-14

Goujon, C., Moncorgé, O., Bauby, H., Doyle, T., Ward, C. C., Schaller, T., Hué, S., Barclay, W. S., Schulz, R., & Malim, M. H. (2013). Human MX2 is an interferon-induced post-entry

inhibitor of HIV-1 infection. *Nature*, *502*(7472), 559–562. https://doi.org/10.1038/nature12542

Goujon, C., Schaller, T., Galão, R. P., Amie, S. M., Kim, B., Olivieri, K., Neil, S. J., & Malim, M. H. (2013). Evidence for IFNα-induced, SAMHD1-independent inhibitors of early HIV-1 infection. *Retrovirology*, *10*(1), 23. https://doi.org/10.1186/1742-4690-10-23

Graham, C., Seow, J., Huettner, I., Khan, H., Kouphou, N., Acors, S., Winstone, H., Pickering, S., Galao, R. P., Dupont, L., Lista, M. J., Jimenez-Guardeño, J. M., Laing, A. G., Wu, Y., Joseph, M., Muir, L., van Gils, M. J., Ng, W. M., Duyvesteyn, H. M. E., … Doores, K. J. (2021). Neutralization potency of monoclonal antibodies recognizing dominant and subdominant epitopes on SARS-CoV-2 Spike is impacted by the B.1.1.7 variant. *Immunity*, *54*(6), 1276-1289.e6. https://doi.org/10.1016/j.immuni.2021.03.023

Guadalupe, M., Reay, E., Sankaran, S., Prindiville, T., Flamm, J., McNeil, A., & Dandekar, S. (2003). Severe CD4 $^{+}$ T-Cell Depletion in Gut Lymphoid Tissue during Primary Human Immunodeficiency Virus Type 1 Infection and Substantial Delay in Restoration following Highly Active Antiretroviral Therapy. *Journal of Virology*, *77*(21), 11708–11717. https://doi.org/10.1128/JVI.77.21.11708-11717.2003

Gumbs, S. B. H., Berdenis van Berlekom, A., Kübler, R., Schipper, P. J., Gharu, L., Boks, M. P., Ormel, P. R., Wensing, A. M. J., de Witte, L. D., & Nijhuis, M. (2022). Characterization of HIV-1 Infection in Microglia-Containing Human Cerebral Organoids. *Viruses*, *14*(4), 829. https://doi.org/10.3390/v14040829

Gupta, R. K., & Mlcochova, P. (2022). Cyclin D3 restricts SARS-CoV-2 envelope incorporation into virions and interferes with viral spread. *The EMBO Journal*, *41*(22), e111653. https://doi.org/10.15252/embj.2022111653

Gupta, R. K., Peppa, D., Hill, A. L., Gálvez, C., Salgado, M., Pace, M., McCoy, L. E., Griffith, S. A., Thornhill, J., Alrubayyi, A., Huyveneers, L. E. P., Nastouli, E., Grant, P., Edwards, S. G., Innes, A. J., Frater, J., Nijhuis, M., Wensing, A. M. J., Martinez-Picado, J., & Olavarria, E. (2020). Evidence for HIV-1 cure after CCR5Δ32/Δ32 allogeneic haemopoietic stem-cell transplantation 30 months post analytical treatment interruption: a case report. *The Lancet. HIV*, *7*(5), e340–e347. https://doi.org/10.1016/S2352-3018(20)30069-2

Hallenberger, S., Bosch, V., Angliker, H., Shaw, E., Klenk, H. D., & Garten, W. (1992). Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature*, *360*(6402), 358–361. https://doi.org/10.1038/360358a0

Hamer, D. (2004). Can HIV be Cured? Mechanisms of HIV Persistence and Strategies to Combat It. *Current HIV Research*, *2*(2), 99–111. https://doi.org/10.2174/1570162043484915

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S., & Merigan, T. C. (1996). A Trial Comparing Nucleoside Monotherapy with Combination Therapy in HIV-Infected Adults with CD4 Cell Counts from 200 to 500 per Cubic Millimeter. *New England Journal of Medicine*, *335*(15), 1081–1090. https://doi.org/10.1056/NEJM199610103351501

Han, F., Li, X., Song, D., Jiang, S., Xu, Q., & Zhang, Y. (2015). SCNT versus iPSCs: proteins and small molecules in reprogramming. *The International Journal of Developmental Biology*, *59*(4-5–6), 179–186. https://doi.org/10.1387/ijdb.150042fh

Haque, M., Lei, F., Xiong, X., Ren, Y., Peng, H.-Y., Wang, L., Kumar, A., Das, J. K., & Song, J. (2021). Stem Cell-Derived Viral Antigen-Specific T Cells Suppress HIV Replication and

PD-1 Expression on CD4+ T Cells. *Viruses*, *13*(5), 753. https://doi.org/10.3390/v13050753
Hare, S., Gupta, S. S., Valkov, E., Engelman, A., & Cherepanov, P. (2010). Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature*, *464*(7286), 232–236. https://doi.org/10.1038/nature08784
Hargrave, A., Mustafa, A. S., Hanif, A., Tunio, J. H., & Hanif, S. N. M. (2021). Current Status of HIV-1 Vaccines. *Vaccines*, *9*(9). https://doi.org/10.3390/vaccines9091026
Hatziioannou, T., Perez-Caballero, D., Cowan, S., & Bieniasz, P. D. (2005). Cyclophilin Interactions with Incoming Human Immunodeficiency Virus Type 1 Capsids with Opposing Effects on Infectivity in Human Cells. *Journal of Virology*, *79*(1), 176–183. https://doi.org/10.1128/JVI.79.1.176-183.2005
Hecht, F. M., Hartogensis, W., Bragg, L., Bacchetti, P., Atchison, R., Grant, R., Barbour, J., & Deeks, S. G. (2010). HIV RNA level in early infection is predicted by viral load in the transmission source. *AIDS*, *24*(7), 941–945. https://doi.org/10.1097/QAD.0b013e328337b12e
Heurtier, V., Owens, N., Gonzalez, I., Mueller, F., Proux, C., Mornico, D., Clerc, P., Dubois, A., & Navarro, P. (2019). The molecular logic of Nanog-induced self-renewal in mouse embryonic stem cells. *Nature Communications*, *10*(1), 1109. https://doi.org/10.1038/s41467-019-09041-z
Hiatt, J., Hultquist, J. F., McGregor, M. J., Bouhaddou, M., Leenay, R. T., Simons, L. M., Young, J. M., Haas, P., Roth, T. L., Tobin, V., Wojcechowskyj, J. A., Woo, J. M., Rathore, U., Cavero, D. A., Shifrut, E., Nguyen, T. T., Haas, K. M., Malik, H. S., Doudna, J. A., … Krogan, N. J. (2022). A functional map of HIV-host interactions in primary human T cells. *Nature Communications*, *13*(1), 1752. https://doi.org/10.1038/s41467-022-29346-w
Higaki, K., Hirao, M., Kawana-Tachikawa, A., Iriguchi, S., Kumagai, A., Ueda, N., Bo, W., Kamibayashi, S., Watanabe, A., Nakauchi, H., Suzuki, K., & Kaneko, S. (2018). Generation of HIV-Resistant Macrophages from IPSCs by Using Transcriptional Gene Silencing and Promoter-Targeted RNA. *Molecular Therapy - Nucleic Acids*, *12*, 793–804. https://doi.org/10.1016/j.omtn.2018.07.017
Hirsch, V. M., Olmsted, R. A., Murphey-Corb, M., Purcell, R. H., & Johnson, P. R. (1989). An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature*, *339*(6223), 389–392. https://doi.org/10.1038/339389a0
Hladik, F., Liu, H., Speelmon, E., Livingston-Rosanoff, D., Wilson, S., Sakchalathorn, P., Hwangbo, Y., Greene, B., Zhu, T., & McElrath, M. J. (2005). Combined Effect of CCR5-Δ32 Heterozygosity and the CCR5 Promoter Polymorphism −2459 A/G on CCR5 Expression and Resistance to Human Immunodeficiency Virus Type 1 Transmission. *Journal of Virology*, *79*(18), 11677–11684. https://doi.org/10.1128/JVI.79.18.11677-11684.2005
Holmes, D. (2012). FDA paves the way for pre-exposure HIV prophylaxis. *Lancet (London, England)*, *380*(9839), 325. https://doi.org/10.1016/s0140-6736(12)61235-5
Horton, H., Frank, I., Baydo, R., Jalbert, E., Penn, J., Wilson, S., McNevin, J. P., McSweyn, M. D., Lee, D., Huang, Y., De Rosa, S. C., & McElrath, M. J. (2006). Preservation of T Cell Proliferation Restricted by Protective HLA Alleles Is Critical for Immune Control of HIV-1 Infection. *The Journal of Immunology*, *177*(10), 7406–7415. https://doi.org/10.4049/jimmunol.177.10.7406
Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115. https://doi.org/10.1186/gb-2013-14-10-r115

Hrecka, K., Hao, C., Gierszewska, M., Swanson, S. K., Kesik-Brodacka, M., Srivastava, S., Florens, L., Washburn, M. P., & Skowronski, J. (2011). Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature*, *474*(7353), 658–661. https://doi.org/10.1038/nature10195

Huang, Y., Paxton, W. A., Wolinsky, S. M., Neumann, A. U., Zhang, L., He, T., Kang, S., Ceradini, D., Jin, Z., Yazdanbakhsh, K., Kunstman, K., Erickson, D., Dragon, E., Landau, N. R., Phair, J., Ho, D. D., & Koup, R. A. (1996). The role of a mutant CCR5 allele in HIV–1 transmission and disease progression. *Nature Medicine*, *2*(11), 1240–1243. https://doi.org/10.1038/nm1196-1240

Hurley, J. H., & Emr, S. D. (2006). THE ESCRT COMPLEXES: Structure and Mechanism of a Membrane-Trafficking Network. *Annual Review of Biophysics and Biomolecular Structure*, *35*(1), 277–298. https://doi.org/10.1146/annurev.biophys.35.040405.102126

Hurley, J. H., & Hanson, P. I. (2010). Membrane budding and scission by the ESCRT machinery: it's all in the neck. *Nature Reviews Molecular Cell Biology*, *11*(8), 556–566. https://doi.org/10.1038/nrm2937

Hussein, S. M., Batada, N. N., Vuoristo, S., Ching, R. W., Autio, R., Närvä, E., Ng, S., Sourour, M., Hämäläinen, R., Olsson, C., Lundin, K., Mikkola, M., Trokovic, R., Peitz, M., Brüstle, O., Bazett-Jones, D. P., Alitalo, K., Lahesmaa, R., Nagy, A., & Otonkoski, T. (2011). Copy number variation and selection during reprogramming to pluripotency. *Nature*, *471*(7336), 58–62. https://doi.org/10.1038/nature09871

Hütter, G., Nowak, D., Mossner, M., Ganepola, S., Müssig, A., Allers, K., Schneider, T., Hofmann, J., Kücherer, C., Blau, O., Blau, I. W., Hofmann, W. K., & Thiel, E. (2009). Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *The New England Journal of Medicine*, *360*(7), 692–698. https://doi.org/10.1056/NEJMoa0802905

Iizuka-Koga, M., Asashima, H., Ando, M., Lai, C.-Y., Mochizuki, S., Nakanishi, M., Nishimura, T., Tsuboi, H., Hirota, T., Takahashi, H., Matsumoto, I., Otsu, M., & Sumida, T. (2017). Functional Analysis of Dendritic Cells Generated from T-iPSCs from CD4+ T Cell Clones of Sjögren's Syndrome. *Stem Cell Reports*, *8*(5), 1155–1163. https://doi.org/10.1016/j.stemcr.2017.04.010

Ingram, Z., Fischer, D. K., & Ambrose, Z. (2021). Disassembling the Nature of Capsid: Biochemical, Genetic, and Imaging Approaches to Assess HIV-1 Capsid Functions. *Viruses*, *13*(11), 2237. https://doi.org/10.3390/v13112237

International Committee on Taxonomy of Viruses. (2022). *Virus Taxonomy: 2022 Release*.

International HIV Controllers Study, Pereyra, F., Jia, X., McLaren, P. J., Telenti, A., de Bakker, P. I. W., Walker, B. D., Ripke, S., Brumme, C. J., Pulit, S. L., Carrington, M., Kadie, C. M., Carlson, J. M., Heckerman, D., Graham, R. R., Plenge, R. M., Deeks, S. G., Gianniny, L., Crawford, G., … Zhao, M. (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science (New York, N.Y.)*, *330*(6010), 1551–1557. https://doi.org/10.1126/science.1195271

Isel, C., Lanchy, J. M., Le Grice, S. F., Ehresmann, C., Ehresmann, B., & Marquet, R. (1996). Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA3Lys. *The EMBO Journal*, *15*(4), 917–924.

Isogai, S., Yamamoto, N., Hiramatsu, N., Goto, Y., Hayashi, M., Kondo, M., & Imaizumi, K. (2018). Preparation of Induced Pluripotent Stem Cells Using Human Peripheral Blood

Monocytes. *Cellular Reprogramming*, *20*(6), 347–355. https://doi.org/10.1089/cell.2018.0024

Iwamoto, Y., Seki, Y., Taya, K., Tanaka, M., Iriguchi, S., Miyake, Y., Nakayama, E. E., Miura, T., Shioda, T., Akari, H., Takaori-Kondo, A., & Kaneko, S. (2021). Generation of macrophages with altered viral sensitivity from genome-edited rhesus macaque iPSCs to model human disease. *Molecular Therapy - Methods & Clinical Development*, *21*, 262–273. https://doi.org/10.1016/j.omtm.2021.03.008

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., & Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, *331*(6153), 280–283. https://doi.org/10.1038/331280a0

Jakobsen, M. R., Bak, R. O., Andersen, A., Berg, R. K., Jensen, S. B., Jin, T., Laustsen, A., Hansen, K., Østergaard, L., Fitzgerald, K. A., Xiao, T. S., Mikkelsen, J. G., Mogensen, T. H., & Paludan, S. R. (2013). IFI16 senses DNA forms of the lentiviral replication cycle and controls HIV-1 replication. *Proceedings of the National Academy of Sciences*, *110*(48). https://doi.org/10.1073/pnas.1311669110

James, L. C., Keeble, A. H., Khan, Z., Rhodes, D. A., & Trowsdale, J. (2007). Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proceedings of the National Academy of Sciences*, *104*(15), 6200–6205. https://doi.org/10.1073/pnas.0609174104

Jerebtsova, M., Kumari, N., Xu, M., Melo, G. B. A. de, Niu, X., Jeang, K.-T., & Nekhai, S. (2012). HIV-1 Resistant CDK2-Knockdown Macrophage-Like Cells Generated from 293T Cell-Derived Human Induced Pluripotent Stem Cells. *Biology*, *1*(2), 175–195. https://doi.org/10.3390/biology1020175

Jia, F., Wilson, K. D., Sun, N., Gupta, D. M., Huang, M., Li, Z., Panetta, N. J., Chen, Z. Y., Robbins, R. C., Kay, M. A., Longaker, M. T., & Wu, J. C. (2010). A nonviral minicircle vector for deriving human iPS cells. *Nature Methods*, *7*(3), 197–199. https://doi.org/10.1038/nmeth.1426

Jimenez-Guardeño, J. M., Apolonia, L., Betancor, G., & Malim, M. H. (2019). Immunoproteasome activation enables human TRIM5α restriction of HIV-1. *Nature Microbiology*, *4*(6), 933–940. https://doi.org/10.1038/s41564-019-0402-0

Johannesson, B., Sagi, I., Gore, A., Paull, D., Yamada, M., Golan-Lev, T., Li, Z., LeDuc, C., Shen, Y., Stern, S., Xu, N., Ma, H., Kang, E., Mitalipov, S., Sauer, M. V., Zhang, K., Benvenisty, N., & Egli, D. (2014). Comparable Frequencies of Coding Mutations and Loss of Imprinting in Human Pluripotent Cells Derived by Nuclear Transfer and Defined Factors. *Cell Stem Cell*, *15*(5), 634–642. https://doi.org/10.1016/j.stem.2014.10.002

Johannsdottir, H. K., Mancini, R., Kartenbeck, J., Amato, L., & Helenius, A. (2009). Host Cell Factors and Functions Involved in Vesicular Stomatitis Virus Entry. *Journal of Virology*, *83*(1), 440–453. https://doi.org/10.1128/JVI.01864-08

Jolly, C., Mitar, I., & Sattentau, Q. J. (2007). Requirement for an Intact T-Cell Actin and Tubulin Cytoskeleton for Efficient Assembly and Spread of Human Immunodeficiency Virus Type 1. *Journal of Virology*, *81*(11), 5547–5560. https://doi.org/10.1128/JVI.01469-06

Jouvenet, N., Zhadina, M., Bieniasz, P. D., & Simon, S. M. (2011). Dynamics of ESCRT protein recruitment during retroviral assembly. *Nature Cell Biology*, *13*(4), 394–401. https://doi.org/10.1038/ncb2207

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., & Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature*, *458*(7239), 771–775. https://doi.org/10.1038/nature07864

Kalpana, G. V., Marmon, S., Wang, W., Crabtree, G. R., & Goff, S. P. (1994). Binding and Stimulation of HIV-1 Integrase by a Human Homolog of Yeast Transcription Factor SNF5. *Science*, *266*(5193), 2002–2006. https://doi.org/10.1126/science.7801128

Kambal, A., Mitchell, G., Cary, W., Gruenloh, W., Jung, Y., Kalomoiris, S., Nacey, C., McGee, J., Lindsey, M., Fury, B., Bauer, G., Nolta, J. A., & Anderson, J. S. (2011). Generation of HIV-1 Resistant and Functional Macrophages From Hematopoietic Stem Cell–derived Induced Pluripotent Stem Cells. *Molecular Therapy*, *19*(3), 584–593. https://doi.org/10.1038/mt.2010.269

Kane, M., Yadav, S. S., Bitzegeio, J., Kutluay, S. B., Zang, T., Wilson, S. J., Schoggins, J. W., Rice, C. M., Yamashita, M., Hatziioannou, T., & Bieniasz, P. D. (2013). MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature*, *502*(7472), 563–566. https://doi.org/10.1038/nature12653

Kang, H., Minder, P., Park, M. A., Mesquitta, W.-T., Torbett, B. E., & Slukvin, I. I. (2015). CCR5 Disruption in Induced Pluripotent Stem Cells Using CRISPR/Cas9 Provides Selective Resistance of Immune Cells to CCR5-tropic HIV-1 Virus. *Molecular Therapy - Nucleic Acids*, *4*, e268. https://doi.org/10.1038/mtna.2015.42

Kao, S.-Y., Calman, A. F., Luciw, P. A., & Peterlin, B. M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature*, *330*(6147), 489–493. https://doi.org/10.1038/330489a0

Karami, Z., Moradi, S., Eidi, A., Soleimani, M., & Jafarian, A. (2022). Induced pluripotent stem cells: Generation methods and a new perspective in COVID-19 research. *Frontiers in Cell and Developmental Biology*, *10*, 1050856. https://doi.org/10.3389/fcell.2022.1050856

Kaslow, R. A., Carrington, M., Apple, R., Park, L., Muñoz, A., Saah, A. J., Goedert, J. J., Winkler, C., O'Brien, S. J., Rinaldo, C., Detels, R., Blattner, W., Phair, J., Erlich, H., & Mann, D. L. (1996). Influence of combinations of human major histocompatibility complex genes on the course of HIV–1 infection. *Nature Medicine*, *2*(4), 405–411. https://doi.org/10.1038/nm0496-405

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., & Rinaldo, C. R. (1987). The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology*, *126*(2), 310–318. https://doi.org/10.1093/aje/126.2.310

Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., Sun, C., Grayson, T., Wang, S., Li, H., Wei, X., Jiang, C., Kirchherr, J. L., Gao, F., Anderson, J. A., Ping, L.-H., Swanstrom, R., Tomaras, G. D., Blattner, W. A., … Shaw, G. M. (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences*, *105*(21), 7552–7557. https://doi.org/10.1073/pnas.0802203105

Keele, B. F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M. L., Bibollet-Ruche, F., Chen, Y., Wain, L. V, Liegeois, F., Loul, S., Ngole, E. M., Bienvenue, Y., Delaporte, E., Brookfield, J. F. Y., Sharp, P. M., Shaw, G. M., Peeters, M., & Hahn, B. H. (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science (New York, N.Y.)*, *313*(5786), 523–526. https://doi.org/10.1126/science.1126531

Kelleher, A. D., Long, C., Holmes, E. C., Allen, R. L., Wilson, J., Conlon, C., Workman, C., Shaunak, S., Olson, K., Goulder, P., Brander, C., Ogg, G., Sullivan, J. S., Dyer, W., Jones, I., McMichael, A. J., Rowland-Jones, S., & Phillips, R. E. (2001). Clustered Mutations in HIV-1 Gag Are Consistently Required for Escape from Hla-B27–Restricted Cytotoxic T

Lymphocyte Responses. *The Journal of Experimental Medicine*, *193*(3), 375–386. https://doi.org/10.1084/jem.193.3.375

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, *50*(9), 1219–1224. https://doi.org/10.1038/s41588-018-0183-z

Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D., Casale, F. P., Culley, O. J., Danecek, P., Faulconbridge, A., Harrison, P. W., Kathuria, A., McCarthy, D., McCarthy, S. A., Meleckyte, R., Memari, Y., Moens, N., … Gaffney, D. J. (2017). Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*, *546*(7658), 370–375. https://doi.org/10.1038/nature22403

Kim, K., Dauphin, A., Komurlu, S., McCauley, S. M., Yurkovetskiy, L., Carbone, C., Diehl, W. E., Strambio-De-Castillia, C., Campbell, E. M., & Luban, J. (2019). Cyclophilin A protects HIV-1 from restriction by human TRIM5α. *Nature Microbiology*, *4*(12), 2044–2051. https://doi.org/10.1038/s41564-019-0592-5

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M. J., Ji, H., Ehrlich, L. I. R., Yabuuchi, A., Takeuchi, A., Cunniff, K. C., Hongguang, H., Mckinney-Freeman, S., Naveiras, O., Yoon, T. J., Irizarry, R. A., Jung, N., … Daley, G. Q. (2010). Epigenetic memory in induced pluripotent stem cells. *Nature*, *467*(7313), 285–290. https://doi.org/10.1038/nature09342

Kim, K.-P., Thurston, A., Mummery, C., Ward-van Oostwaard, D., Priddle, H., Allegrucci, C., Denning, C., & Young, L. (2007). Gene-specific vulnerability to imprinting variability in human embryonic stem cell lines. *Genome Research*, *17*(12), 1731–1742. https://doi.org/10.1101/gr.6609207

Kimbrel, E. A., & Lanza, R. (2015). Current status of pluripotent stem cells: moving the first therapies to the clinic. *Nature Reviews Drug Discovery*, *14*(10), 681–692. https://doi.org/10.1038/nrd4738

Kingsley, LawrenceA., Kaslow, R., Rinaldo, CharlesR., Detre, K., Odaka, N., Vanraden, M., Detels, R., Polk, B. F., Chmiel, J., Kelsey, SherylF., Ostrow, D., & Visscher, B. (1987). Risk factors for seroconversion to human immunodeficiency virus among male homosexuals. *The Lancet*, *329*(8529), 345–349. https://doi.org/10.1016/S0140-6736(87)91725-9

Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., & Steitz, T. A. (1992). Crystal Structure at 3.5 Å Resolution of HIV-1 Reverse Transcriptase Complexed with an Inhibitor. *Science*, *256*(5065), 1783–1790. https://doi.org/10.1126/science.1377403

König, R., Zhou, Y., Elleder, D., Diamond, T. L., Bonamy, G. M. C., Irelan, J. T., Chiang, C., Tu, B. P., De Jesus, P. D., Lilley, C. E., Seidel, S., Opaluch, A. M., Caldwell, J. S., Weitzman, M. D., Kuhen, K. L., Bandyopadhyay, S., Ideker, T., Orth, A. P., Miraglia, L. J., … Chanda, S. K. (2008). Global Analysis of Host-Pathogen Interactions that Regulate Early-Stage HIV-1 Replication. *Cell*, *135*(1), 49–60. https://doi.org/10.1016/j.cell.2008.07.032

Koppensteiner, H., Brack-Werner, R., & Schindler, M. (2012). Macrophages and their relevance in Human Immunodeficiency Virus Type I infection. *Retrovirology*, *9*(1), 82. https://doi.org/10.1186/1742-4690-9-82

Kulkarni, S., Lied, A., Kulkarni, V., Rucevic, M., Martin, M. P., Walker-Sperling, V., Anderson, S. K., Ewy, R., Singh, S., Nguyen, H., McLaren, P. J., Viard, M., Naranbhai, V., Zou, C., Lin, Z., Gatanaga, H., Oka, S., Takiguchi, M., Thio, C. L., … Carrington, M. (2019). CCR5AS lncRNA

variation differentially regulates CCR5, influencing HIV disease outcome. *Nature Immunology*, *20*(7), 824–834. https://doi.org/10.1038/s41590-019-0406-1

Laguette, N., Sobhian, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Ségéral, E., Yatim, A., Emiliani, S., Schwartz, O., & Benkirane, M. (2011). SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*, *474*(7353), 654–657. https://doi.org/10.1038/nature10117

Lama, J., Mangasarian, A., & Trono, D. (1999). Cell-surface expression of CD4 reduces HIV-1 infectivity by blocking Env incorporation in a Nef- and Vpu-inhibitable manner. *Current Biology : CB*, *9*(12), 622–631. https://doi.org/10.1016/s0960-9822(99)80284-x

Lane, J., McLaren, P. J., Dorrell, L., Shianna, K. V., Stemke, A., Pelak, K., Moore, S., Oldenburg, J., Alvarez-Roman, M. T., Angelillo-Scherrer, A., Boehlen, F., Bolton-Maggs, P. H. B., Brand, B., Brown, D., Chiang, E., Cid-Haro, A. R., Clotet, B., Collins, P., Colombo, S., … Fellay, J. (2013). A genome-wide association study of resistance to HIV infection in highly exposed uninfected individuals with hemophilia A. *Human Molecular Genetics*, *22*(9), 1903–1910. https://doi.org/10.1093/hmg/ddt033

Lepelley, A., Louis, S., Sourisseau, M., Law, H. K. W., Pothlichet, J., Schilte, C., Chaperot, L., Plumas, J., Randall, R. E., Si-Tahar, M., Mammano, F., Albert, M. L., & Schwartz, O. (2011). Innate Sensing of HIV-Infected Cells. *PLoS Pathogens*, *7*(2), e1001284. https://doi.org/10.1371/journal.ppat.1001284

Li, Q., Duan, L., Estes, J. D., Ma, Z.-M., Rourke, T., Wang, Y., Reilly, C., Carlis, J., Miller, C. J., & Haase, A. T. (2005). Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells. *Nature*, *434*(7037), 1148–1152. https://doi.org/10.1038/nature03513

Li, Q., Skinner, P. J., Ha, S.-J., Duan, L., Mattila, T. L., Hage, A., White, C., Barber, D. L., O'Mara, L., Southern, P. J., Reilly, C. S., Carlis, J. V., Miller, C. J., Ahmed, R., & Haase, A. T. (2009). Visualizing Antigen-Specific and Infected Cells in Situ Predicts Outcomes in Early Viral Infection. *Science*, *323*(5922), 1726–1729. https://doi.org/10.1126/science.1168676

Li, Y., Kar, A. K., & Sodroski, J. (2009). Target Cell Type-Dependent Modulation of Human Immunodeficiency Virus Type 1 Capsid Disassembly by Cyclophilin A. *Journal of Virology*, *83*(21), 10951–10962. https://doi.org/10.1128/JVI.00682-09

Liao, J. K., Seto, M., & Noma, K. (2007). Rho kinase (ROCK) inhibitors. *Journal of Cardiovascular Pharmacology*, *50*(1), 17–24. https://doi.org/10.1097/FJC.0b013e318070d1bd

Lim, E. S., Fregoso, O. I., McCoy, C. O., Matsen, F. A., Malik, H. S., & Emerman, M. (2012). The Ability of Primate Lentiviruses to Degrade the Monocyte Restriction Factor SAMHD1 Preceded the Birth of the Viral Accessory Protein Vpx. *Cell Host & Microbe*, *11*(2), 194–204. https://doi.org/10.1016/j.chom.2012.01.004

Lindwasser, O. W., Chaudhuri, R., & Bonifacino, J. S. (2007). Mechanisms of CD4 downregulation by the Nef and Vpu proteins of primate immunodeficiency viruses. *Current Molecular Medicine*, *7*(2), 171–184. https://doi.org/10.2174/156652407780059177

Lister, R., Pelizzola, M., Kida, Y. S., Hawkins, R. D., Nery, J. R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., Downes, M., Yu, R., Stewart, R., Ren, B., Thomson, J. A., Evans, R. M., & Ecker, J. R. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*, *471*(7336), 68–73. https://doi.org/10.1038/nature09798

Little, S. J., McLean, A. R., Spina, C. A., Richman, D. D., & Havlir, D. V. (1999). Viral Dynamics of Acute HIV-1 Infection. *The Journal of Experimental Medicine*, *190*(6), 841–850. https://doi.org/10.1084/jem.190.6.841

Liu, C., Perilla, J. R., Ning, J., Lu, M., Hou, G., Ramalho, R., Himes, B. A., Zhao, G., Bedwell, G. J., Byeon, I.-J., Ahn, J., Gronenborn, A. M., Prevelige, P. E., Rousso, I., Aiken, C., Polenova, T., Schulten, K., & Zhang, P. (2016). Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nature Communications*, *7*(1), 10714. https://doi.org/10.1038/ncomms10714

Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., MacDonald, M. E., Stuhlmann, H., Koup, R. A., & Landau, N. R. (1996). Homozygous Defect in HIV-1 Coreceptor Accounts for Resistance of Some Multiply-Exposed Individuals to HIV-1 Infection. *Cell*, *86*(3), 367–377. https://doi.org/10.1016/S0092-8674(00)80110-5

Llano, M., Saenz, D. T., Meehan, A., Wongthida, P., Peretz, M., Walker, W. H., Teo, W., & Poeschla, E. M. (2006). An Essential Role for LEDGF/p75 in HIV Integration. *Science*, *314*(5798), 461–464. https://doi.org/10.1126/science.1132319

Lo Sardo, V., Ferguson, W., Erikson, G. A., Topol, E. J., Baldwin, K. K., & Torkamani, A. (2017). Influence of donor age on induced pluripotent stem cells. *Nature Biotechnology*, *35*(1), 69–74. https://doi.org/10.1038/nbt.3749

Loh, Y.-H., Agarwal, S., Park, I.-H., Urbach, A., Huo, H., Heffner, G. C., Kim, K., Miller, J. D., Ng, K., & Daley, G. Q. (2009). Generation of induced pluripotent stem cells from human blood. *Blood*, *113*(22), 5476–5479. https://doi.org/10.1182/blood-2009-02-204800

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, J., Pan, Q., Rong, L., Liu, S.-L., & Liang, C. (2011). The IFITM Proteins Inhibit HIV-1 Infection. *Journal of Virology*, *85*(5), 2126–2137. https://doi.org/10.1128/JVI.01531-10

Luo, Y., Kanai, M., Choi, W., Li, X., Sakaue, S., Yamamoto, K., Ogawa, K., Gutierrez-Arcelus, M., Gregersen, P. K., Stuart, P. E., Elder, J. T., Forer, L., Schönherr, S., Fuchsberger, C., Smith, A. V., Fellay, J., Carrington, M., Haas, D. W., Guo, X., … Raychaudhuri, S. (2021). A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nature Genetics*, *53*(10), 1504–1516. https://doi.org/10.1038/s41588-021-00935-7

Lyles, R. H., Muñoz, A., Yamashita, T. E., Bazmi, H., Detels, R., Rinaldo, C. R., Margolick, J. B., Phair, J. P., & Mellors, J. W. (2000). Natural History of Human Immunodeficiency Virus Type 1 Viremia after Seroconversion and Proximal to AIDS in a Large Cohort of Homosexual Men. *The Journal of Infectious Diseases*, *181*(3), 872–880. https://doi.org/10.1086/315339

Ma, H., Morey, R., O'Neil, R. C., He, Y., Daughtry, B., Schultz, M. D., Hariharan, M., Nery, J. R., Castanon, R., Sabatini, K., Thiagarajan, R. D., Tachibana, M., Kang, E., Tippner-Hedges, R., Ahmed, R., Gutierrez, N. M., Van Dyken, C., Polat, A., Sugawara, A., … Mitalipov, S. (2014). Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature*, *511*(7508), 177–183. https://doi.org/10.1038/nature13551

Mackelprang, R. D., Bamshad, M. J., Chong, J. X., Hou, X., Buckingham, K. J., Shively, K., deBruyn, G., Mugo, N. R., Mullins, J. I., McElrath, M. J., Baeten, J. M., Celum, C., Emond, M. J., & Lingappa, J. R. (2017). Whole genome sequencing of extreme phenotypes identifies variants in CD101 and UBE2V1 associated with increased risk of sexually

acquired HIV-1. *PLOS Pathogens*, *13*(11), e1006703. https://doi.org/10.1371/journal.ppat.1006703

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z., & Engelborghs, Y. (2003). LEDGF/p75 Is Essential for Nuclear and Chromosomal Targeting of HIV-1 Integrase in Human Cells. *Journal of Biological Chemistry*, *278*(35), 33528–33539. https://doi.org/10.1074/jbc.M303594200

Mali, P., Chou, B.-K., Yen, J., Ye, Z., Zou, J., Dowey, S., Brodsky, R. A., Ohm, J. E., Yu, W., Baylin, S. B., Yusa, K., Bradley, A., Meyers, D. J., Mukherjee, C., Cole, P. A., & Cheng, L. (2010). Butyrate Greatly Enhances Derivation of Human Induced Pluripotent Stem Cells by Promoting Epigenetic Remodeling and the Expression of Pluripotency-Associated Genes. *Stem Cells*, *28*(4), 713–720. https://doi.org/10.1002/stem.402

Malim, M. H. (2009). APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1517), 675–687. https://doi.org/10.1098/rstb.2008.0185

Malim, M. H., Hauber, J., Le, S.-Y., Maizel, J. V., & Cullen, B. R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature*, *338*(6212), 254–257. https://doi.org/10.1038/338254a0

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., & Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*, *424*(6944), 99–103. https://doi.org/10.1038/nature01709

Marchetto, M. C. N., Yeo, G. W., Kainohana, O., Marsala, M., Gage, F. H., & Muotri, A. R. (2009). Transcriptional Signature and Memory Retention of Human-Induced Pluripotent Stem Cells. *PLoS ONE*, *4*(9), e7076. https://doi.org/10.1371/journal.pone.0007076

Marshall, H. M., Ronen, K., Berry, C., Llano, M., Sutherland, H., Saenz, D., Bickmore, W., Poeschla, E., & Bushman, F. D. (2007). Role of PSIP1/LEDGF/p75 in Lentiviral Infectivity and Integration Targeting. *PLoS ONE*, *2*(12), e1340. https://doi.org/10.1371/journal.pone.0001340

Martin, M. P., Gao, X., Lee, J.-H., Nelson, G. W., Detels, R., Goedert, J. J., Buchbinder, S., Hoots, K., Vlahov, D., Trowsdale, J., Wilson, M., O'Brien, S. J., & Carrington, M. (2002). Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nature Genetics*, *31*(4), 429–434. https://doi.org/10.1038/ng934

Martinez-Picado, J., Prado, J. G., Fry, E. E., Pfafferott, K., Leslie, A., Chetty, S., Thobakgale, C., Honeyborne, I., Crawford, H., Matthews, P., Pillay, T., Rousseau, C., Mullins, J. I., Brander, C., Walker, B. D., Stuart, D. I., Kiepiela, P., & Goulder, P. (2006). Fitness Cost of Escape Mutations in p24 Gag in Association with Control of Human Immunodeficiency Virus Type 1. *Journal of Virology*, *80*(7), 3617–3623. https://doi.org/10.1128/JVI.80.7.3617-3623.2006

Martin-Serrano, J., Zang, T., & Bieniasz, P. D. (2001). HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nature Medicine*, *7*(12), 1313–1319. https://doi.org/10.1038/nm1201-1313

Martinson, J. J., Chapman, N. H., Rees, D. C., Liu, Y.-T., & Clegg, J. B. (1997). Global distribution of the CCR5 gene 32-basepair deletion. *Nature Genetics*, *16*(1), 100–103. https://doi.org/10.1038/ng0597-100

Masuyama, N., Kuronita, T., Tanaka, R., Muto, T., Hirota, Y., Takigawa, A., Fujita, H., Aso, Y., Amano, J., & Tanaka, Y. (2009). HM1.24 Is Internalized from Lipid Rafts by Clathrin-mediated Endocytosis through Interaction with α-Adaptin. *Journal of Biological Chemistry*, *284*(23), 15927–15941. https://doi.org/10.1074/jbc.M109.005124

Matreyek, K. A., & Engelman, A. (2011). The Requirement for Nucleoporin NUP153 during Human Immunodeficiency Virus Type 1 Infection Is Determined by the Viral Capsid. *Journal of Virology*, *85*(15), 7818–7827. https://doi.org/10.1128/JVI.00325-11

Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M. C., & Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, *464*(7290), 927–931. https://doi.org/10.1038/nature08858

Mattapallil, J. J., Douek, D. C., Hill, B., Nishimura, Y., Martin, M., & Roederer, M. (2005). Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. *Nature*, *434*(7037), 1093–1097. https://doi.org/10.1038/nature03501

McCune, J. M. (2001). The dynamics of CD4+ T-cell depletion in HIV disease. *Nature*, *410*(6831), 974–979. https://doi.org/10.1038/35073648

McLaren, P. J., Coulonges, C., Bartha, I., Lenz, T. L., Deutsch, A. J., Bashirova, A., Buchbinder, S., Carrington, M. N., Cossarizza, A., Dalmau, J., De Luca, A., Goedert, J. J., Gurdasani, D., Haas, D. W., Herbeck, J. T., Johnson, E. O., Kirk, G. D., Lambotte, O., Luo, M., … Fellay, J. (2015). Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. *Proceedings of the National Academy of Sciences*, *112*(47), 14658–14663. https://doi.org/10.1073/pnas.1514867112

McLaren, P. J., Coulonges, C., Ripke, S., van den Berg, L., Buchbinder, S., Carrington, M., Cossarizza, A., Dalmau, J., Deeks, S. G., Delaneau, O., De Luca, A., Goedert, J. J., Haas, D., Herbeck, J. T., Kathiresan, S., Kirk, G. D., Lambotte, O., Luo, M., Mallal, S., … Fellay, J. (2013). Association Study of Common Genetic Variants and HIV-1 Acquisition in 6,300 Infected Cases and 7,200 Controls. *PLoS Pathogens*, *9*(7), e1003515. https://doi.org/10.1371/journal.ppat.1003515

McLaren, P. J., Porreca, I., Iaconis, G., Mok, H. P., Mukhopadhyay, S., Karakoc, E., Cristinelli, S., Pomilla, C., Bartha, I., Thorball, C. W., Tough, R. H., Angelino, P., Kiar, C. S., Carstensen, T., Fatumo, S., Porter, T., Jarvis, I., Skarnes, W. C., Bassett, A., … Fellay, J. (2023). Africa-specific human genetic variation near CHD1L associates with HIV-1 load. *Nature*, *620*(7976), 1025–1030. https://doi.org/10.1038/s41586-023-06370-4

McLaren, P. J., Pulit, S. L., Gurdasani, D., Bartha, I., Shea, P. R., Pomilla, C., Gupta, N., Gkrania-Klotsas, E., Young, E. H., Bannert, N., Del Amo, J., Gill, M. J., Gilmour, J., Kellam, P., Kelleher, A. D., Sönnerborg, A., Wolinsky, S. M., Zangerle, R., Post, F. A., … Fellay, J. (2017). Evaluating the Impact of Functional Genetic Variation on HIV-1 Control. *The Journal of Infectious Diseases*, *216*(9), 1063–1069. https://doi.org/10.1093/infdis/jix470

McLaren, P. J., Ripke, S., Pelak, K., Weintrob, A. C., Patsopoulos, N. A., Jia, X., Erlich, R. L., Lennon, N. J., Kadie, C. M., Heckerman, D., Gupta, N., Haas, D. W., Deeks, S. G., Pereyra, F., Walker, B. D., & de Bakker, P. I. W. (2012). Fine-mapping classical HLA variation associated with durable host control of HIV-1 infection in African Americans. *Human Molecular Genetics*, *21*(19), 4334–4347. https://doi.org/10.1093/hmg/dds226

Mehandru, S., Poles, M. A., Tenner-Racz, K., Horowitz, A., Hurley, A., Hogan, C., Boden, D., Racz, P., & Markowitz, M. (2004). Primary HIV-1 Infection Is Associated with Preferential Depletion of CD4+ T Lymphocytes from Effector Sites in the Gastrointestinal Tract. *The Journal of Experimental Medicine*, *200*(6), 761–770. https://doi.org/10.1084/jem.20041196

Mehle, A., Strack, B., Ancuta, P., Zhang, C., McPike, M., & Gabuzda, D. (2004). Vif Overcomes the Innate Antiviral Activity of APOBEC3G by Promoting Its Degradation in the

Ubiquitin-Proteasome Pathway. *Journal of Biological Chemistry*, *279*(9), 7792–7798. https://doi.org/10.1074/jbc.M313093200

Melikyan, G. B. (2008). Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm. *Retrovirology*, *5*(1), 111. https://doi.org/10.1186/1742-4690-5-111

Mellors, J. W., Rinaldo, C. R., Gupta, P., White, R. M., Todd, J. A., & Kingsley, L. A. (1996). Prognosis in HIV-1 Infection Predicted by the Quantity of Virus in Plasma. *Science*, *272*(5265), 1167–1170. https://doi.org/10.1126/science.272.5265.1167

Merigan, T., Amato, D., Balsley, J., Power, M., Price, W., Benoit, S., Perez-Michael, A., Brownstein, A., Kramer, A., & Brettler, D. (1991). Placebo-controlled trial to evaluate zidovudine in treatment of human immunodeficiency virus infection in asymptomatic patients with hemophilia. NHF-ACTG 036 Study Group. *Blood*, *78*(4), 900–906. https://doi.org/10.1182/blood.V78.4.900.900

Merkle, F. T., Ghosh, S., Kamitaki, N., Mitchell, J., Avior, Y., Mello, C., Kashin, S., Mekhoubad, S., Ilic, D., Charlton, M., Saphier, G., Handsaker, R. E., Genovese, G., Bar, S., Benvenisty, N., McCarroll, S. A., & Eggan, K. (2017). Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature*, *545*(7653), 229–233. https://doi.org/10.1038/nature22312

Michael, N. L., Chang, G., Loum, L. G., Mascola, J. R., Dondero, D., Birx, D. L., & Sheppard, H. W. (1997). The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression. *Nature Medicine*, *3*(3), 338–340. https://doi.org/10.1038/nm0397-338

Migueles, S. A., Laborico, A. C., Shupert, W. L., Sabbaghian, M. S., Rabin, R., Hallahan, C. W., Baarle, D. Van, Kostense, S., Miedema, F., McLaughlin, M., Ehler, L., Metcalf, J., Liu, S., & Connors, M. (2002). HIV-specific CD8+ T cell proliferation is coupled to perforin expression and is maintained in nonprogressors. *Nature Immunology*, *3*(11), 1061–1068. https://doi.org/10.1038/ni845

Migueles, S. A., Osborne, C. M., Royce, C., Compton, A. A., Joshi, R. P., Weeks, K. A., Rood, J. E., Berkley, A. M., Sacha, J. B., Cogliano-Shutta, N. A., Lloyd, M., Roby, G., Kwan, R., McLaughlin, M., Stallings, S., Rehm, C., O'Shea, M. A., Mican, J., Packard, B. Z., … Connors, M. (2008). Lytic Granule Loading of CD8+ T Cells Is Required for HIV-Infected Cell Elimination Associated with Immune Control. *Immunity*, *29*(6), 1009–1021. https://doi.org/10.1016/j.immuni.2008.10.010

Min, A. K., Javidfar, B., Missall, R., Doanman, D., Durens, M., Vil, S. S., Masih, Z., Graziani, M., Mordelt, A., Marro, S., de Witte, L., Chen, B. K., Swartz, T. H., & Akbarian, S. (2023). HIV-1 infection of genetically engineered iPSC-derived central nervous system-engrafted microglia in a humanized mouse model. *BioRxiv : The Preprint Server for Biology*. https://doi.org/10.1101/2023.04.26.538461

Mitchell, R. S., Katsura, C., Skasko, M. A., Fitzpatrick, K., Lau, D., Ruiz, A., Stephens, E. B., Margottin-Goguet, F., Benarous, R., & Guatelli, J. C. (2009). Vpu Antagonizes BST-2–Mediated Restriction of HIV-1 Release via β-TrCP and Endo-Lysosomal Trafficking. *PLoS Pathogens*, *5*(5), e1000450. https://doi.org/10.1371/journal.ppat.1000450

Miyoshi, N., Ishii, H., Nagano, H., Haraguchi, N., Dewi, D. L., Kano, Y., Nishikawa, S., Tanemura, M., Mimori, K., Tanaka, F., Saito, T., Nishimura, J., Takemasa, I., Mizushima, T., Ikeda, M., Yamamoto, H., Sekimoto, M., Doki, Y., & Mori, M. (2011). Reprogramming of Mouse and Human Cells to Pluripotency Using Mature MicroRNAs. *Cell Stem Cell*, *8*(6), 633–638. https://doi.org/10.1016/j.stem.2011.05.001

Müller, F.-J., Schuldt, B. M., Williams, R., Mason, D., Altun, G., Papapetrou, E. P., Danner, S., Goldmann, J. E., Herbst, A., Schmidt, N. O., Aldenhoff, J. B., Laurent, L. C., & Loring, J. F. (2011). A bioinformatic assay for pluripotency in human cells. *Nature Methods*, *8*(4), 315–317. https://doi.org/10.1038/nmeth.1580

Narsinh, K. H., Jia, F., Robbins, R. C., Kay, M. A., Longaker, M. T., & Wu, J. C. (2011). Generation of adult human induced pluripotent stem cells using nonviral minicircle DNA vectors. *Nature Protocols*, *6*(1), 78–88. https://doi.org/10.1038/nprot.2010.173

Nehme, R., & Barrett, L. E. (2020). Using human pluripotent stem cell models to study autism in the era of big data. *Molecular Autism*, *11*(1), 21. https://doi.org/10.1186/s13229-020-00322-9

Neil, S. J. D., Zang, T., & Bieniasz, P. D. (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature*, *451*(7177), 425–430. https://doi.org/10.1038/nature06553

Ni, Z., Knorr, D. A., Bendzick, L., Allred, J., & Kaufman, D. S. (2014). Expression of Chimeric Receptor CD4 ζ by Natural Killer Cells Derived from Human Pluripotent Stem Cells Improves In Vitro Activity but Does Not Enhance Suppression of HIV Infection In Vivo. *Stem Cells*, *32*(4), 1021–1031. https://doi.org/10.1002/stem.1611

Ni, Z., Knorr, D. A., Clouser, C. L., Hexum, M. K., Southern, P., Mansky, L. M., Park, I.-H., & Kaufman, D. S. (2011). Human Pluripotent Stem Cells Produce Natural Killer Cells That Mediate Anti-HIV-1 Activity by Utilizing Diverse Cellular Mechanisms. *Journal of Virology*, *85*(1), 43–50. https://doi.org/10.1128/JVI.01774-10

Nishimura, T., Kaneko, S., Kawana-Tachikawa, A., Tajima, Y., Goto, H., Zhu, D., Nakayama-Hosoya, K., Iriguchi, S., Uemura, Y., Shimizu, T., Takayama, N., Yamada, D., Nishimura, K., Ohtaka, M., Watanabe, N., Takahashi, S., Iwamoto, A., Koseki, H., Nakanishi, M., … Nakauchi, H. (2013). Generation of Rejuvenated Antigen-Specific T Cells by Reprogramming to Pluripotency and Redifferentiation. *Cell Stem Cell*, *12*(1), 114–126. https://doi.org/10.1016/j.stem.2012.11.002

Nishino, K., Toyoda, M., Yamazaki-Inoue, M., Fukawatase, Y., Chikazawa, E., Sakaguchi, H., Akutsu, H., & Umezawa, A. (2011). DNA Methylation Dynamics in Human Induced Pluripotent Stem Cells over Time. *PLoS Genetics*, *7*(5), e1002085. https://doi.org/10.1371/journal.pgen.1002085

Novembre, J., Galvani, A. P., & Slatkin, M. (2005). The Geographic Spread of the CCR5 Δ32 HIV-Resistance Allele. *PLoS Biology*, *3*(11), e339. https://doi.org/10.1371/journal.pbio.0030339

Obimbo, E. M., Wamalwa, D., Richardson, B., Mbori-Ngacha, D., Overbaugh, J., Emery, S., Otieno, P., Farquhar, C., Bosire, R., Payne, B. L., & John-Stewart, G. (2009). Pediatric HIV-1 in Kenya: Pattern and Correlates of Viral Load and Association With Mortality. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *51*(2), 209–215. https://doi.org/10.1097/QAI.0b013e31819c16d8

Ocwieja, K. E., Sherrill-Mix, S., Mukherjee, R., Custers-Allen, R., David, P., Brown, M., Wang, S., Link, D. R., Olson, J., Travers, K., Schadt, E., & Bushman, F. D. (2012). Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Research*, *40*(20), 10345–10355. https://doi.org/10.1093/nar/gks753

OhAinle, M., Helms, L., Vermeire, J., Roesch, F., Humes, D., Basom, R., Delrow, J. J., Overbaugh, J., & Emerman, M. (2018). A virus-packageable CRISPR screen identifies

host factors mediating interferon inhibition of HIV. *ELife*, *7*. https://doi.org/10.7554/eLife.39823

Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., Kikuchi, H., Yoshida, H., Tanaka, A., Suyama, M., & Arima, T. (2014). Genome-Wide Analysis of DNA Methylation Dynamics during Early Human Development. *PLoS Genetics*, *10*(12), e1004868. https://doi.org/10.1371/journal.pgen.1004868

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K., Shibata, T., Kunisada, T., Takahashi, M., Takahashi, J., Saji, H., & Yamanaka, S. (2011). A more efficient method to generate integration-free human iPS cells. *Nature Methods*, *8*(5), 409–412. https://doi.org/10.1038/nmeth.1591

Ono, A., Ablan, S. D., Lockett, S. J., Nagashima, K., & Freed, E. O. (2004). Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proceedings of the National Academy of Sciences*, *101*(41), 14889–14894. https://doi.org/10.1073/pnas.0405596101

Ormel, P. R., Vieira de Sá, R., van Bodegraven, E. J., Karst, H., Harschnitz, O., Sneeboer, M. A. M., Johansen, L. E., van Dijk, R. E., Scheefhals, N., Berdenis van Berlekom, A., Ribes Martínez, E., Kling, S., MacGillavry, H. D., van den Berg, L. H., Kahn, R. S., Hol, E. M., de Witte, L. D., & Pasterkamp, R. J. (2018). Microglia innately develop within cerebral organoids. *Nature Communications*, *9*(1), 4167. https://doi.org/10.1038/s41467-018-06684-2

Pache, L., König, R., & Chanda, S. K. (2011). Identifying HIV-1 host cell factors by genome-scale RNAi screening. *Methods*, *53*(1), 3–12. https://doi.org/10.1016/j.ymeth.2010.07.009

Panganiban, A. T., & Fiore, D. (1988). Ordered Interstrand and Intrastrand DNA Transfer During Reverse Transcription. *Science*, *241*(4869), 1064–1069. https://doi.org/10.1126/science.2457948

Papapetrou, E. P., & Sadelain, M. (2011). Generation of transgene-free human induced pluripotent stem cells with an excisable single polycistronic vector. *Nature Protocols*, *6*(9), 1251–1273. https://doi.org/10.1038/nprot.2011.374

Park, R. J., Wang, T., Koundakjian, D., Hultquist, J. F., Lamothe-Molina, P., Monel, B., Schumann, K., Yu, H., Krupzcak, K. M., Garcia-Beltran, W., Piechocka-Trocha, A., Krogan, N. J., Marson, A., Sabatini, D. M., Lander, E. S., Hacohen, N., & Walker, B. D. (2017). A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nature Genetics*, *49*(2), 193–203. https://doi.org/10.1038/ng.3741

Pelak, K., Goldstein, D. B., Walley, N. M., Fellay, J., Ge, D., Shianna, K. V., Gumbs, C., Gao, X., Maia, J. M., Cronin, K. D., Hussain, S. K., Carrington, M., Michael, N. L., & Weintrob, A. C. (2010). Host Determinants of HIV-1 Control in African Americans. *The Journal of Infectious Diseases*, *201*(8), 1141–1149. https://doi.org/10.1086/651382

Pereira, L. A., Bentley, K., Peeters, A., Churchill, M. J., & Deacon, N. J. (2000). A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Research*, *28*(3), 663–668. https://doi.org/10.1093/nar/28.3.663

Pertel, T., Hausmann, S., Morger, D., Züger, S., Guerra, J., Lascano, J., Reinhard, C., Santoni, F. A., Uchil, P. D., Chatel, L., Bisiaux, A., Albert, M. L., Strambio-De-Castillia, C., Mothes, W., Pizzato, M., Grütter, M. G., & Luban, J. (2011). TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature*, *472*(7343), 361–365. https://doi.org/10.1038/nature09976

Pettit, S. C., Moody, M. D., Wehbie, R. S., Kaplan, A. H., Nantermet, P. V, Klein, C. A., & Swanstrom, R. (1994). The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *Journal of Virology*, *68*(12), 8017–8027. https://doi.org/10.1128/jvi.68.12.8017-8027.1994

Pezeshkian, N., Groves, N. S., & van Engelenburg, S. B. (2019). Single-molecule imaging of HIV-1 envelope glycoprotein dynamics and Gag lattice association exposes determinants responsible for virus incorporation. *Proceedings of the National Academy of Sciences*, *116*(50), 25269–25277. https://doi.org/10.1073/pnas.1910008116

Piatak, M., Saag, M. S., Yang, L. C., Clark, S. J., Kappes, J. C., Luk, K.-C., Hahn, B. H., Shaw, G. M., & Lifson, J. D. (1993). High Levels of HIV-1 in Plasma During All Stages of Infection Determined By Competitive PCR. *Science*, *259*(5102), 1749–1754. https://doi.org/10.1126/science.8096089

Pick, M., Stelzer, Y., Bar-Nur, O., Mayshar, Y., Eden, A., & Benvenisty, N. (2009). Clone- and Gene-Specific Aberrations of Parental Imprinting in Human Induced Pluripotent Stem Cells. *Stem Cells*, *27*(11), 2686–2690. https://doi.org/10.1002/stem.205

Pickering, S., Betancor, G., Galão, R. P., Merrick, B., Signell, A. W., Wilson, H. D., Kia Ik, M. T., Seow, J., Graham, C., Acors, S., Kouphou, N., Steel, K. J. A., Hemmings, O., Patel, A., Nebbia, G., Douthwaite, S., O'Connell, L., Luptak, J., McCoy, L. E., … Edgeworth, J. D. (2020). Comparative assessment of multiple COVID-19 serological technologies supports continued evaluation of point-of-care lateral flow assays in hospital and community healthcare settings. *PLOS Pathogens*, *16*(9), e1008817. https://doi.org/10.1371/journal.ppat.1008817

Plantier, J.-C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D. L., & Simon, F. (2009). A new human immunodeficiency virus derived from gorillas. *Nature Medicine*, *15*(8), 871–872. https://doi.org/10.1038/nm.2016

Pollpeter, D., Parsons, M., Sobala, A. E., Coxhead, S., Lang, R. D., Bruns, A. M., Papaioannou, S., McDonnell, J. M., Apolonia, L., Chowdhury, J. A., Horvath, C. M., & Malim, M. H. (2017). Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G. *Nature Microbiology*, *3*(2), 220–233. https://doi.org/10.1038/s41564-017-0063-9

Purcell, D. F., & Martin, M. A. (1993). Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *Journal of Virology*, *67*(11), 6365–6378. https://doi.org/10.1128/jvi.67.11.6365-6378.1993

Purohit, V., Roques, B. P., Kim, B., & Bambara, R. A. (2007). Mechanisms That Prevent Template Inactivation by HIV-1 Reverse Transcriptase RNase H Cleavages. *Journal of Biological Chemistry*, *282*(17), 12598–12609. https://doi.org/10.1074/jbc.M700043200

Quinn, T. C., Wawer, M. J., Sewankambo, N., Serwadda, D., Li, C., Wabwire-Mangen, F., Meehan, M. O., Lutalo, T., & Gray, R. H. (2000). Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1. *New England Journal of Medicine*, *342*(13), 921–929. https://doi.org/10.1056/NEJM200003303421303

Raha, T., Cheng, S. W. G., & Green, M. R. (2005). HIV-1 Tat Stimulates Transcription Complex Assembly through Recruitment of TBP in the Absence of TAFs. *PLoS Biology*, *3*(2), e44. https://doi.org/10.1371/journal.pbio.0030044

Ramsuran, V., Naranbhai, V., Horowitz, A., Qi, Y., Martin, M. P., Yuki, Y., Gao, X., Walker-Sperling, V., Del Prete, G. Q., Schneider, D. K., Lifson, J. D., Fellay, J., Deeks, S. G., Martin, J. N., Goedert, J. J., Wolinsky, S. M., Michael, N. L., Kirk, G. D., Buchbinder, S., …

Carrington, M. (2018). Elevated *HLA-A* expression impairs HIV control through inhibition of NKG2A-expressing cells. *Science*, *359*(6371), 86–90. https://doi.org/10.1126/science.aam8825

Rankovic, S., Varadarajan, J., Ramalho, R., Aiken, C., & Rousso, I. (2017). Reverse Transcription Mechanically Initiates HIV-1 Capsid Disassembly. *Journal of Virology*, *91*(12). https://doi.org/10.1128/JVI.00289-17

Rey, D. (2011). Post-exposure prophylaxis for HIV infection. *Expert Review of Anti-Infective Therapy*, *9*(4), 431–442. https://doi.org/10.1586/eri.11.20

Richardson, B. A., Mbori-Ngacha, D., Lavreys, L., John-Stewart, G. C., Nduati, R., Panteleeff, D. D., Emery, S., Kreiss, J. K., & Overbaugh, J. (2003). Comparison of Human Immunodeficiency Virus Type 1 Viral Loads in Kenyan Women, Men, and Infants during Primary and Early Infection. *Journal of Virology*, *77*(12), 7120–7123. https://doi.org/10.1128/JVI.77.12.7120-7123.2003

Richman, D. D., Wrin, T., Little, S. J., & Petropoulos, C. J. (2003). Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proceedings of the National Academy of Sciences*, *100*(7), 4144–4149. https://doi.org/10.1073/pnas.0630530100

Rivera, T., Zhao, Y., Ni, Y., & Wang, J. (2020). Human-Induced Pluripotent Stem Cell Culture Methods Under cGMP Conditions. *Current Protocols in Stem Cell Biology*, *54*(1), e117. https://doi.org/10.1002/cpsc.117

Rosa, A., Chande, A., Ziglio, S., De Sanctis, V., Bertorelli, R., Goh, S. L., McCauley, S. M., Nowosielska, A., Antonarakis, S. E., Luban, J., Santoni, F. A., & Pizzato, M. (2015). HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature*, *526*(7572), 212–217. https://doi.org/10.1038/nature15399

Rouhani, F. J., Zou, X., Danecek, P., Badja, C., Amarante, T. D., Koh, G., Wu, Q., Memari, Y., Durbin, R., Martincorena, I., Bassett, A. R., Gaffney, D., & Nik-Zainal, S. (2022). Substantial somatic genomic variation and selection for BCOR mutations in human induced pluripotent stem cells. *Nature Genetics*, *54*(9), 1406–1416. https://doi.org/10.1038/s41588-022-01147-3

Rugg-Gunn, P. J., Ferguson-Smith, A. C., & Pedersen, R. A. (2007). Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. *Human Molecular Genetics*, *16*(R2), R243–R251. https://doi.org/10.1093/hmg/ddm245

Ryoo, J., Choi, J., Oh, C., Kim, S., Seo, M., Kim, S.-Y., Seo, D., Kim, J., White, T. E., Brandariz-Nuñez, A., Diaz-Griffero, F., Yun, C.-H., Hollenbaugh, J. A., Kim, B., Baek, D., & Ahn, K. (2014). The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. *Nature Medicine*, *20*(8), 936–941. https://doi.org/10.1038/nm.3626

Salazar-Gonzalez, J. F., Salazar, M. G., Keele, B. F., Learn, G. H., Giorgi, E. E., Li, H., Decker, J. M., Wang, S., Baalwa, J., Kraus, M. H., Parrish, N. F., Shaw, K. S., Guffey, M. B., Bar, K. J., Davis, K. L., Ochsenbauer-Jambor, C., Kappes, J. C., Saag, M. S., Cohen, M. S., … Shaw, G. M. (2009). Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *Journal of Experimental Medicine*, *206*(6), 1273–1289. https://doi.org/10.1084/jem.20090378

Sawyer, S. L., Wu, L. I., Emerman, M., & Malik, H. S. (2005). Positive selection of primate *TRIM5* α identifies a critical species-specific retroviral restriction domain. *Proceedings of the National Academy of Sciences*, *102*(8), 2832–2837. https://doi.org/10.1073/pnas.0409853102

Schaller, T., Ocwieja, K. E., Rasaiyaah, J., Price, A. J., Brady, T. L., Roth, S. L., Hué, S., Fletcher, A. J., Lee, K., KewalRamani, V. N., Noursadeghi, M., Jenner, R. G., James, L. C., Bushman, F. D., & Towers, G. J. (2011). HIV-1 Capsid-Cyclophilin Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency. *PLoS Pathogens*, *7*(12), e1002439. https://doi.org/10.1371/journal.ppat.1002439

Schlaeger, T. M., Daheron, L., Brickler, T. R., Entwisle, S., Chan, K., Cianci, A., DeVine, A., Ettenger, A., Fitzgerald, K., Godfrey, M., Gupta, D., McPherson, J., Malwadkar, P., Gupta, M., Bell, B., Doi, A., Jung, N., Li, X., Lynes, M. S., … Daley, G. Q. (2015). A comparison of non-integrating reprogramming methods. *Nature Biotechnology*, *33*(1), 58–63. https://doi.org/10.1038/nbt.3070

Schneidewind, A., Brockman, M. A., Yang, R., Adam, R. I., Li, B., Le Gall, S., Rinaldo, C. R., Craggs, S. L., Allgaier, R. L., Power, K. A., Kuntzen, T., Tung, C.-S., LaBute, M. X., Mueller, S. M., Harrer, T., McMichael, A. J., Goulder, P. J. R., Aiken, C., Brander, C., … Allen, T. M. (2007). Escape from the Dominant HLA-B27-Restricted Cytotoxic T-Lymphocyte Response in Gag Is Associated with a Dramatic Reduction in Human Immunodeficiency Virus Type 1 Replication. *Journal of Virology*, *81*(22), 12382–12393. https://doi.org/10.1128/JVI.01543-07

Schröder, A. R. W., Shinn, P., Chen, H., Berry, C., Ecker, J. R., & Bushman, F. (2002). HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell*, *110*(4), 521–529. https://doi.org/10.1016/S0092-8674(02)00864-4

Schulte, B., Selyutina, A., Opp, S., Herschhorn, A., Sodroski, J. G., Pizzato, M., & Diaz-Griffero, F. (2018). Localization to detergent-resistant membranes and HIV-1 core entry inhibition correlate with HIV-1 restriction by SERINC5. *Virology*, *515*, 52–65. https://doi.org/10.1016/j.virol.2017.12.005

Seki, T., Yuasa, S., Oda, M., Egashira, T., Yae, K., Kusumoto, D., Nakata, H., Tohyama, S., Hashimoto, H., Kodaira, M., Okada, Y., Seimiya, H., Fusaki, N., Hasegawa, M., & Fukuda, K. (2010). Generation of Induced Pluripotent Stem Cells from Human Terminally Differentiated Circulating T Cells. *Cell Stem Cell*, *7*(1), 11–14. https://doi.org/10.1016/j.stem.2010.06.003

Selyutina, A., Persaud, M., Simons, L. M., Bulnes-Ramos, A., Buffone, C., Martinez-Lopez, A., Scoca, V., Di Nunzio, F., Hiatt, J., Marson, A., Krogan, N. J., Hultquist, J. F., & Diaz-Griffero, F. (2020). Cyclophilin A Prevents HIV-1 Restriction in Lymphocytes by Blocking Human TRIM5α Binding to the Viral Core. *Cell Reports*, *30*(11), 3766-3777.e6. https://doi.org/10.1016/j.celrep.2020.02.100

Seow, J., Graham, C., Merrick, B., Acors, S., Pickering, S., Steel, K. J. A., Hemmings, O., O'Byrne, A., Kouphou, N., Galao, R. P., Betancor, G., Wilson, H. D., Signell, A. W., Winstone, H., Kerridge, C., Huettner, I., Jimenez-Guardeño, J. M., Lista, M. J., Temperton, N., … Doores, K. J. (2020). Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nature Microbiology*, *5*(12), 1598–1607. https://doi.org/10.1038/s41564-020-00813-8

Shah, S., Alexaki, A., Pirrone, V., Dahiya, S., Nonnemacher, M. R., & Wigdahl, B. (2014). Functional properties of the HIV-1 long terminal repeat containing single-nucleotide polymorphisms in Sp site III and CCAAT/enhancer binding protein site I. *Virology Journal*, *11*, 92. https://doi.org/10.1186/1743-422X-11-92

Sheehy, A. M., Gaddis, N. C., Choi, J. D., & Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, *418*(6898), 646–650. https://doi.org/10.1038/nature00939

Sheehy, A. M., Gaddis, N. C., & Malim, M. H. (2003). The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nature Medicine*, *9*(11), 1404–1407. https://doi.org/10.1038/nm945

Shi, Y., Inoue, H., Wu, J. C., & Yamanaka, S. (2017). Induced pluripotent stem cell technology: a decade of progress. *Nature Reviews. Drug Discovery*, *16*(2), 115–130. https://doi.org/10.1038/nrd.2016.245

Shoeman, R. L., Höner, B., Stoller, T. J., Kesselmeier, C., Miedel, M. C., Traub, P., & Graves, M. C. (1990). Human immunodeficiency virus type 1 protease cleaves the intermediate filament proteins vimentin, desmin, and glial fibrillary acidic protein. *Proceedings of the National Academy of Sciences*, *87*(16), 6336–6340. https://doi.org/10.1073/pnas.87.16.6336

Shoeman, R. L., Hüttermann, C., Hartig, R., & Traub, P. (2001). Amino-terminal Polypeptides of Vimentin Are Responsible for the Changes in Nuclear Architecture Associated with Human Immunodeficiency Virus Type 1 Protease Activity in Tissue Culture Cells. *Molecular Biology of the Cell*, *12*(1), 143–154. https://doi.org/10.1091/mbc.12.1.143

Siliciano, J. D., Kajdas, J., Finzi, D., Quinn, T. C., Chadwick, K., Margolick, J. B., Kovacs, C., Gange, S. J., & Siliciano, R. F. (2003). Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nature Medicine*, *9*(6), 727–728. https://doi.org/10.1038/nm880

Silva, S. S., Rowntree, R. K., Mekhoubad, S., & Lee, J. T. (2008). X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proceedings of the National Academy of Sciences*, *105*(12), 4820–4825. https://doi.org/10.1073/pnas.0712136105

Silvestri, G., Sodora, D. L., Koup, R. A., Paiardini, M., O'Neil, S. P., McClure, H. M., Staprans, S. I., & Feinberg, M. B. (2003). Nonpathogenic SIV infection of sooty mangabeys is characterized by limited bystander immunopathology despite chronic high-level viremia. *Immunity*, *18*(3), 441–452. https://doi.org/10.1016/s1074-7613(03)00060-8

Smith, J. S., & Roth, M. J. (1992). Specificity of human immunodeficiency virus-1 reverse transcriptase-associated ribonuclease H in removal of the minus-strand primer, tRNA(Lys3). *The Journal of Biological Chemistry*, *267*(21), 15071–15079.

Solis, M., Nakhaei, P., Jalalirad, M., Lacoste, J., Douville, R., Arguello, M., Zhao, T., Laughrea, M., Wainberg, M. A., & Hiscott, J. (2011). RIG-I-Mediated Antiviral Signaling Is Inhibited in HIV-1 Infection by a Protease-Mediated Sequestration of RIG-I. *Journal of Virology*, *85*(3), 1224–1236. https://doi.org/10.1128/JVI.01635-10

Sood, C., Marin, M., Chande, A., Pizzato, M., & Melikyan, G. B. (2017). SERINC5 protein inhibits HIV-1 fusion pore formation by promoting functional inactivation of envelope glycoproteins. *Journal of Biological Chemistry*, *292*(14), 6014–6026. https://doi.org/10.1074/jbc.M117.777714

Sowd, G. A., Serrao, E., Wang, H., Wang, W., Fadel, H. J., Poeschla, E. M., & Engelman, A. N. (2016). A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proceedings of the National Academy of Sciences*, *113*(8). https://doi.org/10.1073/pnas.1524213113

Stadtfeld, M., & Hochedlinger, K. (2010). Induced pluripotency: history, mechanisms, and applications. *Genes & Development*, *24*(20), 2239–2263. https://doi.org/10.1101/gad.1963910

Stadtfeld, M., Maherali, N., Breault, D. T., & Hochedlinger, K. (2008). Defining Molecular Cornerstones during Fibroblast to iPS Cell Reprogramming in Mouse. *Cell Stem Cell*, *2*(3), 230–240. https://doi.org/10.1016/j.stem.2008.02.001

Stolp, B., & Fackler, O. T. (2011). How HIV Takes Advantage of the Cytoskeleton in Entry and Replication. *Viruses*, *3*(4), 293–311. https://doi.org/10.3390/v3040293

Stremlau, M., Owens, C. M., Perron, M. J., Kiessling, M., Autissier, P., & Sodroski, J. (2004). The cytoplasmic body component TRIM5α restricts HIV-1 infection in Old World monkeys. *Nature*, *427*(6977), 848–853. https://doi.org/10.1038/nature02343

Subramanyam, D., Lamouille, S., Judson, R. L., Liu, J. Y., Bucay, N., Derynck, R., & Blelloch, R. (2011). Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nature Biotechnology*, *29*(5), 443–448. https://doi.org/10.1038/nbt.1862

Sugiura, M., Kasama, Y., Araki, R., Hoki, Y., Sunayama, M., Uda, M., Nakamura, M., Ando, S., & Abe, M. (2014). Induced Pluripotent Stem Cell Generation-Associated Point Mutations Arise during the Initial Stages of the Conversion of These Cells. *Stem Cell Reports*, *2*(1), 52–63. https://doi.org/10.1016/j.stemcr.2013.11.006

Sullivan, S., Stacey, G. N., Akazawa, C., Aoyama, N., Baptista, R., Bedford, P., Bennaceur Griscelli, A., Chandra, A., Elwood, N., Girard, M., Kawamata, S., Hanatani, T., Latsis, T., Lin, S., Ludwig, T. E., Malygina, T., Mack, A., Mountford, J. C., Noggle, S., … Song, J. (2018). Quality control guidelines for clinical-grade human induced pluripotent stem cell lines. *Regenerative Medicine*, *13*(7), 859–866. https://doi.org/10.2217/rme-2018-0095

Sweeney, N., Merrick, B., Pedro Galão, R., Pickering, S., Botgros, A., Wilson, H. D., Signell, A. W., Betancor, G., Tan, M. K. I., Ramble, J., Kouphou, N., Acors, S., Graham, C., Seow, J., MacMahon, E., Neil, S. J. D., Malim, M. H., Doores, K., Douthwaite, S., … Edgeworth, J. D. (2021). Clinical utility of targeted SARS-CoV-2 serology testing to aid the diagnosis and management of suspected missed, late or post-COVID-19 infection syndromes: Results from a pilot service implemented during the first pandemic wave. *PLOS ONE*, *16*(4), e0249791. https://doi.org/10.1371/journal.pone.0249791

Takahashi, K., Okubo, C., Nakamura, M., Iwasaki, M., Kawahara, Y., Tabata, T., Miyamoto, Y., Woltjen, K., & Yamanaka, S. (2022). A stress-reduced passaging technique improves the viability of human pluripotent cells. *Cell Reports Methods*, *2*(2), 100155. https://doi.org/10.1016/j.crmeth.2021.100155

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, *131*(5), 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Takehisa, J., Kraus, M. H., Ayouba, A., Bailes, E., Van Heuverswyn, F., Decker, J. M., Li, Y., Rudicell, R. S., Learn, G. H., Neel, C., Ngole, E. M., Shaw, G. M., Peeters, M., Sharp, P. M., & Hahn, B. H. (2009). Origin and biology of simian immunodeficiency virus in wild-living western gorillas. *Journal of Virology*, *83*(4), 1635–1648. https://doi.org/10.1128/JVI.02311-08

Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E., & Hammer, S. M. (2008). The challenge of HIV-1 subtype diversity. *The New England Journal of Medicine*, *358*(15), 1590–1602. https://doi.org/10.1056/NEJMra0706737

Tchieu, J., Kuoy, E., Chin, M. H., Trinh, H., Patterson, M., Sherman, S. P., Aimiuwu, O., Lindgren, A., Hakimian, S., Zack, J. A., Clark, A. T., Pyle, A. D., Lowry, W. E., & Plath, K. (2010). Female Human iPSCs Retain an Inactive X Chromosome. *Cell Stem Cell*, *7*(3), 329–342. https://doi.org/10.1016/j.stem.2010.06.024

Teque, F., Ye, L., Xie, F., Wang, J., Morvan, M. G., Kan, Y. W., & Levy, J. A. (2020). Genetically-edited induced pluripotent stem cells derived from HIV-1-infected patients on therapy can give rise to immune cells resistant to HIV-1 infection. *AIDS*, *34*(8), 1141–1149. https://doi.org/10.1097/QAD.0000000000002539

The International HIV Controllers Study. (2010). The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation. *Science*, *330*(6010), 1551–1557. https://doi.org/10.1126/science.1195271

Toccafondi, E., Lener, D., & Negroni, M. (2021). HIV-1 Capsid Core: A Bullet to the Heart of the Target Cell. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.652486

Trautmann, L., Janbazian, L., Chomont, N., Said, E. A., Gimmig, S., Bessette, B., Boulassel, M.-R., Delwart, E., Sepulveda, H., Balderas, R. S., Routy, J.-P., Haddad, E. K., & Sekaly, R.-P. (2006). Upregulation of PD-1 expression on HIV-specific CD8+ T cells leads to reversible immune dysfunction. *Nature Medicine*, *12*(10), 1198–1202. https://doi.org/10.1038/nm1482

Trickey, A., May, M. T., Vehreschild, J. J., Obel, N., Gill, M. J., Crane, H. M., Boesecke, C., Patterson, S., Grabar, S., Cazanave, C., Cavassini, M., Shepherd, L., Monforte, A. d. A., van Sighem, A., Saag, M., Lampe, F., Hernando, V., Montero, M., Zangerle, R., … Sterne, J. A. C. (2017). Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *The Lancet HIV*, *4*(8), e349–e356. https://doi.org/10.1016/S2352-3018(17)30066-8

Usami, Y., Wu, Y., & Göttlinger, H. G. (2015). SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature*, *526*(7572), 218–223. https://doi.org/10.1038/nature15400

Van Damme, N., Goff, D., Katsura, C., Jorgenson, R. L., Mitchell, R., Johnson, M. C., Stephens, E. B., & Guatelli, J. (2008). The Interferon-Induced Protein BST-2 Restricts HIV-1 Release and Is Downregulated from the Cell Surface by the Viral Vpu Protein. *Cell Host & Microbe*, *3*(4), 245–252. https://doi.org/10.1016/j.chom.2008.03.001

Vella, S., Giuliano, M., Dally, L. G., Agresti, M. G., Tomino, C., Floridia, M., Chiesi, A., Fragola, V., Moroni, M., & Piazza, M. (1994). Long-term follow-up of zidovudine therapy in asymptomatic HIV infection: results of a multicenter cohort study. The Italian Zidovudine Evaluation Group. *Journal of Acquired Immune Deficiency Syndromes*, *7*(1), 31–38.

Vidya Vijayan, K. K., Karthigeyan, K. P., Tripathi, S. P., & Hanna, L. E. (2017). Pathophysiology of CD4+ T-Cell Depletion in HIV-1 and HIV-2 Infections. *Frontiers in Immunology*, *8*, 580. https://doi.org/10.3389/fimmu.2017.00580

Vigilante, A., Laddach, A., Moens, N., Meleckyte, R., Leha, A., Ghahramani, A., Culley, O. J., Kathuria, A., Hurling, C., Vickers, A., Wiseman, E., Tewary, M., Zandstra, P. W., Durbin, R., Fraternali, F., Stegle, O., Birney, E., Luscombe, N. M., Danovi, D., & Watt, F. M. (2019). Identifying Extrinsic versus Intrinsic Drivers of Variation in Cell Behavior in Human iPSC Lines from Healthy Donors. *Cell Reports*, *26*(8), 2078-2087.e3. https://doi.org/10.1016/j.celrep.2019.01.094

Wada, T. (1998). Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription invitro. *The EMBO Journal*, *17*(24), 7395–7403. https://doi.org/10.1093/emboj/17.24.7395

Wang, M. Q., Huang, Y. L., Huang, J., Zheng, J. L., & Qian, G. X. (2015). RIG-I detects HIV-1 infection and mediates type I interferon response in human macrophages from patients with HIV-1-associated neurocognitive disorders. *Genetics and Molecular Research*, *14*(4), 13799–13811. https://doi.org/10.4238/2015.October.28.42

Wang, Q., Zhang, X., Han, Y., Wang, X., & Gao, G. (2016). M2BP inhibits HIV-1 virion production in a vimentin filaments-dependent manner. *Scientific Reports*, *6*(1), 32736. https://doi.org/10.1038/srep32736

Warren, L., & Lin, C. (2019). mRNA-Based Genetic Reprogramming. *Molecular Therapy*, *27*(4), 729–734. https://doi.org/10.1016/j.ymthe.2018.12.009

Warren, L., Manos, P. D., Ahfeldt, T., Loh, Y.-H., Li, H., Lau, F., Ebina, W., Mandal, P. K., Smith, Z. D., Meissner, A., Daley, G. Q., Brack, A. S., Collins, J. J., Cowan, C., Schlaeger, T. M., & Rossi, D. J. (2010). Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. *Cell Stem Cell*, *7*(5), 618–630. https://doi.org/10.1016/j.stem.2010.08.012

Wawer, M. J., Gray, R. H., Sewankambo, N. K., Serwadda, D., Li, X., Laeyendecker, O., Kiwanuka, N., Kigozi, G., Kiddugavu, M., Lutalo, T., Nalugoda, F., Wabwire-Mangen, F., Meehan, M. P., & Quinn, T. C. (2005). Rates of HIV-1 Transmission per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda. *The Journal of Infectious Diseases*, *191*(9), 1403–1409. https://doi.org/10.1086/429411

Wei, P., Garber, M. E., Fang, S.-M., Fischer, W. H., & Jones, K. A. (1998). A Novel CDK9-Associated C-Type Cyclin Interacts Directly with HIV-1 Tat and Mediates Its High-Affinity, Loop-Specific Binding to TAR RNA. *Cell*, *92*(4), 451–462. https://doi.org/10.1016/S0092-8674(00)80939-3

Wei, X., Decker, J. M., Wang, S., Hui, H., Kappes, J. C., Wu, X., Salazar-Gonzalez, J. F., Salazar, M. G., Kilby, J. M., Saag, M. S., Komarova, N. L., Nowak, M. A., Hahn, B. H., Kwong, P. D., & Shaw, G. M. (2003). Antibody neutralization and escape by HIV-1. *Nature*, *422*(6929), 307–312. https://doi.org/10.1038/nature01470

Wei, Z., Bodnar, B., Zhao, R.-T., Xiao, Q., Saribas, S., Wang, X., Ho, W.-Z., & Hu, W. (2023). Human iPSC-derived brain organoids: A 3D mini-brain model for studying HIV infection. *Experimental Neurology*, *364*, 114386. https://doi.org/10.1016/j.expneurol.2023.114386

Wells, P. M., Doores, K. J., Couvreur, S., Nunez, R. M., Seow, J., Graham, C., Acors, S., Kouphou, N., Neil, S. J. D., Tedder, R. S., Matos, P. M., Poulton, K., Lista, M. J., Dickenson, R. E., Sertkaya, H., Maguire, T. J. A., Scourfield, E. J., Bowyer, R. C. E., Hart, D., … Steves, C. J. (2020). Estimates of the rate of infection and asymptomatic COVID-19 disease in a population sample from SE England. *Journal of Infection*, *81*(6), 931–936. https://doi.org/10.1016/j.jinf.2020.10.011

Wesselschmidt, R. L. (2011). *The Teratoma Assay: An In Vivo Assessment of Pluripotency* (pp. 231–241). https://doi.org/10.1007/978-1-61779-201-4_17

White, T. E., Brandariz-Nuñez, A., Valle-Casuso, J. C., Amie, S., Nguyen, L. A., Kim, B., Tuzova, M., & Diaz-Griffero, F. (2013). The Retroviral Restriction Ability of SAMHD1, but Not Its Deoxynucleotide Triphosphohydrolase Activity, Is Regulated by Phosphorylation. *Cell Host & Microbe*, *13*(4), 441–451. https://doi.org/10.1016/j.chom.2013.03.005

Whitney, J. B., Hill, A. L., Sanisetty, S., Penaloza-MacMaster, P., Liu, J., Shetty, M., Parenteau, L., Cabral, C., Shields, J., Blackmore, S., Smith, J. Y., Brinkman, A. L., Peter, L. E., Mathew, S. I., Smith, K. M., Borducchi, E. N., Rosenbloom, D. I. S., Lewis, M. G., Hattersley, J., … Barouch, D. H. (2014). Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature*, *512*(7512), 74–77. https://doi.org/10.1038/nature13594

Wilgenburg, B. van, Browne, C., Vowles, J., & Cowley, S. A. (2013). Efficient, Long Term Production of Monocyte-Derived Macrophages from Human Pluripotent Stem Cells under Partly-Defined and Fully-Defined Conditions. *PLoS ONE*, *8*(8), e71098. https://doi.org/10.1371/journal.pone.0071098

Wolfs, T. F. W., Zwart, G., Bakker, M., & Goudsmit, J. (1992). HIV-1 genomic rna diversification following sexual and parenteral virus transmission. *Virology*, *189*(1), 103–110. https://doi.org/10.1016/0042-6822(92)90685-I

Wolinsky, S. M., Wike, C. M., Korber, B. T. M., Hutto, C., Parks, W. P., Rosenblum, L. L., Kunstman, K. J., Furtado, M. R., & Muñoz, J. L. (1992). Selective Transmission of Human Immunodeficiency Virus Type-1 Variants from Mothers to Infants. *Science*, *255*(5048), 1134–1137. https://doi.org/10.1126/science.1546316

Wollert, T., & Hurley, J. H. (2010). Molecular mechanism of multivesicular body biogenesis by ESCRT complexes. *Nature*, *464*(7290), 864–869. https://doi.org/10.1038/nature08849

Woltjen, K., Michael, I. P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., Kaji, K., Sung, H.-K., & Nagy, A. (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature*, *458*(7239), 766–770. https://doi.org/10.1038/nature07863

World Health Organisation. (2022). *HIV/AIDS*. Https://Www.Who.Int/Data/Gho/Data/Themes/Hiv-Aids.

Wu, X., Dao Thi, V. L., Huang, Y., Billerbeck, E., Saha, D., Hoffmann, H.-H., Wang, Y., Silva, L. A. V., Sarbanes, S., Sun, T., Andrus, L., Yu, Y., Quirk, C., Li, M., MacDonald, M. R., Schneider, W. M., An, X., Rosenberg, B. R., & Rice, C. M. (2018). Intrinsic Immunity Shapes Viral Resistance of Stem Cells. *Cell*, *172*(3), 423-438.e25. https://doi.org/10.1016/j.cell.2017.11.018

Yamashita, M., & Emerman, M. (2004). Capsid Is a Dominant Determinant of Retrovirus Infectivity in Nondividing Cells. *Journal of Virology*, *78*(11), 5670–5678. https://doi.org/10.1128/JVI.78.11.5670-5678.2004

Yáñez-Muñoz, R. J., Balaggan, K. S., MacNeil, A., Howe, S. J., Schmidt, M., Smith, A. J., Buch, P., MacLaren, R. E., Anderson, P. N., Barker, S. E., Duran, Y., Bartholomae, C., von Kalle, C., Heckenlively, J. R., Kinnon, C., Ali, R. R., & Thrasher, A. J. (2006). Effective gene therapy with nonintegrating lentiviral vectors. *Nature Medicine*, *12*(3), 348–353. https://doi.org/10.1038/nm1365

Yang, B., Chen, K., Zhang, C., Huang, S., & Zhang, H. (2007). Virion-associated Uracil DNA Glycosylase-2 and Apurinic/Apyrimidinic Endonuclease Are Involved in the Degradation of APOBEC3G-edited Nascent HIV-1 DNA. *Journal of Biological Chemistry*, *282*(16), 11667–11675. https://doi.org/10.1074/jbc.M606864200

Yates, A., Stark, J., Klein, N., Antia, R., & Callard, R. (2007). Understanding the slow depletion of memory CD4+ T cells in HIV infection. *PLoS Medicine*, *4*(5), e177. https://doi.org/10.1371/journal.pmed.0040177

Ye, L., Wang, J., Beyer, A. I., Teque, F., Cradick, T. J., Qi, Z., Chang, J. C., Bao, G., Muench, M. O., Yu, J., Levy, J. A., & Kan, Y. W. (2014). Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Δ32 mutation confers resistance to HIV

infection. *Proceedings of the National Academy of Sciences*, *111*(26), 9591–9596. https://doi.org/10.1073/pnas.1407473111

Ye, L., Wang, J., Teque, F., Xie, F., Tan, Y., Kan, Y. W., & Levy, J. A. (2020). Generation of HIV-1-infected patients' gene-edited induced pluripotent stem cells using feeder-free culture conditions. *AIDS (London, England)*, *34*(8), 1127–1139. https://doi.org/10.1097/QAD.0000000000002535

Yeung, M. L., Houzet, L., Yedavalli, V. S. R. K., & Jeang, K.-T. (2009). A Genome-wide Short Hairpin RNA Screening of Jurkat T-cells for Human Proteins Contributing to Productive HIV-1 Replication. *Journal of Biological Chemistry*, *284*(29), 19463–19473. https://doi.org/10.1074/jbc.M109.010033

Yin, X., Langer, S., Zhang, Z., Herbert, K. M., Yoh, S., König, R., & Chanda, S. K. (2020). Sensor Sensibility—HIV-1 and the Innate Immune Response. *Cells*, *9*(1), 254. https://doi.org/10.3390/cells9010254

Yoshizaki, S., Nishi, M., Kondo, A., Kojima, Y., Yamamoto, N., & Ryo, A. (2011). Vaccination with Human Induced Pluripotent Stem Cells Creates an Antigen-Specific Immune Response Against HIV-1 gp160. *Frontiers in Microbiology*, *2*. https://doi.org/10.3389/fmicb.2011.00004

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I. I., & Thomson, J. A. (2009). Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences. *Science*, *324*(5928), 797–801. https://doi.org/10.1126/science.1172482

Yu, J., Li, M., Wilkins, J., Ding, S., Swartz, T. H., Esposito, A. M., Zheng, Y.-M., Freed, E. O., Liang, C., Chen, B. K., & Liu, S.-L. (2015). IFITM Proteins Restrict HIV-1 Infection by Antagonizing the Envelope Glycoprotein. *Cell Reports*, *13*(1), 145–156. https://doi.org/10.1016/j.celrep.2015.08.055

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., & Thomson, J. A. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science*, *318*(5858), 1917–1920. https://doi.org/10.1126/science.1151526

Zeng, C., Waheed, A. A., Li, T., Yu, J., Zheng, Y.-M., Yount, J. S., Wen, H., Freed, E. O., & Liu, S.-L. (2021). SERINC proteins potentiate antiviral type I IFN production and proinflammatory signaling pathways. *Science Signaling*, *14*(700). https://doi.org/10.1126/scisignal.abc7611

Zhang, H., Yang, B., Pomerantz, R. J., Zhang, C., Arunachalam, S. C., & Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature*, *424*(6944), 94–98. https://doi.org/10.1038/nature01707

Zhang, L. Q., MacKenzie, P., Cleland, A., Holmes, E. C., Brown, A. J., & Simmonds, P. (1993). Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *Journal of Virology*, *67*(6), 3345–3356. https://doi.org/10.1128/jvi.67.6.3345-3356.1993

Zhang, Z.-Q., Schuler, T., Zupancic, M., Wietgrefe, S., Staskus, K. A., Reimann, K. A., Reinhart, T. A., Rogan, M., Cavert, W., Miller, C. J., Veazey, R. S., Notermans, D., Little, S., Danner, S. A., Richman, D. D., Havlir, D., Wong, J., Jordan, H. L., Schacker, T. W., … Haase, A. T. (1999). Sexual Transmission and Propagation of SIV and HIV in Resting and Activated CD4 $^{+}$ T Cells. *Science*, *286*(5443), 1353–1357. https://doi.org/10.1126/science.286.5443.1353

Zhao, G., Perilla, J. R., Yufenyuy, E. L., Meng, X., Chen, B., Ning, J., Ahn, J., Gronenborn, A. M., Schulten, K., Aiken, C., & Zhang, P. (2013). Mature HIV-1 capsid structure by cryo-

electron microscopy and all-atom molecular dynamics. *Nature*, *497*(7451), 643–646. https://doi.org/10.1038/nature12162

Zhou, H., Xu, M., Huang, Q., Gates, A. T., Zhang, X. D., Castle, J. C., Stec, E., Ferrer, M., Strulovici, B., Hazuda, D. J., & Espeseth, A. S. (2008). Genome-Scale RNAi Screen for Host Factors Required for HIV Replication. *Cell Host & Microbe*, *4*(5), 495–504. https://doi.org/10.1016/j.chom.2008.10.004

Zhou, M., Halanski, M. A., Radonovich, M. F., Kashanchi, F., Peng, J., Price, D. H., & Brady, J. N. (2000). Tat Modifies the Activity of CDK9 To Phosphorylate Serine 5 of the RNA Polymerase II Carboxyl-Terminal Domain during Human Immunodeficiency Virus Type 1 Transcription. *Molecular and Cellular Biology*, *20*(14), 5077–5086. https://doi.org/10.1128/MCB.20.14.5077-5086.2000

Zhou, W., & Freed, C. R. (2009). Adenoviral Gene Delivery Can Reprogram Human Fibroblasts to Induced Pluripotent Stem Cells. *Stem Cells*, *27*(11), 2667–2674. https://doi.org/10.1002/stem.201

Zhu, T., Mo, H., Wang, N., Nam, D. S., Cao, Y., Koup, R. A., & Ho, D. D. (1993). Genotypic and Phenotypic Characterization of HIV-1 Patients with Primary Infection. *Science*, *261*(5125), 1179–1181. https://doi.org/10.1126/science.8356453

Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B., & Price, D. H. (1997). Transcription elongation factor P-TEFb is required for HIV-1 Tat transactivation in vitro. *Genes & Development*, *11*(20), 2622–2632. https://doi.org/10.1101/gad.11.20.2622

Zila, V., Margiotta, E., Turoňová, B., Müller, T. G., Zimmerli, C. E., Mattei, S., Allegretti, M., Börner, K., Rada, J., Müller, B., Lusic, M., Kräusslich, H.-G., & Beck, M. (2021). Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell*, *184*(4), 1032-1046.e18. https://doi.org/10.1016/j.cell.2021.01.025

Supplementary Files

Supplementary file 2.1: Full detailed version of the Operetta algorithm employed to quantify the percentage of GFP positive cells in the infectivity assay.

Supplementary file 2.2: Full detailed version of the Operetta algorithm employed to quantify the percentage of Gag positive cells in the late gene expression assay.

Supplementary file 4.1: All MACS PBMC sample-derived iPSC lines demonstrating typical iPSC morphological characteristics.

A) Light microscopy images of all MACS iPSC lines. Images taken at 4X magnification and scale bar represents 500µm.

Supplementary file 4.2: Directed differentiation of all MACS iPSC lines into the three germ layers and analysis of OCT-4 expression before and after differentiation.

A+B+C) Immunofluorescence images demonstrating the ability of an iPSC line to differentiate into ectoderm, mesoderm and endoderm lineages respectively. Top left and right images show undifferentiated iPSCs and lineage differentiated cells stained with DAPI respectively. Bottom left and right images display the same undifferentiated iPSCs and lineage differentiated cells stained for that lineage's specific marker respectively. **D)** Immunofluorescence images to show OCT-4 expression before and after lineage specific differentiation. Images taken at 20X magnification and scale bars represent 100µm.

Supplementary file 4.3: G-banded karyotyping analysis of the MACS iPSC line 20722A.

A) Representative image of the karyotype for MACS iPSC line, 20722A, showing a normal male diploid karyotype 46,XY for 20 examined cells at passage 20.

Supplementary file 2.1

Analysis Sequence "20230802 Transduced GFP Linear Classifier"

Input Image	Input
	Flatfield Correction : Advanced Brightfield Correction **Stack Processing :** Maximum Projection **Min. Global Binning :** Dynamic

Filter Image	Input	Method	Output
	Channel : DAPI	**Method :** Sliding Parabola Curvature : 7	Output Image : Sliding Parabola

Find Nuclei	Input	Method	Output
	Channel : Sliding Parabola **ROI :** None	**Method :** C Common Threshold : 0.4 Area : > 30 µm² Splitting Coefficient : 10 Individual Threshold : 0.4 Contrast : > 0.1	Output Population : Nuclei

Calculate Intensity Properties	Input	Method	Output
	Channel : DAPI **Population :** Nuclei **Region :** Nucleus	**Method :** Standard Maximum Minimum	Property Prefix : Intensity Nucleus DAPI

Calculate Morphology Properties	Input	Method	Output
	Population : Nuclei **Region :** Nucleus	**Method :** Standard Area	Property Prefix : Nucleus

Select Population	Input	Method	Output
	Population : Nuclei	**Method :** Filter by Property Intensity Nucleus DAPI Maximum : < 50000 Nucleus Area [µm²] : > 80 Nucleus Area [µm²] : < 1000 Boolean Operations : F1 and F2 and F3	Output Population : Cells

Find Cytoplasm	Input	Method	Output
	Channel : Alexa 647 **Nuclei :** Cells	**Method :** A Individual Threshold : 0.2	

Calculate Intensity Properties (2)	Input	Method	Output
	Channel : EGFP **Population :** Cells **Region :** Cell	**Method :** Standard Standard Deviation Coefficient of Variance Median Sum Maximum Minimum	Property Prefix : Intensity Cell EGFP

Calculate Texture Properties	Input	Method	Output
	Channel : EGFP **Population :** Cells **Region :** Cell	**Method :** SER Features Scale : 1 px Normalization by : Kernel SER Spot SER Hole SER Edge SER Ridge SER Valley SER Saddle SER Bright SER Dark	Property Prefix : Cell EGFP

Select Population (2)	Input	Method	Output
	Population : Cells	**Method :** Linear Classifier Number of Classes : 2 Intensity Nucleus DAPI Maximum Intensity Nucleus DAPI Minimum Nucleus Area [µm²] Intensity Cell EGFP StdDev Intensity Cell EGFP Median Intensity Cell EGFP Maximum Intensity Cell EGFP Minimum Intensity Cell EGFP Sum Intensity Cell EGFP CV [%] Cell EGFP SER Spot 1 px Cell EGFP SER Hole 1 px Cell EGFP SER Edge 1 px Cell EGFP SER Ridge 1 px Cell EGFP SER Valley 1 px Cell EGFP SER Saddle 1 px Cell EGFP SER Bright 1 px Cell EGFP SER Dark 1 px	Output Population A : Transduced Output Population B : Non-Transduced

Define Results	Results
	Method : List of Outputs

Population : Cells
Number of Objects

Population : Transduced
Number of Objects
Intensity Cell EGFP StdDev : StdDev
Intensity Cell EGFP Median : Median

Population : Non-Transduced
Number of Objects
Intensity Cell EGFP StdDev : StdDev
Intensity Cell EGFP Median : Median

Method : Formula Output
Formula : 100*a/b
Population Type : Objects
Variable a : Transduced - Number of Objects
Variable b : Cells - Number of Objects
Output Name : % Transduced

Method : Formula Output
Formula : 100*a/b
Population Type : Objects
Variable a : Non-Transduced - Number of Objects
Variable b : Cells - Number of Objects
Output Name : % Non-Transduced

Object Results
Population : Cells : None
Population : Nuclei : None
Population : Transduced : None
Population : Non-Transduced : None

Acapella version: 5.1.2.129056. Timestamp: 2023-10-12 10:26:53 -0400.

Supplementary file 2.2

Analysis Sequence "20230807 Viral Linear Classiier + /- with actin stain"

Input Image	Input
	Flatfield Correction : Basic Brightfield Correction **Stack Processing :** Maximum Projection **Min. Global Binning :** Dynamic

Filter Image	Input	Method	Output
	Channel : DAPI	**Method :** Sliding Parabola Curvature : 4	Output Image : Sliding Parabola

Find Nuclei	Input	Method	Output
	Channel : Sliding Parabola **ROI :** None	**Method :** C Common Threshold : 0.35 Area : > 30 µm² Splitting Coefficient : 3.5 Individual Threshold : 0.35 Contrast : > 0.05	Output Population : Nuclei

Calculate Morphology Properties	Input	Method	Output
	Population : Nuclei **Region :** Nucleus	**Method :** Standard Roundness	Property Prefix : Nucleus

Calculate Intensity Properties (2)	Input	Method	Output
	Channel : DAPI **Population :** Nuclei **Region :** Nucleus	**Method :** Standard Median	Property Prefix : Intensity Nucleus DAPI

Select Population (3)	Input	Method	Output
	Population : Nuclei	**Method :** Filter by Property Nucleus Roundness : > 0.6 Intensity Nucleus DAPI Median : > 1500 Boolean Operations : F1 and F2	Output Population : Nuclei Selected

Filter Image (2)	Input	Method	Output
	Channel : Alexa 647	**Method :** Texture SER Filter : SER Ridge Scale : 2 µm	Output Image : SER Ridge

		Normalization by : Region Intensity	

Find Cytoplasm	Input	Method	Output
	Channel : Alexa 647 **Nuclei :** Nuclei Selected	**Method :** F Membrane Channel : SER Ridge Individual Threshold : 0.3	

Select Population (2)	Input	Method	Output
	Population : Nuclei Selected	**Method :** Common Filters Remove Border Objects Region : Cell	Output Population : Cells

Calculate Intensity Properties	Input	Method	Output
	Channel : Alexa 555 **Population :** Cells **Region :** Cytoplasm	**Method :** Standard Mean Standard Deviation Coefficient of Variance Median Sum Maximum Minimum	Property Prefix : Intensity Cytoplasm Alexa 555

Calculate Texture Properties	Input	Method	Output
	Channel : Alexa 555 **Population :** Cells **Region :** Cell	**Method :** SER Features Scale : 1 px Normalization by : Kernel SER Spot SER Hole SER Edge SER Ridge SER Valley SER Saddle SER Bright SER Dark	Property Prefix : Cell Alexa 555

Select Population	Input	Method	Output
	Population : Cells	**Method :** Linear Classifier Number of Classes : 2 Nucleus Roundness Intensity Nucleus DAPI Median Intensity Cytoplasm Alexa 555 Mean Intensity Cytoplasm Alexa 555 StdDev Intensity Cytoplasm Alexa 555 Median Intensity Cytoplasm Alexa 555 Maximum	Output Population A : Positive Cells Output Population B : Negative Cells

		Intensity Cytoplasm Alexa 555 Minimum Intensity Cytoplasm Alexa 555 Sum Intensity Cytoplasm Alexa 555 CV [%] Cell Alexa 555 SER Spot 1 px Cell Alexa 555 SER Hole 1 px Cell Alexa 555 SER Edge 1 px Cell Alexa 555 SER Ridge 1 px Cell Alexa 555 SER Valley 1 px Cell Alexa 555 SER Saddle 1 px Cell Alexa 555 SER Bright 1 px Cell Alexa 555 SER Dark 1 px	

Define Results

Results

Method : List of Outputs
Population : Cells
Number of Objects

Population : Positive Cells
Number of Objects

Population : Negative Cells
Number of Objects

Method : Formula Output
Formula : 100*a/b
Population Type : Objects
Variable a : Positive Cells - Number of Objects
Variable b : Cells - Number of Objects
Output Name : % P24+ Cells

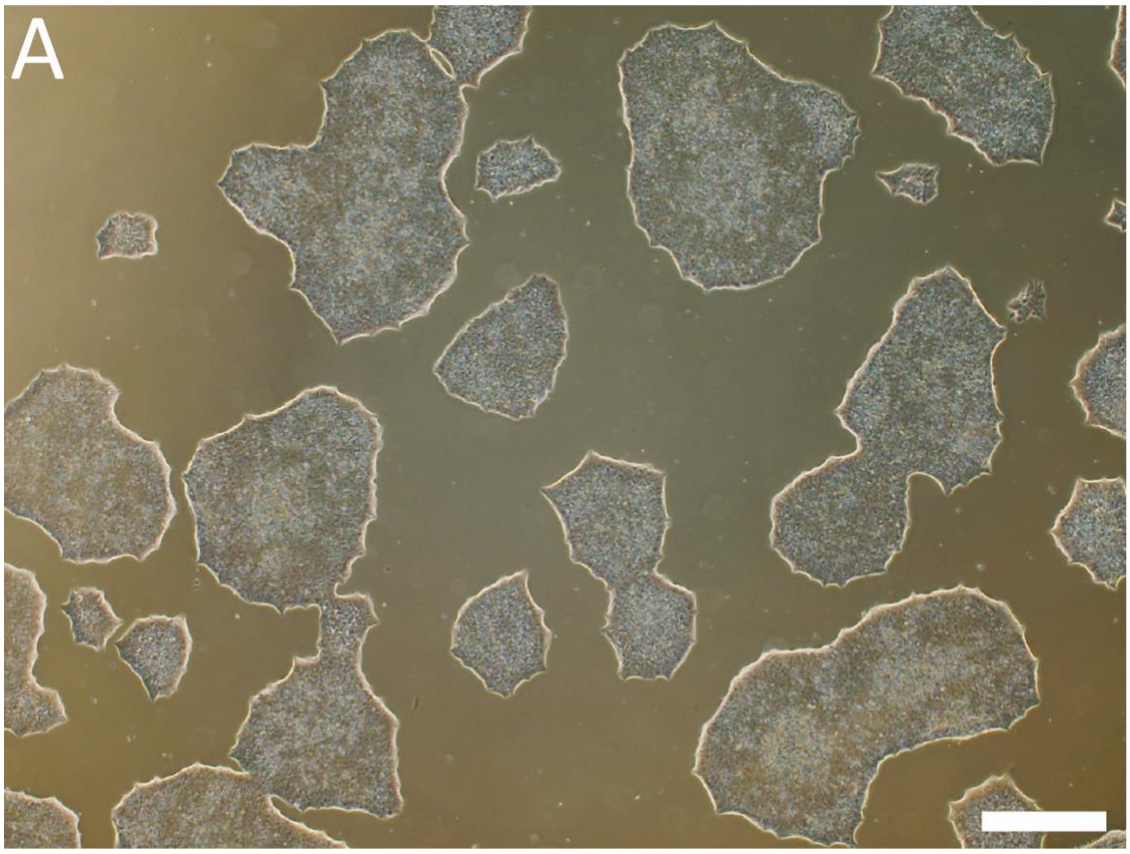
Method : Formula Output
Formula : 100*a/b
Population Type : Objects
Variable a : Negative Cells - Number of Objects
Variable b : Cells - Number of Objects
Output Name : % P24- Cells

Object Results
Population : Cells : None
Population : Nuclei Selected : None
Population : Nuclei : None
Population : Positive Cells : None
Population : Negative Cells : None

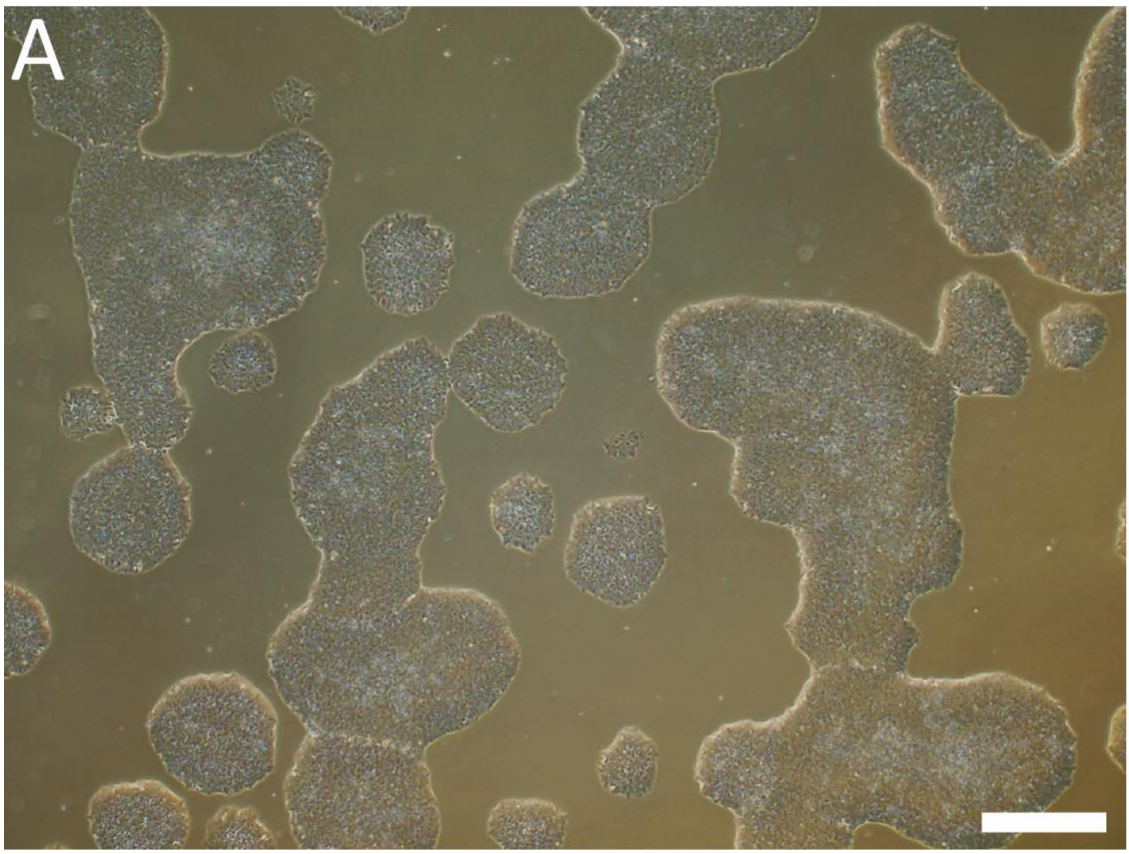
Acapella version: 5.1.2.129056. Timestamp: 2023-10-12 10:28:06 -0400.

Supplementary file 4.1

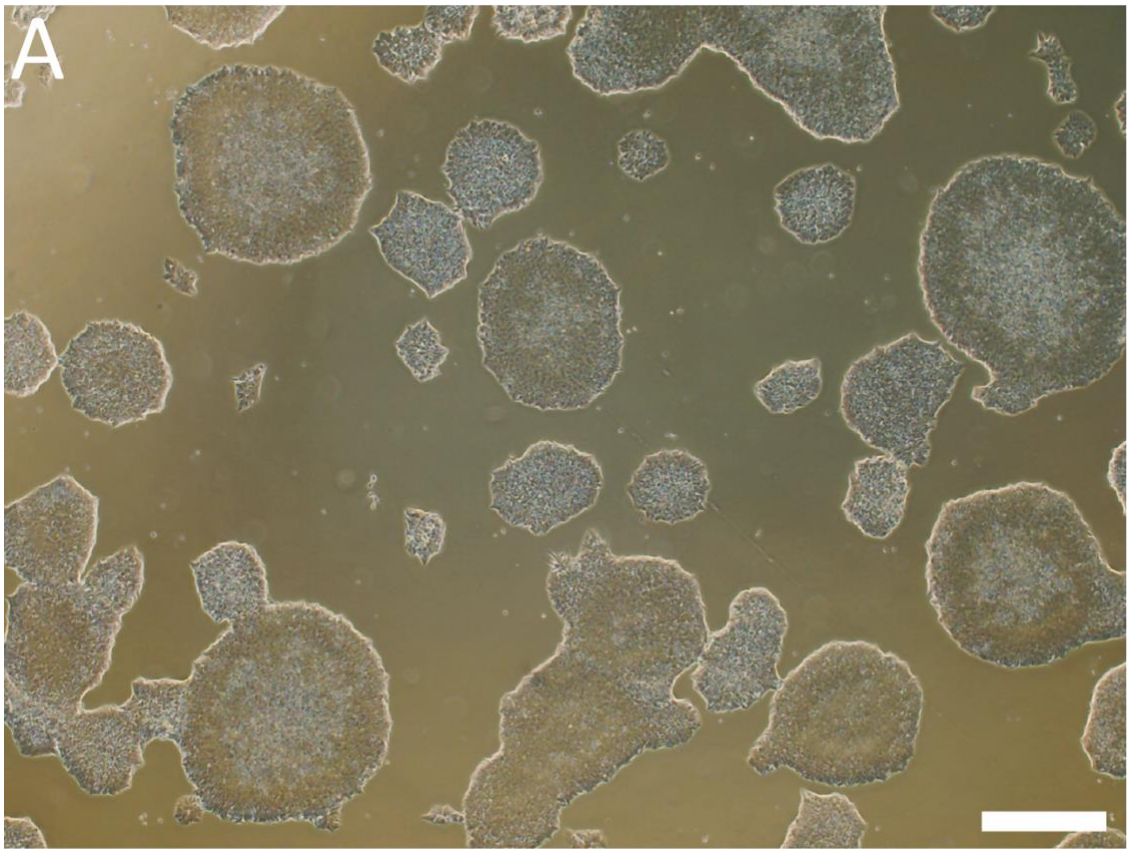
20032A



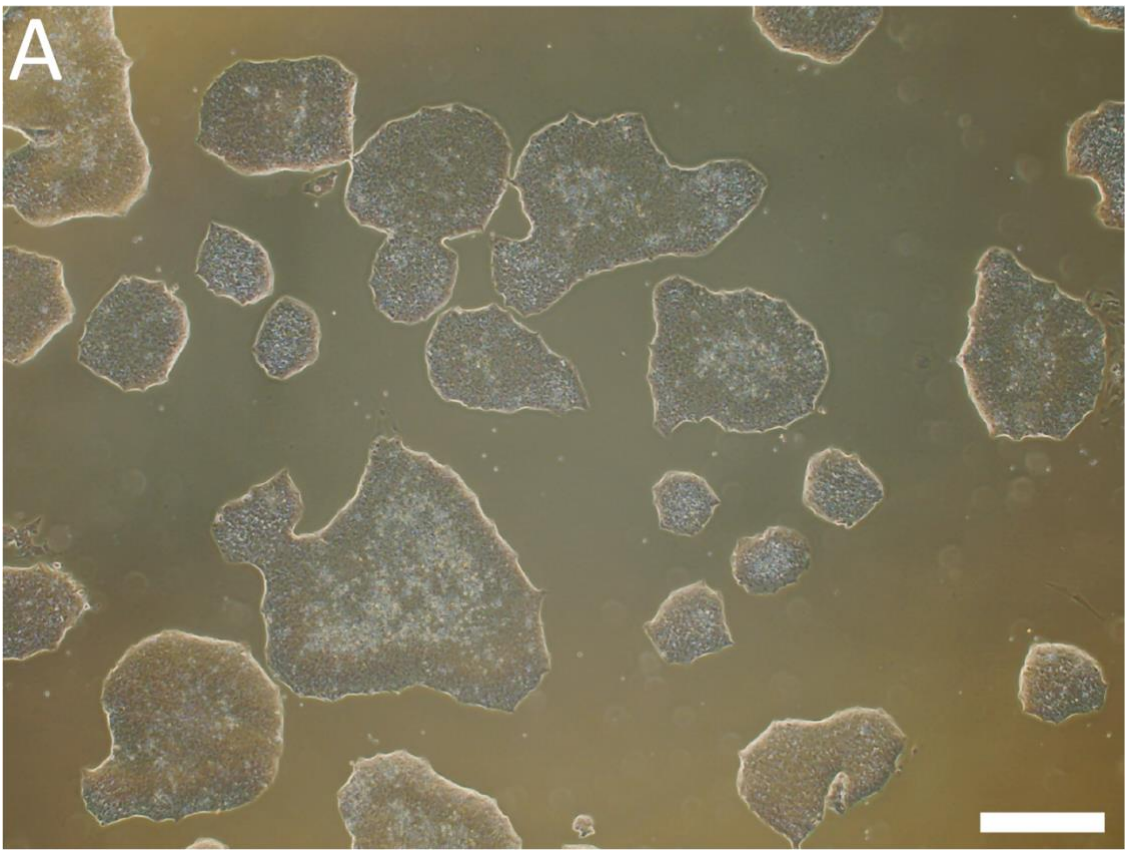
20032B



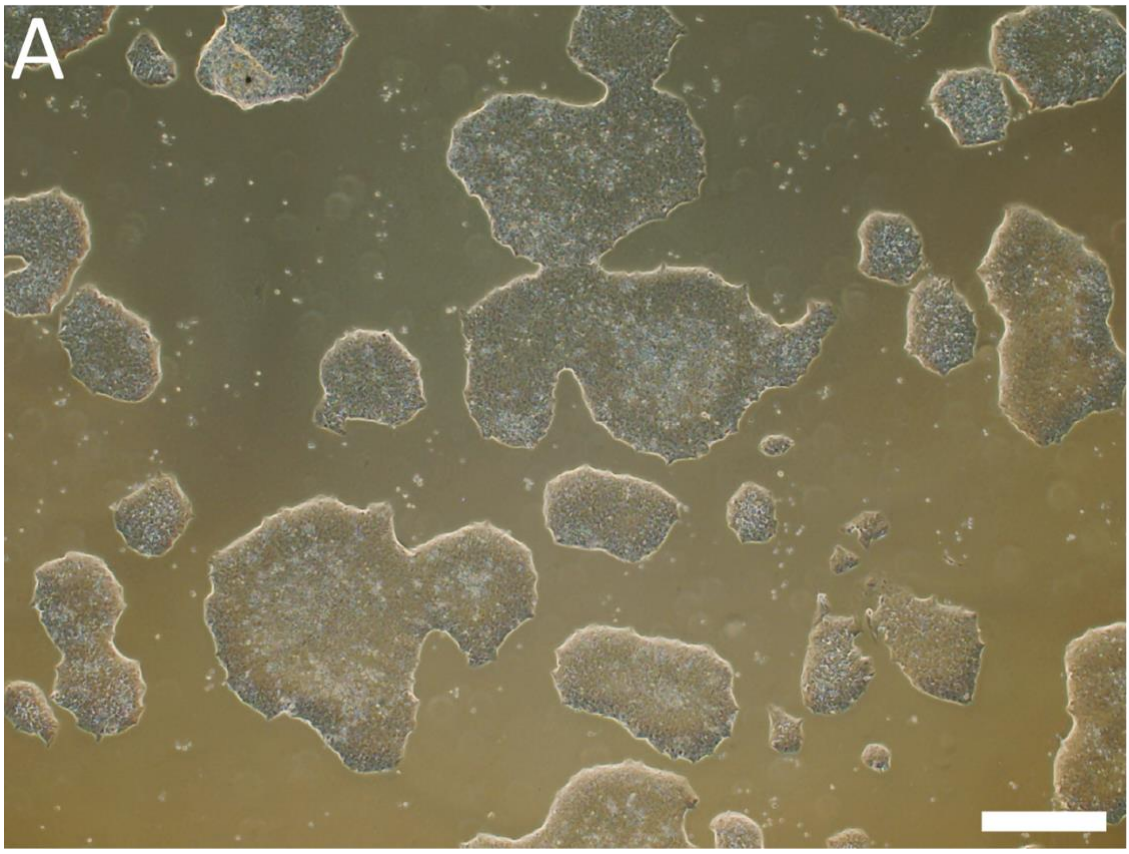
20032C



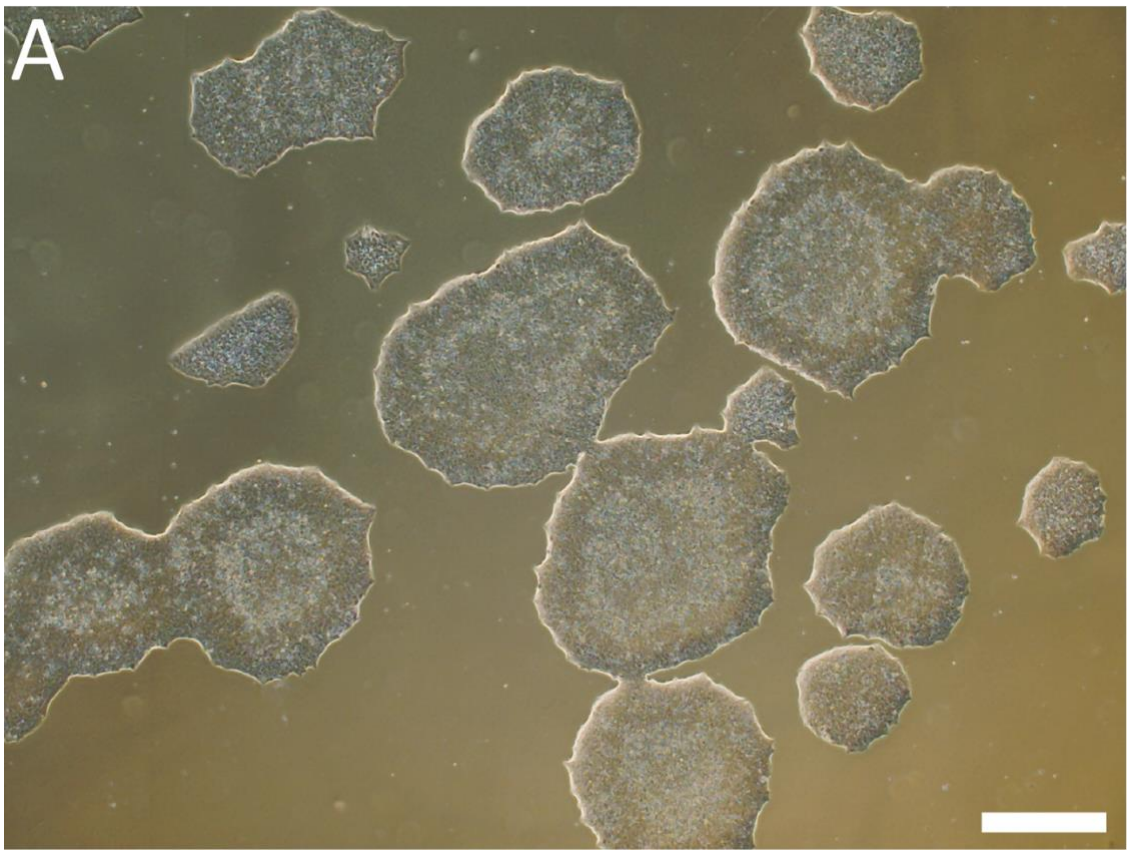
20318A



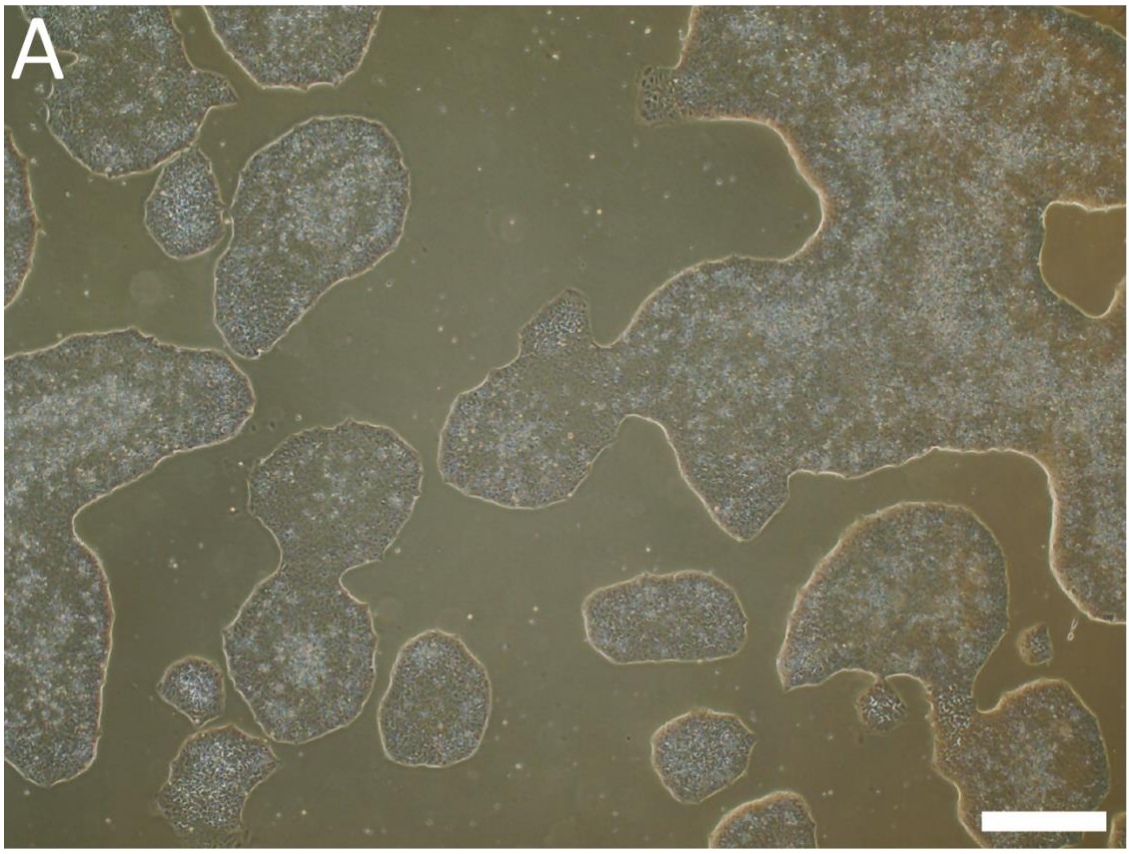
20318B



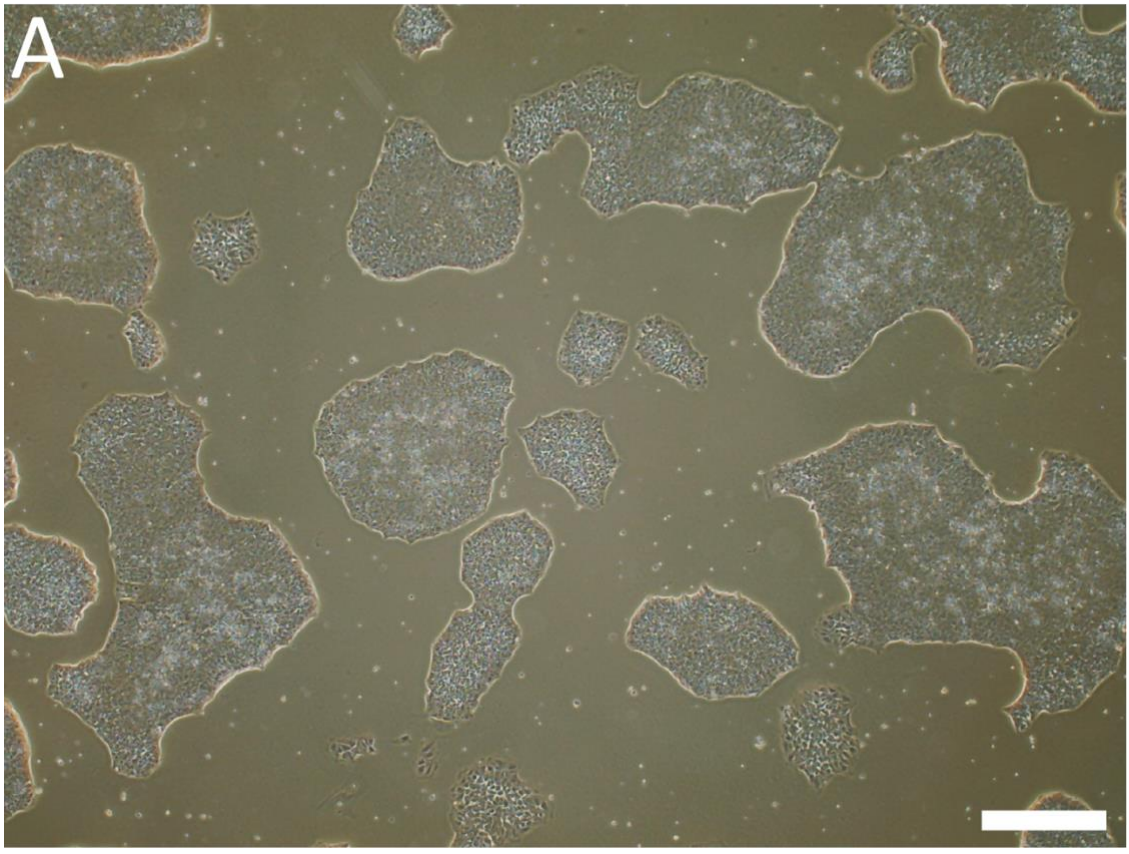
20318C



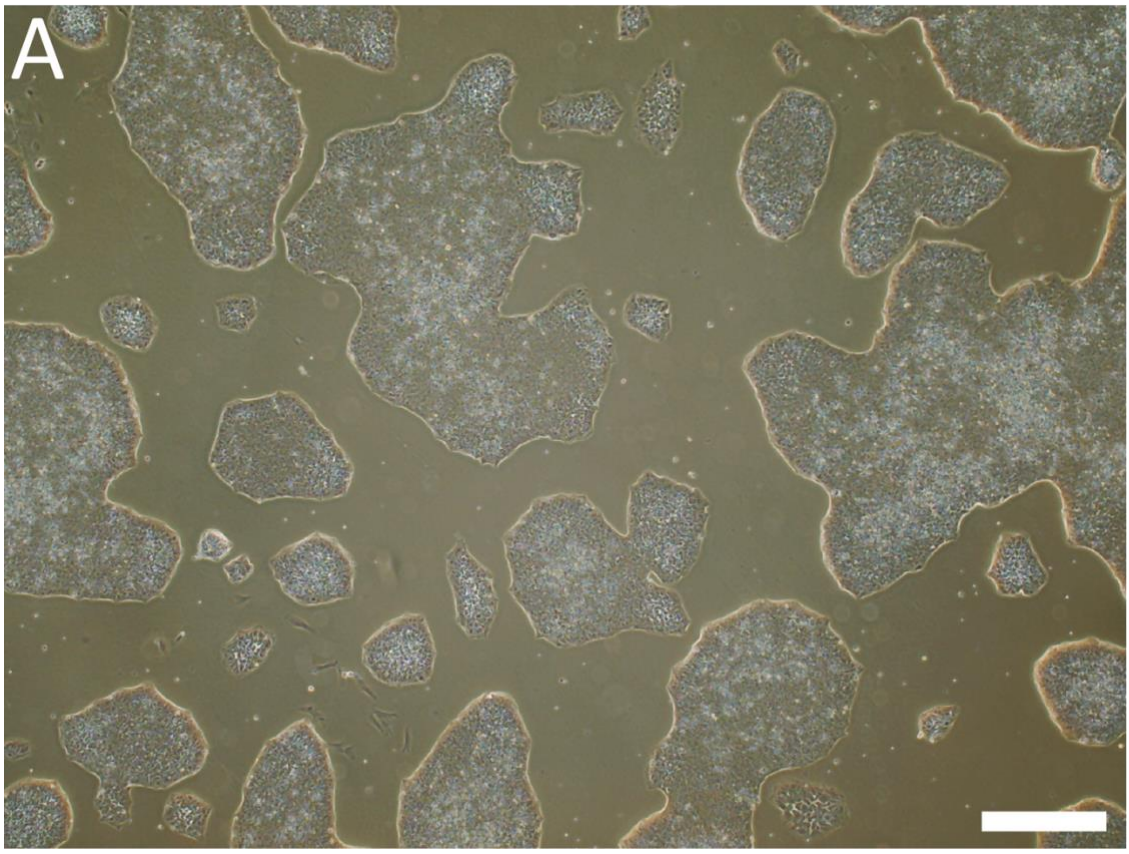
20480A



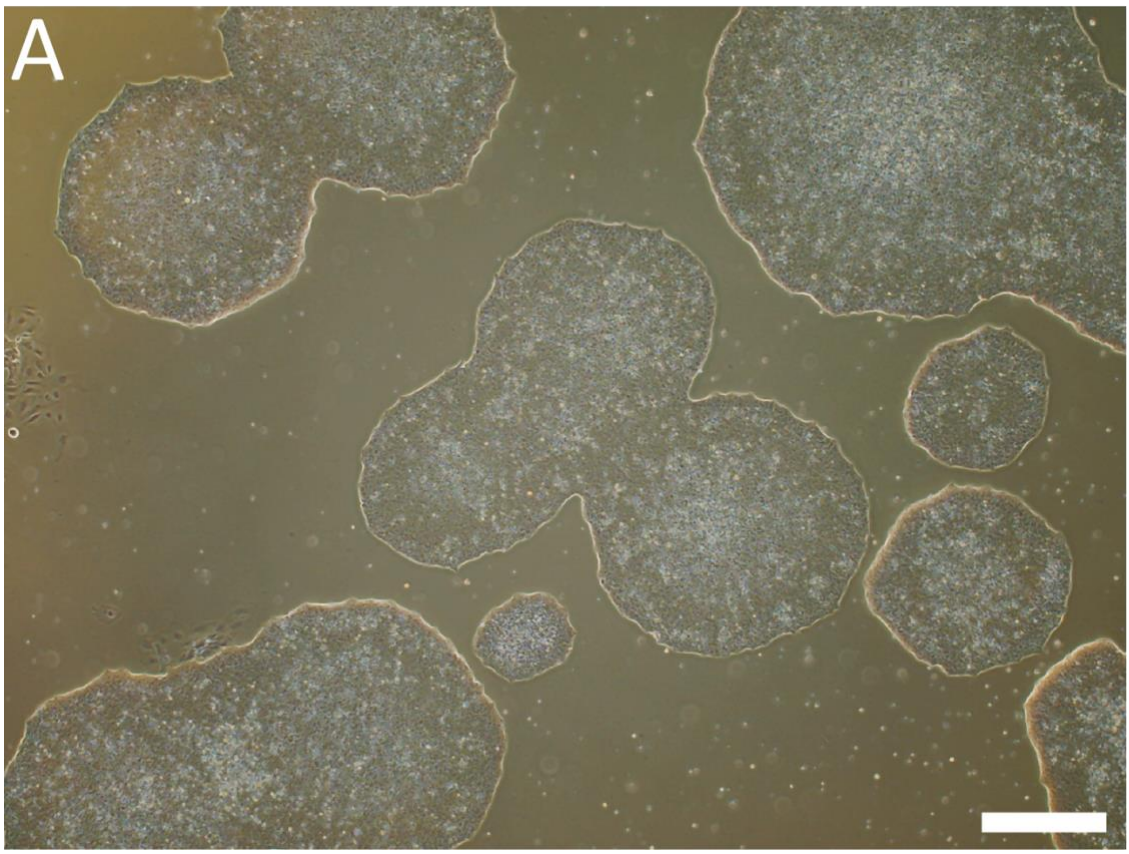
20480B



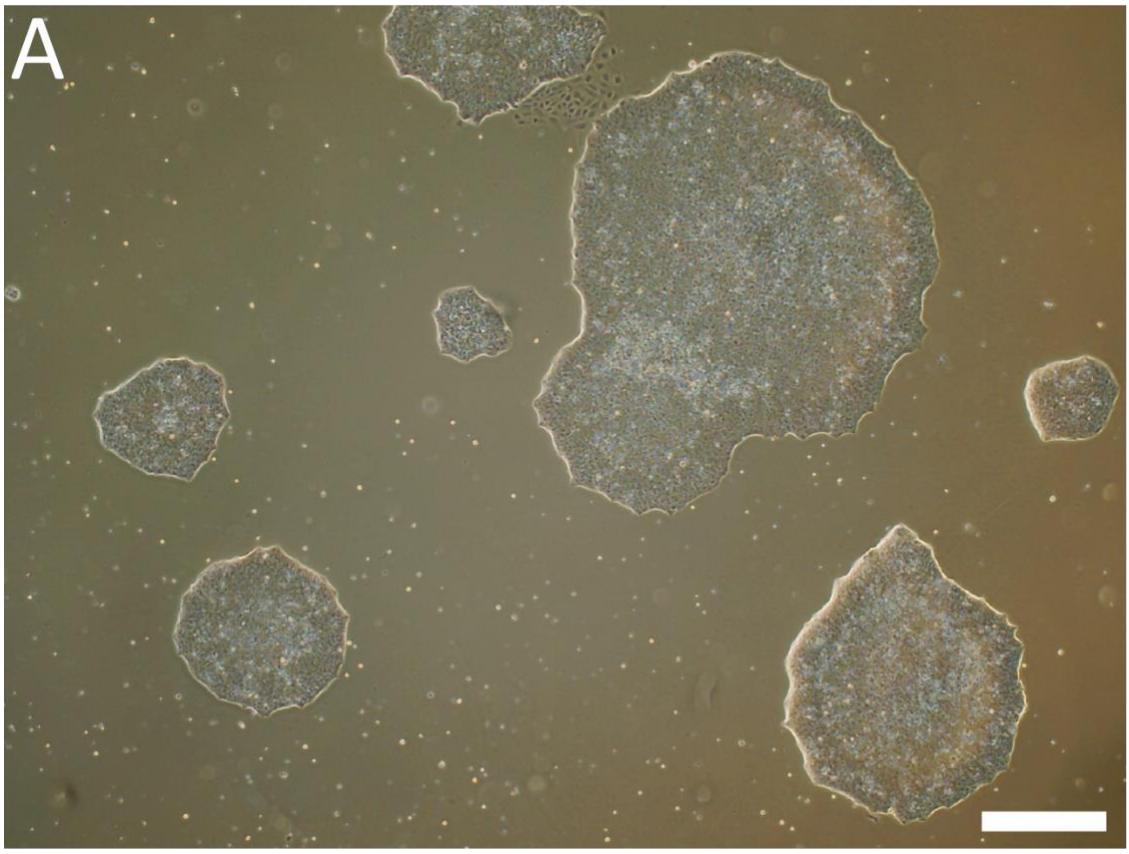
20480C



20543A

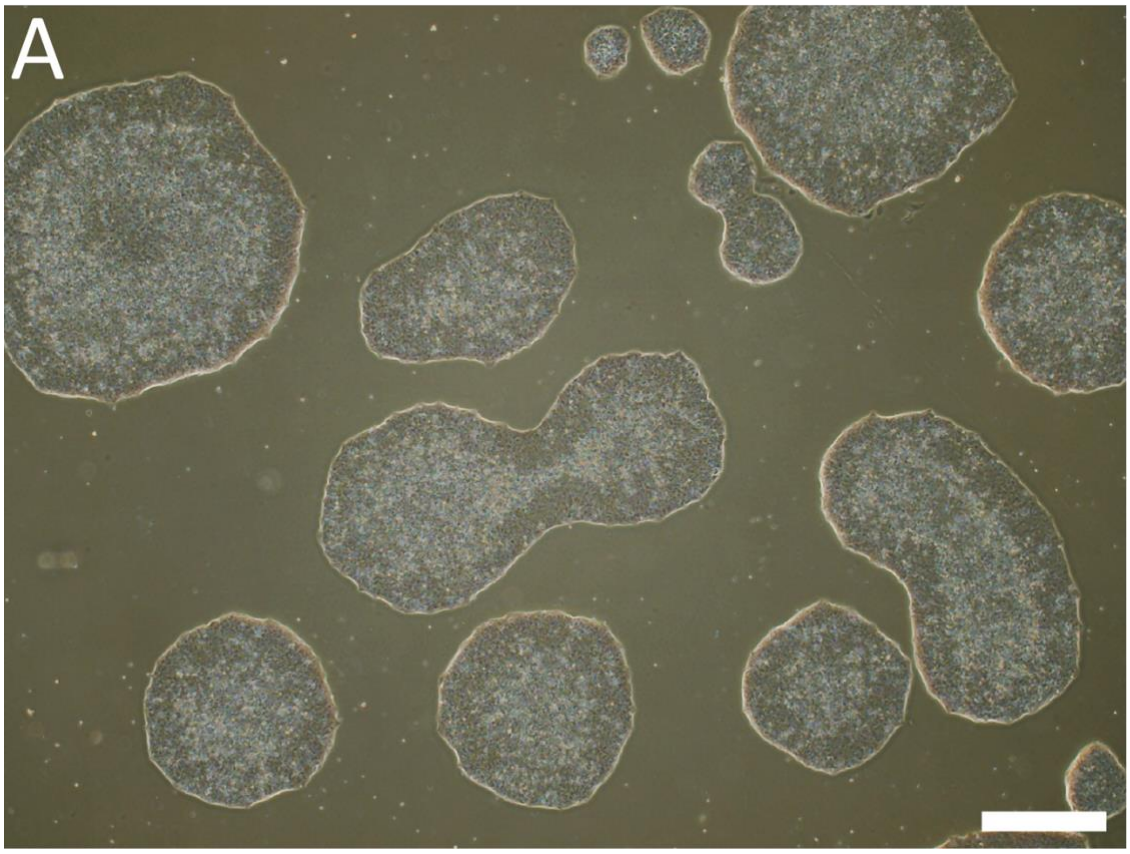


20543B

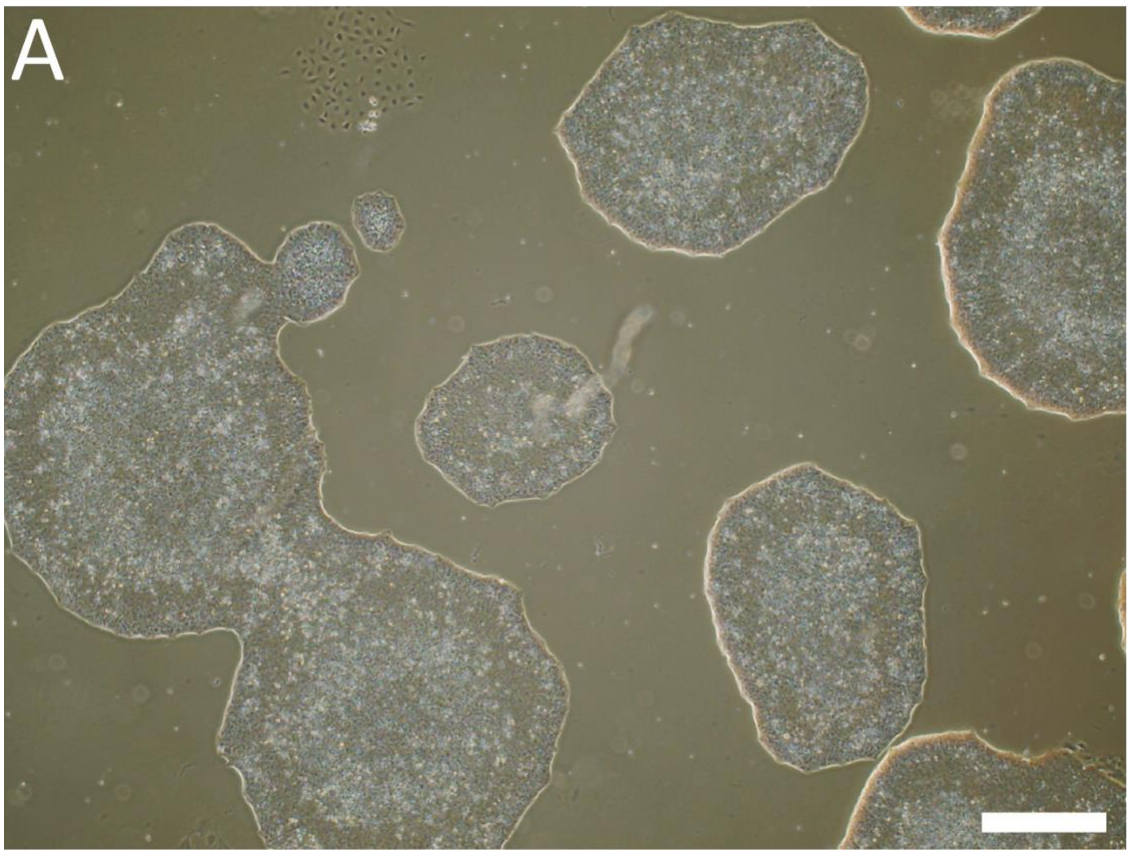


24020A

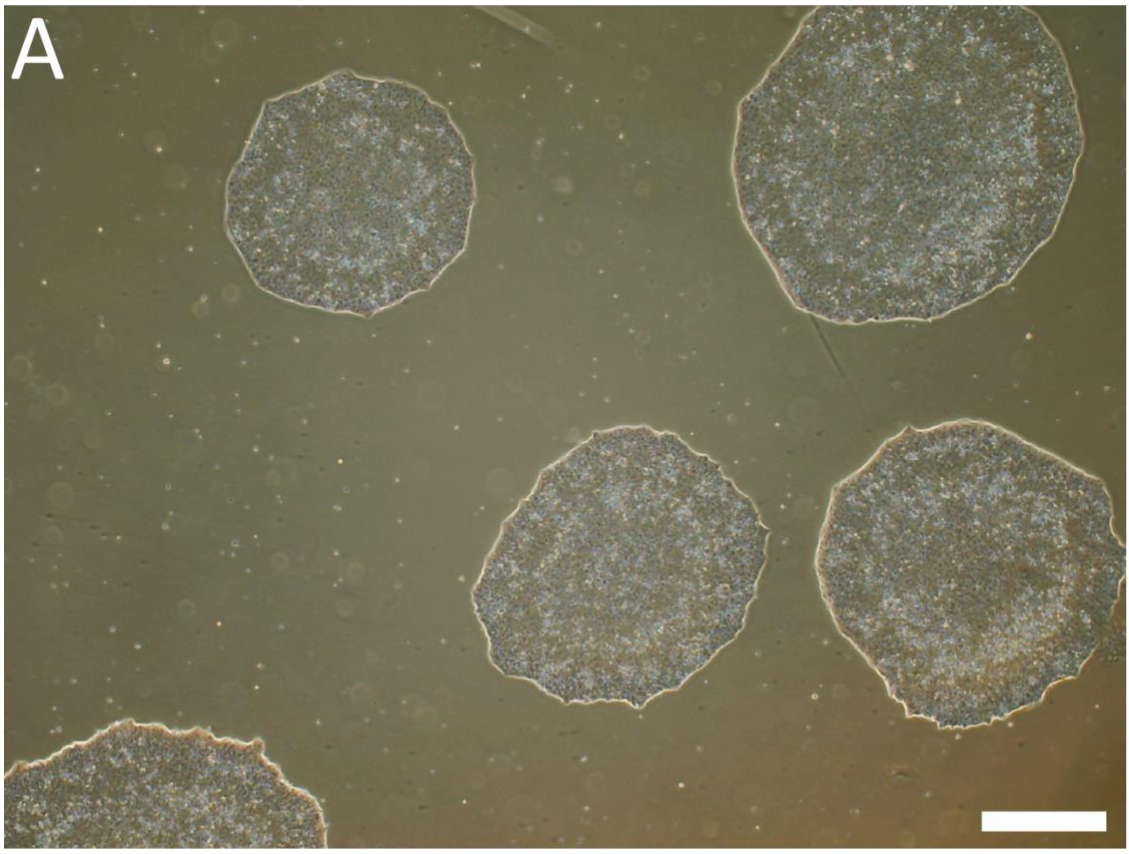
24020B



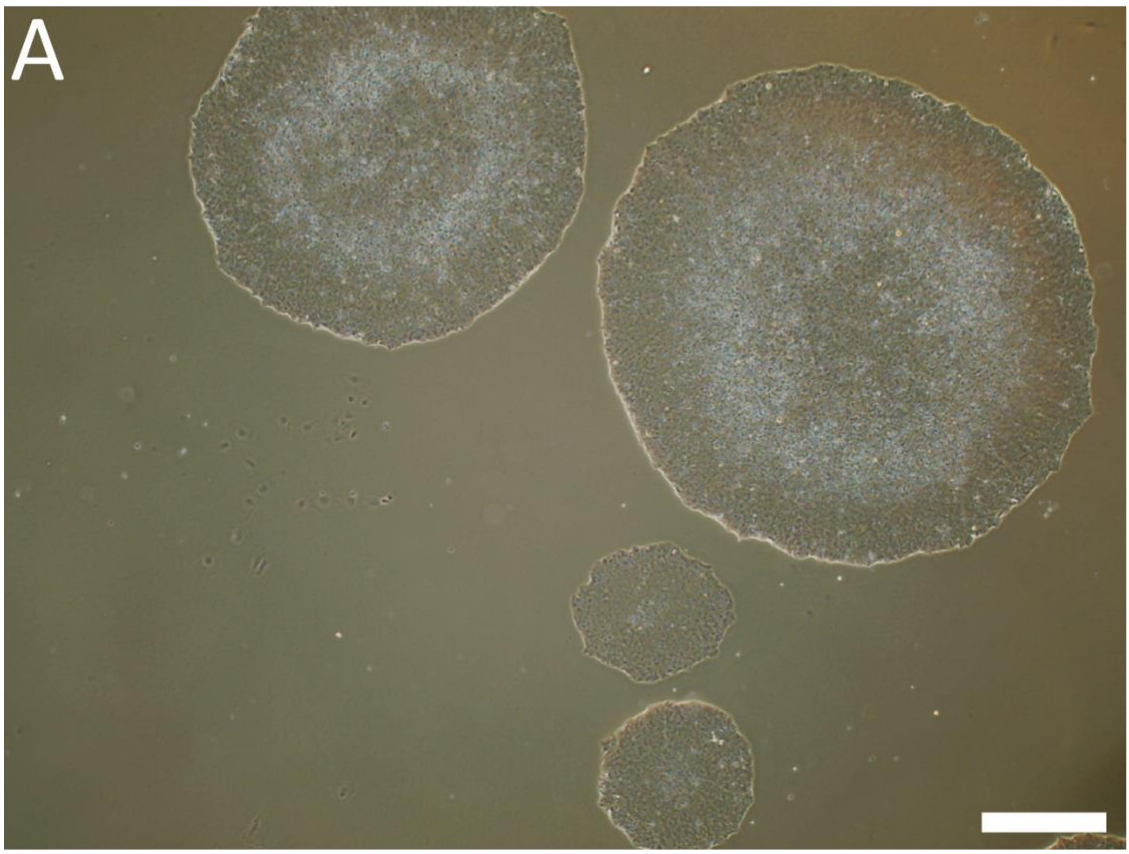
24020C



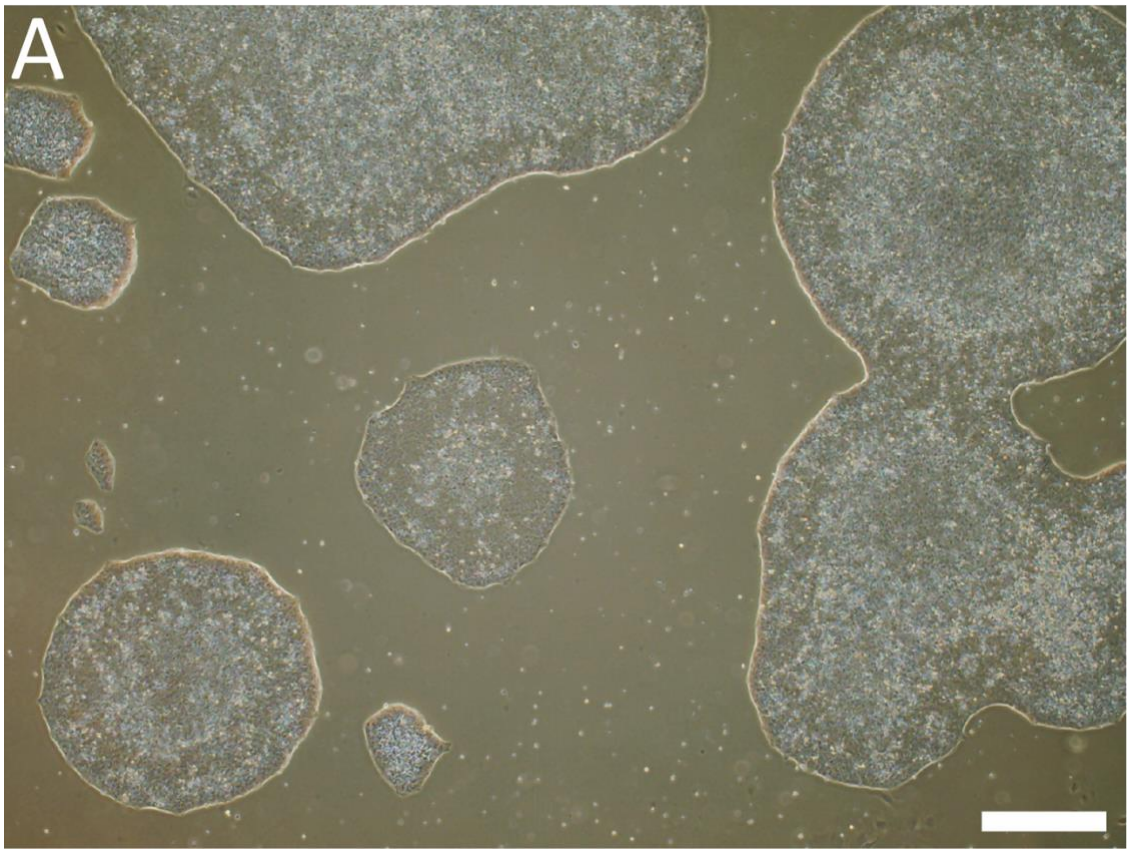
24208A



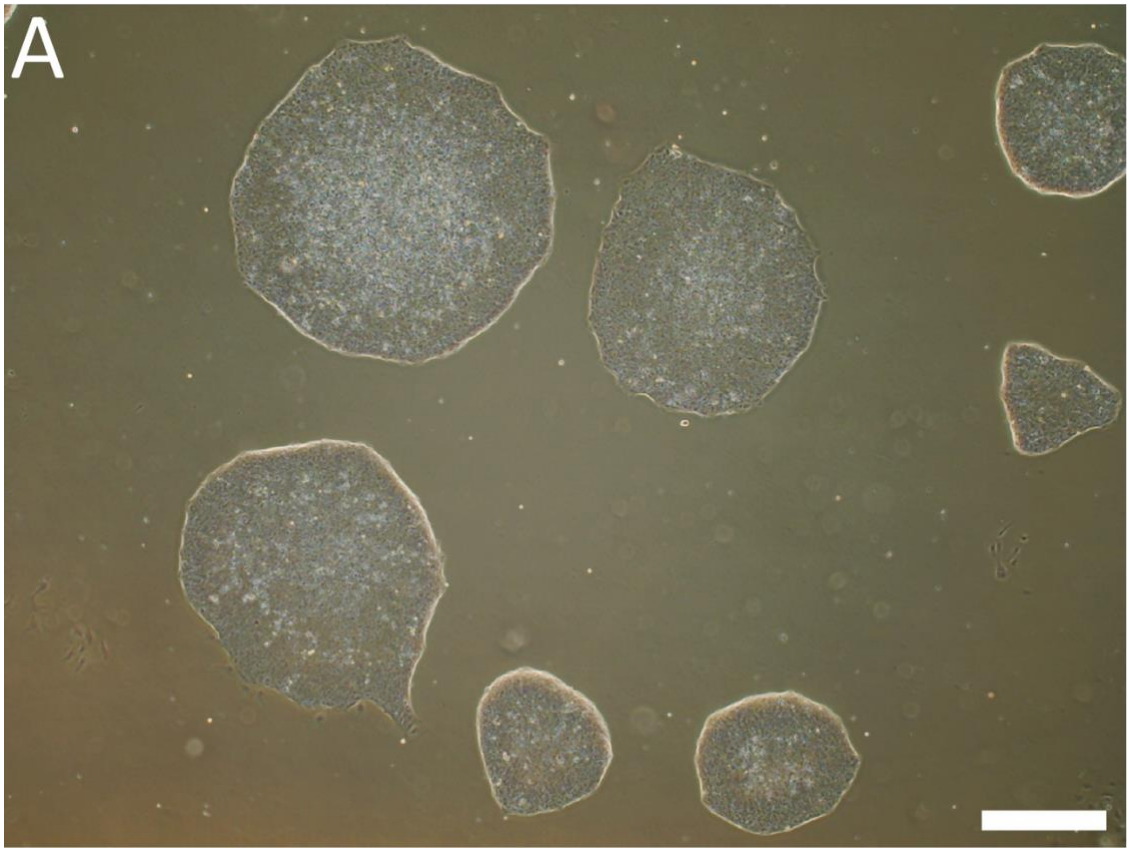
24208B



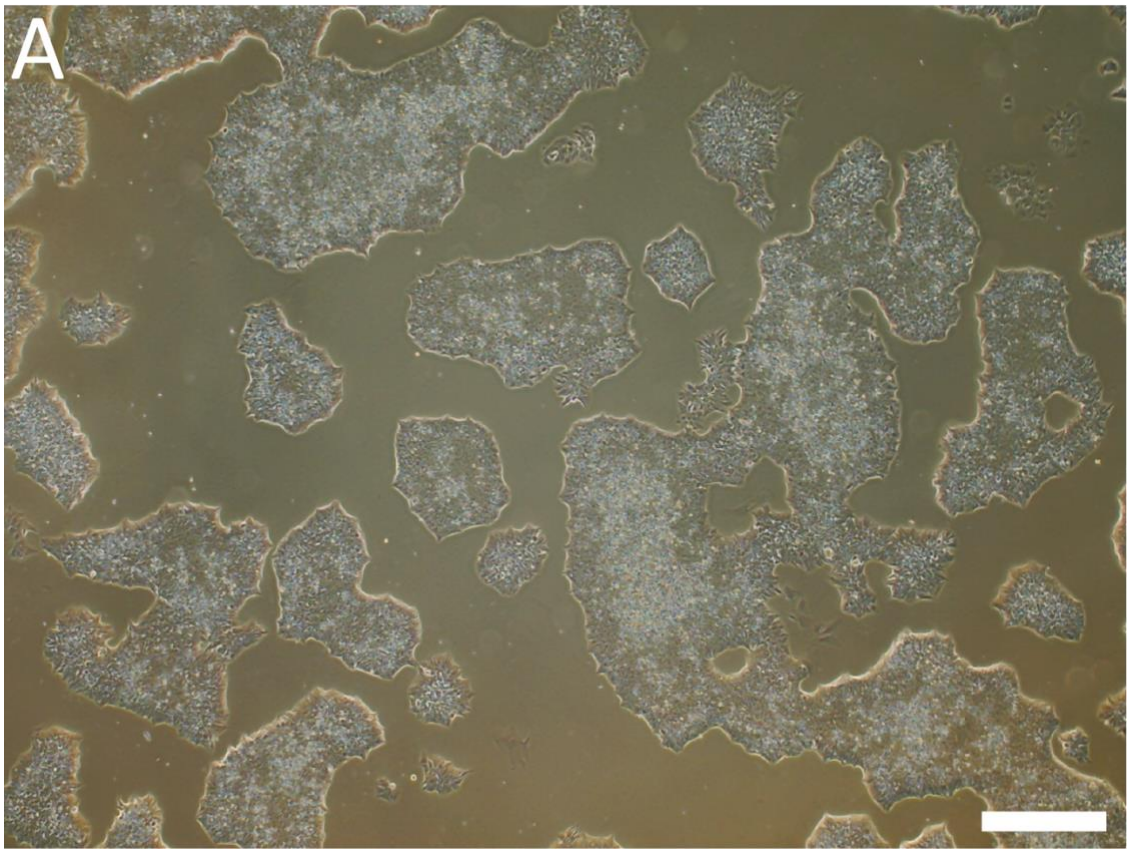
24312A



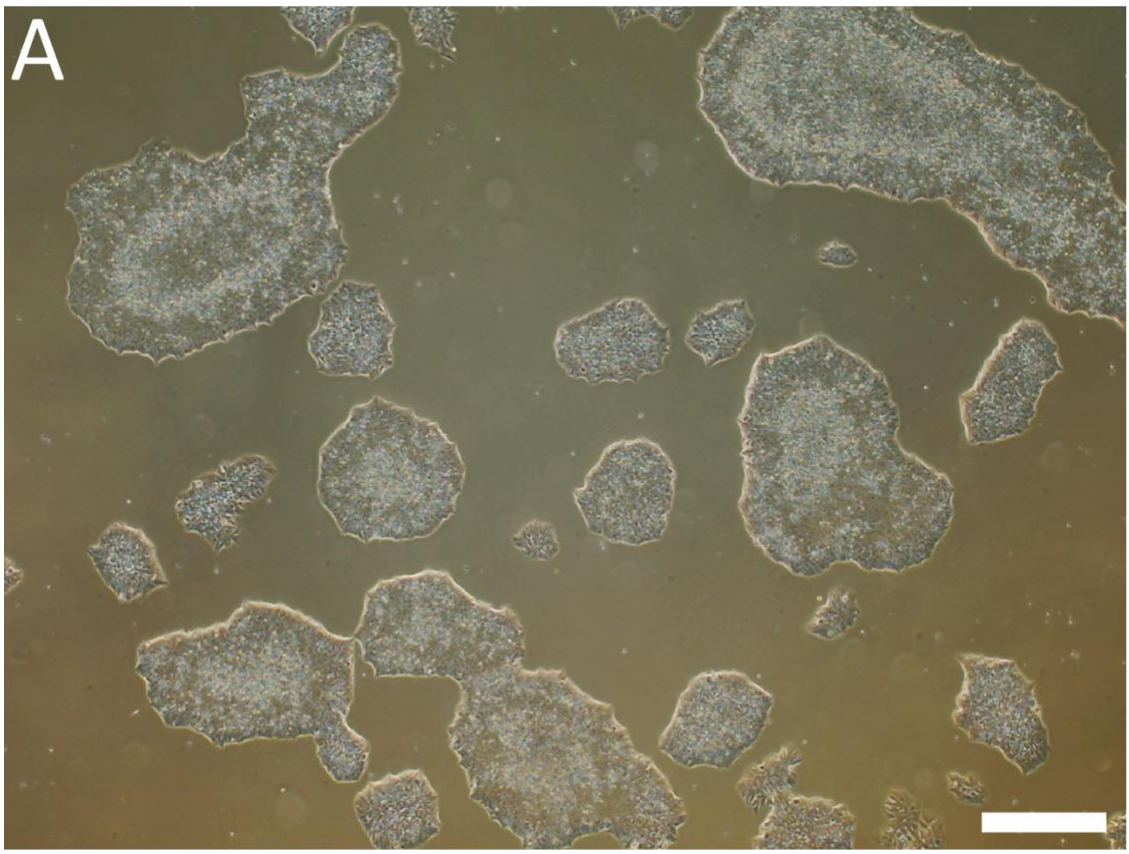
24312B



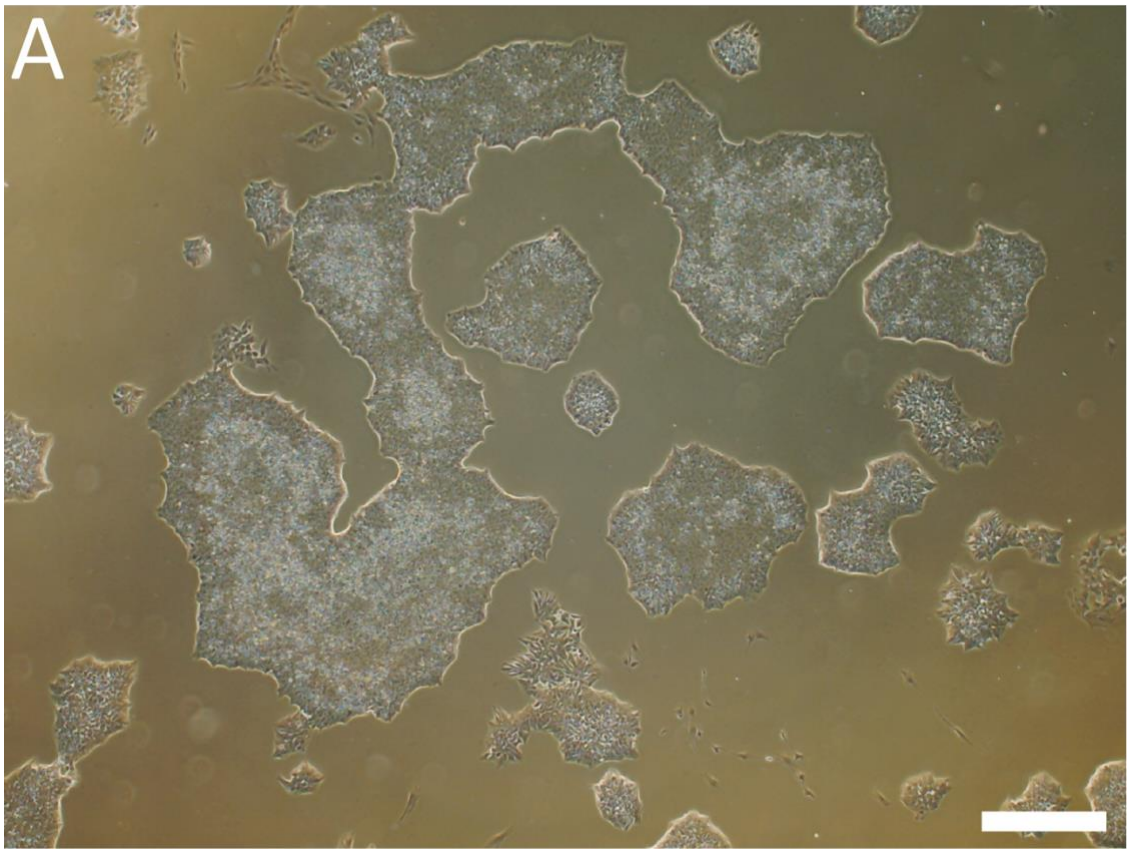
25501A



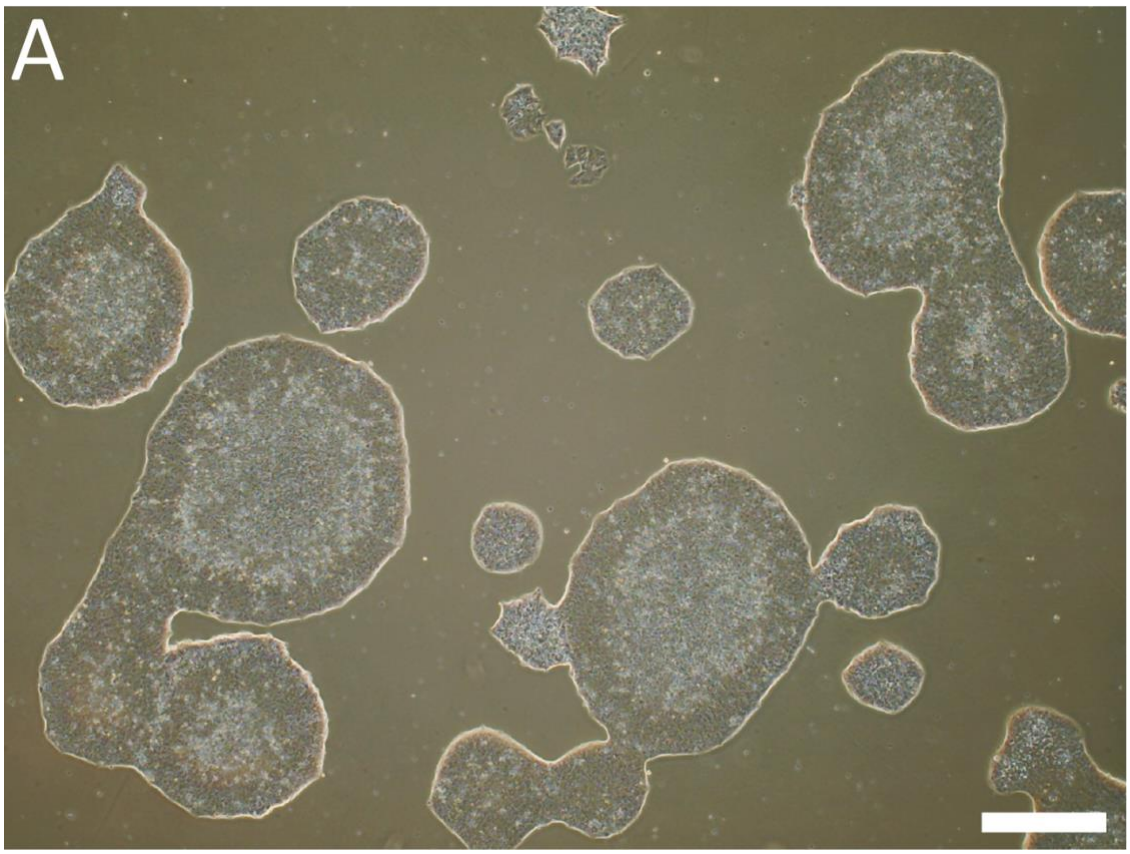
25501B



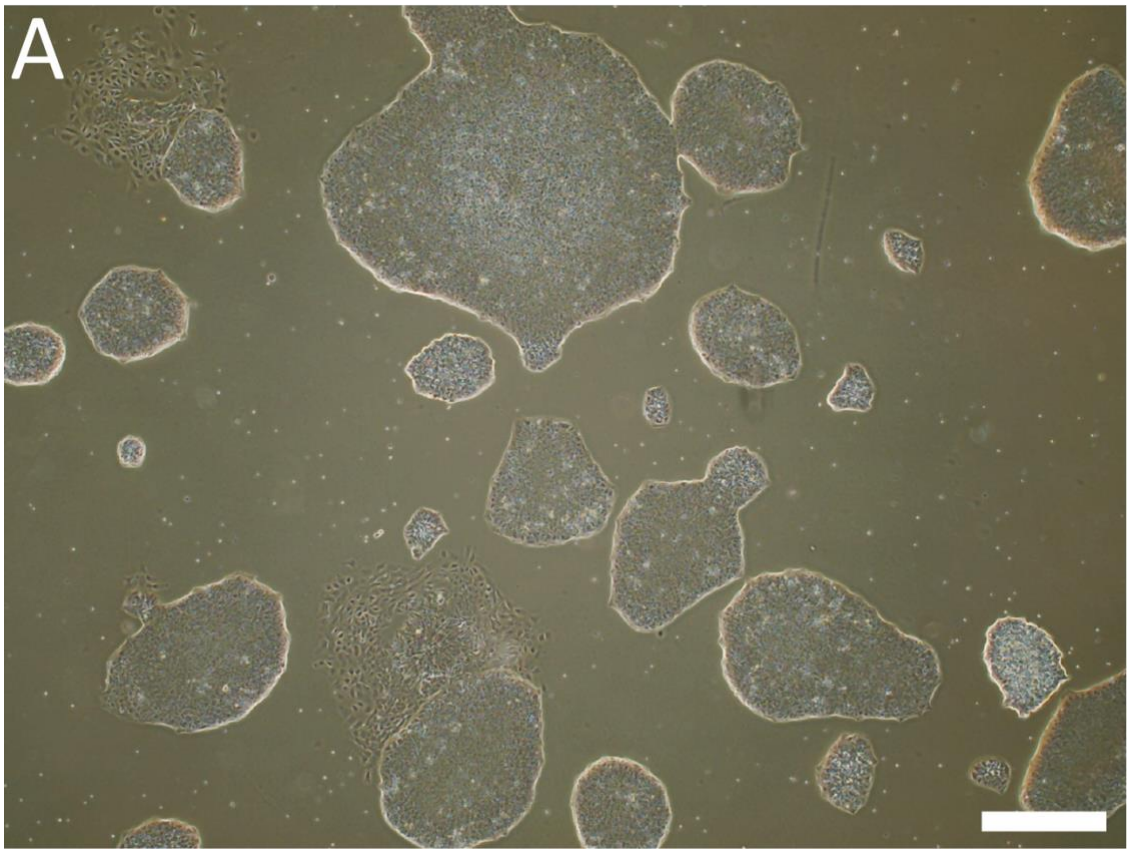
25501C



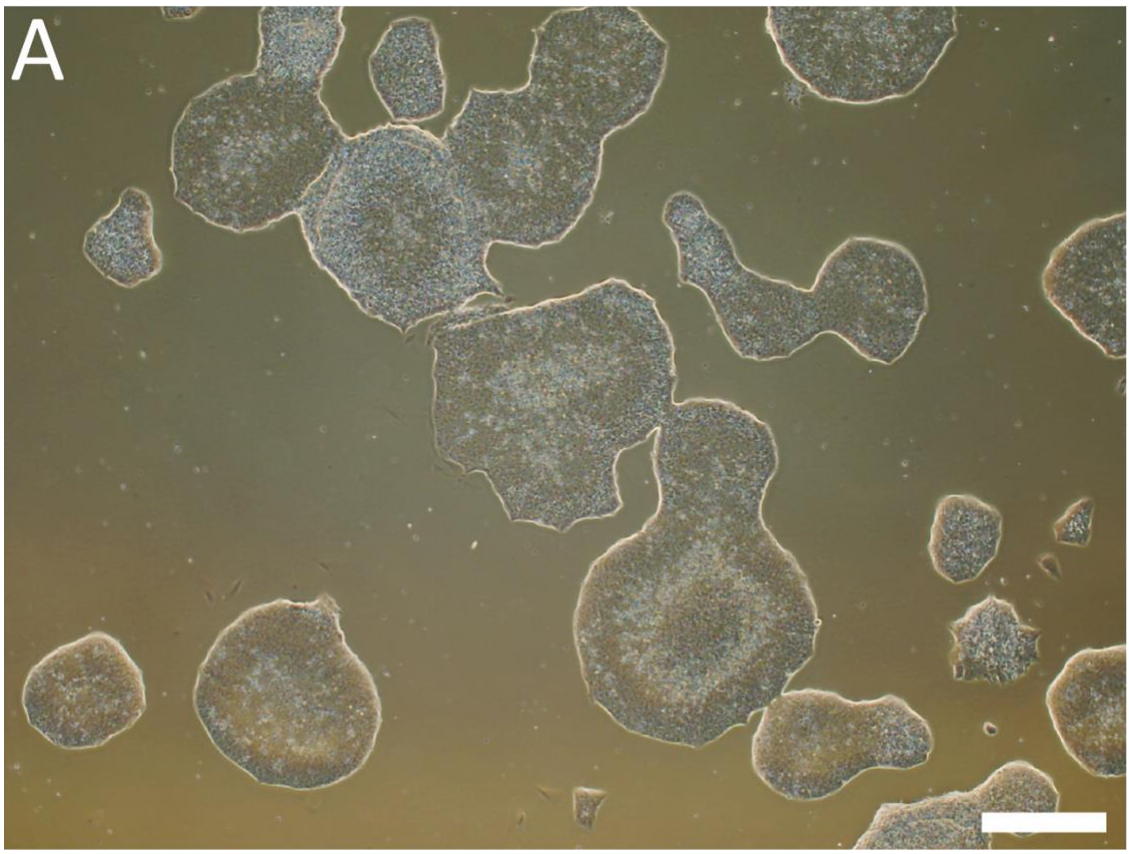
20152A



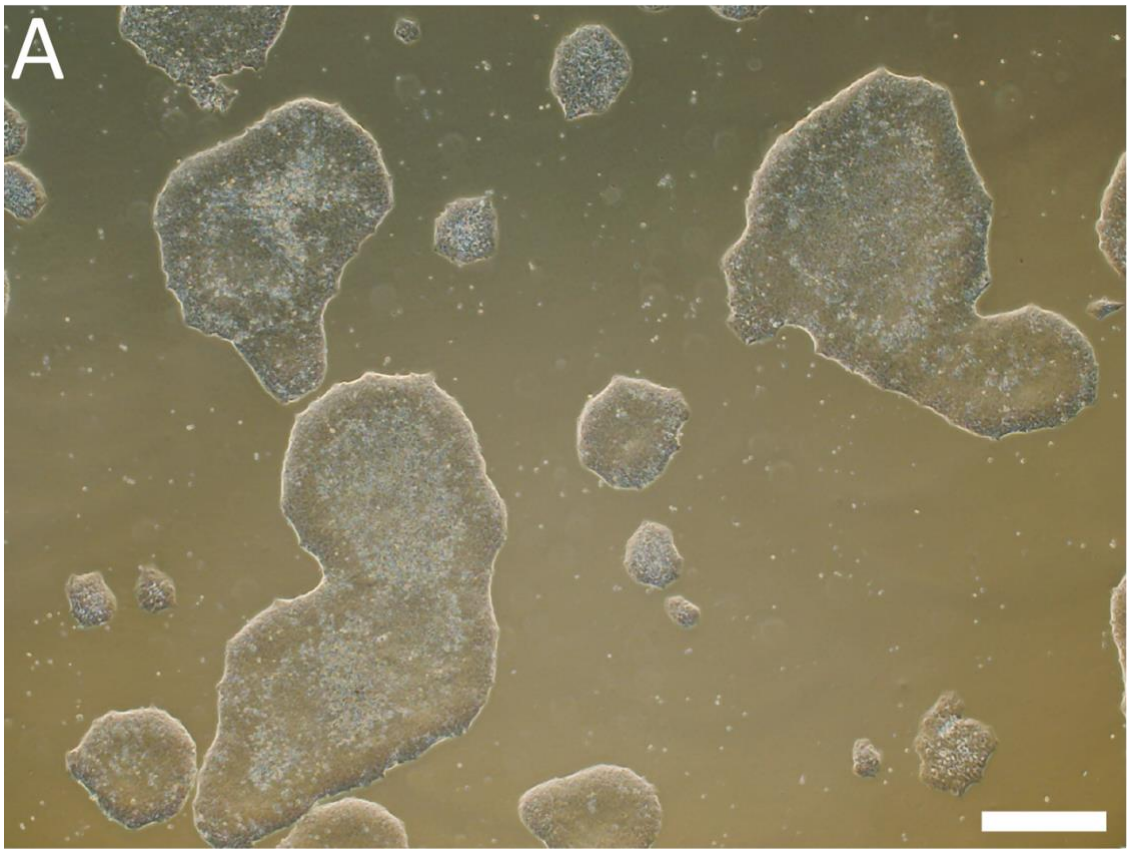
20152B



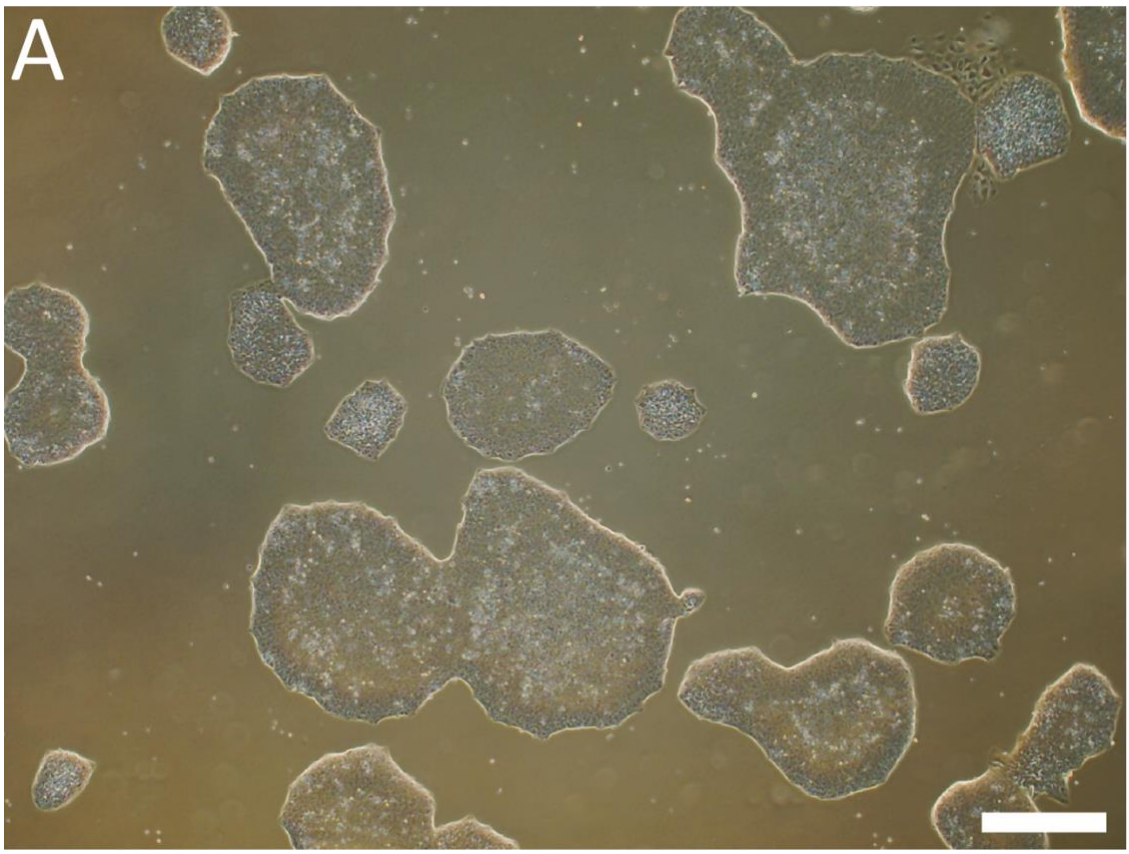
20152C



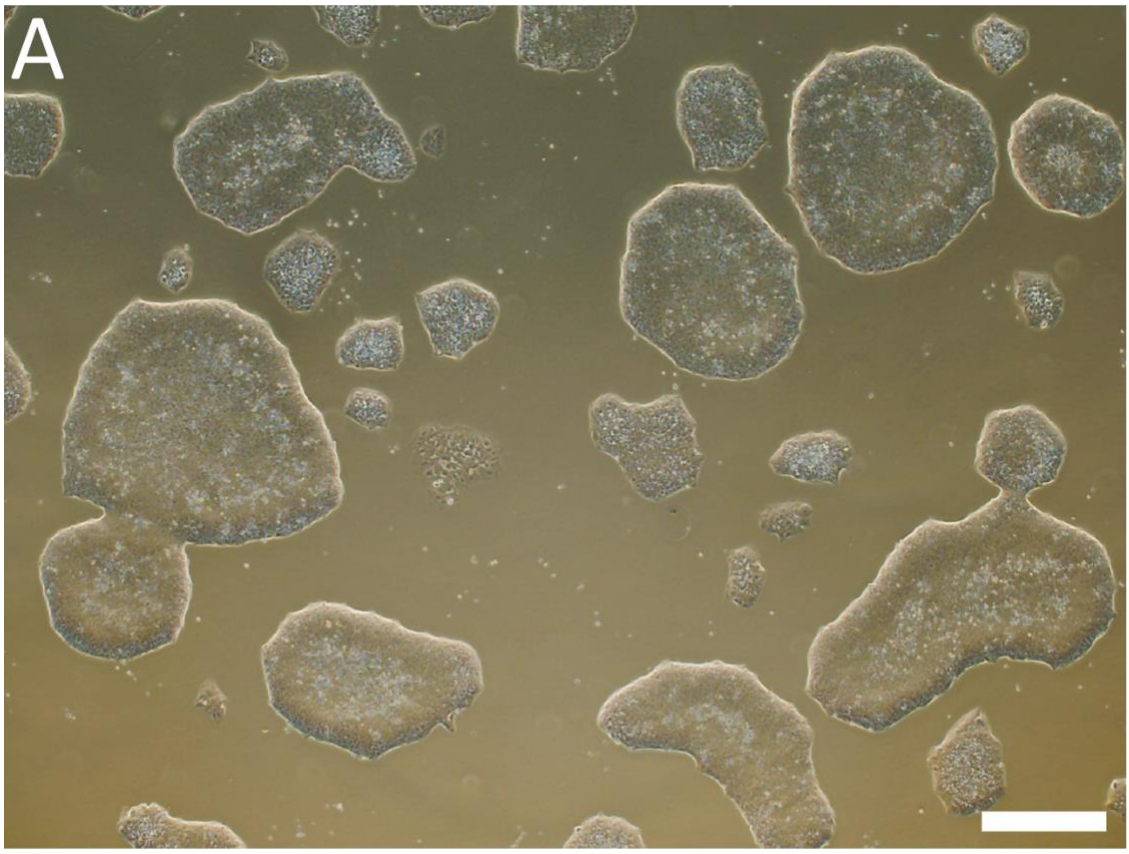
20446A



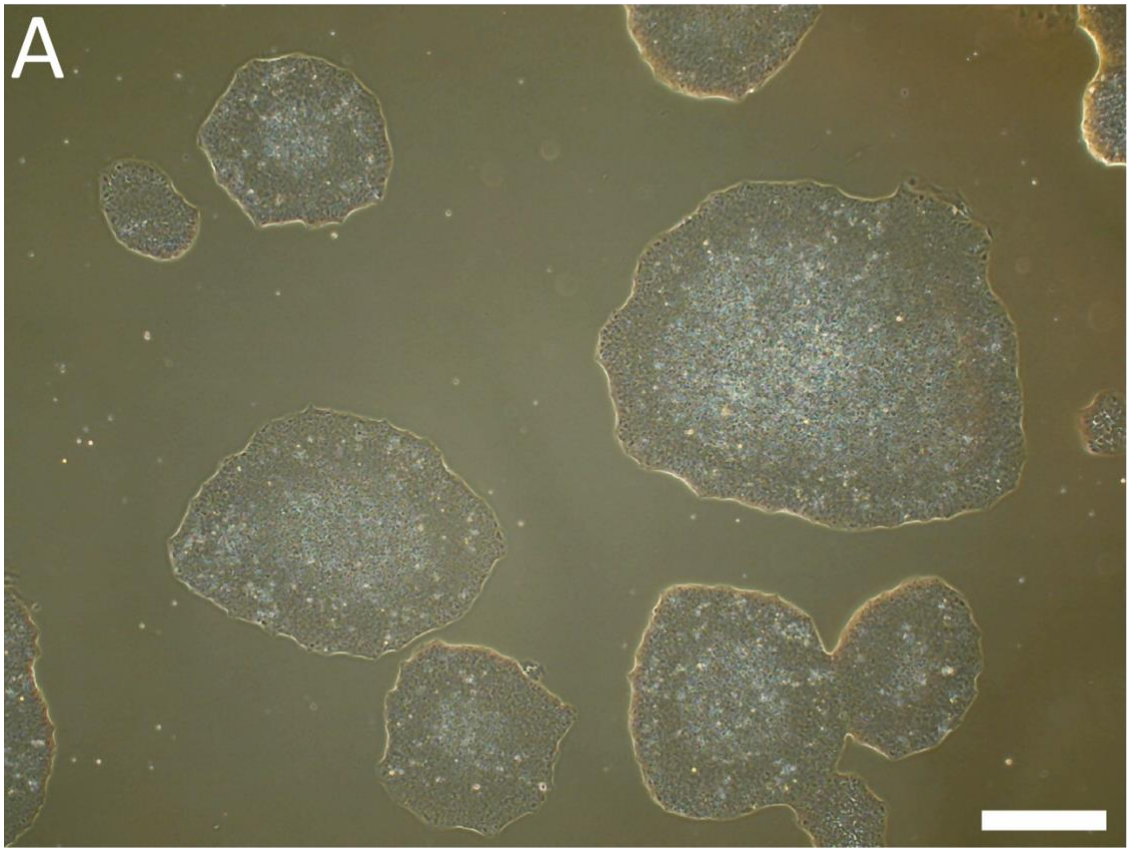
20446B



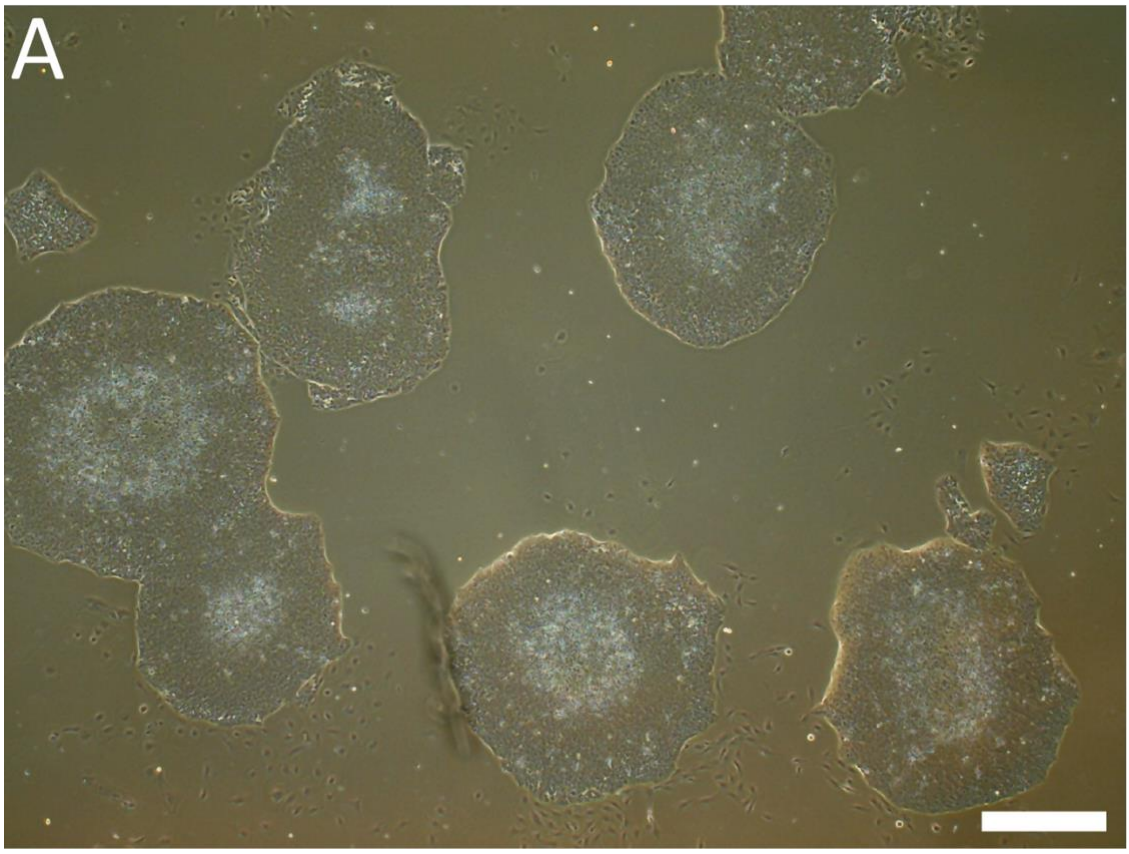
20446C



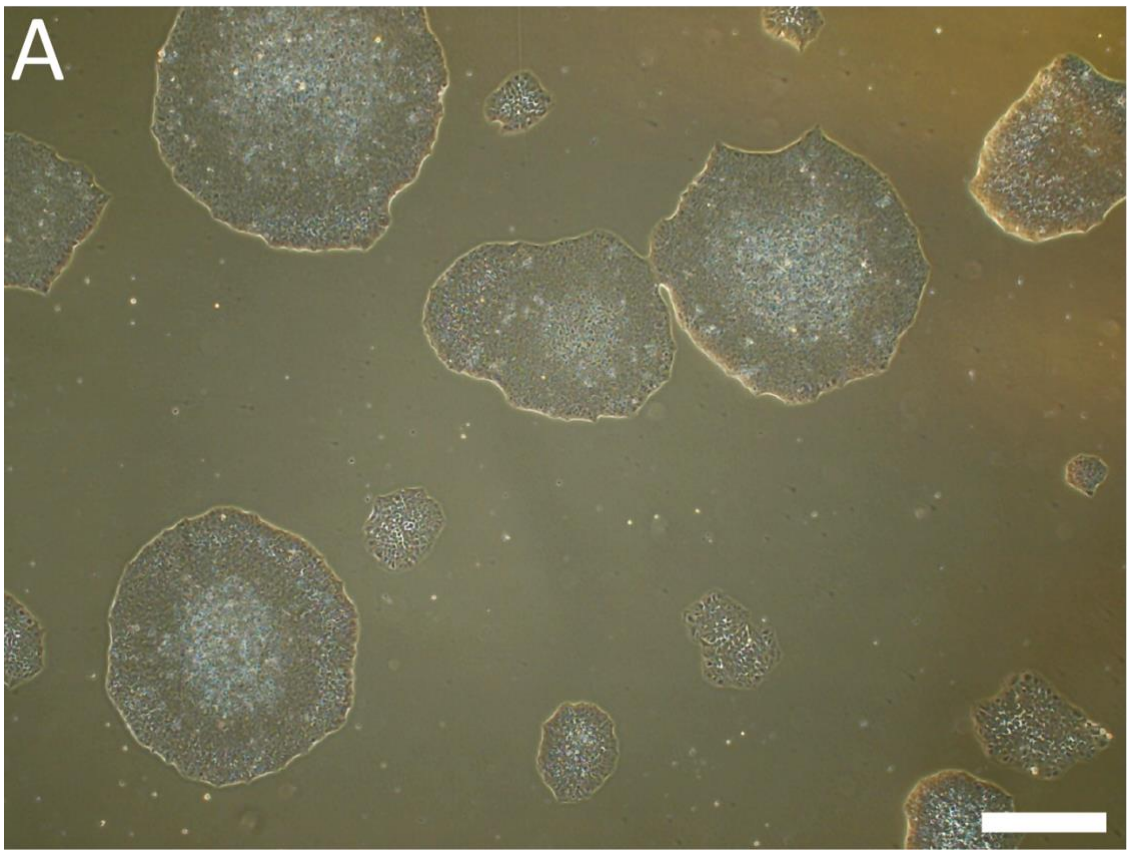
20722A



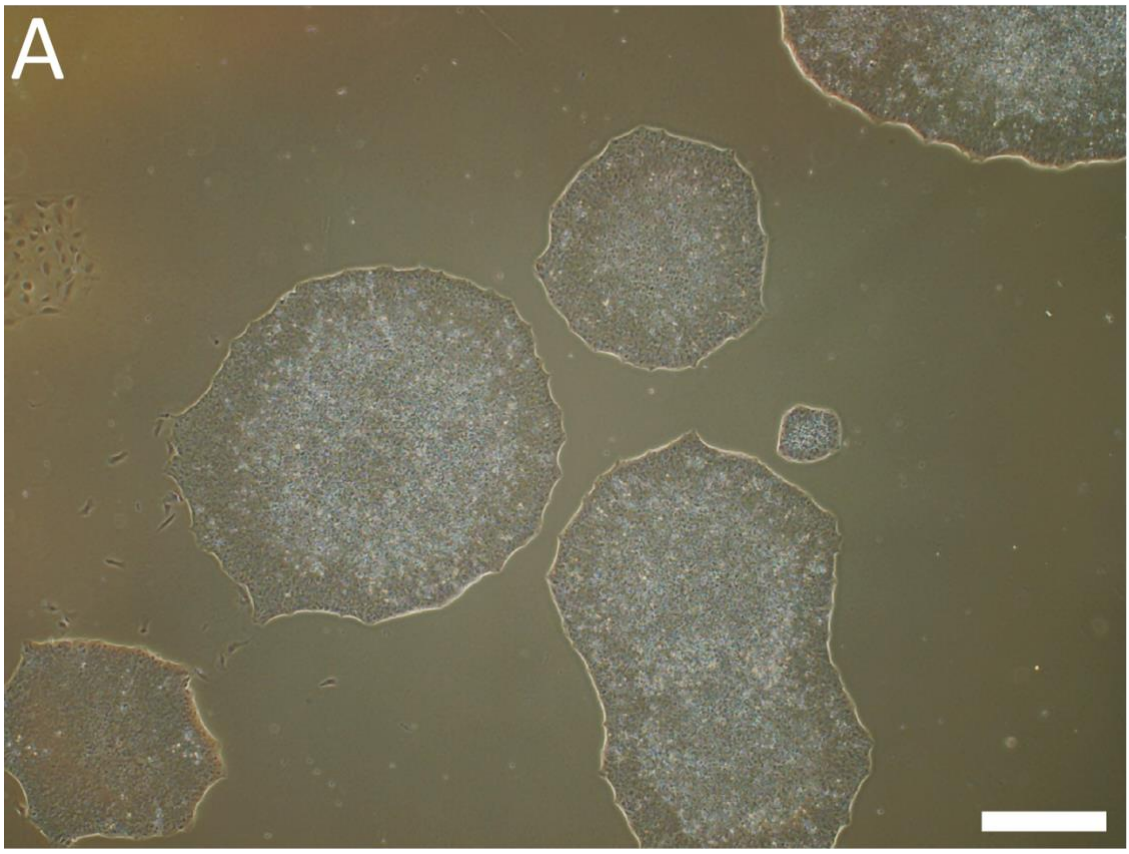
20722B



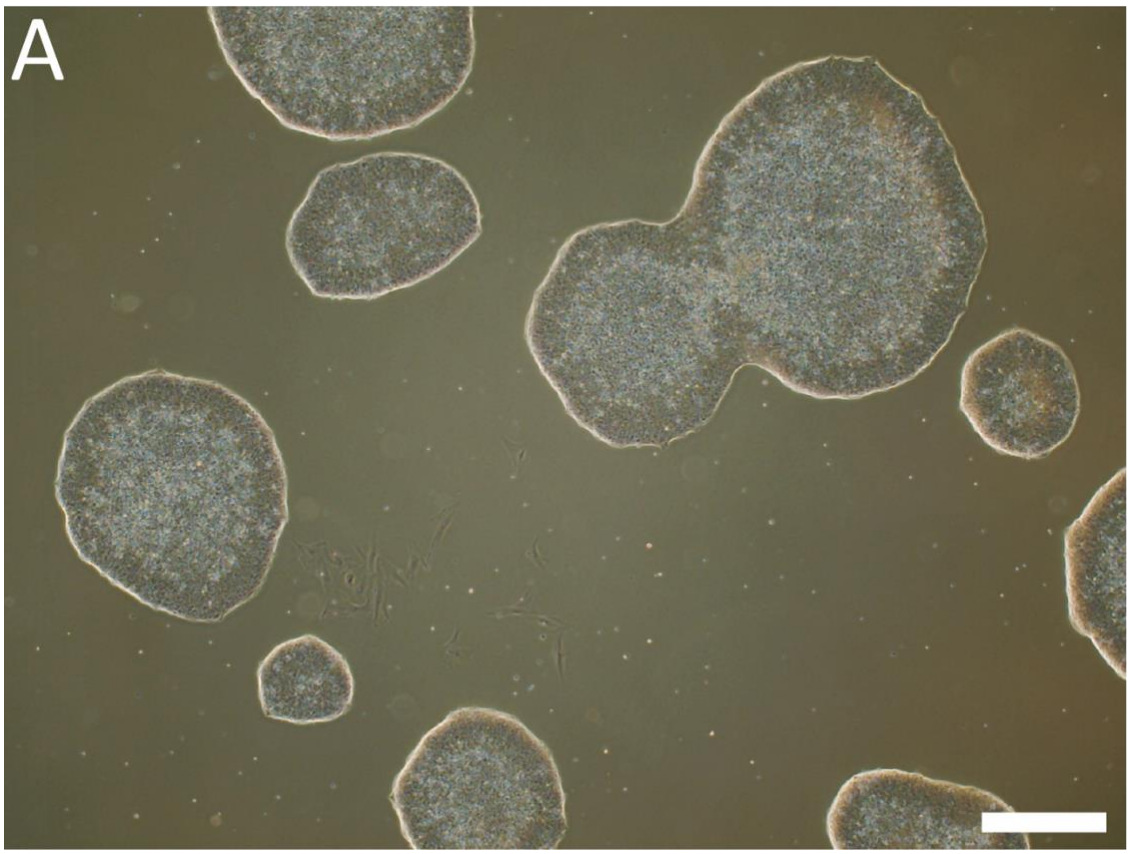
20722C



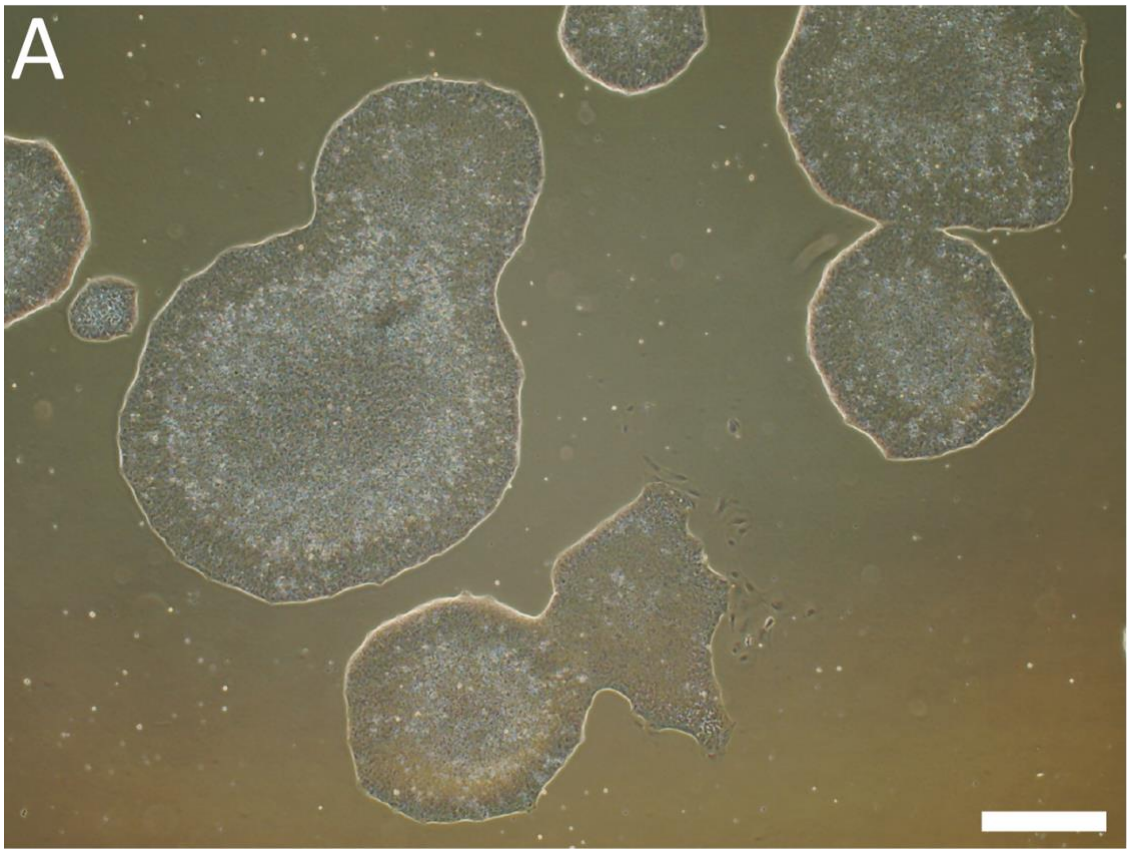
20176A



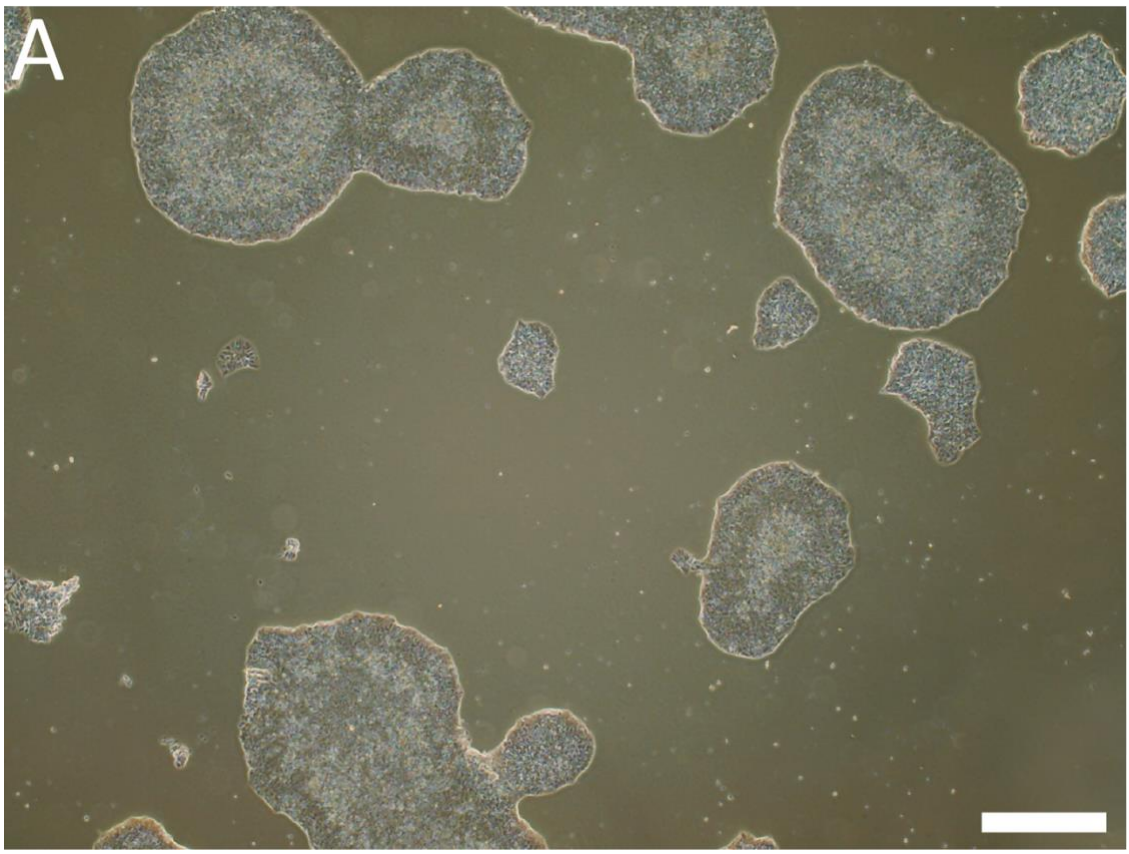
20176B



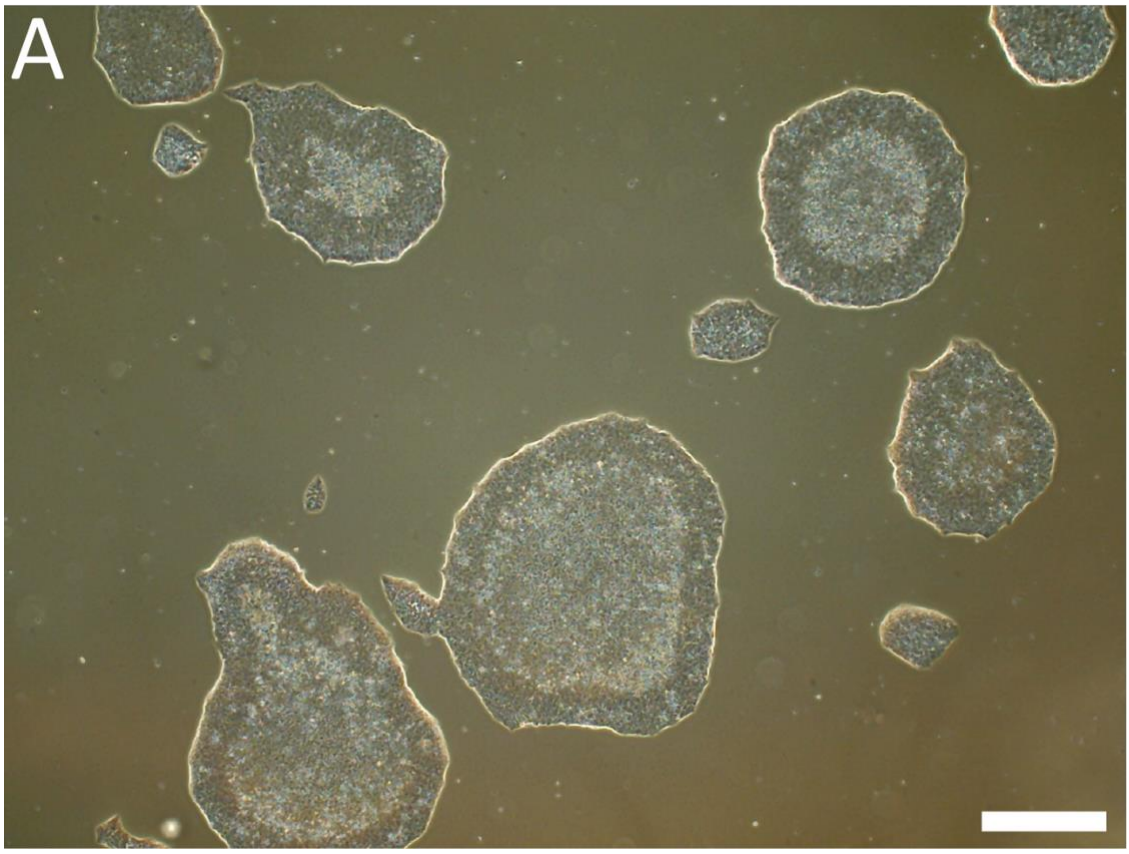
20176C



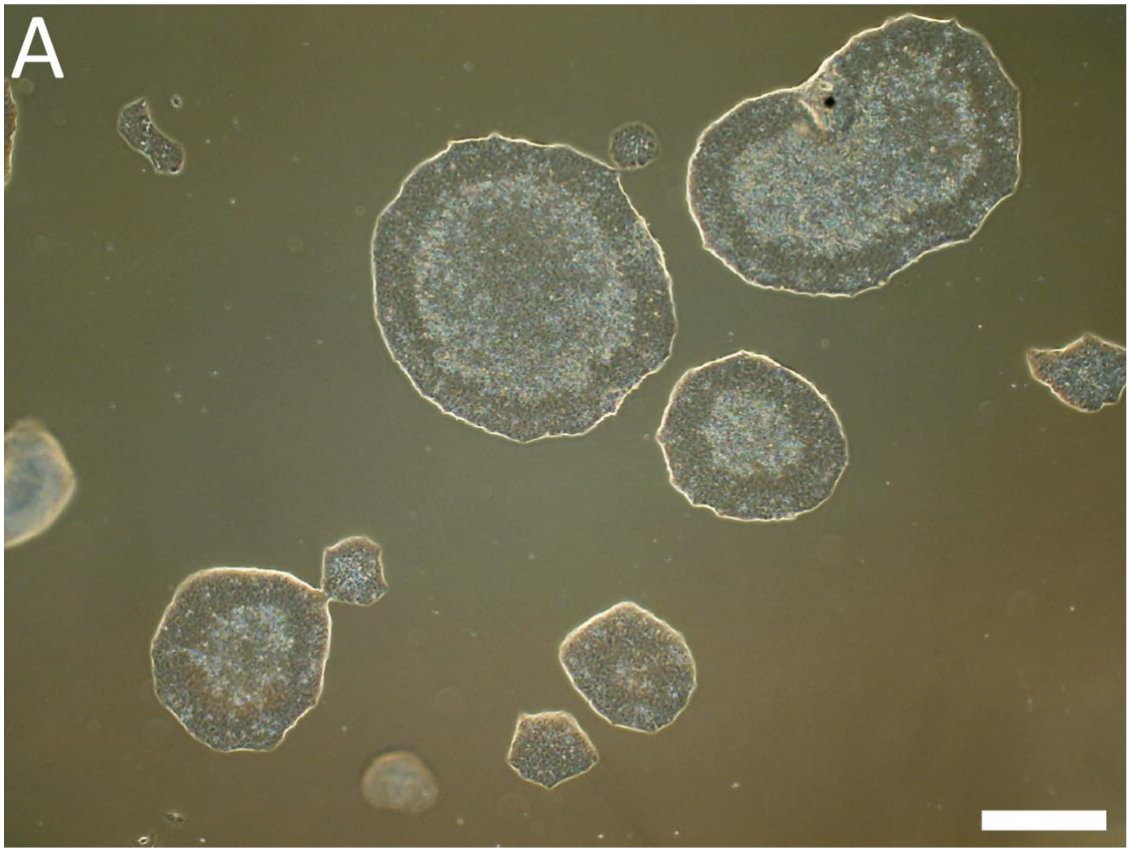
20316A



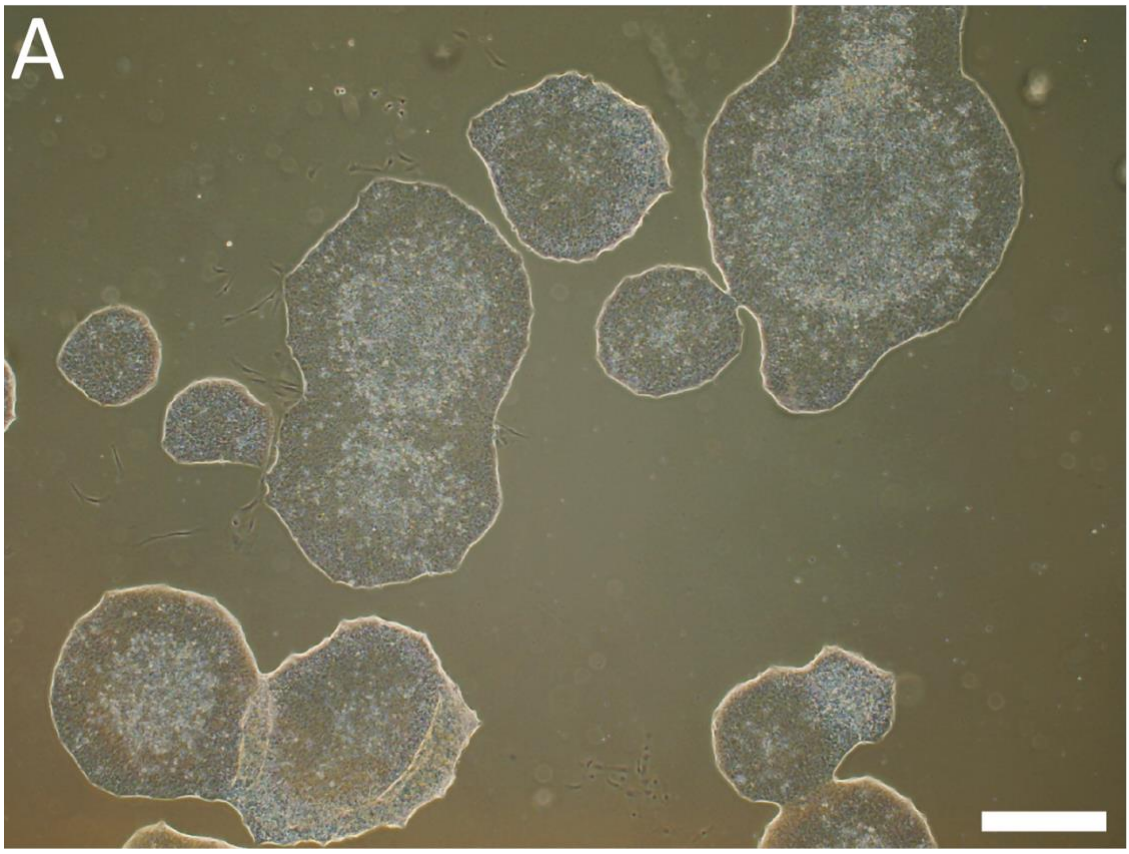
20316B



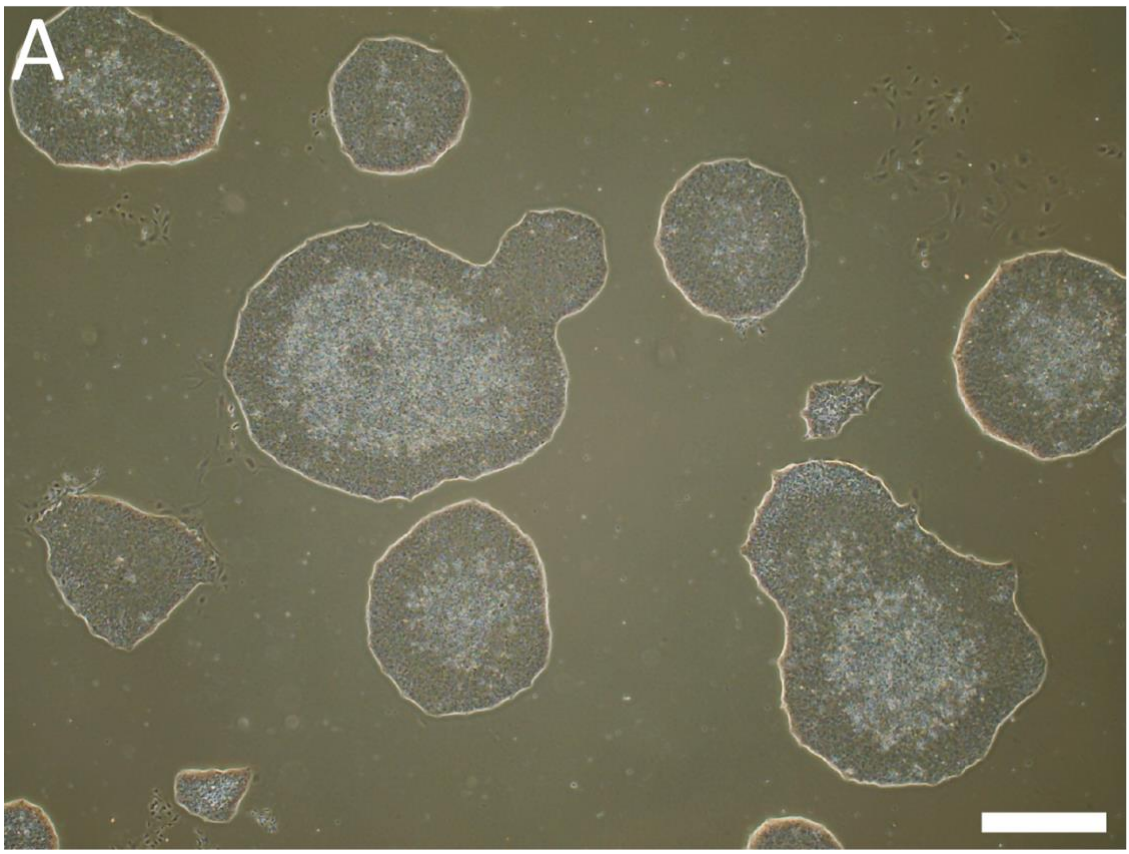
20316C



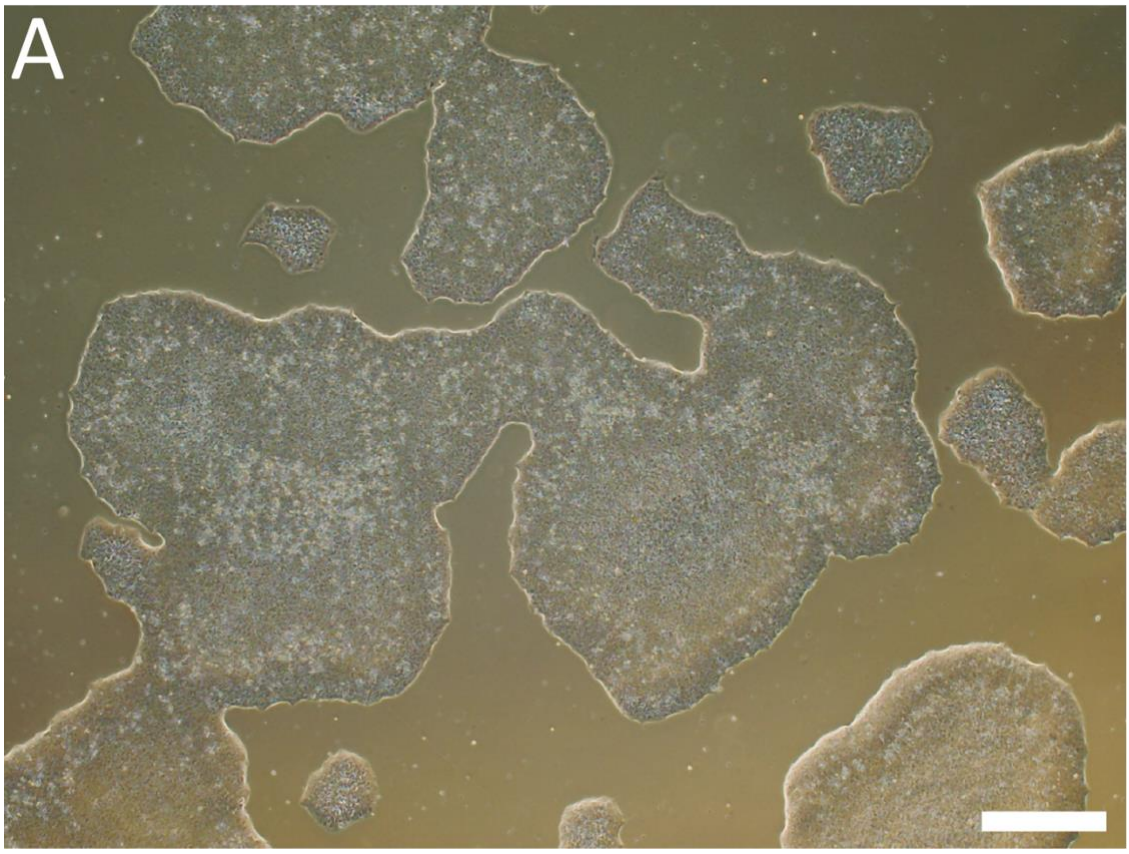
20451A



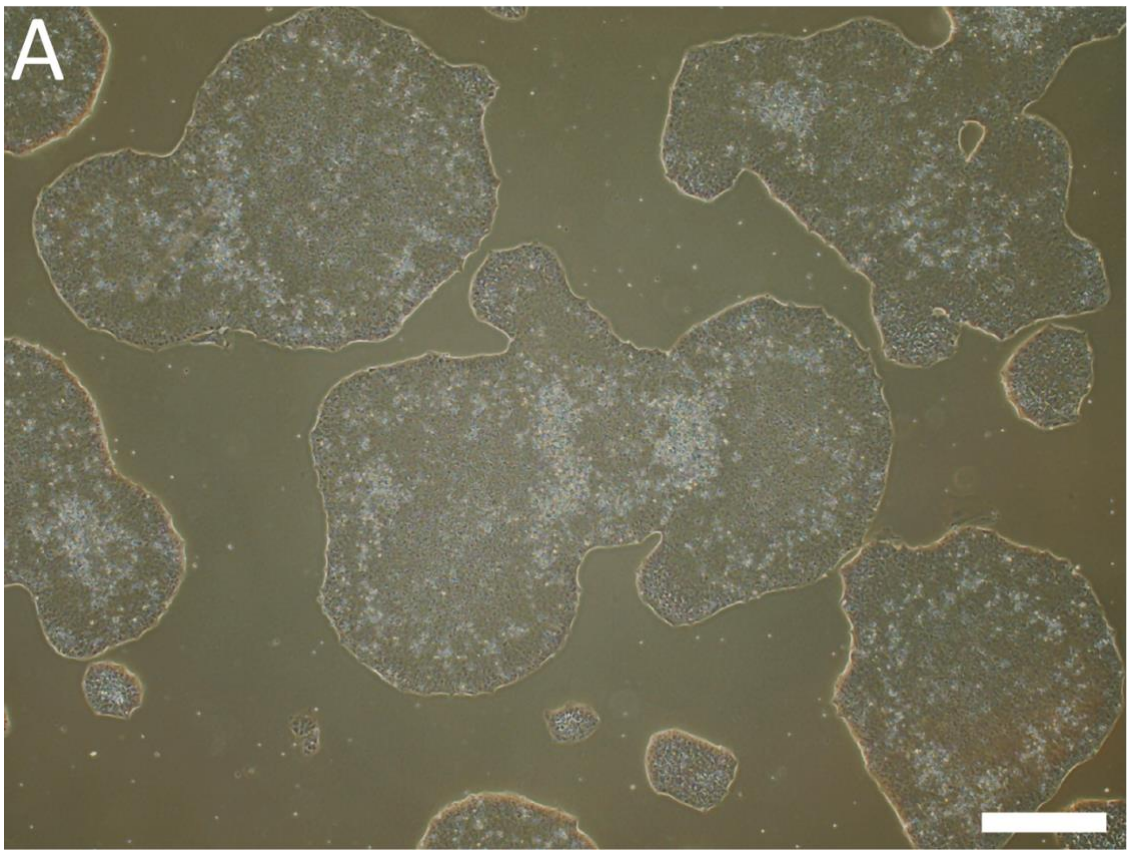
20451B



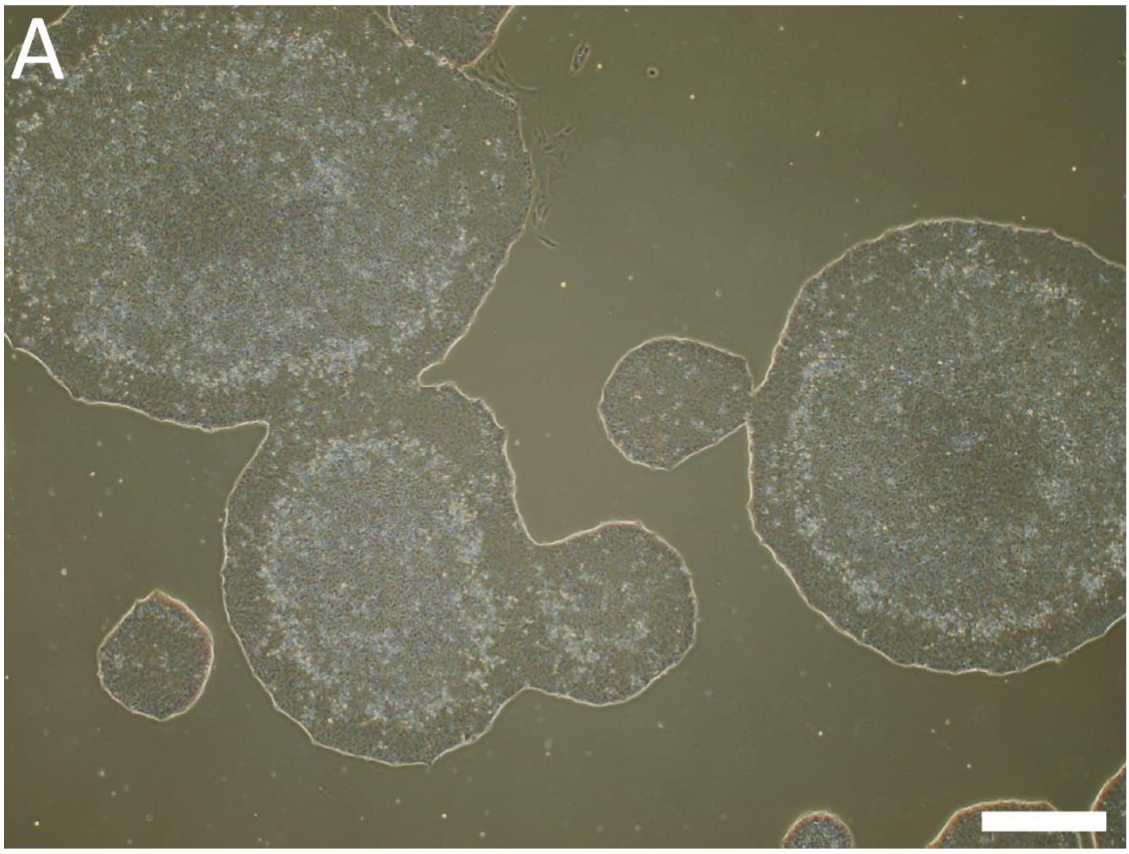
20451C



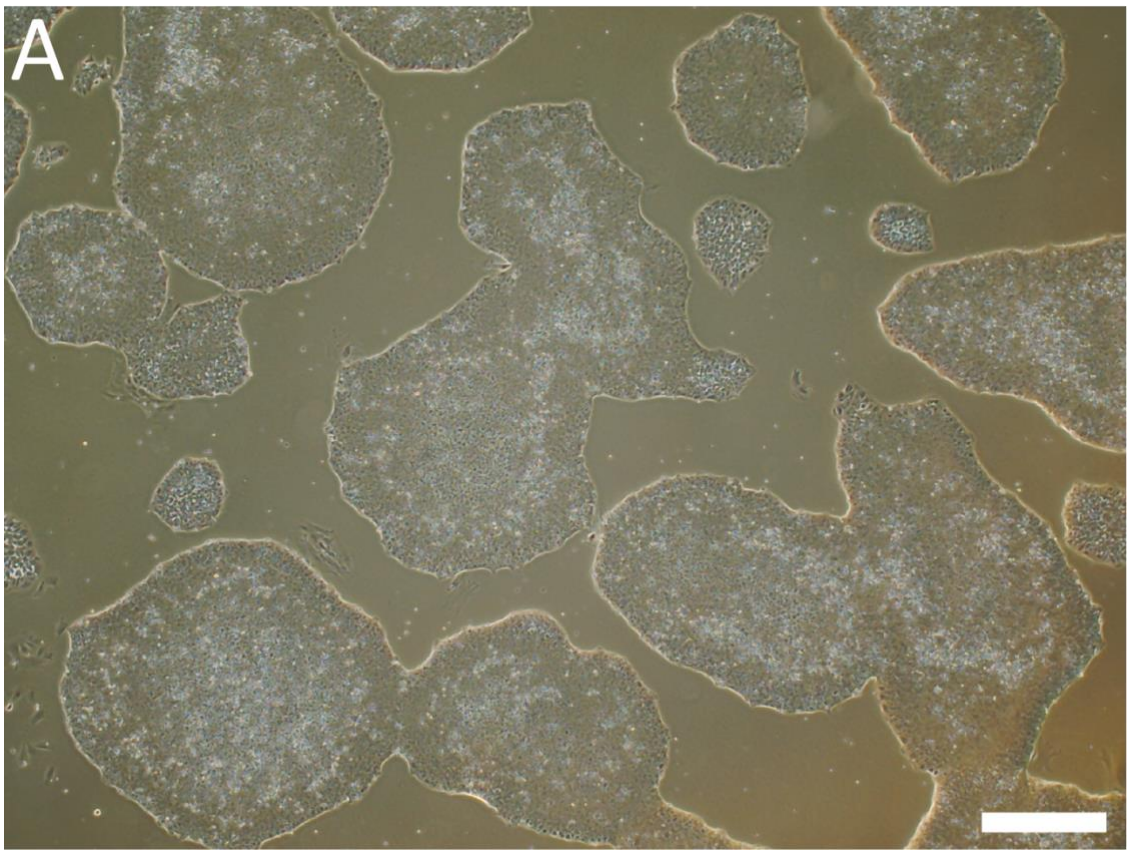
20777A



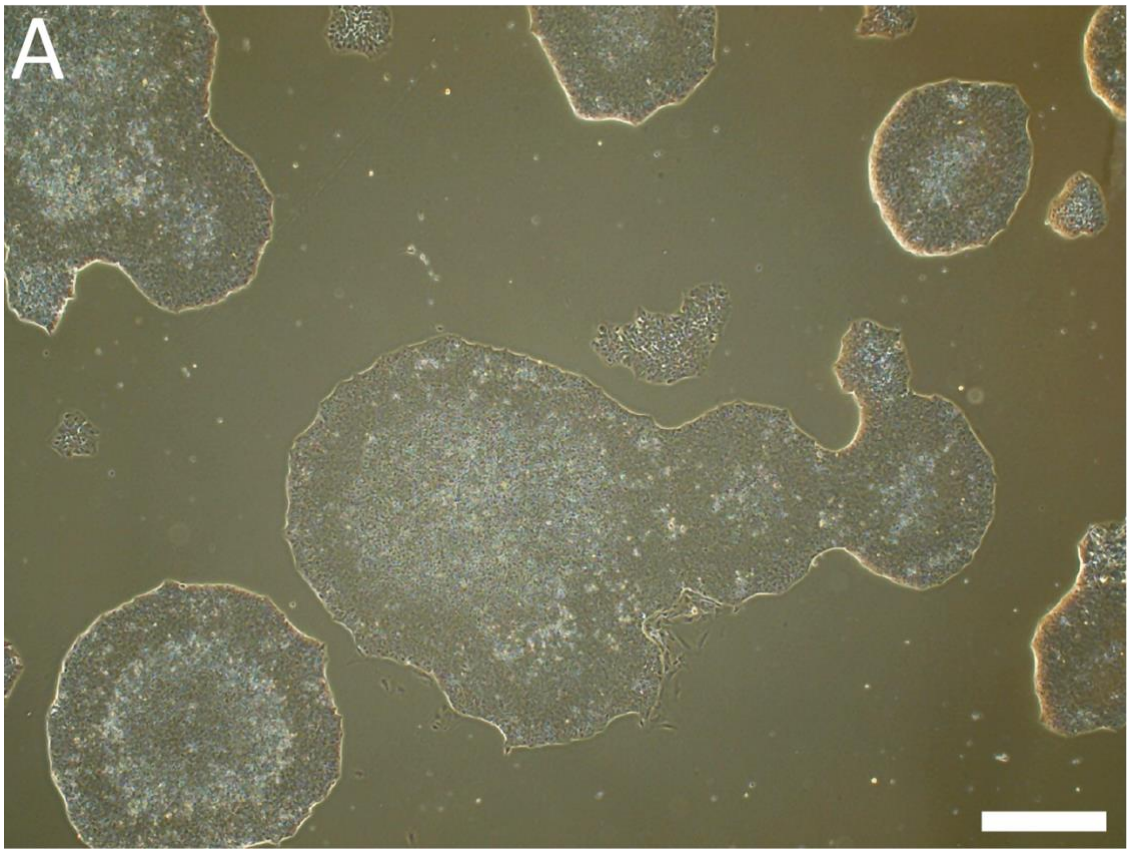
20777B



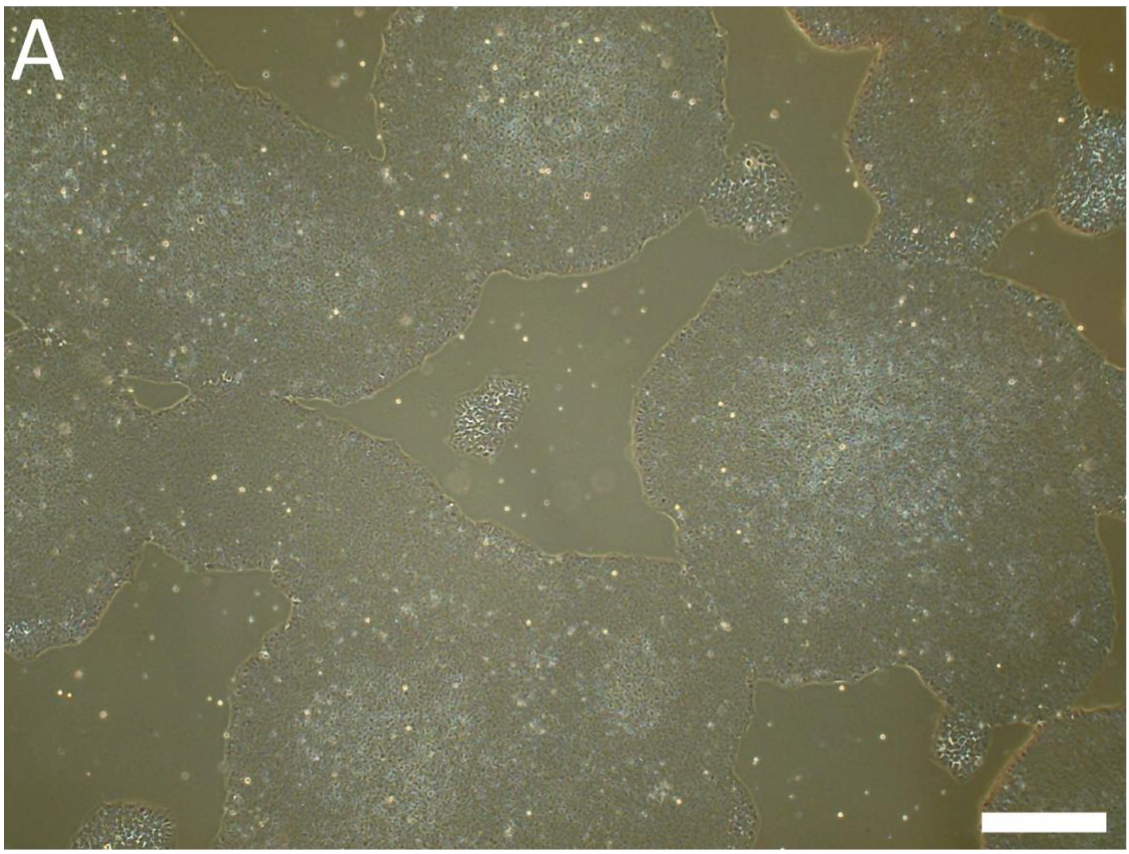
20777C



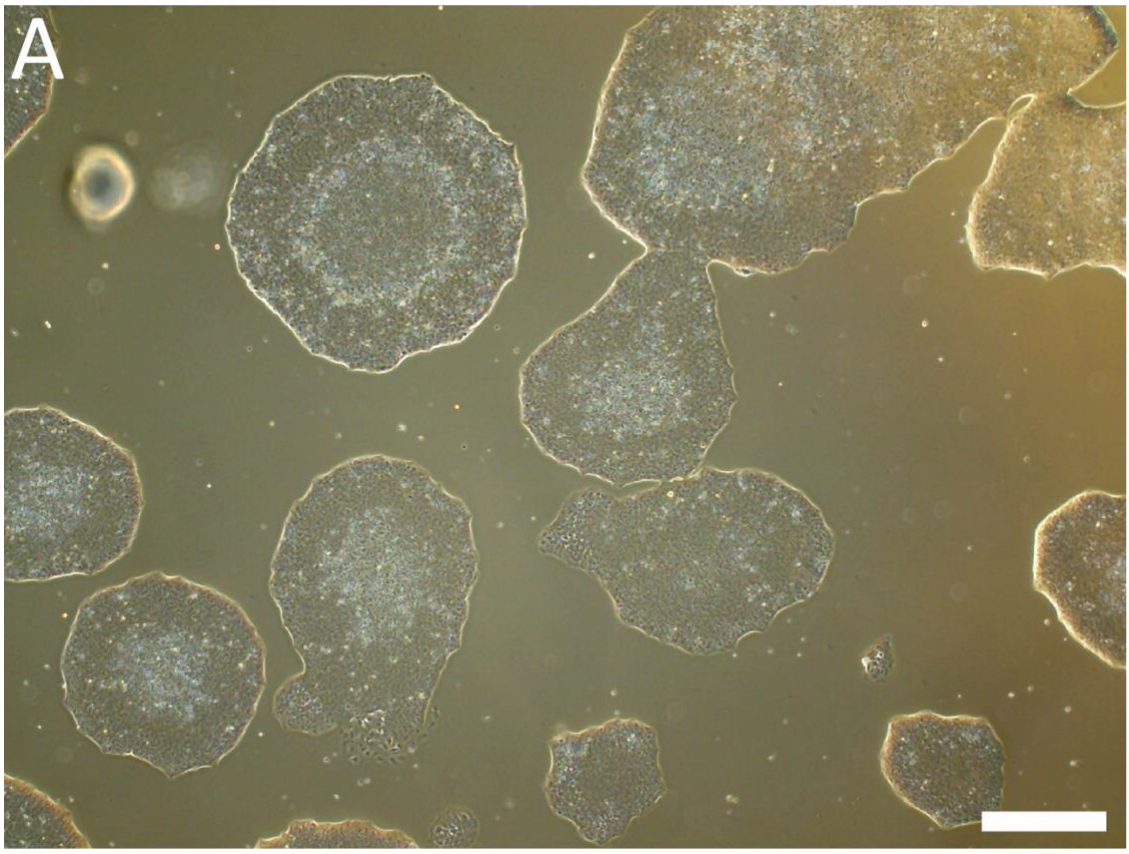
20236A



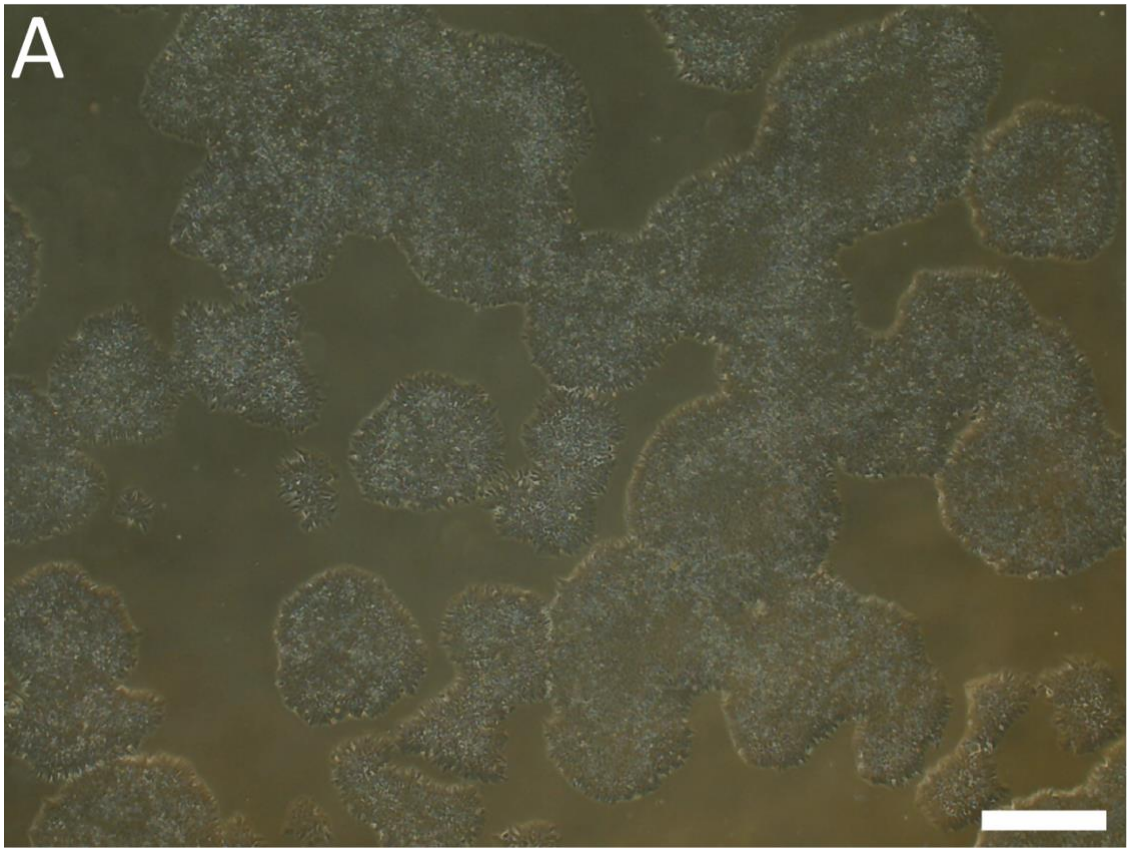
20236B



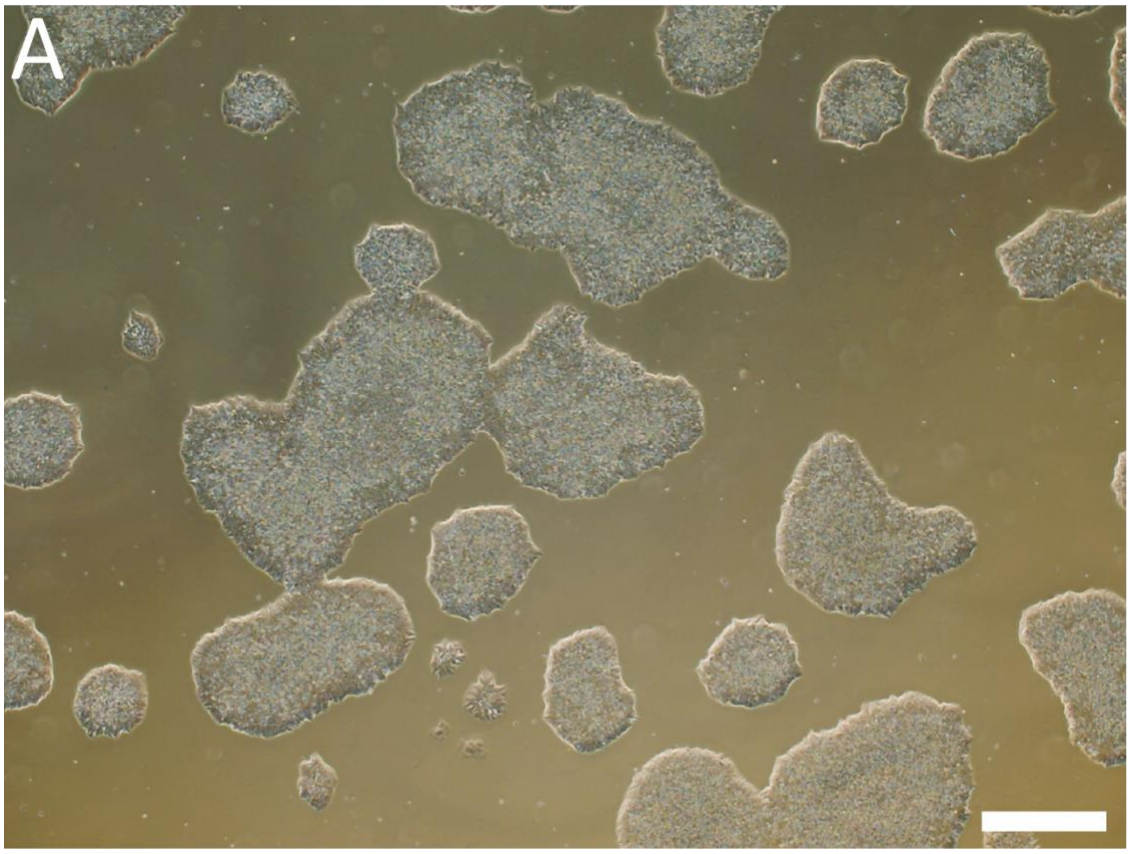
20236C



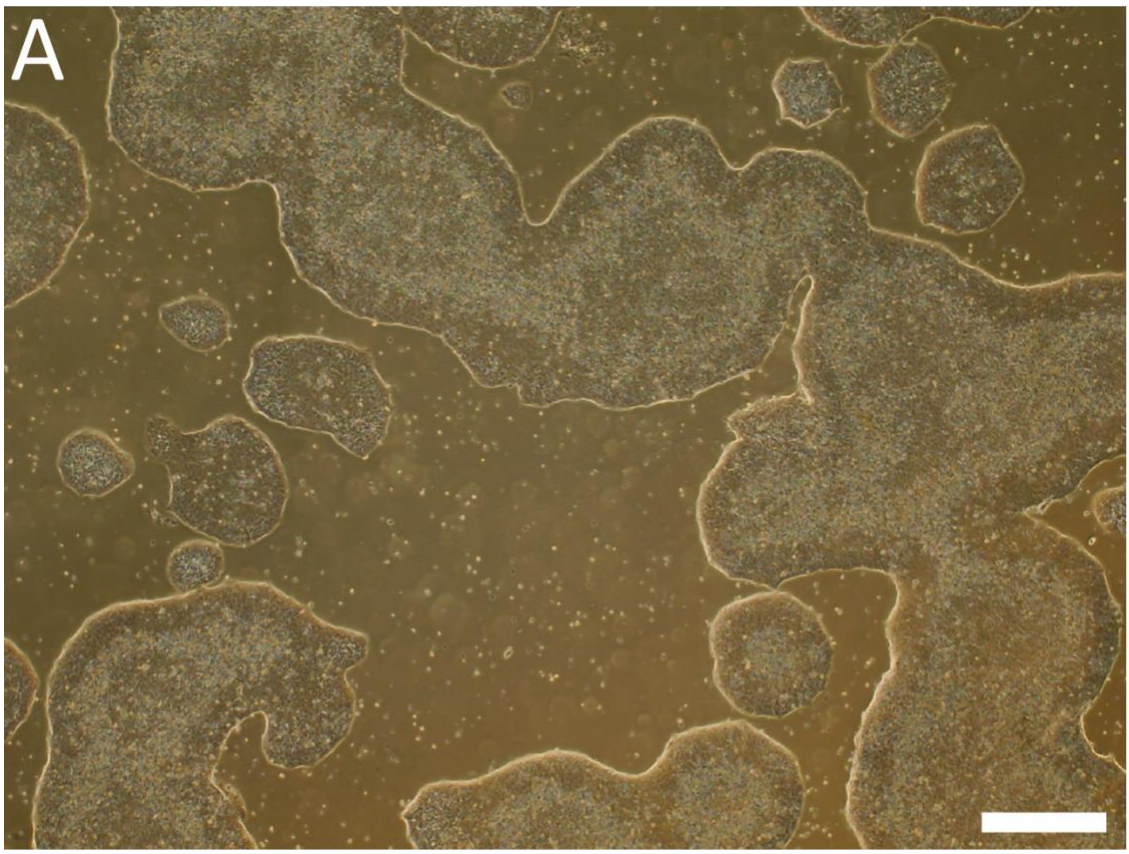
21052A



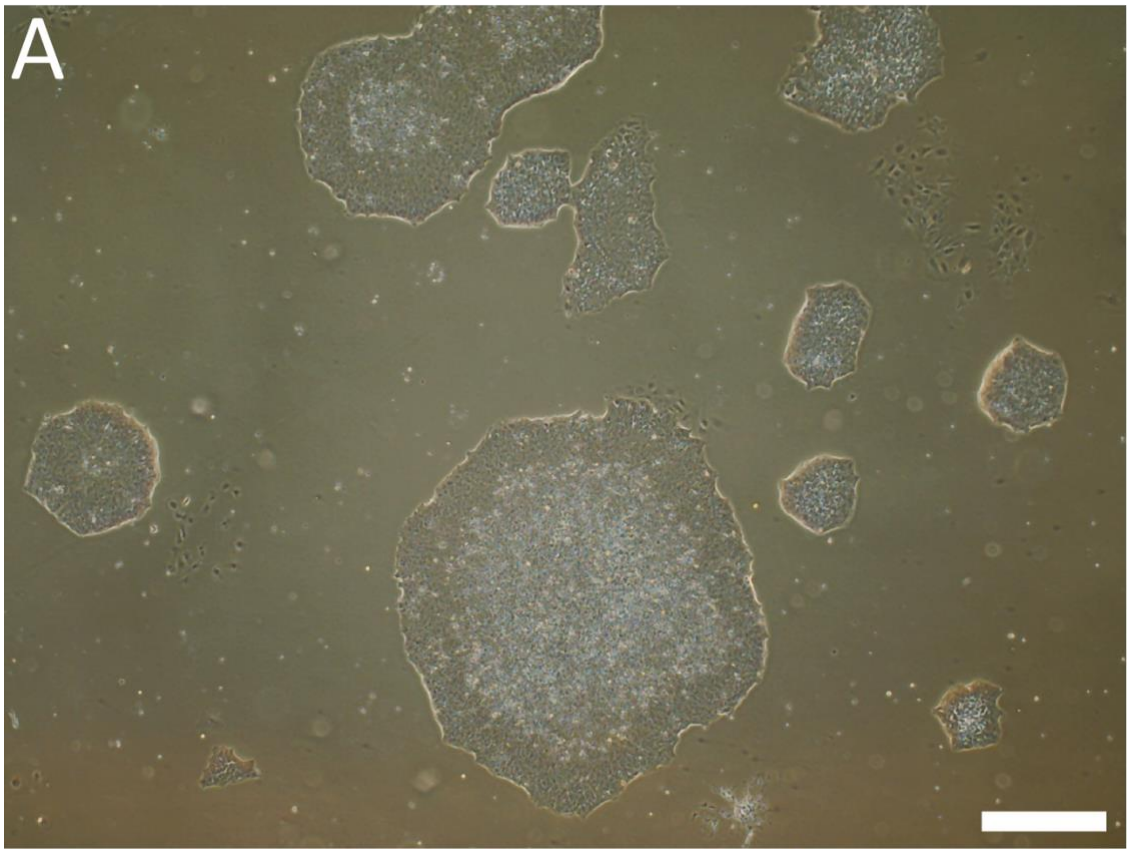
21052B



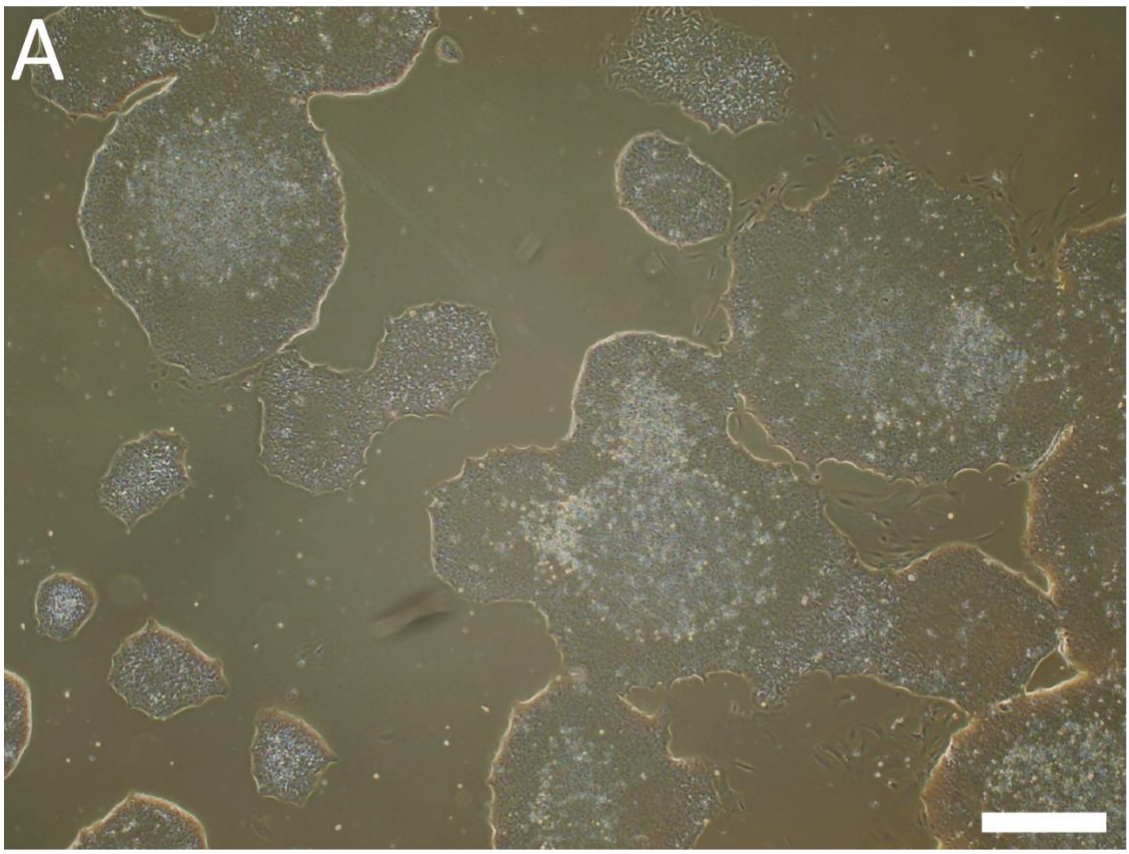
21052C



41624A

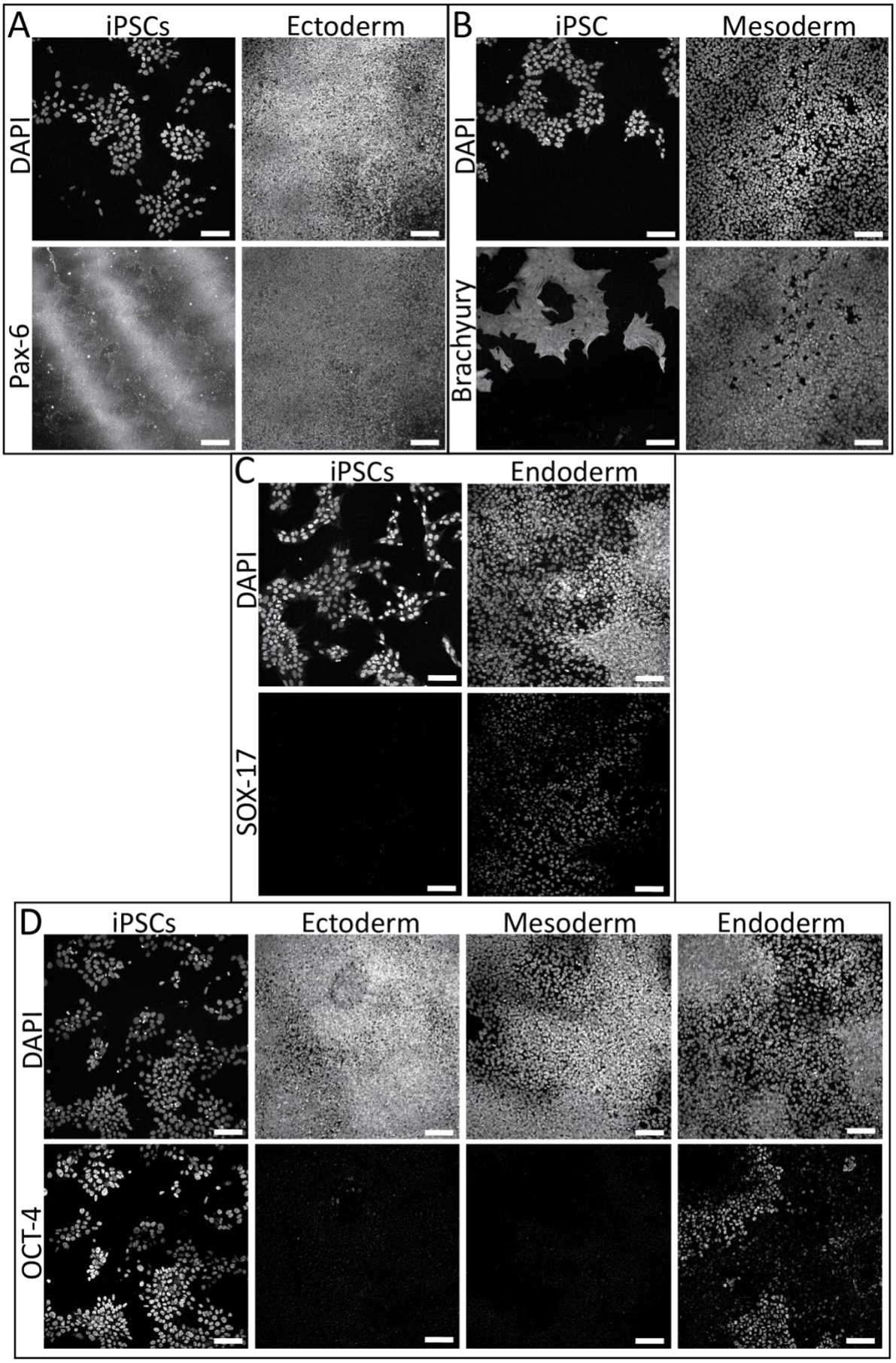


41624B

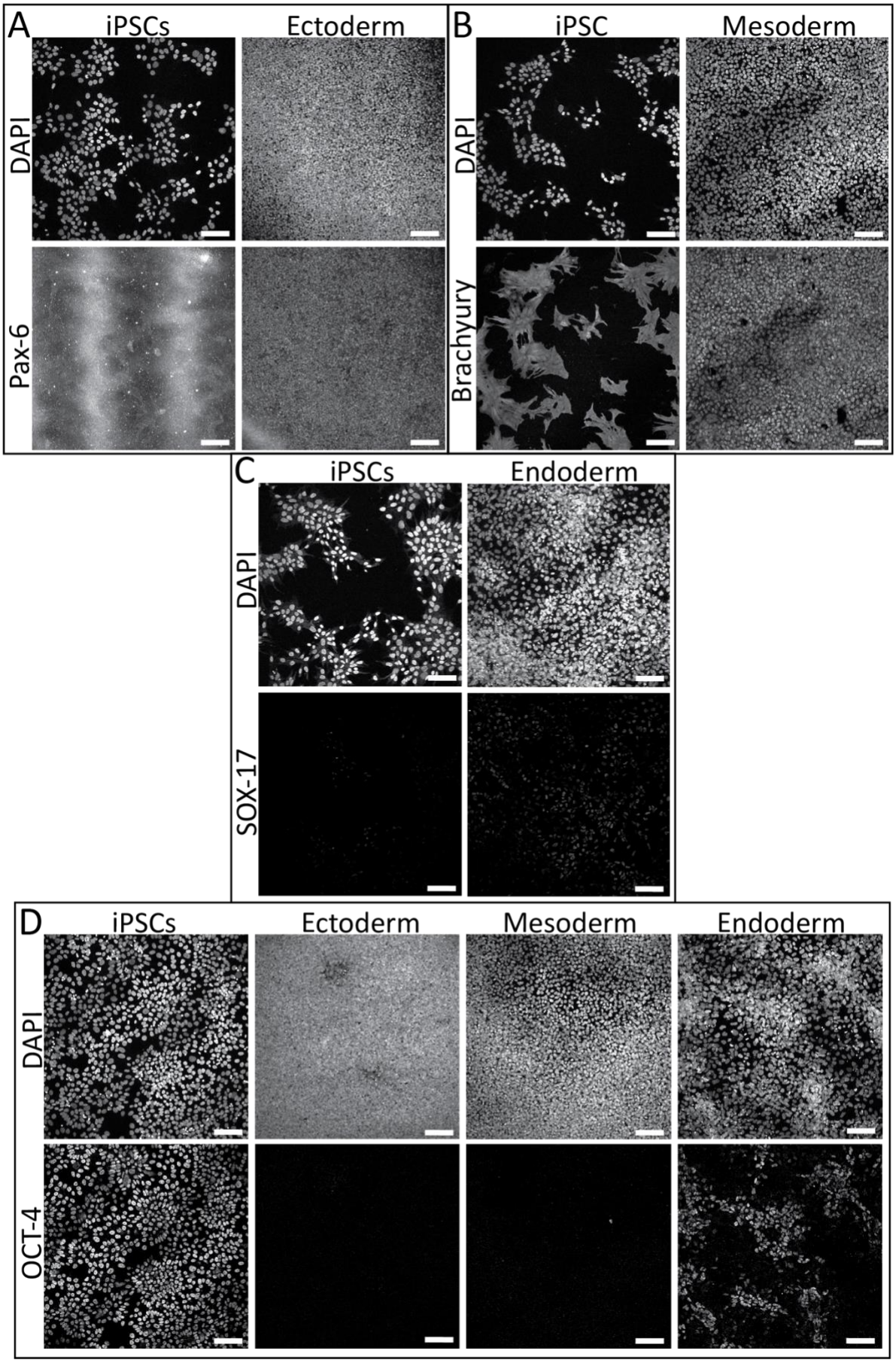


Supplementary file 4.2

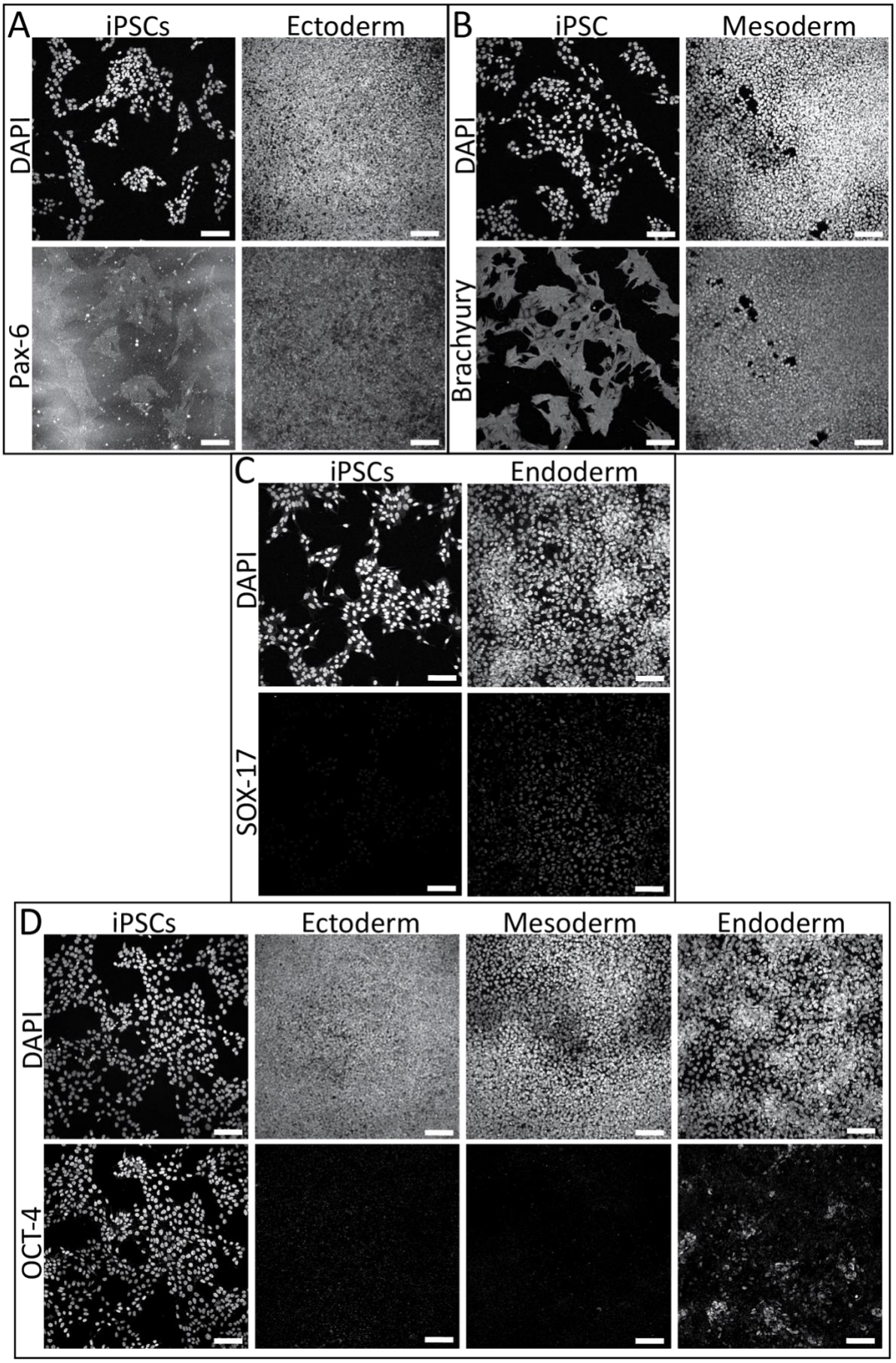
20032A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

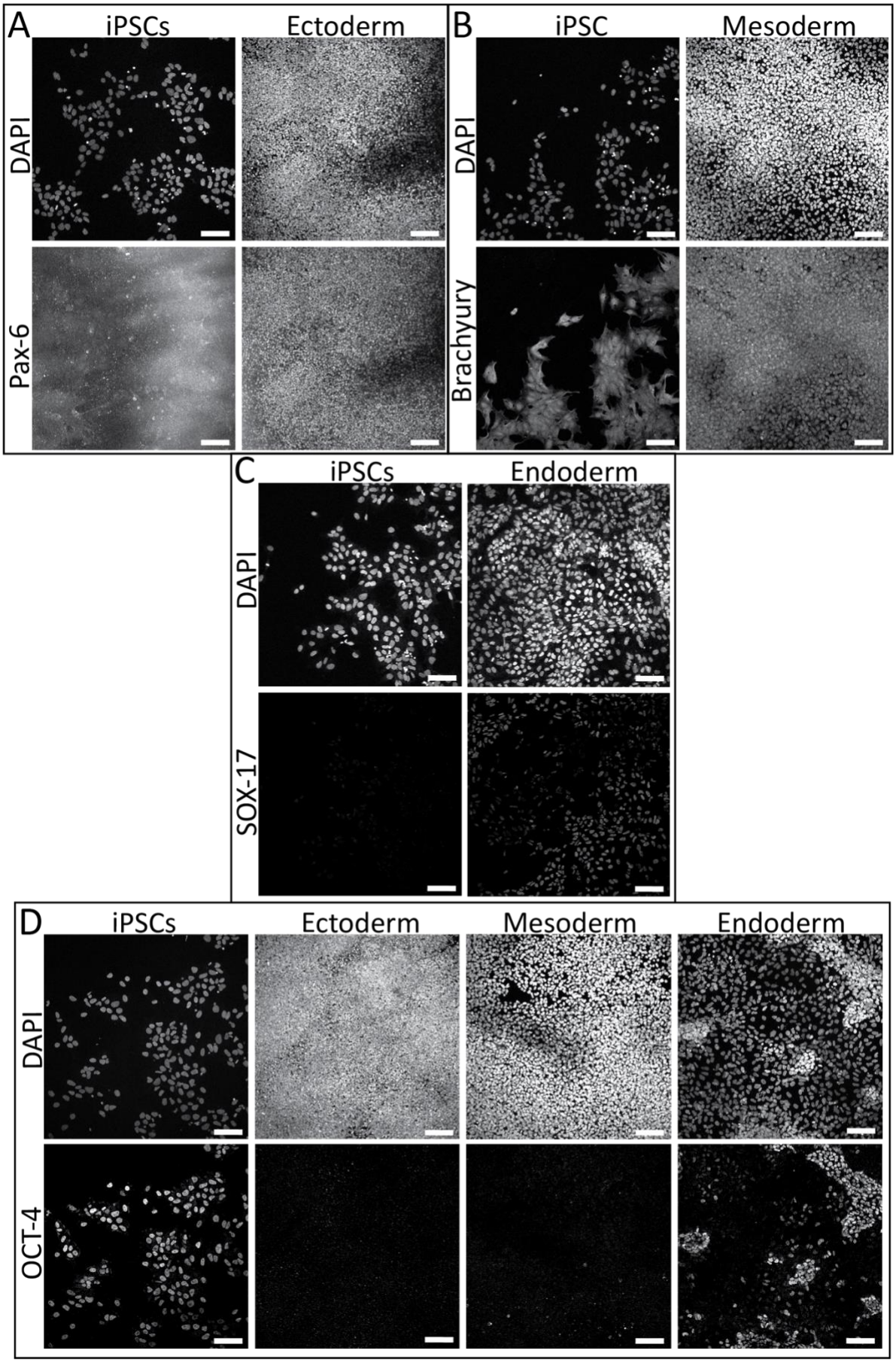
20032B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

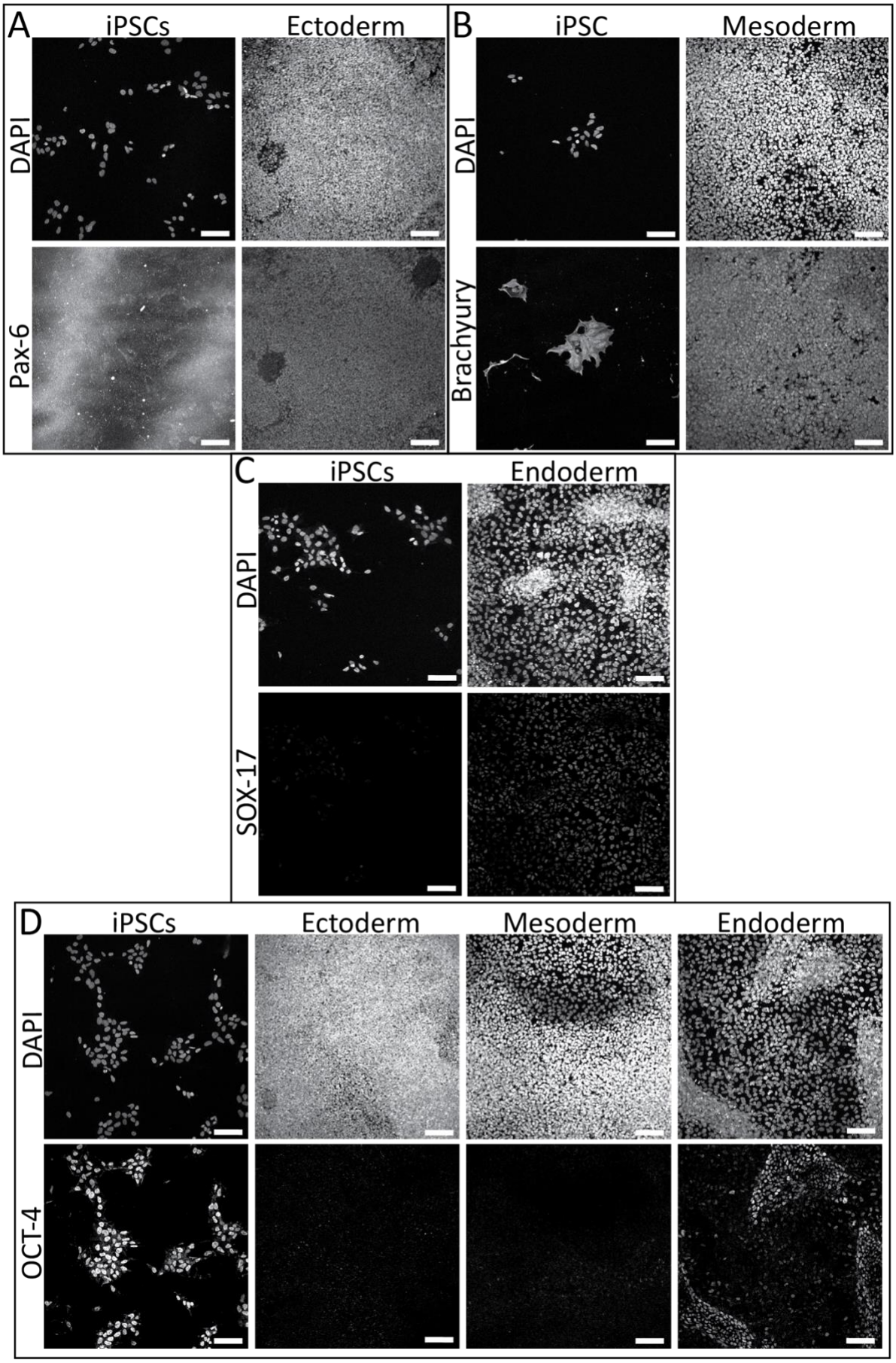
20318A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

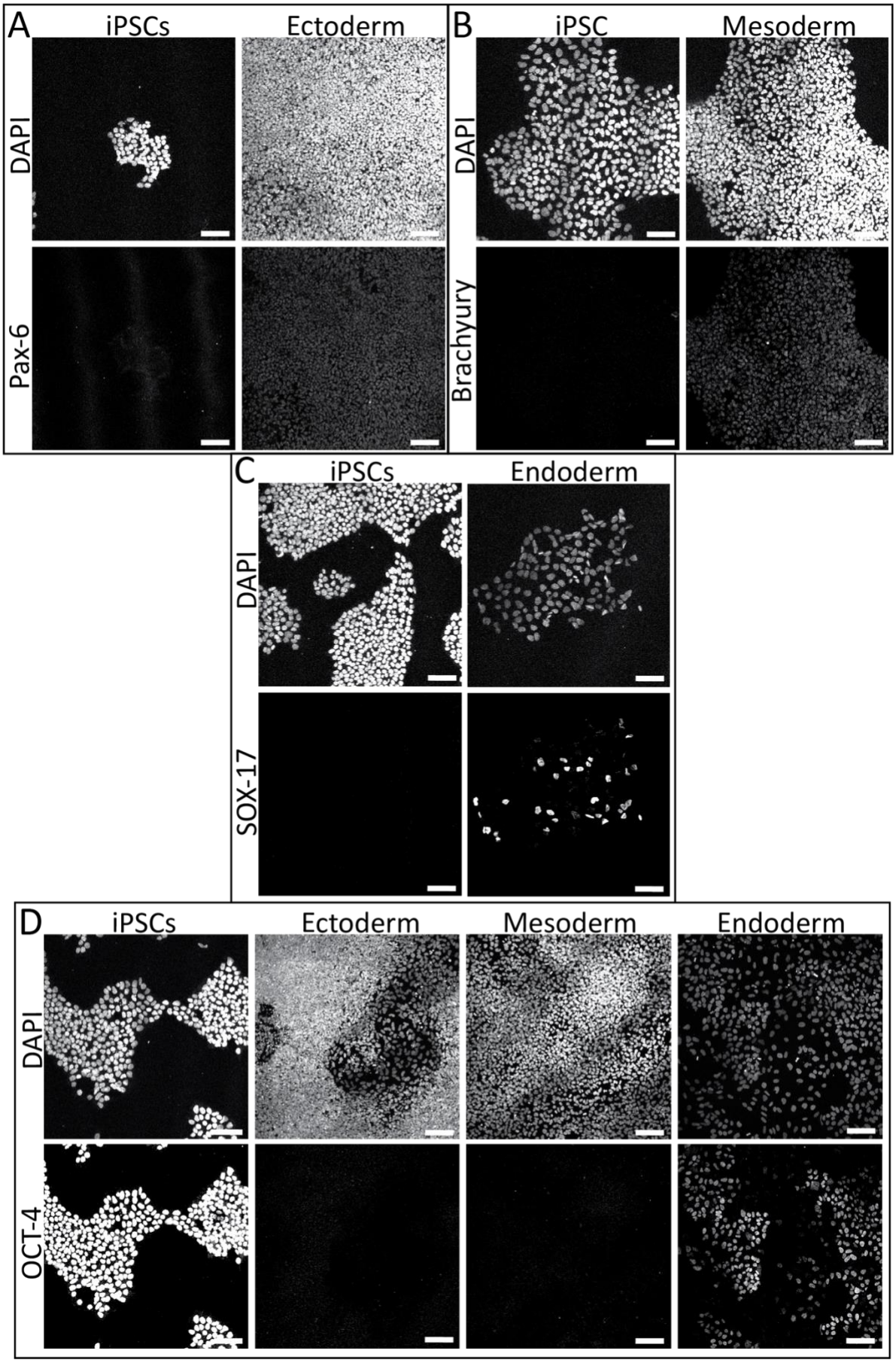
20318B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

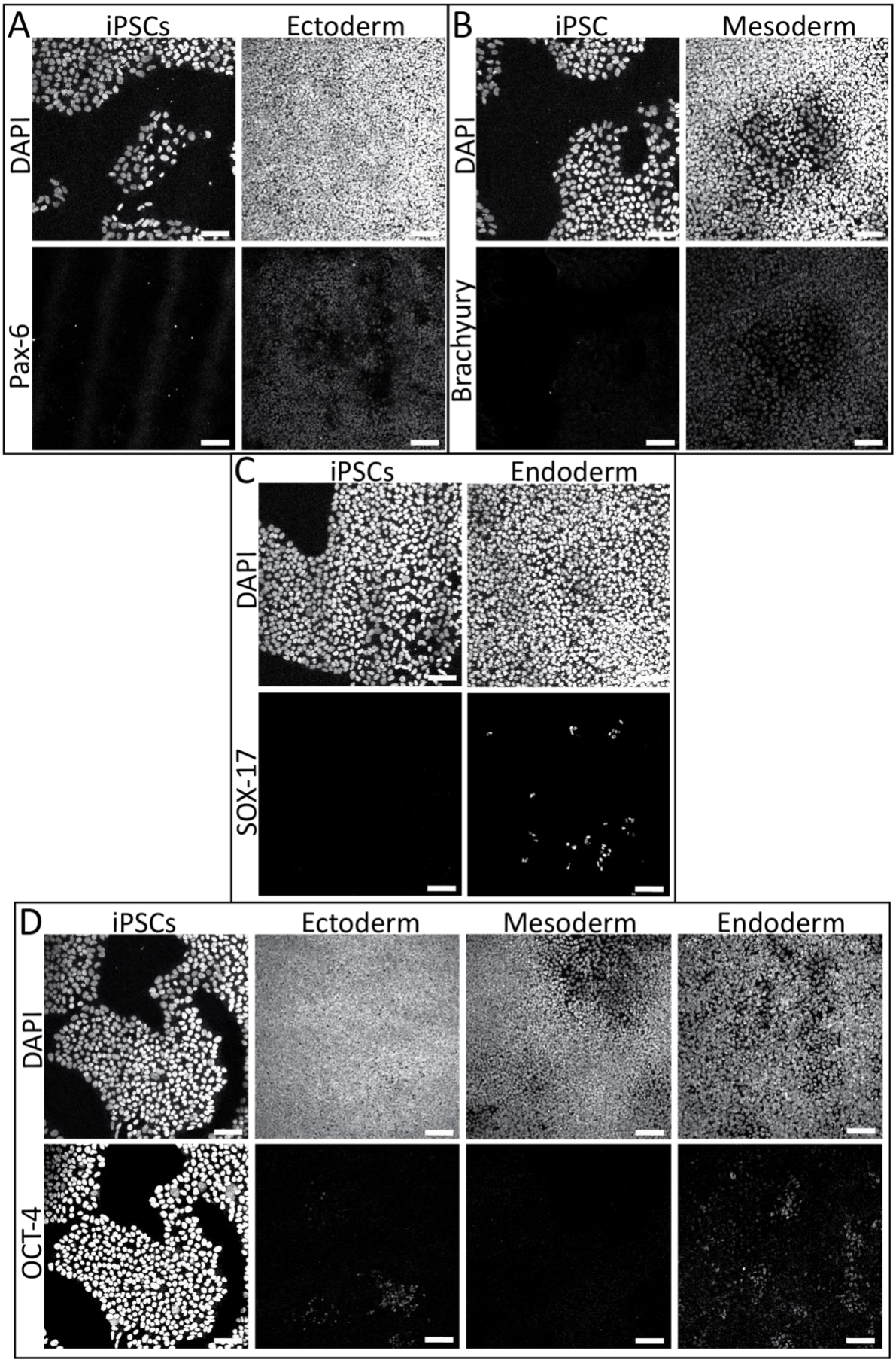
20318C


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

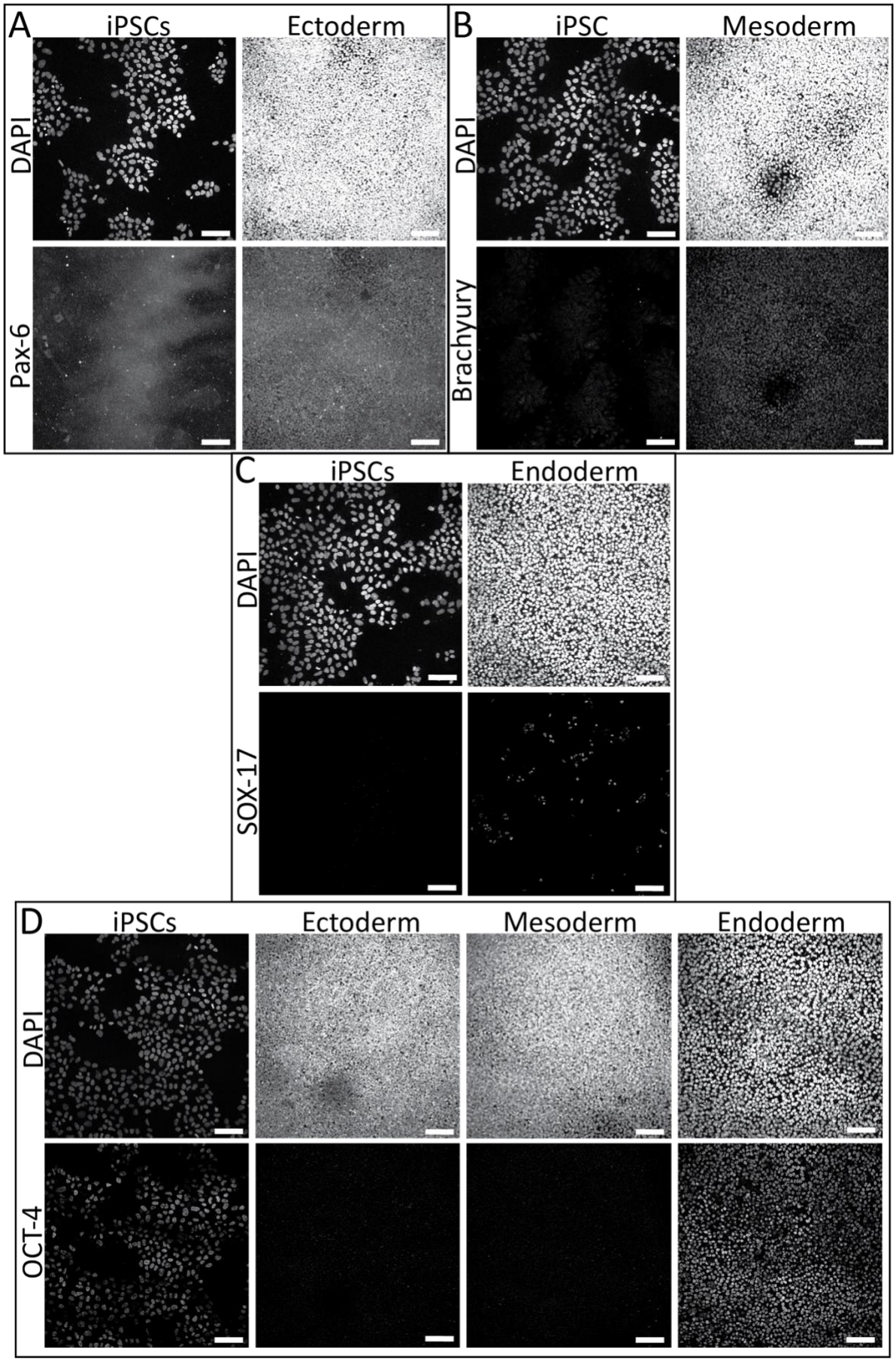
20480A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

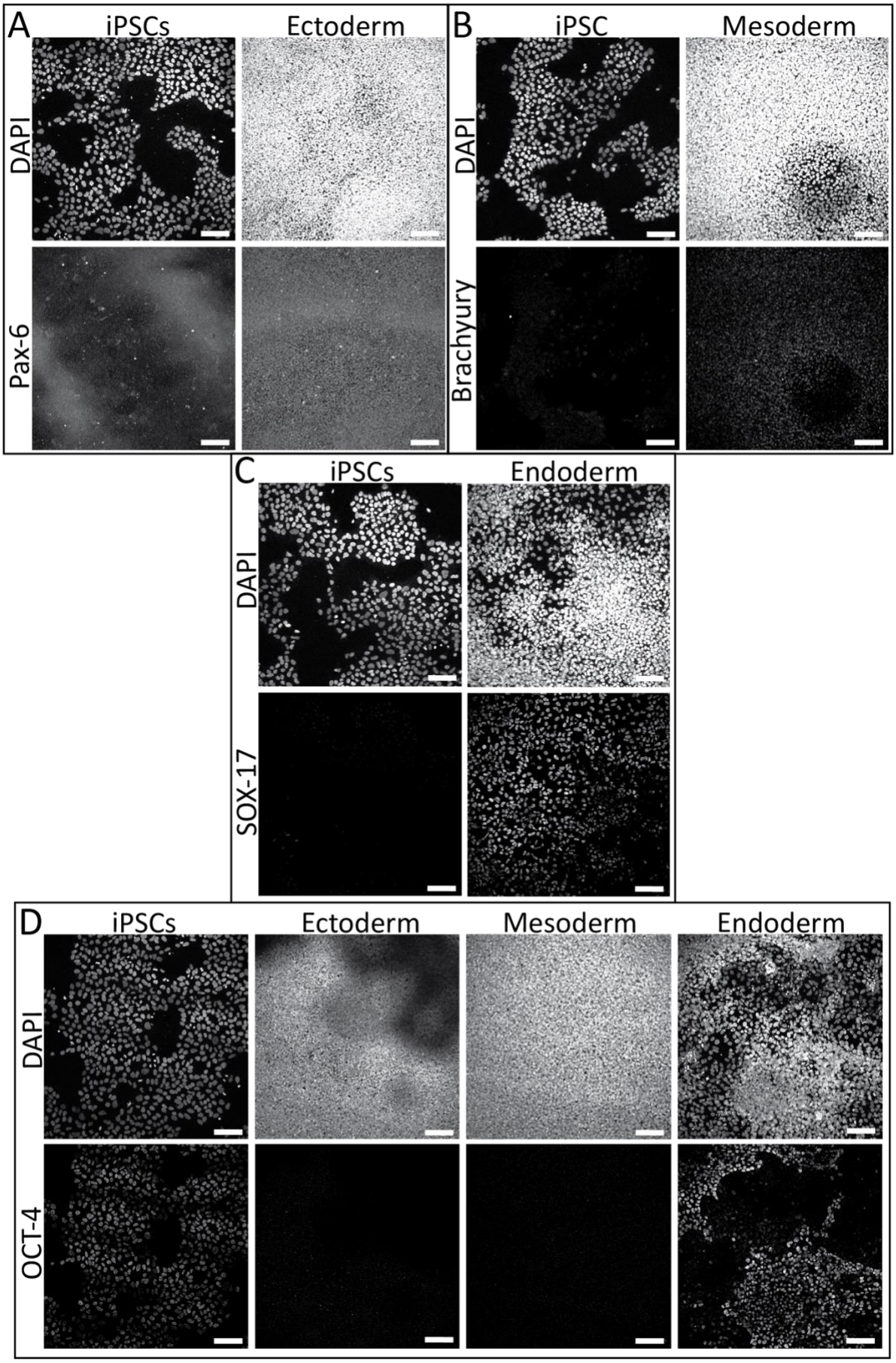
20480B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

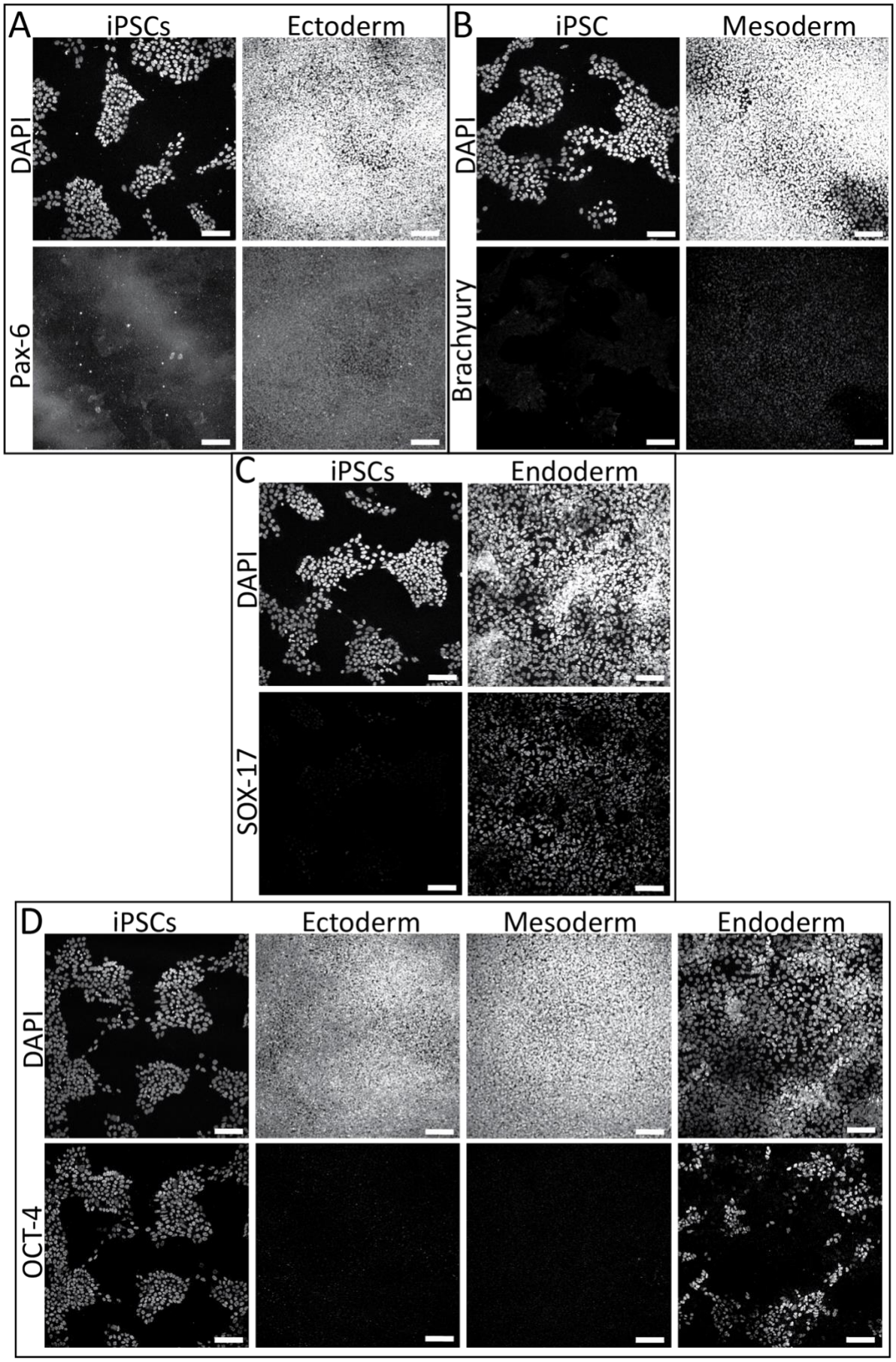
20480C


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

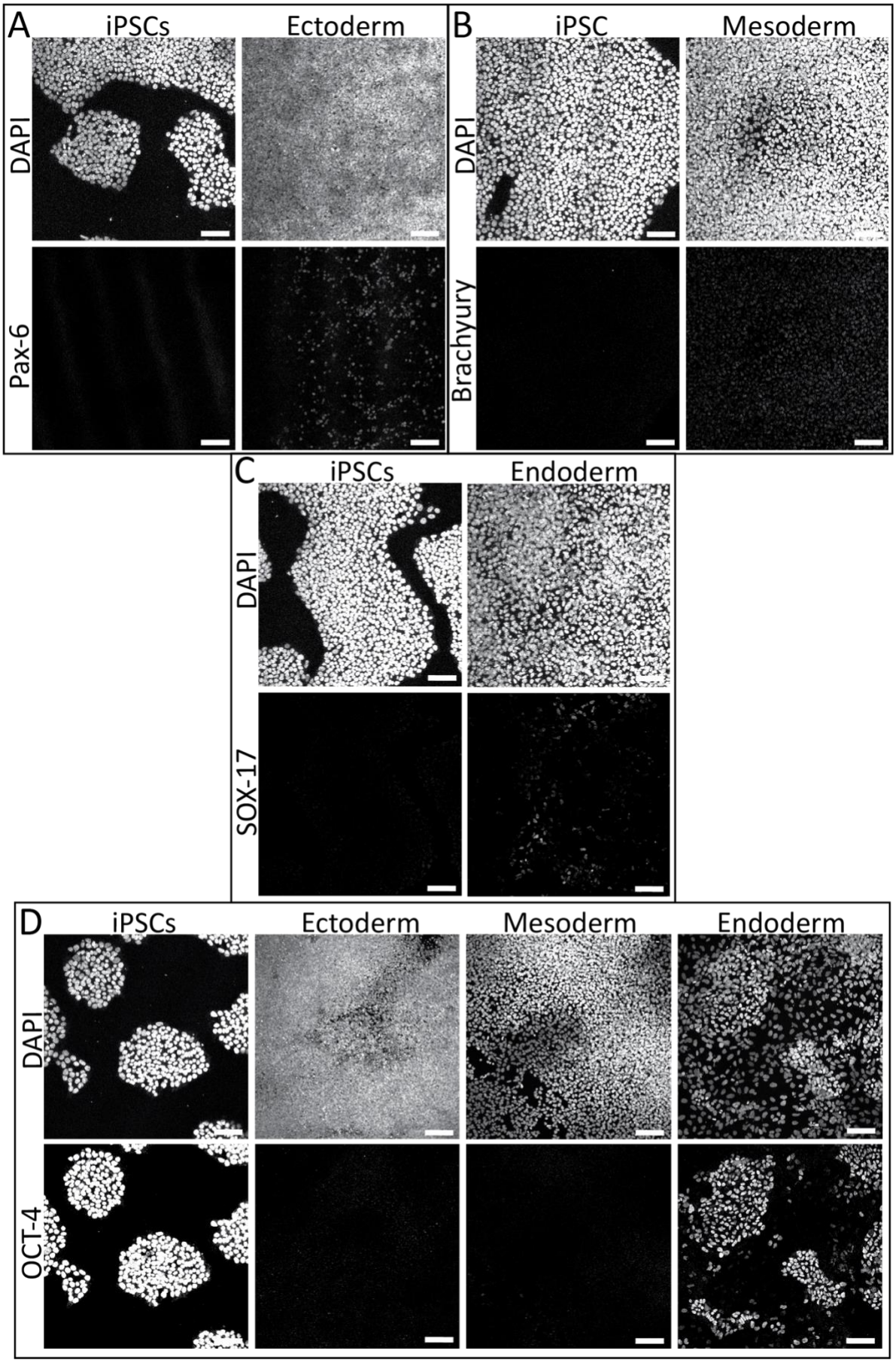
20543A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

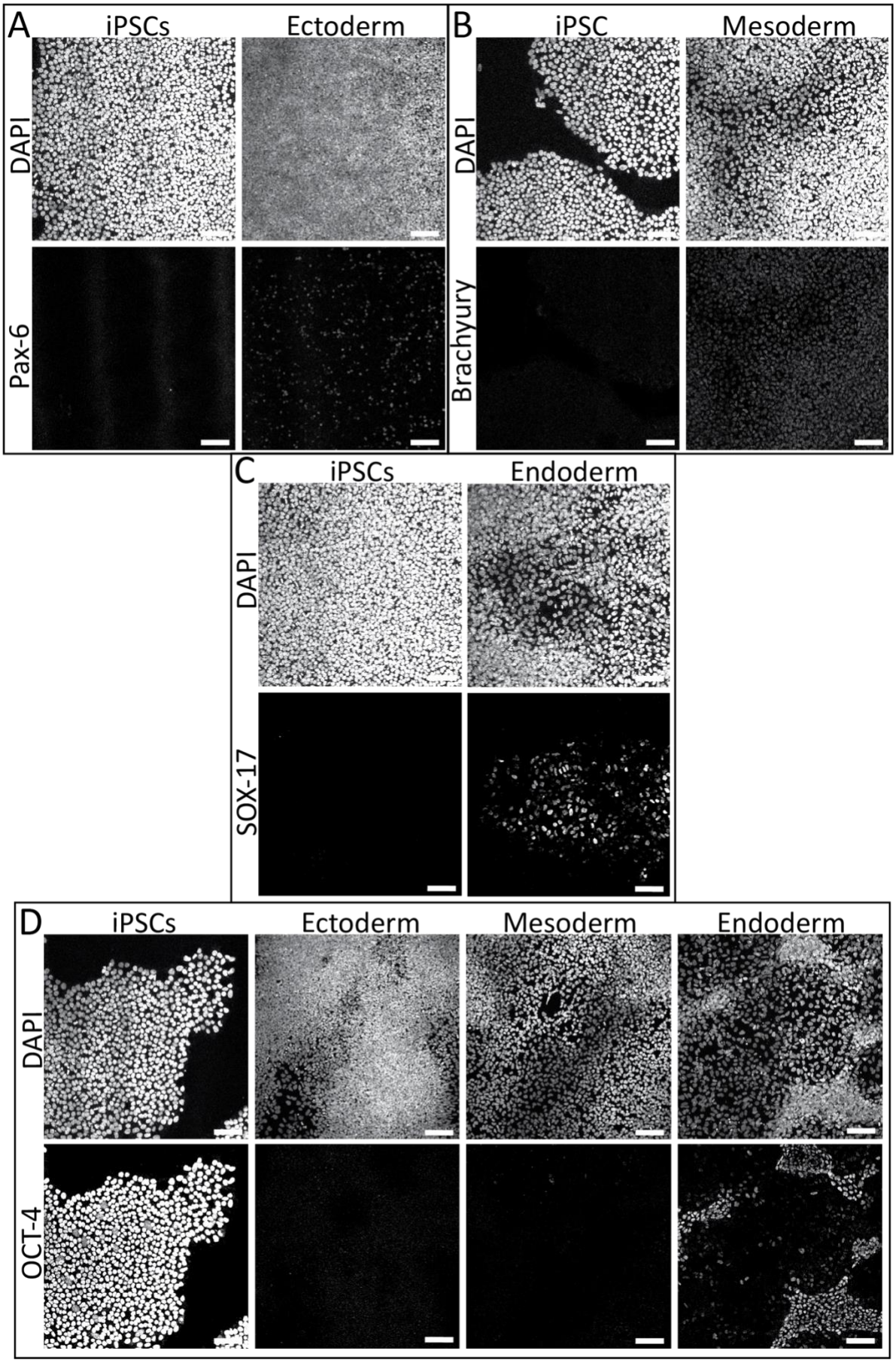
20543B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

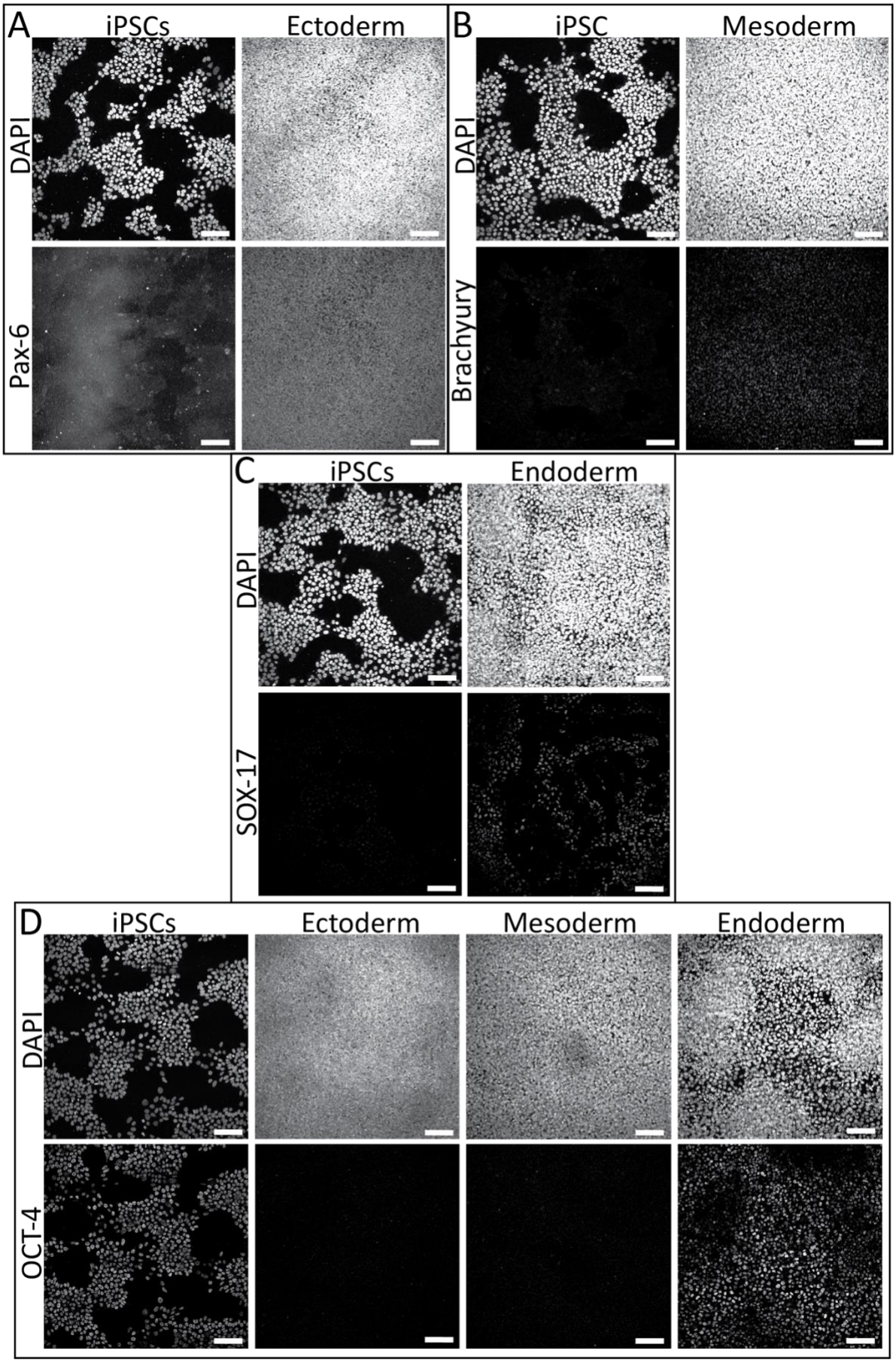
24020A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

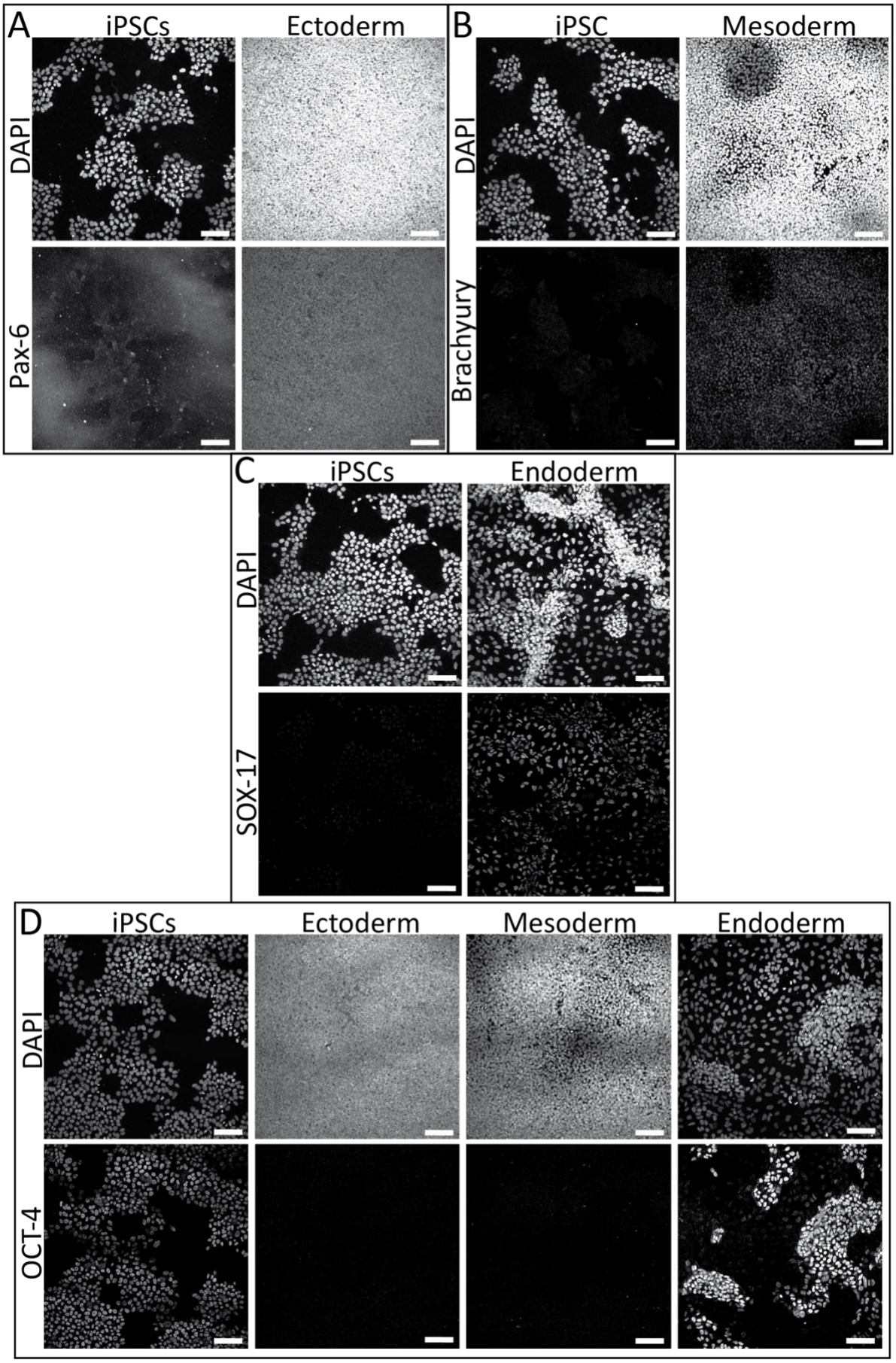
24020B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

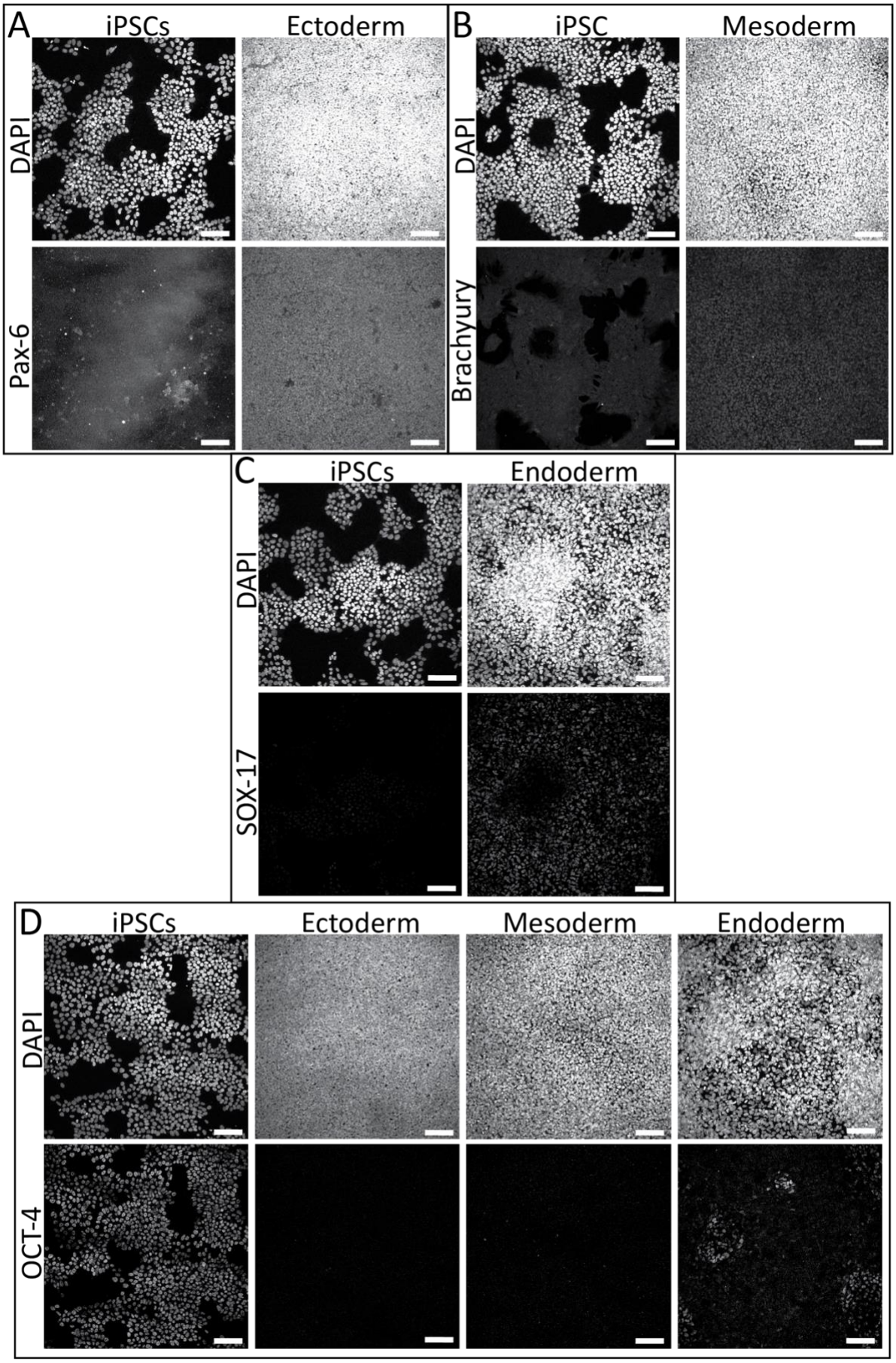
24020C


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

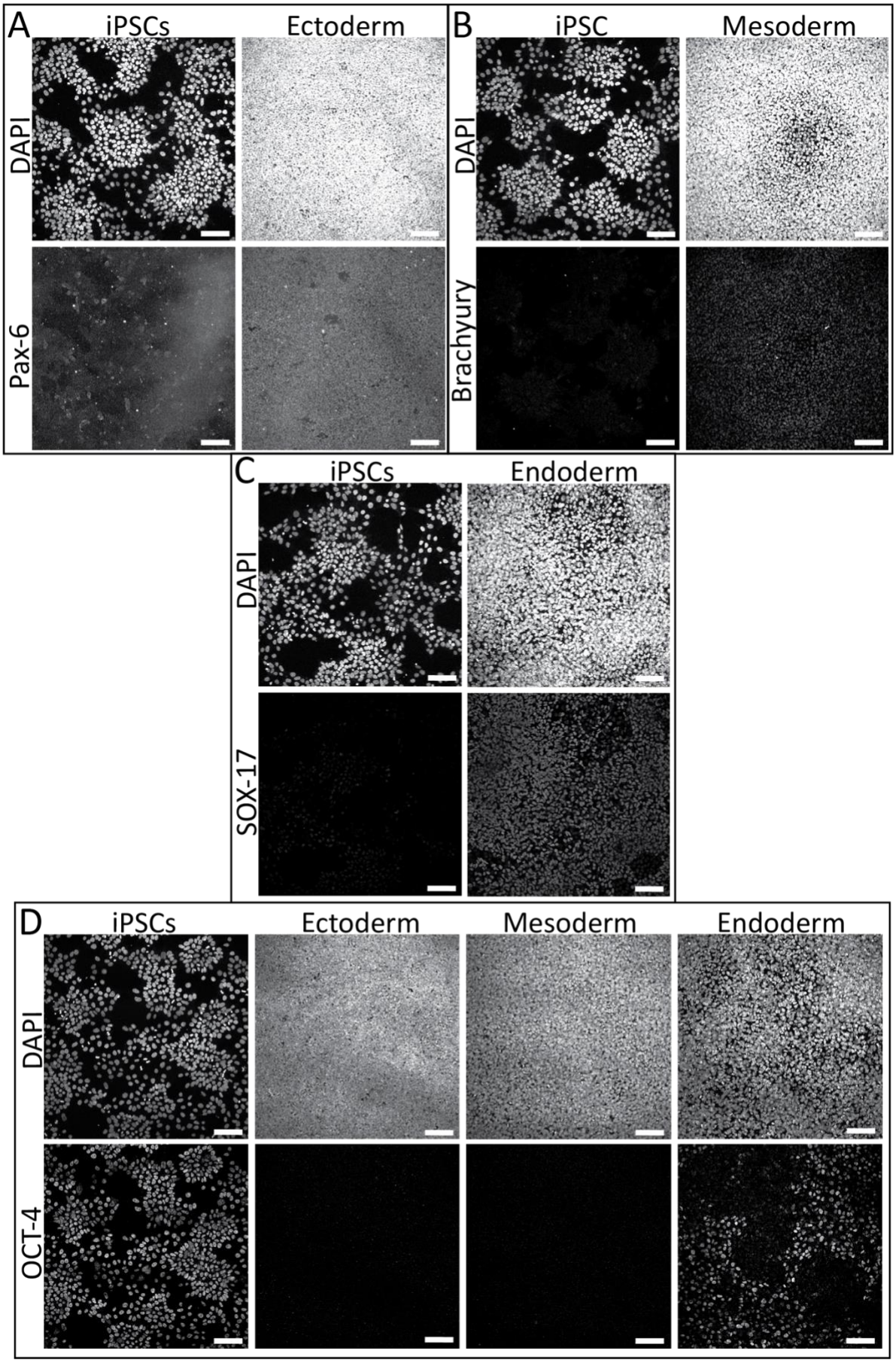
24208A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

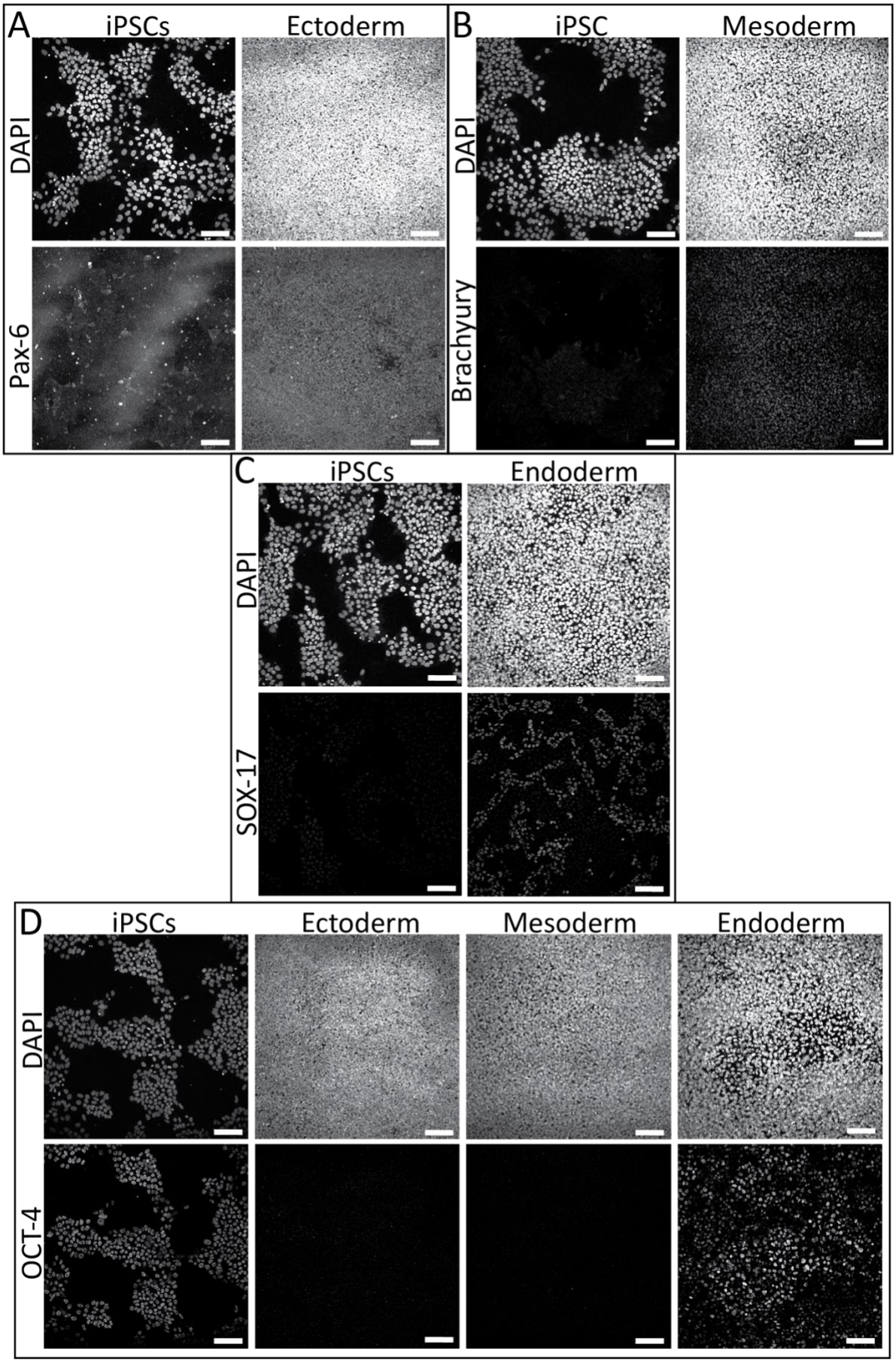
24208B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

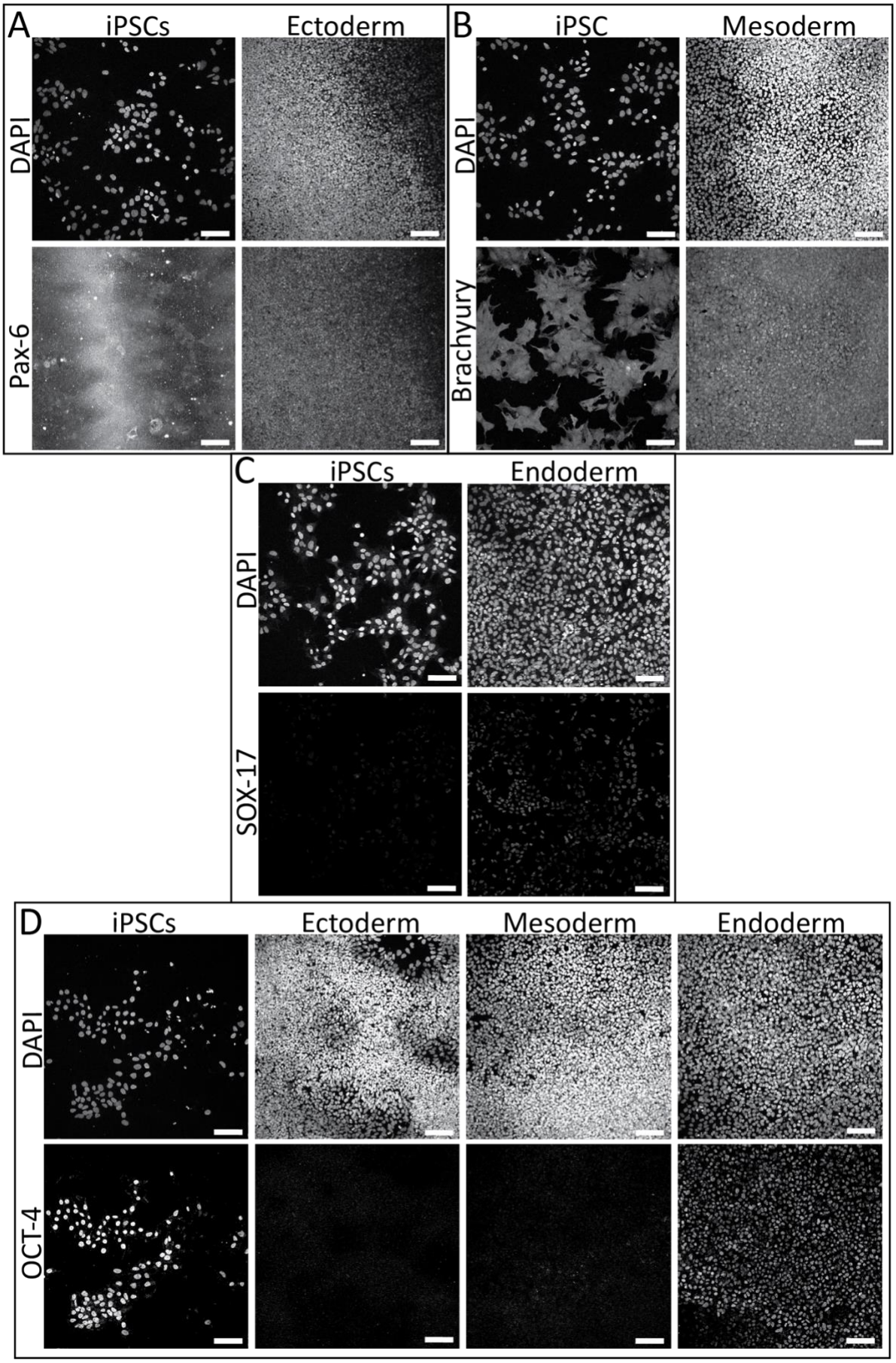
24312A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

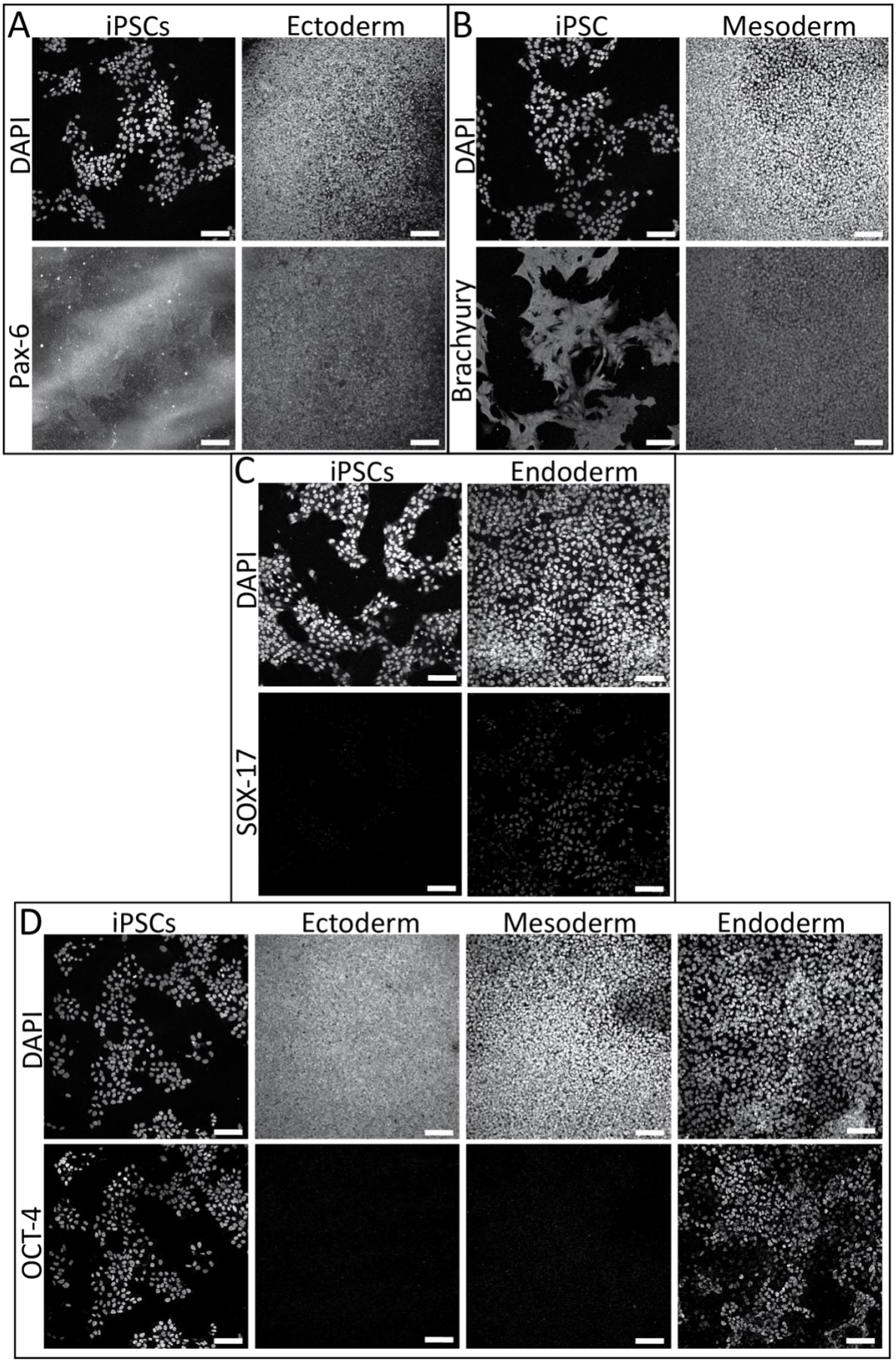
24312B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

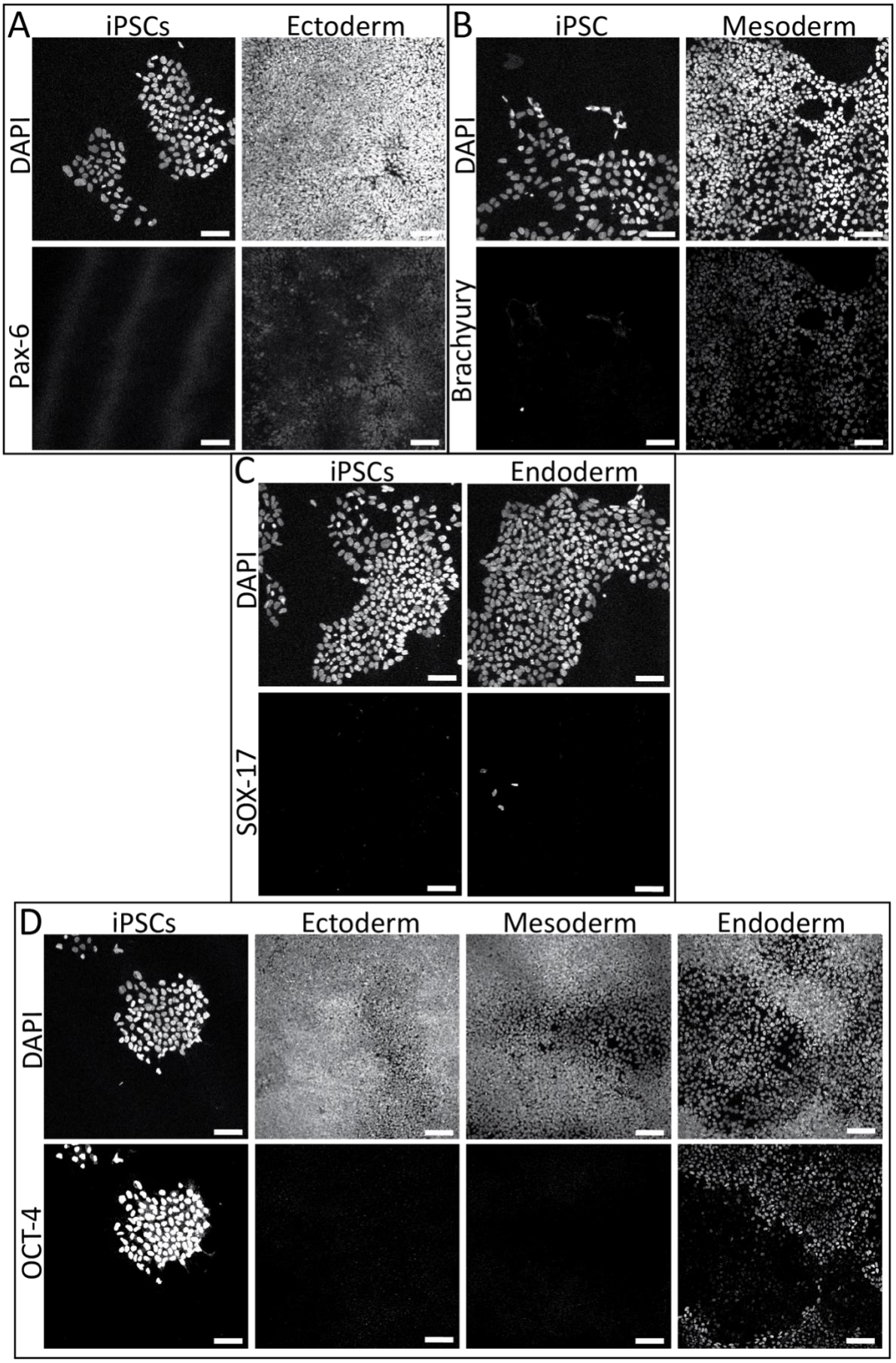
20722A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

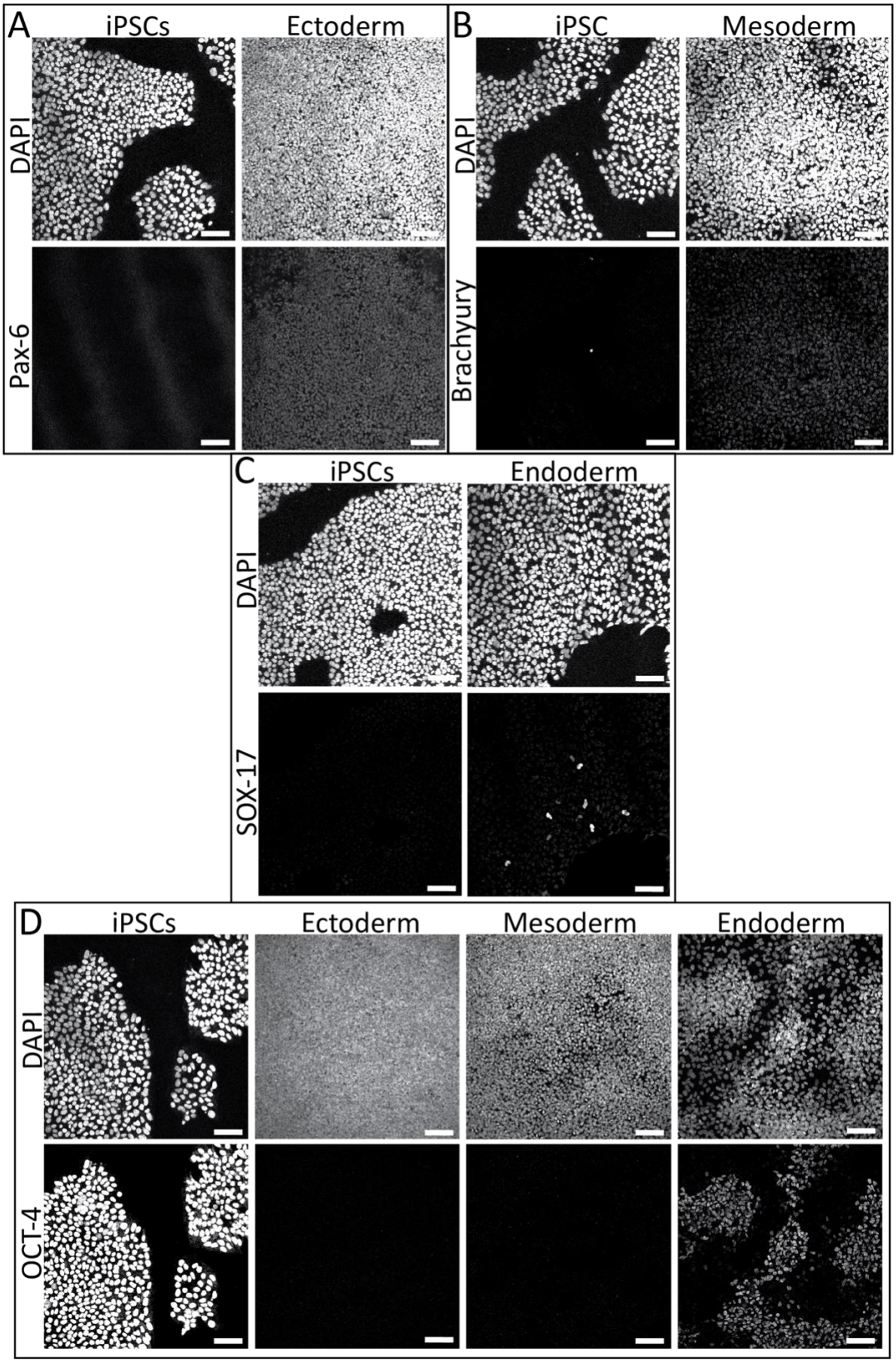
20722B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

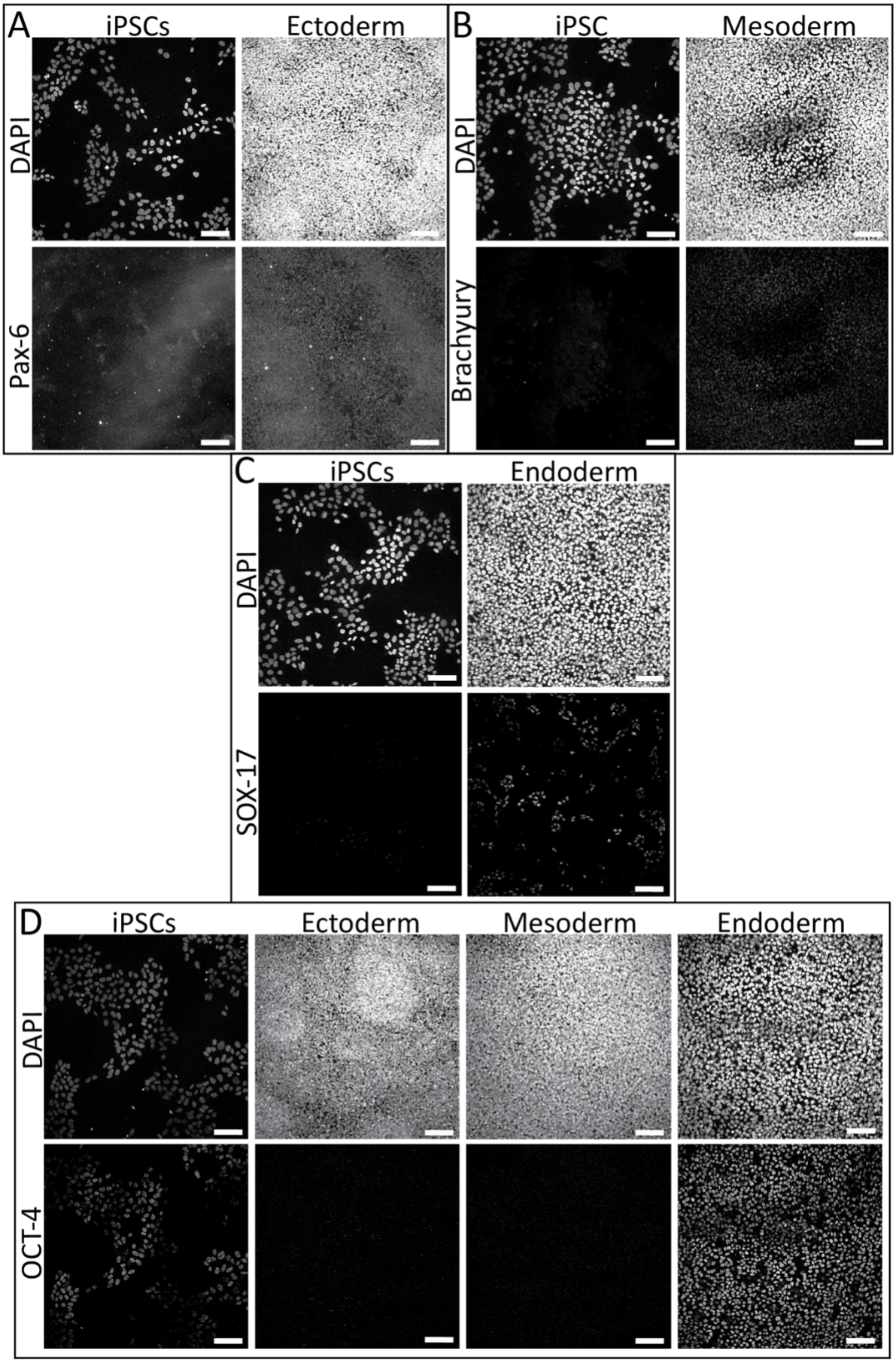
20176A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

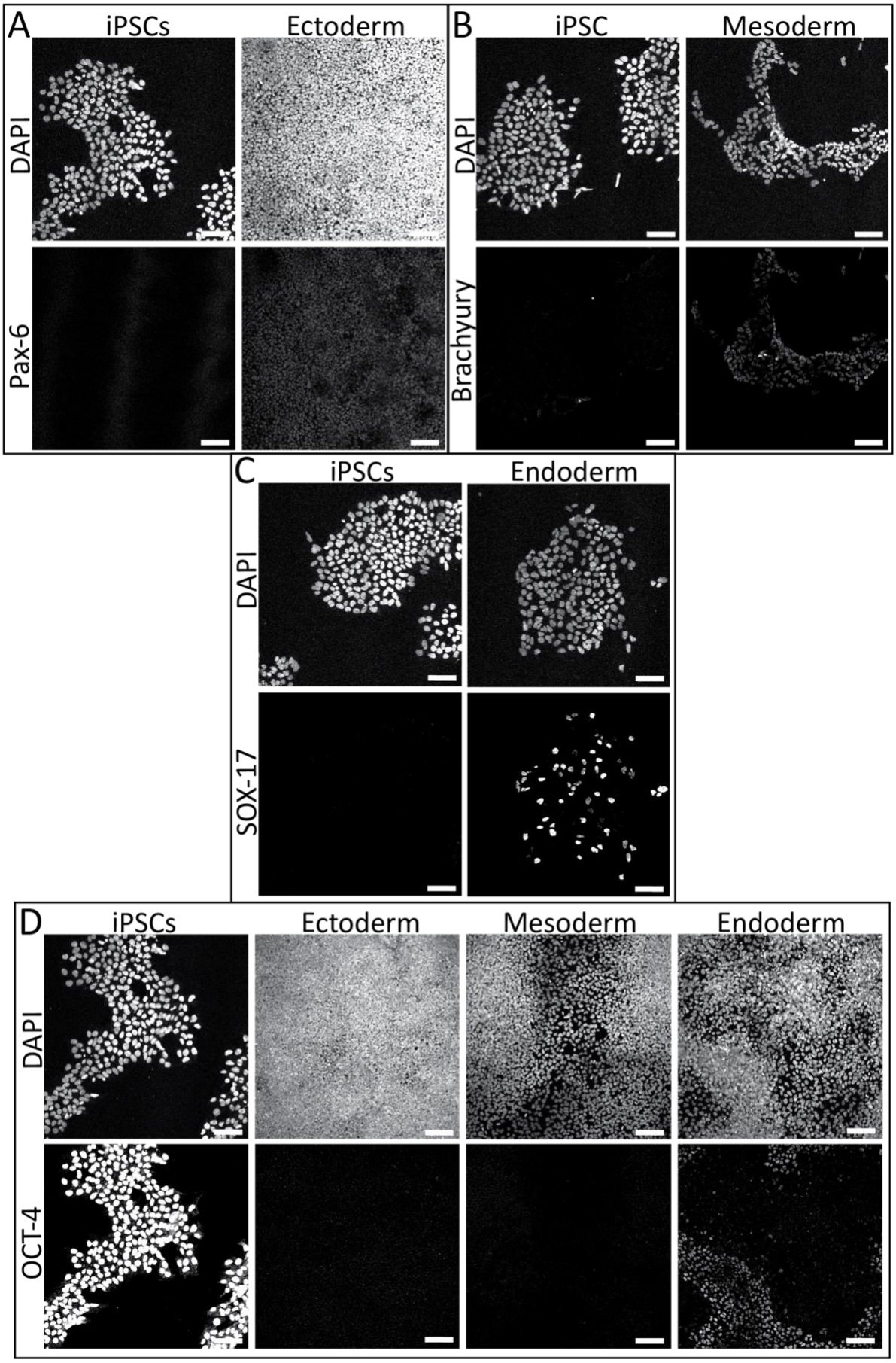
20176B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

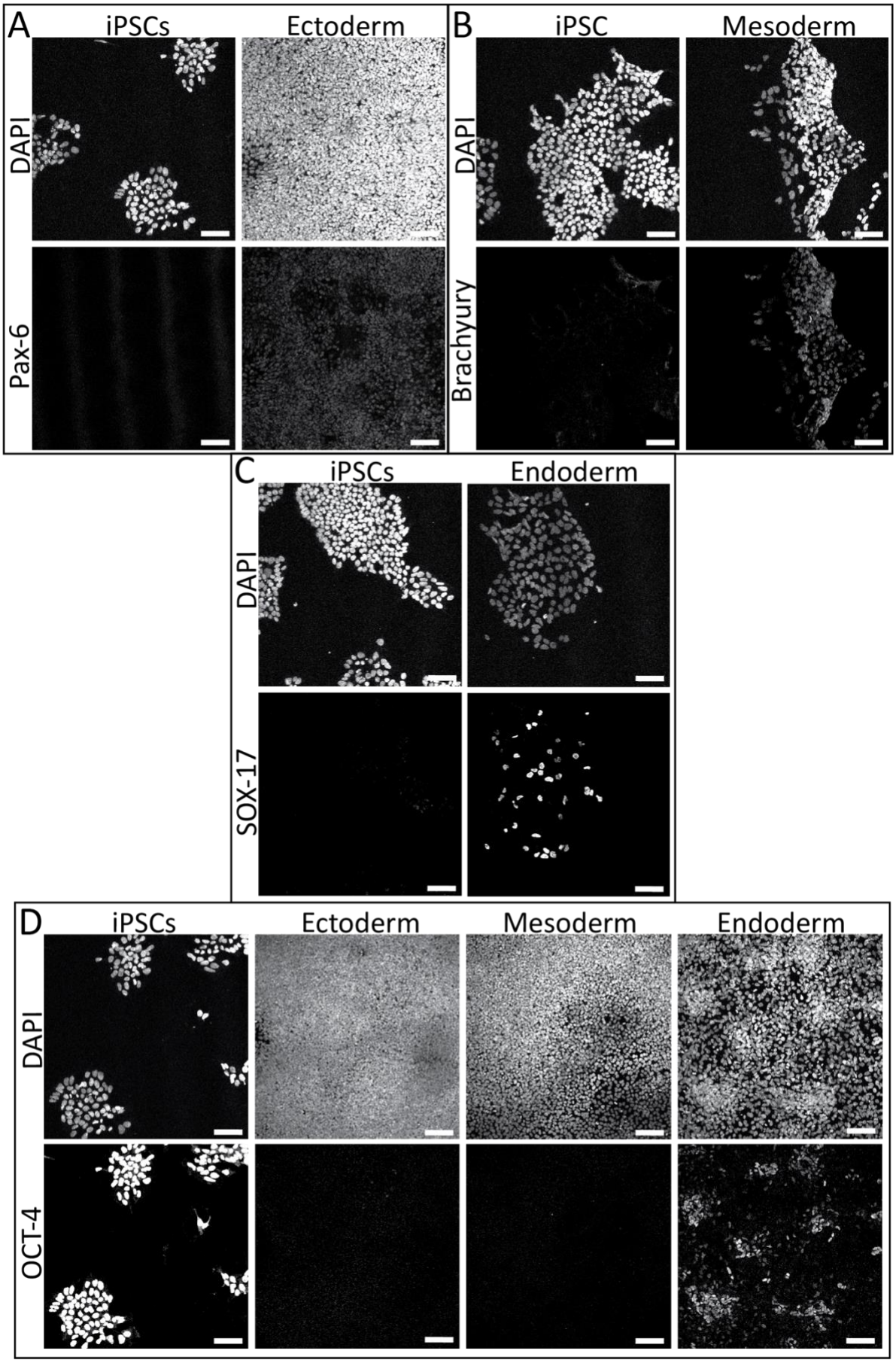
20176C


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

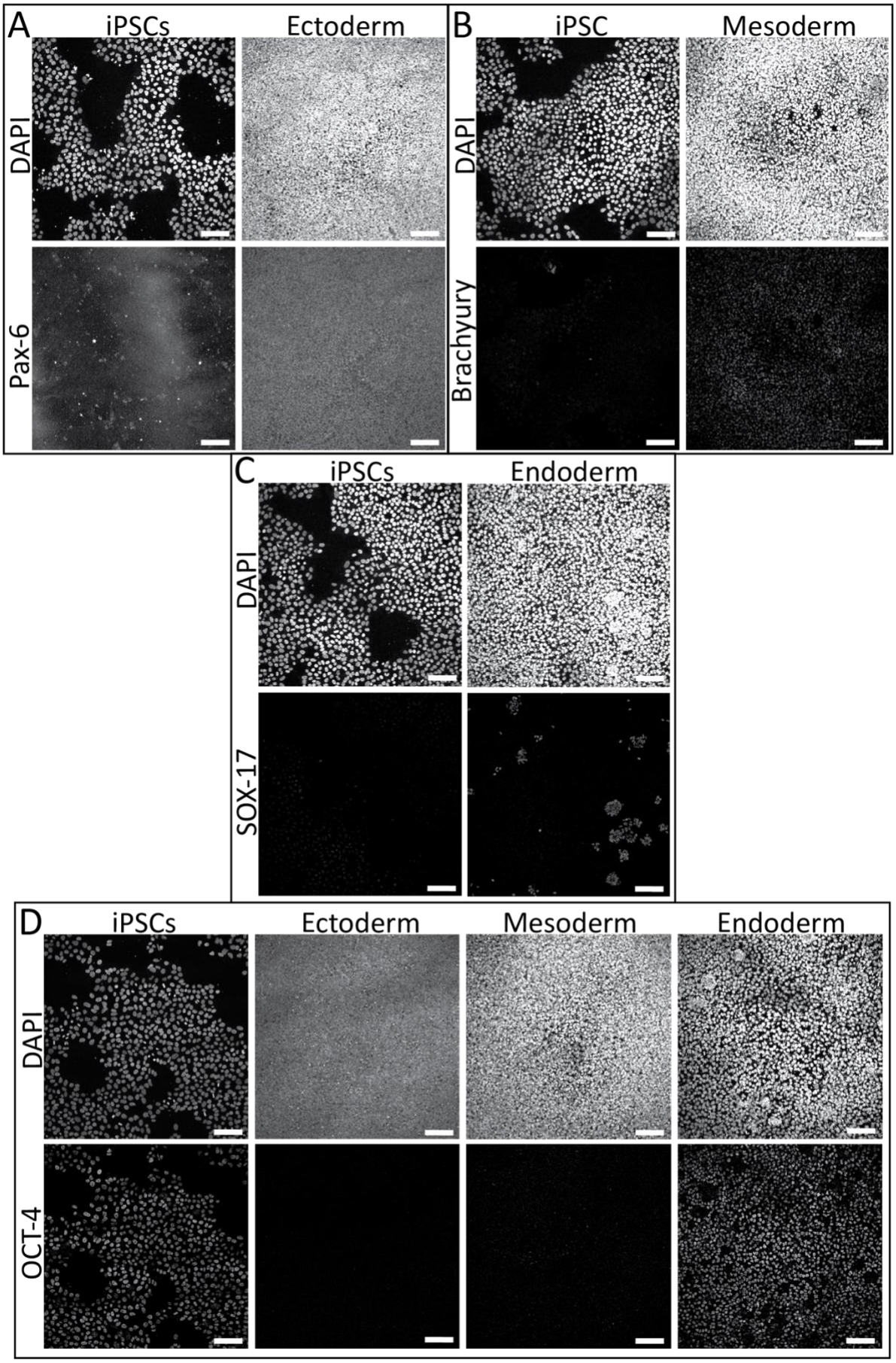
20777A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

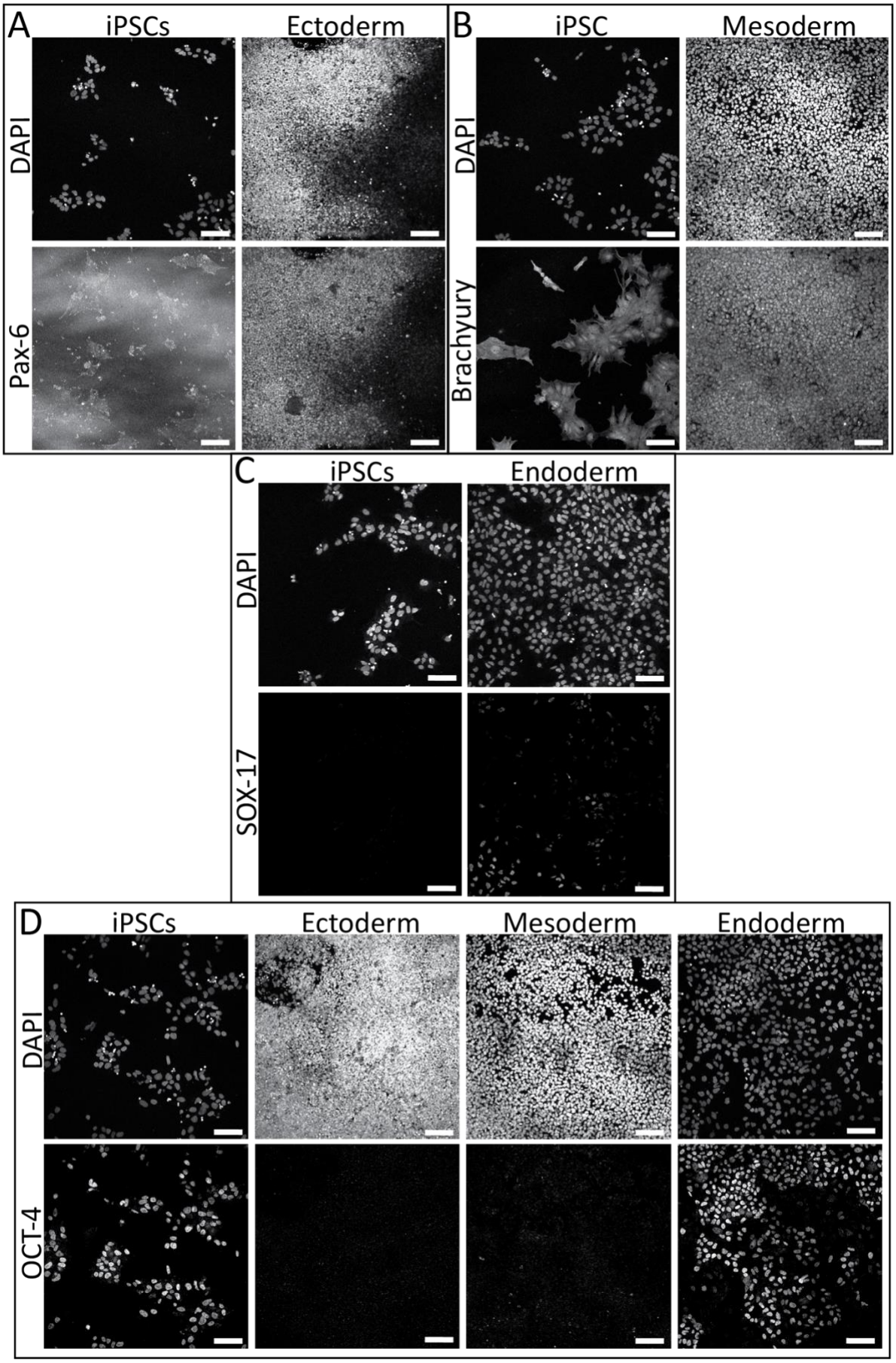
20777B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

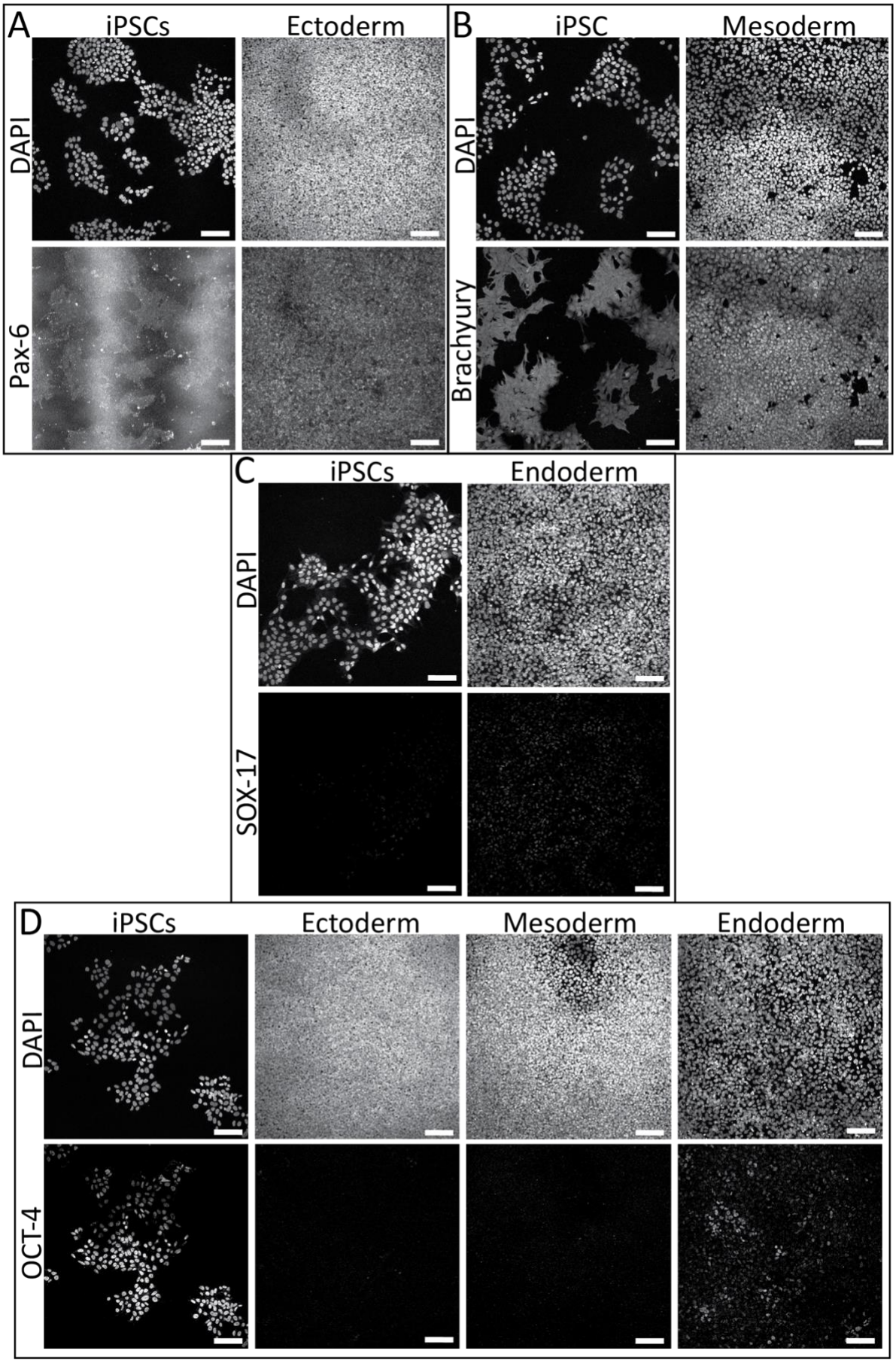
20777C


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

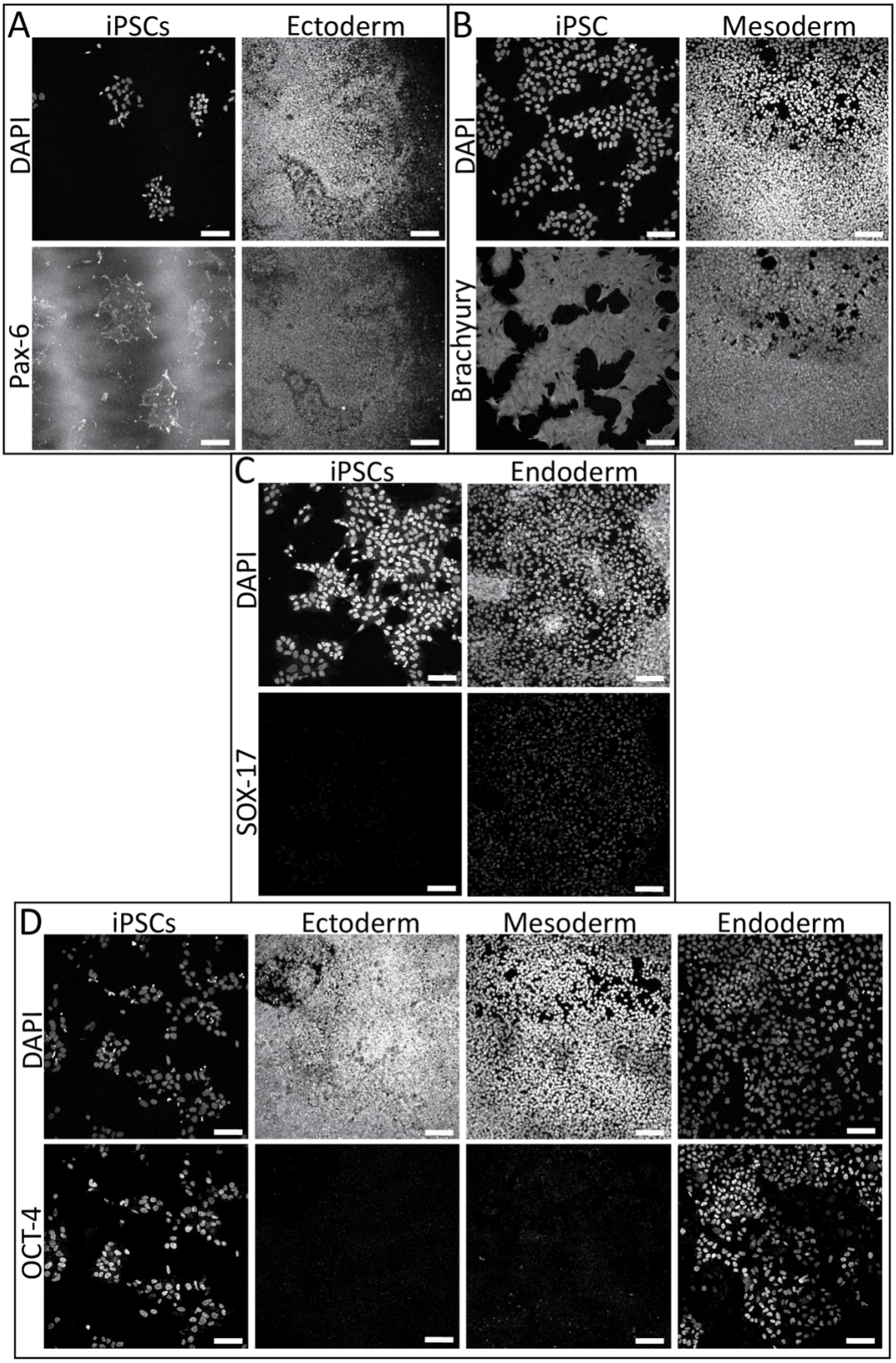
20236A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

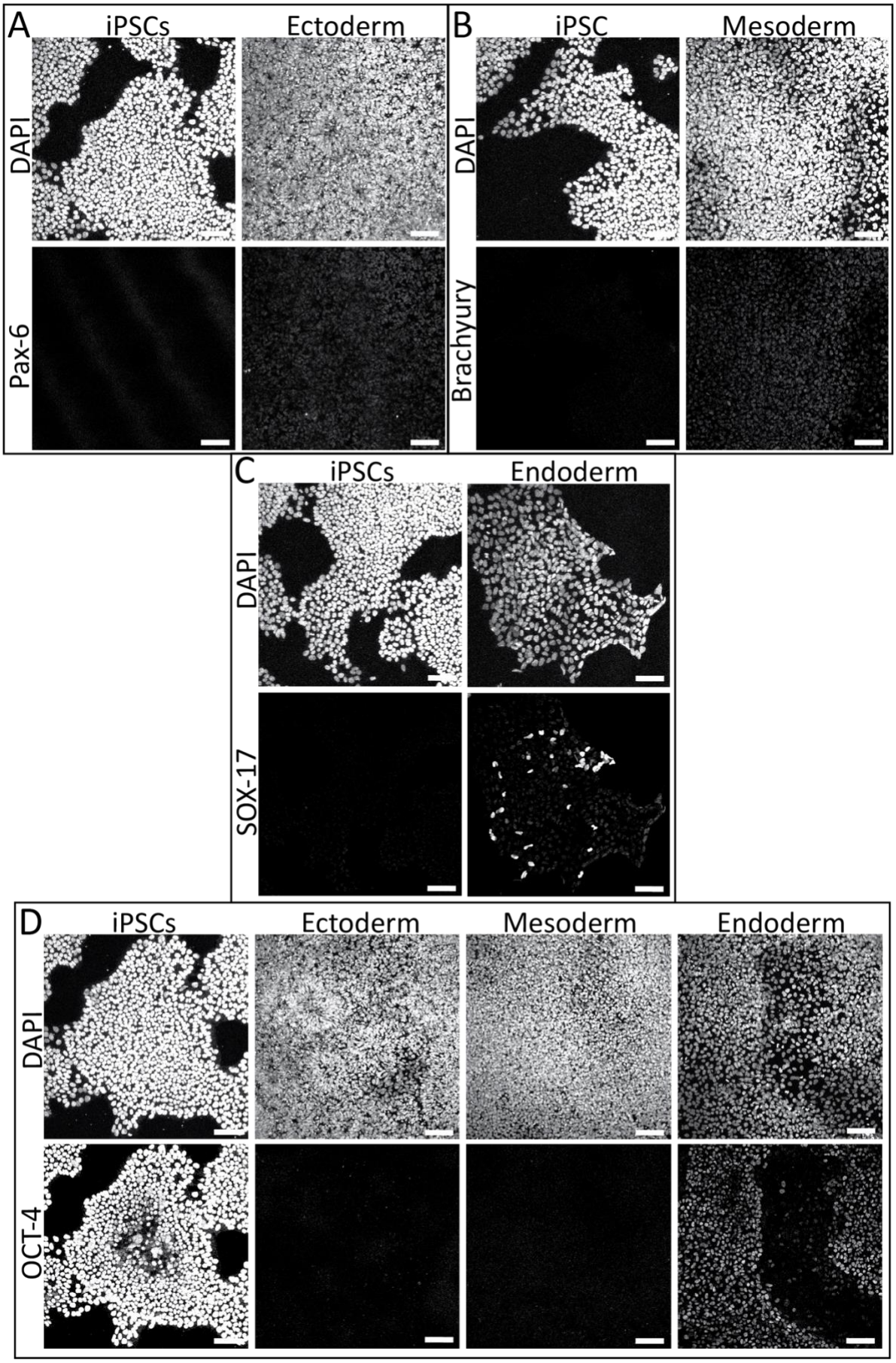
20236B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

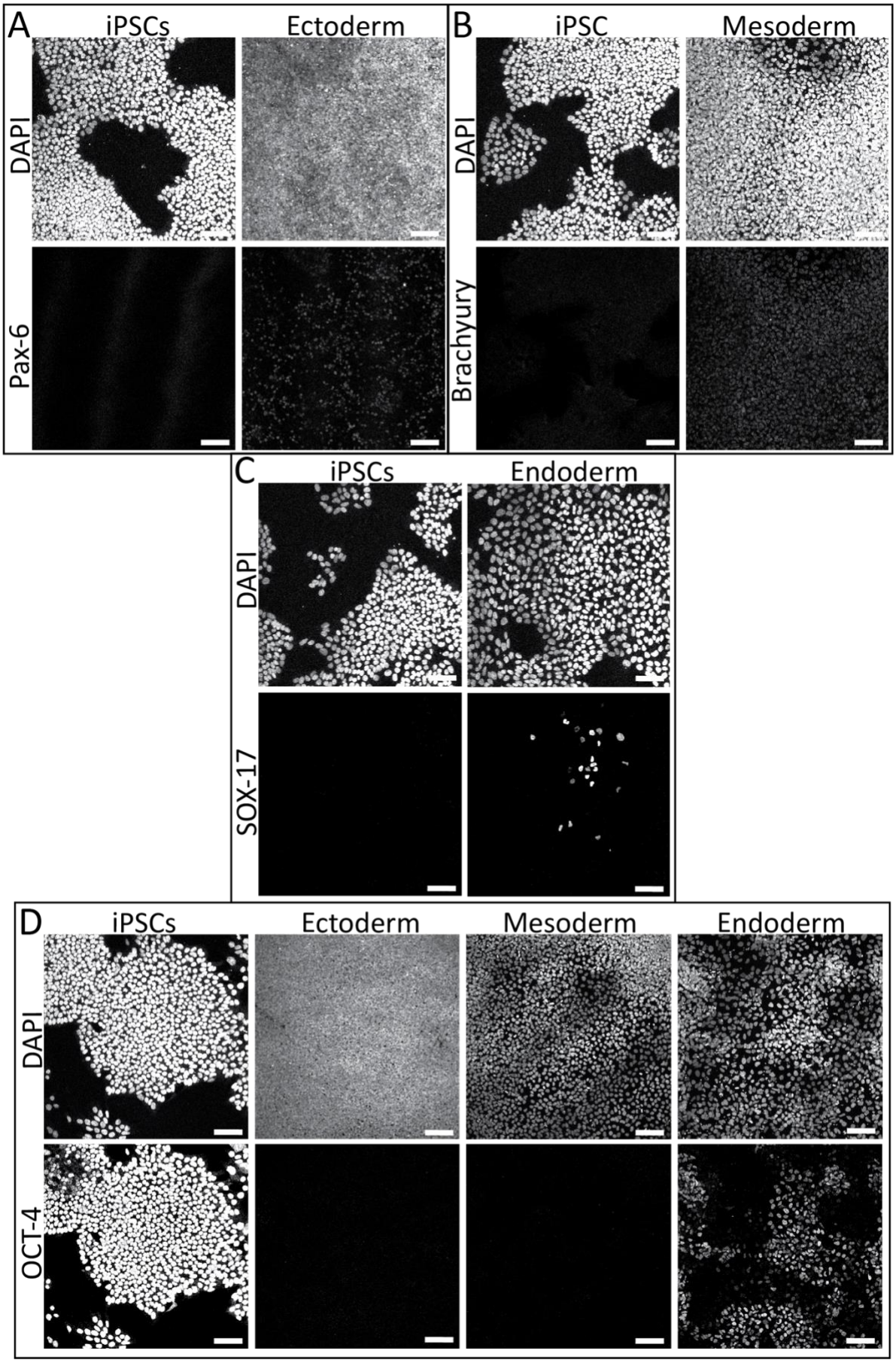
20236C


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

41624A


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

41624B


A
iPSCs
Ectoderm
DAPI
Pax-6
B
iPSC
Mesoderm
DAPI
Brachyury
C
iPSCs
Endoderm
DAPI
SOX-17
D
iPSCs
Ectoderm
Mesoderm
Endoderm
DAPI
OCT-4

Supplementary file 4.3

A