UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Investigations of HIV Latency Through Transcriptomic and Proteomic Profiling

Permalink

https://escholarship.org/uc/item/3p09x3pj

Author

White, Cory Haley

Publication Date

2016

Supplemental Material

https://escholarship.org/uc/item/3p09x3pj#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Investigations of HIV Latency Through Transcriptomic and Proteomic Profiling

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Cory Haley White

Committee in charge:

Professor Douglas D. Richman, Chair
Professor David M. Smith
Professor Celsa A. Spina
Professor Shankar Subramaniam
Professor Christopher H. Woelk
Professor Gene Yeo

2016

Copyright

Cory Haley White, 2016

All rights reserved.

The dissertation of Cory Haley White is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2016

DEDICATION

If I had to choose one person to dedicate this dissertation to, it would have to be my mother. Without her, I would not be anywhere near the person I am now, nor would I have a love of science and mathematics. All throughout my life, she has loved, supported, and pushed me to succeed in whatever I do. Mom, I can never thank you enough. I'm living proof of your accomplishments in life.

To my two brothers and sister in law, Darcy, Jamie, and Ashley, you three are awesome. Darcy, you keep me endlessly entertained with your hilarious antics, particularly all of the fun things you have planned for census workers and telemarketing callers. Jamie, I'm continually amazed at what you can do when you put your mind to it. You have both inspired me all throughout my life. Ashley, you have limitless possibilities ahead of you, I look forward to seeing you grab them.

To my nephews Dillan and Caden. I love you both with all of my heart. Every time I come to visit you, I'm reminded of how much fun you can have and entertainment you bring to others as a toddler. Keep those as you grow older. I will always enjoy spending time with you both, especially as I get to the be the fun uncle.

Finally, to my friends in California, particularly my Tang Soo Do martial arts family, you all have kept me grounded and sane during all of the craziness throughout my 15 years of living in California. To Masters Ricardo Hassan and William Glenn, thank you both for introducing me to martial arts. The discipline and confidence I have gained from that has allowed me to push myself far passed what I once thought were my limitations both in and out of martial arts. To Masters Patrick Marsch and Miguel Soto, thank you Master Marsch for introducing me to Modern Arnis and thank you Master Soto for all of the work you have put into my forms. I will keep those gifts with me wherever I go. Tang Soo!

EPIGRAPH

The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny...'

– Isaac Asimov

TABLE OF CONTENTS

Signature Page iii
Dedication iv
Epigraph v
Table of Contents vi
Glossary and Abbreviations x
List of Supplementary Files xv
List of Figures xvi
List of Tables xvii
Acknowledgements xviii
Vita xxi
Abstract of the Dissertation xxiii

Chapter 1 An Introduction to HIV Latency and the Methods to Study this Disease 1
1.1 HIV - A Global Threat 1
1.2 HIV Latency - The Sleeping Killer 2
1.3 In Vitro Cell Models to Study HIV Latency 3
1.4 Elimination of the Latent HIV Reservoir 4
1.5 Transcriptomic Methods for the Study of HIV Latency and Latency Reversing Agents 11
1.6 Proteomics Methods for the Study of HIV Latency Reversing Agents 14
1.7 Scope of this Dissertation 17

Chapter 2 An Unexpected Finding from an RNA-Seq Analysis of a T_{CM} Model of HIV Latency 18
2.1 Abstract 18
2.2 Introduction 19
2.2.1 The latent HIV model 19
2.2.2 RNA-Seq analysis 20
2.3 Materials and Methods 21
2.3.1 Generation of the DHIV and pLET-LAI plasmids 21
2.3.2 Cell isolation and culture 22
2.3.3 Viral generation and infection 22
2.3.4 Cell activation 22
2.3.5 P24 staining and flow cytometry 23
2.3.6 RNA-Seq data generation and analysis 23
2.4 The RNA-Seq Pipeline 24
2.4.1 Quality control 24

2.4.2 Mapping to a genome or transcriptome reference 25
2.4.3 Quantification of transcripts 26
2.4.4 Differential expression analysis 26
2.4.5 Downstream applications 27
2.4.6 Interchangeability of components 27
2.5 Results 28
2.5.1 Differential expression results 28
2.5.2 Correlation analysis 29
2.6 Discussion 29
2.6.1 Recombination of the DHIV construct with the pLET-LAI plasmid 29
2.6.2 Correlation analysis reveals potential markers of HIV latency 31
2.7 Figures 33
2.8 Tables 39
2.9 Acknowledgements 41
Chapter 3 Transcriptomic Analysis Implicates the p53 Signaling Pathway in the Establishment of HIV Latency in Central Memory CD4 T-Cells 42
3.1 Abstract 42
3.2 Introduction 43
3.3 Materials and Methods 46
3.3.1 Reagents and sample generation 46
3.3.2 RNA isolation and total RNA-Seq data generation 47
3.3.3 Total RNA-Seq data analysis 48
3.3.4 Quantification of spliced transcripts 49
3.3.5 Biological scale normalization (BSN) from resting to activated states 49
3.3.6 Normalization and differential expression within the resting state 50
3.3.7 Functional analysis of differentially expressed genes 52
3.3.8 RT-qPCR validation 52
3.3.9 Flow cytometry and p24 analysis 53
3.3.10 Measurements of HIV DNA with PCR 53
3.4 Results 55
3.4.1 Generation of latently infected cultured T_{CM} cells 55
3.4.2 Cultured T_{CM} cells have a resting and quiescent phenotype 56
3.4.3 Activation of HIV from a latent state 57
3.4.4 Perturbation of host gene expression in latently infected cells 57
3.4.5 Functional analysis of genes perturbed during latency 58
3.4.6 Inhibition of p53 signaling reduces the size of the latent reservoir 59

3.5 Discussion 60
3.5.1 Analysis of host differentially expressed genes and viral transcripts 60
3.5.2 Differential expression reveals dysregulation of multiple genes related to p53 signaling 62
3.5.3 Inhibition of p53 transcriptional activity results in a reduction of latently infected cells and viral integration 64
3.5.4 Limitations and summary conclusions 65
3.6 Figures 67
3.7 Tables 81
3.8 Acknowledgements 83

Chapter 4 Mixed Effects of Suberoylanilide hydroxamic acid (SAHA) on the Host Transcriptome and Proteome and their Implications for HIV Reactivation from Latency 84
4.1 Abstract 84
4.2 Introduction 85
4.3 Materials and Methods 88
4.3.1 CD4 T-cell isolation and SAHA treatment 88
4.3.2 Proteomic and transcriptomic datasets 88
4.3.3 Statistical analyses 89
4.4 Results 90
4.4.1 Proteins and genes modulated by SAHA 90
4.4.2 PINs for proteins and genes modulated by SAHA 91
4.4.3 Functional analysis of proteins and genes modulated by SAHA 92
4.4.4 Validation of gene expression by ddPCR 93
4.5 Discussion 93
4.5.1 SAHA transcriptional and post-transcriptional regulation 93
4.5.2 Known effects of SAHA translate from the RNA to the protein level 95
4.5.3 Effects of SAHA on RNA and proteins with a role in HIV reactivation 95
4.5.4 Conclusions and Implications 98
4.6 Figures 99
4.7 Tables 107
4.8 Supplemental Methods 108
4.8.1 Protein isolation and iTRAQ labeling 108
4.8.2 Peptide pre-fractionation with high-pH reverse-phase C8 chromatography 108
4.8.3 Low pH Ultra HPLC-nano electrospray ionization-Orbitrap MS 109

4.8.4 MS data processing and peptide normalization 109
4.8.5 Ratio testing using the modified permutation method 110
4.8.6 Group testing with limma 111
4.8.7 Gene ontology (GO) analysis using FAIME 112
4.8.8 Droplet digital polymerase chain reaction (ddPCR) .. 113
4.9 Acknowledgements 115
Chapter 5 Dysregulation of Human Endogenous Retroviruses in Primary CD4 T-Cells upon Treatment with SAHA 117
5.1 Abstract 117
5.2 Introduction 118
5.3 Materials and Methods 122
5.3.1 CD4 T-cell isolation and SAHA treatment 122
5.3.2 RNA isolation, preparation, and extraction 123
5.3.3 RNA-Seq data analysis 123
5.3.4 Assessment of HERVs for confirmation 124
5.3.5 Read coverage analysis 124
5.3.6 Droplet digital PCR (ddPCR) analysis 125
5.4 Results 127
5.4.1 Modulation of HERVs upon treatment with SAHA 127
5.4.2 Dose responsive upregulation of selected LTR12 HERV elements in an independent donor set 128
5.4.3 Analysis of HK2, HERV-W, and HERV-FRD expression 129
5.5 Discussion 130
5.5.1 Overview of results 130
5.5.2 Lessons from the study of HERVs with RNA-Seq 131
5.5.3 Implications of HERV dysregulation 132
5.5.4 Conclusions of this work 135
5.6 Figures 137
5.7 Tables 148
5.8 Acknowledgements 150
Chapter 6 Overall Summary and Conclusions 151
6.1 Summary of Results 151
6.2 A Reflection on Models of HIV Latency 153
6.3 The Utility of SAHA in a "Shock and Kill" Strategy for a Cure for HIV 156
6.4 The Importance of Transcriptomic and Proteomic Methods to the Study of HIV Latency 158
6.5 Concluding Statement 160
Bibliography 161

GLOSSARY AND ABBREVIATIONS

1. AIDS – Acquired immune deficiency syndrome
2. ART – Anti-retroviral therapy
3. BLAST – Basic local alignment search tool
4. BLAT – BLAST-like alignment tool
5. BSN – Biological scale normalization
6. Bowtie – Base mapping program used in Tophat
7. bp – base pair
8. cART – Combination antiretroviral therapy
9. *CCR5* – Chemokine C-C motif receptor type 5, required for HIV entry into the cell.
10. CD4 T-cell – T helper cell which expresses the surface protein CD4
11. cDNA – Complementary double stranded DNA synthesized from a template mRNA
12. CPM – Counts per million, a measure for gene expression based upon read count data
13. D1 splice junction – Well known HIV splice site located between the 5' LTR and the *gag* gene
14. D4 splice junction – Well known splice site in HIV located between the *Tat*/*rev* genes and the *Vpu* gene
15. DAPs – Differentially expressed acetylated forms of proteins
16. DART proteins – Dual-affinity re-targeting proteins
17. ddPCR – Droplet digital polymerase chain reaction
18. DEGs – Differentially expressed genes
19. DEPs – Differentially expressed proteins

20. DHIV – Deficient HIV construct utilized in the first iteration of the Bosque and Planelles model
21. DMSO – Dimethyl sulfoxide, the solvent for SAHA
22. DPPs – Differentially expressed phosphorylated forms of proteins
23. EdgeR – Differential expression analysis program utilizing an overdispersed Poisson model to account for biological and technical variability
24. *Env* – Viral gene encoding a protein that forms the viral envelop in retroviruses, including HIV, and is also present in many endogenous retrovirus sequences incorporated into the human genome
25. ERCC – External RNA controls consortium
26. FAIME – Functional analysis of individual microarray expression
27. FASTQ – Read file representing the reads generated by a sequencer and used for mapping to a reference
28. FDR – False discovery rate (Benjamini and Hochberg method)
29. *Gag* – Group specific antigen which codes for the core structural proteins of retroviruses, including HIV-1, and is also present in many endogenous retrovirus sequences incorporated into the human genome
30. GLM – Generalized linear model
31. gMFI – Geometric mean fluorescent intensity
32. GO – Gene ontology
33. HDAC – Histone deacetylase
34. HDACis – Histone deacetylase inhibitors
35. HERV – Human endogenous retrovirus
36. HERVd – HERV database based upon the human reference hg19
37. hg19 reference genome – Human reference genome created in Feb. 2009,

and in use until Dec. 2013

38. hg38 reference genome – Human reference genome created in Dec. 2013 and, as of June 2016, is the most up to date version of the human genome
39. HIV – Human immunodeficiency virus
40. HIV-1_{NL4-3} – A primary HIV-1 strain
41. HMG – High mobility group
42. HMTIs – Histone methyltransferase inhibitors
43. HTSeq – Program for the quantification of output from mapping programs
44. IGB – Integrated genome browser
45. IGV – Integrative genomics viewer
46. in vivo – Latin for "within the living"
47. iTRAQ – Isobaric tags for relative absolute quantification
48. KEGG – Kyoto encyclopedia of genes and genomes
49. LC-MS/MS – Liquid chromatography tandem mass spectrometry
50. LI – Latently infected
51. LIA – Latently infected activated
52. LTR12 – HERV family whose members are highly upregulated by HDACis, such as SAHA
53. Log_2FC – Log base 2 fold changes calculated by first obtaining the fold change and then taking the base 2 logarithm of that fold change
54. LRA – Latency reversing agent
55. LTR – Long terminal repeat
56. mRNA – Messenger RNA
57. MS – Multiply spliced transcripts
58. *Nef* – Negative regulatory factor protein encoded by the HIV-1 virus

59. NER – Nuclear excision repair
60. NK Cells – Natural killer cells that provide a rapid response to viral infection
61. PBMCs – Peripheral blood mononuclear cells
62. PCR – Polymerase chain reaction
63. Pifithrin-α – Reversible inhibitor of p53-mediated apoptosis and p53 dependent gene transcription
64. PIN – Protein interaction network
65. PKC – Protein kinase C
66. pLET-LAI – Plasmid containing the envelop protein provided with the replication deficient HIV construct in the Bosque and Planelles latency model
67. *Pol* – DNA polymerase gene found in retroviruses, including HIV-1, and is also present in many endogenous retrovirus sequences incorporated into the human genome
68. p24 – HIV viral protein, also known as the capsid protein
69. p53 – Tumor protein p53 with roles in DNA repair, growth arrest, and apoptosis
70. p-TEFb – Positive transcription elongation factor
71. RepeatMasker – Program to screen DNA sequences for repeats and low complexity DNA, such as those present in HERV sequences
72. q-value – FDR corrected (Benjamini and Hochberg method) p-value
73. RNA-Seq – Next generation sequencing technology for high throughput analysis of transcriptional output
74. *RPL27* – Gene used to normalize ddPCR results in samples treated with SAHA and, importantly, not affected by treatment with SAHA
75. *Rev* – Transactivating protein essential to the regulation of HIV-1 protein

expression

76. RIN – RNA integrity number
77. RNAse P – Ribonuclease P
78. RPMI – Roswell Park Memorial Institute
79. RT-qPCR – Real time quantitative PCR
80. RUVSeq – Remove unwanted variation Seq program used to generate a vector corresponding to negative control ERCCs to remove unwanted variation arising from nuisance technical sources, such as batch effects
81. SAHA – Suberoylanilide hydroxamic acid (a.k.a. Vorinostat)
82. "Shock and Kill" – Strategy to eliminate the latently HIV infected reservoir by shocking the virus out of latency and killing the infected cell
83. siRNA – Small interfering RNA
84. SS – Single spliced transcripts
85. *Tat* – HIV-1 trans-activating protein
86. T_{CM} – Central memory CD4 T-cell
87. Tophat – Mapping program designed to work with transcriptome data which may include splicing
88. UI – Uninfected
89. UIA – Uninfected activated
90. US – Unspliced transcripts
91. *Vpu* – Viral protein unique gene found in the HIV-1 genome
92. ZFN – Zinc finger nuclease
93. αCD3/αCD28 beads – αCD3/αCD28 Coated magnetic beads to activate Naïve CD4 T-cells

LIST OF SUPPLEMENTARY FILES

Table S3.1: White_Table_S3-1_DEG_list_LI_vs_UI.csv
Table S4.1: White_Table_S4-1_Differentially_expressed_proteins_and_their_post-translationally_modified_forms.xlsx
Table S4.2: White_Table_S4-2_ Differentially_expressed_genes.xlsx
Table S4.3: White_Table_S4-3_Gene_ontology_terms_modulated_by_SAHA.xlsx

LIST OF FIGURES

Figure 1.1: The "Shock and Kill" strategy to facilitate a cure for HIV 8
Figure 2.1: The RNA-Seq pipeline 33
Figure 2.2: Example of high quality sequencing reads 34
Figure 2.3: Example of poor quality sequencing reads 35
Figure 2.4: Tophat segmented search algorithm 36
Figure 2.5: HIV read coverage 37
Figure 2.6: Reads mapping across the deleted region break points 37
Figure 2.7: Recombination of the pLET-LAI plasmid with DHIV 38
Figure 2.8: Candidate marker of HIV latency 38
Figure 3.1: Generation of the Martins *et al*. model 67
Figure 3.2: Average RNA content increases upon activation of CD4 T-cells 68
Figure 3.3: Known markers of CD4 T-cell activation are modulated following CD3/CD28 stimulation 68
Figure 3.4: Abundance of reads mapping to *IL2* 69
Figure 3.5: HIV splicing variants are upregulated following CD3/CD28 stimulation 70
Figure 3.6: Heatmap of differentially expressed genes during HIV latency 71
Figure 3.7: Overlap among four models of HIV latency 72
Figure 3.8: Functional analysis of genes modulated during HIV latency 73
Figure 3.9: Effect of pifithrin-α upon the establishment of latency 76
Figure 4.1: Overlap among differentially expressed proteins and genes 99
Figure 4.2: Protein interaction network (PIN) for combined DEPs, DPPs, DAPs and DEGs 100
Figure 4.3: Expanded PIN 101
Figure 4.4: Scatter-plot of significant GO terms 102
Figure 4.5: Overlap among GO terms modulated by SAHA at the RNA and protein levels 103
Figure 4.6: GO terms significantly modulated by SAHA 104
Figure 4.7: Validation of gene expression by droplet digital PCR (ddPCR) 105
Figure 4.8: The mass spectrometry characterization of the effect of SAHA on *HMGA1* expression 106
Figure 5.1: Overall dysregulation of all HERV elements from HERV families by SAHA treatment 137
Figure 5.2: Expression coverage map of HERV elements upregulated by SAHA across all samples 138
Figure 5.3: Representative image of raw amplitude data generated with ddPCR on the QX100 ddPCR system 141
Figure 5.4: Confirmation of upregulation by ddPCR for selected LTR12 HERV elements 142
Figure 5.5: Dose response curve for select HERV elements 143
Figure 5.6: Expression coverage map of non-dysregulated HERVs 144

LIST OF TABLES

Table 1.1: Latency reversing agents (LRAs) 9
Table 2.1: A large number of DEGs appear to be responding to HIV active infection 39
Table 2.2: HIV read mapping statistics 39
Table 2.3: Top 20 DEGs correlating to HIV reads sorted by correlation 40
Table 2.4: Potential markers of HIV latency 40
Table 3.1: Overlapping upregulated genes between two primary CD4 T-cells models of HIV latency 81
Table 3.2: Genes responsive to p53 activity 82
Table 4.1: Proteins with a known role in HIV reactivation, up- or downregulated as the result of SAHA treatment 107
Table 5.1: Primers designed to confirm HERV upregulation with ddPCR 148
Table 5.2: Dysregulated HERV elements after SAHA treatment 149
Table 5.3: HERV elements chosen for confirmation 149

ACKNOWLEDGEMENTS

I would first like to thank my dissertation advisor, Dr. Douglas Richman, for his mentorship. Dr. Richman has been a superb scientific role model who I wish to emulate in the future. His advice and guidance throughout my time at UCSD has been invaluable during my graduate studies. I would also like to thank Dr. Christopher (Topher) Woelk for the time and effort he has dedicated to my studies. Topher, you've done an amazing job in balancing my dissertation work with enough freedom to explore and make it my own, but with enough supervision and oversight to keep me focused. There's nothing more reassuring and inspiring then someone that is willing to accompany you into the overwhelming depths of scientific study.

I would also like to thank the members of my dissertation committee for the time they have dedicated towards my training and examination; Dr. David M. Smith, Dr. Celsa A. Spina, Dr. Gene Yeo, and Dr. Shankar Subramaniam. In particular, I would like to also thank Dr. Spina for her fruitful collaboration on projects relating to the mixed effects of the latency reversing agent, SAHA, (included as Chapter 4) and upon the studies of latent models of HIV (included as Chapter 3).

Finally, I feel that I have received the best possible training during my time at UCSD, which was achievable due in a large part to the talented scientists I was fortunate enough to collaborate with during my studies. The following is a list of those without whom this work would not have been possible.

Dr. Nadejda Beliakova-Bethell (Nadia) from UCSD, for her collaborations in nearly every one of my projects. Nadia, your knowledge regarding biological systems and their analyses is unparalleled and I can't thank you enough for the insights you have given me throughout this work. Dr. Bastiaan Moesker from the University of Southampton, for his collaboration and creativity on my projects. Bastiaan, working

with you on these projects has been a wonderful experience. The way you approach a scientific question has complimented my own approach and encouraged me to think about these questions in an entirely different manner. Dr. Alberto Bosque and Dr. Vicente Planelles from the University of Utah. Thank you both for the knowledge you have provided me regarding cell models and for your collaboration on my projects. Steven Lada, from UCSD for his efforts pertaining to droplet digital PCR in this work. Dr. Brian Reardon for graciously providing his microarray data for an analysis upon the effects of SAHA treatment. Dr. Harvey Johnston and Dr. Spiro Garbis from the University of Southampton for their collaboration upon my work with proteomics. Finally, I would also like to thank the members of the Woelk group from Southampton that made my visit there a true delight; Michael Breen, Akul Singhania, Ashley Heinson, and Jeongmin Woo. I had a fantastic time during my visit and look forward to seeing all of you in the future. Throughout someone's life, they can never list enough people to whom they owe their thanks. Naming all of them may very well require more pages than a dissertation itself. As such, if I have missed any of you, please accept my sincerest apologies and know that you have my deepest gratitude.

Chapter 2 is adapted from **White, C.H.**, Beliakova-Bethell, N., Bosque, A., Richman, D.D., Planelles, V., Woelk, C.H., *An Unexpected Finding from an RNA-Seq Analysis of a T_{CM} Model of HIV Latency*, a collaborative unpublished work. The dissertation author was the primary author of this work and responsible for the development of the next generation sequencing pipeline, the detection and analysis of reads from HIV and including the deleted region, the quality control of the data, the application of this pipeline to sequencing data from this model, the analysis and interpretation of the transcriptome results, and writing the manuscript.

Chapter 3 is adapted from material in submission to Cell Host and Microbe, **White, C.H.**, Moesker, B., Beliakova-Bethell, N., Martins, L., Spina, C.A., Margolis, D.M., Richman, D.R., Planelles, V., Bosque, A., Woelk, C.H., *Transcriptomic analysis implicates the p53 signaling pathway in the establishment of HIV-1 latency in central memory CD4 T-cells,* (2016). The dissertation author was the primary author of this paper and was responsible for designing the study, developing the next generation sequencing pipeline to function with host and viral RNA along with ERCC negative controls, both biological and statistical analysis of the data, interpretation of the results, and writing the manuscript.

Chapter 4 is adapted from, **White, C.H.**, Johnston, H.E., Moesker, B., Manousopoulou A., Margolis D.M., Richman D.D., Spina C.A., Garbis S.D., Woelk, C.H., Beliakova-Bethell, N. (2015), *Mixed effects of suberoylanilide hydroxamic acid (SAHA) on the host transcriptome and proteome and their implications for HIV reactivation from latency.* Antiviral Res 123: 78-85. The dissertation author was the primary author of this paper, and was responsible for developing the proteomics pipeline, developing the program used for proteomics analysis (modified permutation method), the statistical analysis of all transcriptomic and proteomic data, the GO and pathway analysis with FAIME, and writing the manuscript.

Chapter 5 is adapted from material in preparation for submission, **White, C.H.**, Beliakova-Bethell, N., Lada, S., Richman, D.D., Woelk, C.H., *Dysregulation of human endogenous retroviruses in primary CD4 T-Cells upon treatment with SAHA*. The dissertation author was the primary author of this paper and responsible for designing the study, the development of the next generation sequencing pipeline to study HERVs, the analysis of transcriptomic and HERV data, interpreting the results, and writing the manuscript.

VITA

2005	Bachelor of Science in Physics, California Institute of Technology, Pasadena, California
2006-2009	Research Technician, Department of Cell and Molecular Genetics, House Ear Institute, Los Angeles, California
2009-2011	Research Associate. V., Department of Cell and Molecular Genetics, House Ear Institute, Los Angeles, California
2016	Doctor of Philosophy in Bioinformatics and Systems Biology, University of California, San Diego, California

PUBLICATIONS

White C.H., Ohmen J.D., Sheth S., Zebboudj A.F., McHugh R.K., Hoffman L.F., Lusis A.J., Davis R.C., Friedman R.A., *Genome-wide screening for genetic loci associated with noise-induced hearing loss.* Mamm Genome. 20:207-213. (NIHMS 153403). PMID:19337678. (2009).

Perez-Santiago J., Diaz-Alarcia R., Callado L.F., Zhang J.X., Chana G., **White C.H.**, Glatt S.J., Tsuang M.T., Everall I.P., Meana J.J., Woelk C.H., *A combined analysis of microarray gene expression studies of the human prefrontal cortex identifies genes implicated in schizophrenia.* J Psychiatr Res. Nov; 46(11):1464-74. PMID:22954356. (2012).

Tanaka K., Eskin A., Chareyre F., Jessen W.J., Manent J., Niwa-Kawakita M., Chen R., **White C.H.**, Vitte J., Jaffer Z.M., Nelson S.F., Rubenstein A.E., Giovannini M., *Therapeutic potential of HSP90 inhibition for neurofibromatosis Type 2.* Clin Cancer Res. July 15; 19(14):3856-70. PMID:23714726. (2013).

Ohmen J.D., **White C.H.**, Li X., Wang J., Fisher L.M., Derebery M.J., Friedman R.,A., *Genetic evidence for ethnic diversity in susceptibility to meniere's disease.* OtolNeurotolo. Sep;34(7): 1336-41. PMID:23598705. (2013).

Massanella M., Sighania A., Beliakova-Bethell N., Pier R., Lada S.M., **White C.H.**, Perez-Santiago J., Blacno J., Richman D.D., Little S.J., Woelk C.H., *Differential gene expression in HIV-infected individuals following ART.* Antiviral Res. Nov; 100(2):420-8. PMID:23933117. (2013).

Beliakova-Bethell N., Massanella M., **White C.H.**, Lada S.M., Du P., Vaida F., Blanco J., Spina C.A., Woelk C.H., *The effect of cell subset isolation method on gene expression in leukocytes.* Cytometry A. Jan;85(1):94-104. PMID:24115734. (2014).

Bonczkowski P., De Spiegelaere W., Bosque A., **White C.H.**, Van Nuffel A., Kiselinova M., Trypsteen W., Witkowski W., Vermeire J., Verhasselt B., Martins L., Woelk C.H., Planelles V., Vandkerckhove L., *Replication competent virus as an*

important source of bias in HIV latency models utilizing single round viral constructs. Retrovirology. Aug 21;11(1):70. PMID:25142072. (2014).

Fransen E., Bonneux S., Corneveaux J.J., Schrauwen I., Di Berardino F., **White C.H.**, Ohmen J.D., Van de Heyning P., Ambrosetti U., Huentelman M.J., Van Camp G., Friedman R.A., *Genome-wide association analysis demonstrates the highly polygenic character of age-related hearing impairment.* Eur J Hum Genet. Jan;23(1):110-5. PMID:24939585. (2015).

Schrauwen I., Hasin-Brumshtein Y., Corneveaux J.J., Ohmen J., **White C.H.**, Allen A.N., Lusis A.J., Van Camp G., Huentelman M.J., Friedman R.A., *A comprehensive catalogue of the coding and non-coding transcripts of the human inner ear.* Hearing Research Mar; 333:266-74. PMID:26341477. (2015).

White C.H., Johnston H.E., Moesker B., Manousopoulou A., Margolis D.M., Richman D.D., Spina C.A., Garbis S.D., Woelk C.H., Beliakova-Bethell N., *Mixed effects of suberoylanilide hydroxamic acid (SAHA) on the host transcriptome and proteome and their implications for HIV reactivation from latency.* Antiviral Res., Sep 4; 123:78-85. PMID:26343910. (2015).

Myint H., **White C.H.**, Ohmen J.D., Li X., Wang J., Lavinsky J., Salehi P., Crow A.L., Ohyama T., Friedman R.A., *Large-scale phenotyping of noise-induced hearing loss in 100 strains of mice.* Hearing Research Feb; 332:113-20. PMID:26706709. (2016).

White C.H., Moesker B., Ciuffi A., Beliakova-Bethell N., *Systems biology applications to study mechanisms of HIV latency and reactivation.* World J Clin Infect Dis; In press. (2016).

Breen, M., **White C.H.**, Shekhtman, T., Lin, K., Looney, D., Woelk, C.H., Kelsoe, J., *Lithium responsive genes and gene networks in bipolar disorder patient derived lymphoblastoid cell lines*. The Pharmacogenomics Journal. In press. (2016)

White, C.H., Moesker, B., Beliakova-Bethell, N., Martins, L., Spina, C.A., Margolis, D.M., Richman, D.R., Planelles, V., Bosque, A., Woelk, C.H., *Transcriptomic analysis implicates the p53 signaling pathway in the establishment of HIV-1 latency in central memory CD4 T-cells.* In submission to PLOS Pathogens. (2016).

FIELDS OF STUDY

Major Field: Bioinformatics and Systems Biology

Professors: Dr. Douglas D. Richman and Dr. Christopher H. Woelk

ABSTRACT OF THE DISSERTATION

Investigations of HIV Latency Through Transcriptomic and Proteomic Profiling

by

Cory Haley White

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California, San Diego 2016

Professor Douglas D. Richman, Chair

The human immunodeficiency virus (HIV) has been a global threat for over three decades despite treatments, preventative measures, and public awareness. HIV hides from the immune system by entering a latent state within CD4 T-cells and removal of therapy results in a resurgence of infection. Therapies to combat HIV were designed to shock the virus from latency and kill infected cells, i.e. "Shock and Kill", and are often studied in cell models due to the paucity of infected cells and no method to isolate them from HIV infected individuals. Transcriptomic and proteomic methods have proven key to studying HIV latency and demonstrate methods by which the "Shock and Kill" strategy may be improved for HIV eradication.

First, two RNA-Seq studies of models of HIV latency are presented (**Chapters 2** and **3**). Through host transcriptome analysis in the "Bosque and Planelles T_{CM} model of HIV latency", it was demonstrated that host gene expression was reflective of the response to active instead of latent infection. It was shown that fully infectious virus was reconstituted through recombination between the deficient HIV construct and pLET-LAI plasmid. Analysis of host and virus transcripts in the refined "Martins *et al*." model, demonstrated it represented a form of latency. Furthermore, genes relating to p53 signaling were dysregulated and inhibition of p53 resulted in reduction of the total percentage of latently infected cells.

Next, the effect of the histone deacetylase inhibitor SAHA, a latency reversing agent, was analyzed (**Chapters 4** and **5**). This revealed SAHA dysregulated genes in a way that was counterproductive to HIV reactivation. Furthermore, analysis of HERV elements dysregulated by SAHA demonstrated elements from multiple families were upregulated with a specificity for those from LTR12.

In summary, this work demonstrated the importance of transcriptomic and proteomic analysis to the study of HIV latency models, led to a revision of the first model, and the importance of the p53 signaling in latency for the second. This work also demonstrated the effect of SAHA treatment on CD4 T-cells, which may potentially explain why SAHA has proven somewhat ineffective in clinical trials to reactivate HIV from latency.

Chapter 1

An Introduction to HIV Latency and the Methods to Study this Disease

1.1 HIV - A Global Threat

Acquired immune deficiency syndrome (AIDS) is a spectrum of conditions resulting from infection with the human immunodeficiency virus (HIV) [1]. Cases of AIDS have been documented since before the 1970s, and entry of the virus into the human population is believed to have occurred as early as the 1920s [2]. Entry is thought to have happened via zoonotic transfer from chimpanzees in Africa through consumption of primate meat infected with the simian immunodeficiency virus (SIV) [3]. However, the modern epidemic was officially documented on June 5th, 1981 [4] when five patients were treated for *Pneumocystis carinii*, a rare form of pneumonia, and an opportunistic infection that targets individuals with a weakened immune system [5]. Since then, HIV has spread across the world to every habitable continent and current estimates put the total number of infected individuals at 36.9 million while 25.3 million have died due to AIDS related illnesses [6]. While significant progress has been made regarding both prevention and treatment of this disease, an

estimated 2.1 million people worldwide were infected in 2014 [4]. In addition, HIV places a considerable financial burden on both the individual and the healthcare system of a country as HIV infected people require daily therapy with expensive drug regimens. The per person cost of treating HIV can reach upwards of $23,000 per year [7]. Despite over 30 years of research, anti-retroviral therapy (ART), and public awareness policies, over 50% of infected individuals may be unaware of their condition, and only 41% have access to treatment [6]. Due to these factors, HIV/AIDS remains a global health threat and will likely continue to remain so well into the 21st century.

1.2 HIV Latency - The Sleeping Killer

A major hindrance to a cure for HIV is the persistence of the latent viral reservoir [8], primarily composed of CD4 T-cells with long half-lives that may persist throughout the lifetime of the individual [9, 10]. The composition of this reservoir is an active area of study, and other cell types such as macrophages, monocytes, and astrocytes have been suggested as contributing to the persistent reservoir [11, 12]. Even stem cells and progenitor cells have been suggested as potential reservoirs for latent HIV [13]. There is even disagreement about what constitutes latency for these long lived CD4 T-cells [14]. One view is that these cells are completely dormant and quiescent [15] while another view is that persistent infection is facilitated by low levels of viral replication that continues despite ART [16, 17].

While ART has been extremely effective at suppressing HIV replication in cells with integrated viral DNA [18], small populations of these cells remain relatively unaffected [19]. It has also been shown that interruption of treatment for HIV infected individuals with no detectable viral load for years after ART initiation results in a high

viral load in as little as two weeks [13, 20, 21]. Even people on ART who have suppressed the virus for decades still have detectable viral DNA, which results in rebound if treatment is interrupted [22]. This demonstrates that the latently infected pool of cells can re-infect other healthy cells in the absence of ART, causing a resurgence of the disease. This resurgence of HIV would need to be halted during ART interruption to be considered a functional cure, while a complete elimination of this viral reservoir would be required for a sterilizing cure.

Why would an HIV cure be beneficial? While ART treatment has resulted in a significant reduction in both morbidity and mortality for those infected with HIV [18], they must remain on this treatment for their entire lives. The virus within the HIV infected individual may develop resistance to treatment [23], or the person may develop co-morbidities such as chronic inflammation or low CD4 T-cell count, which increases susceptibility to opportunist infections [24]. In addition, HIV infected individuals are often subject to the early onset of age related symptoms such as cardiovascular disease, obesity [25], loss of muscle mass [26], bone density loss [27], and cognitive impairment [28]. The cause of these symptoms is currently not well understood, but could be a result of the insult to the immune system during acute or chronic infection [29], or may be associated with lifelong exposure to ART [30]. For all of these reasons, eradication of HIV is highly desirable, but will not be achieved solely through suppression of viral replication using ART [31, 32].

1.3 In Vitro Cell Models to Study HIV Latency

The study of HIV latency is complicated by the genetic heterogeneity of both the host and virus [33], along with the paucity of latently infected cells (1 in 10^6) in an HIV infected individual on ART [19]. The study of HIV latency *in vivo* has been further

hindered by the lack of any phenotypic markers to distinguish latently infected from uninfected cells [34]. To address these issues, a number of *in vitro* models of HIV latency have been developed [34]. Early models utilized immortalized cell lines [35], but cell line models can develop host gene mutations during growth, and are often limited by the clonal nature of the initial viral integration site [34]. Furthermore, cell line models often have a defect which allows them to be immortalized [36], hence they are not a wild type cell, and may not be the best *in vivo* representation of cells infected with HIV. More recently, several laboratories have developed primary CD4 T-cell models which may better reflect *in vivo* latency [8]. These primary cell models provide investigators with the ability to study mechanisms central to HIV latency and the effect of drug compounds upon latently infected cells. Finally, primary cell models have the crucial benefit of generating a much higher number of latently infected cells (5-10% latently infected cells) compared to the number that can be isolated from an HIV infected individual (1 in 10^6) for subsequent analyses. Of the potential primary cell models available, the Bosque and Planelles model of HIV latency [37] was initially chosen for studies presented in this dissertation (**Chapter 2**). This model is a primary central memory T-cell (T_{CM}) model of HIV latency that utilizes a single round reporter virus deficient in the envelope (*env*) gene. The advantages of this model were the purportedly high proportion of latently infected cells (80-90%), a virus that resembled wild type HIV_{NL4-3}, and a single round of replication that does not generate productively infected cells.

1.4 Elimination of the Latent HIV Reservoir

Numerous methods have been proposed to eradicate or inactivate the latent HIV reservoir (ART intensification, Zinc-finger nucleases [ZFNs], homing

endonucleases, "Shock and Kill", among others). Several of these strategies, often evaluated in model systems of HIV latency, have had only partial success. Early studies examined the intensification of ART with the assumption that latency is maintained through low levels of HIV replication in the presence of drug therapy [38]. However, intensification of ART has not resulted in a significant reduction in the viral reservoir [31].

Another set of methods for the elimination of the viral reservoir is based upon gene editing strategies. For example, ZFNs have been proposed as a method by which integrated viral DNA may be targeted and deleted [39]. These molecules may direct DNA cleavage to specific sequences and have been utilized in gene editing techniques previously to control HIV *in vivo* in mouse models [40]. One major difficulty with these techniques is the heterogeneity of clinically relevant HIV strains [33]. As an individual progresses with HIV infection, this heterogeneity may hinder the ability of ZFNs to effectively target all strains present after infection. Another potential cure method involving ZFNs, is *CCR5* gene editing [41]. In this method, ZFNs are used to disrupt the *CCR5* gene sequences such that the cells become resistant to HIV infection [42]. This method was tested on a small cohort of 12 HIV infected individuals by transfusion of CD4 T-cells that had the *CCR5* receptor rendered dysfunctional. No adverse reactions were reported relating to the modified cells. Additionally, six of these patients underwent a 12 week treatment interruption, and for the four that completed this, viral load decreased from peak levels after ART removal. The additional two patients reinitiated ART due to high viral load and were unable to complete the treatment interruption study. As these methods are still in their infancy, long-term effects of these gene editing techniques, and their effectiveness at clearing the latent reservoir are still unknown. In addition to ZFNs,

engineered homing endonucleases have been proposed by Dr. Keith Jerome and colleagues, as a functional cure by which proviral DNA may be targeted and altered to delete enough integrated viral DNA to render the virus incapable of replication and subsequent infection of neighboring cells should viral activation occur [43].

While the previous methods mentioned show promise for the elimination of this latently infected reservoir, the primary strategy for the elimination or significant reduction of reservoir size is to shock HIV out of latency and then kill the infected cell (i.e., "Shock and Kill" strategy) [44] (**Figure 1.1**). In this strategy a latency reversing agent (LRA) is used to provide the "Shock" and activate HIV replication. Continued ART is used to block the subsequent infection of neighboring cells. The cell may then die on its own, or the immune system, combined with a potential immune system primer (e.g. therapeutic vaccine [45], Bcl2 inhibitors [46], or LRAs such as ALT-803 that potentially synergize with CD8 T-cells [47]), is utilized to "Kill" the infected cell, thereby clearing the latently infected pool from the individual.

For the "Shock" portion, a major drawback is that the LRA must be successful at affecting a large portion of cells in the latent reservoir for a functional cure. Estimates indicate that a sustained 1,000 to 10,000-fold reduction in the size of this reservoir may be required to prevent a resurgence of HIV during treatment interruption [48]. Therefore, if treatments with LRAs fail to activate a part of the viral reservoir that is capable of generating fully functional virus, then the HIV infected individual will not be cured.

Several classes of compounds have been utilized to "Shock" HIV out of latency including; cytokines [49], bromodomain inhibitors [50], disulfram [51], protein kinase C (PKC) agonists [52], histone methyltransferase inhibitors (HMTIs) [53], and histone deacetylase inhibitors (HDACis) [54] (**Table 1.1**). HDACis have been

extensively studied, well characterized, and approved as anti-cancer (suberoylanilide hydroxamic acid, SAHA [55] and Romidepsin, RMD [56]), or anti-epilepsy drugs (valproic acid) [57].

Unfortunately, studies have not reported a significant post-treatment reduction in the latent HIV reservoir [44, 58, 59]. The multiplicity of molecular mechanisms involved in latency control suggests that a combination approach will likely be required to achieve a degree of reactivation required for recognition of the infected cell by the immune system [60, 61, 62]. Indeed, some of the LRAs tested in combination strategies demonstrated synergy for HIV reactivation [63, 64, 65, 66, 67]. Given the many options for LRAs and combination treatments, the most well studied HDACi, SAHA [68, 69, 70, 71], was chosen for studies upon the effect of HDACi treatment on the host transcriptome (**Chapters 4 and 5**).

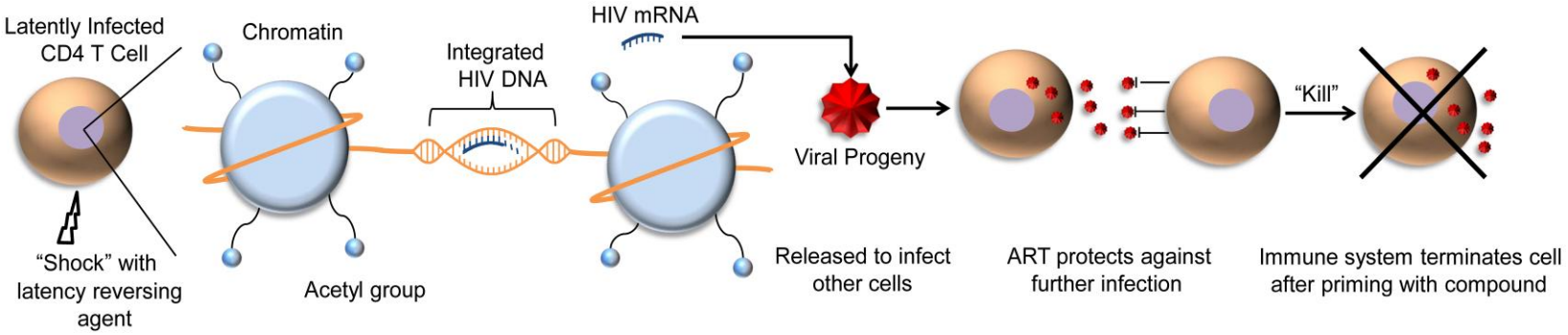

Figure 1.1: The "Shock and Kill" strategy to facilitate a cure for HIV. Primary strategy for the elimination of the latently infected reservoir. A compound is utilized to "Shock" the HIV virus into replication from a latently infected cell, which would allow for integrated HIV DNA to produce mRNA to be incorporated into virions. These would then attempt to infect other cells, but ART prevents this. The infected cell is then recognized by the immune system and killed, possibly with the aid of an adjuvant therapy.

Table 1.1: Latency reversing agents (LRAs). Multiple latency reversing agents have been utilized to shock a cell out of latency with varying levels of success. Below is a list of compounds proposed or utilized along with the method by which they assist in HIV reactivation from latency.

Type	Compounds	Citation	Function
Cytokines	Combination IL-2, IL-6, and TNF-α	Chun *et al.* 1998 [49]	Cytokines modulate levels of HIV expression by activating CD4 T-cells latently infected with HIV. This leads to an increase in transcription of the cell and subsequent activation of HIV.
Bromodomain inhibitors	JQ1	Li *et al*. 2013 [72]	Brd4 competitively binds to p-TEFb thereby reducing *Tat* binding to p-TEFb. Inhibiting Brd4 binding results in an increase of HIV transcription.
	I-BET	Boehm *et al*. 2013 [50]	
	I-BET151	Boehm *et al*. 2013 [50]	
	MS417	Boehm *et al*. 2013 [50]	
	OTX015	Lu *et al*. 2016 [73]	
p-TEFb releasing agents	HMBA	Contreras *et al*. 2007 [74]	These agents activate the Akt pathway leading to phosphorylation of HEXIM1 and release of p-TEFB, which is recruited to the HIV promoter to stimulate HIV transcription.
	Disulfram	Doyon *et al*. 2013 [51]	
PKC agonists	Prostratin	Williams *et al*. 2004 [52]	PKCs phosphorylate IκBs which then break down to release NF-κB, a protein essential for HIV replication.
	Bryostatin-1	Dias *et al*. 2015 [75]	
	Ingenol	Jiang *et al*. 2014 [76]	
	phorbol-13-stearate	Marquez *et al. 2008* [77]	
HMTIs	Chaetocin	Bernhard *et al*. 2011 [78]	Inhibits histone methylation that represses transcription by altering the affinity of transcription factors for their binding sites or by hindering the binding of basal transcriptional machinery (Pol II).
	BIX-01294	Bouchat *et al*. 2012 [53]	
HDACis	SAHA	Archin *et al*. 2009 [54]	Inhibits histone de-acetylases that remove an acetyl group from histones. This allows chromatin to remain in a relaxed and transcriptionally favorable state.
	Panobinostat	Palmer *et al*. 2014 [79]	
	Romidepsin	Wei *et al*. 2014 [80]	
	Entinostat	Wightman *et al*. 2013 [81]	
	Givinostat	Banga *et al*. 2015 [82]	
	Belinostat	Rasmussen *et al*. 2013 [83]	
	Valproic Acid	Routy *et al*. 2012 [84]	

For the "Kill" portion of this strategy, it was initially believed that the immune system could clear the actively infected cell after exposure to a "Shock" compound. Unfortunately, this proved difficult as the immune system may be unable to perform this function alone [85], and instead require a priming second step to clear the virus producing cells [86]. One approach to solve this problem is to utilize natural killer (NK) cells. Through the use of antibodies raised against the envelope of the infecting virus strain, NK cells may be primed to recognize CD4 T-cells replicating HIV once the virus is shocked out of latency [87]. A second method, which may not be solely dependent upon the person's immune system, is to utilize a protein that targets cells actively replicating virus to assist in their destruction [88, 89]. Specifically, dual-affinity re-targeting (DART) proteins can direct cytolysis of latently infected HIV cells. Both of these methods, when paired with a latency reversing agent, have the potential to clear the latent HIV reservoir.

The success of the "Shock and Kill" treatment strategy may depend upon multiple rounds of treatment with activating or killing compounds in order to destroy all of the latently infected cells with functional integrated provirus (sterilizing cure), or destroy a sufficient number of them to prevent viral rebound in the lifetime of the individual (functional cure). If activation misses even a small fraction of these cells with functional integrated provirus, which may be the case in clinical trials with SAHA [58, 90], a resurgence of HIV when ART is discontinued will likely follow. It may also require a combination of compounds (e.g. HDACis such as SAHA, Romidepsin, Panobinostat, mixed with Bromodomain inhibitors, or combining multiple LRAs) as a single compound may not reach adequate levels of HIV reactivation. Indeed, treatment with SAHA alone dysregulates many genes that are counterproductive for

HIV activation [91] (**Chapter 4**), but utilizing this drug in a combination strategy may be more successful for reactivating HIV.

1.5 Transcriptomic Methods for the Study of HIV Latency and Latency Reversing Agents

Over the past decade, the study of viral infection and HIV latency have benefited greatly from the advent of novel technologies [92, 93]. Transcriptomic methods in particular are capable of providing a wealth of information on models of HIV latency and pathogenesis, not only on host gene modulation, but also on changes in transcripts produced by HIV. Two primary methods for high throughput transcriptomic analysis are microarrays and RNA-Seq [94]. While the majority of studies published involving HIV latency research have utilized microarray technologies, RNA-Seq technology has recently become more common due to the improved detection of rare variants and ability to analyze reads from the HIV virus [95, 96, 97].

Microarrays are an early transcriptomic technology used to study the expression of a large number of genes in a single experiment [98]. In this technology, probes are arrayed on a solid support and used to assay mRNA present in a sample through hybridization of complementary strands [99]. To quantify expression differences, either a less expensive and more flexible single color, or the more expensive yet potentially more robust (due to the internal reference) dual color microarray is used. In the dual colored microarray, equal amounts of cDNA are generated from each condition through reverse transcription of mRNA and labeled with the Cy5 and Cy3 fluorophores which fluoresce at red and green wavelengths

respectively [100]. By measuring the relative intensity of each probe on the plate, one obtains a measure of expression for all genes on the microarray. In the single color microarray, either Cy5 or Cy3 is used alone, and the absolute intensity is utilized to quantify the hybridization level with the labeled target [101]. The single color microarray, HumanHT-12 v3 Expression BeadChip from Illumina contains over 48,000 probes for over 25,000 annotated genes and is an ideal low cost microarray for transcriptomic studies (**Chapter 4**).

Techniques to analyze microarrays have been well established to facilitate straightforward and expedient analysis of data. In addition, they have been shown to be highly accurate regarding differential gene expression for abundant transcripts [102, 103]. However, this technology has some limitations. Microarrays rely on specific oligonucleotide sequences present, which are specific to the set of genes on the platform. Should new genes be discovered, they may not be found in microarray data. Additionally, while the existence and direction of expression changes are reliable with microarrays, the magnitude of change may not be reliable [104], potentially due to probe saturation.

RNA-Seq is a newer technology to study the transcriptome, often referred to as “next generation” in comparison to microarrays [105]. In this technology, RNA is isolated from a sample and abundant ribosomal RNAs are depleted (Total RNA-Seq) to preserve the non-coding RNA fraction or polyadenylated RNAs are enriched (mRNA-Seq) prior to random fragmentation for sequencing. These fragments are then ligated to adaptors and run across a flow cell which randomly binds these single strand fragments. Then nucleotides and polymerase are added to generate the complimentary strand and double stranded bridges which bend over and reattach to the flow cell. A denaturizing compound is then used to break the double strand into

two single strands, again doubling the number of single stranded fragments. This process is repeated multiple times to create clusters of identical singled stranded DNA, which are then sequenced using the sequence by synthesis method [106]. One last double strand is assembled with labeled nucleotides that contain a fluorescent probe which is measured by the sequencer upon synthesis of the complimentary strand. These nucleotides are added one base at a time and fluoresce if used by polymerase to create double stranded DNA, which is measured by the sequencer and used to generate a sequence for each cluster. For this work, the HiSeq2000 instrument from Illumina was chosen.

RNA-Seq technologies and the methods for their analysis have improved significantly in the past decade and reduced in cost to the point where they are a reasonable, if not superior, alternative to microarrays [94]. This technology also has several advantages over microarrays, such as an improved ability to detect low abundance transcripts (e.g. rare variants) [103]. Furthermore, unlike microarrays, RNA-Seq does not require species specific probes, and can detect novel transcripts that are not present in current microarrays. Additionally, the lack of species specific probes in RNA-Seq allows for the simultaneous detection of both host and viral genes without having to create a custom microarray. This attribute is particularly important for work with HIV infected individuals or latently infected cell models as the detection of HIV transcripts can reveal and explain many exciting details that may be missed with only a microarray experiment (**Chapters 2 and 3**).

For experiments involving primary CD4 T-cell HIV latency models (**Chapters 2 and 3**), viral transcript detection was thought to be important to both validate the latent model, as well as determine if HIV produced low levels of reads in the latent state. For this reason, RNA-Seq was chosen for transcriptomic analysis.

Experiments involving human endogenous retroviral (HERV) elements (**Chapter 5**) also required the use of RNA-Seq, as the full complement of HERV elements in databases used are not present on a microarray platform. For RNA-Seq experiments, Tophat [107] was the mapping program of choice along with HTSeq [108] for counting, and EdgeR [109] for differential expression analysis. Tophat was chosen due to the ability of the program to work with both a transcriptomic dataset, as well as use a built in algorithm to map transcript data back to a human genome reference. EdgeR was chosen due to the ability of this program to work with paired data as experiments with HIV latency are compared to an uninfected sample derived from the same blood donor. For the purposes of experiments involving the study of HDACis upon host gene expression (**Chapter 4**), the Human HT-12 v3 Expression BeadChip, a microarray available from Illumina, was chosen due to the high number of probes (over 48,000) covering the RefSeq [110] and UniGene [111] annotated genes, the low sample input requirement, and the low cost of the chip [112].

1.6 Proteomics Methods for the Study of HIV Latency Reversing Agents

Proteomics profiling has revealed an incredible amount of information about the mechanisms of disease and response to drugs [91, 113, 114]. This analysis method is more downstream when compared to transcriptomics, and measures the ultimate effector molecule, the protein. While it is not currently possible to detect every protein, proteomics profiling technologies have advanced greatly in recent years and are now able to detect several thousand proteins [91, 115, 116]. Combining proteomic and transcriptomic approaches complement one another and

provide a wide lens with which to interrogate and interpret biological phenomena such as the off-target effects of latency reversing agents (**Chapter 4**).

Early in the history of proteomics research, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) was used whereby differentially expressed spots were excised from a gel and identified by mass spectrometry. However, this method had issues with reproducibility because of difficulty in the detection of low abundant proteins [117]. The 2D-PAGE method required a large amount of time to excise gels for individual proteins, and was therefore limited in the number of proteins that could be quantified in a reasonable amount of time. To increase the number of proteins that may be quantified quickly, additional methods were developed. Isotope Coded Affinity Tags (ICAT) utilizes both a light and heavy isotope in separate samples which are then combined for analysis after which the ratio of signal intensity is used to determine relative abundances [118]. However, ICAT relies on iodoacetamide to modify cysteine residues present on the peptide and therefore proteins which lack cysteine residues remain hidden [115]. A third method known as stable isotope labeling by amino acids in cell culture (SILAC) involves the incorporation of a non-radioactive isotopically labeled amino acid in place of the original amino acid [119]. Cells are grown in a medium with this modified labeled amino acid which is utilized for quantification of proteins. However, this requires the development of cell lines that incorporate this modified amino acid, hence this method is not applicable to the study of primary CD4 T-cells collected from human subjects.

Two recent methods that have become central to quantitative proteomics research are the label-based approach, isobaric tags for relative absolute quantification (iTRAQ) [120], and label-free approaches [121]. In iTRAQ, the isobaric tag reacts with the primary amines of peptides. Each tag contains a separate

reporter ion with a distinct mass over charge ratio. Samples are pooled together and then analyzed with liquid chromatography tandem mass spectrometry (LC-MS/MS) to determine the mass to charge ratio of the ions present. The attached tag is then used to determine the relative quantification of the peptide and corresponding protein. Peptide and/or protein identification can then be determined by comparing the spectrograph to a known database of proteins such as Uniprot [122]. In label-free approaches, quantification is based not upon stable isotopes, but rather upon spectral counting (the total number of spectra identified for a protein) and/or intensity-based measurements (peptide peak area or peak height) from a precursor signal [116, 123, 124]. In this technology, samples are individually processed and subjected to LC-MS/MS analysis.

These newer technologies have their benefits and drawbacks. Both may be implemented after cell lysis and do not require the input of a stable isotope during cell culture making them ideal for the study of HIV latency when not working with cell lines. Both methods are also not limited to proteins with specific amino acid residues or those that are amenable to gel electrophoresis. Therefore, these methods have access to a much larger portion of the proteome, including post-translational modifications, compared with older methods. However, differences in the method of quantification (labeled tag quantification or tag-free quantification) must be taken into account during study design. Label-free approaches do not require the isobaric tags and may seem to be less expensive, as their quantification may be done by observing spectra. However, label free methods may not be multiplexed, whereas iTRAQ may be multiplexed up to 8-10 samples [125]. This can reduce the total run time and potential cost for the experiment, and removes possible batch effects that may be present in label free approaches where each sample must be processed separately.

Some studies report that label free methods have increased accuracy in relative protein abundances in complex mixtures of proteins [125, 126]. However, iTRAQ appears to have access to a larger segment of the proteome, particularly when labeled peptides are pre-fractionated [125].

For studies involving treatment with the HDACi SAHA (**Chapter 4**), a large number of genes were previously shown to be dysregulated by SAHA [127]. Hence a high coverage of the proteome was required. Additionally, to further increase the confidence in proteomic results, 8-paired samples were utilized for this analysis. For these reasons, iTRAQ was chosen over other methods as a high throughput technique for protein quantification, for its higher coverage, and for its multiplexing capabilities to avoid batch effects.

1.7 Scope of this Dissertation

This dissertation represents 4 years of research. During this time, the first iteration of the Bosque and Planelles primary cell model of HIV latency [37] was characterized and analyzed (**Chapter 2**). This work led to a revision of this model, the Martins *et al*. model [128] which was also characterized and analyzed (**Chapter 3**). Analysis of this new model identified the p53 pathway to have a role in the establishment of latency. Then, the effect of SAHA upon CD4 T-cells in the "Shock and Kill" treatment was analyzed. It was determined that several off-target effects of SAHA may be inhibitory for HIV reactivation out of latency (**Chapter 4**). Finally, the effect of SAHA on the expression of HERV elements was examined, and it was found that several hundred HERV elements, many of which belonged to the LTR12 HERV family, were upregulated in the human genome (**Chapter 5**).

Chapter 2

An Unexpected Finding from an RNA-Seq Analysis of a T_{CM} Model of HIV Latency

2.1 Abstract

The primary central memory T-cell (T_{CM}) model developed in the laboratory of Dr. Vicente Planelles, henceforth referred to as the Bosque and Planelles model, utilizes a single round reporter virus. This virus is based upon an engineered viral vector that contains a 580 bp deletion in the envelope (*env*) gene, and is thought to be incapable of replication after the initial round of infection. An RNA-Seq analysis pipeline was developed to study this model and demonstrated that despite the deficient *env* gene, this virus was capable of replication due to recombination with the pLET-LAI plasmid containing the *env* sequence provided during infection. This is evident from the presence of reads from a total RNA-Seq experiment that mapped to the region of the HIV genome that was believed to be deleted. In addition, the transcriptional profile of the latency model, as compared to uninfected paired samples, was reflective of a productively infected state. Finally, a potential method to separate transcriptomic results of latent infection from active infection is presented.

Due to these results, caution is urged in the future use of models which utilize deficient HIV constructs to seed the latent reservoir. It is further recommended that future researchers implement RNA-Seq in their analysis of models of HIV latency due the ability of this technology to detect low abundant transcripts such as HIV RNA that may be present in the model, and to confirm that both host and virus gene expression are truly reflective of a latent state.

2.2 Introduction

2.2.1 The latent HIV model

The recent interest in a cure for HIV latency led to the development of cell models recapitulating viral latency *in vitro* [34]. The T_{CM} model, published by Bosque and Planelles [37], is a widely used model that is based upon *in vitro* differentiated T_{CM} cells. This model employs a replication deficient HIV (DHIV) virus with a 580bp deleted region in the *env* gene. The *env* protein is then provided during infection *in trans* from a plasmid (pLET-LAI), to allow for integration, viral gene expression, and viral protein expression. This deletion of the *env* gene in DHIV was previously thought to make this virus incapable of replication once it had been integrated into the cell. However, when RNA-Seq analysis was applied to this model of HIV latency, it was discovered that the virus was indeed replication competent, and that previous accepted wisdom regarding this model should be re-evaluated.

A pipeline was developed for total RNA-Seq analysis and applied to transcriptomic data generated from this model. It was demonstrated that host genes within this model appeared to be reflective of a productively infected state with several interferon related genes being dysregulated. It was further shown that these genes were highly correlated to the number of reads mapping to HIV. As these

results were quite unexpected, an examination of the viral transcriptome was performed. RNA-Seq reads in the data were found to map to the region of the HIV genome that was believed to have been deleted, hence it was concluded that the virus was capable of replication, due to recombination with the pLET-LAI plasmid.

These results indicated that what was previously believed to be latently infected cells is likely a mix of latently infected and productively infected cells that were producing fully replication competent progeny. This progeny and subsequent infection was likely responsible for large proportions of cells that were erroneously considered to be latently infected and reported for this T_{CM} model [129]. These results suggested that caution should be used with this model, and that RNA-Seq should be utilized in the study of this and other models of HIV latency to better analyze expression of host and virus genes to confirm a latent state.

2.2.2 RNA-Seq analysis

RNA-Seq has become a powerful tool in transcriptomic research. It has become increasingly popular in HIV latency studies due to several advantages over microarray technologies [94] (**Chapter 1**) such as increased sensitivity to lowly expressed transcripts, a wider dynamic range [130], higher reproducibility, and the ability to detect novel splice variants [103]. An additional benefit is the ability to re-evaluate data when new genes, long non-coding RNAs, miRNAs, etc. are added to a database or discovered de novo. Finally, a particularly important benefit when evaluating models of HIV latency is the ability to detect transcripts from multiple species or genome references in the same experiment, such as transcripts from human cellular RNA and the HIV virus (**Chapters 2 and 3**) [95]. While microarray pipelines have been well tested and established, RNA-Seq pipelines are still under

development. Due to this, the field is replete with various tools for read mapping [107, 131, 132, 133], transcript quantification [108, 134], differential expression analysis [109, 135, 136, 137], and network analysis [138, 139]. However, there is not much consensus on the best mixture of these tools for RNA-Seq analysis, and they are updated frequently to implement algorithms newly discovered or modified. Additionally, tools based upon previous analysis of microarray results are often adapted to RNA-Seq datasets [140, 141]. Due to the multitude of choices available for each step, a pipeline for the analysis of RNA-Seq data was developed and applied to the study of the Bosque and Planelles T_{CM} model of HIV latency.

2.3 Materials and Methods

2.3.1 Generation of the DHIV and pLET-LAI plasmids

Generation of the DHIV plasmid has been previously described [142], and was performed by collaborators at the University of Utah. In short, a fragment between two BglII restriction endonuclease sites located at nucleotides 7032 and 7612 in the HIV-1 NL4-3 sequence was cut and the ends re-ligated. This generated a HIV-1_{NL4-3} viral vector, DHIV, containing a 580 base-pair deletion within the gp120-coding region, and rendered the downstream portion of the *env* gene out of frame.

The pLET-LAI construct (a plasmid containing the wild type *env* sequence from HIV-1_{LAI}) was generated as previously described [143]. In short, complete LTR, *Tat* and *env* genes from HIV-1 were inserted into plasmid pUC18. The SalI-XhoI fragment containing *env* was ligated downstream of BglII-NarI fragment containing the LTR sequence. The noncoding AvaI-BglII fragment from hepatitis B virus was ligated 3' to the *env* gene to provide for poly(A), splicing acceptor sequences and termination sequences.

2.3.2 Cell isolation and culture

Peripheral blood mononuclear cells (PBMCs) were isolated by collaborators at the university of Utah. Naïve CD4 T-cells were isolated by MACS microbead-negative sorting using the naïve T-cell isolation kit (Miltenyi Biotec, Auburn, CA). In vitro differentiated central memory T-cells were generated as previously described [129]. Briefly, naïve CD4 T-cells were cultured for 3 days in non-polarizing conditions, i.e. RPMI (Roswell Park Memorial Institute) medium supplemented with 1 µg/mL anti-IL-4, 2 µg/mL anti-IL-12, 10 ng/mL TGF-β (all Peprotech) and αCD3/αCD28 microbeads (Invitrogen). After three days, magnetic beads were removed with DynaMag™ Spin Magnet (Life Technologies), cells were collected, counted and resuspended in fresh RPMI with 30IU/mL IL-2 at 1 million cells in 1 mL of RPMI medium before seeding in 96 well plates. Daily medium change was performed for 4 additional days. The central memory T-cells were infected at day 7 post-isolation.

2.3.3 Viral generation and infection

Virus stocks were prepared by collaborators at the University of Utah through calcium phosphate transfection [144] of HEK293T cells according to the manufacturer's instructions (Life Technologies). Cells were infected by spinoculation. 1×10^6 cells were infected with 500ng/mL p24 during 2 hours and converted to p24 values as previously described [145].

2.3.4 Cell activation

Cell activation was performed to activate HIV and confirm a latent state in the T_{CM} model. Cells were stimulated αCD3/αCD28 microbeads (Dynabeads®

Dynabeads CD3/CD28 T-cell Expander, Dynal/Invitrogen) for 72 hours in the presence of IL-2 (1 bead per cell) before readout by flow cytometry.

2.3.5 P24 staining and flow cytometry

To quantify levels of HIV infected cells, intracellular p24 staining was performed by collaborators at the University of Utah. Cells (5×10^5) were fixed and permeabilized with Cytofix/Cytoperm during 30 minutes at 4ºC (BD Biosciences, San Diego, CA). Afterwards, cells were washed with Perm/Wash Buffer (BD Biosciences) with a 1:40 dilution of anti-p24 monoclonal antibody AG3.0 (Cat. No. 4121) obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (originally from Dr. Jonathan Allan). After 30 min of staining, the primary antibody was washed away and the cells were suspended in 100 µl PBS with 1:100 Alexa Fluor 488 goat anti-mouse IgG (H+L) antibody (Invitrogen). After staining, the cells were spinoculated and suspended in PBS after removing the antibody solution to be analyzed by flow cytometry using FACSCalibur (BD).

2.3.6 RNA-Seq data generation and analysis

Total RNA was isolated from the T_{CM} model derived cells from 4 healthy donors and deemed of sufficient quality for Total RNA-Seq analysis (RNA integrity number > 7 as assessed by Agilent Bioanalyzer 2100, Agilent, Santa Clara, CA, USA). For Total RNA-Seq analysis, the Ribozero protocol (Illumina, CA, USA) was used to deplete ribosomal RNAs. The remaining RNA was DNase treated and prepared using the Illumina TrueSeq™ RNA Sample Preparation Kit for 50 bp paired-end sequencing with the Illumina HiSeq 2000. Resulting reads were first checked for quality and then mapped to the human genome (build hg19) using TopHat [107] and

counted using HTSeq [108]. Differential expression analysis of human transcripts was performed with the program EdgeR [109] in the R statistical computing environment [146]. All reads were then mapped to the HIV genome using TopHat, counted using HTSeq, and visualized using the Integrative Genome Viewer (IGV) [147] to produce a coverage map of reads along the HIV genome. More details and the rationale for the choices regarding the RNA-Seq pipeline developed for the study of this and other models of HIV latency is given in **Section 2.4**.

2.4 The RNA-Seq Pipeline

The RNA-Seq pipeline (**Figure 2.1**) was developed with several thoughts in mind: versatility in the mapping algorithm to function with either a transcriptome index or genome reference, speed in the overall design, ease of use through the development of scripts, and the necessity of output at each step to be applicable to numerous differential expression programs and other analysis tools.

2.4.1 Quality control

The pipeline begins with the quality control of data output from the sequencer. Several quality metrics must be assessed before proceeding to the mapping stage. If such metrics are provided by the sequencing company (i.e. Expression Analysis Inc.), they may be assessed; otherwise one must perform their own quality control. FastQC [148] was chosen for quality control analysis when not provided by the sequencing company. Presented is an example of high quality reads from this experiment (**Figure 2.2**) generated from the HiSeq2000 with read statistics (average, median, minimum quality, 90th percentile, and interquartile range) well above the recommended minimum quality score of 20. Also shown is an example of low quality

reads for comparison (**Figure 2.3**). For low quality reads, trimming the read to the first 50-100 bp allows enough of the read to remain for mapping purposes and ensures accuracy of base pair calls.

2.4.2 Mapping to a genome or transcriptome reference

The choice of the mapping program is often dictated by the type of experiment performed. For example, when analyzing transcriptomic output where splicing is present, an ideal tool is Tophat [107]. When splicing is not present or not considered important, Bowtie [132] is a faster low-memory use alternative which produces similar output. Two algorithms in the Tophat program are implemented when mapping to a genome reference. The first is to create a transcriptome reference using a gene transfer file (.gtf), which contains the chromosome and positional information for each gene of interest, and then map to this reference. A second algorithm allows Tophat to map to the transcriptome for genes not present in the transcript index (**Figure 2.4**). Tophat first splits a read into 25 basepair (bp) segments (adjustable in the settings) and determines where each segment maps. If read segments map to the same chromosome, but span a pre-defined distance, a pileup island may be formed. Tophat puts these islands together to form potential splice junction where the program then maps reads to each splice junction. Junctions which have more than 10 reads mapping to them are considered true junctions, and are stored as a transcript mapping. Once mapping has concluded, reads must be sorted by either read name or read coordinates based upon the quantification method chosen. If the mapping program doesn't perform sorting, this can be done with Samtools [149].

2.4.3 Quantification of transcripts

Quantification of transcript abundances is done with the use of a gene transfer file which contains the coordinates of each gene of interest. HTSeq [108] was chosen because several downstream differential expression programs (DESeq, EdgeR, EBSeq, etc.) require count data. HTSeq was run in *union* mode which requires a read be present in only one gene (not overlapping with a second gene) to be counted towards the gene of interest. Overlapping reads in this count method are discarded.

2.4.4 Differential expression analysis

Count data output from HTSeq is used to perform differential expression testing. Many programs are available for this process such as DESeq [135], EBSeq [136], EdgeR [109], etc. which are implemented in the R statistical computing environment [146]. EdgeR was chosen for differential expression due to the ability to study a design where the groups were paired. Data is first normalized and filtered. Then, EdgeR utilizes an overdispersed Poisson model which accounts for both biological and technical variability, coupled with an empirical Bayes method to moderate the degree of overdispersion. Afterwards, the data is fit to a generalized linear model (GLM). Finally, a GLM likelihood ratio test is performed to determine if the coefficient representing the contrast between conditions of interest is equal to 0, indicating no differential expression. An additional advantage of EdgeR is that it is less conservative in calling differential expression compared to DESeq [150, 151], and therefore has a lower false negative rate.

2.4.5 Downstream applications

The final step of this pipeline is functional analyses of differentially expressed genes. These involve gene ontology (GO) analysis, heatmaps, and network analysis, among others (**Chapters 3 and 4**). These methods were not performed upon this data due to the finding that this model of HIV latency contained actively replicating virus, but were instead developed in lab for future studies.

2.4.6 Interchangeability of components

This pipeline utilizes programs that are interchangeable with others performing similar tasks, but are more applicable to the study in question or the sequencer utilized (**Figure 2.1**). For example, if one is using the Ion PGM Sequencer from Thermo Fisher Scientific, read lengths vary considerably compared to the Illumina HiSeq 2000. As such, a mapping program that incorporates a varying level of mismatches based upon read length may be more appropriate compared to utilizing the same level of mismatches for all reads. The program TMAP [152], designed for use with output from this machine, can be easily incorporated into this pipeline in place of Tophat or other mapping programs. For differential expression analysis programs that require fragments per kilomillion bases per million mapped reads (FPKMs) instead of count data, one can use the exonic lengths of each gene from the GenomicFeatures package [153] available from Bioconductor [154] to convert count data to FPKMs. If one wishes to be more strict in differential expression calls, DESeq [135] may be substituted in place of EdgeR [109]. Changes to the pipeline may be made and re-visited should another new program be deemed more suitable at a later date.

2.5 Results

2.5.1 Differential expression results

The pipeline described previously (**Chapter 2, Section 2.4**) was used to study RNA-Seq data from the T_{CM} model of HIV latency developed by Bosque and Planelles [37]. Eight samples were obtained from four donors and used to generate latently infected (LI) T_{CM} cells and paired uninfected (UI, i.e. mock-infected), controls. These cells then had RNA extracted and subjected to RNA-Seq analysis. LI and UI samples were compared to identify host genes dysregulated during latency. A total of 93 differentially expressed genes (FDR = 0.05) were identified, 45 of which were downregulated and 48 upregulated. Of note was the presence of several genes related to the interferon antiviral response, such as Interferon-Induced Protein 44-like (*IFI44l*), MX Dynamin-Like GTPase 1 (*MX1*), 2'-5'-oligoadenylate synthetase 1 (*OAS1*), Hect And RLD Domain Containing E3 Ubiquitin Protein Ligase 5 (*HERC5*), among others (**Table 2.1**).

As this interferon response was not expected in a truly latent HIV model, it was suspected that actively replicating virus was present. To test this, all reads were mapped to the full HIV genome, including the deleted sequence which encodes for the envelope protein. Reads were quantified for both the full HIV genome and the deleted region (**Table 2.2**). These results demonstrated a widely varying number of reads mapping to the HIV genome (Average 39947, s.d. +/- 53710). Further noted are reads that mapped to the region that was believed to be deleted, between 7032-7612 bp. Reads from this region varied in abundance in a similar fashion to the variation in total HIV read count. Read coverage was then determined for the full HIV genome (**Figure 2.5**). Upon closer examination of the start and end of this deleted region, overlapping reads were noted (**Figure 2.6**) . This result indicated that reads

span both sections of the recombination point, which demonstrated that reads were not being expressed from the vector alone, but rather from the integrated wild type virus which resulted from vector recombination.

2.5.2 Correlation analysis

As these differential expression results were quite unexpected, correlation analysis was utilized in the hopes of separating potential markers of HIV latency from clear signals of an active infection. First, genes with a high correlation (absolute correlation greater than 0.8) between gene expression and HIV read count were identified (**Table 2.3**), and removed leaving 52 genes for later analysis. Many of the genes removed were previously identified as ones that respond to interferon signaling (**Table 2.1**). Interestingly, the percentage of supposedly latently infected cells did not correlate with the total HIV reads produced in each sample (Pearson correlation -0.56, p-value 0.43). This suggested that correlating the gene expression results to percentage of infected cells might have been capable of identifying actual markers of latent infection once genes whose expression was correlating to HIV read level were removed. Accordingly, the remaining genes expression levels were then correlated to the total infected percentage measured. This produced five potential markers that had low HIV read correlation, but high negative correlation to HIV latency percentage (**Table 2.4**). No markers with high positive correlation were noted.

2.6 Discussion

2.6.1 Recombination of the DHIV construct with the pLET-LAI plasmid

An RNA-Seq pipeline was used to study the Bosque and Planelles T_{CM} model of HIV latency that utilizes a replication deficient virus (DHIV) [129]. These results

indicated that the defective vectors used to seed the latent reservoir in this model were recombining to form a replication competent wild-type virus (**Figures 2.5 and 2.6**). This finding raises several concerns regarding this model and its use as a primary model of HIV latency. Not only were reads detected from the deleted region in DHIV, they were detected at similar levels compared to reads across the rest of the HIV genome. This would suggest that viral recombination was not a rare event that produces a small contamination, but rather that a significant number of reads mapping to HIV were from this replication competent strain. This brings into question the measures of the total percentage of latently infected cells in this model (ranging from 80-90%, **Table 2.2**). The actual percentage of latently infected cells may be much lower than reported due to the actively replicating virus and presence of the productively infected cells. As such, other research groups that utilize this model must remain cautious of the interpretation of their results. It also indicates that researchers creating models of HIV latency that rely on multiple vectors with deletions in the HIV genome should ensure that recombination is not occurring to reconstitute wild type virus. RNA-Seq or PCR based analysis that utilizes primers and probes within or across the splice sites of a deleted regions may be used to assess whether such recombination is occurring.

In support of these findings, collaborators at the University of Ghent in Belgium demonstrated the presence of the full length *env* sequence in viral DNA [155]. Specifically, electrophoretic analysis was performed upon the DHIV plasmids demonstrating that regions not deleted amplified normally, while the deleted *env* region did not, indicating that the deletion was present. Cells were then infected with only the DHIV vector without the pLET-LAI plasmid containing the *env* sequence, which does not result in p24 expression. PCR was performed upon DNA isolated

from DHIV-only infected cells resulting in no amplification of this deleted *env* region. Only virus generated with *env*-deficient DHIV and the pLET-LAI plasmid containing the *env* sequence produced a positive single band with PCR on this deleted *env* region. This further demonstrates that a recombination event occurred between the vectors used to seed latent infection in this model (**Figure 2.7**).

2.6.1 Correlation analysis reveals potential markers of HIV latency

Despite these limitations, correlation analysis was utilized in an attempt to identify gene expression biomarkers of latency. This method could remove or order genes based upon their correlation to HIV read count, thereby reducing the noise present in this type of data. Latently infected cells are likely present in this model and their signal can potentially be found within the analysis of the data once confounding signals from actively replicating virus are removed. To this end, correlation analysis was used with the results from this model to demonstrate dysregulation of 5 genes which showed low HIV read correlation, but high negative correlation to latency percentage (**Table 2.4**). Of these genes, early growth response 1 (*EGR1*) was found previously to be dysregulated during latency in cell line models [35]. This gene encodes for a zinc finger nuclear protein that is involved in the transcriptional regulation of cell growth and is differentially expressed in response to growth factors and stress stimuli [156]. In addition, *EGR1* is downregulated in several tumor types [157] and is correlated with tumor formation [158], making it a likely tumor suppressor gene. Upregulation of *EGR1* leading to growth arrest may create an ideal scenario for viral replication, which would inhibit or reduce latency leading to an actively infected state. Therefore, the reduction of *EGR1* leading to growth increase and reduced viral replication may assist in sending the virus into a latent state (**Figure**

2.8). Interestingly, no genes with positive correlation to latent infection were identified in this method. One might initially expect markers of latency would likely be positively correlated with the latent infection percentage. However, if a particular gene is downregulated during latency, then the lower expression of this gene might result in higher latency percentage. In that case, the correlation would be highly negative as was the case with *EGR1*. Given what is now known regarding this model however, caution should be taken in the interpretation of these results. The correlation method utilized to study this model was implemented in an attempt to reduce the signal from productively infected cells. Correlating to total infected percentages as a proxy for latently infected percentages assumes that these measurements are proportional to one another. Any results from this method when applied to this model will require confirmation in other models of HIV latency, as was done in this work with *EGR1*.

Ultimately, this data led the Planelles group to revise their model of HIV latency [155]. In this new revised model, ART is utilized along with a wild type HIV-1_{NL4-3} virus to more accurately represent HIV latency *in vivo*. RNA-Seq profiling was utilized to evaluate the latent state of HIV in this revised model, and to elucidate mechanisms which are central to HIV latency establishment (**Chapter 3**).

2.7 Figures

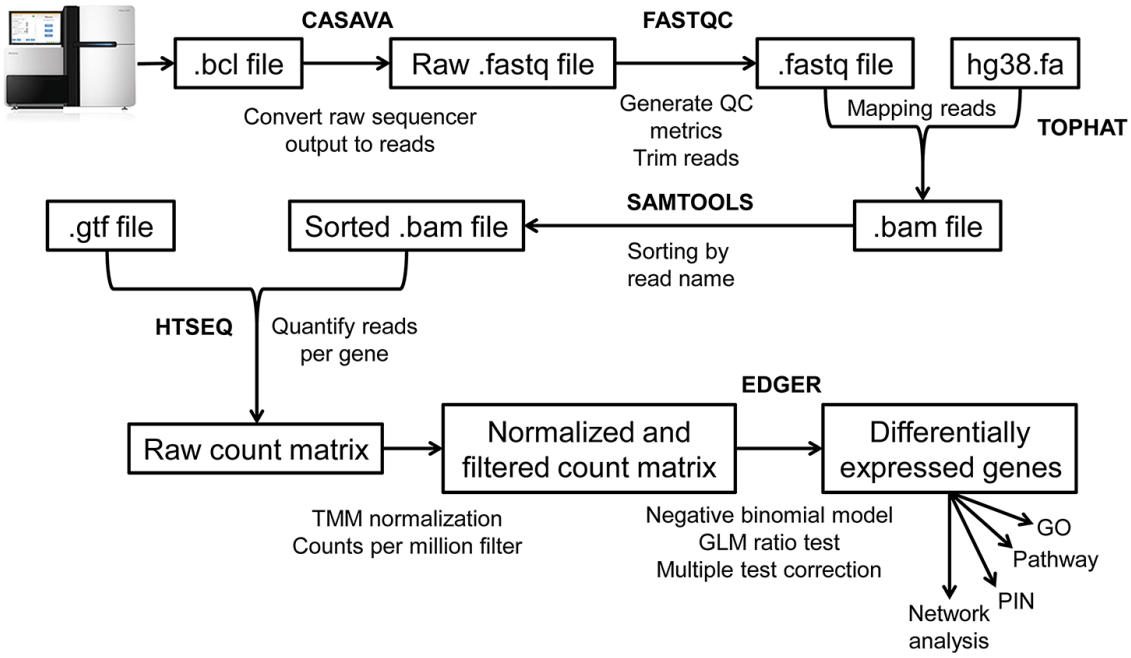

Figure 2.1: The RNA-Seq pipeline. This pipeline was generated to study transcriptome data from models of HIV latency. In bold are the names of programs that may be used to complete a particular step. In boxes are the output files produced after each step or the input file required for the next. Generation of the reads was done with the HiSeq2000 and converting raw sequencer reads was done by the company that performed the sequencing. The mapping through quantification steps were done using the Linux operating system (Ubuntu 12.10 or 14.04), while normalization and differential expression testing was done in the R statistical computing environment. For mapping and differential expression testing, the programs Tophat and EdgeR were chosen respectively. Abbreviations are defined as follows: TMM – trimmed mean of M-values normalization method, GLM – generalized linear model, GO – gene ontology, PIN – protein interaction network.

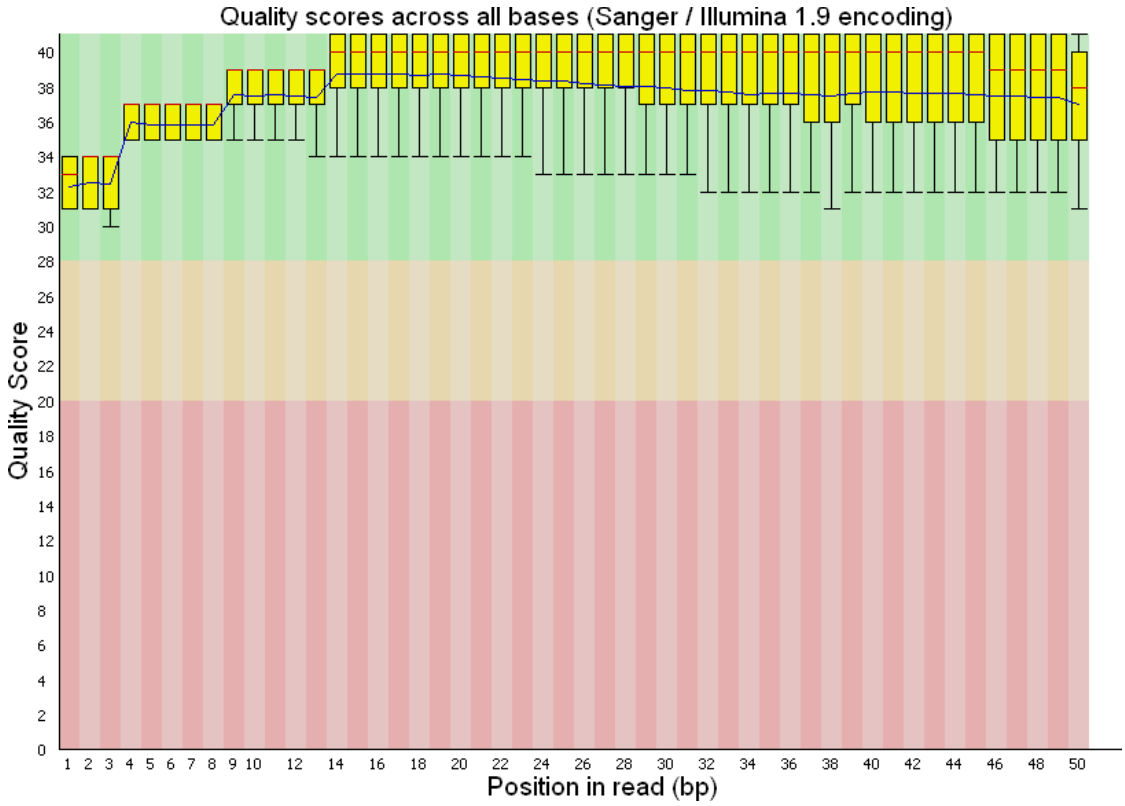

Figure 2.2: Example of high quality sequencing reads. A single sample was used to generate this image. Red bar – Median quality, Blue line – Average quality, Yellow bar – 25% and 75% quartile, Whiskers – 10% and 90% interval. Quality score is calculated as $Q = -10log_{10}(e)$ where e is the probability of an incorrect base call. Hence a quality of 10, 20, and 30 results in an accuracy of 90%, 99%, and 99.9% respectively.

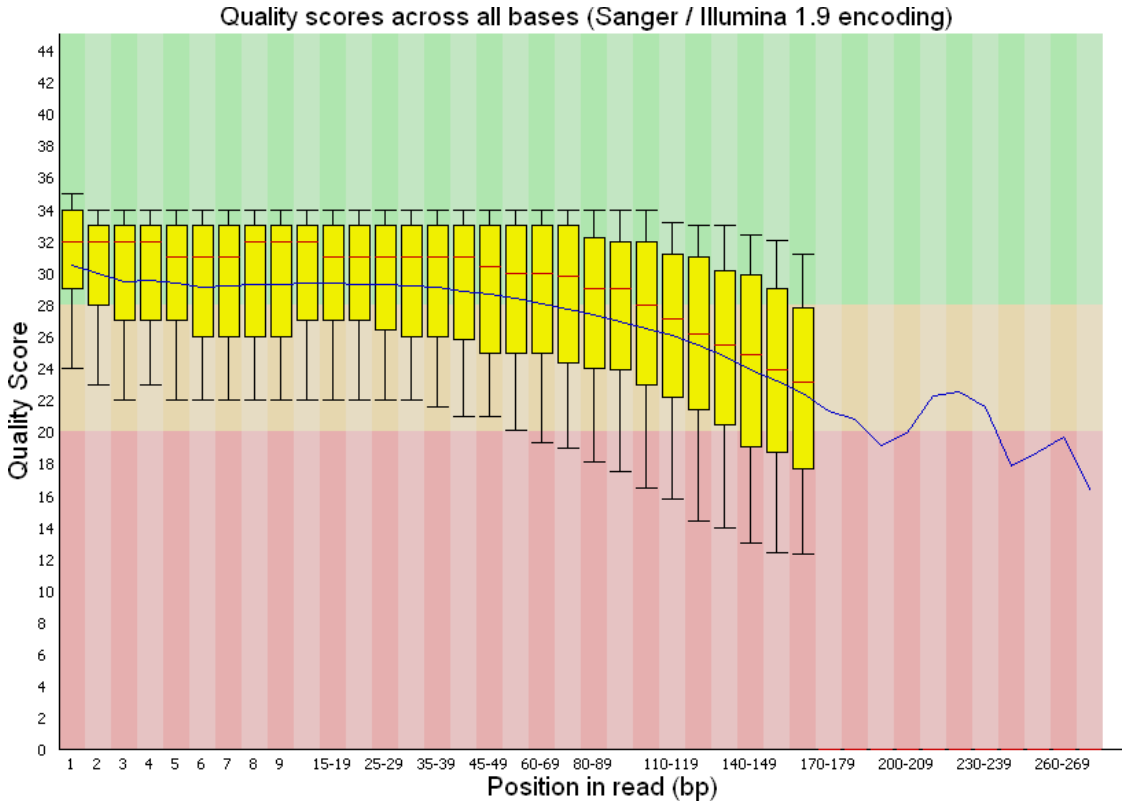

Figure 2.3: Example of poor quality sequencing reads. A single sample, not utilized for experiments presented in this dissertation, was used to generate reads with an in house ion Torrent PGM sequencer presented here for illustration. Trimming all reads to 100bp and filtering reads which have a large number of poor quality bases is warranted with this raw read data as the average quality score drops below the recommended level of 20.

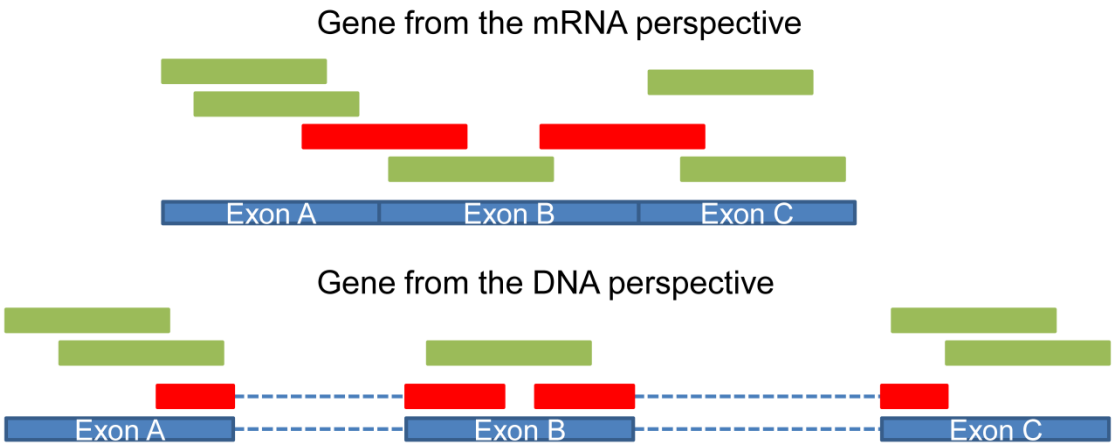

Figure 2.4: Tophat segmented search algorithm. The segmented search was applied in all sequencing studies. This algorithm breaks a read into 25bp segments and then attempts to map those segments to the reference genome. If a mapping is found, the program attempts to map the other segment to the same chromosome within a certain base pair distance specified by the user. Once pileup islands are identified, the program attempts to map across junctions by merging islands into a separate reference and mapping to it. Should enough reads map to the new merged reference, it is identified as a true junction and reported in the output file. Green coloring indicates unbroken reads, red indicates reads crossing a splice junction, blue indicates exons of the gene. Dotted blue lines indicate the junctions between read segments and between exons.

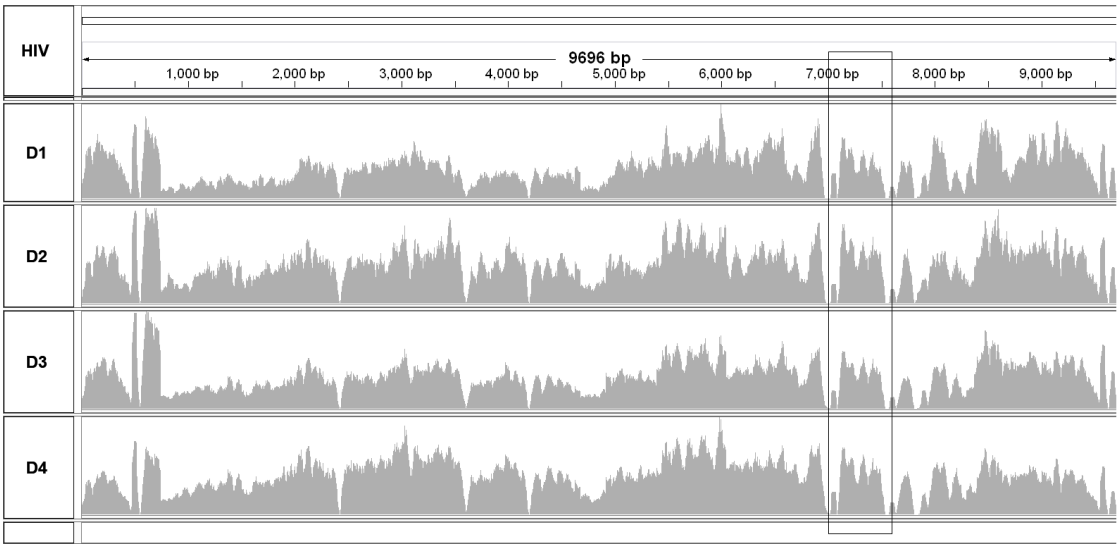

Figure 2.5: HIV read coverage. All reads mapping to the HIV genome were utilized to generate a coverage map for all four infected samples and viewed with the integrative genome viewer [147]. The boxed region starting at ~7000bp is the region which is deleted in the DHIV construct. This image is adapted from Bonczkowski *et al*. (2014), *Replication competent virus as an important source of bias in HIV latency models utilizing single round viral constructs*. Retrovirology 11: 70.

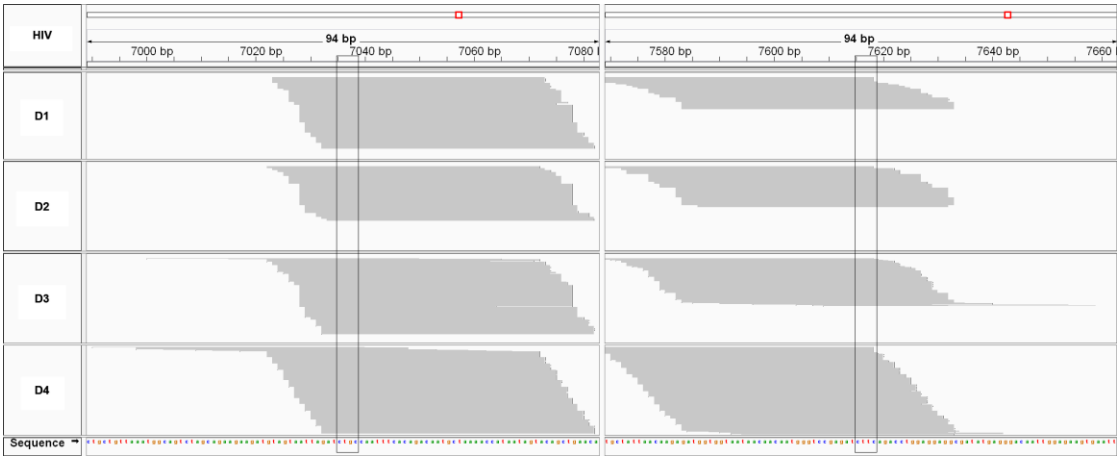

Figure 2.6: Reads mapping across the deleted region break points. (Left) Reads mapping across the start of the deleted region in the infected samples. (Right) Reads mapping across the end of the deleted region. The aligned file (.bam file) was filtered to only show reads crossing these break points. The boxed in region is the junction between the DHIV construct and the deleted region. The red box indicates the zoomed in portion of the HIV genome. This image is adapted from Bonczkowski *et al*. (2014), *Replication competent virus as an important source of bias in HIV latency models utilizing single round viral constructs*. Retrovirology 11: 70.

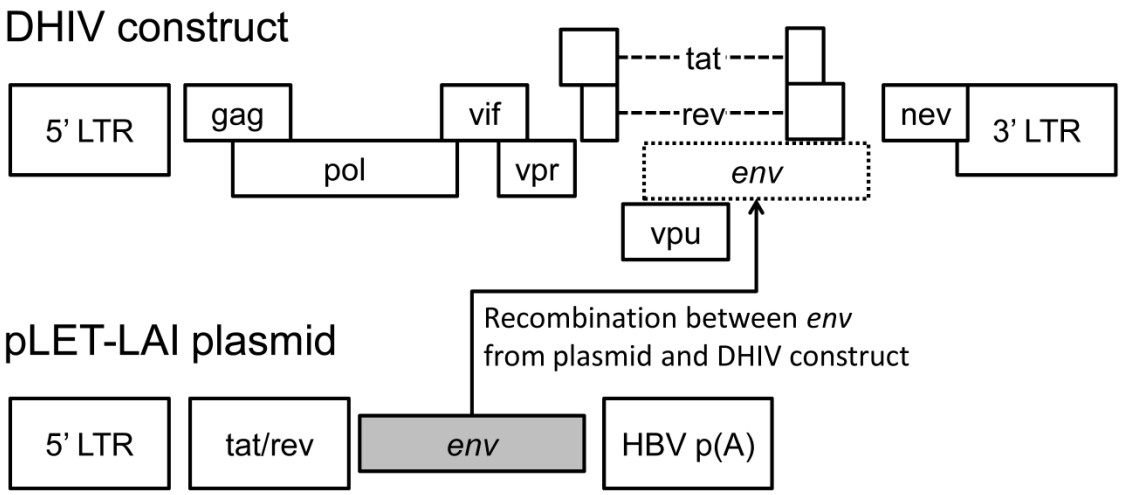

Figure 2.7: Recombination of the pLET-LAI plasmid with DHIV. Represented is the DHIV construct and the pLET-LAI plasmid. The dotted box in DHIV is the *env* gene which is deleted in DHIV. That same region is shaded in the pLET-LAI plasmid. Recombination occurred where DHIV incorporated the *env* sequence from the plasmid to create fully infectious progeny. This image is inspired and adapted from Bosque, A., and Planelles, V. (2011). *Studies of HIV-1 latency in an ex vivo model that uses primary central memory T cells.* Methods 53: 54-61.

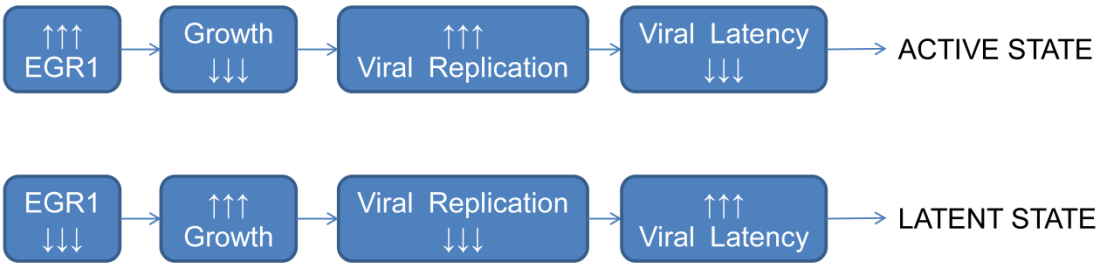

Figure 2.8: Candidate marker of HIV latency. *EGR1* was downregulated in this model of HIV latency, and has been previously shown to be involved in HIV latency [35]. Upregulation of *EGR1* can lead to an active state as growth is suppressed, leading to viral replication. Downregulation of this gene may lead to a growth increase, a reduction in viral replication, and an increase in latency.

2.8 Tables

Table 2.1: A large number of DEGs appear to be responding to HIV active infection. All DEGs were curated for a relationship to the interferon anti-viral response, and then sorted based upon fold change. The q-value represents the FDR corrected p-value from differential expression testing with EdgeR, and adjusted for multiple comparisons using all DEGs. Significance is set at $q < 0.05$.

Genes	Fold Changes	q-value
IFI44L	3.240	1.85E-04
IFIT1	3.092	5.97E-07
IFI44	1.831	4.66E-03
OAS1	1.830	5.14E-04
IFIT3	1.789	1.96E-02
MX1	1.718	3.87E-04
OASL	1.718	1.52E-03
ISG15	1.681	1.47E-03
IRF8	1.664	2.14E-02
IFI6	1.592	1.30E-02
IFIT2	1.571	1.62E-02
HERC5	1.494	3.67E-02
USP18	1.471	3.67E-02

Table 2.2: HIV read mapping statistics. Mapping to the whole HIV genome revealed a significant number of reads mapping to the supposedly deleted region. A large amount of variation was detected in the number of HIV reads, but the average percentage of reads mapping to the deleted region remained stable. The percentage of latently infected cells is also reported.

Sample	D1	D2	D3	D4	Avg.	s.d.
HIV Reads	6220	6051	28583	118932	39947	53710
Reads in Deleted Region	447	449	1997	7790	2671	3490
% of Delete Region Reads	7.2	7.4	7.0	6.5	7.03	0.39
% Latently Infected Cells	92.3	88.9	81.6	83.7	86.6	4.87

Table 2.3: Top 20 DEGs correlating to HIV reads sorted by correlation. Asterisks next to the gene indicate that it is involved with interferon gamma signaling. Pearson correlation was performed between the expression level of the gene in the latently infected (LI) samples and the number of HIV reads mapping to that LI sample.

Genes	Correlation to HIV reads
HLA-DQA1	1.000
PNPLA7	0.998
HKDC1	0.996
KDM4B	0.995
*OAS1**	0.993
STC2	-0.988
C12orf68	0.986
*IFI6**	0.986
ADPRH	0.985
HLA-DRB5	0.985
ATF3	0.983
*OASL**	0.983
*MX1**	0.980
*IFI44L**	0.979
GADD45A	0.973
LOC100130231	0.953
GJB6	0.946
*HERC5**	0.944
*IFI44**	0.939
GPI	-0.935

Table 2.4: Potential markers of HIV latency. Correlation analysis (Pearson correlation) was used identify genes with low correlation (< 0.5) to the number of HIV reads, but high absolute correlation (> |0.75|) to latency percentage. Presented are genes which passed this filter and sorted by absolute correlation. Fold change values presented are LI/UI.

Genes	Fold Change	Correlation to Latency	Correlation to HIV reads
EGR1	-1.81	-0.941	0.312
PTCH2	0.71	-0.862	0.113
HTR2B	-1.74	-0.860	0.440
ADNP	-1.36	-0.816	0.464
GOLGA8A	-1.50	-0.757	0.477

2.9 Acknowledgements

The dissertation author would like to thank the National Institute of Health (AI104282), the San Diego Veterans Medical Research Foundation, and the Department of Veterans Affairs (VA) for supporting this work. The dissertation author also acknowledges the use of the computer facilities and associated support services at the University of California, San Diego. The views expressed in this chapter are those of the dissertation author and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government. The sponsors of this research were not involved in the study design, collection or interpretation of the data, manuscript preparation, or the decision to submit the article for publication.

This chapter is adapted from **White, C.H.**, Beliakova-Bethell, N., Bosque, A., Richman, D.D., Planelles, V., Woelk, C.H., *An Unexpected Finding from an RNA-Seq Analysis of a T_{CM} Model of HIV Latency*, a collaborative unpublished work. The dissertation author was the primary author and responsible for the detection and analysis of reads from the deleted region of HIV, the development of the next generation sequencing pipeline, quality control of the data, the application of this pipeline to sequencing data from this model, the analysis and interpretation of transcriptome results, and writing the manuscript.

Chapter 3

Transcriptomic Analysis Implicates the p53 Signaling Pathway in the Establishment of HIV Latency in Central Memory CD4 T-Cells

3.1 Abstract

The search for a cure for HIV has been greatly hindered by the presence of a viral reservoir that persists despite anti-retroviral therapy (ART). At present, HIV cure strategies are impeded by the limited understanding of transcriptomic changes in CD4 T-cells upon the establishment of latency, and by the paucity of latently infected cells that may be derived from HIV infected individuals for study. To examine the signaling pathways and viral determinants of latency and reactivation, several models of HIV latency have been developed and characterized. The cultured primary central memory CD4 T-cell (T_{CM}) model of HIV latency, henceforth referred to as the Martins *et al*. model, is reflective of HIV latency as it incorporates wild type virus (HIV_{NL4-3}) and ART to suppress virus replication. RNA-Seq was used to examine this model, and an investigation of the host and viral gene expression in the resting and activated

conditions indicated that the resting condition was reflective of a latent state. Comparison of the host transcriptome between the uninfected and latently infected conditions of this model demonstrated 826 differentially expressed genes, many of which were related to p53 signaling and suggested that this pathway may be integral to the establishment of HIV latency. Inhibition of the transcriptional activity of p53 by pifithrin-α during HIV infection reduced the establishment of latency. Further study in pifithrin-α treated samples suggested that this decrease in latently infected cells was caused by a reduction of integrated HIV. In summary, this T_{CM} model was validated as reflective of HIV latency, and may be utilized for screening for latency reversing agents to be investigated in shock and kill approaches to cure HIV, for searching for cellular markers of latency, and for demonstrating key aspects of the establishment of latency such as the involvement of the p53 signaling pathway.

3.2 Introduction

A major obstacle to the eradication of HIV is the persistence of the latent viral reservoir [8]. While anti-retroviral therapy (ART) has been extremely effective at suppressing viral replication, it has not eradicated this virus reservoir [19]. Upon the removal of ART, HIV emerges from the latent state with concomitant disease progression [20, 21]. The low frequency of latently infected cells in the HIV patient (1 in 10^6 resting CD4 T-cells) complicates the study of this viral reservoir *in vivo* [19, 159, 160]. This has prompted the development of models of HIV latency based on chronically infected cell lines and primary human CD4 T-cells [34]. To obtain an accurate representation of HIV latency *in vivo*, it is essential to fully characterize the different models of HIV latency.

Transcriptome profiling by microarrays or RNA-Seq allows the simultaneous evaluation of transcriptional activity for either very large subsets of genes or the entire genome within a sample, thus providing a comprehensive analysis of the biological condition of the cell population at any given time. These technologies are becoming important for the study of HIV latency, particularly in the search for biomarkers of HIV latency [94] and for treatment effects with latency reversing agents [91]. Krishnan and Zeichner [35], utilized cDNA spotted microarrays to compare cell lines chronically infected with HIV (i.e., ACH-2, J1.1, and U1) to identify 32 genes that were differentially regulated across cell lines. The laboratory of Fabio Romerio used Agilent microarrays to profile latently infected and uninfected conditions from four donors from their primary CD4 T-cell latency model, and a gene encoding a surface receptor, *CD2*, was identified to be enriched in latently infected cells [161].

RNA-Seq is the current state of the art technology with respect to transcriptomics and is thought to exhibit greater specificity and dynamic range compared to microarrays [103]. The first RNA-Seq study of a primary CD4 T-cell model of latency incorporated a green fluorescent protein expressing virus [95]. When samples from a single donor were profiled over time, a large number of genes were identified as dysregulated during the latent phase (N=227) and were associated with chemokine receptors, cytokine signaling, and general immune responses. Previous work (**Chapter 2**) used RNA-Seq to profile latently infected and uninfected samples from 4 donors from the first iteration of this cultured primary central memory CD4 T-cell model of HIV latency, the Bosque and Planelles model [37]. This study demonstrated the defective vectors used to seed the latent reservoir in this model were recombining to reconstitute actively replicating HIV. This observation led to the revision of this model by incorporating wild type virus (HIV_{NL4-3}) and anti-retroviral

therapy (ART) to suppress virus replication [128], further referred to as "Martins *et al*." in this dissertation.

The purpose of this study was to characterize the Martins *et al*. model of HIV latency [128] by RNA-Seq to confirm whether this model reflects a latent state and to identify mechanisms involved in the establishment of HIV latency. Using biological scale normalization (BSN) with spike-in controls from the External RNA Controls Consortium (ERCCs), an increase in HIV transcription and a shift from unspliced to multiply spliced transcripts was demonstrated after activation, suggesting this model reflects a non-induced latent state as has been demonstrated in aviremic patients [162, 163]. Furthermore, analysis of cellular markers of activation and quiescence indicated that generated cultured T_{CM} cells were in a quiescent and resting state even when cultured in the presence of continuous IL-2. Comparison of latently infected cells to uninfected cells identified differential expression of genes in the p53 signaling pathway. Using pifithrin-α, an inhibitor of the transcriptional activity of p53 [164], a reduction in the number of latently infected cells in the Martins *et al*. model was demonstrated. These results indicated that the this model represents a suitable model to study HIV latency, as it reflects a quiescent state of the cells with an accumulation of unspliced HIV transcripts. Both characteristics are reflective of the HIV latent state in aviremic patients [162, 163]. Moreover, activation of the transcriptional activity of p53 by HIV may be involved in the efficient establishment of latency in resting CD4 T-cells.

3.3 Methods

3.3.1 Reagents and sample generation

Nelfinavir was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH; raltegravir from Merck & Company, Inc.; human IL-2 from Dr. Maurice Gately, Hoffmann-La Roche Inc. [51]; HIV-1_{NL4-3} from Dr. Malcolm Martin [52]. Pifithrin-α was obtained from Santa Cruz Biotechnology.

Sample preparation and generation of infected T_{CM} cells was fully described previously [128]. Briefly, peripheral blood mononuclear cells (PBMCs) were isolated from 4 healthy donors following IRB-approved protocol no. #67637 (University of Utah) or obtained from the Gulf Coast Regional Blood Center (Houston, TX). Naïve CD4 T-cells were isolated and then activated using human αCD3/αCD28-coated magnetic beads (one bead per cell, Thermo Fisher Scientific, Cat. No. #11131D) in the presence of αIL-4, (1 µg/mL; Peprotech; Cat. No. 500-P24) αIL-12, (2 µg/mL; Peprotech, Cat No. 500-P154G), and tumor growth factor (TGF)-β1 (10 µg/mL; Peprotech; Cat. No. 100-21) for 3 days. Cells were expanded in medium containing human IL-2 (30 IU/mL) for a further 4 days and then infected (or mock-infected) with HIV-1_{NL4-3} by spinoculation at a multiplicity of infection of 0.1. After infection, cells were further cultured in IL-2 for 3 days, subjected to crowding in round bottom plates in the presence of IL-2 for another 3 days, and then cultured for a further 4 days in the presence of IL-2 and ART (nelfinavir, 0.5 µM; raltegravir, 1.0 µM). Any productively infected cells remaining were then removed by magnetic isolation of CD4+ cells using the Dynabeads CD4+ isolation kit (Thermo Fisher Scientific, Cat. No. #11551D). At this stage, samples were taken for the latently infected (LI) condition. Uninfected (UI) cells were cultured under the same conditions and collected at the same time as LI cells. Additional cell aliquots were subjected to

reactivation with αCD3/αCD28-coated magnetic beads in the presence of ART for 2 days for the uninfected activated (UIA) and latently infected activated (LIA) conditions. For pifithrin-α treated samples, pifithrin-α was added 3 days post infection (day 10) during the crowding stage and replenished at day 13 until just prior to reactivation at 10 days post infection (day 17).

3.3.2 RNA isolation and total RNA-Seq data generation

Total RNA was extracted from 16 T_{CM} cell samples from 4 donors for the 4 conditions (UI, LI, UIA and LIA) using the RNeasy Plus Mini Kit (QIAGEN, Cat. No. 74134) according to manufacturer's instructions, with the addition of an on-column DNase treatment (RNase-free DNase Set, QIAGEN, Cat. No. 79254). RNA integrity (RIN) values of samples were on average 9.9 (± 0.1 s.d.) as determined using a Bioanalyzer 2100 (Agilent Technologies, CA, USA) and RNA concentration was measured by Nanodrop 2000 (Thermo Fisher Scientific). To account for transcriptional amplification, synthetic RNA standards from the External RNA Controls Consortium (ERCC RNA Spike-In Mix 1, Ambion, CA, USA, Cat. No. 4456740) were spiked into total RNA isolated from each sample. T_{CM} cells in each sample were counted in quadruplicate and ERCC spike-ins were added at 1 µL per million cells (1:100 dilution). RNA-Seq libraries were prepared from 100ng of total RNA using the TruSeq Stranded Total RNA Library Prep kit (Illumina, CA, USA) after depletion of cytoplasmic and mitochondrial ribosomal RNA with Ribo-Zero Gold (Epicentre, WI, USA). All libraries were sequenced to a read depth of >75 million reads using the Illumina HiSeq2000 to generate 100bp paired-end reads.

3.3.3 Total RNA-Seq data analysis

FASTQ files from the HiSeq2000 instrument were provided by Expression Analysis (Q^2 Solutions), NC, USA and assessed for quality using the Fastx-Toolkit suite available from the Hannon lab [165]. Reads were then mapped to the hg38 (human genome released on December 24th, 2013) using Tophat (version 2.0.13) [107]. Default settings were used with the options -bowtie1 and --no-coverage-search. Mapped reads were then sorted by name and converted to sequence aligned map files (.sam files) using samtools [149]. Sam files were then counted against the Gencode annotation (version 21) [166] with HTSeq v0.6.1p2 [108] using the default settings. All level three annotations were removed from Gencode prior to counting to remove the possibility of assigning reads to genes not verified as stringently. Any gene which did not achieve at least 1 count per million (CPM) reads in at least 4 samples (either resting or activated) was removed. Reads were then separately mapped to the HIV genome (Genebank accession number AF324493) with Tophat utilizing the following alterations from default settings: Max segment intron length of 10000, max coverage intron length of 10000, 0 segment mismatches, 0 mismatches, and no coverage search. These options were selected to reduce the possibility of a read being erroneously assigned to the HIV genome due to possible sequence similarity to hg38. Reads were then counted against the HIV genome using HTSeq as described previously.

Finally, all reads were mapped to the 92 ERCCs using bowtie with 0 mismatches and the --best option (reference available from Thermo Fisher Scientific [167]). Any ERCC that did not achieve greater than 5 reads in at least 4 samples was removed leaving 67 ERCCs. Reads that mapped to the ERCC reference were then re-mapped to hg38 and HIV to demonstrate that these reads mapped solely to

ERCCs and were not cross mapping to any other reference. All raw FASTQ files for this experiment are available through the Gene Expression Omnibus accession number (GSE81810).

3.3.4 Quantification of spliced transcripts

To determine HIV read splicing, a method developed by Mohammadi and *et al.* [95] was utilized. This method counts reads that pass through two HIV splice junctions, D1 (directly after the LTR region), and D4 (splice junction between *Tat-rev* and *Vpu*) that define multiply spliced (MS), single spliced (SS), and unspliced (US) transcripts. Reads passing through the junction D1 belong to US transcripts. Reads which align to the left of this junction, but are broken at D1 and align to other sections of HIV, correspond to either SS or MS transcripts (SS+MS). Reads overlapping the D4 junction correspond to reads from either US or SS transcripts (US+SS). Finally, reads broken at the D4 junction correspond to reads from MS transcripts. The percentage of reads from US transcripts is estimated by obtaining the fraction of reads that pass through the D1 junction. The percentage of reads from MS transcripts is estimated by obtaining the fraction of reads that break at the D4 junction. The percentage of reads from SS transcripts is estimated by subtracting the estimated percentages of US and MS reads from 100%.

3.3.5 Biological scale normalization (BSN) from resting to activated states

Due to the large difference in cellular RNA content between the activated and resting states, BSN normalization described previously [168] was utilized to adjust read levels based upon internal ERCC spike in controls. 11 ERCCs (ERCC_00130, ERCC_00002, ERCC_00096, ERCC_00074, ERCC_00004, ERCC_00113,

ERCC_00046, ERCC_00171, ERCC_00136, ERCC_00003, and ERCC_00108) with read counts above 1000 across all samples were chosen for this analysis as low abundance ERCCs were not readily detected in activated samples (data not shown). For each ERCC, an average was calculated across the resting state only. This average was then divided by each ERCC count in individual donors resulting in an ERCC scale factor based upon the resting state for each sample. These scale factors were then averaged across all 11 ERCCs to obtain a sample specific scaling factor. These scaling factors were then used in the BSN normalization method described. First, the raw counts in each state were summed together to generate a raw library size for each sample. This library size was averaged to generate a common library size and then converted to a pseudo library size for each donor by multiplication with the previous scale factors (Equation 1). Then, for each gene and sample, the read concentration is calculated (Equation 2).

$$\boldsymbol{X_{ij}} = CommonLibrarySize * ScaleFactor \quad (1)$$

$$\boldsymbol{C_{ij}} = \frac{\boldsymbol{R_{ij}}}{\boldsymbol{R_j}} \quad (2)$$

Where $\mathbf{C_{ij}}$ is the concentration of transcript **i** relative to sample **j**, $\mathbf{R_{ij}}$ is the read count of gene **i** in sample **j**, and $\mathbf{R_j}$ is the total number of reads in sample **j**. The final step is to assign the previously calculated pseudo reads to genes based upon their concentration to obtain a scaled normalized dataset (Equation 3).

$$\boldsymbol{P_{ij}} = \boldsymbol{C_{ij}} * \boldsymbol{X_{ij}} \quad (3)$$

3.3.6 Normalization and differential expression within the resting state

Since the resting state has no drastic changes in total RNA, RUVSeq [169] was utilized to remove unwanted variation across samples. This method is

appropriate within the resting dataset because it removes variation based upon sample differences not attributable to differing conditions, but rather due to technical variation in sample prep and processing. RUVSeq adjusts for nuisance technical effects by performing factor analysis on the ERCC spike in controls. This program creates a generalized linear model where the read counts are regressed onto known covariates of interest (e.g. the covariate of treatment effect) and factors of unwanted variation of unknown origin. RUVSeq then calculates a vector of unwanted variation based upon the expression of the negative controls (ERCCs in this experiment) that may be used in linear modeling approaches for differential gene expression. This vector is then used as a blocking variable where the unwanted variation is pushed onto this variable that is then removed from the comparison of states (uninfected vs. latently infected in this case). ERCC results that pass filtering were combined with host gene mapping results and normalized using upper quartile normalization available in RUVSeq.

Following read count normalization, differential expression was performed with EdgeR [170]. This program utilizes an overdispersed Poisson model which accounts for both biological and technical variability coupled with an empirical Bayes method to moderate the degree of overdispersion. Afterwards the data is fit to a generalized linear model (GLM) and a GLM likelihood ratio test is performed to determine if the coefficient representing the contrast between the conditions of interest is equal to 0 indicating no differential expression. As this was a paired design, the donor number was used as a blocking variable. Additionally, the vector of unwanted variation provided by RUVSeq was also used as a blocking variable. Significance for differential expression was set at an FDR corrected p-value < 0.05.

3.3.7 Functional analysis of differentially expressed genes

Pathway analysis was performed with Toppgene [171] using the functional analysis enrichment tool, ToppFun, with the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways selected. Pathway images were generated from the KEGG Pathway Database [172]. Fold change data were $\log_2$ transformed, colored, and overlaid upon the p53 signaling pathway. A protein interaction network (PIN) was generated with Metacore^{TM} and visualized through Cytoscape v2.8.3 [173]. For PIN construction, genes were filtered using a $\log_2$ fold change of 0.5 between latently infected and uninfected cells. Read pileup figures were generated with the Integrated Genome Browser [174]. Venn diagrams were constructed using Venny 2.1.0 [175].

3.3.8 RT-qPCR validation

Total RNA was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) following the manufacturer's instructions. RT-qPCR was performed for each target using TaqMan Universal PCR Master Mix (AmpErase UNG) on the ABI 7900HT Fast Real-Time PCR System (Applied Biosystems, CA, USA). Thermal cycling conditions were 95˚C for 10 minutes followed by 40 cycles of 95˚C for 15 seconds and 60˚C for 1 minute. Changes in gene expression were calculated using the $2^{-\Delta\Delta CT}$ method with the spike-in ERCC control, ERCC_00130, as the normalizer. Targets validated by RT-qPCR were *GADD45A, MDM2, FAS, TNFRSF10B, TP53I3, BBC, IL2, KLF2*, and *HEXIM1* (Taqman Assay IDs Hs00169255_m1, Hs01066930_m1, Hs00163653_m1, Hs00366278_m1, Hs00936520_m1, Hs00248075_m1, Hs00174114_m1, Hs00360439_g1, Hs00538918_s1, respectively). In addition, levels of unspliced Gag, multiply spliced *Tat-Rev* and poly-adenylated HIV transcripts were measured using custom Taqman

Gene Expression Assays using primer and probe sequences described previously [176, 177] (http://www.bio-protocol.org/e1492).

3.3.9 Flow cytometry and p24 analysis

For the detection of surface *FAS* (CD95) expression, cells were stained with FAS/CD95 Antibody (DX2), FITC conjugate (Molecular Probes™). For the detection of surface *TNFRS10B* (DR5) expression, cells were stained with CD262 (DR5, TRAIL-R2) Antibody (DJR2-4 (7-8)), APC conjugate (Biolegend®). Cells were also were stained with a viability dye (Fixable Viability Dye eFluor 450, Affymetrix, eBioscience, San Diego, CA). For the dual detection of CD4 and HIV-1 p24Gag, cells were first stained with the viability dye (Fixable Viability Dye eFluor 450), followed by staining with CD4 antibody (S3.5), APC conjugate (Molecular Probes™). After staining, cells were fixed, permeabilized, and stained for HIV-1 p24Gag as previously described [128]. In all experiments, HIV p24Gag negative staining regions were set with uninfected cells treated in parallel. Flow cytometry was performed with a BD FacsCanto II flow cytometer using FACSDiva acquisition software (Becton Dickinson, Mountain View, CA). Data were analyzed with Flow Jo (TreeStar Inc, Ashland, OR).

3.3.10 Measurements of HIV DNA with PCR

DNA from $2x10^6$ cells was isolated using DNeasy Blood and Tissue Kit (Qiagen). DNA was quantified using NanoDrop 1000 (Thermo Fisher Scientific) and 10 ng of total DNA was subject to PCR to analyze total HIV DNA (Forward: 5'-CATGTTTTCAGCATTATCAGAAGGA-3'; Reverse: 5'-TGCTTGATGTCCCCCCACT-3'; Probe: 5'-FAM-CCACCCCACAAGATTTAAATACCATGCTT-BHQ1-3'); HIV-1 2-

LTR (Forward: 5'-GTGCCCGTCTGTTGTGTGACT-3'; Reverse: 5'CTTGTCTTCTTTGGGAGTGAATTAGC-3'; Probe: 5'-FAM-TCCACACTGACTAAAAGGGTCTGAGGGATCTCTTB-BHQ1-3'); and RNase P (TaqMan® RNase P Detection Reagents Kit, Thermo Fisher Scientific). Reactions were carried out in a LightCycler® 480 (Roche) using PCR Master Mix (2X) (Thermo Fisher Scientific). Final concentration of primers and probe for total HIV DNA and 2-LTR was 400 nM and 0.2 pmol/μL respectively. Serial dilutions of pNL43 (NIH AIDS Reagent Program) or pU3U5 [178] (kindly provided by Dr. Ward De Spiegelaere and Dr. Linos Vandekerckhove) were used for a molecular standard curve for total HIV DNA or 2-LTR, respectively. Serial dilutions of male genomic DNA was used for a molecular standard curve for RNase P (TaqMan® RNase P Detection Reagents Kit, Thermo Fisher Scientific).

Genomic DNA was subjected to nested quantitative Alu-LTR PCR for integrated provirus as previously described with some modifications [179]. For the first reaction, 250 ng of total DNA was amplified using Platinum Taq DNA polymerase (Invitrogen). Reactions were carried with 1.5 mM of $MgCl_2$, 200 μM dNTPs, 400 nM of Alu164 primer (5'-TCCCAGCTACTCGGGGAGGCTGAGG-3') and 400 nM of PBS primer (5'-TTTCAAGTCCCTGTTCGGGCGCCA-3'). Amplifications were performed in a MultiGene Optimax (Labnet International, Inc) with the following parameters: 1) 94C 5 min; 2) 18x 94C 30 sec, 66C 30 sec, 72C 5 min; 3) 72C 10 min. PCR samples were subject to a 1/10 dilution in water. 2 μL of the diluted sample was subject to qPCR reactions in a LightCycler® 480 (Roche) using PCR Master Mix (2X) (Thermo Fisher Scientific). Final concentration of primers (AE989-2 5'-CTCTGGCTAACTAGGGAACCCAC-3'; AE990-2 5'-CTGACTAAAAGGGTCTGAGGGATCTC-3') and probe (5'-FAM-

TTAAGCCTCAATAAAGCTTGCCTTGAGTGC-BHQ1-3') was 400 nM and 0.2 pmol/µL respectively. A serial dilution of pcDNA3.1-LTR was used for a molecular standard curve. pcDNA3.1-LTR was generated by cloning the 3' LTR from NL43 into pcDNA3.1.

3.4 Results

3.4.1 Generation of latently infected cultured T_{CM} cells

To dissect the viral and cellular status from the Martins *et al.* model of HIV latency [128], gene expression data was generated by RNA-Seq for a total of 16 T_{CM} samples representing the following 4 conditions from 4 different donors: uninfected (UI), latently infected (LI), uninfected and activated (UIA), and latently infected and activated (LIA). Briefly, naïve CD4 T-cells from four healthy donors were isolated and activated for three days in conditions that generate central memory CD4 T-cells (**Figure 3.1A**) [37]. After activation, cells were expanded in the presence of IL-2 and infected at day 7 with HIV-1_{NL4-3} at a low multiplicity of infection that rendered 3-7% of cells infected at day 10 (**Figure 3.1B**, Day 10). Cells were then densely cultured in a 96 well round plate to increase the efficiency of infection for an additional three-day period (**Figure 3.1B**, Day 13). At that time point, cells were diluted and cultured for 4 extra days in the presence of IL-2 and ART. When using the previous iteration of this model, "Bosque and Planelles", cells become resting even in the presence of IL-2 [129]. At day 17, CD4 T-cells were isolated using magnetic bead sorting. This strategy was chosen because productively infected cells not only express HIV-1 gag (p24), but also downregulate CD4 expression on the cell surface due to the expression of the accessory genes *Nef* and *Vpu* (**Figure 3.1B**, Day 17) [180, 181]. This procedure eliminates any productively infected (p24 positive) cell, as well as any

CD4 negative cells present in the culture. As such, cells that are analyzed by RNA-Seq are CD4 positive cells which do not express detectable levels of viral antigens and represent uninfected (UI) or latently infected (LI) cells. These cells were further activated with αCD3/αCD28 beads in the presence of ART to generate uninfected and activated (UIA) or latently infected and activated (LIA) cells.

3.4.2 Cultured T_{CM} cells have a resting and quiescent phenotype

Initially, the resting and activated conditions were compared (UI vs. UIA and LI vs. LIA) to analyze the activation status of the resting cells. When comparing RNA content, transcriptional amplification had occurred (**Figure 3.2**). Traditional normalization procedures for transcriptomic data do not account for transcriptional activation [182, 183]. Therefore, BSN using ERCC spike-in control RNAs, was used to normalize RNA-Seq data across activation conditions [168]. A number of gene expression markers of CD4 T-cell activation were modulated following activation in both the LIA and UIA conditions (**Figure 3.3**). Notably, *IL2* and components of its receptor (*IL2RA*, *IL2RB*, *IL2RG*) were upregulated upon activation, as were members of the NFκB complex (*NFκB1*, *NFκB2*, *REL, RELA*, *RELB*), and *CD28* itself. The *KLF2* gene, which is highly upregulated in quiescent CD4 T-cell lymphocytes, but repressed during activation [184, 185], was significantly downregulated as expected [185]. The modulation of *IL2* and *KLF2* upon activation was confirmed by RT-qPCR (**Figure 3.3**), and, as an example, reads mapping to *IL2* were examined for Donor 1 and showed an absence of reads in the UI condition and a strong signal for each exon following activation (**Figure 3.4**). In summary, the cultured T_{CM} cells in this model have the phenotypic characteristics of a quiescent T-cell, and stimulation with αCD3/αCD28 beads modulates known markers of CD4 T-cell activation as expected.

3.4.3 Activation of HIV from a latent state

After confirming that the cultured T_{CM} cells were representative of a quiescent CD4 T-cell, and that the cells responded to αCD3/αCD28 beads appropriately, the effect of activation on HIV transcription was evaluated. HIV transcription in the resting and activated states was compared after BSN. Activation induced a global upregulation of HIV reads from the resting to the activated conditions (average 6.6 fold change, s.d. ± 3.6, *t*-test *p*-value = 0.0394) with an increase in all major splicing groups: US, SS, and MS (**Figure 3.5A**). A significant increase of US, MS, and polyadenylated HIV transcript reads upon activation was confirmed by RT-qPCR (**Figure 3.5B**). In summary, examination of HIV reads and splicing further confirmed that the resting state in the Martins *et al*. model reflected activation of transcription from an HIV latent state.

3.4.4 Perturbation of host gene expression in latently infected cells

The UI and LI samples were compared to identify 826 DEGs **(See supplementary file "White_Table_S3-1_DEG_list_LI_vs_UI.csv")**, 275 of which were downregulated and 551 of which were upregulated (FDR = 0.05). The top 100 DEGs presented in the heatmap (**Figure 3.6**) demonstrated the consistency of gene expression across donors. The DEGs identified in the Martins *et al*. model were compared to other primary CD4 T-cell [95, 161] and cell line models [35]. Although no up- or downregulated DEGs were identified in common across all four models of HIV latency (**Figures 3.7**), greater overlap was identified when comparing primary cell models in a pairwise fashion, in particular, with upregulated genes observed between the Martins *et al.* model [128] and the model used in Iglesias-Ussel *et al.* [161] (**Table 3.1**).

3.4.5 Functional analysis of genes perturbed during latency

In order to develop a better understanding of genes perturbed during latency in the Martins *et al*. model, the 826 genes differentially expressed between the LI and UI conditions were subjected to pathway and protein interaction (PIN) analysis. The only KEGG pathway that attained significance for over-representation of DEGs was the "*p53 signaling pathway*" (FDR corrected *p*-value = 5.5E-06). The results of differential gene expression were overlaid on the *p53 signaling pathway* and revealed that multiple threads related to *apoptosis* and *DNA repair and damage prevention* were upregulated in latently infected cells [172] (**Figure 3.8A**).

A PIN constructed using genes differentially expressed during latency also revealed a number of genes related to p53 activity (**Figure 3.8** and **Table 3.2**). This PIN contained two major hubs (*AR* and *MDM2*) and one minor hub (*TNFRSF10B a.k.a. DR5* or *TRAIL-R2*), which are all related to p53 activity. For example, *MDM2* facilitates negative feedback to the p53 signaling pathway by degrading p53 and the upregulation of this gene in the LI condition may be indicative of prior p53 activity [122]. *AR* is negatively regulated by p53 signaling [186] and, correspondingly, is downregulated in the LI condition. *TNFRSF10B*, a gene involved in programmed cell death, is also upregulated and known to be induced by p53 activity in response to DNA damaging agents [187].

Several genes (*BBC3*, *FAS, GADD45A, HEXIM1, MDM2, TNFRSF10B* and *TP53I3*) selected from the *p53 signaling pathway* and the PIN were subjected to RT-qPCR analysis (**Figure 3.8C**). The upregulation of transcripts of *BBC3, GADD45A, MDM2, TP53I3*, and downregulation of *HEXIM1*, was confirmed as significant by RT-qPCR. The direction of fold change was validated for the cell surface markers *FAS* and *TNFRSF10B,* which were significantly upregulated with the RNA-Seq data

(**Figure 3.8D**). The cell surface expression of these markers was interrogated in an independent donor set by flow cytometry, but only *FAS* (CD95) was confirmed as significantly upregulated (fold change of 1.47 ± 0.20, **Figure 3.8F**). In summary, functional analysis of genes perturbed during latency using pathway and PIN analysis identified dysregulation of genes associated with p53 activity.

3.4.6 Inhibition of p53 signaling reduces the size of the latent reservoir

In order to further investigate the possible role of p53 transcriptional activity in HIV latency, pifithrin-α, a well-characterized inhibitor of the transcriptional activity of p53 [164], was added to CD4 T-cells from four additional donors during the establishment of latency in the Martins *et al*. model (**Figure 3.9A**). Inhibition of p53 had no effect on HIV replication, since p24 levels exhibited no difference between treated and untreated samples at days 13 and 17 (**Figure 3.9B and 3.9C**). However, inhibition of p53 using pifithrin-α resulted in a reduction (average 32%, s.d. 8%) of the number of cells producing p24 after reactivation with αCD3/αCD28 beads (**Figure 3.9B and 3.9D**). Interestingly, when pifithrin-α was added only during the reactivation step, there was no effect upon viral reactivation (**Figure 3.9E**). These results suggest that inhibition of p53 may have an effect on the establishment of latency. To test this hypothesis, integrated HIV was analyzed by Alu-PCR in latently infected cells treated or untreated with pifithrin-α. Although not significant at the $p < 0.05$ cut off, there was a strong trend towards the reduction of integrated HIV in pifithrin-α treated samples (**Figure 3.9F**). This trend is further demonstrated in the correlation between integrated copies of HIV and the percentage of p24 positive cells after reactivation with αCD3/αCD28 beads (**Figure 3.9G**). Moreover, there were no differences between treated or untreated samples when cell death, total HIV DNA, or 2-LTR

copies were analyzed (**Figure 3.9H through 3.9L**). In summary, these studies with pifithrin-α confirmed that p53 signaling may play an important role in the establishment of HIV latency.

3.5 Discussion

3.5.1 Analysis of host differentially expressed genes and viral transcripts

In this study, RNA-Seq was utilized to characterize the Martins *et al*. model of HIV latency [128], which incorporates a replication-competent virus (HIV-1_{NL4-3}) and ART to suppress HIV replication. RNA-Seq analysis demonstrated that the resting condition in this model reflects a quiescent and a latent state when compared to the activated state both at the cellular and viral level. An increase in HIV transcription was noted in all donors after activation (**Figure 3.5A**), indicative of a departure from a latent state [188]. When examining the percentage of transcripts, a significant shift from US towards MS transcripts was noted suggesting an increase in early HIV transcripts (*Vpr*, *Tat*, *Rev*, and *Nef*). This result is in agreement with previous reports showing an accumulation of MS transcripts in resting CD4 T-cells isolated from aviremic HIV patients [162, 189]. This increase in HIV transcription was not due to abortive transcripts since polyadnylated viral transcripts, which reflect fully elongated and correctly processed HIV mRNA, were significantly upregulated after activation with αCD3/αCD28 beads.

By comparing expression profiles from uninfected (UI) to those from latently infected (LI) cells, a total of 826 differentially expressed genes were identified. While there was only minimal overlap in genes dysregulated during latency when comparing the Martins *et al*. model to published data from three additional models of HIV latency (**Figure 3.7**), there appeared to be greater overlap between primary CD4 T-cell

models compared to models based on cell lines. Several reasons can account for these differences. First, this overlap might have been greater if the same transcriptomic technologies had been utilized; Iglesias-Ussel et al. [161] used the Agilent-012391 Whole Human Genome Oligo Microarray (G4112A) to profile gene expression compared to RNA-Seq in this study. Second, the RNA-Seq study of Mohammadi et al. [95] analyzed samples from only a single donor over a latent time course. Greater overlap between primary HIV latency models may occur when better powered transcriptomic studies (i.e. multiple donors) are performed using the state-of-the-art technology (i.e. RNA-Seq). Different models of latency vary in their intrinsic sensitivity to LRAs and this may also explain the low overlap within genes across different models found in this study [34].

Given these overlap findings with data published from other models of HIV latency; it is not surprising that the previous markers discovered in the first iteration, the Bosque and Planelles model of HIV latency (**Chapter 2**) were not identified as differentially expressed in the revised Martins *et al*. model. This revised version is quite different from the first iteration, hence overlap would not be expected. The Bosque and Planelles model used a deficient HIV construct that recombined with the pLET-LAI plasmid to produce productively infected cells and resulted in a large infected percentage (80-90%) with an unknown percentage of latently infected cells. The Martins *et al.* model utilized wild type virus with ART and had a much lower infected percentage in the presence of uninfected cells (3-5%). It is also possible that markers identified in the first model are not true markers of latency. Correlation analysis was utilized in an attempt to remove the effect of productive infection (**Chapter 2**), but the success of this method is uncertain, and only five DEGs remained after this analysis. It may also very well be possible that the first model

analyzed is actually a better representation of active infection while the second model analyzed in this chapter truly reflects a form of HIV latency.

3.5.2 Differential expression reveals dysregulation of multiple genes related to p53 signaling

Functional analysis of DEGs identified between the latently infected when compared to uninfected conditions clearly implicated the transcriptional activation of genes by p53 (**Figure 3.8A through 3.8C** and **Table 3.2**). For example, the genes *ACTA2*, *BBC3 (a.k.a. Puma), DDB2, DRAM1, FAS, FDXR, GADD45A, PRDM1, RHOC, TNFRSF10B,* and *TP53I3* are known to be induced by p53 [187, 190, 191, 192, 193, 194, 195, 196, 197], and were upregulated in the latently infected condition, which is in agreement with previous studies showing the activation of the p53 pathway mediated by HIV [198, 199, 200, 201]. Of these genes, *FAS* was further confirmed at the protein level (**Figure 3.8E and 3.8F**). In addition to upregulated p53 related genes, activation of p53 by genotoxic stress has been shown to result in downregulation of *AR* expression [186] which was also downregulated in the T_{CM} model (**Figure 3.8B**). In summary, the modulation of these genes was consistent with prior transcriptional activation of p53 [198, 199, 200, 201].

Interestingly, p53 related a gene that may facilitate the persistence of HIV latency, *PRDM1* (a.k.a Blimp-1), was upregulated in the latently infected condition. This gene binds to the p53 promoter and represses p53 activity in the regulation of normal cell growth [197]. *PRDM1* further represses basal and Tat-mediated HIV transcription by binding to an interferon stimulated response element within the HIV provirus [202]. This binding can repress transcriptional elongation, which may be a mechanism by which HIV is transcriptionally silenced to facilitate a latent state.

The p53 protein itself is highly regulated and its activity must be tightly controlled to allow normal cellular functioning [203]. To maintain homeostasis, p53 will not only activate genes that enhance and stabilize its activity, but also genes that repress its activity through negative feedback loops. Further evidence of prior p53 activation was demonstrated by the upregulation of genes which act to degrade p53 following activation (**Table 3.2**). For instance, *MDM2* was upregulated in the latently infected condition (fold change = 1.503) and mediates ubiquitination and breakdown of p53 resulting in the inhibition of p53-mediated apoptosis [122]. Several other genes upregulated in the latently infected condition, are also involved in the degradation or inhibition of p53: *BCL3, CCNG1, LIF,* and *PTK2* [122, 204, 205, 206]. For example, *PTK2* recruits *MDM2* to p53 and promotes degradation of p53 [122], while *BCL3* is induced by DNA damage and is required for the induction of *MDM2* [204]. The downregulation of *HEXIM1* and *NR4A1* in the latently infected condition may also contribute to a negative feedback loop with respect to p53 activity. The protein product of *HEXIM1* inhibits the ubiquitination of p53 by *MDM2* [207] and thus its downregulation in this study would facilitate p53 degradation. Additionally, the protein product of *NR4A1* interacts with p53 to block ubiquitination and degradation induced by *MDM2* [208]. Downregulation of *NR4A1* may allow *MDM2* to break down p53. Therefore, the latently infected (LI) condition in the Martins *et al*. model of HIV latency exhibited evidence of prior p53 activation and subsequent negative feedback of this signaling pathway.

3.5.3 Inhibition of p53 transcriptional activity results in a reduction of latently infected cells and viral integration

The identification of p53 related genes modulated during HIV infection led to the hypothesis that this pathway may be important for the establishment of latency. Experiments with the p53 inhibitor, pifithrin-α, demonstrated that inhibition of this pathway did not affect viral replication or apoptosis, but did limit the number of cells that could be reactivated from latency (**Figure 3.9D, and 3.9H through 3.9L**). Alu-PCR analysis indicated that p53 activation is required for successful integration of HIV (**Figure 3.9F and 3.9G**). The p53 gene is not only involved in apoptosis and cell cycle arrest, but is also involved in the activation of DNA repair mechanisms [209]. A number of p53-responsive genes identified in this study (*BBC3*, *DDB2*, *GADD45A*, *FDXR*, *PCNA* and *XPC*) are related to radiation-induced DNA damage [210] and point towards involvement of the nucleotide excision repair (NER) pathway, a process that recognizes and removes helix-distorting DNA lesions from the genome [211, 212]. For example, GADD45A has been shown to bind to UV-damaged DNA where it is believed to modify DNA accessibility within chromatin [213]. Although other studies have primarily implicated other DNA repair pathways such as non-homologous end joining and base excision repair in HIV DNA integration [214, 215], these results suggest the p53-responsive genes that are components of the NER pathway may also play a role in the establishment of latency. Interestingly, a previous siRNA screening to characterize DNA repair factors involved in HIV integration identified that siRNA knockdown of DDB2, a component of the NER pathway, resulted in a large reduction (71.3%) of HIV infectivity [214]. Further studies in this pathway will be needed to completely understand the role of p53 in the establishment of HIV latency.

Finally, these results are in agreement with previous studies showing the activation of the p53 pathway mediated by HIV [198, 199, 200, 201].

The importance of p53 signaling may not be confined to only the Martins *et al*. model as several genes related to the p53 pathway were also identified in the overlap with the T-cell model examined by Iglesias-Ussel and colleagues [161]. Specifically, the p53 related genes *ACTA2, BBC3, DDB2, DRAM1, FDXR, GADD45A,* and *TNFRSF10B* were significantly upregulated in both models (**Table 3.1**). It will be interesting to know if inhibiting the transcriptional activity of p53 in this model also affects the establishment of latency. This may suggest that some mechanisms involved in the establishment of latency may not be model dependent, but rather a more universal phenomenon across primary CD4 T-cell models of HIV latency.

3.5.4 Limitations and summary conclusions

The present study allowed the comparison of gene expression between the LI and UI conditions of the Martins *et al*. model of HIV latency, which may represent potential biomarkers of latency. However, the search for these biomarkers in this study was somewhat limited by the relatively low proportion of latently infected cells in the LI condition (mean 2.92%, s.d. $\pm$ 0.71%) that were in a large background of uninfected cells. Therefore, a gene specific to latently infected cells must be upregulated greater than 20 fold in order for it to be identified as being upregulated by 2-fold when comparing the LI and UI conditions. Furthermore, bystander effects could be contributing to the signal of differential gene expression between the LI and UI conditions, whereby signals emanating from the small proportion of latently infected cells (e.g. cytokines and chemokines) may be perturbing gene expression in uninfected cells in the LI condition. Regardless of these limitations, gene expression

markers of HIV latency undoubtedly exist within the 826 DEGs identified between the UI and LI conditions as demonstrated in this work by the importance of p53 signaling to HIV latency.

In summary, RNA-Seq was used to demonstrate that the Martins *et al*. model of HIV latency [128], is truly reflective of a latent state and that p53 signaling may be involved in the establishment of latency. Therefore, this model has been validated for screening latency reversing agents (LRAs) for shock and kill approaches to an HIV cure [44] (**Chapter 1, Section 1.4**) and may also be useful for identifying gene expression biomarkers of HIV latency. It would be beneficial to subject all primary CD4 T-cell models of HIV latency [34] to RNA-Seq analysis in statistically powered studies (i.e. > 3 donors) so that similarities and differences across models may be dissected. Finding similar genes across models will lead to the identification of gene expression biomarkers of HIV latency that may be used to isolate latently infected cells from HIV infected subjects and utilized in innovative cure strategies (e.g., radioimmunotherapy [216]), or killing latently infected cells by means of immunotoxins [217]. The profiling of other HIV latency models on the omics scale will lead to the validation or modification of these models, which will undoubtedly result in a better understanding of both *in vivo* latency and the therapies that can be used to facilitate a cure for HIV.

3.6 Figures

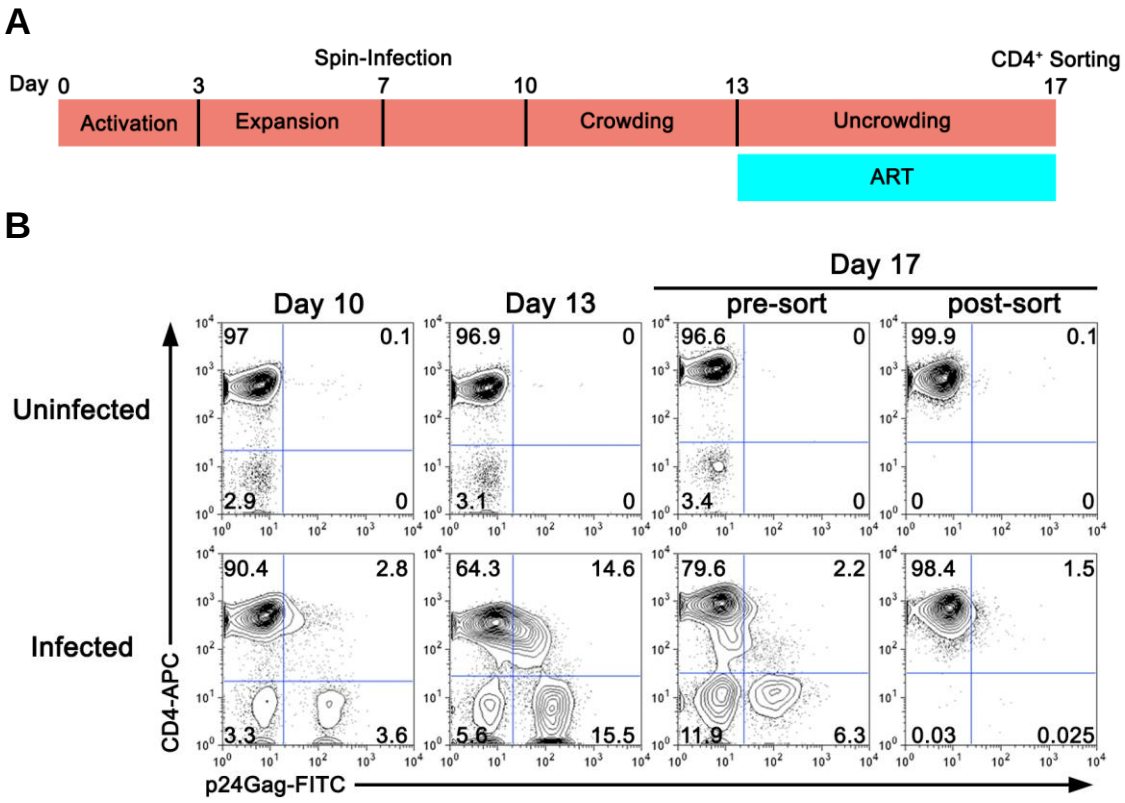

Figure 3.1: Generation of the Martins *et al*. model. (A) Timeline for the generation of latently infected cells indicating key events in the development of this model. (B) Analysis of infection in a representative donor. At the time points indicated, CD4 expression and HIV-1 p24Gag expression was measured by flow cytometry. The x-axis indicates HIV-1 p24Gag staining while the y-axis indicates CD4 staining. At day 17, cells were sorted based upon surface CD4 expression. Staining of cells pre- and post-sorting is indicated. Numbers in the plots are as follows: uninfected non-activated (top left), infected non-activated (bottom left), uninfected activated (top right), and infected activated (bottom right) cells. APC and FITC are the fluorochromes that were utilized in flow cytometry.

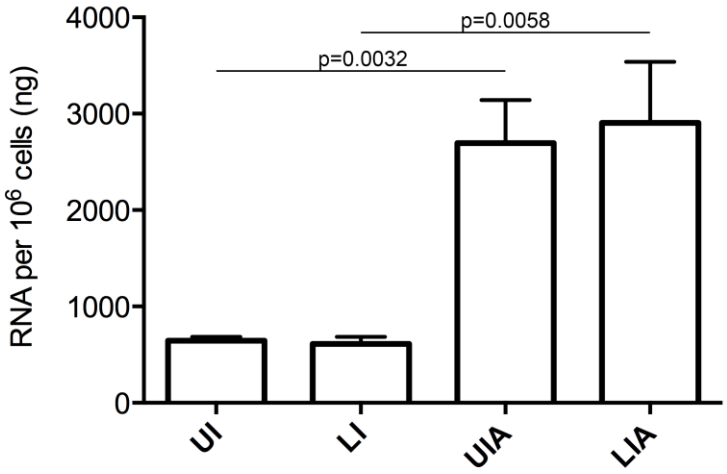

Figure 3.2: Average RNA content increases upon activation of CD4 T-cells. Total RNA was extracted from resting (UI, LI) and activated (UIA, LIA) CD4 T-cells and quantified by Nanodrop. Error bars indicate standard deviation measurements across donors. Significant increase in total RNA was determined with a paired *t*-test.

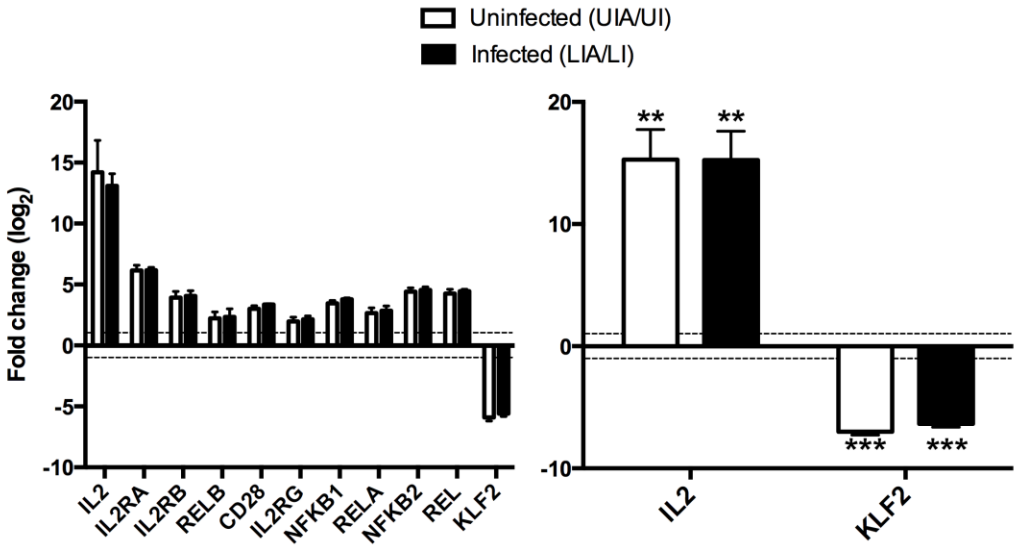

Figure 3.3: Known markers of CD4 T-cell activation are modulated following CD3/CD28 stimulation. The fold change of a panel of host gene markers of CD4 T-cell activation was assessed pre- and post-stimulation with αCD3/αCD28 beads by (left) RNA-Seq and (right) RT-qPCR. Fold changes were calculated for both uninfected (white bars) and latently infected (black bars) cells. Fold changes for all genes assessed by RNA-Seq were significant at an FDR corrected p-value < 0.05. Fold changes for *IL2* and *KLF2* assessed by RT-qPCR were significant in a paired *t*-test at the following levels: **p < 0.01 ***p < 0.001. Fold change is plotted on the $\log_2$ scale with error bars representing standard deviations. The horizontal dotted line in each figure indicates a $\log_2$ fold change cut off of one (i.e., actual fold change cut off of two). Abbreviations are as follows: *UI*, uninfected; *UIA*, uninfected activated; *LI*, latently infected; *LIA*, latently infected activated.

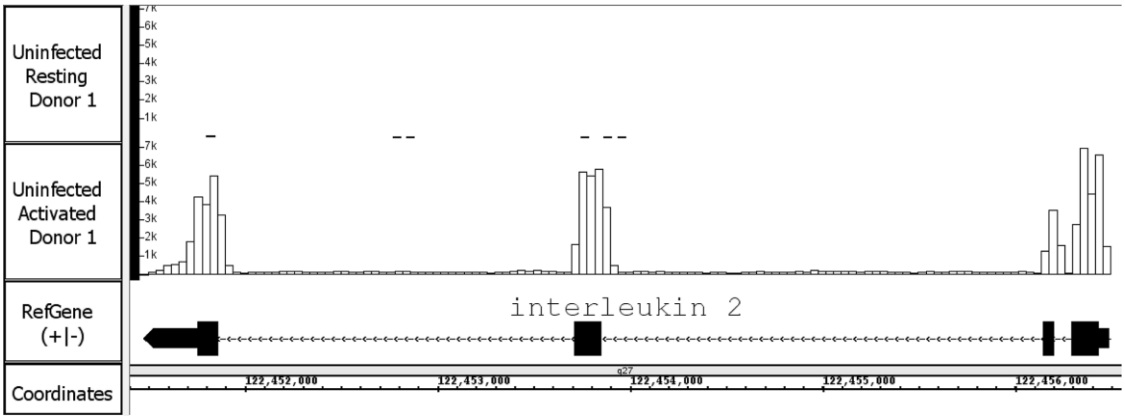

Figure 3.4: Abundance of reads mapping to *IL2*. Reads were mapped to the interleukin 2 gene in the uninfected resting (UI) and uninfected activated (UIA) conditions for donor 1 and were visualized using the Integrated Genome Browser [174]. Read coverage in the uninfected resting donor were extremely low in only a few locations indicated by bars in the top panel. The coordinates are based upon the hg38 human reference genome. The numbering on the left indicates the read coverage.

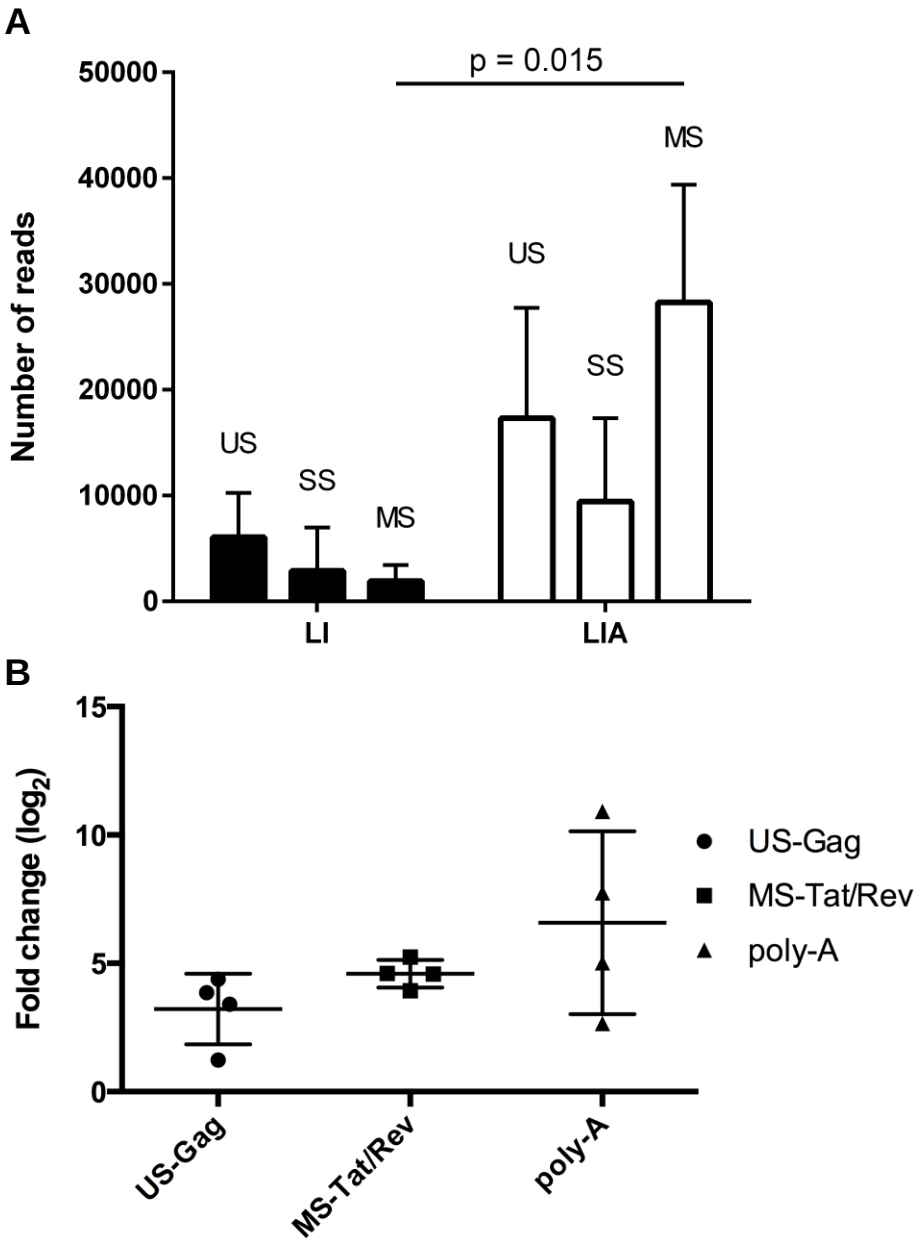

Figure 3.5: HIV splicing variants are upregulated following CD3/CD28 stimulation. (A) The abundance of HIV unspliced (US), singly spliced (SS), and multiply spliced (MS) transcripts in the LI and LIA conditions was estimated by mapping reads over the two major splice sites D1 and D4 (see Methods, **Section 3.3.4** for details). The significance of increases in the number of reads upon activation was assessed using a one-tailed paired *t*-test. (B) Fold changes upon activation of US RNA (US-*Gag*), MS RNA (MS-*Tat/Rev*) and poly-A RNA were determined by RT-qPCR. Significance of fold changes was confirmed using a one-tailed paired *t*-test and all RT-qPCR measurements were considered significant at the $p < 0.001$ level. Fold change is plotted on the $\log_2$ scale with error bars representing the standard deviation.

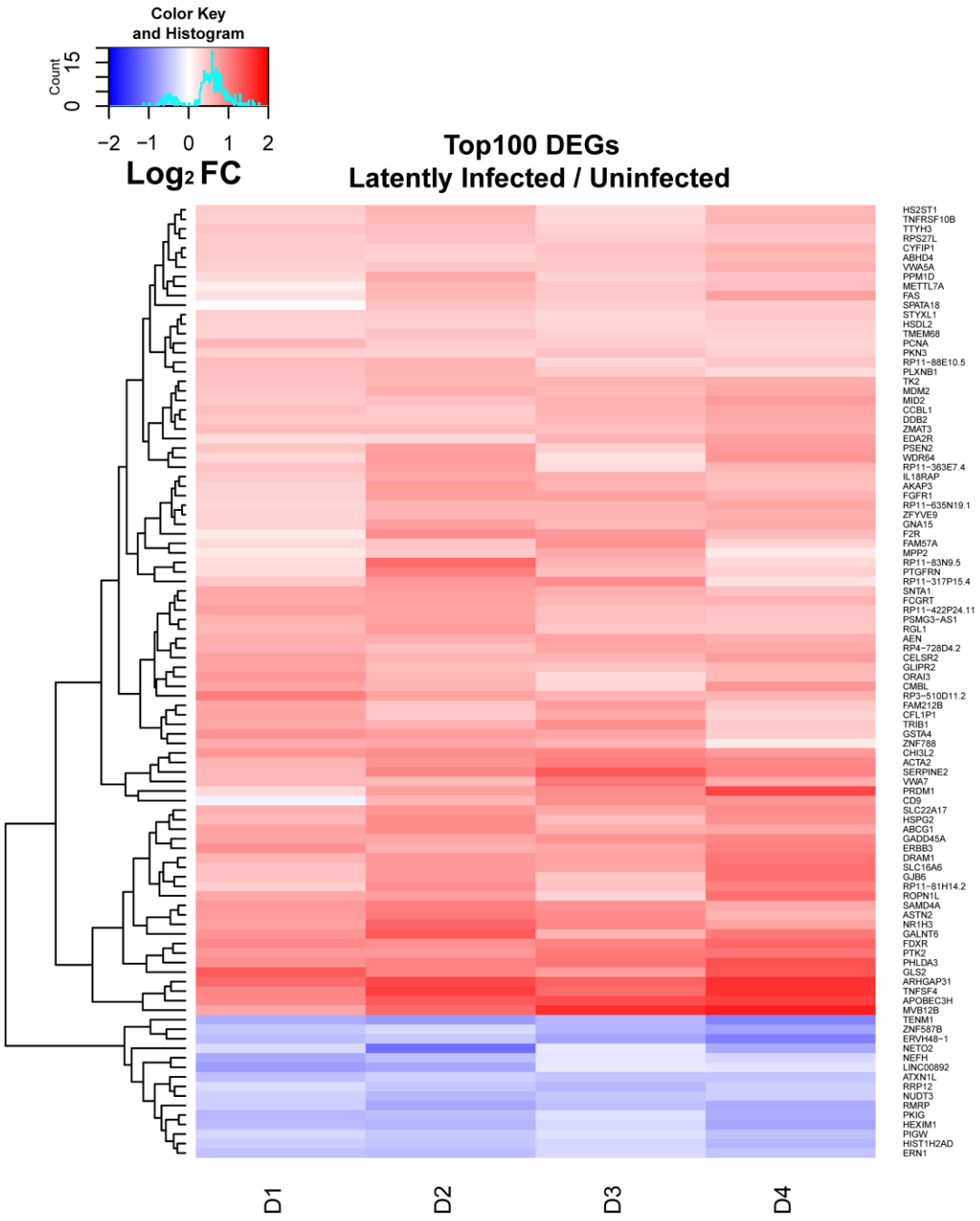

Figure 3.6: Heatmap of differentially expressed genes during HIV latency. Heatmap showing the top 100 differentially expressed genes (from a total of 826) with smallest p-values (FDR corrected) when comparing LI and UI conditions in a paired analysis. Log_2 fold change values were used to generate this heatmap where red represents upregulation and blue downregulation, according to the scale bar on the top left. The dendrogram represents how genes cluster based upon fold changes across the donors D1 through D4. This dendrogram was constructed by hierarchical clustering with distances calculated using the Euclidean method and then clustering these distances with the "complete" method.

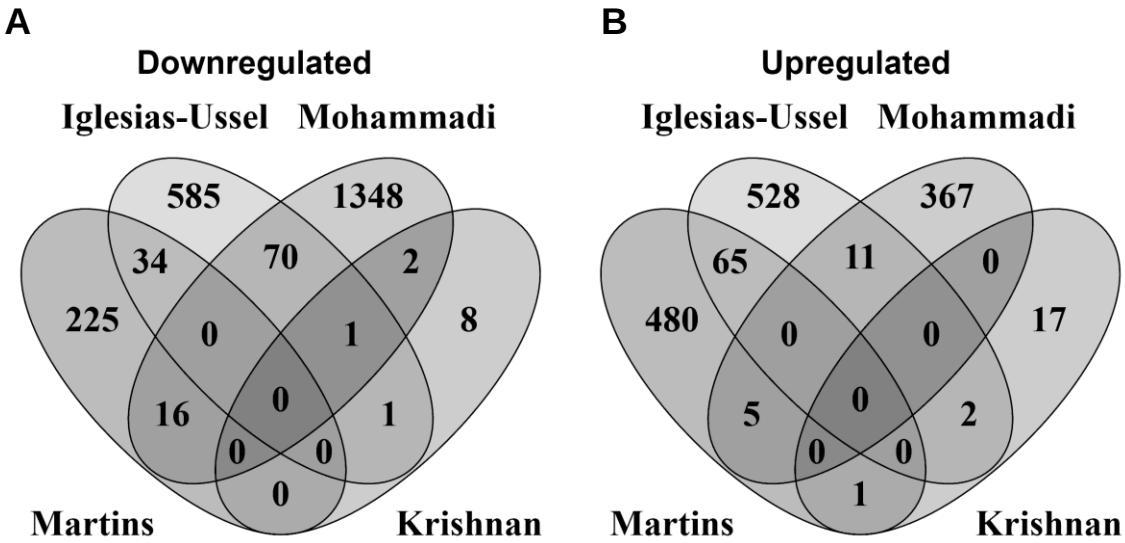

Figure 3.7: Overlap among four models of HIV latency. Venn diagram demonstrating overlap across four models of HIV latency. All differentially expressed genes were separated into (A) downregulated and (B) upregulated genes, and then compared across the following models: Martins *et al.* 2015 [128] (the RNA-Seq results from a cultured T_{CM} model of HIV latency for four donors presented here), Iglesias-Ussel *et al.* 2013 [161] (microarray results generated from a primary CD4 T-cell model for four donors), Mohammadi *et al.* 2014 [95] (RNA-Seq results from a primary CD4 T-cell model for a single donor), and Krishnan and Zeichner 2004 [35] (microarray results generated from HIV infected cell line models).

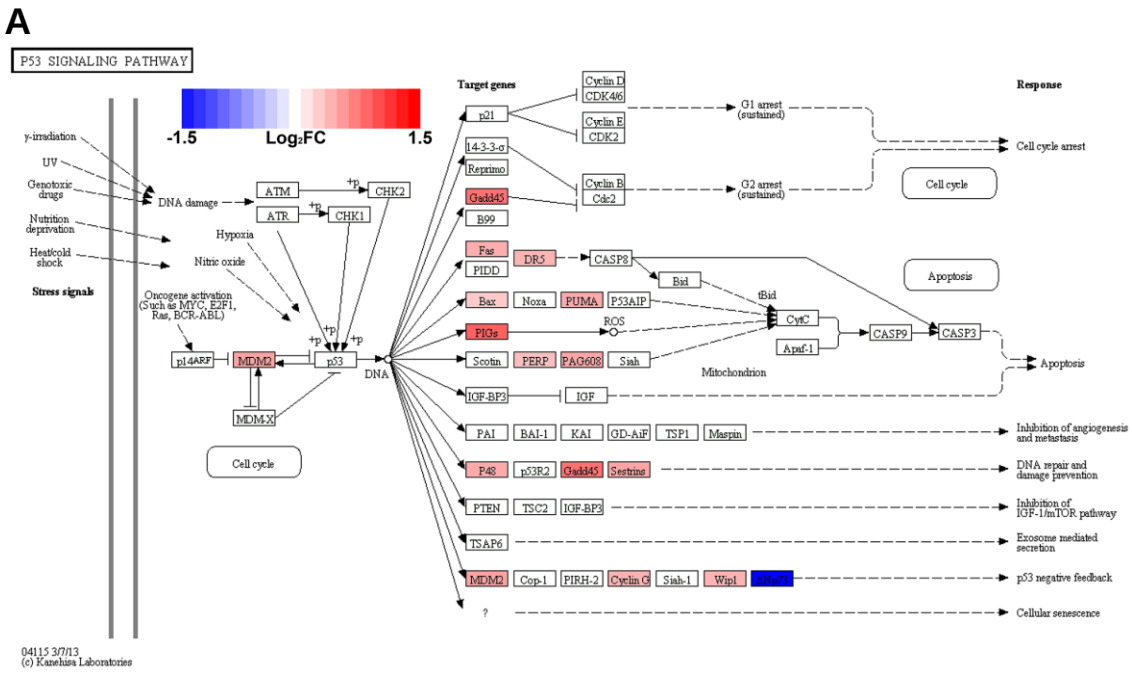

Figure 3.8: Functional analysis of genes modulated during HIV latency. (A) Gene expression changes overlaid upon the KEGG pathway, *p53 signaling pathway*. All differentially expressed genes from this pathway were included. (B) Protein interaction network (PIN) generated using all differentially expressed genes (N = 836) after a $\log_2$FC > 0.5 filter was applied (N = 64). Furthermore, any single nodes not attached to a main hub were removed for clarity. Abbreviations for the type of interaction are as follows: B – Binding, C – Cleavage, CM – Covalent Modification, GR – Group Relation, +P – Phosphorylation, TR – Transcription Regulation, T – Transformation. Edge coloring of Green indicates activation, Red indicates inhibition, Black indicates unspecified, and Blue is a special color utilized for a group relation. Arrows indicate the direction of this interaction. The color gradient of $\log_2$ fold changes for both (A) and (B) is included with red indicating upregulation while blue indicates downregulation. (C) RT-qPCR validation of a panel of p53 related genes modulated in HIV latency. Fold change is plotted on the $\log_2$ scale with error bars representing standard deviation. Significance for RT-qPCR and RNA-Seq results was determined with a paired *t*-test (*p < 0.05, **p < 0.01, ***p < 0.001) and EdgeR (FDR = 0.05) respectively. (D) Dot plot of counts per million (CPM) data from RNA-Seq results with p-values calculated with EdgeR. (E) Histograms of DR5 and CD95 surface expression from a representative donor. The red line indicates untreated controls while the blue line indicates pifithrin-α treated cells. Geometric mean intensity fluoresce (gMIF) is indicated. The grey histogram represents the unstained control. (F) Surface expression analysis as measured by FACS for DR5 (6 donors) or CD95 (5 donors). Significance was determined with a paired *t*-test (p-values provided).

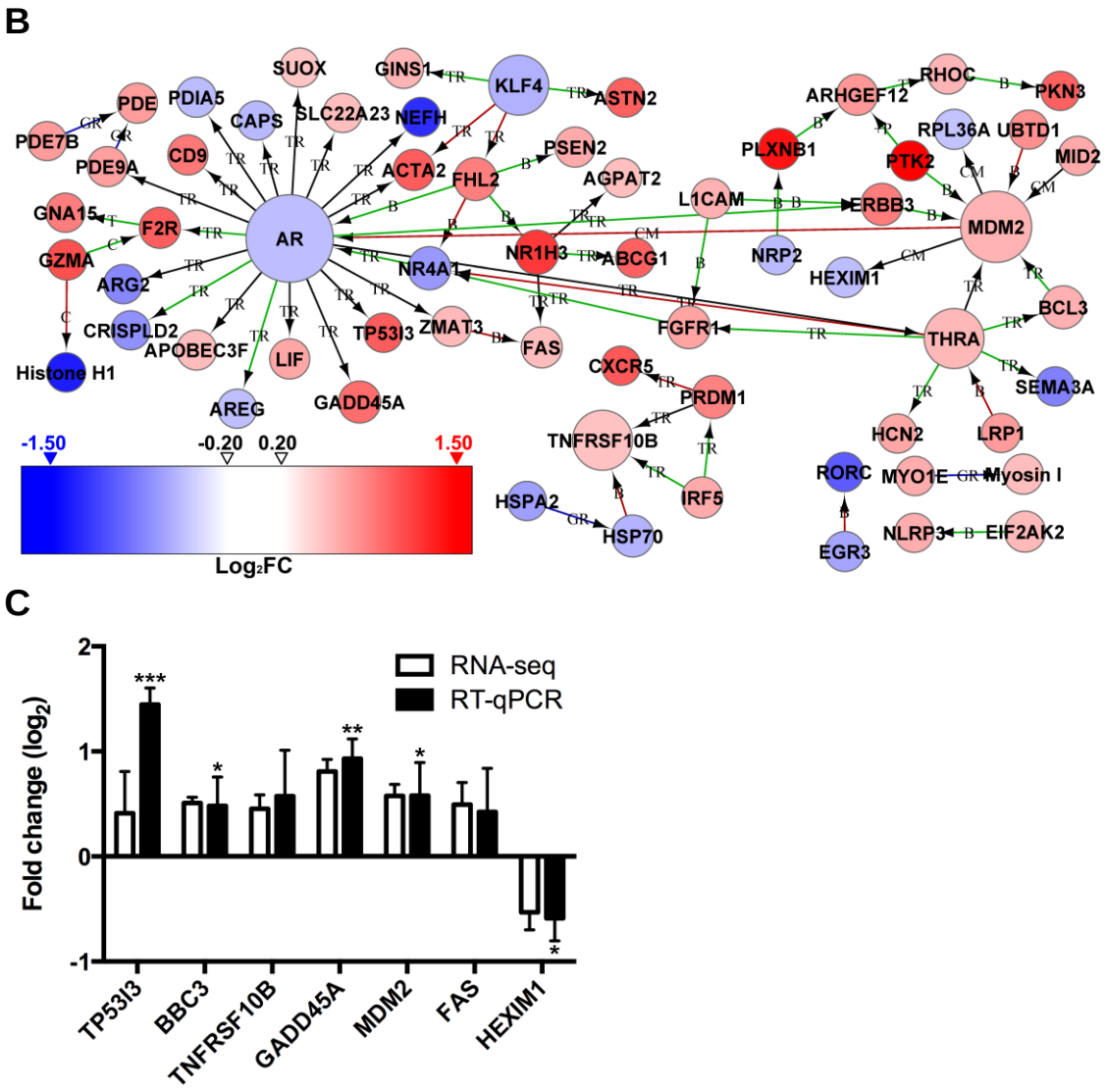

Figure 3.8: Functional analysis of genes modulated during HIV latency, continued.

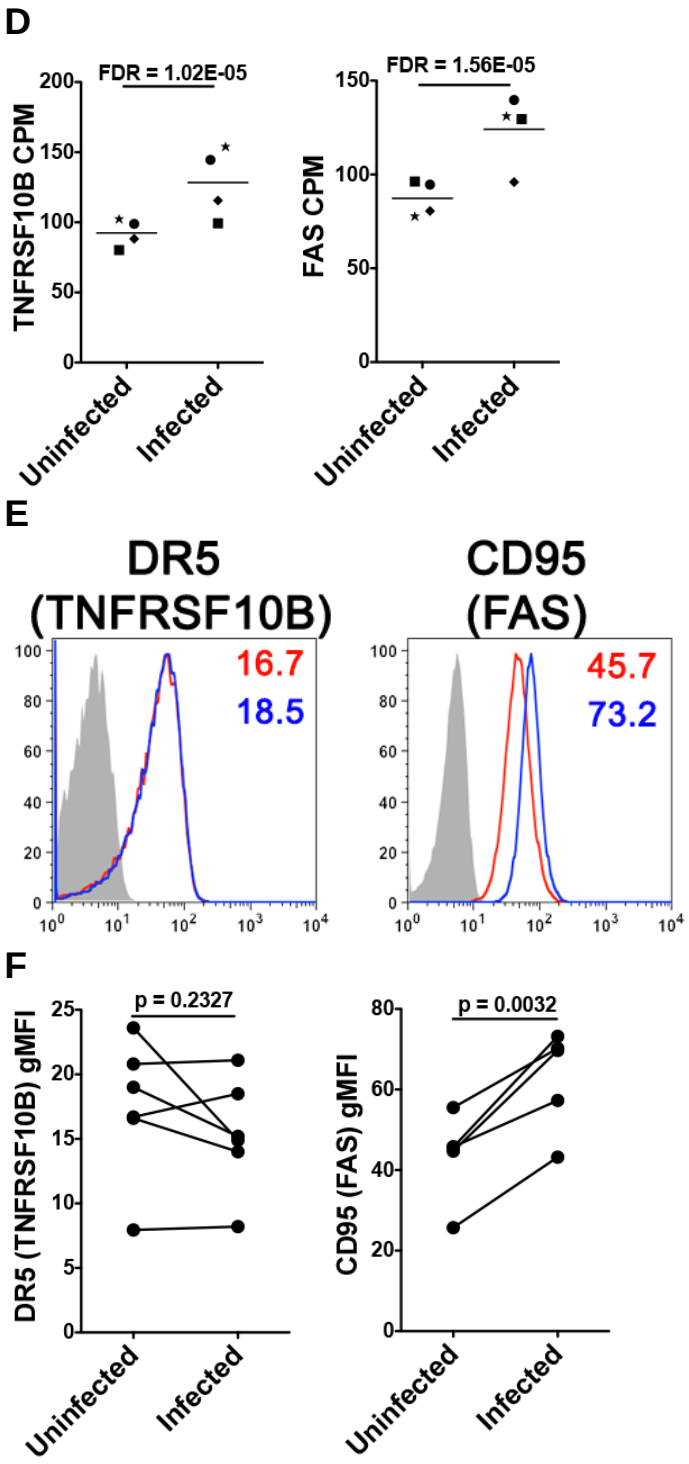

Figure 3.8: Functional analysis of genes modulated during HIV latency, continued.

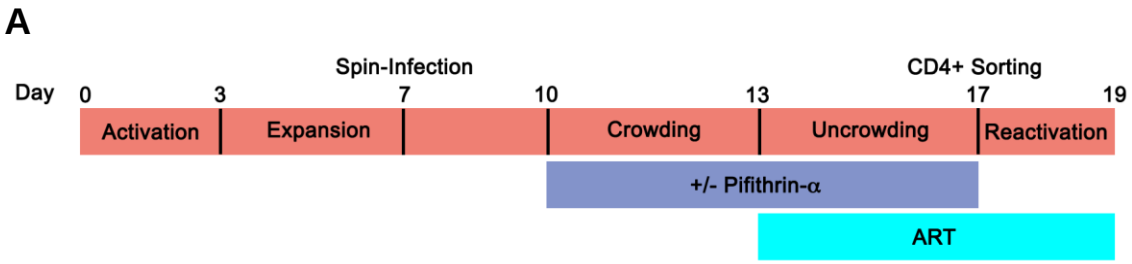

Figure 3.9: Effect of pifithrin-α upon the establishment of latency. (A) Timeline of the experimental procedure, including pifithrin-α treatment which was added to the culture at day 10 and replaced at day 13 until just prior to CD4+ sorting on day 17. (B) Representative flow cytometry plot with CD4 and p24Gag staining from a single representative donor indicating the percentage of uninfected non-activated (top left), infected non-activated (bottom left), uninfected activated (top right), and infected activated (bottom right) cells in all 3 treatment conditions from days 10 through 19. Dot plot of the effect of pifithrin-α treatment upon the percentage of latently infected cells producing p24 for (C) days 13 (left) and 17 (right) without reactivation. Four additional donors were tested and indicated by the different symbols in all dot plots. (D) Dot plot measuring percentage of cells producing p24 at day 19 after reactivation with pifithrin-α added during the crowding stage, day 10, indicated by blue coloring, and (E) added only during reactivation, day 17, indicated by salmon coloring. (F) Relative integrated copies of HIV as measured by Alu-PCR for the uninfected, infected, and infected pifithrin-α treated conditions. (G) Correlation between integrated copies and the percentage of p24 positive cells after reactivation. Blue coloring indicates pifithrin-α treated samples while black indicates normal infected samples. (H) Cell death as measured by the viability dye, Fixable Viability Dye eFluor 450 from Affymetrix at day 13 and day 17 in non-treated and treated with pifithrin-α. (I) HIV Gag copies and (J) 2-LTR copies per copies of the endogenous reference gene, RNAse P. (K) Correlation of total HIV DNA and (L) 2-LTR circles with the percentage of p24 positive cells after reactivation with αCD3/αCD28 beads. Black symbols represent untreated samples, while blue symbols represent cells treated with pifithrin-α from day 10 to day 17. No measures in (G) through (L) were considered significant. All p-values generated in (C) through (F) and (H) through (J) were obtained with a paired student's *t*-test while the correlation presented in (G), (K), and (L) was Pearson correlation with the p-value calculated based upon the Pearson's product moment correlation coefficient which follows a *t* distribution.

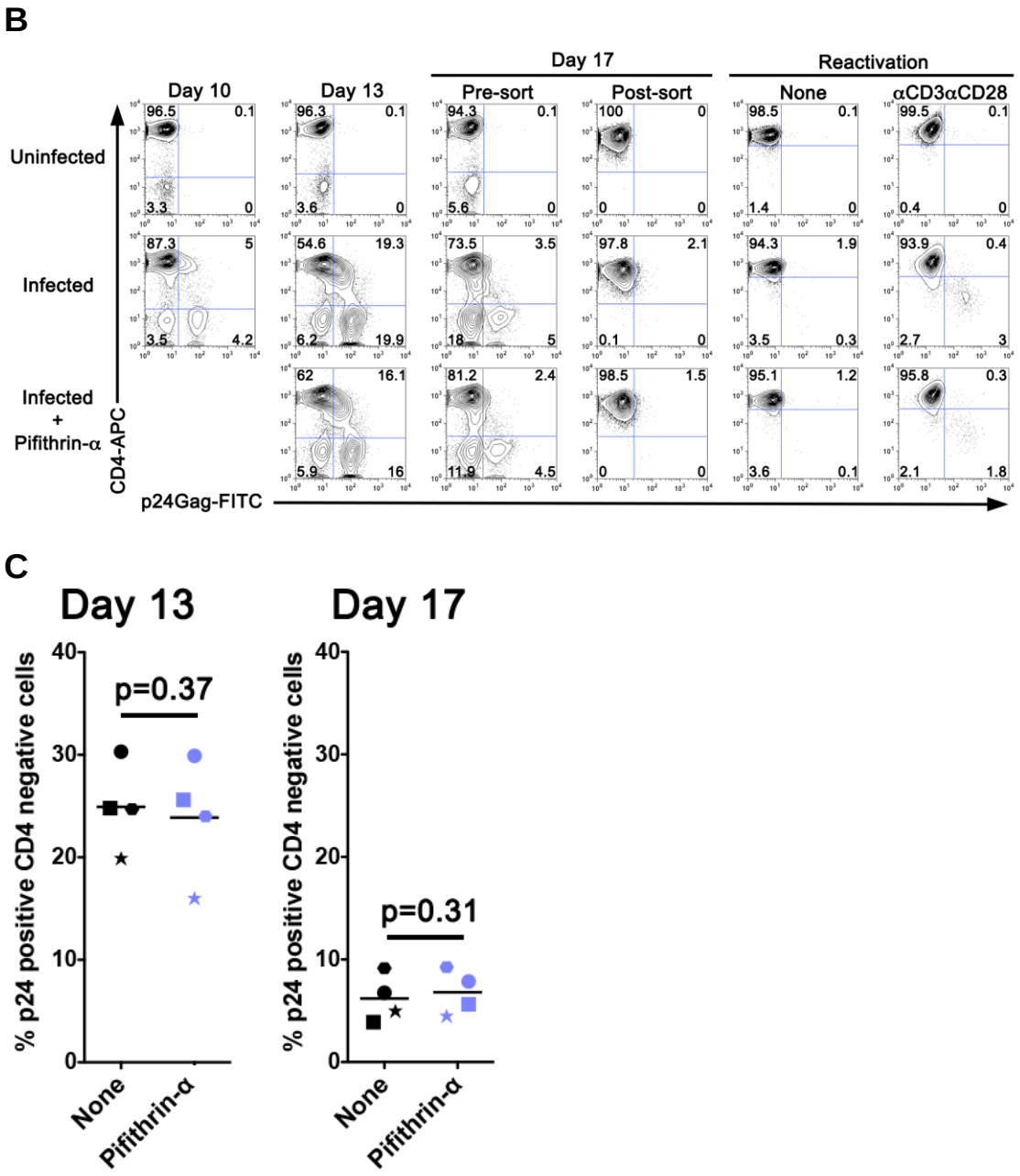

Figure 3.9: Effect of pifithrin-α upon the establishment of latency, continued.

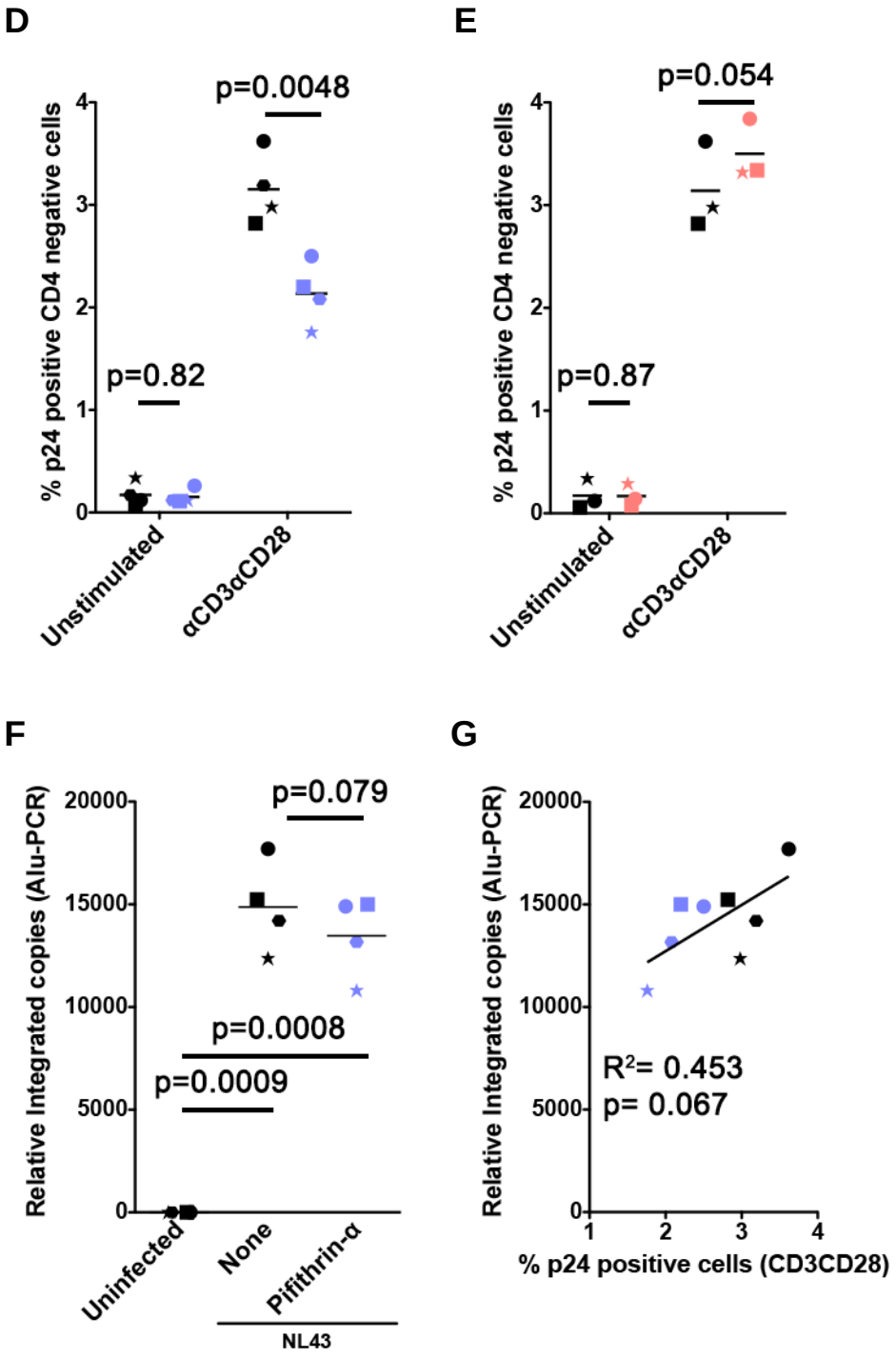

Figure 3.9: Effect of pifithrin-α upon the establishment of latency, continued.

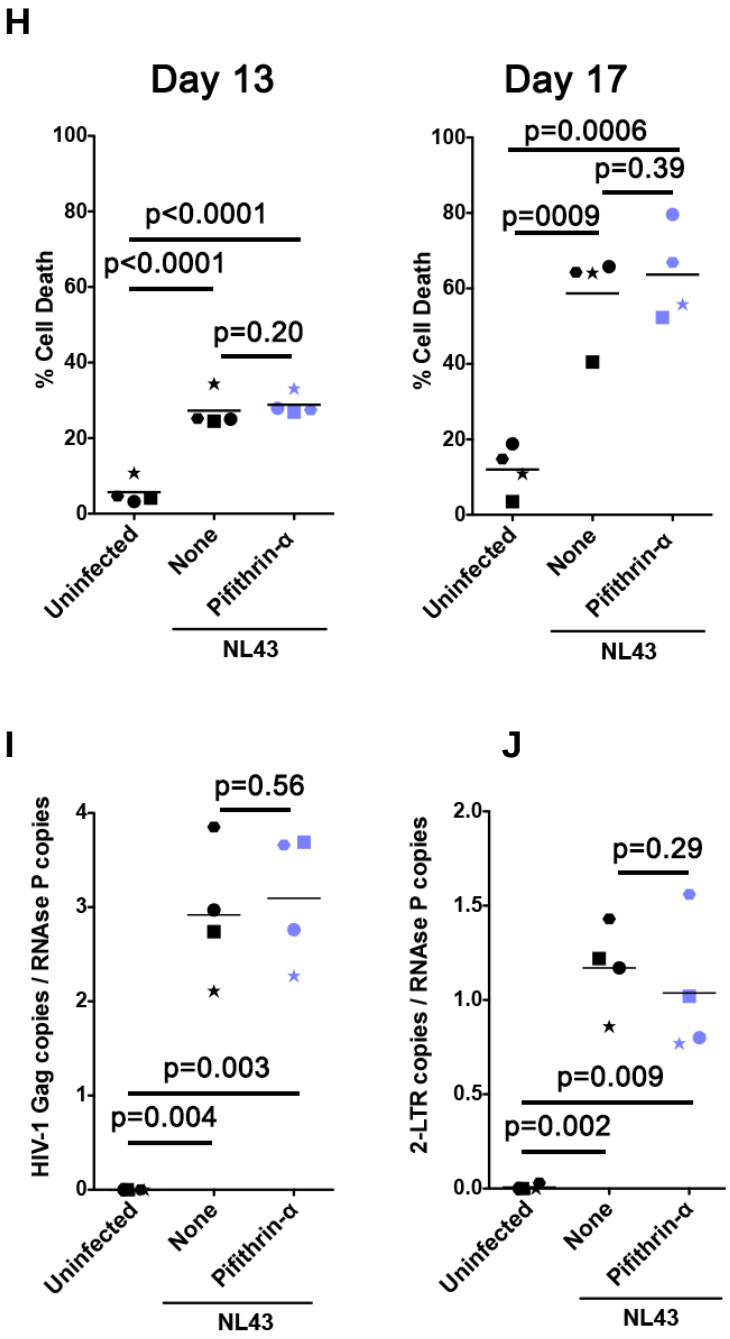

Figure 3.9: Effect of pifithrin-α upon the establishment of latency, continued.

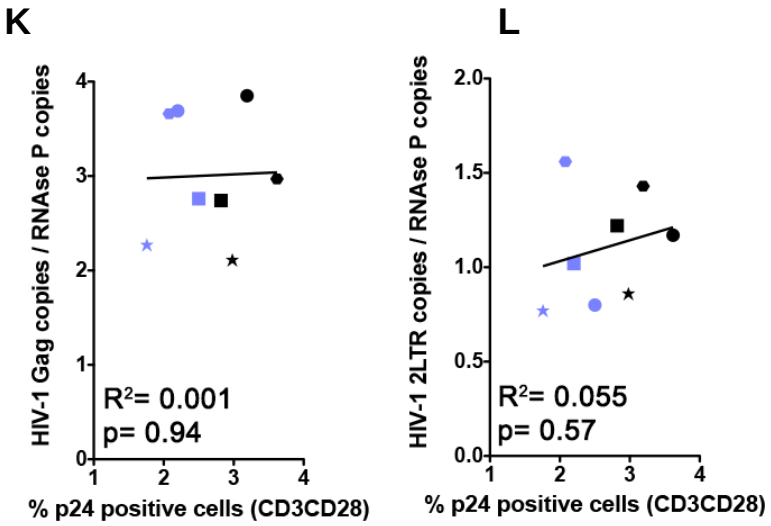

Figure 3.9: Effect of pifithrin-α upon the establishment of latency, continued.

3.7 Tables

Table 3.1: Overlapping upregulated genes between two primary CD4 T-cells models of HIV latency. This table demonstrates genes which overlap in significance and direction between the T_{CM} model, Martins *et al*. [128], and the model examined in Iglesias-Ussel *et al.* [161]. Asterisks indicate involvement with p53 signaling.

Upregulated				Downregulated	
ABHD4	*DDB2**	*KIF9-AS1*	*RGL1*	*ATF4*	*NAA15*
*ACTA2**	*DNAJC1*	*MAN2A2*	*RNFT1*	*BOP1*	*NOL6*
AEN	*DRAM1**	*MICAL1*	*RPL39L*	*C10ORF2*	*NOP16*
ALDH4A1	*FDXR**	*MPZL1*	*RTN4RL1*	*C1QBP*	*PDIA5*
APOBEC3H	*FOXP4*	*MSRB2*	*SAMD4A*	*C2CD2L*	*PIGW*
ASTN2	*GADD45A**	*NBPF14*	*SIDT2*	*CROCCP2*	*POGLUT1*
ATF7IP2	*GAMT*	*NEAT1*	*SLC22A23*	*GNL3*	*PTMA*
AZIN2	*GBP2*	*NR1H3*	*SMIM3*	*HDGFRP3*	*RORC*
BAX	*GNA15*	*NRBP2*	*SNTA1*	*HNRNPDL*	*RSL1D1*
*BBC3**	*GSS*	*PCBP4*	*SPATA18*	*HSP90AB1*	*SFPQ*
C11orf24	*GSTM3*	*PDE7B*	*SYTL3*	*IFRD2*	*STT3A*
C12orf5	*GZMA*	*PHLDA3*	*TM7SF3*	*ING5*	*TRIB2*
C7orf50	*IDNK*	*PHPT1*	*TNFRSF10B**	*KLF2*	*TSR1*
CACNB3	*IFI44*	*PMEPA1*	*TRIAP1*	*MEPCE*	*TWISTNB*
CD70	*IFIT2*	*PSMG3-AS1*	*TRIB1*	*METTL1*	*UBE2S*
CELSR2	*IL18RAP*	*RAB15*	*VWCE*	*MIDN*	*USP46*
CKLF				*MYC*	*ZNF335*

Table 3.2: Genes responsive to p53 activity. This table shows genes involved in p53 activity that are dysregulated during latency identified in the PIN. Highlighted in bold are the central hub genes from the PIN. Fold change was calculated as LI/UI and determined to be significant with FDR corrected p-value (q-value) < 0.05.

Symbol	Full Gene Name	FC	q-value	Citation
MDM2	**MDM2 proto-oncogene, E3 Ubiquitin Protein Ligase**	**1.503**	**8.688E-06**	**Uniprot [122]**
BCL3	B-Cell CLL/Lymphoma 3	1.517	4.338E-02	Kashatus *et al.* 2006 [204]
HEXIM1	Hexamethylene Bis-Acetamide Inducible 1	-1.449	3.251E-05	Lew *et al.* 2012 [207]
PTK2	Protein Tyrosine Kinase 2	3.060	1.748E-07	Uniprot [122]
RHOC	Ras Homolog Family Member C	1.494	2.900E-04	Croft *et al.* 2011 [192]
UBTD1	Ubiquitin Domain Containing 1	1.698	1.893E-03	Zhang *et al.* 2015 [218]
NRP2	Neuropilin 2	-1.464	1.234E-03	Arakawa *et al.* 2005 [219]
AR	**Androgen Receptor**	**-1.453**	**1.796E-02**	**Alimirah *et al.* 2007 [186]**
ACTA2	Actin, Alpha 2, Smooth Muscle, Aorta	2.033	1.946E-08	Comer *et al.* 1998 [190]
AREG	Amphiregulin	-1.439	4.697E-02	Taira *et al.* 2014 [220]
FAS	Fas Cell Surface Death Receptor	1.455	1.534E-05	Kim *et al.* 1999 [221]
FHL2	Four and a Half LIM Domains 2	1.803	3.223E-03	Xu *et al.* 2014 [222]
GADD45A	Growth Arrest and DNA-Damage-Inducible, Alpha	1.924	1.054E-10	Jin *et al.* 2003 [193]
NR4A1	Nuclear Receptor Subfamily 4, Group A, Member 1	-1.656	3.151E-02	Zhao *et al.* (2006) [208]
TP53I3	Tumor Protein p53 Inducible Protein 3	2.039	7.704E-04	Entrez Gene Summary [196]
TNFRSF10B	**Tumor Necrosis Factor Receptor Superfamily, Member 10b**	**1.425**	**1.007E-05**	**Takimoto *et al.* (2000) [187]**
CXCR5	Chemokine (C-X-C Motif) Receptor 5	2.073	4.235E-04	Mitkin *et al.* (2015) [223]
IRF5	Interferon Regulatory Factor 5	1.548	9.405E-03	Mori *et al.* (2002) [224]
PRDM1	PR Domain Containing 1, with ZNF Domain	1.785	1.290E-06	Yan *et al.* (2015) [197]

3.8 Acknowledgements

The dissertation author and co-authors for this article in preparation for publication would like to thank the following organizations for supporting this work: the Martin Delaney CARE Collaboratory (AI096113), the Bill and Melinda Gates Foundation through the Grand Challenges Explorations initiative (OPP1045955), the Genomics and Sequencing Core at the UCSD Center for AIDS Research (AI36214), the San Diego Veterans Medical Research Foundation, the Department of Veterans Affairs (VA), the James G. Pendleton Foundation, the Bioinformatics Core at the University of Southampton, as well as the NIH grants (AI08019 and AI104282) and VA grant BX001160. The authors also acknowledge the use of the IRIDIS High Performance Computing Facility and associated support services at the University of Southampton. The views expressed in this chapter and article are those of the dissertation author and authors of the article in preparation for publication respectively, and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government. The sponsors of this research were not involved in the study design, collection or interpretation of the data, manuscript preparation, or the decision to submit the article for publication.

This chapter is adapted from **White, C.H.,** Moesker, B., Beliakova-Bethell, N., Martins, L., Spina, C.A., Margolis, D.M., Richman, D.D., Planelles, V., Bosque, A., Woelk, C.H., *Transcriptomic analysis implicates the p53 signaling pathway in the establishment of HIV-1 latency in central memory CD4 T-cells*, in preparation for submission. The dissertation author was the primary author on this paper and responsible for designing the study, developing the RNA-Seq analysis pipeline, statistical analysis of data, interpreting the results, and writing the manuscript.

Chapter 4

Mixed Effects of Suberoylanilide hydroxamic acid (SAHA) on the Host Transcriptome and Proteome and their Implications for HIV Reactivation from Latency

4.1 Abstract

Suberoylanilide hydroxamic acid (SAHA) has been assessed in clinical trials as part of a “Shock and Kill” strategy to cure HIV infected individuals. While this drug was effective at inducing expression of HIV RNA (“shock”), treatment with SAHA did not result in a reduction of reservoir size (“kill”). A combined analysis of the effects of SAHA was used on the host transcriptome and proteome to dissect its mechanisms of action that may explain its limited success in “Shock and Kill” strategies. CD4 T-cells from HIV seronegative donors were treated with 1 μM SAHA or its solvent

dimethyl sulfoxide (DMSO) for 24 hours. Protein expression and post-translational modifications were measured with iTRAQ proteomics using ultra high-precision two-dimensional liquid chromatography - tandem mass spectrometry. Gene expression was assessed by Illumina microarrays. Using the *limma* package in the R computing environment, 185 proteins, 18 phosphorylated forms, 4 acetylated forms and 2,982 genes were identified, whose expression was modulated by SAHA. A protein interaction network integrating these 4 data types identified the HIV transcriptional repressor *HMGA1* to be upregulated by SAHA at the transcript, protein, and acetylated protein levels. Further functional category assessment of proteins and genes modulated by SAHA identified gene ontology terms related to NFκB signaling, protein folding and autophagy, which are all relevant to HIV reactivation. In summary, SAHA modulated numerous host cell transcripts, proteins and post-translational modifications of proteins, which would be expected to have very mixed effects on the induction of HIV-specific transcription and protein function. Proteome profiling highlighted a number of potential counter-regulatory effects of SAHA with respect to viral induction, which transcriptome profiling alone would not have identified. These observations could lead to a more informed selection and design of other HDACi with a more refined targeting profile, and prioritization of latency reversing agents of other classes to be used in combination with SAHA to achieve more potent induction of HIV expression.

4.2 Introduction

The persistent cellular reservoir of HIV provirus is a major obstacle to a cure [225]. The “Shock and Kill” treatment strategy has been envisioned as a controlled induction of virus reactivation in the presence of combination antiretroviral therapy

(cART) to reveal latently infected cells for immune system recognition and destruction [226]. The histone deacetylase (HDAC) inhibitor (HDACi) suberoylanilide hydroxamic acid (SAHA), an FDA-approved compound for treatment of cutaneous T-cell lymphoma [227], has been used in clinical trials to reactivate HIV to reduce the size of the latent reservoir [44, 58, 90]. Exposure to HDACis is tightly associated with histone hyperacetylation and chromatin decondensation, which provides a transcriptionally favorable environment for HIV reactivation [228]. SAHA was effective at inducing HIV RNA expression in most patients on cART with suppressed viremia [44, 58]; however, treatment with SAHA did not result in a reduction of reservoir size [58, 90]. SAHA is also used as a synergistic agent to screen *in vitro* for other latency reversing agents (LRAs); therefore, limitations of its activity require further elucidation. Understanding potential counter-regulatory effects of SAHA on HIV reactivation will guide the selection of modifications of this compound and prioritization of LRAs in combination therapies.

SAHA was well tolerated in HIV infected patients [44, 58, 90], and an *in vitro* treatment of primary CD4 T-cells with a physiological concentration of SAHA elicited only modest effects on gene expression [229]. However, genes induced by SAHA may specifically regulate the state of HIV latency [230], so that the net effect of SAHA on these genes results in insufficient viral induction to kill a cell. The function of non-histone targets of HDACs (e.g. chaperone protein HSP90) may also be modulated by SAHA [231]. SAHA binds not only HDACs, but other proteins as well [232], opening the possibility for direct regulation of additional targets. A non-histone effect of SAHA relevant to HIV reactivation was previously demonstrated by the Peterlin group [233]. In this case, SAHA promoted HIV reactivation by causing the release of positive transcription elongation factor (p-TEFb) from its inactive complex, which is required

for Tat-mediated transcriptional elongation. Ultimately, multiple steps in the HIV replication cycle must be successfully completed to reveal the infected cell to the immune system. These include cell signaling leading to proper assembly of transcription factors on the long terminal repeat (LTR), transcription, RNA splicing, RNA nuclear export, protein translation, and membrane trafficking. Systems-wide studies would enhance our understanding of complex effects of SAHA on key cellular pathways and processes required for HIV reactivation.

RNA expression profiling by microarrays and RNA-Seq technology have been the foremost strategies for identifying genome-wide effects of a disease or a treatment. Studies using SAHA demonstrated downregulation of a subset of genes [68, 229, 234], which is consistent with the existence of the secondary mechanisms of action and cannot be explained by chromatin decondensation. Transcriptomic methods are sensitive and capable of detecting the majority of the annotated genes; however, gene expression studies do not uncover the effects at the functional (protein) level. Liquid chromatography - mass spectrometry (LC-MS) based proteomics methods may be used to confirm functionality of transcripts. In addition, despite currently being less sensitive than transcriptomics, LC-MS proteomics can identify the endophenotypic effects not otherwise reflected in the transcriptome including the occurrence of post-translationally modified proteins. Therefore non-targeted quantitative iTRAQ proteomics experiments were performed by ultra-high precision two-dimensional LC-MS using human primary CD4 T-cells treated with SAHA. By combining proteomic and transcriptomic datasets, integrated data analysis was performed for a more complete characterization of the secondary effects of SAHA. Based on protein function established in published literature, it is proposed

that some of the observed effects of SAHA may have relevance to HIV reactivation, i.e. enhance or inhibit HIV transcription.

4.3 Materials and Methods

4.3.1 CD4 T-cell isolation and SAHA treatment

Healthy donor volunteers provided written informed consent using a protocol approved by UCSD IRB. Primary CD4 T-cells were isolated and cultured as described previously [229]. All CD4 T-cell samples had > 98% purity and < 5% activation (HLA-DR+), as assessed by flow cytometry. Prior to treatment, cell concentration was adjusted to 2.5 million per mL with fresh medium. Cells were treated with 1 µM SAHA or its solvent dimethyl sulfoxide (DMSO) and plated into 6-well tissue culture plates at 2 mL/well. Following 24 hours of treatment, 10-15 million cells were collected after washing 4 times with 50 mL phosphate buffered saline to remove all traces of serum proteins. Cell pellets were frozen in dry ice/ethanol bath and stored at -80^{o}C until protein isolation. A separate sample set was treated to validate gene expression by a method independent of high throughput profiling, droplet digital polymerase chain reaction (ddPCR) [[235] and Supplementary Methods (**Section 4.8**)].

4.3.2 Proteomic and transcriptomic datasets

Protein was isolated and liquid chromatography - tandem mass spectrometry was performed as described previously [236, 237, 238]. Briefly, 100 µg of protein from each sample was extracted, reduced, alkylated, and proteolysed with trypsin. Peptides were labeled with iTRAQ 8-plex, pooled, and subjected to two dimensions of liquid chromatography, and were characterized with nano-capillary ultra-performance

liquid chromatography hyphenated with a nanospray ionization hybrid LTQ / FT-Obitrap Elite ultra-high resolution mass spectrometry system. Unprocessed raw data files were searched by Proteome Discoverer for native, phosphorylated and acetylated peptides at a peptide false discovery rate of < 1% against the human Uniprot proteome. For more details on protein preparation and proteomics, please refer to Supplementary Methods (**Section 4.8**). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [239] via the PRIDE partner repository (PXD002150). Transcriptomic data that assessed the effects of SAHA (1 µM) in CD4 T-cells was obtained from the previously published SAHA dose responsive Illumina microarray dataset [240] at the Gene Expression Omnibus (GSE66994).

4.3.3 Statistical analyses

To identify detected proteins, a statistical processing approach was used that accounts for key mass spectral features to reduce the effects of peptide co-isolation on the resulting iTRAQ reporter ions and thus increase the accuracy of relative protein expression. Proteins were median-normalized and converted to $\log_2$ paired ratios (SAHA/DMSO). Protein expression values were obtained by averaging peptide intensity values and ratio-filtered. Filtering was not performed on acetylated or phosphorylated forms due to their intrinsic low abundance. Determination of differentially expressed proteins (DEPs), their phosphorylated (DPPs) and acetylated (DAPs) forms, and genes (DEGs) was done using *limma* [241]. Genes were further filtered on fold change ($|\log_2 FC| > 1$) for the protein interaction network (PIN). The PIN was constructed using Metacore^{TM} from GeneGo, Inc. and visualized with Cytoscape [242]. Node colour was subdivided into sections using the

MultiColoredNodes package [243]. Gene Ontology (GO) analysis was performed using Functional Analysis of Individual Microarray Expression (FAIME) [244]. Gene membership for each GO term was determined with BioMart [245] using the Ensemble 78 Genes database and the GRCh38 Dataset [246]. GO term differential expression between the SAHA and DMSO control conditions was determined using a paired Student's *t*-test. In a discovery driven approach, proteins and protein GO terms with a nominal p-value (p) < 0.05 were considered significant. For genes and gene GO terms, a false discovery rate-corrected p-value (FDR) < 0.05 [247] was considered significant. Genes and proteins are referred to by official gene symbol except where noted. Please refer to the Supplementary Methods (**Section 4.8**) for details of data analysis procedures.

4.4 Results

4.4.1 Proteins and genes modulated by SAHA

The quantitative proteomics profiled 1,547 proteins, identifying 185 DEPs, 18 DPPs and 4 DAPs (p < 0.05) between CD4 T-cells from 4 donors treated with SAHA or the DMSO control (**Supplementary File "White_Table_S4-1_Differentially_expressed_proteins_and_their_post-translationally_modified_forms.xlsx"**). The identification of DPPs and DAPs was possible thanks to the in-depth and orthogonal two-dimensional liquid chromatographic separation of the tryptic peptides followed by their ultra-high resolution mass spectrometry. However, their non-targeted detection suggests that these *in vivo* modified proteins had higher abundance relative to other *in vivo* modified proteins not detected in this study that typically require prior enrichment for their analysis [236]. To compare the effect of SAHA treatment between the proteome and transcriptome, microarray gene expression data were

selected from a previous SAHA dose responsive study [240]. A paired analysis identified 2,982 genes modulated by SAHA (FDR = 0.05) in CD4 T-cells from 6 donors (**See Supplementary File White_Table_S4-2_Differentially_expressed_genes**.xlsx for the complete list of DEGs at the probe level). The modulation of a large number of these genes was confirmed at the protein level with 56 up- and 49 downregulated at both levels (**Figure 4.1)**. Even though an order of magnitude more DEGs were identified compared to DEPs, there was near complete agreement in the direction of modulation by SAHA when overlapped at the RNA and protein levels.

4.4.2 PINs for proteins and genes modulated by SAHA

It was hypothesized that the most important genes modulated by SAHA would be affected at several different levels (i.e., DEGs, DEPs, DPPs and DAPs). To integrate the 4 datasets, they were superimposed onto a PIN (**Figures 4.2 and 4.3**). For visualization purposes, only genes and proteins with 5 or more connections are presented in **Figure 4.2**, whereas the complete PIN is presented in **Figure 4.3.** These PINs revealed that high mobility group (HMG) AT-hook 1 (*HMGA1*) was upregulated at the RNA (DEG), protein (DEP) and acetylated protein (DAP) levels. Heat shock protein 70 (HSP70) was represented in the PIN by 2 genes, *HSPA1A*, upregulated at the protein level, and *HSPA2*, upregulated at the RNA level. V-ets avian erythroblastosis virus E26 oncogene homolog 1 (*ETS1*) was downregulated by SAHA at the protein level, including total and phosphorylated (pS294) forms. A number of well-connected genes were modulated at the protein level only, e.g. lymphoid enhancer-binding factor 1 (*LEF1*), lysine (K)-specific demethylase 1A (*KDM1A*), and inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase

beta (*IKBKB*). Another set of well-connected genes was regulated by SAHA only at the RNA level, e.g. v-MYC avian myelocytomatosis viral oncogene homolog (*MYC*), enhancer of zeste homolog 2 (*EZH2*), activator protein 1 (*AP-1*), and nuclear receptor coactivator 3 (*NCOA3*). While well-connected genes modulated at the protein level by SAHA may have regulatory roles for other genes, the role of genes modulated by SAHA at the RNA level would need further confirmation by more sensitive methods of protein detection.

4.4.3 Functional analysis of proteins and genes modulated by SAHA

To better understand the biological processes modulated by the well-connected proteins in the PIN, DEPs and DEGs were subjected to GO analysis using FAIME [244]. An order of magnitude fewer proteins was detected by quantitative proteomics compared to genes detected by microarrays. Therefore, protein representation across the GO terms was assessed. A good correlation (R^2=0.96) was observed between the number of genes and proteins mapping to all the GO terms (**Figure 4.4**). The exceptions were the GO terms *Plasma membrane proteins* and *Integral components of membrane*, which were under-represented for proteins, consistent with their hydrophobic properties and low abundance that result in poor extraction [248]. Overall, 1,484 GO terms were upregulated and 935 downregulated by SAHA at the RNA level. At the protein level, 119 GO terms were upregulated and 146 downregulated. There was good overlap in GO terms (N=40 up and N=20 down) that were significantly modulated by SAHA at the RNA and protein levels (**Figure 4.5 and Supplementary File White_Table_S4-3_Gene_ontology_terms_modulated_by_SAHA.xlsx)**. Terms related to chromatin regulation, *Negative regulation of chromatin silencing* and *Nuclear euchromatin*, were upregulated by SAHA (**Figure**

4.6). Terms related to histone acetylation and histone acetyltransferase complex were downregulated, as was observed previously with a lower dose of SAHA [229]. *Positive regulation of T cell proliferation* and *Positive regulation of T cell activation* were downregulated. Importantly, a number of terms functionally relevant to HIV replication were identified. *Regulation of I-kappaB kinase/NF-kappaB signaling*, *Protein binding involved in protein folding*, *Chaperone mediated protein folding requiring cofactor*, and *Autophagic vacuole* were upregulated by SAHA.

4.4.4 Validation of gene expression by ddPCR

Six donors, different from the ones who participated in profiling studies, were recruited. Three genes, whose expression was modulated by SAHA both at the RNA and protein levels, were independently validated using ddPCR. Two of the selected genes were upregulated (*HMGA1* and *ASF1A*) and one downregulated (*AES*) by SAHA. All three genes were significantly modulated by SAHA as determined by ddPCR (**Figure 4.7**), in the same direction as in the microarray and quantitative proteomics studies.

4.5 Discussion

4.5.1 SAHA transcriptional and post-transcriptional regulation

Quantitative proteomics identified 185 proteins significantly modulated by SAHA. Over half of these proteins (56%) appear to be regulated at the transcriptional level since their corresponding transcripts were also modulated by SAHA (**Figure 4.1**). The remaining proteins were not modulated at the RNA level and would not have been detected in a transcriptomics approach, demonstrating the added value of a proteomics approach. Modulation of a protein may be a result of the function of

another protein whose transcript was upregulated. For example, upregulation of proteins required for translation (e.g. translation initiation factor *EIF5B* [249]), may result in an increase of translation from existing messenger RNAs [250]. Protein expression may also be regulated post-translationally via activity of other proteins modulated by SAHA. For example, 3 proteins that regulate ubiquitination state were modulated by SAHA: E3 ubiquitin ligase *DTX3L*, E2 ubiquiting-conjugating enzyme *UBE2H*, and a deubiquitinase *USP13*. It is also possible that the changes in protein expression were the result of earlier transient changes in gene expression, which were not captured in the present study. In addition, even though enrichment for post-translationally modified peptides was not performed, changes in expression of 4 acetylated and 18 phosphorylated proteins after SAHA treatment were detected. Altogether, these data are consistent with the idea that SAHA may have much broader secondary effects beyond chromatin modification than previously recognized.

Among the detected 1,547 proteins, 260 were differentially expressed only at the RNA level, and not at the protein level, consistent with the idea that transcriptional effects of SAHA do not always translate into protein production [251]. However, this conclusion should be interpreted with caution due to the smaller sample size used in the proteomics part of the present study (N=4 for the protein vs. N=6 for gene expression analyses). Since cells from different biological donors were used for proteomics and transcriptomics studies, it is also possible that some of the variation between identified DEPs and DEGs was the result of donor-to-donor differences in response to SAHA. However, donor-to-donor variation did not likely play a large role since whenever proteins and transcripts were both detected as differentially expressed, they were modulated in the same direction (**Figure 4.1**), and expression

of selected genes was confirmed by an independent method in an independent cohort (**Figure 4.7**).

4.5.2 Known effects of SAHA translate from the RNA to the protein level

Identification of GO terms related to previously recognized effects of SAHA at the RNA and protein levels gives confidence that the chosen proteomics methodology provides reliable data. For example, upregulation of genes encoding histones, but downregulation of genes encoding components of the acetyltransferase complexes was observed previously by transcriptomics [229]. Downregulation of acetyltransferases suggests potential mechanisms by which cells attempt to regain control of acetylation following removal by SAHA of their ability to control acetylation through HDACs. Another process known to be downregulated by SAHA was T-cell activation [240, 251]. Both *Positive regulation of T cell activation*, and *Positive regulation of T cell proliferation*, were downregulated by SAHA at the RNA and protein levels (**Figure 4.6**).

4.5.3 Effects of SAHA on RNA and proteins with a role in HIV reactivation

A number of HMG proteins (*HMGA1*, *HMGN1*, *HMG20B, LEF1*) were modulated by SAHA at the protein and/or RNA levels. Like histones, HMG proteins regulate chromatin dynamics, dependent on post-translational modifications [252]. The most remarkable observation, made possible by using integrated proteomics and transcriptomics data, was the upregulation of *HMGA1* at the RNA and protein levels (**Figure 4.2**), as well as a consistent upregulation of its acetylated form (**Figure 4.8**). Two mechanisms by which *HMGA1* interferes with HIV transcription have been demonstrated. First, it competes with *Tat* for TAR binding and inhibits both basal and

Tat-mediated HIV transcription [253]. Second, *HMGA1* inhibits transcription of host genes, as well as HIV, by recruiting inactive p-TEFb to target promoters [254]. Even though the observed increase of the *HMGA1* protein was relatively small in the present study (1.15-fold), experimental overexpression (20% increase) [253] resulted in a measurable reduction of LTR activity proportionate to levels of *HMGA1* expression. Thus, upregulation of *HMGA1* by SAHA would appear to have an undesirable negative effect with respect to HIV reactivation. The *HMGA1* protein possesses 5 lysine residues that can be acetylated [255]. Of these, acK64 and acK70 have a known function in interferon-beta transcriptional switch in response to viral infection [256]. In the present study, the acK14 form of *HMGA1* was upregulated by SAHA, for which no specific function has yet been demonstrated.

Identification of *HMGA1* (**Figure 4.2**) prompted a more in-depth analysis of individual proteins that were modulated by SAHA. A literature search on DEPs and "HIV" or "HIV latency" was performed to determine whether any other proteins modulated by SAHA, besides *HMGA1*, have a role in activation or repression of HIV transcription (**Table 4.1**). Three of the proteins that were found in this search were modulated by SAHA at the RNA level as well, and were confirmed by ddPCR (**Figure 4.7**). More than half of the identified proteins were not modulated by SAHA at the RNA level. Change in RNA may be transient for some genes, as was noted by Elliott and colleagues [58]. For example, *BRD2* was detected in their study at the RNA level 2 hours post-treatment, while in the present study it was detected at the protein, but not the RNA level, 24 hours post-treatment. Thus, a proteomics approach has added value to the transcriptomic approach by capturing some of the transient effects on genes at the protein level.

GO terms, which may be relevant to HIV reactivation and were modulated by SAHA at the RNA and protein levels, included *Regulation of I-kappaB kinase/NF-kappaB signaling, Protein binding involved in protein folding*, *Chaperone mediated protein folding requiring cofactor,* and *Autophagic vacuole* (**Figure 4.6**). NFκB signaling is well recognized in HIV transcriptional activation [257, 258]. Interestingly, individual genes and proteins significantly upregulated by SAHA and mapping to *Regulation of I-kappaB kinase/NF-kappaB signaling* had opposite effects on NFκB activity. For example, *TNFA*, *F2RL1*/PAR2 (DEGs) and *HSPB1* (DEP) activate NFκB [258, 259, 260] and thus promote HIV reactivation, while *ZFAND6* (DEG) represses NFκB [261]. In addition, *IKBKB* (DEP) phosphorylates the inhibitor in the inhibitor/NFκB complex [262], causing dissociation of the inhibitor and activation of NFκB. Its downregulation by SAHA would thus have a negative effect with respect to HIV reactivation. Protein folding may have a role in HIV reactivation, because sequential actions of HSP70 and HSP90 are required for proper folding and stabilization of cyclin-dependent kinase 9 (CDK9) and assembly of p-TEFb [263]. Recently, the role of autophagy in HDACi-induced HIV reactivation and clearance of infected cells has become of interest. In monocyte-derived macrophages, intracellular HIV was shown to be degraded via canonical autophagy pathway upon reactivation with HDACis [264]. Lysosomal destabilization following HDACi treatment promoted death of HIV infected cells, even with incomplete activation of HIV [265]. These results indicate the importance of autophagy when using SAHA, and other HDACis, for HIV reactivation.

4.5.4 Conclusions and implications

The present study is the first to the authors' knowledge to identify proteins and their post-translationally modified forms modulated by SAHA in human primary CD4 T-cells. Combined with the analysis of induced transcriptomic changes, this study demonstrates global regulatory networks affected by SAHA treatment and enhances the understanding of the secondary mechanisms of SAHA action. Expression of a number of genes and proteins with previously reported roles in HIV transcriptional control was modulated by SAHA; some of these effects would appear to be inhibitory for HIV reactivation. Identification of these counter-regulatory effects of SAHA on HIV induction have the potential to strongly impact selection and modification of HDACis and prioritization of other LRAs for future evaluations and advancement to clinical trials. Better potencies for HIV reactivation of the HDACis Romidepsin [266] and Panobinostat [267] may be due to the lower impact of secondary negative effects possessed by SAHA, which warrants further investigation. Synergistic HIV reactivation when using SAHA with Protein Kinase C (PKC) agonists, such as prostratin and bryostatin [65, 268], may be due to negating adverse effects of SAHA on NFκB signaling pathway by these LRAs. Prostratin and bryostatin are not suitable for use *in vivo* due to side effects or limited availability; however, other PKCs, such as Ingenol derivatives [269, 270], warrant further testing in combination with SAHA using cells from HIV infected patients *ex vivo*. Proteome profiling performed in this study revealed a number of potential counter-regulatory effects of SAHA not present at the transcript level. It is therefore recommend using transcriptomic and proteomic profiling as two complementary techniques, transcript profiling being a more sensitive method, and protein profiling for confirmation of transcripts and detection of protein-specific effects.

4.6 Figures

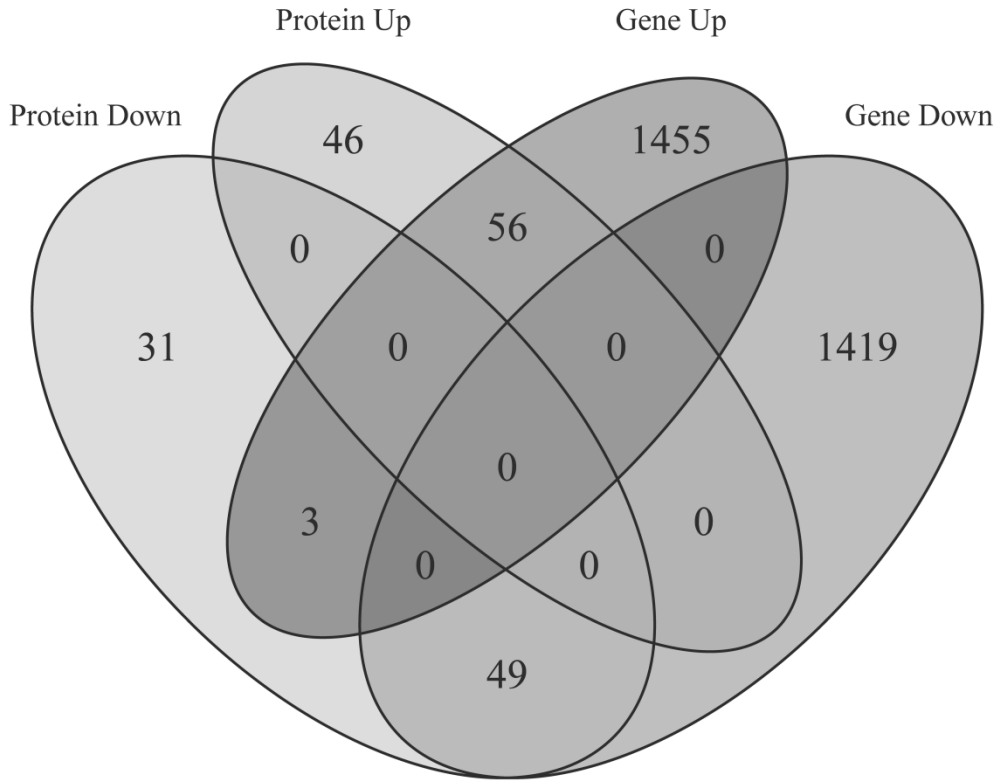

Figure 4.1: Overlap among differentially expressed proteins and genes. DEPs ($p < 0.05$) and DEGs (FDR = 0.05) were identified using linear modeling in R (package limma). The venn diagram was constructed using the VennDiagram package in R. Up- and downregulated genes and proteins are shown.

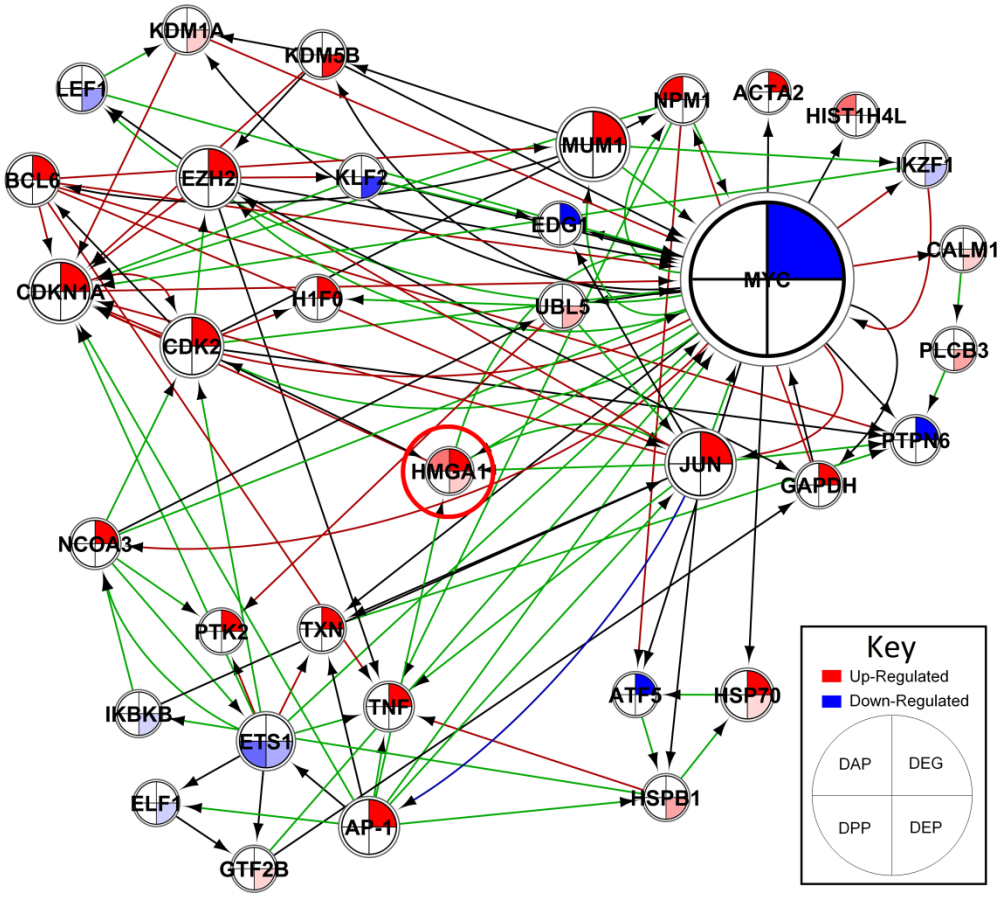

Figure 4.2: Protein interaction network (PIN) for combined DEPs, DPPs, DAPs and DEGs. The PIN was constructed using MetaCore^{TM}, and visualized using Cytoscape. All 185 DEPs, 18 DPPs, 4 DAPs and 368 DEGs (after filtering using FDR = 0.05 and |log2 FC| > 1) were included, removing redundancies. Nodes were color coded according to fold changes ($log_2FC = -1$ to $log_2FC = 1$), in 4 sections corresponding to DEGs, DEPs, DPPs and DAPs (as indicated by the key). Only nodes that had 5 or more connections to other DEPs or DEGs are shown, while nodes with fewer connections were hidden to improve the quality of the image. Several well-connected DEPs and DEGs in the PIN represent transcription factors with a recognized role in HIV transcriptional control. Green and red lines refer to positive and negative regulation, respectively, whereas black lines depict unspecified effects. The red circle highlights HMGA1.

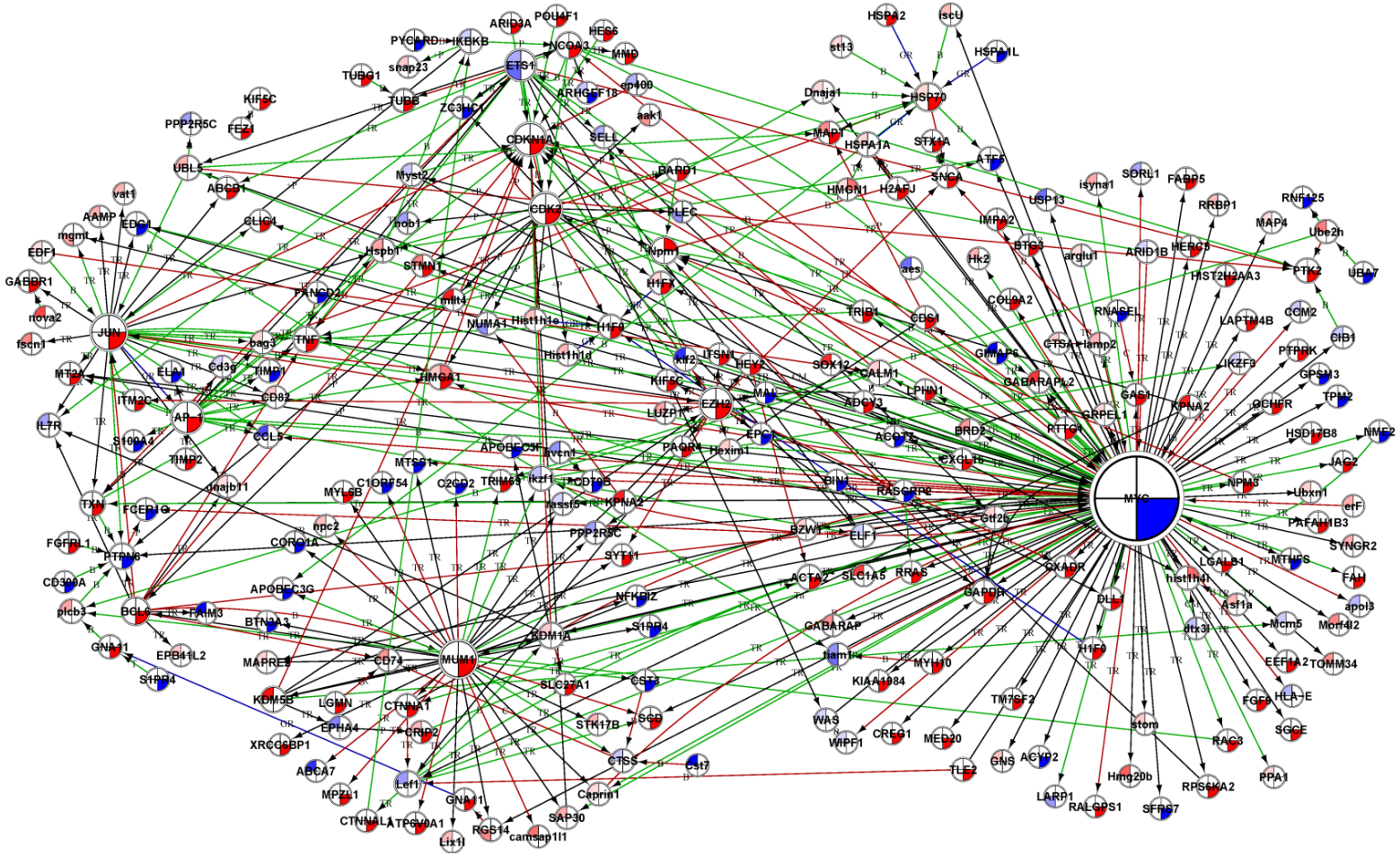

Figure 4.3: Expanded PIN. This image contains all nodes with at least one connection to another node. All DEPs, DPPs, DAPs, and DEGs were filtered using FDR corrected p-values < 0.05 and $|\log_2 FC| > 1$. Edge abbreviations: B – binding, C – cleavage, CM – covalent modifications, +p – phosphorylation, -p – dephosphorylation, T – transformation, TR – transcription regulation, GR - group relation, CS – complex subunit.

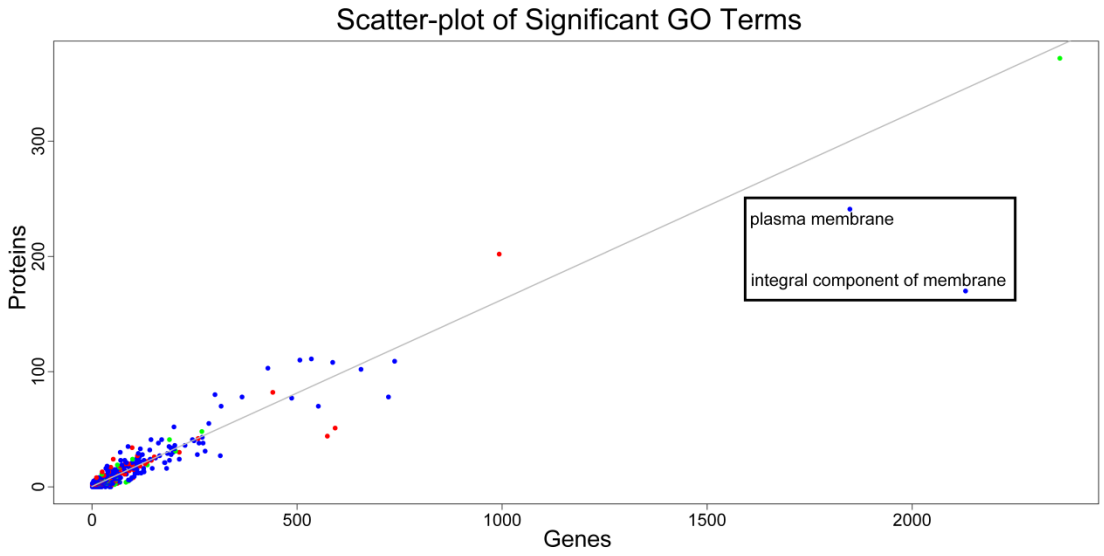

Figure 4.4: Scatter-plot of significant GO terms. The line through the middle is the regression line relating the number of detected proteins to the number of detected genes (y = 0.1622x + 0.3106). This correlation function indicates that for every 10 genes, approximately 2 proteins are expected to map to each GO term. If a GO term is located above the regression line, then it is over-represented for detected proteins; conversely, if a GO term is located underneath the regression line, it is under-represented. Blue, GO terms significant for genes only; Red, GO terms significant for proteins only; Green, GO terms significant for both genes and proteins. GO terms under-represented for detected proteins, “plasma membrane” and “integral component of membrane”, are highlighted in the boxed region.

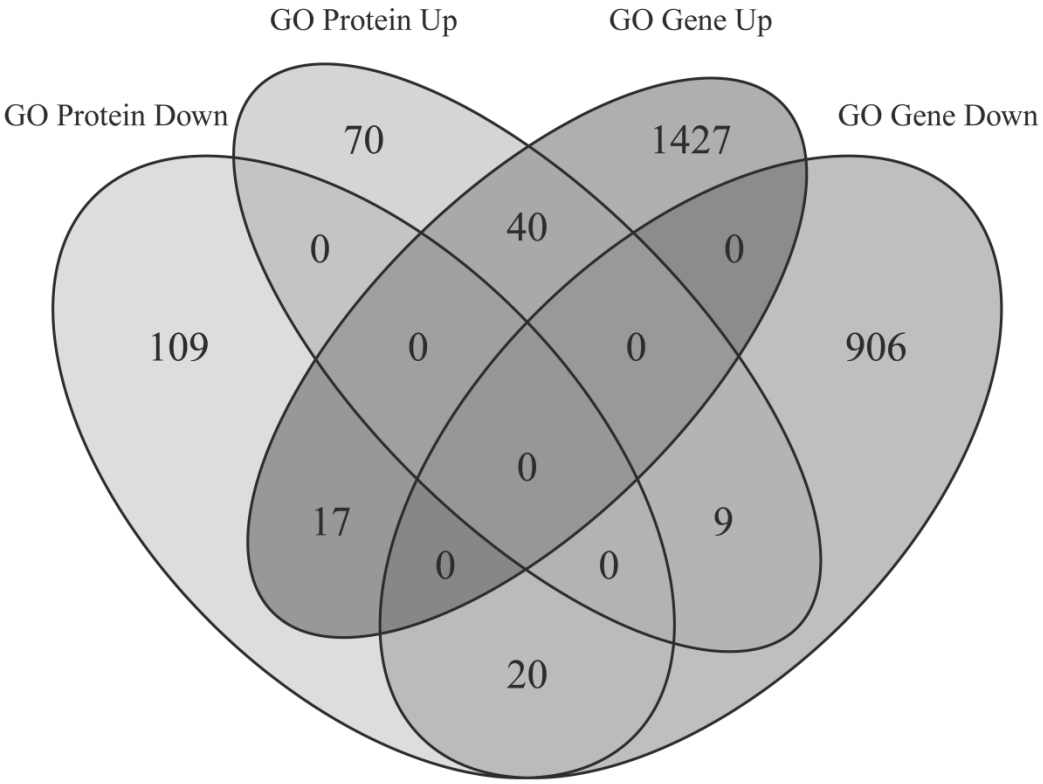

Figure 4.5: Overlap among GO terms modulated by SAHA at the RNA and protein levels. GO terms modulated by SAHA were identified using FAIME. The venn diagram was constructed using the VennDiagram package in R. Up- and down-regulated GO terms are shown.

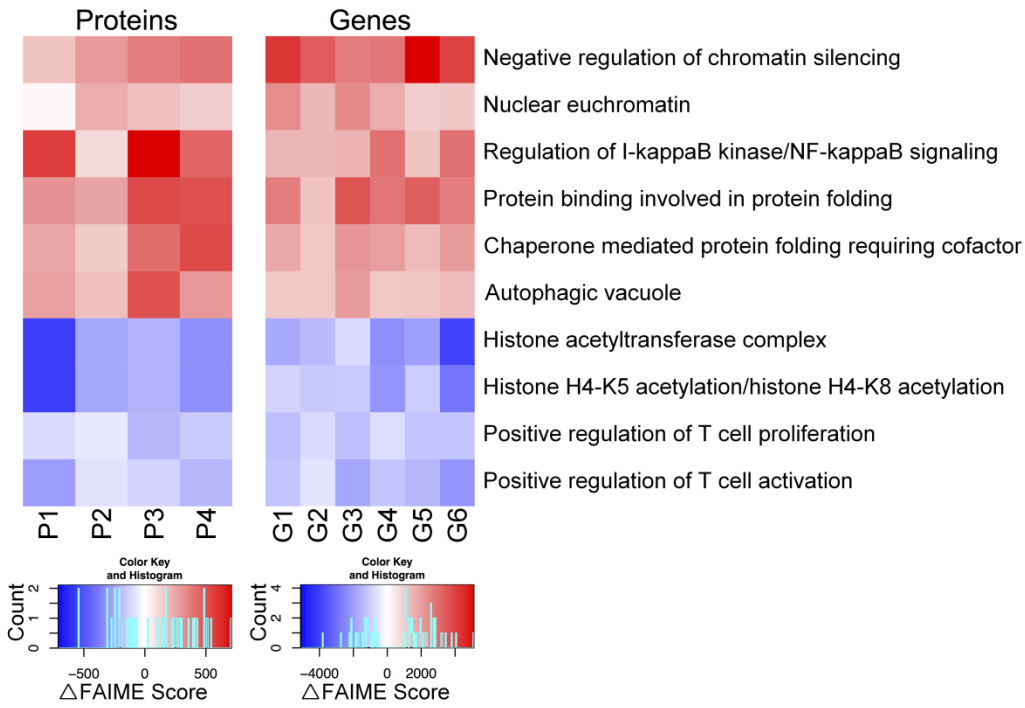

Figure 4.6: GO terms significantly modulated by SAHA. GO terms modulated by SAHA were identified using FAIME. Terms related to known effects of SAHA and effects of SAHA relevant to HIV reactivation are shown. The heatmap represents values obtained by subtraction of a GO term average FAIME score for DMSO control from a GO term average FAIME score for SAHA (ΔFAIME Score). Red, the difference in FAIME score is greater than 0, and GO term is upregulated as the result of SAHA treatment. Blue, the difference in FAIME score is less than 0, and GO term is downregulated as the result of SAHA treatment. Count refers to the number of cells in the heatmap with the indicated difference in FAIME scores between SAHA and DMSO controls. P1 through P4 indicate samples used for protein analysis; G1 through G6 indicate samples used for gene expression analysis. The same genes and proteins were represented in the GO terms Histone H4-K5 acetylation and Histone H4-K8 acetylation, so these terms are depicted by a single row on the heatmap.

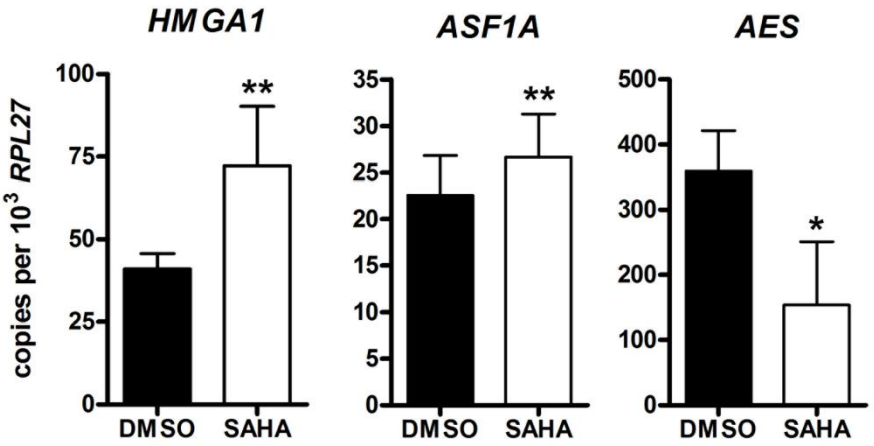

Figure 4.7: Validation of gene expression by droplet digital PCR (ddPCR). The number of molecules of each target mRNA was normalized to the number of molecules of the normalizer mRNA (RPL27) in one nanogram of total RNA (expressed as copies per thousand RPL27 molecules). Normality of the distribution and equality of variance in the DMSO and SAHA treated groups were assessed in the R computing environment, and either t-test (*HMGA1* and *ASF1A*) or Wilcoxon signed rank test (AES) were performed to assess the difference of expression induced by SAHA. The experiment was performed with cells from 6 independent donors. Error bars represent standard deviation. $^{**}p < 0.01$; $^{*}0.01 < p < 0.05$.

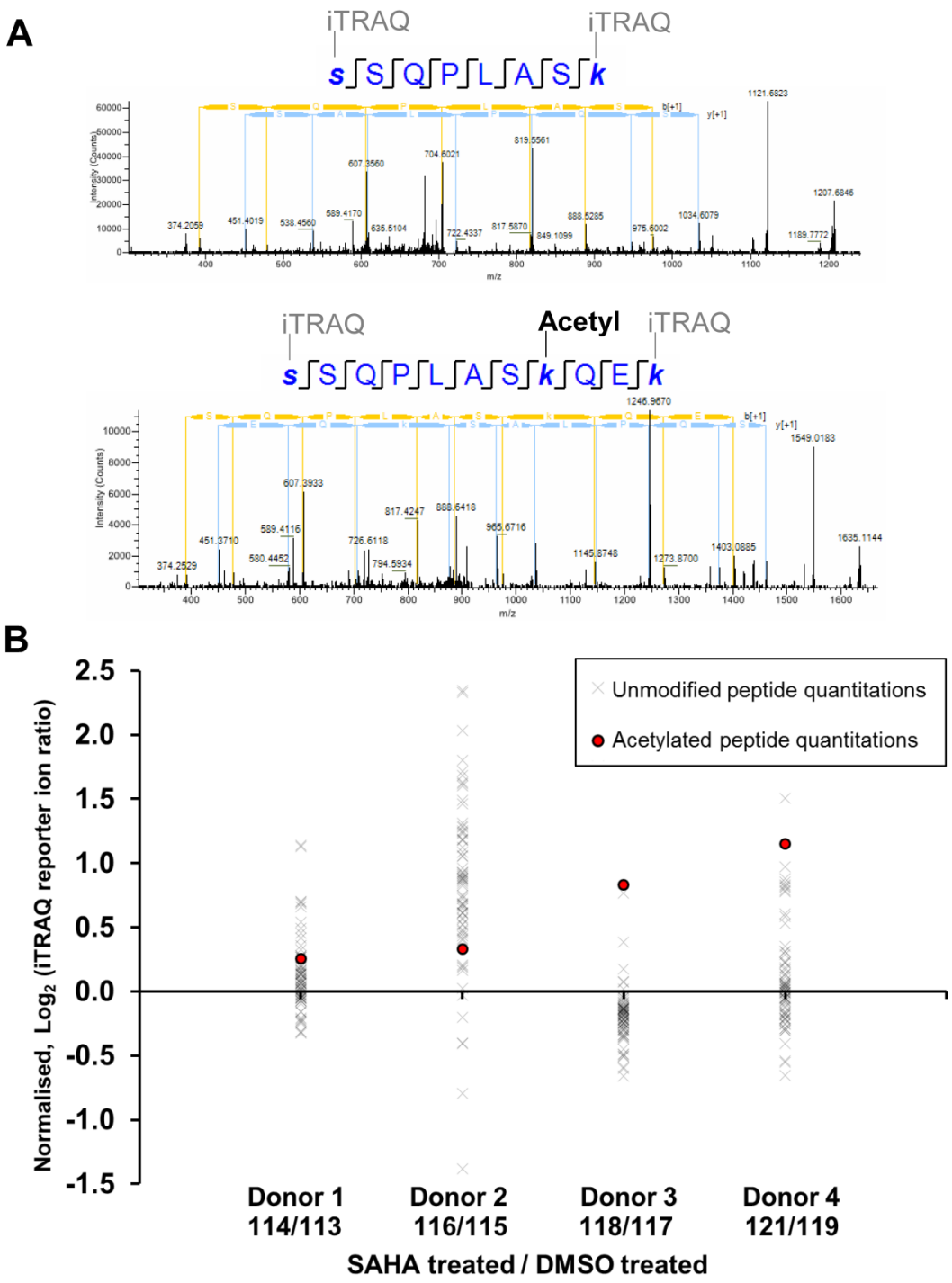

Figure 4.8: The mass spectrometry characterization of the effect of SAHA on *HMGA1* expression. (A) Spectra demonstrating the identification of unmodified and acetylated peptides uniquely matching *HMGA1*. (B) Relative expression ratios detected by iTRAQ for all detected peptide spectrum matches to *HMGA1*. Ratios demonstrate the differential expression induced by SAHA treatment across the 4 donors, relative to DMSO control treatment.

4.7 Tables

Table 4.1: Proteins with a known role in HIV reactivation, up- or downregulated as the result of SAHA treatment. An implied effect of SAHA treatment on HIV transcription, based on the function of a given protein and the direction of its modulation by SAHA, is indicated in the rightmost column. Plus sign, possible positive effect on HIV transcription, e.g., resulting from upregulation of an activator or downregulation of a repressor. Minus sign, possible negative effect on HIV transcription, e.g., resulting from downregulation of an activator or upregulation of a repressor. Asterisk (*) indicates that for the heat shock protein 70, one gene (HSPA1A) was modulated by SAHA at the protein level only, while another gene (HSPA2) was modulated only at the transcript level.

Symbol	Gene name	Reference(s)	Transcript modulated	Effect on HIV transcription
HSPA1A	Heat shock 70 kDa protein 1A	O'Keeffe *et al*. 2000 [263]	No/yes*	+
KDM1A	Lysine (K)-specific demethylase 1A	Sakane *et al*. 2011 [271]	No	+
ETS1	v-ets avian erythroblastosis virus E26 oncogene homolog 1	Bassuk *et al*. 1997 [272], Posada *et al*. 2000 [273], Sieweke *et al*. 1998 [274], and Yang *et al*. 2009 [275]	No	−
LEF1	Lymphoid enhancer-binding factor 1	Sheridan *et al*. 1995 [276]	No	−
IKBKB	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	Mercurio *et al*. 1997 [277] Nabel *et al*. 1987 [257]	No	−
HMGA1	High mobility group AT-hook 1	Eilebrecht *et al*. 2014 [254] Eilebrecht *et al*. 2013 [253]	Yes	−
ASF1A	Anti-silencing function 1A histone chaperone	Gallastegui *et al*. 2011 [278]	Yes	−
BRD2	Bromodomain containing 2	*Boehm et al*. 2013 [50]	No	−
BRDT	Bromodomain, testis-specific	*Bisgrove et al*. 2007 [279]	No	−
AES	Amino-terminal enhancer of split	*Tetsuka et al*. 2000 [280]	Yes	+
ARID1B	AT rich interactive domain 1B (SWI1-like)	Mahmoudi 2012 [281] Rafati *et al*. 2011 [282]	No	+

4.8 Supplemental Methods

4.8.1 Protein isolation and iTRAQ labeling

Proteins were isolated from SAHA or DMSO treated samples for 4 donors (N=8) as described previously [236, 237, 238]. Cells were lysed in 0.5 M triethylammonium bicarbonate (TEAB, Sigma-Aldrich) and 0.05% sodium dodecyl sulfate (SDS) with pulse probe sonication. Protein concentration was determined using bicinchoninic acid (BCA) assay (Thermo-Pierce). One hundred micrograms of protein from each sample were reduced with 50mM tris (2-carboxyethyl) phosphine (TCEP) for 1 hour at 60°C and alkylated with 200mM methyl methanethiosulfonate (MMTS) for 15 minutes at room temperature, as per the manufacturer's instructions (ABSciex). Proteins were digested in the dark at room temperature overnight with 3 µg proteomics grade trypsin (Roche). Peptides were labeled with isobaric tags for relative and absolute quantitation (iTRAQ) 8-plex, as per the manufacturer's instructions. Peptides were assigned to the following labels: 113; Donor 1 DMSO, 114; Donor 1 SAHA, 115; Donor 2 DMSO, 116; Donor 2 SAHA, 117; Donor 3 DMSO, 118; Donor 3 SAHA, 119; Donor 4 DMSO, 121; Donor 4 SAHA. Reactions were terminated with 8 µl of 5% hydroxylamine (Sigma-Aldrich) for 15 minutes at room temperature.

4.8.2 Peptide pre-fractionation with high-pH reverse-phase C8 chromatography

Labeled peptides were lyophilized and serially reconstituted in 100 µl of 2% acetonitrile (ACN, Fisher), 0.1% NH4OH (Sigma-Aldrich). Peptides were resolved using a C8 column (150 x 3 mm (inner diameter (ID)), 3.5 µm particle size) (XBridge, Waters) at 300 µl/min with a LC-20AD system (Shimazdu). Peptide fractions were collected in a peak-dependent manner over a 100 minute, high pH gradient: Mobile

phase A (aqueous); 0.1% NH4OH, mobile phase B (organic); 99.9% ACN, 0.1% NH4OH (0-10 mins; 2% B, 10-20 mins 5% B, 20-90 mins; 5-85% B, 90-100 mins; 85% B). All other conditions were as described previously [237].

4.8.3 Low pH Ultra HPLC-nano electrospray ionization-Orbitrap MS

Approximately 1 µg of peptides were loaded by a Dionex Ultimate 3000 (Thermo Fisher Scientific) at 20 µl/min for 4 minutes onto a C18 PepMap100 trapping cartridge (5mm × 500 µm ID, 5 µm particle) (Thermo Fisher Scientific) in mobile phase of 2% ACN, 0.1% formic acid (FA). Seventy peptide fractions were eluted at 300 nl/min over a gradient of 2-35% (105 mins) and 35-85 % (25 mins) organic phase (95% ACN, 5% DMSO, 0.1% FA) in aqueous phase (2% ACN, 5% DMSO, 0.1% FA). Peptides were resolved on an Acclaim PepMap 100 column (C18 75 µm × 50 cm, 3 µm particle) retrofitted to a PicoTip nESI emitter (New Objective). Electrospray ionization was conducted at 2.4 kV and precursor ions were characterized with an Obitrap Elite –Velos pro mass spectrometer (Thermo Fisher Scientific) at 120,000 mass resolution. The top 15 +2 and +3 precursor ions per MS scan (minimum intensity 1000) were characterized by HCD (30,000 mass resolution, 1.2 Da isolation window, 40 keV normalized collision energy) and CID (ion trap MS, 2 Da isolation window, 35 keV) with a dynamic exclusion (± 5 ppm) of 200 seconds.

4.8.4 MS data processing and peptide normalization

Target-decoy searching was performed with Proteome Discoverer version 1.4.1.14 software using SequestHT, with Percolator used to estimate false discovery rate (FDR) with a threshold of ≤ 0.01. Spectra were searched for fully tryptic peptides allowing 1 missed cleavage, a precursor tolerance of 10 ppm and a minimum peptide

length of 6. Spectra were first searched allowing for a single variable modification of oxidation (M), iTRAQ (Y) or deamidation (N,Q), with Methylthio (C) and iTRAQ (K and N-terminus) set as fixed modifications. Spectra not reaching a match of < 1% FDR were then searched for iTRAQ (K), oxidation (M), acetylation (K) and phosphorylation (S,T,Y) as variable modifications (maximum of 2), with methylthio (C) and iTRAQ (N-terminus) as fixed modifications. Fragment ion mass tolerances of 0.02 Da for the FT-acquired HCD spectra and 0.5 Da for the CID spectra. Spectra were searched against the UniProtKB SwissProt human proteome (March 2014). Reporter ions were extracted with a tolerance of 20 ppm. Peptide filtering was performed to remove any peptides not unique to a single protein. The peptide level data was then normalized through median normalization.

4.8.5 Ratio testing using the modified permutation method

The permutation method adopted from Nguyen and colleagues [283] has the benefit of not being reliant on the data having any particular distribution, but rather on the availability of sufficient data to represent general protein expression. 1000 iterations were used because real p-values are estimated at this level [284]. This method is described below.

1). Set S as the sample of all peptide $\log_2$ ratios from the experiment.

2). Set the number of permutations as B (1000 for this experiment).

3). For protein *j* in the number of proteins from one to M:

a) Calculate the number of peptides n_j of protein *j*.

b) Calculate the protein log ratio $\boldsymbol{L_j} = \sum_{i=1}^{\boldsymbol{n_j}} \frac{\boldsymbol{l_i}}{\boldsymbol{n_j}}$ for protein *j*.

c) Randomly sample without replacement n_i ratios from set S to

obtain a subset of ratios equal in size to number of peptides in protein *j*.

d) Calculate the random peptide log-ratio L_R of this subset as $\boldsymbol{L_R = \sum_{i=1}^{n_j} \frac{l_i}{n_j}}$.

e) Estimate the p-value for protein *j* by evaluating $\boldsymbol{p_{value} = \frac{\sum_{r=1}^{B} I\{|L_j| > |L_r|\} + 1}{B+1}}$ where $\boldsymbol{I\{|L_j| > |L_r|\} = 1}$ if $\boldsymbol{|L_j| > |L_r|}$ is true and $\boldsymbol{I\{|L_j| > |L_r|\} = 0}$ otherwise.

i) One is added to both numerator and denominator to avoid a p-value of zero and a $\log_{10}$(p-value) of infinity. The lowest p-value possible will be on the order of 1/B.

4). Store p-values for each protein and output the ratios and corresponding p-values.

Any protein ratios that did not have a p-value of 0.05 in at least half of the samples (2 samples in this experiment) were filtered out.

4.8.6 Group testing with limma

The package *limma* [241] is available from Bioconductor in R. Proteins which did not pass ratio testing were filtered to reduce the number of false positives. Peptide level data was normalized using median normalization and filtered to remove any peptides which were not unique to a single protein. This peptide level data was then $\log_2$ transformed and averaged to give one averaged $\log_2$ intensity value for each protein. Gene expression intensity values from the microarrays were used at the probe level. *Limma* was used to fit a linear model to each protein or gene, and a contrast representing the difference between the DMSO and SAHA conditions was

used to generate the empirical Bayes moderated *t*-statistic. This statistic tests if the contrast is equal to 0 indicating no difference between the DMSO and SAHA conditions for a protein or a gene.

4.8.7 Gene ontology (GO) analysis using FAIME

The FAIME algorithm was described previously [244]. Briefly, gene expression was ranked in descending order for each sample according to an exponential weighting scheme. Let $\boldsymbol{r_{g,s}}$ be the expression rank for every gene $\boldsymbol{g} \in \boldsymbol{G}$ within the sample set $\boldsymbol{s}$. Let $|\boldsymbol{G}|$ denote the total number of genes present in the dataset. The weighted rank $\boldsymbol{w_{g,s}}$ may be calculated according to equation 1.

$$\boldsymbol{w_{g,s}} = \boldsymbol{r_{g,s}} * \boldsymbol{e}^{\frac{\boldsymbol{r_{g,s}}}{|\boldsymbol{G}|}} \quad (1)$$

The normalized centroid ($\boldsymbol{NC}$) of this set was then calculated by summing the weighted expression of gene elements in the gene set. For every GO term, there is a gene subset denoted as $\boldsymbol{GO_i}$ whose genes satisfy $\boldsymbol{g} \in \boldsymbol{GO_i}$ and a complementary set $\boldsymbol{G/GO_i}$ of genes present in the experiment, but not found within the GO term of interest. The normalized centroid of each set is calculated according to equations 2 and 3.

$$\boldsymbol{NC(GO_i, s)} = \frac{\boldsymbol{1}}{\boldsymbol{|GO_i|}} \sum_{\boldsymbol{g \in GO_i}} \boldsymbol{w_{g,s}} \quad (2)$$

$$\boldsymbol{NC(G/GO_i, s)} = \frac{\boldsymbol{1}}{\boldsymbol{|GO_i|}} \sum_{\boldsymbol{g \in G/GO_i}} \boldsymbol{w_{g,s}} \quad (3)$$

Finally, a FAIME Score is calculated by subtracting equation 3 from equation 2 (equation 4).

$$\boldsymbol{F_{GO_i,s}} = \boldsymbol{NC(GO_i, s)} - \boldsymbol{NC(G/GO_i, s)} \quad (4)$$

The FAIME score is calculated for gene and protein results separately to generate a gene and protein FAIME score for every GO term. FAIME scores were

then filtered based upon a standard deviation filter. Any results which had less than a 10% relative standard deviation (equation 5) across all samples in the experiment were removed. Any term for which only one gene or protein was present was also removed to reduce false positives.

$$\%\boldsymbol{RSD} = \frac{\boldsymbol{Stdev}}{\boldsymbol{Mean}} * \mathbf{100} \quad (5)$$

FAIME Score differences were calculated from the significant results by averaging results in each group and subtracting the DMSO control score from the SAHA score. Heatmaps of the FAIME difference scores were generated using the *gplots* [285] package in R.

4.8.8 Droplet digital polymerase chain reaction (ddPCR)

Cells were treated with DMSO or 1 µM SAHA for 24 hours. RNA was isolated using Qiagen RNeasy kit. Three genes, whose expression was modulated by SAHA both at the RNA and the protein levels (*HMGA1*, *ASF1A*, *AES*), were selected for validation by ddPCR. *RPL27*, whose expression was not responsive to SAHA treatment [229], was used as a normalizer. The following TaqMan Gene Expression assays were used: Hs00852949_g1 for *HMGA1*, Hs01011627_m1 for *ASF1A*, Hs01081013_g1 for *AES*, and HS03044961_g1 for *RPL27*. One nanogram of target RNA was used for each reaction and performed in duplicate. RNA was reverse transcribed and cDNA amplified using ddPCR Supermix for Probes (No dUTP) (Bio-Rad, Inc., Hercules, CA). Droplet formation was performed using the QX-100 Droplet Generator (BioRad Inc., Hercules, CA) following the manufacturer's instructions. Droplets were transferred to a 96-well PCR plate which were heat-sealed with foil and amplified to endpoint using the following cycle: 95^{o}C 10 minutes (1 cycle), 94^{o}C 30 seconds and 60^{o}C 1 minute (45 cycles), 98^{o}C 10 minutes (1 cycle), 4^{o}C until droplet

reading. Droplets were read using QX100 Droplet Reader and analyzed using the QuantaSoft software (Bio-Rad, Inc., Hercules, CA) that determines concentration of target cDNA as copies per microliter from the fraction of positive droplets using Poisson statistics, from which the total number of copies per reaction can be obtained. The number of copies of each mRNA molecule was normalized to the number of *RPL27* copies by calculating mRNA per one thousand copies of *RPL27*. Statistical analyses described below were performed in the R computing environment. The normality of the distribution in each DMSO- and SAHA- treated group for each gene was verified using the Shapiro test (*shapiro.test* function). The equal variance test was performed using the function *var.test*. Based on the results of these tests, *t*-test with unequal variance was performed for *HMGA1*, *t*-test with equal variance was performed for *ASF1A*, and Wilcoxon signed rank test was performed for *AES* (*t.test* and *wilcox.test* functions). All tests were one-sided tests due to knowledge about the direction of change of expression induced by SAHA from the high throughput experiments. Plots were produced using GraphPad Prism version 4 for Windows, GraphPad Software, L Jolla California USA, www.graphpad.com.

4.9 Acknowledgements

The dissertation author and co-authors for this published work gratefully acknowledge all volunteers at UCSD, who contributed blood, and Merck and Co., Inc., who graciously provided SAHA used in this study. The authors are indebted to Mr. Roger Allsopp and Mr. Derek Coates for their enthusiasm, fund raising and vision in promoting the FT-MS proteomics platform at the University of Southampton. Thanks to Dr. Xunli Zhang for the kind use of the high-performance liquid chromatography system. The authors further acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work. We thank the PRIDE team for the proteomics data processing and repository assistance. This research was supported by the grants from the National Institute of Health, Martin Delaney Collaboratory of AIDS Researchers for Eradication (CARE, U19 AI 096113), through the research infrastructure provided by the UCSD Center for AIDS Research (CFAR, AI 36214) and the James B. Pendleton Charitable Trust. The Wessex Cancer Trust and Medical Research, UK, Hope for Guernsey and the University of Southampton 'Annual Adventures in Research' Grant, made it possible to establish the proteomics infrastructure and its use for this study. This material is based upon work supported in part by the Department of Veterans Affairs (VA), Veterans Health Administration, Office of Research and Development. The views expressed in this chapter and published article are those of the dissertation author and authors of the published work respectively, and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government. The sponsors of this research were not involved in the study design, collection or interpretation of the data, manuscript preparation, or the decision to submit the article for publication.

This chapter is adapted from **White, C.H.,** Johnston, H.E., Moesker, B., Manousopoulou, A., Margolis, D.M., Richman, D.D., Spina, C.A., Garbis, S.D., Woelk, C.H., Beliakova-Bethell, N. *Mixed effects of suberoylanilide hydroxamic acid (SAHA) on the host transcriptome and proteome and their implications for HIV reactivation from latency. Antiviral Research. (2015)*. The dissertation author was the primary author on this paper and responsible for developing the proteomics pipeline, developing the program used for proteomics analysis (modified permutation method), statistical analysis of all transcriptomic and proteomic data, GO and pathway analysis with FAIME, and writing the manuscript.

Chapter 5

Dysregulation of Human Endogenous Retroviruses in Primary CD4 T-Cells upon Treatment with SAHA

5.1 Abstract

Current strategies for treatment of HIV with anti-retroviral therapy (ART) have proven extremely effective at suppressing viral replication and prolonging the lifespan of HIV infected individuals. However, ART has proven ineffective at eradicating the latent HIV reservoir, i.e. persistently infected cells that can replicate HIV following ART interruption. One method for the elimination of this reservoir is the "Shock and Kill" strategy, where HIV is activated with histone deacetylase inhibitors (HDACis) such as SAHA (a.k.a. Vorinostat) to induce virus expression in these latently infected cells and then the activated cell either dies or is eliminated by the immune system. The method by which HDACis activate HIV could conceivably upregulate HERVs and may lead to unwanted, or potentially dangerous off-target effects. In this study, RNA-Seq was utilized to examine the transcription of 98,000 HERV elements curated in

the database HERVd in cells treated with a high dose of the HDACi SAHA. It was shown that a number of HERV elements (~1,000) were upregulated upon treatment with a potential specificity of SAHA for the LTR12 family of HERV elements. These findings suggested that unwanted off-target effects of HDACis upon HERV elements should be considered if using SAHA to reactivate the latent HIV provirus.

5.2 Introduction

Anti-retroviral therapy (ART) has proven to be extremely effective at suppressing HIV within infected individuals, and has also drastically decreased HIV related morbidities and mortality [18]. However, a persistent reservoir of cells remains infected despite therapy [19]. A goal of current cure strategies is the elimination of this latently infected reservoir through a "Shock and Kill" approach [44, 286] (**Chapter 1, Section 1.4**). In this approach, compounds, such as histone deacetylase inhibitors (HDACis), are used to reactivate latent HIV, while relying upon ART to prevent the spread of infection. The infected cell then either dies on its own, or the immune system, combined with potential priming agents, kills the latently infected cell, thereby clearing the latently infected reservoir.

The primary mechanism of action of HDACis, such as SAHA, is to inhibit histone deacetylase (HDAC) enzymes which remove an acetyl group from histones, thereby allowing the chromatin they organize to achieve a more relaxed state [287]. This can lead to a more transcriptionally favorable environment in which HIV transcription becomes activated. A secondary mechanism of SAHA treatment may be dysregulation of genes through modulation of transcription factors or alterations of pathways after chromatin relaxation (**Chapter 4**) [91]. Both the primary and secondary mechanisms of action of SAHA may be able to modulate other genome

elements, such as human endogenous retroviruses (HERVs). Chromatin relaxation (**Chapter 1, Section 1.4**) may potentially allow HERVs to be upregulated by SAHA treatment. As this is the primary mechanisms of action of SAHA, this work focused predominately upon upregulated HERV elements.

HERVs arise from retroviral infections of germ line cells, and subsequent incorporation into the host's genome [288]. HERVs entered into the mammalian genome as early as 55 million years ago [289], and as recently as 100,000 years ago [290]. A subset of HERVs, the HERV-K (HML-2) family, comprises roughly 1% of all HERVs and some members have been active and infectious for the past 30 million years [291, 292]. Furthermore, patients with germ line cancers show specific immune reactions to HERV-K (HML-2) antigens [293, 294]. Some members of the HERV-K group code and actively produce retroviral-like particles [295]. In the past, up-regulation of HERVs has been associated with diseases such as schizophrenia, cancer, ALS, and multiple sclerosis [296, 297, 298]. Additionally, HERVs increase genome instability through non-allelic homologous recombination [299]. Finally, HERVs have been demonstrated as upregulated during HIV infection. Specifically, HERV-K (HML-2) is markedly increased after HIV infection [300, 301] due to activation by the HIV encoded trans-activator of transcription (*Tat*) protein. These potential issues of HERV dysregulation, such as increased risk of disease, genome instability, and interactions with HIV, all advocate for an analysis of HERVs upregulated upon treatment with LRAs.

To address the possibility of unwanted HERV dysregulation, Hurst and colleagues used the targeted approach, RT-qPCR, to assess the modulation of specific HERV families (i.e. HERV-K, HERV-W, and HERV-FRD) following HDACi treatment [302]. Through an analysis of the *env* and *pol* genes, this targeted

approach showed that these HERV families were not dysregulated by treatment with 1μM, or even 1mM, of SAHA. It was then concluded that HERV dysregulation was not a barrier to the use of HDACis in HIV cure strategies. While examining RT-qPCR results from portions of the HERV sequence may be appropriate for these three HERVs, it does not provide a comprehensive analysis of all HERV families. In contrast to the notion that HDACis would not dysregulate HERVs, it has been demonstrated previously in human cell lines that HERVs may indeed be upregulated upon treatment with not only SAHA, but other HDACis [303]. Specifically, RT-qPCR analysis demonstrated that LTR12 was not only induced by SAHA treatment (both 1μM and 5μM doses), but genes which had LTR12 inserted in their promoter region (e.g. *TNFRSF10B* and *TP63*) were further upregulated by activation of LTR12 with HDACis.

HERVs comprise nearly 8% of the human genome [304] and their study can be extremely difficult as they are often present in the human genome in thousands of locations as imperfect copies [305]. Therefore it is paramount to not just test for differential expression across a portion of the HERV element, but also to examine that element to determine if reads are centralized upon a particular region of the HERV or if they are spread out over its entirety. Through this, a determination can be made as to whether the HERV sequence identified should be separated into several elements, or if a portion of that sequence is truly part of the HERV based upon expression results. Due to this, RNA-Seq (**Chapter 1, Section 1.5**) is a tool best suited for this challenge as it is not dependent upon prior knowledge regarding the HERV sequence, and provides a picture of expression across the entire HERV sequence. This technology may further keep pace with updates to HERV elements as new ones are added and their sequences modified. Furthermore, as nearly 8% of

the human genome is comprised of HERVs, with over 98,000 separate elements present in various databases [306, 307], RNA-Seq provides a comprehensive analysis which other tools may lack. Due to this, RNA-Seq was selected for the unbiased study of HERV dysregulation upon SAHA treatment.

A further difficulty with the study of HERVs is the categorization of HERV sequences. Classification schemes can differ depending upon the database or algorithm utilized [305, 306, 307]. While one may decide to separate portions of the HERV into different entries, another database may keep them all as one entity. For the purposes of this work, the database HERVd [306] was used to determine HERV element sequences. Similarly, the naming method of HERVd was adopted in that the HERVs (LTR12, HERV9, HERVK, ERVL, etc.) are referred to either as HERV or HERV family in subsequent mention. The specific elements (rv_007357, rv_007420, and rv_010177) present in HERVd as separate unique entities are referred to as HERV elements from a particular HERV family. While this categorization and naming scheme are subject to debate, these were adopted for simplicity in this work. It should also be noted that sequences from HERV families can be identified in hundreds, if not thousands, of locations in the human genome across all chromosomes. This is caused by upwards of millions of years of genome editing, re-insertion of retrotransposons, imperfect sequence duplication, and mutation. As such, some elements from a family (LTR12, ERVL, HERVK, etc) will be upregulated, others from the same family will be downregulated, and some will be unaffected by drug treatment. This phenomenon is quite different from gene dysregulation and must be kept in mind when analyzing results from HERVs.

Presented are the results of an untargeted RNA-Seq screen of over 98,000 HERV elements, many of which demonstrate dysregulation upon treatment with a

high dose (10 µM) of SAHA. A substantial upregulation of several hundred HERV elements was detected, 3 of which were confirmed through droplet digital PCR (ddPCR). It is also demonstrated that multiple HERV elements were not dysregulated by SAHA, suggesting that this drug may have a specificity towards HERV elements from specific families. Finally, it is suggested that researchers studying HERVs and their interaction with drug treatments should utilize RNA-Seq due to the ability of this technology to detect transcription in an unbiased manner from anywhere in the human genome.

5.3 Materials and Methods

5.3.1 CD4 T-cell isolation and SAHA treatment

Frozen primary CD4 T-cells were purchased from AllCells, Inc. (Emeryville, CA) from four different donors and thawed in RPMI + 20% human serum. Dead cells were removed with Viahance (Biophysics Assay Laboratory Inc., Worcester, MA) resulting in populations of CD4 T-cells with > 92% purity and an average activation of 6.7% (HLA-DR+) as assessed by flow cell cytometry. Prior to treatment, cell concentration was adjusted to 2.5 million cells per mL plated into 6-well tissue culture plates at 2 mL/well. Cells from each donor were then treated with SAHA (10 µM) or left untreated (i.e. DMSO solvent only) resulting in 8 samples total. Following 24 hours of treatment, samples were washed twice with 10 mL phosphate buffered saline, resuspended in RLT Plus buffer (Qiagen, Valencia, CA) containing β-mercaptoethanol for RNA extraction.

5.3.2 RNA isolation, preparation, and extraction

Total RNA was extracted from primary CD4 T-cells using the RNeasy Plus Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions, except 1.5 volumes of absolute ethyl alcohol (EtOH) instead of an equal volume of 70% EtOH, to ensure binding of all RNAs to the column. RNA integrity was assessed using the Agilent Bioanalyzer 2100 and RNA integrity numbers (RINs) of samples were on average 8.9 (s.d. $\pm$ 0.29), which was suitable for RNA-Seq experiments. RNA-Seq libraries were prepared from total RNA using the TruSeq^{TM} Stranded Total RNA library Prep kit (Illumina, CA, USA) after depletion of cytoplasmic and mitochondrial ribosomal RNA. All libraries were sequenced to a depth of 100 million reads using the Illumina HiSeq2000 to generate 50bp paired-end library reads (See **Chapters 1 and 2** for more details regarding RNA-Seq analysis).

5.3.3 RNA-Seq data analysis

FASTQ files for each sample were mapped to a reference of all 98,000 HERV elements downloaded from the HERVd database (2012 release) [306]. As HERV sequences are present in multiple locations in the genome with several complete and partial duplications, it was decided that HERV element mapping would be done using only uniquely mapping reads. Paired ends were mapped separately with Bowtie [132] utilizing the -m 1 setting to only report mappings with a single location in the HERVd reference. Quantification was done through a custom gene transfer file which treated each HERV element as a single exon. Non-specific filtering was used to remove elements that did not achieve at least 1 CPM in half of the samples. Differential expression was performed using EdgeR [109], which uses an overdispersed Poisson model to account for both biological and technical variability.

HERV elements were identified as differentially expressed (FDR corrected p-value < 0.05). Pileup pictures were generated with the use of Bamview [308] and the Integrative Genomics Viewer (IGV) [147]. Mapping to the human genome (hg19) was done to view dysregulation of HERV elements in the UCSC genome browser [309], as the HERVd reference is not available within this tool. This mapping was performed with Tophat [107], using the default settings with the no-coverage-search option selected.

5.3.4 Assessment of HERVs for confirmation

All significant HERV elements were filtered using a $\log_2$FC > 3 (actual FC > 8) to focus results upon those highly upregulated. Read pileup pictures for these individual HERV elements were generated with IGV [147] and visually assessed to determine a set for confirmation. Poorly expressed HERV elements, and those with reads generated from only a small portion of the HERV element were removed. Additionally, extremely short sequences (< 500bp) were removed as HERV elements with larger sequences provide more regions along the element for primer and probe development. HERV elements present within genes were also excluded from consideration as their upregulation could conceivably be due to gene dysregulation instead of HERV element upregulation alone.

5.3.5 Read coverage analysis

Read coverage was created by converting all mapped bam files to wiggle files with the program Pyicos [310]. The wiggle files were generated by averaging the coverage results in a 40bp window and then uploaded to the UCSC genome browser [309]. The HERV sequences selected for confirmation were searched against the

hg19 reference using BLAT [311] (BLAST-like alignment tool) to identify their location. The UCSC table browser [312] was then used to generate a custom track for the identified HERV elements.

To confirm that the HERVs presented in Hurst *et al.* [302] were not dysregulated upon SAHA treatment, the primer and probe sequences presented in their supplemental information was used to perform a BLAT search for those same locations in the hg19 reference. Once identified, through matching both primer and probe sequences, the corresponding locations were viewed in the UCSC genome browser with tracks corresponding to the mapped RNA-Seq results generated from this study. Additionally, this location was then viewed in the RepeatMasker track [307] to confirm the presence of the HERV referred to in Hurst *et al*. [302]. For some HERVs, multiple locations had equal sequence matching scores as determined by a BLAT search. One representative location in hg19 was chosen for illustration of HERV upregulation, while other locations of equal matching score were confirmed to have a similar expression levels (data not shown).

5.3.6 Droplet digital PCR (ddPCR) analysis

Custom TaqMan qPCR Gene Expression Assays were used in a ddPCR approach to assess differential expression as previously described [92]. Briefly, the QX100 ddPCR system (Bio-Rad, Hercules, CA) was utilized to confirm the expression of three HERV elements, rv_007357, rv_007420, and rv_010177. In ddPCR, RNA within thousands of uniform oil emulsion droplets undergo reverse transcription and end-point PCR reaction. The number of droplets which contain template RNA are then used to calculate the total number of target RNA molecules present using Poisson statistics [313]. Five nanograms of RNA in 20 µL ddPCR reaction volume

were used for each target in duplicate. Copy numbers were pooled across duplicates, normalized to *RPL27* due to constant expression of this gene across DMSO and SAHA treated samples [314], and used to calculate fold changes between conditions. Significance testing was done with a paired *t*-test, and results were deemed significant if p-value < 0.05. A small value (0.01 copies per µL) was added to each measured HERV element at the start of the analysis to avoid division by 0 when calculating fold changes.

Primers were designed based upon read pileup locations in each of the HERV elements to be confirmed (**Table 5.1**) by obtaining a 300 bp region along the element and subjecting that region to analysis with the PrimerQuest Tool [315] available from Integrated DNA Technologies. For larger HERV elements (rv_007357 and rv_007420), two regions were chosen for confirmation in case one primer and probe set failed. One primer and probe set (rv_007420 2nd set), demonstrated amplification in the non-template negative control, which may have be due to sequence similarity of that region of the HERV element to bacterial DNA, RNA present in water, or RNA present in the ddPCR master mix itself. Another possibility is primer dimerization of the primer and probe sequence leading to amplification. This effect was not pursued further due to having another region to confirm this HERV element. For the smaller HERV element (rv_010177), only a single primer and probe set was chosen. As many HERV sequences are partially replicated across the genome in multiple locations, these primers and probes were searched using BLAT to confirm they would amplify RNA from only the HERV element of interest, and not from other copies of the element inserted elsewhere in the genome.

5.4 Results

5.4.1 Modulation of HERVs upon treatment with SAHA

RNA-Seq data was generated from eight samples (untreated and SAHA treated paired samples from 4 donors) and mapped to a reference of HERV elements from the database HERVd [306]. After mapping and quantification, a paired analysis approach identified that a total of 2101 HERV elements comprising 120 unique HERV families were dysregulated upon treatment with SAHA. Of these elements, 1095 (115 families) were upregulated and 1007 (102 families) were downregulated (**Table 5.2**). As this was a particularly high dose of SAHA, results were filtered based upon an absolute $Log_2FC > 3$ leaving 814 HERV elements (451 upregulated from 81 families and 363 downregulated from 82 families). Plotted in a histogram are the dysregulated elements from each HERV family which demonstrated that upregulation in particular is not entirely random across all HERV families, but appears to have a high specificity towards the HERV elements from the LTR12 family (**Figure 5.1**).

Upregulation of HERVs following treatment was the primary concern due to the possibility that HERV elements may be able to re-integrate into oncogenes [316], or that upregulated HERVs may be able to recombine with HIV [317, 318], hence the focus of future analyses was solely upon upregulated HERV elements. Three HERV elements differentially expressed upon SAHA treatment were selected for ddPCR confirmation (**Table 5.3**) based upon upregulation along large sections of the HERV element, a length greater than 500 bp, consistent upregulation across all donors, and the HERV element not being present within a gene which could potentially be dysregulated by SAHA.

The locations of these HERV elements were identified in the hg19 human reference genome with BLAT. Then, all RNA-Seq reads were mapped to the hg19

human reference genome. Mapped data was used to generate a wiggle file for upload to the UCSC genome browser. A custom track for these three HERV elements, along with a track for each sample, was created and examined to confirm upregulation as determined by previous differential expression analysis. The HERV elements rv_007357, rv_007420, and rv_010177 all demonstrated significant upregulation (**Figure 5.2**). Expression of these elements was then confirmed with ddPCR in the original sample set, and analyzed for dysregulation. A representative image of the DMSO and SAHA conditions is generated from the QX100 ddPCR system and shown to demonstrate raw count data which is used to obtain Log_2FC values (**Figures 5.3A and 5.3B**). These results were then averaged across donors to generate a bar plot which demonstrated significant HERV element upregulation (**Figure 5.4**). This upregulation was deemed significant by a paired *t*-test (p-value < 0.05) for each HERV element tested.

5.4.2 Dose responsive upregulation of selected LTR12 HERV elements in an independent donor set

HERV dysregulation was demonstrated at the 10 µM SAHA level, which is close to the peak serum level for an intravenous dose [319]. However, this dose is considered a high concentration and not achievable with oral dosages as used in HIV "Shock and Kill" approaches [127, 227]. Therefore, it was determined if treatment with concentrations which are more likely to be achievable in HIV infected individuals treated with SAHA, also result in the same, or lower, degree of upregulation of HERVs. To accomplish this, cells from two additional donors were collected and treated with the following doses of SAHA; 0, 0.34, 1, 3, and 10 µM. RNA was extracted and subjected to ddPCR using the same set of primers for confirmation of

the three HERV elements described previously (**Chapter 5, Section 5.3.6**). For each HERV element, $\log_2$ fold changes were calculated by dividing the normalized HERV element copies at each dose by the copies at 0 µM and then converting all results to log base 2 (**Figure 5.5A and B**). These results demonstrate a dose response to SAHA treatment for each HERV element in both patients. The increase in response to SAHA as measured by fold change appears to be quite large as the dose increases, but even at the minimum dose HERV upregulation was noted for all but one HERV element (rv_010177) in donor 1.

5.4.3 Analysis of HK2, HERV-W, and HERV-FRD expression

Over 2,000 HERV elements were noted as dysregulated (**Table 5.2**), however, the vast majority of HERV elements tested (approximately 96,000) were not dysregulated. In a targeted approach using RT-qPCR, Hurst *et al*. [302] noted that the HERVs; HK2, HERV-W, and HERV-FRD were not dysregulated upon treatment with SAHA at either the 1 µM or even the 1mM level. Therefore it was determined if RNA-Seq reads in this study mapped to these HERVs. All FASTQ files were mapped to the hg19 genome reference and converted to wiggle files for viewing purposes as previously described. The primer and probe sequences utilized in Hurst *et. al.* [302] were obtained and identified in the hg19 genome. These locations where then searched for dysregulation in the RNA-Seq results. No dysregulation was found for all four locations indicating agreement with Hurst et al. (**Figure 5.6**).

5.5 Discussion

5.5.1 Overview of results

The objective of this study was to determine the potential off-target effects on HERVs that may accompany treatment with an HDACi, such as SAHA. As this treatment is used to reactivate HIV in "Shock and Kill" strategies, it is important to consider the off-target effects this drug may have on HERVs, particularly their potential reactivation. These results demonstrated that while the vast majority of HERV elements appear to not be dysregulated by treatment with a high dose of SAHA, a smaller but significant number were heavily upregulated. HERV elements from 115 unique HERV families were upregulated upon treatment suggesting that SAHA can, at least in part, activate portions of these HERVs. Significant upregulation of these HERV elements at a $Log_2FC > 3$ indicated that SAHA may exhibit specificity towards elements from the LTR12 HERV family. When examining specific HERV elements, it was also noted that the entire HERV element was not upregulated suggesting some level of specificity as to the sequence of the HERV element upregulated by SAHA.

Furthermore, when examining select HERV elements in a dose response curve it was noted that these elements were upregulated at lower clinically relevant dosages of SAHA (**Figure 5.5**). This increase appeared to be linear with respect to the logarithm of the dose of SAHA. The results from this dose response curve indicated that current concerns regarding HERV upregulation (i.e. potential re-insertion of upregulated HERV elements elsewhere in the genome or recombination of LTR elements with the HIV sequence) should be examined further. Integration of the HERV element next to proviral DNA (i.e. *gag, pol,* and *env* sequences) from other HERVs or even integrated HIV may create retrotransposons which can lead to

genome instability [320]. Additionally, it is conceivable that HERV LTR elements may potentially fuse with HIV transcripts generating new strains of this virus. However, it should be noted that neither of these possibilities have been reported for HERV elements from the LTR12 family, hence further study upon the interaction between these elements and HIV is merited.

5.5.2 Lessons from the study of HERVs with RNA-Seq

HERVs are particularly difficult to study as their classification schemes, naming conventions, and even membership in a particular group are subject to debate [305, 306, 307]. The benefits of using RNA-Seq to study HERVs is demonstrated throughout this work. HERV sequences are continually being discovered by new algorithms, such as RepeatMasker [307], and re-classified due to improvements in the classification scheme [305]. While one may refer to a particular HERV as a single entity, such as LTR12, it must also be recognized that LTR12, or small portions of it, are present in hundreds of locations in the human genome [306]. RNA-Seq allows one to re-examine data after additions or reclassification of HERVs in a database. Hurst et al. [302] indicated that SAHA treatment does not upregulate HERVs when only a select few sequences were examined. While the results presented here are in agreement with Hurst *et al.* for those three HERVs tested (**Figure 5.6**), more than 2000 other HERV elements were dysregulated with SAHA treatment. This indicates that RNA-Seq analysis is far more capable of determining overall HERV expression when compared to relying solely upon PCR analysis.

This data also demonstrated that several HERV elements had dysregulation in a portion of the sequence and not the full HERV element (**Figure 5.2**). This may be due to imperfect copies, genomic rearrangement, specificity of a drug to that region of

the HERV, or may indicate the database used to identify the HERV element should reclassify it as two or more separate entities. For example, RepeatMasker identified several sequences as separate entities while HERVd considered them one element (**Figure 5.2A**). Upregulation of only a portion of the element from HERVd may be explained by breaking it into more elements. Relying upon PCR alone to identify dysregulation of an entire HERV sequence may miss regions of that HERV which behave differently than the region examined and give an inaccurate assessment of total dysregulation. For example, if the HERV elements examined in this work had been measured by PCR in a location not showing upregulation, one would erroneously conclude that there was no dysregulation. Hence a more comprehensive approach, such as RNA-Seq, that can examine the coverage across the entire HERV element is required to accurately determine dysregulation. Afterwards, ddPCR may be used to confirm RNA-Seq results as was done in this study.

5.5.3 Implications of HERV dysregulation

The three HERV elements presented here and confirmed with ddPCR are all members of the HERV family, LTR12. This family has hundreds of members spread throughout the genome [305], with partial sequences found in all chromosomes when examining HERV elements from this family in HERVd [306]. LTR12 has been previously shown to be lowly expressed in various tissue types with the highest expression in tissue from the adrenal gland and testis [321]. These results are in agreement with this notion, as untreated samples showed low expression of these 3 HERV elements (**Figure 5.2**). Interestingly, it has been reported that LTR12 may drive proapototic p63 isoforms in the male germ line of both humans and apes [322]. The gene *TP63* encodes a member of the p53 family of transcription factors and

plays a role in the development and maintenance of stratified epithelial tissues [196], and in apoptosis [323], and also protects the female germ line against DNA damage [324]. LTR12 was integrated into the region upstream of *TP63* between 10 and 15 million years ago, during primate evolution and demonstrates strong promoter activity. It is suspected that activation of *TP63* by LTR12 in the testis provides protection against DNA damage [322]. Upregulation of this HERV has further been studied in the context of other death receptor genes such as *TNFRSF10B* (a.k.a. death receptor 5, DR5) [303]. A targeted approach demonstrated that HDACis such as SAHA can upregulate LTR12 associated with the promoters of these genes. This can further upregulate *TNFRSF10B* transcripts as LTR12 is found 757 bp upstream of the transcription start site of exon 1 [303]. Indeed, HDACi treatment for cancer eliminates the cancerous cells and may very well be brought about, in part, by upregulation of death receptor genes such as *TP63* and *TNFRSF10B*. The results from an untargeted genome-wide approach presented in this work are in agreement with the notion that the HDACi SAHA may upregulate LTR12, or portions of it, and may potentially lead to apoptosis in various cells.

Further concerns regarding upregulation of HERV elements is the potential for those elements to be activated creating retrotransposons which may be able to re-insert in other portions of the genome or insert directly into oncogenes increasing the risk of cancer [316]. Integration of the HERV element next to proviral DNA (i.e. *gag, pol,* and *env* sequences) from other HERVs or even integrated HIV may create retrotransposons which can lead to genome instability [299, 320]. A final concern is the potential for HERV elements to interact with integrated HIV. HIV has demonstrated a significant number of recombinations through the past century [317] and it is theoretically possible that HIV may recombine with HERVs to produce a new

strain of this virus. A method to screen for HERV and HIV recombination would be to examine fusion transcripts between HERV elements and HIV. While it is conceivable that HERVs may potentially recombine with HIV or expression of HERVs can lead to the creation of retrotransposons, these potential interactions between HIV and HERV elements from the LTR12 family have not been demonstrated or well studied. Further research into the interaction of upregulated HERVs presented in these results is warranted to rule out these possibilities as it was very clear that treatment with SAHA upregulated multiple elements at a high does and elements from the LTR12 family at clinically relevant doses of SAHA.

A final observation regarding HERV element dysregulation is the noticeable downregulation of various elements from HERV families, such as ERVL (**Figure 5.1**). These findings were not pursued in this work, but merit mention as this form of dysregulation is not consistent with the theoretical method by which SAHA can interact with HERVs (upregulation of the HERV through chromatin relaxation, **Chapter 1, Section 1.4**). It is entirely possible that HERV elements are integrated into genes downregulated by SAHA treatment, thereby explaining these results. However, it is also possible that SAHA could interact directly with HERVs. Indeed, various drugs therapies such as interferon beta (IFN-β) therapy are known to mediate the downregulation of HERV-H and HERV-W [325]. SAHA dysregulates thousands of genes (**Chapter 4**), therefore is it conceivable that SAHA can alter gene expression in a way that reduces HERV expression either through direct interaction with the HERV or interaction with genes which can regulate HERV expression.

5.5.4 Conclusions of this work

A wealth of information regarding HERV dysregulation upon SAHA treatment was provided by this analysis, much of which will generate future projects and discoveries. LTR12 in known to be integrated into hundreds of locations across the human genome. Therefore, it would be advantageous to determine if HERVs from this family are integrated next to any pro-viral sequences, or elements which are known for being active retrotransposons. LTR12 could act as a promoter for these elements potentially creating genome instability. Additionally, it is not currently known if sequences from elements of the HERV family LTR12 are capable of recombining with HIV to potentially generate or unleash new strains of this virus. This work did not assess this possibility, but merits mention for future research into the effect of HDACis, such as SAHA, upon HERVs in HIV infected individuals.

Other HDACis (Romidepsin, Panobinostat, etc., **Chapter 1, Section 1.4**) are capable of upregulating HERV elements from the LTR12 family [303], but it is not known exactly which elements from LTR12 or other HERV families are targeted for dysregulation by various HDACis. This work demonstrated that a number of elements from this family are upregulated, but there are some from this same family which are also significantly downregulated upon treatment with SAHA (**Figure 5.1**). This may be due to sequence variation between HERV elements from the same family, or to differences in the location of integration for HERV elements. Further study which examines the sequence similarity and integration sites of these elements is required.

This research may further lead to a potential method to organize HERV elements into specific families based upon expression as opposed to algorithms searching for sequence similarity. It is clear from this work that SAHA upregulates

HERVs from the family LTR12, but this upregulation is not consistent across the entire element (**Figure 5.2**). An argument could be made that these elements should be separated based upon their expression after treatment or that expression should be used to assist in their categorization, as this expression differs vastly across the whole HERV element sequence.

Finally, this research provides a new analysis method by which one can examine dysregulation for a large number of HERV sequences in a small number of samples, and then better target certain HERV elements of interest in additional samples without having to subject them to potentially expensive RNA-Seq methods. The similarity between Log_2FCs measured by RNA-Seq and ddPCR demonstrates the benefit of utilizing RNA-Seq to choose primer and probe locations for HERV sequences, and may ultimately streamline the analysis of HERVs and focus research upon HERV families more relevant to the treatment utilized.

5.6 Figures

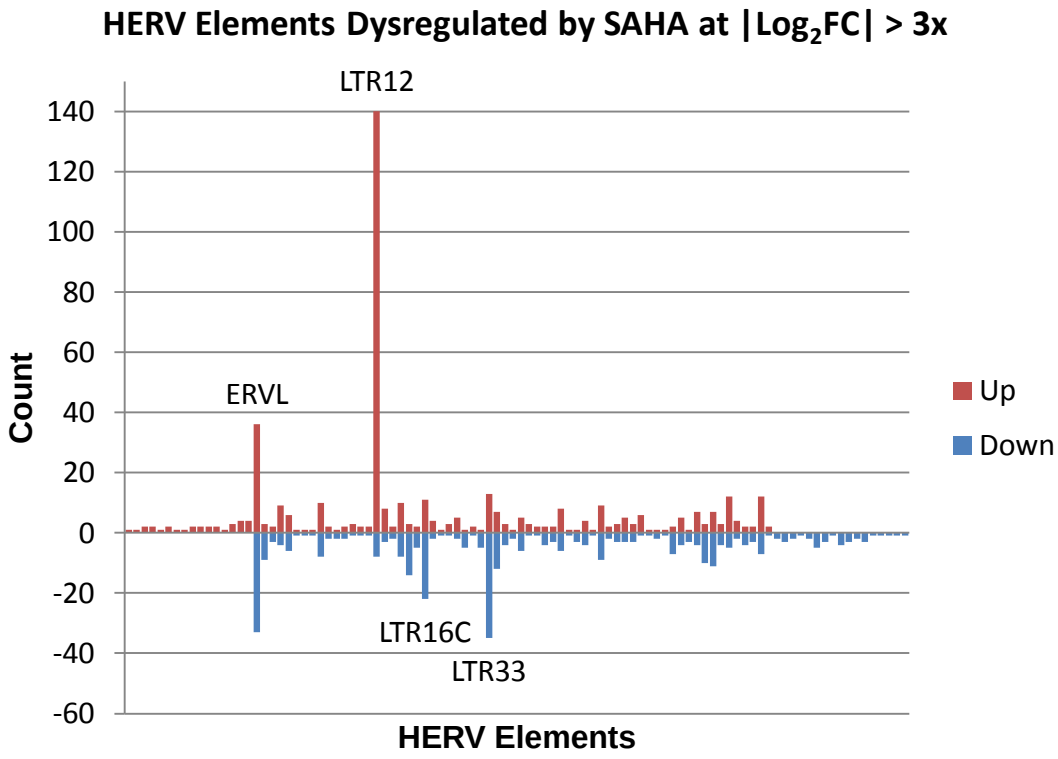

Figure 5.1: Overall dysregulation of all HERV elements from HERV families by SAHA treatment. Counts from upregulated elements were combined into HERV families (81 total) and were plotted above, while downregulated elements merged into HERV families (82 total) were plotted below. Each element used to generate the total count of elements present in the HERV family had an absolute $Log_2FC > 3$. HERV families which had the highest number of dysregulated elements are labeled, the largest being the HERV family LTR12.

A

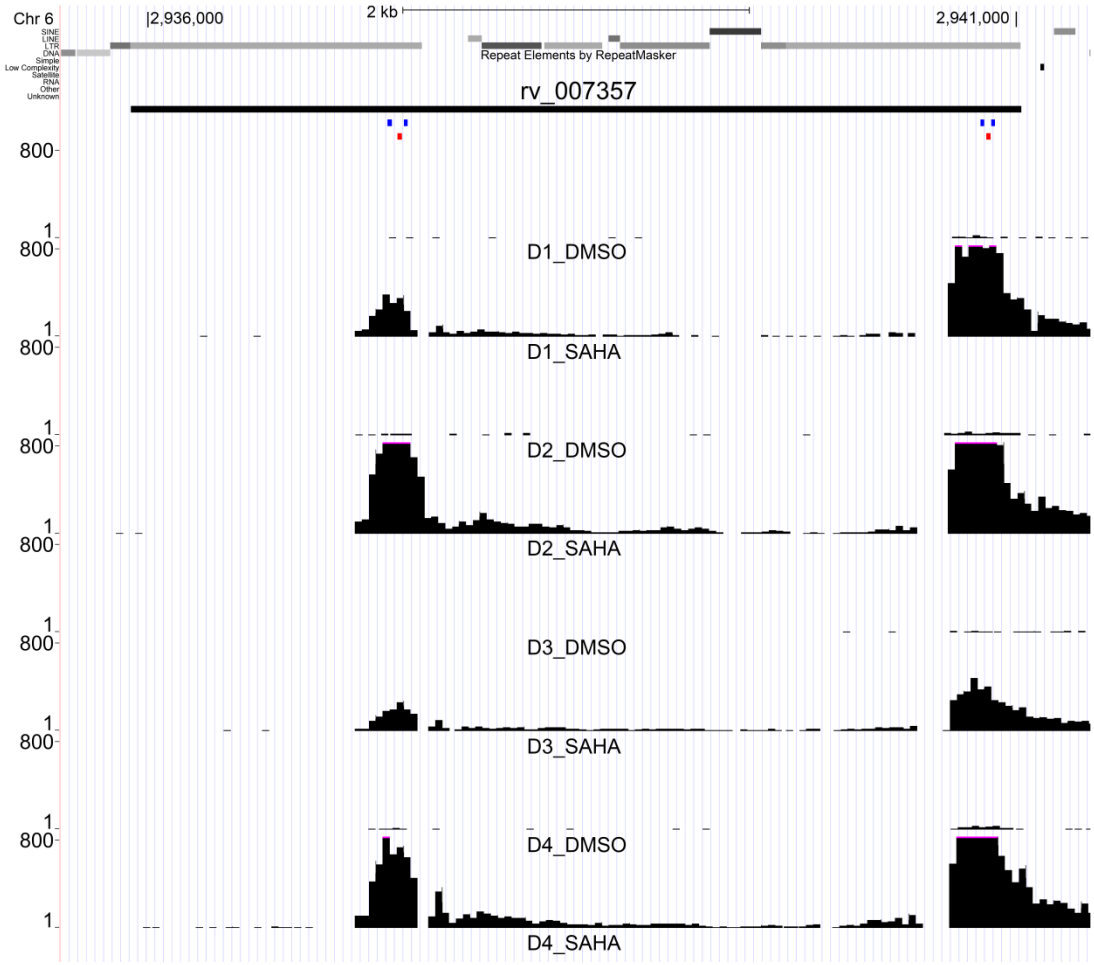

Figure 5.2: Expression coverage map of HERV elements upregulated by SAHA across all samples. The HERV elements rv_007357 (A), rv_007420 (B), and rv_010177 (C) were measured by converting mapped read files to wiggle files and then visualized with the UCSC genome browser. The bar at the top of the image represents the scale bar while the numbers on the top left and right represent the hg19 positional information. The top track in all three images is the RepeatMasker track, and shows the region for which the elements are present in the HERVd database. RepeatMasker considers this region multiple HERV elements (A and B) or agrees with HERVd in that it is a single element (C). Below this track is the full HERV element from HERVd. Shown in blue and red bars are the primer and probe locations respectively, which were utilized in ddPCR confirmation. For the two longer HERV elements (A and B), primers and probes were developed for two regions, the left most set labeled as set 1 and the rightmost set labeled as set 2, while for the shorter HERV element (C), a single primer and probe sequence was utilized. The donor tracks created from all mapped read data is presented with the pink line in the donor tracks (A and B) indicating the coverage peak was higher than the chosen range. Numbering on the left represents the maximum range of coverage.

B

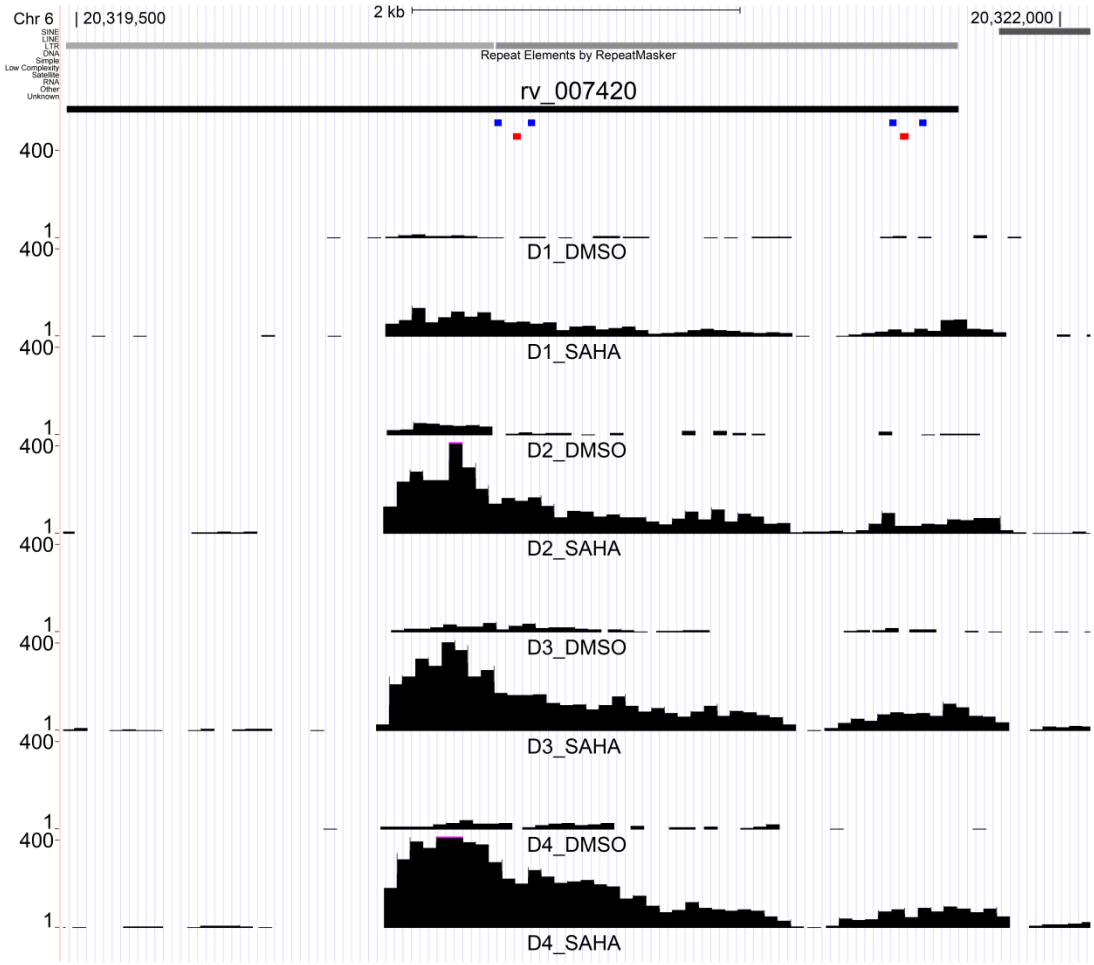

Figure 5.2: Expression coverage map of HERV elements upregulated by SAHA across all samples, continued.

C

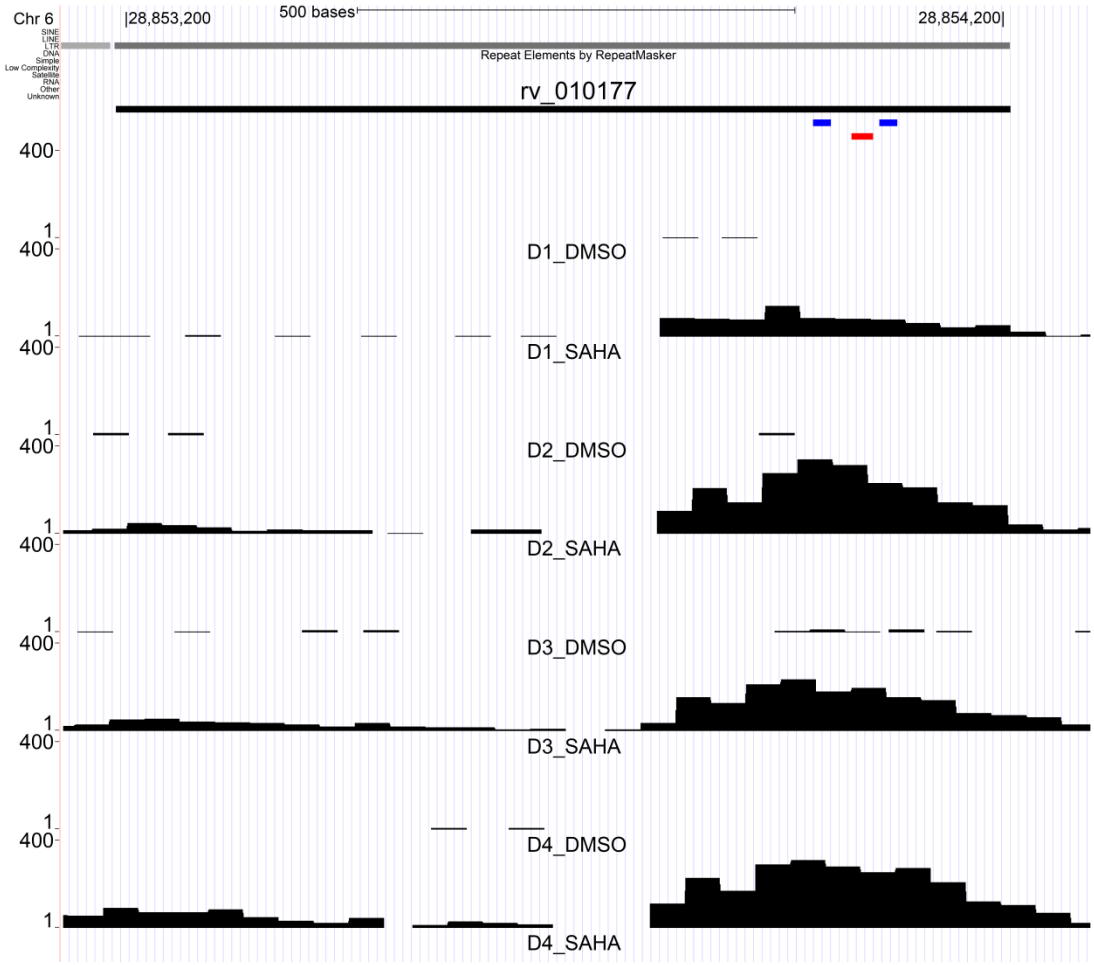

Figure 5.2: Expression coverage map of HERV elements upregulated by SAHA across all samples, continued.

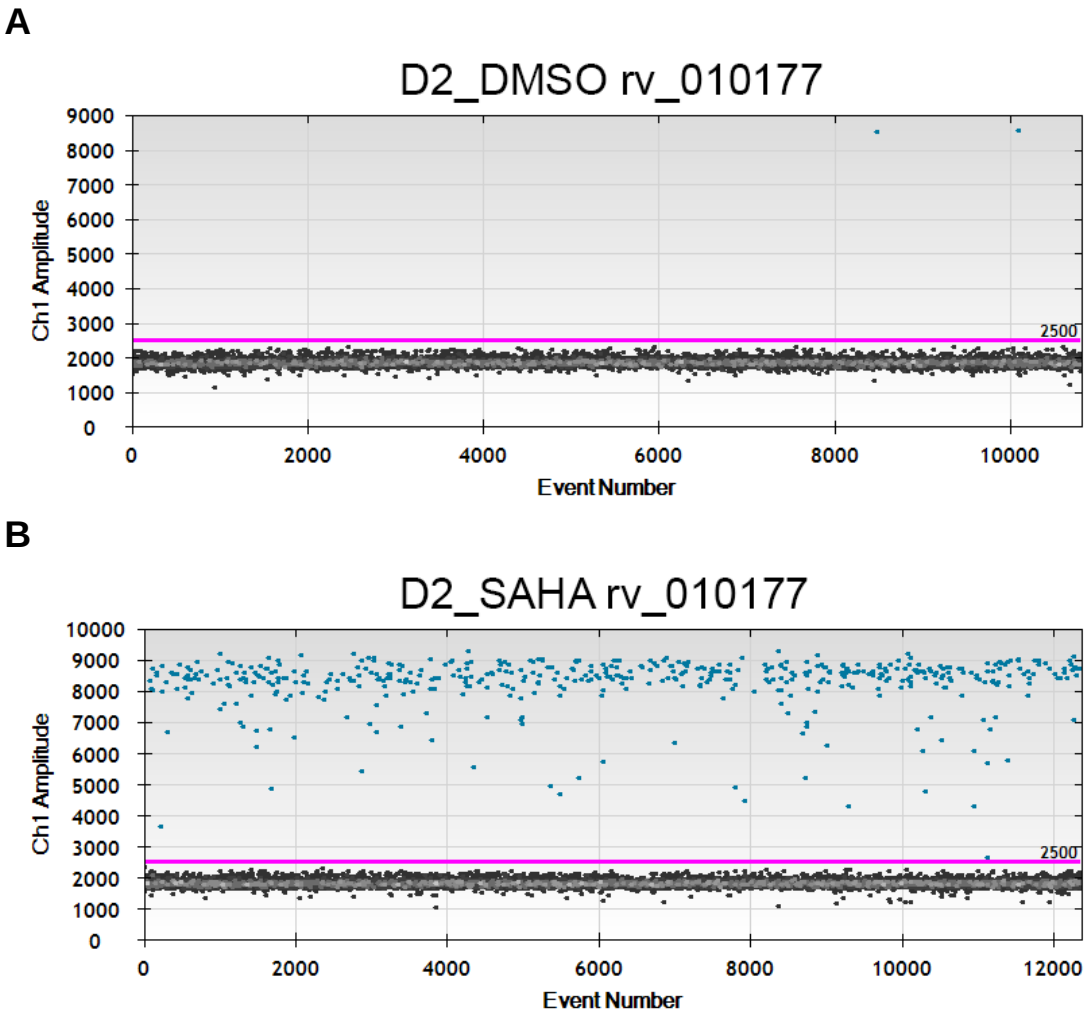

Figure 5.3: Representative image of raw amplitude data generated with ddPCR on the QX100 ddPCR system. Channel amplitude (y-axis) and event number (x-axis) are utilized to determine if an emulsion PCR droplet containing the desired RNA sequence amplified the primer and probe set for the DMSO (A), and SAHA (B) conditions. The threshold is set above the background fluorescence (black dots) level for each sample. Poisson statistics are then utilized to convert these count events (blue dots) to counts per µL for use in comparison between DMSO and SAHA treated conditions.

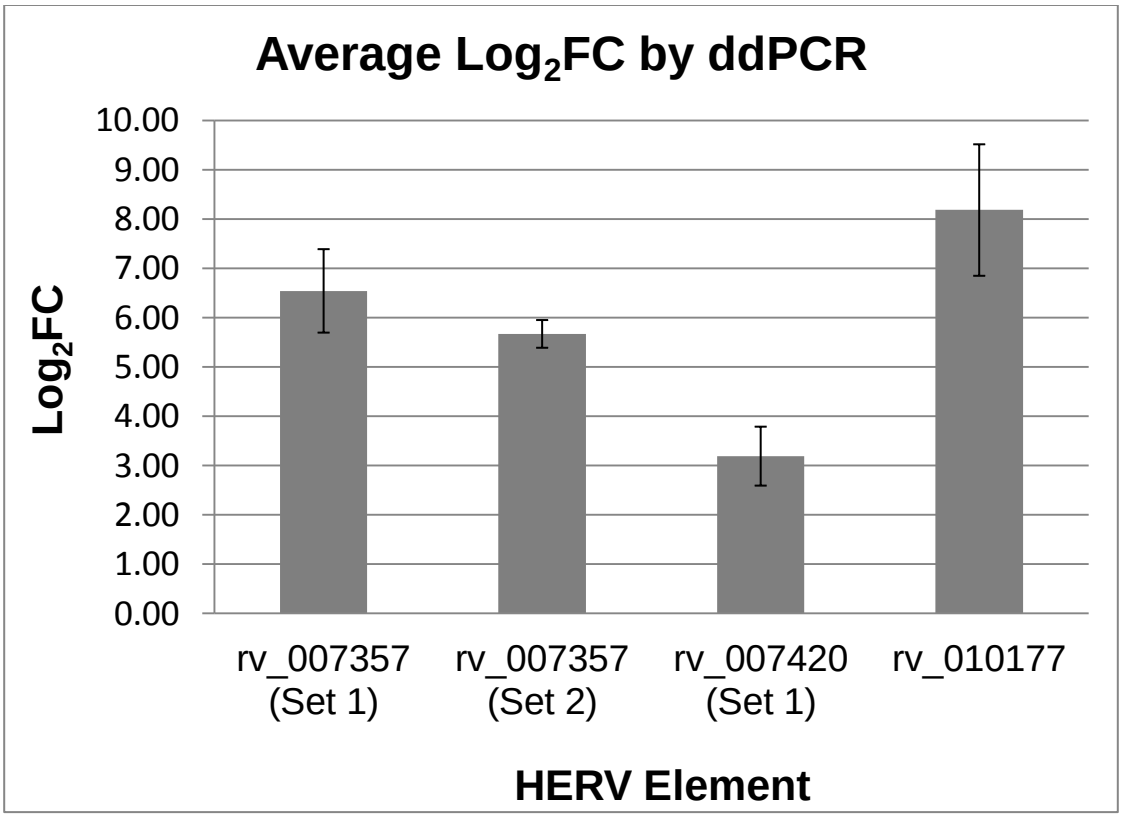

Figure 5.4: Confirmation of upregulation by ddPCR for selected LTR12 HERV elements. Log_2 fold changes between SAHA and DMSO treated conditions were averaged across all 4 donors and the results of each HERV element are presented. Error bars represent standard deviations across donors. The terms "Set 1" and "Set 2" indicate the primer and probe sequence utilized when two locations where chosen.

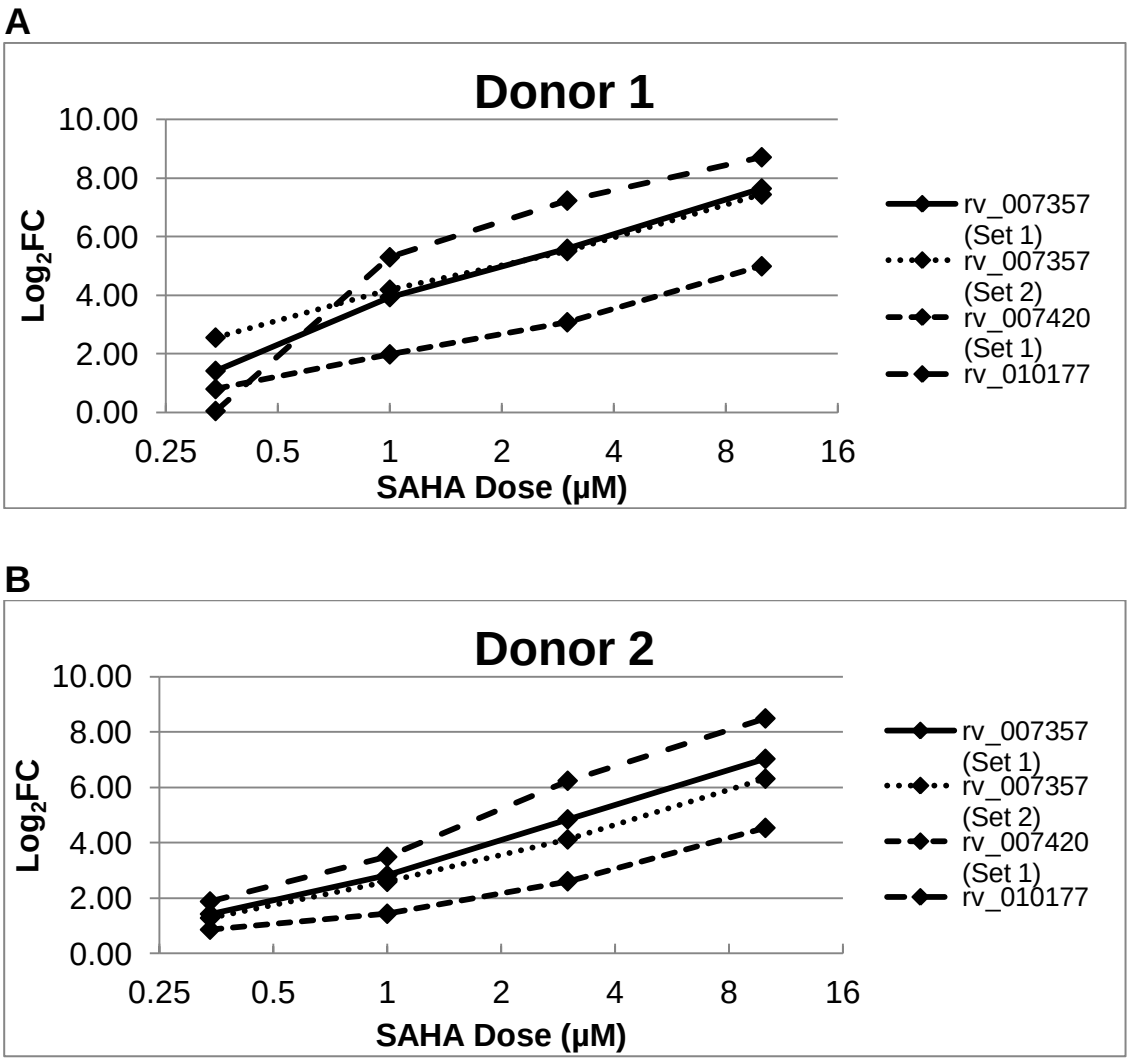

Figure 5.5: Dose response curve for select HERV elements. Log_2FC results for (A) donor 1 and (B) donor 2 for all four HERV primer and probe sets. The y-axis represents Log_2FCs that were generated by dividing the normalized HERV element copies at each dose by the number of copies at the 0 µM dose. The x-axis is the SAHA dosage plotted on a logarithmic scale. The terms "Set 1" and "Set 2" indicate the different primer and probes used to confirm a single HERV element.

A

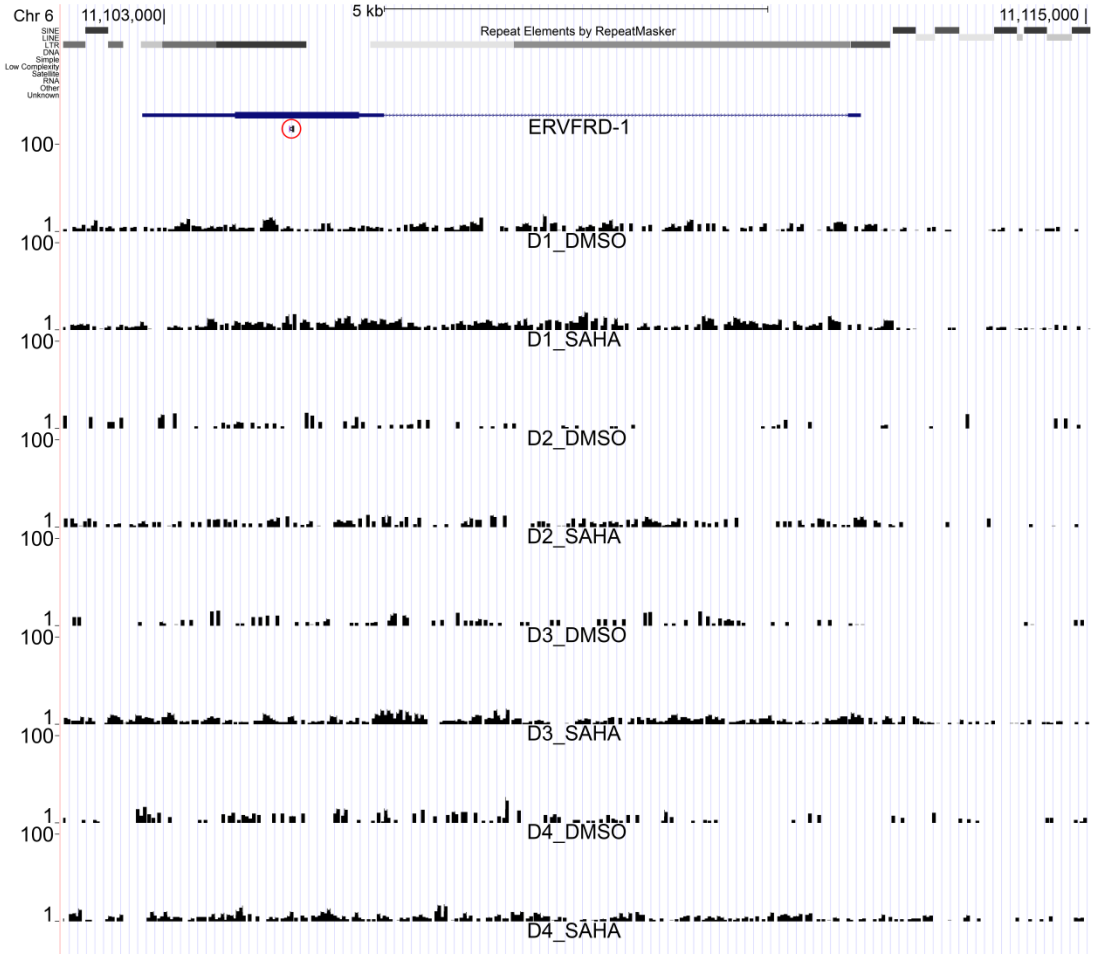

Figure 5.6: Expression coverage map of non-dysregulated HERVs. The HERVs (A) HERV-FRD, (B) HERV-W, (C) HK2 *env*, and (D) HK2 *pol* were examined by generating a coverage map with mapped read files. HERV locations were determined from the primer and probe sequences of Hurst *et al*. [302] All genomic locations are given in hg19 coordinates with a scale bar on top. Numbering on the left indicates the maximum scaling of each coverage track. The top track is the RepeatMasker track and the large black bar represents the HERV noted, except in the case of HERV-FRD (A). For this HERV, the RepeatMasker track divides this region into several segments and, because the RefSeq track contains this HERV, it is presented below for only (A). The red circle in each graph indicates the location of the primer and probe sequences from Hurst *et al* 2016.

B

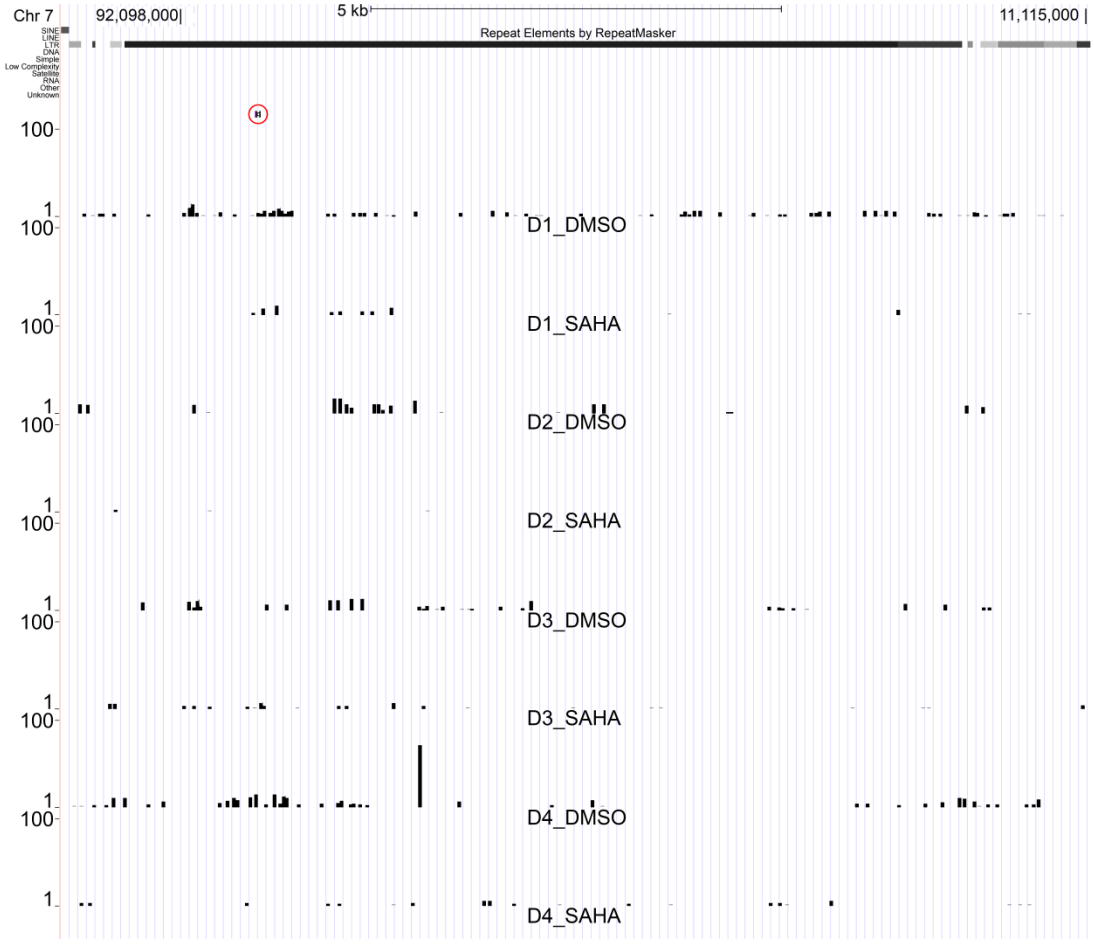

Figure 5.6: Expression coverage map of non-dysregulated HERVs, continued.

C

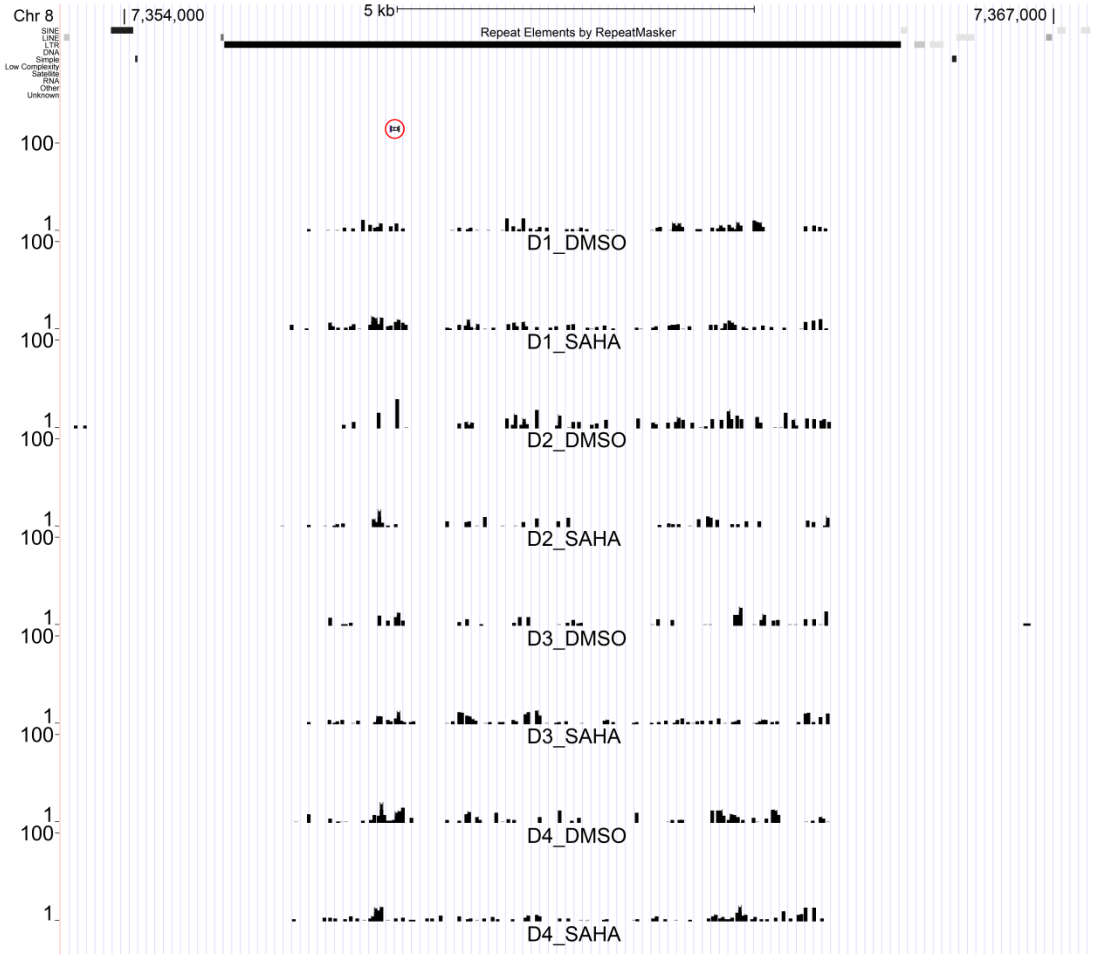

Figure 5.6: Expression coverage map of non-dysregulated HERVs, continued.

D

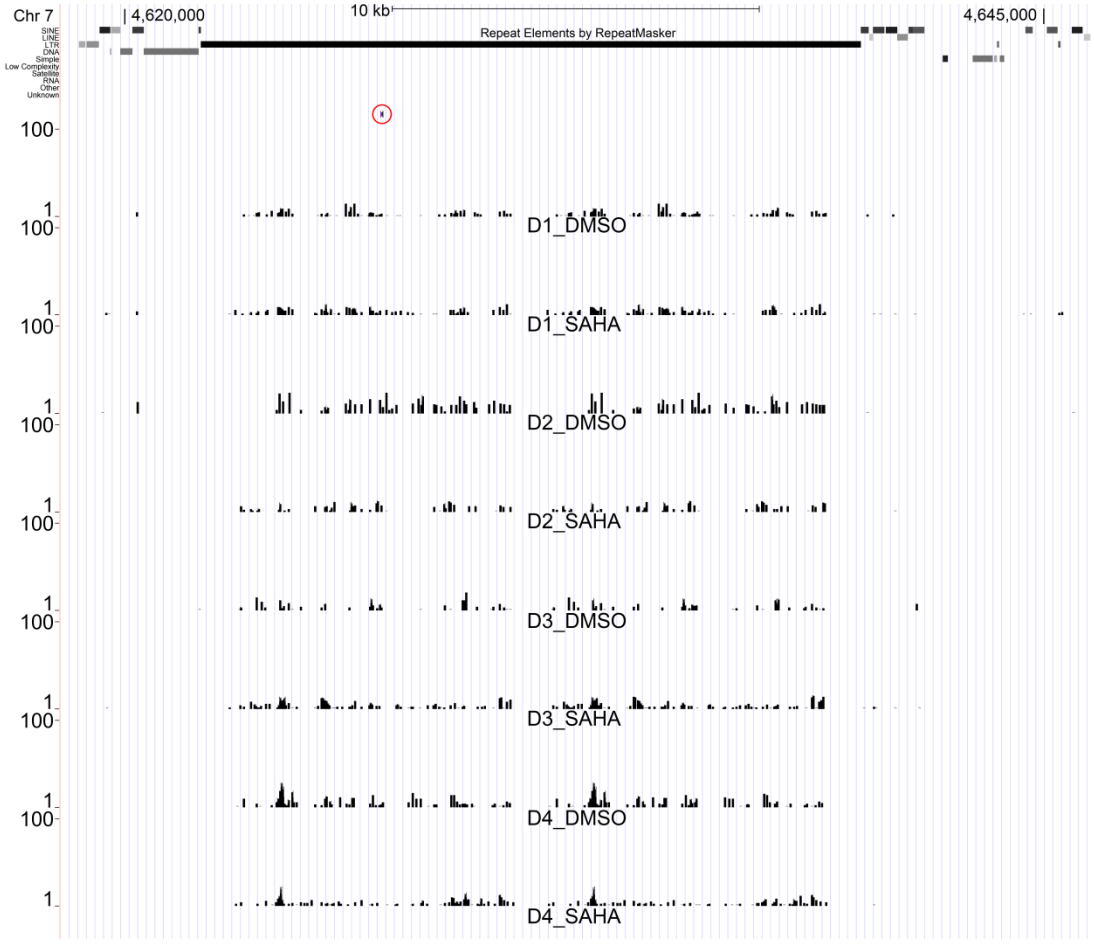

Figure 5.6: Expression coverage map of non-dysregulated HERVs, continued.

5.7 Tables

Table 5.1: Primers designed to confirm HERV upregulation with ddPCR. Five primer and probe pairs were created for confirmation of HERV elements. Primer pair #4 (rv_007420 Set 2) was ultimately discarded due to amplification in the negative control. The Amplicon length is the total length of the primers and probes including nucleotides in between. Length represents the length of the sequence, and GC% is the percentage of G and C nucleotides in the primer or probe sequence.

rv_007357 Set 1	Sequence (Total Amplicon Length: 114bp)	Length	GC%
Forward	GAGCGTATGGCGTTATGTAGTT(Sense)	22	45.5
Probe	TTGAGCCGATGAGATCGCTAAGCC(Sense)	24	54.0
Reverse	AGCGGTATGTCCTCCCTTTA(AntiSense)	20	50.0
rv_007357 Set 2	Sequence (Total Amplicon Length: 102bp)	Size	GC%
Forward	GGAGGAACGAAACACTCATCT(Sense)	21	47.6
Probe	TGCAACTTTCACAGAGTCGTCTCACC(AntiSense)	26	50.0
Reverse	CGTCTCACCCACTTCAGAAA(AntiSense)	20	50.0
rv_007420 Set 1	Sequence (Total Amplicon Length: 124bp)	Size	GC%
Forward	GGTAGTGAGAGAGAACGGTATG(Sense)	22	50.0
Probe	TCCTCTGCTCATTCTGGTTGTGCT(Sense)	24	50.0
Reverse	CTAAAGAGCTCCCACGGTATAG(AntiSense)	22	50.0
rv_007420 Set 2	Sequence (Total Amplicon Length: 114bp)	Size	GC%
Forward	GCTTGACTCTTGAGCCAGTAA(Sense)	21	47.6
Probe	ACCAGAAGGAGGAAACTCCGAACAC(Sense)	25	52.0
Reverse	CCGGAGTTTGTTCCTTCTGATA(AntiSense)	22	45.5
rv_010177	Sequence (Total Amplicon Length: 96bp)	Size	GC%
Forward	ACTCCAGACACACCGTCTTA(Sense)	20	50.0
Probe	ATTGGTAGCTTTCCCGAGTCAGCG(Sense)	24	54.0
Reverse	TCATTCCATTCAGGTGGGTTC(AntiSense)	21	47.6

Table 5.2: Dysregulated HERV elements after SAHA treatment. Dysregulated HERV elements were binned based upon total fold change. In parenthesis are the number of unique HERV families for these elements as determined by HERVd annotation. Log_2 fold changes were obtained by calculating the fold change as SAHA treated over DMSO treated and then taking the log base 2.

Dysregulation	Total	$Log_2FC > 1$	$Log_2FC > 2$	$Log_2FC > 3$
Upregulated	1095 (115)	1011	691	451
Downregulated	1007 (102)	938	642	363

Table 5.3: HERV elements chosen for confirmation. The names of the HERV elements are based upon the HERVd annotation, while the family to which they belong is also presented. Positional information is based upon the hg19 human reference. Log_2FC and FDR corrected p-values (q-values) are presented based upon differential expression analysis performed with EdgeR.

HERV Element	rv_007357	rv_007420	rv_010177
Family	LTR12	LTR12	LTR12
Strand	+	+	+
Chr	6	6	6
Start (bp)	2,935,901	20,318,970	28,853,189
End (bp)	2,941,031	20,321,689	28,854,209
Length (bp)	5130	2719	1020
Differential Expression			
Log_2FC	5.590	3.112	7.162
q-value	1.76E-63	2.36E-22	1.35E-18

5.8 Acknowledgements

The dissertation author and co-authors of this work would like to thank the CARE Collaboratory (AI096113) for supporting this work. The authors also acknowledge the use of the computer facilities and associated support services at the University of California, San Diego.

The views expressed in this chapter are those of the dissertation author and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government. The sponsors of this research were not involved in the study design, collection or interpretation of the data, manuscript preparation, or the decision to submit the article for publication.

This chapter is adapted from **White, C.H.**, Beliakova-Bethell, N., Lada, S., Richman, D.D., Woelk, C.H., *Dysregulation of human endogenous retroviruses in primary CD4 T-Cells upon treatment with SAHA*, in preparation for submission. The dissertation author was the primary author on this paper and responsible for designing the study, development of the next generation sequencing pipeline to study HERVs, analysis of the data, interpreting the results, and writing the manuscript.

Chapter 6

Overall Summary and Conclusions

6.1 Summary of Results

The overreaching theme of the research in this dissertation was the application of omics technologies (e.g. proteomics, microarray, and RNA-Seq) to facilitate a better understanding of HIV latency and strategies to cure this persistent viral infection. First, a pipeline for the analysis of RNA-Seq data was developed and utilized to study the Bosque and Planelles primary T_{CM} cell model of HIV latency [37] (**Chapter 2**). Reads mapping to the deleted region in the defective HIV virus (DHIV) used in this model suggested that this construct had recombined with the pLET-LAI plasmid, thereby reconstituting a wild type virus capable of active replication. This work established that closer examination of models utilizing defective vectors was merited and that care should be taken when interpreting results from similar models of HIV latency. Ultimately, this research led to a revision of the Bosque and Planelles model of HIV latency, the Martins *et al*. model [128], whereby replication competent HIV-1_{NL4-3} was utilized along with ART to better represent *in vivo* latency. The RNA-Seq pipeline was then used to characterize the transcriptional profile of both host and virus for this revised new model [128] (**Chapter 3**). In this chapter, potential DEGs for

latency were identified, many of which were related to the p53 signaling pathway. The host transcriptomic profile was then compared to that of three other models of HIV latency and established that minimal overlap was present. Finally, the relationship between p53 signaling and the establishment of latency was examined by the inhibition of the transcriptional activity of p53 with pifithrin-α. This inhibition resulted in a statistically significant reduction in the establishment of latency that was correlated to the level of HIV integration, as measured by Alu-PCR, suggesting that this inhibition may interfere with viral integration. These results demonstrated the importance of RNA-Seq as a technology to study models of HIV latency and shed light upon the potential mechanisms of latency establishment.

Latency models such, as those described previously (**Chapters 2 and 3**), are often used to screen LRAs for "Shock and Kill" strategies (**Chapter 1, Section 1.4**). One such agent is the HDACi, SAHA, which was studied using both transcriptomic and proteomic methods (**Chapters 4 and 5**). It was found that SAHA modulated several thousand genes and hundreds of proteins (**Chapter 4**), many of which have a deleterious effect upon HIV activation from latency (e.g. upregulation of genes which compete with *Tat* for TAR binding, upregulation of genes which compete with p-TEFb binding, and downregulation of genes required for HIV activation). This may explain, at least in part, why such treatments have had little success in "Shock and Kill" strategies to eliminate the latent HIV viral reservoir [58, 90]. It was further determined that SAHA also dysregulates HERV elements (**Chapter 5**). In particular, SAHA was shown to upregulate HERV elements from numerous families, with a specificity towards those from LTR12. This may lead to deleterious effects during treatment as the dysregulation of HERV elements has previously been associated with various diseases including cancer, shown to upregulate genes when integrated into the

promoter region of the gene [303], and to increase genome instability through recombination [299]. However, it should be noted that these deleterious effects of HERV upregulation were not demonstrated in this work, nor is it currently known if upregulation of elements from LTR12 can lead to genome instability or recombination with HIV. Overall, the results from omics studies using SAHA suggested that other HDACis may be better suited for HIV reactivation, or that SAHA should be paired with another LRA with a different mechanism of action to accomplish the "Shock" portion of this strategy without the unwanted side effects of host gene and HERV dysregulation.

What follows is a brief discussion of the studies performed in support of this dissertation, their limitations, and how omics technologies have been used to characterize models of HIV latency and identify potential behaviors of LRAs in "Shock and Kill" strategies for a cure for HIV.

6.2 A Reflection on Models of HIV Latency

The study of HIV latency is hindered by the paucity of latently infected cells in an HIV infected individual on ART [19], the lack of phenotypic markers to identify latently infected cells [34], and the genetic heterogeneity of both host and virus [33]. For these reasons, models of HIV latency have been developed that have provided a glimpse into the mechanisms by which HIV hides from the immune system during latency, potential biomarkers of latency, and the pathways through which various drug compounds may activate HIV while in a latent state to facilitate a cure for HIV through the "Shock and Kill" strategy [34, 35, 95, 128, 129, 161]. Presented in this dissertation are two models of HIV latency, the first is the Bosque and Planelles model [129], and the second is the revised Martins *et al*. model [128]. These models

were analyzed through the use of RNA-Seq and identified (1) an issue with the first, Bosque and Planelles model, in that the replication deficient HIV virus (DHIV) was able to recombine with the plasmid thereby reconstituting a fully replication competent virus, (2) potential biomarkers of HIV latency, and (3) a role for the p53 signaling pathway in the establishment of latency within the 2nd model, Martins *et al*.

Interestingly, the differentially expressed genes identified in the Bosque and Planelles model of latency (**Chapter 2**) were not identified as potential biomarkers in the revised version of this model, Martins *et al*. (**Chapter 3**). However, this is likely due to active infection present in the first model. The Bosque and Planelles model may be a more accurate representation of acute HIV infection, and potentially the beginnings of the establishment of latency, hence the biomarkers found may be those of acute infection instead of latent infection. The Martins *et al*. model may be a better representation of *in vivo* latency for an HIV infected individual undergoing treatment with ART.

When examining multiple models of HIV latency, no overall overlap in upregulated or downregulated genes was found. Only 5.4% of the total upregulated genes and 5.7% of the total downregulated genes were shared between at least two models (**Chapter 3, Section 3.4.5**). This raises a question as to whether the markers identified are indeed actual biomarkers of HIV latency, or if they are specific to the model studied. Even in models which utilized primary CD4 T-cells, the methodology for model generation, the cell cycle status upon infection, and the viral vector utilized may have an effect upon the biomarkers identified. Therefore, a large number of markers may be specific towards only one model. Additionally, the technologies used to examine these models (proteomics methods, microarrays, RNA-Seq, etc.) may differ in their ability to measure transcripts and proteins, which can result in less

overlap from differentially expressed genes (DEGs) and proteins (DEPs). Despite these obstacles, some identified DEGs and DEPs, as well as pathways, may be shared across models (**Chapter 3, Section 3.4.5**). For example, differential gene expression relating to p53 signally identified in the Martins *et al*. model (**Chapter 3**) is also found in the model utilized by Iglesias-Ussel and colleagues [161]. This would suggest that a pathway approach, in addition to the identification of individual markers, would be highly beneficial when comparing these models. Overlaying the HIV latency signature from several models upon a protein interaction network (PIN) would allow one to identify sections and potential pathways shared and demonstrate overlap at the sub-network level. Sections of the PIN which are not shared may be representative of model specific effects that may have captured one aspect of the multiple variables present during HIV latency, such as the maturation stages of infected cells, the cell cycle status (resting or dividing), and the changing conditions for the infected individual undergoing treatment with ART.

In summary these results from RNA-Seq studies of the models of HIV latency studied demonstrated the importance of confirming that HIV infected cells produce the expected host and viral transcripts, that caution should be taken when interpreting results from models that use deficient HIV constructs, and the importance of RNA-Seq analysis in the study of HIV latency. It is greatly recommended that in the future, researchers utilize RNA-Seq to examine both the host and virus transcriptome in more highly powered studies (> 3 samples). It is further suggested that researchers examining comparisons between models of HIV latency utilize pathway information and interaction networks to identify overlapping themes which may be missed by looking at only overlapping genes or proteins. This methodology will allow for a better

analysis of latency models, thereby leading to a more complete understanding of HIV latency.

6.3 The Utility of SAHA in a "Shock and Kill" Strategy for a Cure for HIV

An important observation from this work is the effect of SAHA treatment on primary CD4 T-cells (**Chapters 4 and 5**). SAHA treatment, while initially utilized to activate HIV from latency, has been shown to also modulate a number of genes which may have a counter effect upon HIV reactivation by either downregulating genes which assist in HIV activation (e.g. *LEF1* and *IKBKB*), or upregulating repressors of HIV transcription (e.g. *HMGA1*, *BRD2*, and *HEXIM1*) (**Chapter 4**). This study was initially performed without HIV infection to avoid confounding the results with genes which activate in response to viral replication. However, now that genes which have a counter effect upon HIV reactivation have been identified, the next stage is to examine the behavior of these genes within models of HIV latency. Future studies would also include testing if counteracting the potentially inhibitory effects of SAHA may result in more robust HIV reactivation. If a gene required to successfully shock HIV out of latency is downregulated by SAHA, a second compound could be added to counteract the negative effect of the HDACi upon HIV reactivation. If a gene which represses HIV activation is upregulated by SAHA, then knock down of such a gene through small interfering RNAs (siRNAs) or inhibition of the gene product may enhance HIV activation. For example, it is hypothesized that utilizing siRNAs against *HMGA1*, along with SAHA, may result in a higher level of HIV reactivation than SAHA treatment alone. Additionally, other HDACis such as Romidepsin [80]

and Panobinostat [267] have improved HIV activation. This may be due to more limited off-target effects on genes which impede HIV activation when compared to SAHA, and warrants study utilizing transcriptomic and proteomic methods in a similar manner as presented in this work for SAHA.

In addition to the effect on genes, which may lead to inhibition of HIV activation, SAHA appeared to potentially dysregulate multiple HERV elements across a large number of HERV families (**Chapter 5**). It was previously believed that HERV dysregulation upon SAHA treatment was not a concern due to the lack of upregulation of the 3 HERVs: HK2, HERV-W, and HERV-FRD [302]. However, while the results from the present studies confirm the findings of those HERVs (**Chapter 5, Section 5.4.3**), it was found that multiple elements from 100+ families are upregulated, and that elements from the HERV family LTR12 were upregulated far more frequently compared to other HERV families. This suggests that SAHA my preferentially target this HERV family for upregulation. Additionally, it is conceivable that HERV elements derived from LTR12 may act as a promoter for full-length proviral DNA integrated after LTR12 elements leading to potential reverse transcription and increased genome instability [299]. While this possibility hasn't been tested in this work, the results presented demonstrate that further analysis of the potential side effects upon HERV dysregulation by treatment with SAHA or other HDACis is warranted.

It has been shown previously that LTR12 is integrated upstream, and acts as a promoter for the genes *TNFSRF10B* and *TP63* [303]. Furthermore, it was also shown that activation of LTR12 with SAHA (18 hour treatment) can lead to upregulation of both genes in cell lines [303, 322]. However, no differential expression for either gene or their protein products was found in primary CD4 T-cells

treated for 24 hours with a 1μM dose of SAHA (**Chapter 4**) despite finding upregulation of multiple elements from the LTR12 family in another study (**Chapter 5**). These differing results may be due to several possibilities. The effect of SAHA on gene expression may be transient. For proteomic analysis the lack of dysregulation may be due the inability of iTRAQ to detect those proteins in an untargeted approach. While iTRAQ was able to identify proteins from over 6,000 unique genes, this leaves around 14,000 genes whose proteins were undetected [166]. A more targeted approach to examine protein products of these two genes is required to assess their differential expression at the protein level upon treatment with SAHA. A third possibility is that the cell type has an effect upon the activation of these genes by LTR12 upregulation. Upregulation of *TNFSRF10B* and *TP63* was noted in cell lines, particularly those derived from testicular cancer [303, 322]. However, the work presented in this dissertation examined primary CD4 T-cells. Lack of dysregulation of both of these genes in this work may suggest a cell type specific effect. Additional study of the integration sites of these HERV elements from LTR12 and a targeted analysis of the proteins derived from the genes *TNFSRF10B* and *TP63*, and other genes for which LTR12 elements may act as promoters, is required to determine the effect of SAHA treatment upon the interaction between HERV elements and the transcriptome and proteome of CD4 T-cells.

6.4 The Importance of Transcriptomic and Proteomic Methods to the Study of HIV Latency

The importance of RNA-Seq to the study of HIV latency is demonstrated throughout chapters 1, 2, 3, and 5. This technology was key to the identification of

the recombination event that produced the replication competent virus in the Bosque and Planelles model of HIV latency (**Chapter 2**), which was previously thought not to be possible in this model. This technology also allowed an analysis of HIV transcripts (**Chapters 2 and 3**) which may have remained hidden in other targeted transcriptomic methods such as microarrays or PCR. RNA-Seq further facilitated the use of ERCC spike in controls (**Chapter 3**), which significantly reduced the unwanted variation from nuisance sources and allowed for increased power with a small sample size. Finally, RNA-Seq provided a far more comprehensive and detailed picture of HERV dysregulation by examining all HERV elements in the human genome, and identified dysregulation across large portions of the HERV sequence, both of which were lacking in a previous analysis [302].

Another observation is the importance of proteomic research in the study of HIV latency. Without the analysis techniques developed in this dissertation to study protein results generated from the iTRAQ LC-MS method, several proteins would have remained unrealized as affecting the activation of HIV during SAHA treatment (**Chapter 4, Section 4.5.1**). Nearly half of the proteins identified as differentially expressed did not have a corresponding RNA transcript identified as such. Furthermore, a post transcriptionally modified (e.g. acetylated) form of the protein product of *HMGA1*, was identified as significantly differentially expressed, a finding that would have remained hidden without these proteomics methods.

Due to these findings, it is highly encouraged that in the future, researchers utilize an approach of combined analysis of data generated by transcriptomic and proteomic methods. The cost to incorporate both methods will be offset by the benefits and additional insights generated. Total RNA-Seq is highly suggested as the transcriptomic method of choice because it provides access to information such as

viral transcription, splicing patterns of transcripts from both host and virus, potential dysregulation of elements such as HERVs which are not normally present on a microarray or are too cumbersome to analyze comprehensively with PCR based methods. Furthermore, proteomic screens can identify dysregulated elements missed at the transcript level and provide a more comprehensive picture of the biological systems studied. Integration of data obtained by both of these technologies can be costly and challenging, but the results presented here demonstrate that this will be important for the identification of strategies for the elimination of the latent HIV viral reservoir.

6.5 Concluding Statement

This dissertation has focused upon using omics technologies to describe model systems of HIV latency and to characterize the effects of compounds used to facilitate a cure for HIV. Although numerous challenges stand in the way of this cure (multiple cell reservoirs, clear definitions of latency, difficulties in the identification and targeting of latently infected cells for termination, and difficult to access regions of the body for drug therapy), the power of these approaches is expected to accelerate the search for a cure. It will be of great interest to revisit the HIV cure debate in the next decade to determine how far we have come in the eradication of HIV from infected individuals.

Bibliography

1. Sharp, P.M. and B.H. Hahn, *Origins of HIV and the AIDS pandemic.* Cold Spring Harb Perspect Med, 2011. **1**(1): p. a006841.

2. Faria, N.R., A. Rambaut, M.A. Suchard, G. Baele, T. Bedford, M.J. Ward, A.J. Tatem, J.D. Sousa, N. Arinaminpathy, J. Pepin, D. Posada, M. Peeters, O.G. Pybus, and P. Lemey, *HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations.* Science, 2014. **346**(6205): p. 56-61.

3. Hahn, B.H., G.M. Shaw, K.M. De Cock, and P.M. Sharp, *AIDS as a zoonosis: scientific and public health implications.* Science, 2000. **287**(5453): p. 607-14.

4. UNAIDS, *UNAIDS: How AIDS Changed Everything*. 2015.

5. Aliouat-Denis, C.M., M. Chabe, C. Demanche, M. Aliouat el, E. Viscogliosi, J. Guillot, L. Delhaes, and E. Dei-Cas, *Pneumocystis species, co-evolution and pathogenic power.* Infect Genet Evol, 2008. **8**(5): p. 708-26.

6. UNAIDS, *UNAIDS: Fact Sheet: 2014 statistics.* 2015.

7. Gebo, K.A., J.A. Fleishman, R. Conviser, J. Hellinger, F.J. Hellinger, J.S. Josephs, P. Keiser, P. Gaist, and R.D. Moore, *Contemporary costs of HIV healthcare in the HAART era.* AIDS, 2010. **24**(17): p. 2705-15.

8. Pace, M.J., L. Agosto, E.H. Graf, and U. O'Doherty, *HIV reservoirs and latency models.* Virology, 2011. **411**(2): p. 344-54.

9. Finzi, D., J. Blankson, J.D. Siliciano, J.B. Margolick, K. Chadwick, T. Pierson, K. Smith, J. Lisziewicz, F. Lori, C. Flexner, T.C. Quinn, R.E. Chaisson, E. Rosenberg, B. Walker, S. Gange, J. Gallant, and R.F. Siliciano, *Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of*

HIV-1, even in patients on effective combination therapy. Nat Med, 1999. **5**(5): p. 512-7.

10. Strain, M.C., H.F. Gunthard, D.V. Havlir, C.C. Ignacio, D.M. Smith, A.J. Leigh-Brown, T.R. Macaranas, R.Y. Lam, O.A. Daly, M. Fischer, M. Opravil, H. Levine, L. Bacheler, C.A. Spina, D.D. Richman, and J.K. Wong, *Heterogeneous clearance rates of long-lived lymphocytes infected with HIV: intrinsic stability predicts lifelong persistence.* Proc Natl Acad Sci U S A, 2003. **100**(8): p. 4819-24.

11. Alexaki, A., Y. Liu, and B. Wigdahl, *Cellular reservoirs of HIV-1 and their role in viral persistence.* Curr HIV Res, 2008. **6**(5): p. 388-400.

12. Watters, S.A., P. Mlcochova, and R.K. Gupta, *Macrophages: the neglected barrier to eradication.* Curr Opin Infect Dis, 2013. **26**(6): p. 561-6.

13. Carter, C.C., A. Onafuwa-Nuga, L.A. McNamara, J.t. Riddell, D. Bixby, M.R. Savona, and K.L. Collins, *HIV-1 infects multipotent progenitor cells causing cell death and establishing latent cellular reservoirs.* Nat Med, 2010. **16**(4): p. 446-51.

14. Blankson, J.N., D. Persaud, and R.F. Siliciano, *The challenge of viral reservoirs in HIV-1 infection.* Annu Rev Med, 2002. **53**: p. 557-93.

15. Matreyek, K.A., I. Oztop, E.O. Freed, and A. Engelman, *Viral latency and potential eradication of HIV-1.* Expert Rev Anti Infect Ther, 2012. **10**(8): p. 855-7.

16. Dahl, V. and S. Palmer, *Establishment of drug-resistant HIV-1 in latent reservoirs.* J Infect Dis, 2009. **199**(9): p. 1258-60.

17. DeMaster, L.K., X. Liu, D.J. VanBelzen, B. Trinite, L. Zheng, L.M. Agosto, S.A. Migueles, M. Connors, L. Sambucetti, D.N. Levy, A.O. Pasternak, and U. O'Doherty, *A Subset of CD4/CD8 Double-Negative T Cells Expresses HIV Proteins in Patients on Antiretroviral Therapy.* J Virol, 2015. **90**(5): p. 2165-79.

18. Palella, F.J., Jr., K.M. Delaney, A.C. Moorman, M.O. Loveless, J. Fuhrer, G.A. Satten, D.J. Aschman, and S.D. Holmberg, *Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators.* N Engl J Med, 1998. **338**(13): p. 853-60.

19. Chun, T.W., L. Stuyver, S.B. Mizell, L.A. Ehler, J.A. Mican, M. Baseler, A.L. Lloyd, M.A. Nowak, and A.S. Fauci, *Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy.* Proc Natl Acad Sci U S A, 1997. **94**(24): p. 13193-7.

20. Davey, R.T., Jr., N. Bhat, C. Yoder, T.W. Chun, J.A. Metcalf, R. Dewar, V. Natarajan, R.A. Lempicki, J.W. Adelsberger, K.D. Miller, J.A. Kovacs, M.A. Polis, R.E. Walker, J. Falloon, H. Masur, D. Gee, M. Baseler, D.S. Dimitrov, A.S. Fauci, and H.C. Lane, *HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression.* Proc Natl Acad Sci U S A, 1999. **96**(26): p. 15109-14.

21. Imamichi, H., K.A. Crandall, V. Natarajan, M.K. Jiang, R.L. Dewar, S. Berg, A. Gaddam, M. Bosche, J.A. Metcalf, R.T. Davey, Jr., and H.C. Lane, *Human immunodeficiency virus type 1 quasi species that rebound after discontinuation of highly active antiretroviral therapy are similar to the viral quasi species present before initiation of therapy.* J Infect Dis, 2001. **183**(1): p. 36-50.

22. Besson, G.J., C.M. Lalama, R.J. Bosch, R.T. Gandhi, M.A. Bedison, E. Aga, S.A. Riddler, D.K. McMahon, F. Hong, and J.W. Mellors, *HIV-1 DNA decay dynamics in blood during more than a decade of suppressive antiretroviral therapy.* Clin Infect Dis, 2014. **59**(9): p. 1312-21.

23. WHO, *Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: a multicentre retrospective cohort study.* Lancet Infect Dis, 2016.

24. Deeks, S.G., *Immune dysfunction, inflammation, and accelerated aging in patients on antiretroviral therapy.* Top HIV Med, 2009. **17**(4): p. 118-23.

25. Falutz, J., *Growth hormone and HIV infection: contribution to disease manifestations and clinical implications.* Best Pract Res Clin Endocrinol Metab, 2011. **25**(3): p. 517-29.

26. Buehring, B., E. Kirchner, Z. Sun, and L. Calabrese, *The frequency of low muscle mass and its overlap with low bone mineral density and lipodystrophy in individuals with HIV--a pilot study using DXA total body composition analysis.* J Clin Densitom, 2012. **15**(2): p. 224-32.

27. Yin, M.T., C.A. Zhang, D.J. McMahon, D.C. Ferris, D. Irani, I. Colon, S. Cremers, and E. Shane, *Higher rates of bone loss in postmenopausal HIV-infected women: a longitudinal study.* J Clin Endocrinol Metab, 2012. **97**(2): p. 554-62.

28. Watkins, C.C. and G.J. Treisman, *Cognitive impairment in patients with AIDS - prevalence and severity.* HIV AIDS (Auckl), 2015. **7**: p. 35-47.

29. Seng, R., C. Goujard, E. Krastinova, P. Miailhes, S. Orr, J.M. Molina, M. Saada, L. Piroth, C. Rouzioux, and L. Meyer, *Influence of lifelong cumulative HIV viremia on long-term recovery of CD4+ cell count and CD4+/CD8+ ratio among patients on combination antiretroviral therapy.* AIDS, 2015. **29**(5): p. 595-607.

30. Friis-Moller, N., C.A. Sabin, R. Weber, A. d'Arminio Monforte, W.M. El-Sadr, P. Reiss, R. Thiebaut, L. Morfeldt, S. De Wit, C. Pradier, G. Calvo, M.G. Law, O. Kirk, A.N. Phillips, and J.D. Lundgren, *Combination antiretroviral therapy and the risk of myocardial infarction.* N Engl J Med, 2003. **349**(21): p. 1993-2003.

31. Gandhi, R.T., L. Zheng, R.J. Bosch, E.S. Chan, D.M. Margolis, S. Read, B. Kallungal, S. Palmer, K. Medvik, M.M. Lederman, N. Alatrakchi, J.M. Jacobson, A. Wiegand, M. Kearney, J.M. Coffin, J.W. Mellors, and J.J. Eron, *The effect of raltegravir intensification on low-level residual viremia in HIV-infected patients on antiretroviral therapy: a randomized controlled trial.* PLoS Med, 2010. **7**(8).

32. Lafeuillade, A., A. Assi, C. Poggi, C. Bresson-Cuquemelle, E. Jullian, and C. Tamalet, *Failure of combined antiretroviral therapy intensification with maraviroc and raltegravir in chronically HIV-1 infected patients to reduce the viral reservoir: the IntensHIV randomized trial.* AIDS Res Ther, 2014. **11**(1): p. 33.

33. Ross, L., M. Johnson, R. DeMasi, Q. Liao, N. Graham, M. Shaefer, and M. St Clair, *Viral genetic heterogeneity in HIV-1-infected individuals is associated with increasing use of HAART and higher viremia.* AIDS, 2000. **14**(7): p. 813-9.

34. Spina, C.A., J. Anderson, N.M. Archin, A. Bosque, J. Chan, M. Famiglietti, W.C. Greene, A. Kashuba, S.R. Lewin, D.M. Margolis, M. Mau, D. Ruelas, S. Saleh, K. Shirakawa, R.F. Siliciano, A. Singhania, P.C. Soto, V.H. Terry, E. Verdin, C. Woelk, S. Wooden, S. Xing, and V. Planelles, *An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and*

resting CD4+ T cells from aviremic patients. PLoS Pathog, 2013. **9**(12): p. e1003834.

35. Krishnan, V. and S.L. Zeichner, *Host cell gene expression during human immunodeficiency virus type 1 latency and reactivation and effects of targeting genes that are differentially expressed in viral latency.* J Virol, 2004. **78**(17): p. 9458-73.

36. Holt, S.E., W.E. Wright, and J.W. Shay, *Regulation of telomerase activity in immortal cell lines.* Mol Cell Biol, 1996. **16**(6): p. 2932-9.

37. Bosque, A. and V. Planelles, *Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells.* Blood, 2009. **113**(1): p. 58-65.

38. Buzon, M.J., M. Massanella, J.M. Llibre, A. Esteve, V. Dahl, M.C. Puertas, J.M. Gatell, P. Domingo, R. Paredes, M. Sharkey, S. Palmer, M. Stevenson, B. Clotet, J. Blanco, and J. Martinez-Picado, *HIV-1 replication and immune dynamics are affected by raltegravir intensification of HAART-suppressed subjects.* Nat Med, 2010. **16**(4): p. 460-5.

39. Qu, X., P. Wang, D. Ding, L. Li, H. Wang, L. Ma, X. Zhou, S. Liu, S. Lin, X. Wang, G. Zhang, L. Liu, J. Wang, F. Zhang, D. Lu, and H. Zhu, *Zinc-finger-nucleases mediate specific and efficient excision of HIV-1 proviral DNA from infected and latently infected human T cells.* Nucleic Acids Res, 2013. **41**(16): p. 7771-82.

40. Holt, N., J. Wang, K. Kim, G. Friedman, X. Wang, V. Taupin, G.M. Crooks, D.B. Kohn, P.D. Gregory, M.C. Holmes, and P.M. Cannon, *Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo.* Nat Biotechnol, 2010. **28**(8): p. 839-47.

41. Yi, G., J.G. Choi, P. Bharaj, S. Abraham, Y. Dang, T. Kafri, O. Alozie, M.N. Manjunath, and P. Shankar, *CCR5 Gene Editing of Resting CD4(+) T Cells by Transient ZFN Expression From HIV Envelope Pseudotyped Nonintegrating Lentivirus Confers HIV-1 Resistance in Humanized Mice.* Mol Ther Nucleic Acids, 2014. **3**: p. e198.

42. Tebas, P., D. Stein, W.W. Tang, I. Frank, S.Q. Wang, G. Lee, S.K. Spratt, R.T. Surosky, M.A. Giedlin, G. Nichol, M.C. Holmes, P.D. Gregory, D.G. Ando, M. Kalos, R.G. Collman, G. Binder-Scholl, G. Plesa, W.T. Hwang, B.L. Levine,

and C.H. June, *Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV.* N Engl J Med, 2014. **370**(10): p. 901-10.

43. Aubert, M., B.Y. Ryu, L. Banks, D.J. Rawlings, A.M. Scharenberg, and K.R. Jerome, *Successful targeting and disruption of an integrated reporter lentivirus using the engineered homing endonuclease Y2 I-AniI.* PLoS One, 2011. **6**(2): p. e16825.

44. Archin, N.M., A.L. Liberty, A.D. Kashuba, S.K. Choudhary, J.D. Kuruc, A.M. Crooks, D.C. Parker, E.M. Anderson, M.F. Kearney, M.C. Strain, D.D. Richman, M.G. Hudgens, R.J. Bosch, J.M. Coffin, J.J. Eron, D.J. Hazuda, and D.M. Margolis, *Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy.* Nature, 2012. **487**(7408): p. 482-5.

45. Anderson, J.L., R. Fromentin, G.M. Corbelli, L. Ostergaard, and A.L. Ross, *Progress towards an HIV cure: update from the 2014 International AIDS Society Symposium.* AIDS Res Hum Retroviruses, 2015. **31**(1): p. 36-44.

46. Cummins, N.W., A.M. Sainski, H. Dai, S. Natesampillai, Y.P. Pang, G.D. Bren, M.C. de Araujo Correia, R. Sampath, S.A. Rizza, D. O'Brien, J.D. Yao, S.H. Kaufmann, and A.D. Badley, *Prime, Shock, and Kill: Priming CD4 T Cells from HIV Patients with a BCL-2 Antagonist before HIV Reactivation Reduces HIV Reservoir Size.* J Virol, 2016. **90**(8): p. 4032-48.

47. Jones, R.B., S. Mueller, R. O'Connor, K. Rimpel, D.D. Sloan, D. Karel, H.C. Wong, E.K. Jeng, A.S. Thomas, J.B. Whitney, S.Y. Lim, C. Kovacs, E. Benko, S. Karandish, S.H. Huang, M.J. Buzon, M. Lichterfeld, A. Irrinki, J.P. Murry, A. Tsai, H. Yu, R. Geleziunas, A. Trocha, M.A. Ostrowski, D.J. Irvine, and B.D. Walker, *A Subset of Latency-Reversing Agents Expose HIV-Infected Resting CD4+ T-Cells to Recognition by Cytotoxic T-Lymphocytes.* PLoS Pathog, 2016. **12**(4): p. e1005545.

48. Hill, A.L., D.I. Rosenbloom, F. Fu, M.A. Nowak, and R.F. Siliciano, *Predicting the outcomes of treatment to eradicate the latent reservoir for HIV-1.* Proc Natl Acad Sci U S A, 2014. **111**(37): p. 13475-80.

49. Chun, T.W., D. Engel, S.B. Mizell, L.A. Ehler, and A.S. Fauci, *Induction of HIV-1 replication in latently infected CD4+ T cells using a combination of cytokines.* J Exp Med, 1998. **188**(1): p. 83-91.

50. Boehm, D., V. Calvanese, R.D. Dar, S. Xing, S. Schroeder, L. Martins, K. Aull, P.C. Li, V. Planelles, J.E. Bradner, M.M. Zhou, R.F. Siliciano, L. Weinberger,

E. Verdin, and M. Ott, *BET bromodomain-targeting compounds reactivate HIV from latency via a Tat-independent mechanism.* Cell Cycle, 2013. **12**(3): p. 452-62.

51. Doyon, G., J. Zerbato, J.W. Mellors, and N. Sluis-Cremer, *Disulfiram reactivates latent HIV-1 expression through depletion of the phosphatase and tensin homolog.* AIDS, 2013. **27**(2): p. F7-F11.

52. Williams, S.A., L.F. Chen, H. Kwon, D. Fenard, D. Bisgrove, E. Verdin, and W.C. Greene, *Prostratin antagonizes HIV latency by activating NF-kappaB.* J Biol Chem, 2004. **279**(40): p. 42008-17.

53. Bouchat, S., J.S. Gatot, K. Kabeya, C. Cardona, L. Colin, G. Herbein, S. De Wit, N. Clumeck, O. Lambotte, C. Rouzioux, O. Rohr, and C. Van Lint, *Histone methyltransferase inhibitors induce HIV-1 recovery in resting CD4(+) T cells from HIV-1-infected HAART-treated patients.* AIDS, 2012. **26**(12): p. 1473-82.

54. Archin, N.M., A. Espeseth, D. Parker, M. Cheema, D. Hazuda, and D.M. Margolis, *Expression of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid.* AIDS Res Hum Retroviruses, 2009. **25**(2): p. 207-12.

55. Prince, H.M., M.J. Bishton, and S.J. Harrison, *Clinical studies of histone deacetylase inhibitors.* Clin Cancer Res, 2009. **15**(12): p. 3958-69.

56. Prince, H.M. and M. Dickinson, *Romidepsin for cutaneous T-cell lymphoma.* Clin Cancer Res, 2012. **18**(13): p. 3509-15.

57. Goldenberg, M.M., *Overview of drugs used for epilepsy and seizures: etiology, diagnosis, and treatment.* P T, 2010. **35**(7): p. 392-415.

58. Elliott, J.H., F. Wightman, A. Solomon, K. Ghneim, J. Ahlers, M.J. Cameron, M.Z. Smith, T. Spelman, J. McMahon, P. Velayudham, G. Brown, J. Roney, J. Watson, M.H. Prince, J.F. Hoy, N. Chomont, R. Fromentin, F.A. Procopio, J. Zeidan, S. Palmer, L. Odevall, R.W. Johnstone, B.P. Martin, E. Sinclair, S.G. Deeks, D.J. Hazuda, P.U. Cameron, R.-P. Sékaly, and S.R. Lewin, *Activation of HIV Transcription with Short-Course Vorinostat in HIV-Infected Patients on Suppressive Antiretroviral Therapy.* PLoS Pathog., 2014. **10**(11): p. e1004473.

59. Rasmussen, T.A., M. Tolstrup, C.R. Brinkmann, R. Olesen, C. Erikstrup, A. Solomon, A. Winckelmann, S. Palmer, C. Dinarello, M. Buzon, M. Lichterfeld,

S.R. Lewin, L. Ostergaard, and O.S. Sogaard, *Panobinostat, a histone deacetylase inhibitor, for latent-virus reactivation in HIV-infected patients on suppressive antiretroviral therapy: a phase 1/2, single group, clinical trial.* Lancet HIV, 2014. **1**(1): p. e13-21.

60. Margolis, D.M. and D.J. Hazuda, *Combined approaches for HIV cure.* Curr Opin HIV AIDS, 2013. **8**(3): p. 230-5.

61. De Crignis, E. and T. Mahmoudi, *HIV eradication: combinatorial approaches to activate latent viruses.* Viruses, 2014. **6**(11): p. 4581-608.

62. Mohammadi, P., A. Ciuffi, and N. Beerenwinkel, *Dynamic models of viral replication and latency.* Curr Opin HIV AIDS, 2015. **10**(2): p. 90-5.

63. Tripathy, M.K., M.E. McManamy, B.D. Burch, N.M. Archin, and D.M. Margolis, *H3K27 Demethylation at the Proviral Promoter Sensitizes Latent HIV to the Effects of Vorinostat in Ex Vivo Cultures of Resting CD4+ T Cells.* J Virol, 2015. **89**(16): p. 8392-405.

64. Perez, M., A.G. de Vinuesa, G. Sanchez-Duffhues, N. Marquez, M.L. Bellido, M.A. Munoz-Fernandez, S. Moreno, T.P. Castor, M.A. Calzado, and E. Munoz, *Bryostatin-1 synergizes with histone deacetylase inhibitors to reactivate HIV-1 from latency.* Curr HIV Res, 2010. **8**(6): p. 418-29.

65. Laird, G.M., C.K. Bullen, D.I.S. Rosenbloom, A.R. Martin, A.L. Hill, C.M. Durand, J.D. Siliciano, and R.F. Siliciano, *Ex vivo analysis identifies effective HIV-1 latency–reversing drug combinations.* J Clin Invest., 2015. **125**(5): p. 1901-1912.

66. Reuse, S., M. Calao, K. Kabeya, A. Guiguen, J.S. Gatot, V. Quivy, C. Vanhulle, A. Lamine, D. Vaira, D. Demonte, V. Martinelli, E. Veithen, T. Cherrier, V. Avettand, S. Poutrel, J. Piette, Y. de Launoit, M. Moutschen, A. Burny, C. Rouzioux, S. De Wit, G. Herbein, O. Rohr, Y. Collette, O. Lambotte, N. Clumeck, and C. Van Lint, *Synergistic activation of HIV-1 expression by deacetylase inhibitors and prostratin: implications for treatment of latent infection.* PLoS One, 2009. **4**(6): p. e6093.

67. Burnett, J.C., K.I. Lim, A. Calafi, J.J. Rossi, D.V. Schaffer, and A.P. Arkin, *Combinatorial latency reactivation for HIV-1 subtypes and variants.* J Virol, 2010. **84**(12): p. 5958-74.

68. Wozniak, M.B., R. Villuendas, J.R. Bischoff, C.B. Aparicio, J.F. Martínez Leal, P. de La Cueva, M.E. Rodriguez, B. Herreros, D. Martin-Perez, M.I. Longo, M. Herrera, M.Á. Piris, and P.L. Ortiz-Romero, *Vorinostat interferes with the signaling transduction pathway of T-cell receptor and synergizes with phosphoinositide-3 kinase inhibitors in cutaneous T-cell lymphoma.* Haematologica, 2010. **95**(4): p. 613-621.

69. LaBonte, M.J., P.M. Wilson, W. Fazzone, S. Groshen, H.J. Lenz, and R.D. Ladner, *DNA microarray profiling of genes differentially regulated by the histone deacetylase inhibitors vorinostat and LBH589 in colon cancer cell lines.* BMC Med Genomics, 2009. **2**: p. 67.

70. Mann, B.S., J.R. Johnson, M.H. Cohen, R. Justice, and R. Pazdur, *FDA approval summary: vorinostat for treatment of advanced primary cutaneous T-cell lymphoma.* Oncologist, 2007. **12**(10): p. 1247-52.

71. Kavanaugh, S.M., L.A. White, and J.M. Kolesar, *Vorinostat: A novel therapy for the treatment of cutaneous T-cell lymphoma.* Am J Health Syst Pharm, 2010. **67**(10): p. 793-7.

72. Li, Z., J. Guo, Y. Wu, and Q. Zhou, *The BET bromodomain inhibitor JQ1 activates HIV latency through antagonizing Brd4 inhibition of Tat-transactivation.* Nucleic Acids Res, 2013. **41**(1): p. 277-87.

73. Lu, P., X. Qu, Y. Shen, Z. Jiang, P. Wang, H. Zeng, H. Ji, J. Deng, X. Yang, X. Li, H. Lu, and H. Zhu, *The BET inhibitor OTX015 reactivates latent HIV-1 through P-TEFb.* Sci Rep, 2016. **6**: p. 24100.

74. Contreras, X., M. Barboric, T. Lenasi, and B.M. Peterlin, *HMBA releases P-TEFb from HEXIM1 and 7SK snRNA via PI3K/Akt and activates HIV transcription.* PLoS Pathog, 2007. **3**(10): p. 1459-69.

75. Diaz, L., M. Martinez-Bonet, J. Sanchez, A. Fernandez-Pineda, J.L. Jimenez, E. Munoz, S. Moreno, S. Alvarez, and M.A. Munoz-Fernandez, *Bryostatin activates HIV-1 latent expression in human astrocytes through a PKC and NF-kB-dependent mechanism.* Sci Rep, 2015. **5**: p. 12442.

76. Jiang, G., E.A. Mendes, P. Kaiser, S. Sankaran-Walters, Y. Tang, M.G. Weber, G.P. Melcher, G.R. Thompson, 3rd, A. Tanuri, L.F. Pianowski, J.K. Wong, and S. Dandekar, *Reactivation of HIV latency by a newly modified Ingenol derivative via protein kinase Cdelta-NF-kappaB signaling.* AIDS, 2014. **28**(11): p. 1555-66.

77. Marquez, N., M.A. Calzado, G. Sanchez-Duffhues, M. Perez, A. Minassi, A. Pagani, G. Appendino, L. Diaz, M.A. Munoz-Fernandez, and E. Munoz, *Differential effects of phorbol-13-monoesters on human immunodeficiency virus reactivation.* Biochem Pharmacol, 2008. **75**(6): p. 1370-80.

78. Bernhard, W., K. Barreto, A. Saunders, M.S. Dahabieh, P. Johnson, and I. Sadowski, *The Suv39H1 methyltransferase inhibitor chaetocin causes induction of integrated HIV-1 without producing a T cell response.* FEBS Lett, 2011. **585**(22): p. 3549-54.

79. Palmer, C.S. and S.M. Crowe, *Panobinostat clinical trial highlights the challenges towards an HIV cure.* Lancet HIV, 2014. **1**(3): p. e103.

80. Wei, D.G., V. Chiang, E. Fyne, M. Balakrishnan, T. Barnes, M. Graupe, J. Hesselgesser, A. Irrinki, J.P. Murry, G. Stepan, K.M. Stray, A. Tsai, H. Yu, J. Spindler, M. Kearney, C.A. Spina, D. McMahon, J. Lalezari, D. Sloan, J. Mellors, R. Geleziunas, and T. Cihlar, *Histone deacetylase inhibitor romidepsin induces HIV expression in CD4 T cells from patients on suppressive antiretroviral therapy at concentrations achieved by clinical dosing.* PLoS Pathog, 2014. **10**(4): p. e1004071.

81. Wightman, F., H.K. Lu, A.E. Solomon, S. Saleh, A.N. Harman, A.L. Cunningham, L. Gray, M. Churchill, P.U. Cameron, A.E. Dear, and S.R. Lewin, *Entinostat is a histone deacetylase inhibitor selective for class 1 histone deacetylases and activates HIV production from latently infected primary T cells.* AIDS, 2013. **27**(18): p. 2853-62.

82. Banga, R., F.A. Procopio, M. Cavassini, and M. Perreau, *In Vitro Reactivation of Replication-Competent and Infectious HIV-1 by Histone Deacetylase Inhibitors.* J Virol, 2015. **90**(4): p. 1858-71.

83. Rasmussen, T.A., O. Schmeltz Sogaard, C. Brinkmann, F. Wightman, S.R. Lewin, J. Melchjorsen, C. Dinarello, L. Ostergaard, and M. Tolstrup, *Comparison of HDAC inhibitors in clinical development: effect on HIV production in latently infected cells and T-cell activation.* Hum Vaccin Immunother, 2013. **9**(5): p. 993-1001.

84. Routy, J.P., C.L. Tremblay, J.B. Angel, B. Trottier, D. Rouleau, J.G. Baril, M. Harris, S. Trottier, J. Singer, N. Chomont, R.P. Sekaly, and M.R. Boulassel, *Valproic acid in association with highly active antiretroviral therapy for reducing systemic HIV-1 reservoirs: results from a multicentre randomized clinical study.* HIV Med, 2012. **13**(5): p. 291-6.

85. Ke, R., S.R. Lewin, J.H. Elliott, and A.S. Perelson, *Modeling the Effects of Vorinostat In Vivo Reveals both Transient and Delayed HIV Transcriptional Activation and Minimal Killing of Latently Infected Cells.* PLoS Pathog, 2015. **11**(10): p. e1005237.

86. Spivak, A.M., A. Andrade, E. Eisele, R. Hoh, P. Bacchetti, N.N. Bumpus, F. Emad, R. Buckheit, 3rd, E.F. McCance-Katz, J. Lai, M. Kennedy, G. Chander, R.F. Siliciano, J.D. Siliciano, and S.G. Deeks, *A pilot study assessing the safety and latency-reversing activity of disulfiram in HIV-1-infected adults on antiretroviral therapy.* Clin Infect Dis, 2014. **58**(6): p. 883-90.

87. Karn, J. *Supercharging NK Cells for Cure; Conference on Retroviruses and Opportunistic Infections*. in *Conference on Retroviruses and Opportunistic Infections*. 2016. Boston, Massachusetts.

88. Sloan, D.D., C.Y. Lam, A. Irrinki, L. Liu, A. Tsai, C.S. Pace, J. Kaur, J.P. Murry, M. Balakrishnan, P.A. Moore, S. Johnson, J.L. Nordstrom, T. Cihlar, and S. Koenig, *Targeting HIV Reservoir in Infected CD4 T Cells by Dual-Affinity Re-targeting Molecules (DARTs) that Bind HIV Envelope and Recruit Cytotoxic T Cells.* PLoS Pathog, 2015. **11**(11): p. e1005233.

89. Sung, J.A., J. Pickeral, L. Liu, S.A. Stanfield-Oakley, C.Y. Lam, C. Garrido, J. Pollara, C. LaBranche, M. Bonsignori, M.A. Moody, Y. Yang, R. Parks, N. Archin, B. Allard, J. Kirchherr, J.D. Kuruc, C.L. Gay, M.S. Cohen, C. Ochsenbauer, K. Soderberg, H.X. Liao, D. Montefiori, P. Moore, S. Johnson, S. Koenig, B.F. Haynes, J.L. Nordstrom, D.M. Margolis, and G. Ferrari, *Dual-Affinity Re-Targeting proteins direct T cell-mediated cytolysis of latently HIV-infected cells.* J Clin Invest, 2015. **125**(11): p. 4077-90.

90. Archin, N.M., R. Bateson, M.K. Tripathy, A.M. Crooks, K.-H. Yang, N.P. Dahl, M.F. Kearney, E.M. Anderson, J.M. Coffin, M.C. Strain, D.D. Richman, K.R. Robertson, A.D. Kashuba, R.J. Bosch, D.J. Hazuda, J.D. Kuruc, J.J. Eron, and D.M. Margolis, *HIV-1 Expression Within Resting CD4+ T Cells After Multiple Doses of Vorinostat.* J Infect Dis., 2014. **210**(5): p. 728-735.

91. White, C.H., H.E. Johnston, B. Moesker, A. Manousopoulou, D.M. Margolis, D.D. Richman, C.A. Spina, S.D. Garbis, C.H. Woelk, and N. Beliakova-Bethell, *Mixed effects of suberoylanilide hydroxamic acid (SAHA) on the host transcriptome and proteome and their implications for HIV reactivation from latency.* Antiviral Res, 2015. **123**: p. 78-85.

92. Massanella, M., A. Singhania, N. Beliakova-Bethell, R. Pier, S.M. Lada, C.H. White, J. Perez-Santiago, J. Blanco, D.D. Richman, S.J. Little, and C.H.

Woelk, *Differential gene expression in HIV-infected individuals following ART.* Antiviral Res, 2013. **100**(2): p. 420-8.

93. Wie, S.H., P. Du, T.Q. Luong, S.E. Rought, N. Beliakova-Bethell, J. Lozach, J. Corbeil, R.S. Kornbluth, D.D. Richman, and C.H. Woelk, *HIV downregulates interferon-stimulated genes in primary macrophages.* J Interferon Cytokine Res, 2013. **33**(2): p. 90-5.

94. White, C.H., B. Moesker, A. Ciuffi, and N. Beliakova-Bethell, *Systems biology applications to study mechanisms of human immunodeficiency virus latency and reactivation.* World Journal of Clinical Infectious Diseases, 2016.

95. Mohammadi, P., J. di Iulio, M. Munoz, R. Martinez, I. Bartha, M. Cavassini, C. Thorball, J. Fellay, N. Beerenwinkel, A. Ciuffi, and A. Telenti, *Dynamics of HIV latency and reactivation in a primary CD4+ T cell model.* PLoS Pathog, 2014. **10**(5): p. e1004156.

96. Sherrill-Mix, S., M.K. Lewinski, M. Famiglietti, A. Bosque, N. Malani, K.E. Ocwieja, C.C. Berry, D. Looney, L. Shan, L.M. Agosto, M.J. Pace, R.F. Siliciano, U. O'Doherty, J. Guatelli, V. Planelles, and F.D. Bushman, *HIV latency and integration site placement in five cell-based models.* Retrovirology, 2013. **10**: p. 90.

97. Saayman, S.M., D.C. Lazar, T.A. Scott, J.R. Hart, M. Takahashi, J.C. Burnett, V. Planelles, K.V. Morris, and M.S. Weinberg, *Potent and Targeted Activation of Latent HIV-1 Using the CRISPR/dCas9 Activator Complex.* Mol Ther, 2016. **24**(3): p. 488-98.

98. Bumgarner, R., *Overview of DNA microarrays: types, applications, and their future.* Curr Protoc Mol Biol, 2013. **Chapter 22**: p. Unit 22 1.

99. Singh, A. and N. Kumar, *A REVIEW ON DNA MICROARRAY TECHNOLOGY.* IJCRR, 2013. **5**(22): p. 01-05.

100. Tang, T., N. Francois, A. Glatigny, N. Agier, M.H. Mucchielli, L. Aggerbeck, and H. Delacroix, *Expression ratio evaluation in two-colour microarray experiments is significantly improved by correcting image misalignment.* Bioinformatics, 2007. **23**(20): p. 2686-91.

101. Oberthuer, A., D. Juraeva, L. Li, Y. Kahlert, F. Westermann, R. Eils, F. Berthold, L. Shi, R.D. Wolfinger, M. Fischer, and B. Brors, *Comparison of*

performance of one-color and two-color gene-expression analyses in predicting clinical endpoints of neuroblastoma patients. Pharmacogenomics J, 2010. **10**(4): p. 258-66.

102. Yao, B., S.N. Rakhade, Q. Li, S. Ahmed, R. Krauss, S. Draghici, and J.A. Loeb, *Accuracy of cDNA microarray methods to detect small gene expression changes induced by neuregulin on breast epithelial cells.* BMC Bioinformatics, 2004. **5**: p. 99.

103. Zhao, S., W.P. Fung-Leung, A. Bittner, K. Ngo, and X. Liu, *Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells.* PLoS One, 2014. **9**(1): p. e78644.

104. Draghici, S., P. Khatri, A.C. Eklund, and Z. Szallasi, *Reliability and reproducibility issues in DNA microarray measurements.* Trends Genet, 2006. **22**(2): p. 101-9.

105. Illumina. *An Introduction to Next-Generation Sequencing Technology*. 2016; Available from: http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.

106. Chen, C.Y., *DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present.* Front Microbiol, 2014. **5**: p. 305.

107. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.

108. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.

109. Robinson, M., D.J. McCarthy, Y. Chen, A. Lun, and G.K. Smyth. *edgeR: differential expression analysis of digital gene expression data*. 2012; Available from: http://www.bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/edgeR.pdf.

110. Pruitt K., B.G., Tatusova T., *The Reference Sequence (RefSeq) Database. 2002 Oct 9 [Updated 2012 Apr 6]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 18. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21091/*. 2002.

111. Wagner, L. and R. Agarwala, *UniGene. 2013 Nov 14. In: The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-. Available from: http://www.ncbi.nlm.nih.gov/books/NBK169437/*.

112. Ilumina. *HumanHT-12 v3 Expression BeadChip*. 2016; Available from: http://www.illumina.com/Documents/products/datasheets/datasheet_humanht_12.pdf.

113. Zeidan, B., T.R. Jackson, S.E. Larkin, R.I. Cutress, G.R. Coulton, M. Ashton-Key, N. Murray, G. Packham, V. Gorgoulis, S.D. Garbis, and P.A. Townsend, *Annexin A3 is a mammary marker and a potential neoplastic breast cell therapeutic target.* Oncotarget, 2015. **6**(25): p. 21421-7.

114. Aggarwal, S. and A.K. Yadav, *Dissecting the iTRAQ Data Analysis.* Methods Mol Biol, 2016. **1362**: p. 277-91.

115. Fuller, H.R. and G.E. Morris, *Quantitative Proteomics Using iTRAQ Labeling and Mass Spectrometry, Integrative Proteomics, Dr. Hon-Chiu Leung (Ed.), ISBN: 978-953-51-0070-6, InTech, Available from: http://www.intechopen.com/books/integrative-proteomics/quantitative-proteomics-using-itraq-labeling-andmass-spectrometry.* 2012.

116. Wang, H., S. Alvarez, and L.M. Hicks, *Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two Chlamydomonas reinhardtii strains of interest for biofuels engineering.* J Proteome Res, 2012. **11**(1): p. 487-501.

117. Issaq, H. and T. Veenstra, *Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives.* Biotechniques, 2008. **44**(5): p. 697-8, 700.

118. Gygi, S.P., B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, and R. Aebersold, *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.

119. Ong, S.E., B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, and M. Mann, *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.

120. Wiese, S., K.A. Reidegeld, H.E. Meyer, and B. Warscheid, *Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research.* Proteomics, 2007. **7**(3): p. 340-50.

121. Megger, D.A., T. Bracht, H.E. Meyer, and B. Sitek, *Label-free quantification in clinical proteomics.* Biochim Biophys Acta, 2013. **1834**(8): p. 1581-90.

122. UniProt, C., *UniProt: a hub for protein information.* Nucleic Acids Res, 2015. **43**(Database issue): p. D204-12.

123. Bantscheff, M., S. Lemeer, M.M. Savitski, and B. Kuster, *Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present.* Anal Bioanal Chem, 2012. **404**(4): p. 939-65.

124. Wang, H., Y. Li, L. Yang, B. Yu, P. Yan, M. Pang, X. Li, H. Yang, G. Zheng, J. Xie, and R. Guo, *Mass spectrometry-based, label-free quantitative proteomics of round spermatids in mice.* Mol Med Rep, 2014. **10**(4): p. 2009-24.

125. Latosinska, A., K. Vougas, M. Makridakis, J. Klein, W. Mullen, M. Abbas, K. Stravodimos, I. Katafigiotis, A.S. Merseburger, J. Zoidakis, H. Mischak, A. Vlahou, and V. Jankowski, *Comparative Analysis of Label-Free and 8-Plex iTRAQ Approach for Quantitative Tissue Proteomic Analysis.* PLoS One, 2015. **10**(9): p. e0137048.

126. Trinh, H.V., J. Grossmann, P. Gehrig, B. Roschitzki, R. Schlapbach, U.F. Greber, and S. Hemmi, *iTRAQ-Based and Label-Free Proteomics Approaches for Studies of Human Adenovirus Infections.* Int J Proteomics, 2013. **2013**: p. 581862.

127. Reardon, B., N. Beliakova-Bethell, C.A. Spina, A. Singhania, D.M. Margolis, D.R. Richman, and C.H. Woelk, *Dose-responsive gene expression in suberoylanilide hydroxamic acid-treated resting CD4+ T cells.* AIDS, 2015. **29**(17): p. 2235-44.

128. Martins, L.J., P. Bonczkowski, A.M. Spivak, W. De Spiegelaere, C.L. Novis, A.B. DePaula-Silva, E. Malatinkova, W. Typsteen, A. Bosque, L. Vanderkerckhove, and V. Planelles, *Modeling HIV-1 Latency in Primary T Cells Using a Replication-Competent Virus.* AIDS Res Hum Retroviruses, 2015.

129. Bosque, A. and V. Planelles, *Studies of HIV-1 latency in an ex vivo model that uses primary central memory T cells.* Methods, 2011. **53**(1): p. 54-61.

130. Malone, J.H. and B. Oliver, *Microarrays, deep sequencing and the true measure of the transcriptome.* BMC Biol, 2011. **9**: p. 34.

131. Homer, N., B. Merriman, and S.F. Nelson, *BFAST: an alignment tool for large scale genome resequencing.* PLoS One, 2009. **4**(11): p. e7767.

132. Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg, *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.

133. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

134. Trapnell, C., B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, and L. Pachter, *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotechnol, 2010. **28**(5): p. 511-5.

135. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.

136. Leng, N., J.A. Dawson, J.A. Thomson, V. Ruotti, A.I. Rissman, B.M. Smits, J.D. Haag, M.N. Gould, R.M. Stewart, and C. Kendziorski, *EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments.* Bioinformatics, 2013. **29**(8): p. 1035-43.

137. Trapnell, C., D.G. Hendrickson, M. Sauvageau, L. Goff, J.L. Rinn, and L. Pachter, *Differential analysis of gene regulation at transcript resolution with RNA-seq.* Nat Biotechnol, 2013. **31**(1): p. 46-53.

138. Bandyopadhyay, S., R. Kelley, and T. Ideker, *Discovering regulated networks during HIV-1 latency and reactivation.* Pac Symp Biocomput, 2006: p. 354-66.

139. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. **9**: p. 559.

140. Simon, R., A. Lam, M.C. Li, M. Ngan, S. Menenzes, and Y. Zhao, *Analysis of gene expression data using BRB-ArrayTools.* Cancer Inform, 2007. **3**: p. 11-7.

141. Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth, *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.

142. Andersen, J.L., J.L. DeHart, E.S. Zimmerman, O. Ardon, B. Kim, G. Jacquot, S. Benichou, and V. Planelles, *HIV-1 Vpr-induced apoptosis is cell cycle dependent and requires Bax but not ANT.* PLoS Pathog, 2006. **2**(12): p. e127.

143. Langlade-Demoyen, P., F. Michel, A. Hoffenbach, E. Vilmer, G. Dadaglio, F. Garicia-Pons, C. Mayaud, B. Autran, S. Wain-Hobson, and F. Plata, *Immune recognition of AIDS virus antigens by human and murine cytotoxic T lymphocytes.* J Immunol, 1988. **141**(6): p. 1949-57.

144. Zhu, Y., H.A. Gelbard, M. Roshal, S. Pursell, B.D. Jamieson, and V. Planelles, *Comparison of cell cycle arrest, transactivation, and apoptosis induced by the simian immunodeficiency virus SIVagm and human immunodeficiency virus type 1 vpr genes.* J Virol, 2001. **75**(8): p. 3791-801.

145. Vermeire, J., E. Naessens, H. Vanderstraeten, A. Landi, V. Iannucci, A. Van Nuffel, T. Taghon, M. Pizzato, and B. Verhasselt, *Quantification of reverse transcriptase activity by real-time PCR as a fast and accurate method for titration of HIV, lenti- and retroviral vectors.* PLoS One, 2012. **7**(12): p. e50859.

146. R Core Team. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* 2013; Available from: http://www.R-project.org.

147. Robinson, J.T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, and J.P. Mesirov, *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.

148. Babraham Bioinformatics. *FastQC*. 2014; Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

149. Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing, *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

150. Seyednasrollah, F., A. Laiho, and L.L. Elo, *Comparison of software packages for detecting differential expression in RNA-seq studies.* Brief Bioinform, 2015. **16**(1): p. 59-70.

151. Soneson, C. and M. Delorenzi, *A comparison of methods for differential expression analysis of RNA-seq data.* BMC Bioinformatics, 2013. **14**: p. 91.

152. Homer, N. and B. Merriman. *TMAP: the Torrent Mapping Alignment Program*. Available from: https://github.com/iontorrent/TMAP.

153. Lawrence, M., W. Huber, H. Pages, P. Aboyoun, M. Carlson, R. Gentleman, M.T. Morgan, and V.J. Carey, *Software for computing and annotating genomic ranges.* PLoS Comput Biol, 2013. **9**(8): p. e1003118.

154. Gentleman, R.C., V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A.J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J.Y. Yang, and J. Zhang, *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, 2004. **5**(10): p. R80.

155. Bonczkowski, P., W. De Spiegelaere, A. Bosque, C.H. White, A. Van Nuffel, E. Malatinkova, M. Kiselinova, W. Trypsteen, W. Witkowski, J. Vermeire, B. Verhasselt, L. Martins, C.H. Woelk, V. Planelles, and L. Vandekerckhove, *Replication competent virus as an important source of bias in HIV latency models utilizing single round viral constructs.* Retrovirology, 2014. **11**: p. 70.

156. Ragione, F.D., V. Cucciolla, V. Criniti, S. Indaco, A. Borriello, and V. Zappia, *p21Cip1 gene expression is modulated by Egr1: a novel regulatory mechanism involved in the resveratrol antiproliferative effect.* J Biol Chem, 2003. **278**(26): p. 23360-8.

157. Levin, W.J., M.F. Press, R.B. Gaynor, V.P. Sukhatme, T.C. Boone, P.T. Reissmann, R.A. Figlin, E.C. Holmes, L.M. Souza, and D.J. Slamon, *Expression patterns of immediate early transcription factors in human non-small cell lung cancer. The Lung Cancer Study Group.* Oncogene, 1995. **11**(7): p. 1261-9.

158. Huang, R.P., Y. Fan, I. de Belle, C. Niemeyer, M.M. Gottardis, D. Mercola, and E.D. Adamson, *Decreased Egr-1 expression in human, mouse and rat mammary cells and tissues correlates with tumor formation.* Int J Cancer, 1997. **72**(1): p. 102-9.

159. Finzi, D., M. Hermankova, T. Pierson, L.M. Carruth, C. Buck, R.E. Chaisson, T.C. Quinn, K. Chadwick, J. Margolick, R. Brookmeyer, J. Gallant, M. Markowitz, D.D. Ho, D.D. Richman, and R.F. Siliciano, *Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy.* Science, 1997. **278**(5341): p. 1295-300.

160. Wong, J.K., M. Hezareh, H.F. Gunthard, D.V. Havlir, C.C. Ignacio, C.A. Spina, and D.D. Richman, *Recovery of replication-competent HIV despite prolonged suppression of plasma viremia.* Science, 1997. **278**(5341): p. 1291-5.

161. Iglesias-Ussel, M., C. Vandergeeten, L. Marchionni, N. Chomont, and F. Romerio, *High levels of CD2 expression identify HIV-1 latently infected resting memory CD4+ T cells in virally suppressed subjects.* J Virol, 2013. **87**(16): p. 9148-58.

162. Hermankova, M., J.D. Siliciano, Y. Zhou, D. Monie, K. Chadwick, J.B. Margolick, T.C. Quinn, and R.F. Siliciano, *Analysis of human immunodeficiency virus type 1 gene expression in latently infected resting CD4+ T lymphocytes in vivo.* J Virol, 2003. **77**(13): p. 7383-92.

163. Lassen, K.G., K.X. Ramyar, J.R. Bailey, Y. Zhou, and R.F. Siliciano, *Nuclear retention of multiply spliced HIV-1 RNA in resting CD4+ T cells.* PLoS Pathog, 2006. **2**(7): p. e68.

164. Komarov, P.G., E.A. Komarova, R.V. Kondratov, K. Christov-Tselkov, J.S. Coon, M.V. Chernov, and A.V. Gudkov, *A chemical inhibitor of p53 that protects mice from the side effects of cancer therapy.* Science, 1999. **285**(5434): p. 1733-7.

165. Hannon Lab. *FASTX Toolkit*. Available from: http://hannonlab.cshl.edu/fastx_toolkit/index.html.

166. Harrow, J., A. Frankish, J.M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B.L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J.M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigo, and T.J. Hubbard, *GENCODE: the reference human genome annotation for The ENCODE Project.* Genome Res, 2012. **22**(9): p. 1760-74.

167. ThermoFisher Scientific. *ERCC Reference*. 2015; Available from: https://tools.thermofisher.com/content/sfs/manuals/cms_095047.txt.

168. Aanes, H., C. Winata, L.F. Moen, O. Ostrup, S. Mathavan, P. Collas, T. Rognes, and P. Alestrom, *Normalization of RNA-sequencing data from samples with varying mRNA levels.* PLoS One, 2014. **9**(2): p. e89158.

169. Risso, D., J. Ngai, T.P. Speed, and S. Dudoit, *Normalization of RNA-seq data using factor analysis of control genes or samples.* Nat Biotechnol, 2014. **32**(9): p. 896-902.

170. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.

171. Chen, J., E.E. Bardes, B.J. Aronow, and A.G. Jegga, *ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W305-11.

172. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.

173. Shannon, P., A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.

174. Nicol, J.W., G.A. Helt, S.G. Blanchard, Jr., A. Raja, and A.E. Loraine, *The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets.* Bioinformatics, 2009. **25**(20): p. 2730-1.

175. Oliveros, J.C., *Venny. An interactive tool for comparing lists with Venn's diagrams*. 2007-2015.

176. Massanella, M., Gianella, S., Lada, S. M., Richman, D. D. and Strain, M. C. , *Quantification of Total and 2-LTR (Long terminal repeat) HIV DNA, HIV RNA and Herpesvirus DNA in PBMCs.* Bio-protocol, 2015. **5**(11): p. e1492.

177. Vargas-Meneses, M.V., M. Massanella, C.C. Ignacio, and S. Gianella, *Quantification of HIV RNA and Human Herpesvirus DNA in Seminal Plasma.* Bio-protocol, 2015. **5**(9): p. e1465.

178. Cherepanov, P., D. Surratt, J. Toelen, W. Pluymers, J. Griffith, E. De Clercq, and Z. Debyser, *Activity of recombinant HIV-1 integrase on mini-HIV DNA.* Nucleic Acids Res, 1999. **27**(10): p. 2202-10.

179. Butler, S.L., M.S. Hansen, and F.D. Bushman, *A quantitative assay for HIV DNA integration in vivo.* Nat Med, 2001. **7**(5): p. 631-4.

180. Garcia, J.V. and A.D. Miller, *Serine phosphorylation-independent downregulation of cell-surface CD4 by nef.* Nature, 1991. **350**(6318): p. 508-11.

181. Willey, R.L., F. Maldarelli, M.A. Martin, and K. Strebel, *Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4.* J Virol, 1992. **66**(12): p. 7193-200.

182. Loven, J., D.A. Orlando, A.A. Sigova, C.Y. Lin, P.B. Rahl, C.B. Burge, D.L. Levens, T.I. Lee, and R.A. Young, *Revisiting global gene expression analysis.* Cell, 2012. **151**(3): p. 476-82.

183. Coate, J.E. and J.J. Doyle, *Variation in transcriptome size: are we getting the message?* Chromosoma, 2015. **124**(1): p. 27-43.

184. Wu, J. and J.B. Lingrel, *Kruppel-like factor 2, a novel immediate-early transcriptional factor, regulates IL-2 expression in T lymphocyte activation.* J Immunol, 2005. **175**(5): p. 3060-6.

185. Haaland, R.E., W. Yu, and A.P. Rice, *Identification of LKLF-regulated genes in quiescent CD4+ T lymphocytes.* Mol Immunol, 2005. **42**(5): p. 627-41.

186. Alimirah, F., R. Panchanathan, J. Chen, X. Zhang, S.M. Ho, and D. Choubey, *Expression of androgen receptor is negatively regulated by p53.* Neoplasia, 2007. **9**(12): p. 1152-9.

187. Takimoto, R. and W.S. El-Deiry, *Wild-type p53 transactivates the KILLER/DR5 gene through an intronic sequence-specific DNA-binding site.* Oncogene, 2000. **19**(14): p. 1735-43.

188. Siliciano, R.F. and W.C. Greene, *HIV latency.* Cold Spring Harb Perspect Med, 2011. **1**(1): p. a007096.

189. Lassen, K.G., J.R. Bailey, and R.F. Siliciano, *Analysis of human immunodeficiency virus type 1 transcriptional elongation in resting CD4+ T cells in vivo.* J Virol, 2004. **78**(17): p. 9105-14.

190. Comer, K.A., P.A. Dennis, L. Armstrong, J.J. Catino, M.B. Kastan, and C.C. Kumar, *Human smooth muscle alpha-actin gene is a transcriptional target of the p53 tumor suppressor protein.* Oncogene, 1998. **16**(10): p. 1299-308.

191. Crighton, D., S. Wilkinson, J. O'Prey, N. Syed, P. Smith, P.R. Harrison, M. Gasco, O. Garrone, T. Crook, and K.M. Ryan, *DRAM, a p53-induced modulator of autophagy, is critical for apoptosis.* Cell, 2006. **126**(1): p. 121-34.

192. Croft, D.R., D. Crighton, M.S. Samuel, F.C. Lourenco, J. Munro, J. Wood, K. Bensaad, K.H. Vousden, O.J. Sansom, K.M. Ryan, and M.F. Olson, *p53-mediated transcriptional regulation and activation of the actin cytoskeleton regulatory RhoC to LIMK2 signaling pathway promotes cell survival.* Cell Res, 2011. **21**(4): p. 666-82.

193. Jin, S., L. Mazzacurati, X. Zhu, T. Tong, Y. Song, S. Shujuan, K.L. Petrik, B. Rajasekaran, M. Wu, and Q. Zhan, *Gadd45a contributes to p53 stabilization in response to DNA damage.* Oncogene, 2003. **22**(52): p. 8536-40.

194. Liu, G. and X. Chen, *The ferredoxin reductase gene is regulated by the p53 family and sensitizes cells to oxidative stress-induced apoptosis.* Oncogene, 2002. **21**(47): p. 7195-204.

195. Nakano, K. and K.H. Vousden, *PUMA, a novel proapoptotic gene, is induced by p53.* Mol Cell, 2001. **7**(3): p. 683-94.

196. Maglott, D., J. Ostell, K.D. Pruitt, and T. Tatusova, *Entrez Gene: gene-centered information at NCBI.* Nucleic Acids Res, 2005. **33**(Database issue): p. D54-8.

197. Yan, J., J. Jiang, C.A. Lim, Q. Wu, H.H. Ng, and K.C. Chin, *BLIMP1 regulates cell growth through repression of p53 transcription.* Proc Natl Acad Sci U S A, 2007. **104**(6): p. 1841-6.

198. Castedo, M., J.L. Perfettini, M. Piacentini, and G. Kroemer, *p53-A pro-apoptotic signal transducer involved in AIDS.* Biochem Biophys Res Commun, 2005. **331**(3): p. 701-6.

199. Genini, D., D. Sheeter, S. Rought, J.J. Zaunders, S.A. Susin, G. Kroemer, D.D. Richman, D.A. Carson, J. Corbeil, and L.M. Leoni, *HIV induces lymphocyte apoptosis by a p53-initiated, mitochondrial-mediated mechanism.* FASEB J, 2001. **15**(1): p. 5-6.

200. Imbeault, M., M. Ouellet, and M.J. Tremblay, *Microarray study reveals that HIV-1 induces rapid type-I interferon-dependent p53 mRNA up-regulation in human primary CD4+ T cells.* Retrovirology, 2009. **6**: p. 5.

201. Perfettini, J.L., T. Roumier, M. Castedo, N. Larochette, P. Boya, B. Raynal, V. Lazar, F. Ciccosanti, R. Nardacci, J. Penninger, M. Piacentini, and G. Kroemer, *NF-kappaB and p53 are the dominant apoptosis-inducing transcription factors elicited by the HIV-1 envelope.* J Exp Med, 2004. **199**(5): p. 629-40.

202. Kaczmarek Michaels, K., M. Natarajan, Z. Euler, G. Alter, G. Viglianti, and A.J. Henderson, *Blimp-1, an intrinsic factor that represses HIV-1 proviral transcription in memory CD4+ T cells.* J Immunol, 2015. **194**(7): p. 3267-74.

203. Hock, A.K., A.M. Vigneron, S. Carter, R.L. Ludwig, and K.H. Vousden, *Regulation of p53 stability and function by the deubiquitinating enzyme USP42.* EMBO J, 2011. **30**(24): p. 4921-30.

204. Kashatus, D., P. Cogswell, and A.S. Baldwin, *Expression of the Bcl-3 proto-oncogene suppresses p53 activation.* Genes Dev, 2006. **20**(2): p. 225-35.

205. Ohtsuka, T., H. Ryu, Y.A. Minamishima, A. Ryo, and S.W. Lee, *Modulation of p53 and p73 levels by cyclin G: implication of a negative feedback regulation.* Oncogene, 2003. **22**(11): p. 1678-87.

206. Yu, H., X. Yue, Y. Zhao, X. Li, L. Wu, C. Zhang, Z. Liu, K. Lin, Z.Y. Xu-Monette, K.H. Young, J. Liu, Z. Shen, Z. Feng, and W. Hu, *LIF negatively regulates tumour-suppressor p53 through Stat3/ID1/MDM2 in colorectal cancers.* Nat Commun, 2014. **5**: p. 5218.

207. Lew, Q.J., Y.L. Chia, K.L. Chu, Y.T. Lam, M. Gurumurthy, S. Xu, K.P. Lam, N. Cheong, and S.H. Chao, *Identification of HEXIM1 as a positive regulator of p53.* J Biol Chem, 2012. **287**(43): p. 36443-54.

208. Zhao, B.X., H.Z. Chen, N.Z. Lei, G.D. Li, W.X. Zhao, Y.Y. Zhan, B. Liu, S.C. Lin, and Q. Wu, *p53 mediates the negative regulation of MDM2 by orphan receptor TR3.* EMBO J, 2006. **25**(24): p. 5703-15.

209. Smith, M.L. and Y.R. Seo, *p53 regulation of DNA excision repair pathways.* Mutagenesis, 2002. **17**(2): p. 149-56.

210. Budworth, H., A.M. Snijders, F. Marchetti, B. Mannion, S. Bhatnagar, E. Kwoh, Y. Tan, S.X. Wang, W.F. Blakely, M. Coleman, L. Peterson, and A.J. Wyrobek, *DNA repair and cell cycle biomarkers of radiation exposure and inflammation stress in human blood.* PLoS One, 2012. **7**(11): p. e48619.

211. Fitch, M.E., I.V. Cross, and J.M. Ford, *p53 responsive nucleotide excision repair gene products p48 and XPC, but not p53, localize to sites of UV-irradiation-induced DNA damage, in vivo.* Carcinogenesis, 2003. **24**(5): p. 843-50.

212. Spivak, G., *Nucleotide excision repair in humans.* DNA Repair (Amst), 2015. **36**: p. 13-8.

213. Carrier, F., P.T. Georgel, P. Pourquier, M. Blake, H.U. Kontny, M.J. Antinore, M. Gariboldi, T.G. Myers, J.N. Weinstein, Y. Pommier, and A.J. Fornace, Jr., *Gadd45, a p53-responsive stress protein, modifies DNA accessibility on damaged chromatin.* Mol Cell Biol, 1999. **19**(3): p. 1673-85.

214. Espeseth, A.S., R. Fishel, D. Hazuda, Q. Huang, M. Xu, K. Yoder, and H. Zhou, *siRNA screening of a targeted library of DNA repair factors in HIV infection reveals a role for base excision repair in HIV integration.* PLoS One, 2011. **6**(3): p. e17612.

215. Skalka, A.M. and R.A. Katz, *Retroviral DNA integration and the DNA damage response.* Cell Death Differ, 2005. **12 Suppl 1**: p. 971-8.

216. Dadachova, E., M.C. Patel, S. Toussi, C. Apostolidis, A. Morgenstern, M.W. Brechbiel, M.K. Gorny, S. Zolla-Pazner, A. Casadevall, and H. Goldstein, *Targeted killing of virally infected cells by radiolabeled antibodies to viral proteins.* PLoS Med, 2006. **3**(11): p. e427.

217. Rawlings, S.A., F. Alonzo, 3rd, L. Kozhaya, V.J. Torres, and D. Unutmaz, *Elimination of HIV-1-Infected Primary T Cell Reservoirs in an In Vitro Model of Latency.* PLoS One, 2015. **10**(5): p. e0126917.

218. Zhang, X.W., X.F. Wang, S.J. Ni, W. Qin, L.Q. Zhao, R.X. Hua, Y.W. Lu, J. Li, G.P. Dimri, and W.J. Guo, *UBTD1 induces cellular senescence through an UBTD1-Mdm2/p53 positive feedback loop.* J Pathol, 2015. **235**(4): p. 656-67.

219. Arakawa, H., *p53, apoptosis and axon-guidance molecules.* Cell Death Differ, 2005. **12**(8): p. 1057-65.

220. Taira, N., T. Yamaguchi, J. Kimura, Z.G. Lu, S. Fukuda, S. Higashiyama, M. Ono, and K. Yoshida, *Induction of amphiregulin by p53 promotes apoptosis via control of microRNA biogenesis in response to DNA damage.* Proc Natl Acad Sci U S A, 2014. **111**(2): p. 717-22.

221. Kim, J.M., Y.D. Yoon, and B.K. Tsang, *Involvement of the Fas/Fas ligand system in p53-mediated granulosa cell apoptosis during follicular development and atresia.* Endocrinology, 1999. **140**(5): p. 2307-17.

222. Xu, J., J. Zhou, M.S. Li, C.F. Ng, Y.K. Ng, P.B. Lai, and S.K. Tsui, *Transcriptional regulation of the tumor suppressor FHL2 by p53 in human kidney and liver cells.* PLoS One, 2014. **9**(8): p. e99359.

223. Mitkin, N.A., C.D. Hook, A.M. Schwartz, S. Biswas, D.V. Kochetkov, A.M. Muratova, M.A. Afanasyeva, J.E. Kravchenko, A. Bhattacharyya, and D.V. Kuprash, *p53-dependent expression of CXCR5 chemokine receptor in MCF-7 breast cancer cells.* Sci Rep, 2015. **5**: p. 9330.

224. Mori, T., Y. Anazawa, M. Iiizumi, S. Fukuda, Y. Nakamura, and H. Arakawa, *Identification of the interferon regulatory factor 5 gene (IRF-5) as a direct target for p53.* Oncogene, 2002. **21**(18): p. 2914-8.

225. Richman, D.D., D.M. Margolis, M. Delaney, W.C. Greene, D. Hazuda, and R.J. Pomerantz, *The challenge of finding a cure for HIV infection.* Science, 2009. **323**(5919): p. 1304-1307.

226. Ylisastigui, L., N.M. Archin, G. Lehrman, R.J. Bosch, and D.M. Margolis, *Coaxing HIV-1 from resting CD4 T cells: histone deacetylase inhibition allows latent viral expression.* AIDS, 2004. **18**(8): p. 1101-8.

227. Mann, B.S., J.R. Johnson, M.H. Cohen, R. Justice, and R. Pazdur, *FDA Approval Summary: Vorinostat for Treatment of Advanced Primary Cutaneous T-Cell Lymphoma.* Oncologist, 2007. **12**(10): p. 1247-1252.

228. Matalon, S., T.A. Rasmussen, and C.A. Dinarello, *Histone deacetylase inhibitors for purging HIV-1 from latent reservoir.* Mol Med., 2011. **17**(5-6): p. 466-472.

229. Beliakova-Bethell, N., J. Zhang, A. Singhania, V. Lee, V. Terry, D.D. Richman, C.A. Spina, and C.H. Woelk, *Suberoylanilide hydroxamic acid induces limited changes in the transcriptome of primary CD4+ T cells.* AIDS, 2013. **27**(1): p. 29-37.

230. Shirakawa, K., L. Chavez, S. Hakre, V. Calvanese, and E. Verdin, *Reactivation of latent HIV by histone deacetylase inhibitors.* Trends Microbiol., 2013. **21**(6): p. 277-285.

231. Choudhary, C., C. Kumar, F. Gnad, M.L. Nielsen, M. Rehman, T.C. Walther, J.V. Olsen, and M. Mann, *Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions.* Science, 2009. **325**(5942): p. 834-840.

232. Bantscheff, M., C. Hopf, M.M. Savitski, A. Dittmann, P. Grandi, A.-M. Michon, J. Schlegl, Y. Abraham, I. Becher, G. Bergamini, M. Boesche, M. Delling, B. Dumpelfeld, D. Eberhard, C. Huthmacher, T. Mathieson, D. Poeckel, V. Reader, K. Strunk, G. Sweetman, U. Kruse, G. Neubauer, N.G. Ramsden, and G. Drewes, *Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes.* Nat Biotech., 2011. **29**(3): p. 255-265.

233. Contreras, X., M. Schweneker, C.-S. Chen, J.M. McCune, S.G. Deeks, J. Martin, and B.M. Peterlin, *Suberoylanilide Hydroxamic Acid Reactivates HIV from Latently Infected Cells.* J Biol Chem., 2009. **284**(11): p. 6782-6789.

234. LaBonte, M., P. Wilson, W. Fazzone, S. Groshen, H.-J. Lenz, and R. Ladner, *DNA microarray profiling of genes differentially regulated by the histone deacetylase inhibitors vorinostat and LBH589 in colon cancer cell lines.* BMC Med Genomics, 2009. **2**: p. 67.

235. Beliakova-Bethell, N., M. Massanella, C. White, S.M. Lada, P. Du, F. Vaida, J. Blanco, C.A. Spina, and C.H. Woelk, *The effect of cell subset isolation method on gene expression in leukocytes.* Cytometry Part A, 2014. **85A**(1): p. 94-104.

236. Papachristou, E.K., T.I. Roumeliotis, A. Chrysagi, C. Trigoni, E. Charvalos, P.A. Townsend, K. Pavlakis, and S.D. Garbis, *The Shotgun Proteomic Study of the Human ThinPrep Cervical Smear Using iTRAQ Mass-Tagging and 2D LC-FT-Orbitrap-MS: The Detection of the Human Papillomavirus at the Protein Level.* J Proteome Res., 2013. **12**(5): p. 2078-2089.

237. Al-Daghri, N.M., O.S. Al-Attas, H.E. Johnston, A. Singhania, M.S. Alokail, K.M. Alkharfy, S.H. Abd-Alrahman, S.I. Sabico, T.I. Roumeliotis, A. Manousopoulou-Garbis, P.A. Townsend, C.H. Woelk, G.P. Chrousos, and S.D. Garbis, *Whole Serum 3D LC-nESI-FTMS Quantitative Proteomics Reveals Sexual Dimorphism in the Milieu Intérieur of Overweight and Obese Adults.* J Proteome Res., 2014. **13**(11): p. 5094-5105.

238. Manousopoulou, A., J. Woo, C.H. Woelk, H.E. Johnston, A. Singhania, C. Hawkes, S.D. Garbis, and R.O. Carare, *Are you also what your mother eats? Distinct proteomic portrait as a result of maternal high-fat diet in the cerebral cortex of the adult mouse.* Int J Obes, 2015.

239. Vizcaíno, J.A., E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, J.A. Dianes, Z. Sun, T. Farrah, N. Bandeira, P.-A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R.J. Chalkley, H.-J. Kraus, J.P. Albar, S. Martinez-Bartolomé, R. Apweiler, G.S. Omenn, L. Martens, A.R. Jones, and H. Hermjakob, *ProteomeXchange provides globally co-ordinated proteomics data submission and dissemination.* Nat Biotechnol., 2014. **32**(3): p. 223-226.

240. Reardon, B., N. Beliakova-Bethell, C.A. Spina, J.X. Zhang, A. Singhania, D.D. Richman, and C.H. Woelk, *Dose-dependent gene expression changes in suberoylanilide hydroxamic acid (SAHA) treated resting CD4 positive T cells.* AIDS, 2015: p. In press.

241. Smyth, K., *Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.* Stat Appl Genet Mol Biol., 2004. **3**(1): p. 1-25.

242. Shannon, P., A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, *Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.* Genome Res., 2003. **13**(11): p. 2498-2504.

243. Warsow, G., B. Greber, S.S.I. Falk, C. Harder, M. Siatkowski, S. Schordan, A. Som, N. Endlich, H. Schöler, D. Repsilber, K. Endlich, and G. Fuellen, *ExprEssence - Revealing the essence of differential experimental data in the*

context of an interaction/regulation network. BMC Syst Biol., 2010. **4**: p. 164-164.

244. Yang, X., K. Regan, Y. Huang, Q. Zhang, J. Li, T.Y. Seiwert, E.E.W. Cohen, H.R. Xing, and Y.A. Lussier, *Single Sample Expression-Anchored Mechanisms Predict Survival in Head and Neck Cancer.* PLoS Comput Biol., 2012. **8**(1): p. e1002350.

245. Kasprzyk, A., *BioMart: driving a paradigm change in biological data management.* Database (Oxford), 2011. **2011**: p. bar049.

246. Flicek, P., M.R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C.G. Girón, L. Gordon, T. Hourlier, S. Hunt, N. Johnson, T. Juettemann, A.K. Kähäri, S. Keenan, E. Kulesha, F.J. Martin, T. Maurel, W.M. McLaren, D.N. Murphy, R. Nag, B. Overduin, M. Pignatelli, B. Pritchard, E. Pritchard, H.S. Riat, M. Ruffier, D. Sheppard, K. Taylor, A. Thormann, S.J. Trevanion, A. Vullo, S.P. Wilder, M. Wilson, A. Zadissa, B.L. Aken, E. Birney, F. Cunningham, J. Harrow, J. Herrero, T.J.P. Hubbard, R. Kinsella, M. Muffato, A. Parker, G. Spudich, A. Yates, D.R. Zerbino, and S.M.J. Searle, *Ensembl 2014.* Nucleic Acids Res., 2014. **42**(Database issue): p. D749-D755.

247. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* J R Stat Soc, B MET, 1995. **57**(1): p. 289-300.

248. Helbig, A.O., A.J.R. Heck, and M. Slijper, *Exploring the membrane proteome—Challenges and analytical strategies.* J Proteomics, 2010. **73**(5): p. 868-878.

249. Pestova, T.V., I.B. Lomakin, J.H. Lee, S.K. Choi, T.E. Dever, and C.U.T. Hellen, *The joining of ribosomal subunits in eukaryotes requires eIF5B.* Nature, 2000. **403**(6767): p. 332-335.

250. Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-342.

251. Mohammadi, P., J. di Iulio, M. Muñoz, R. Martinez, I. Bartha, M. Cavassini, C. Thorball, J. Fellay, N. Beerenwinkel, A. Ciuffi, and A. Telenti, *Dynamics of HIV latency and reactivation in a primary CD4+ T cell model.* PLoS Pathog., 2014. **10**(5): p. e1004156.

252. Zhang, Q. and Y. Wang, *HMG Modifications and Nuclear Function.* Biochim Biophys Acta., 2010. **1799**(1-2): p. 28.

253. Eilebrecht, S., E. Wilhelm, B.-J. Benecke, B. Bell, and A.G. Benecke, *HMGA1 directly interacts with TAR to modulate basal and Tat-dependent HIV transcription.* RNA Biol., 2013. **10**(3): p. 436-444.

254. Eilebrecht, S., V. Le Douce, R. Riclet, B. Targat, H. Hallay, B. Van Driessche, C. Schwartz, G. Robette, C. Van Lint, O. Rohr, and A.G. Benecke, *HMGA1 recruits CTIP2-repressed P-TEFb to the HIV-1 and cellular target promoters.* Nucleic Acids Res., 2014. **42**(8): p. 4962-4971.

255. Zhang, Q., K. Zhang, Y. Zou, A. Perna, and Y. Wang, *A Quantitative Study on the in-vitro and in-vivo Acetylation of High Mobility Group A1 Proteins.* J Am Soc Mass Spectrom., 2007. **18**(9): p. 1569-1578.

256. Munshi, N., T. Agalioti, S. Lomardas, M. Merika, G. Chen, and D. Thanos, *Coordination of a transcriptional switch by HMGI(Y) actylation.* Science, 2001. **293**: p. 1133-1136.

257. Nabel, G. and D. Baltimore, *An inducible transcription factor activates expression of human immunodeficiency virus in T cells.* Nature, 1987. **326**(6114): p. 711-713.

258. Osborn, L., S. Kunkel, and G.J. Nabel, *Tumor necrosis factor alpha and interleukin 1 stimulate the human immunodeficiency virus enhancer by activation of the nuclear factor kappa B.* Proc Natl Acad Sci U S A., 1989. **86**(7): p. 2336-2340.

259. Sales, K.U., S. Friis, J.E. Konkel, S. Godiksen, M. Hatakeyama, K.K. Hansen, S.R. Rogatto, R. Szabo, L.K. Vogel, W. Chen, J.S. Gutkind, and T.H. Bugge, *Non-hematopoietic PAR-2 is essential for matriptase-driven pre-malignant progression and potentiation of ras-mediated squamous cell carcinogenesis.* Oncogene, 2015. **34**(3): p. 346-356.

260. Parcellier, A., E. Schmitt, S. Gurbuxani, D. Seigneurin-Berny, A. Pance, A. Chantôme, S. Plenchette, S. Khochbin, E. Solary, and C. Garrido, *HSP27 Is a Ubiquitin-Binding Protein Involved in I-κBα Proteasomal Degradation.* Mol Cell Biol., 2003. **23**(16): p. 5790-5802.

261. Chang, E.-J., J. Ha, S.-S. Kang, Z.H. Lee, and H.-H. Kim, *AWP1 binds to tumor necrosis factor receptor-associated factor 2 (TRAF2) and is involved in TRAF2-mediated nuclear factor-kappaB signaling.* Int J Biochem Cell Biol., 2011. **43**(11): p. 1612-1620.

262. Mercurio, F., H. Zhu, B.W. Murray, A. Shevchenko, B.L. Bennett, J.w. Li, D.B. Young, M. Barbosa, M. Mann, A. Manning, and A. Rao, *IKK-1 and IKK-2: Cytokine-Activated IκB Kinases Essential for NF-κB Activation.* Science, 1997. **278**(5339): p. 860-866.

263. O'Keeffe, B., Y. Fong, D. Chen, S. Zhou, and Q. Zhou, *Requirement for a Kinase-specific Chaperone Pathway in the Production of a Cdk9/Cyclin T1 Heterodimer Responsible for P-TEFb-mediated Tat Stimulation of HIV-1 Transcription.* J Biol Chem., 2000. **275**(1): p. 279-287.

264. Campbell, G.R., R.S. Bruckman, Y.-L. Chu, and S.A. Spector, *Autophagy Induction by Histone Deacetylase Inhibitors Inhibits HIV Type 1.* J Biol Chem., 2015. **290**(8): p. 5028-5040.

265. Stankov, M., C. Suhr, H. Lin, D. Panayotova-Dimitrova, C. Goffinet, and G. Behrens. *Latent HIV-1 reactivation and lysosomal destabilization synergize to host cell death.* in *Program and abstracts of the 22nd Conference on Retroviruses and Opportunistic Infections*. 2015. Seattle, WA.

266. Wei, D.G., V. Chiang, E. Fyne, M. Balakrishnan, T. Barnes, M. Graupe, J. Hesselgesser, A. Irrinki, J.P. Murry, G. Stepan, K.M. Stray, A. Tsai, H. Yu, J. Spindler, M. Kearney, C.A. Spina, D. McMahon, J. Lalezari, D. Sloan, J. Mellors, R. Geleziunas, and T. Cihlar, *Histone Deacetylase Inhibitor Romidepsin Induces HIV Expression in CD4 T Cells from Patients on Suppressive Antiretroviral Therapy at Concentrations Achieved by Clinical Dosing.* PLoS Pathog., 2014. **10**(4): p. e1004071.

267. Rasmussen, T.A., O.S. Søgaard, C. Brinkmann, F. Wightman, S.R. Lewin, J. Melchjorsen, C. Dinarello, L. Østergaard, and M. Tolstrup, *Comparison of HDAC inhibitors in clinical development: Effect on HIV production in latently infected cells and T-cell activation.* Hum Vaccin Immunother., 2013. **9**(5): p. 993-1001.

268. Williams, S., A. Samuel, L.-F. Chen, H. Kwon, D. Fenard, D. Bisgrove, E. Verdin, and W.C. Greene, *Prostratin antagonizes HIV latency by activating NF-κB.* J Biol Chem., 2004. **279**(40): p. 42008-42017.

269. Jiang, G., E.A. Mendes, P. Kaiser, S. Sankaran-Walters, Y. Tang, M.G. Weber, G.P. Melcher, G.R.r. Thompson, A. Tanuri, L.F. Pianowski, J.K. Wong, and S. Dandekar, *Reactivation of HIV latency by a newly modified Ingenol derivative via protein kinase Cδ-NF-κB signaling.* AIDS, 2014. **28**(11): p. 1555-1566.

270. José, D.P., K. Bartholomeeusen, R.D. da Cunha, C.M. Abreu, J. Glinski, T.B.F. da Costa, A.F.M.B. Rabay, L.F.P. Filho, L.W. Dudycz, U. Ranga, B.M. Peterlin, L.F. Pianowski, A. Tanuri, and R.S. Aguiar, *Reactivation of latent HIV-1 by new semi-synthetic ingenol esters.* Virology, 2014. **462-463**: p. 328-339.

271. Sakane, N., H.S. Kwon, S. Pagans, K. Kaehlcke, Y. Mizusawa, M. Kamada, K.G. Lassen, J. Chan, W.C. Greene, M. Schnoelzer, and M. Ott, *Activation of HIV transcription by the viral Tat protein requires a demethylation step mediated by lysine-specific demethylase 1 (LSD1/KDM1).* PLoS Pathog, 2011. **7**(8): p. e1002184.

272. Bassuk, A.G., R.T. Anandappa, and J.M. Leiden, *Physical interactions between Ets and NF-kappaB/NFAT proteins play an important role in their cooperative activation of the human immunodeficiency virus enhancer in T cells.* J Virol, 1997. **71**(5): p. 3563-73.

273. Posada, R., M. Pettoello-Mantovani, M. Sieweke, T. Graf, and H. Goldstein, *Suppression of HIV type 1 replication by a dominant-negative Ets-1 mutant.* AIDS Res Hum Retroviruses, 2000. **16**(18): p. 1981-9.

274. Sieweke, M.H., H. Tekotte, U. Jarosch, and T. Graf, *Cooperative interaction of ets-1 with USF-1 required for HIV-1 enhancer activity in T cells.* EMBO J, 1998. **17**(6): p. 1728-39.

275. Yang, H.C., L. Shen, R.F. Siliciano, and J.L. Pomerantz, *Isolation of a cellular factor that can reactivate latent HIV-1 without T cell activation.* Proc Natl Acad Sci U S A, 2009. **106**(15): p. 6321-6.

276. Sheridan, P.L., C.T. Sheline, K. Cannon, M.L. Voz, M.J. Pazin, J.T. Kadonaga, and K.A. Jones, *Activation of the HIV-1 enhancer by the LEF-1 HMG protein on nucleosome-assembled DNA in vitro.* Genes Dev, 1995. **9**(17): p. 2090-104.

277. Mercurio, F., H. Zhu, B.W. Murray, A. Shevchenko, B.L. Bennett, J. Li, D.B. Young, M. Barbosa, M. Mann, A. Manning, and A. Rao, *IKK-1 and IKK-2:*

cytokine-activated IkappaB kinases essential for NF-kappaB activation. Science, 1997. **278**(5339): p. 860-6.

278. Gallastegui, E., G. Millan-Zambrano, J.M. Terme, S. Chavez, and A. Jordan, *Chromatin reassembly factors are involved in transcriptional interference promoting HIV latency.* J Virol, 2011. **85**(7): p. 3187-202.

279. Bisgrove, D.A., T. Mahmoudi, P. Henklein, and E. Verdin, *Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription.* Proc Natl Acad Sci U S A, 2007. **104**(34): p. 13690-5.

280. Tetsuka, T., H. Uranishi, H. Imai, T. Ono, S. Sonta, N. Takahashi, K. Asamitsu, and T. Okamoto, *Inhibition of nuclear factor-kappaB-mediated transcription by association with the amino-terminal enhancer of split, a Groucho-related protein lacking WD40 repeats.* J Biol Chem, 2000. **275**(6): p. 4383-90.

281. Mahmoudi, T., *The BAF complex and HIV latency.* Transcription, 2012. **3**(4): p. 171-6.

282. Rafati, H., M. Parra, S. Hakre, Y. Moshkin, E. Verdin, and T. Mahmoudi, *Repressive LTR nucleosome positioning by the BAF complex is required for HIV latency.* PLoS Biol, 2011. **9**(11): p. e1001206.

283. Nguyen, H.D., I.A. Wood, and M.M. Hill, *A robust permutation test for quantitative SILAC proteomics experiments.* JIOMICS, 2012. **2**(2): p. 80-93.

284. Ernst, M.D., *Permutation methods: a basis for exact inference.* Statist. Sci., 2004. **19**(4): p. 676-685.

285. Warnes, G.R., B. Bolker, L. Bonebakker, R. Gentleman, W. Huber, A. Liaw, T. Lumley, M. Maechler, A. Magnusson, and S. Moeller, *gplots: Various R programming tools for plotting data.* R package version, 2009. **2**(4).

286. Deeks, S.G., *HIV: Shock and kill.* Nature, 2012. **487**(7408): p. 439-40.

287. Rasmussen, T.A., M. Tolstrup, A. Winckelmann, L. Ostergaard, and O.S. Sogaard, *Eliminating the latent HIV reservoir by reactivation strategies: advancing to clinical trials.* Hum Vaccin Immunother, 2013. **9**(4): p. 790-9.

288. Boeke, J.D. and J.P. Stoye, *Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements.* 1997.

289. Bannert, N. and R. Kurth, *Retroelements and the human genome: new perspectives on an old relation.* Proc Natl Acad Sci U S A, 2004. **101 Suppl 2**: p. 14572-9.

290. Turner, G., M. Barbulescu, M. Su, M.I. Jensen-Seaman, K.K. Kidd, and J. Lenz, *Insertional polymorphisms of full-length endogenous retroviruses in humans.* Curr Biol, 2001. **11**(19): p. 1531-5.

291. Belshaw, R., A.L. Dawson, J. Woolven-Allen, J. Redding, A. Burt, and M. Tristem, *Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity.* J Virol, 2005. **79**(19): p. 12507-14.

292. Hughes, J.F. and J.M. Coffin, *Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution.* Nat Genet, 2001. **29**(4): p. 487-9.

293. Bieda, K., A. Hoffmann, and K. Boller, *Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines.* J Gen Virol, 2001. **82**(Pt 3): p. 591-6.

294. Boller, K., O. Janssen, H. Schuldes, R.R. Tonjes, and R. Kurth, *Characterization of the antibody response specific for the human endogenous retrovirus HTDV/HERV-K.* J Virol, 1997. **71**(6): p. 4581-8.

295. Tönjes, R.R., R. Löwer, K. Boller, J. Denner, B. Hasenmaier, H. Kirsch, H. König, C. Korbmacher, C. Limbach, R. Lugert, R.C. Phelps, J. Scherer, K. Thelen, J. Löwer, and R. Kurth, *HERV-K: The Biologically Most Active Human Endogenous Retrovirus Family.* JAIDS Journal of Acquired Immune Deficiency Syndromes, 1996. **13**: p. S261-S267.

296. Dolei, A. and H. Perron, *The multiple sclerosis-associated retrovirus and its HERV-W endogenous family: a biological interface between virology, genetics, and immunology in human physiology and disease.* J Neurovirol, 2009. **15**(1): p. 4-13.

297. Leboyer, M., R. Tamouza, D. Charron, R. Faucard, and H. Perron, *Human endogenous retrovirus type W (HERV-W) in schizophrenia: a new avenue of*

research at the gene-environment interface. World J Biol Psychiatry, 2013. **14**(2): p. 80-90.

298. Liang, Q., Z. Xu, R. Xu, L. Wu, and S. Zheng, *Expression patterns of non-coding spliced transcripts from human endogenous retrovirus HERV-H elements in colon cancer.* PLoS One, 2012. **7**(1): p. e29950.

299. Campbell, I.M., T. Gambin, P. Dittwald, C.R. Beck, A. Shuvarikov, P. Hixson, A. Patel, A. Gambin, C.A. Shaw, J.A. Rosenfeld, and P. Stankiewicz, *Human endogenous retroviral elements promote genome instability via non-allelic homologous recombination.* BMC Biol, 2014. **12**: p. 74.

300. Bhardwaj, N., F. Maldarelli, J. Mellors, and J.M. Coffin, *HIV-1 infection leads to increased transcription of human endogenous retrovirus HERV-K (HML-2) proviruses in vivo but not to increased virion production.* J Virol, 2014. **88**(19): p. 11108-20.

301. Gonzalez-Hernandez, M.J., M.D. Swanson, R. Contreras-Galindo, S. Cookinham, S.R. King, R.J. Noel, Jr., M.H. Kaplan, and D.M. Markovitz, *Expression of human endogenous retrovirus type K (HML-2) is activated by the Tat protein of HIV-1.* J Virol, 2012. **86**(15): p. 7790-805.

302. Hurst, T., M. Pace, A. Katzourakis, R. Phillips, P. Klenerman, J. Frater, and G. Magiorkinis, *Human endogenous retrovirus (HERV) expression is not induced by treatment with the histone deacetylase (HDAC) inhibitors in cellular models of HIV-1 latency.* Retrovirology, 2016. **13**(1): p. 10.

303. Krönung, S.K., *Role of endogenous retrovirus promoter activity in tumor suppression*. 2015, Göttingen Graduate School for Neurosciences, Biophysics and Molecular Biosciences: Göttingen.

304. Paces, J., A. Pavlicek, and V. Paces, *HERVd: database of human endogenous retroviruses.* Nucleic Acids Res, 2002. **30**(1): p. 205-6.

305. Vargiu, L., P. Rodriguez-Tome, G.O. Sperber, M. Cadeddu, N. Grandi, V. Blikstad, E. Tramontano, and J. Blomberg, *Classification and characterization of human endogenous retroviruses; mosaic forms are common.* Retrovirology, 2016. **13**(1): p. 7.

306. Paces, J., A. Pavlicek, R. Zika, V.V. Kapitonov, J. Jurka, and V. Paces, *HERVd: the Human Endogenous RetroViruses Database: update.* Nucleic Acids Res, 2004. **32**(Database issue): p. D50.

307. Smit, A.F.A., R. Hubley, and P. Green. *RepeatMasker Open-3.0*. 1996-2010; Available from: http://www.repeatmasker.org.

308. Carver, T., U. Bohme, T.D. Otto, J. Parkhill, and M. Berriman, *BamView: viewing mapped read alignment data in the context of the reference sequence.* Bioinformatics, 2010. **26**(5): p. 676-7.

309. Kent, W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler, *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

310. Althammer, S., J. Gonzalez-Vallinas, C. Ballare, M. Beato, and E. Eyras, *Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data.* Bioinformatics, 2011. **27**(24): p. 3333-40.

311. Kent, W.J., *BLAT--the BLAST-like alignment tool.* Genome Res, 2002. **12**(4): p. 656-64.

312. Karolchik, D., A.S. Hinrichs, T.S. Furey, K.M. Roskin, C.W. Sugnet, D. Haussler, and W.J. Kent, *The UCSC Table Browser data retrieval tool.* Nucleic Acids Res, 2004. **32**(Database issue): p. D493-6.

313. Hindson, B.J., K.D. Ness, D.A. Masquelier, P. Belgrader, N.J. Heredia, A.J. Makarewicz, I.J. Bright, M.Y. Lucero, A.L. Hiddessen, T.C. Legler, T.K. Kitano, M.R. Hodel, J.F. Petersen, P.W. Wyatt, E.R. Steenblock, P.H. Shah, L.J. Bousse, C.B. Troup, J.C. Mellen, D.K. Wittmann, N.G. Erndt, T.H. Cauley, R.T. Koehler, A.P. So, S. Dube, K.A. Rose, L. Montesclaros, S. Wang, D.P. Stumbo, S.P. Hodges, S. Romine, F.P. Milanovich, H.E. White, J.F. Regan, G.A. Karlin-Neumann, C.M. Hindson, S. Saxonov, and B.W. Colston, *High-throughput droplet digital PCR system for absolute quantitation of DNA copy number.* Anal Chem, 2011. **83**(22): p. 8604-10.

314. Beliakova-Bethell, N., J.X. Zhang, A. Singhania, V. Lee, V.H. Terry, D.D. Richman, C.A. Spina, and C.H. Woelk, *Suberoylanilide hydroxamic acid induces limited changes in the transcriptome of primary CD4(+) T cells.* AIDS, 2013. **27**(1): p. 29-37.

315. Owczarzy, R., A.V. Tataurov, Y. Wu, J.A. Manthey, K.A. McQuisten, H.G. Almabrazi, K.F. Pedersen, Y. Lin, J. Garretson, N.O. McEntaggart, C.A. Sailor, R.B. Dawson, and A.S. Peek, *IDT SciTools: a suite for analysis and design of nucleic acid oligomers.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W163-9.

316. Blomberg, J., D. Ushameckis, and P. Jern, *Evolutionary Aspects of Human Endogenous Retroviral Sequences (HERVs) and Disease*. Madame Curie Bioscience Database [Internet]. 2000-2013, Austin (TX): Landes Bioscience.

317. Tebit, D.M. and E.J. Arts, *Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease.* Lancet Infect Dis, 2011. **11**(1): p. 45-56.

318. van der Kuyl, A.C., *HIV infection and HERV expression: a review.* Retrovirology, 2012. **9**: p. 6.

319. Zain, J. and O.A. O'Connor, *Targeting histone deacetyalses in the treatment of B- and T-cell malignancies.* Invest New Drugs, 2010. **28 Suppl 1**: p. S58-78.

320. Havecker, E.R., X. Gao, and D.F. Voytas, *The diversity of LTR retrotransposons.* Genome Biol, 2004. **5**(6): p. 225.

321. Svensson, A.C., T. Raudsepp, C. Larsson, A. Di Cristofano, B. Chowdhary, G. La Mantia, L. Rask, and G. Andersson, *Chromosomal distribution, localization and expression of the human endogenous retrovirus ERV9.* Cytogenet Cell Genet, 2001. **92**(1-2): p. 89-96.

322. Beyer, U., J. Moll-Rocek, U.M. Moll, and M. Dobbelstein, *Endogenous retrovirus drives hitherto unknown proapoptotic p63 isoforms in the male germ line of humans and great apes.* Proc Natl Acad Sci U S A, 2011. **108**(9): p. 3624-9.

323. Yang, A., M. Kaghad, Y. Wang, E. Gillett, M.D. Fleming, V. Dotsch, N.C. Andrews, D. Caput, and F. McKeon, *p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities.* Mol Cell, 1998. **2**(3): p. 305-16.

324. Suh, E.K., A. Yang, A. Kettenbach, C. Bamberger, A.H. Michaelis, Z. Zhu, J.A. Elvin, R.T. Bronson, C.P. Crum, and F. McKeon, *p63 protects the female germ line during meiotic arrest.* Nature, 2006. **444**(7119): p. 624-8.

325. Christensen, T., *HERVs in neuropathogenesis.* J Neuroimmune Pharmacol, 2010. **5**(3): p. 326-35.