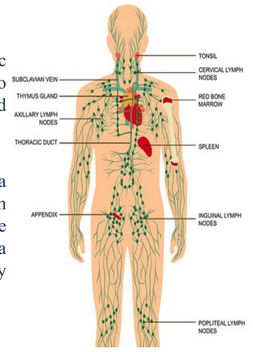
Evaluating Tissue-Specific and Temporal Gene Expression Changes after Simian Immunodeficiency Virus (SIV) Infection in Rhesus Monkeys

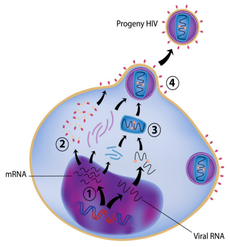
Christina Michalski, Heather A. MacKay, Jeffrey Tang, Nichalle Brito, Rebecca Asiimwe

INTRODUCTION

Human immunodeficiency virus (HIV) has continued to be a public health issue on a global scale and 35 million people have died due to the onset of acquired immunodeficiency syndrome (AIDS)-related diseases (1).



Early diagnosis is crucial as it enables the control of infection via antiretroviral therapy in addition to reducing the risk of transmission (2). It is estimated that half of the patients with HIV worldwide are "late-presenters" due to the asymptomatic nature of the virus (3); a rationale for which more knowledge is needed to understand the early stages of HIV infection, and immune response.



Simian immunodeficiency virus (SIV), a retrovirus infecting nonhuman primates, has similar symptoms and viral life cycle to that of HIV (1,4). Phylogenetic analysis shows that SIV and HIV are ancestrally related and are closely related in viral replication and propagation, making it a good model to understand early development of infection. This project aimed at evaluating tissue-specific transcriptomic changes in the first ten days after vaginal SIV infection in rhesus monkeys.

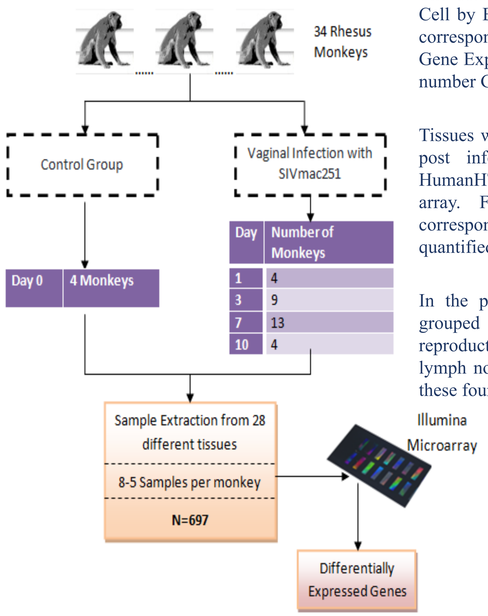
Hypothesis

Temporal tissue-specific transcriptomic changes occur after SIV infection in Rhesus Monkeys.

Objectives

1. Analyze how gene expression changes in various tissues after SIV infection
2. Assess whether transcriptomic changes in blood are the same as in other tissues.
3. Investigate whether transcriptomic changes in lymph nodes vary by lymph node location.

Study Design



The data used in this study was published in Cell by Barouch *et al.* in April 2016 (5). The corresponding dataset is publicly available on Gene Expression Omnibus under the accession number GSE80013.

Tissues were collected at different time points post infection and analyzed by Illumina HumanHT-12 V4.0 expression BeadChip array. For each sample, 47,232 probes corresponding to over 30,000 genes were quantified.

In the publication, the tissue samples were grouped into four categories (blood, female reproductive tract, gastrointestinal tract and lymph nodes) and analysis was performed for these four categories.

Methods

Preliminary Data Inspection

- Imported the Data
- Cleaned the data/subset relevant tissues for analysis
- Identified outliers via. Tissue and sample specific correlation analysis
- Normalized the data via quantile normalization
- Packages used for this analysis: GEOquery, Biobase, NMF, preprocessCore, and ggplot2

How does gene expression change in tissues post-SIV infection?

- Fitted a variety of tissue-specific linear models with different parameters
- Identified the best models for the data by AIC comparison between models with different parameters (e.g. with and without a quadratic of time)
- Conducted gene enrichment analysis to identify the genes and biochemical pathways that are significantly altered with SIV infection via. Entrez annotation, DAVID gene enrichment and KEGG pathway analysis
- Packages used for this analysis: limma, dplyr, lattice, gridExtra, ggplot2, and illuminaHumanv4.db

Are there correlations between changes in blood transcriptomic profiles to tissues?

- Used the tissue-specific linear models fit for question 1 and compared the number of overlapping differentially expressed genes between tissues
- Compared the logFC of overlapping differentially expressed genes between tissues
- Packages used for this analysis: ggplot2, limma, VennDiagram , gridExtra, and dplyr

How does gene expression within the lymph node tissues differ?

- Fitted contrast matrices and linear models between lymph tissues
- Conducted hierarchical clustering and PCA analysis
- Packages used for this analysis: limma, dplyr, ggplot2, stringr, stringi, NMF, and PCA3d

Results

Data cleaning

Prior to analysis, raw data was cleaned and normalized. Tissues with at least 3 biological replicates for each time point were selected, reducing sample size to 231. Presence of outliers was assessed by examining inter-sample correlation within each tissue. Outliers were found in the colon and tonsils samples, leaving 228 samples for final analysis. PreprocessCore was used to log2 normalize the cleaned dataset.

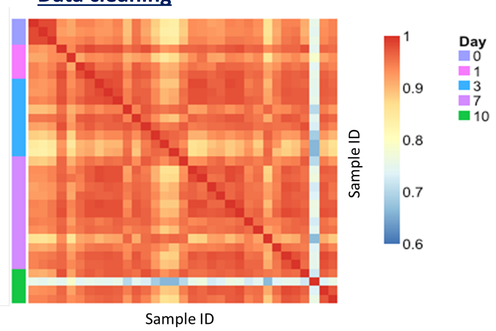


Figure 1: Inter-sample correlation before cleanup of colon samples.

Linear Model Fitting and Validation

Two linear models for each of the seven tissues were fit to identify differentially expressed genes with respect to time. We compared a linear model with and without a quadratic of time to capture non-linear expression trends. To determine the model that best fits the data we compared the AIC values. For most probes, there is no advantage of one model over the other as indicated by a small difference in AIC values. However, there are several probes for which the quadratic model is significantly better (Fig. 2). We therefore chose the quadratic model for all downstream analysis.

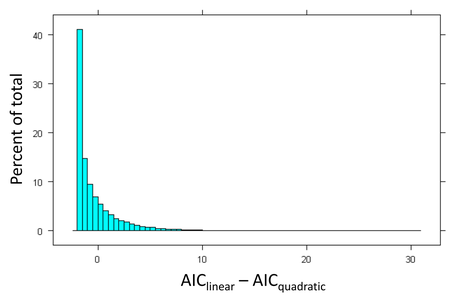


Figure 2: Histogram of the difference in AIC values between the linear and the quadratic model for the jejunum.

Table 1: The number of differentially expressed genes for each tissue using the quadratic models.

Tissue	Number of hits at FDR <= 0.05
Jejunum	5463
Blood	4103
Tonsil	316
Auxiliary Lymph Node	2561
Mesenteric Lymph Node	2063
Genital Pelvic Lymph Node	1886
Colon	595

Gene analysis

Annotation Cluster 1	Enrichment Score: 5.21			Count	P_Value	Benjamini
GOTERM_BP_DIRECT	SRP-dependent cotranslational protein targeting to membrane	RT		24	4.9E-8	1.8E-4
GOTERM_BP_DIRECT	translational initiation	RT		29	1.2E-7	2.2E-4
UP_KEYWORDS	Ribosomal protein	RT		33	2.4E-7	3.7E-5
GOTERM_BP_DIRECT	viral transcription	RT		25	3.7E-7	4.6E-4
KEGG_PATHWAY	Ribosome	RT		28	7.9E-7	2.1E-4
GOTERM_CC_DIRECT	ribosome	RT		30	1.4E-6	8.7E-4
GOTERM_MF_DIRECT	structural constituent of ribosome	RT		36	2.2E-6	2.6E-3
GOTERM_BP_DIRECT	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	RT		24	4.2E-6	3.9E-3
UP_KEYWORDS	Ribonucleoprotein	RT		40	1.4E-5	1.3E-3
GOTERM_BP_DIRECT	translation	RT		37	1.8E-5	1.4E-2
GOTERM_BP_DIRECT	rRNA processing	RT		31	1.2E-4	7.1E-2
GOTERM_CC_DIRECT	cytosolic large ribosomal subunit	RT		14	4.5E-4	9.1E-2
GOTERM_CC_DIRECT	cytosolic small ribosomal subunit	RT		11	1.0E-3	1.2E-1
GOTERM_CC_DIRECT	small ribosomal subunit	RT		8	1.5E-3	1.5E-1

Functional annotation clustering based on enrichment score using DAVID revealed that these differentially expressed genes were most enriched for ribosomal related functions.

This was not far from expected due to the nature of the retrovirus as it requires host cell machinery for viral replication.

Figure 3: The top annotation cluster of differentially expressed probes in the Mesenteric Lymph Node [DAVID and KEGG pathway analysis].

Comparison of transcriptomic changes between tissues

To assess whether transcriptomic changes are the same in blood and other tissues, we first calculated the overlap between DE probes in each tissue and blood (Fig. 4A). For the "common DE probes", we fitted four linear models for each tissue comparing each day to day 0. The resulting fold changes for each probe were compared between tissues (Fig. 4B).

C)

	Intercept	Slope	Adj. R^2
Day 1 vs. 0	0.009	1.10	0.57
Day 3 vs. 0	0.0151	0.92	0.55
Day 7 vs. 0	0.0039	1.11	0.63
Day 10 vs. 0	-0.0085	1.40	0.68

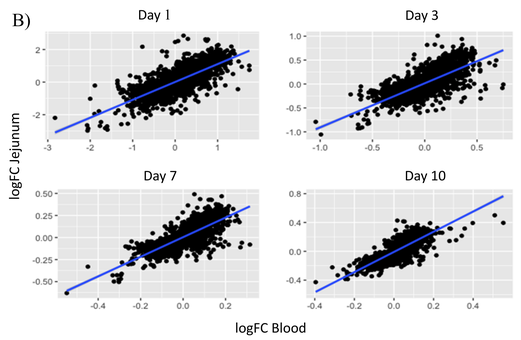


Figure 4: Comparison of DE probes in blood and jejunum. A) Overlap of DE probes. B) Correlation between fold changes of common DE probes. C) Regression statistics from the correlation between fold changes shown in B.

PCA

Unsupervised clustering was performed on the lymph node samples to assess whether they would cluster by location. Samples taken from day 0 (control group) clustered separately from the other time points whereas lymph node location did not influence cluster formation. PCA yielded similar results, indicating no effect of distance from infection site on transcriptomic changes.

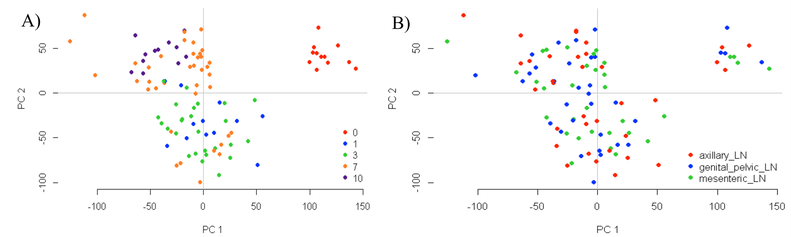


Figure 5: PCA of the lymph node samples. Annotated by time point (A) or lymph tissue type (B).

Conclusion

In line with the findings from Barouch *et al.* we find that significant transcriptomic changes occur as early as day 1 post infection throughout host tissues. This knowledge is crucial for development of treatment as well as vaccines and more effective post-exposure prophylaxis drugs.

References

1. Stevens, D.R. *et al.* (2017). A Global Review of HIV Self-testing: Themes and Implications. AIDS Behavior. doi:10.1007/s10461-017-1707-8.
2. Mocroft, A. *et al.* (2013). Risk factors and outcomes for late presentation for HIV-positive persons in Europe: results from the Collaboration of Observational HIV Epidemiological Research Europe Study (COHERE). PLoS Med 10: e1001510.
3. Levy, I. *et al* (2016). Missed opportunities for earlier diagnosis of HIV in patients who presented with advanced HIV disease: a retrospective cohort study BMJ 6: e012721.
4. Klatt, N.R., *et al.* (2012).Nonpathogenic Simian Immunodeficiency Virus Infections. Cold Spring Harbor Perspectives in Medicine 2: a007153.
5. Barouch, D.H. *et al.* (2016). Rapid Inflammasome Activation Following Mucosal SIV infection of Rhesus Monkeys. Cell 165: 656-667.